VOCAL FOLD PHYSIOLOGY

Biomechanics, Acoustics and Phonatory Control

Edited by
Ingo R. Titze and Ronald C. Scherer

The Denver Center For The Performing Arts

VOCAL FOLD PHYSIOLOGY:

Biomechanics, Acoustics and Phonatory Control

VOCAL FOLD PHYSIOLOGY:

Biomechanics, Acoustics and Phonatory Control

Edited by Ingo R. Titze and Ronald C. Scherer

The Denver Center for the Performing Arts
Denver, Colorado

Proceedings of the International Conference on
Physiology and Biophysics of the Voice,
University of Iowa, Iowa City, Iowa,
May 4-7, 1983

Published by
The Denver Center for the Performing Arts, Inc.
1245 Champa Street
Denver, Colorado 80204

Printed in the United States of America

CONTENTS

PART I. ANATOMY AND NEUROPHYSIOLOGY OF THE LARYNX

Section Editors: L. Malmgren and C. Larson

PART II. VOCAL FOLD KINEMATICS AND TECHNIQUES FOR CLINICAL AND PEDAGOGICAL OBSERVATION

Section Editors: K. Munhall and D. Childers

PART III. BIOMECHANICS, AERODYNAMICS, AND ACOUSTIC INTERACTIONS OF THE LARYNX

Section Editors: P. Milenkovic, R. Scherer, and I. Titze

CONTRIBUTORS

Fariborz Alipour-Haghighi, Ph.D.
Department of Speech Pathology and Audiology
University of Iowa
Iowa City, Iowa 52242

Tirupattur Ananthapadmanabha, Ph.D.
Department of Speech Communication and Music Acoustics
Royal Institute of Technology
KTH S-100 44 Stockholm
SWEDEN

Thomas Baer, Ph.D.
Haskins Laboratories
270 Crown Street
New Haven, Connecticut 06510

James Beauchamp. Ph.D.
2136 Music Building
1114 W. Nevada
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

Arthur H. Benade, Ph.D.
Physics Department
Case Western Reserve University
Cleveland, Ohio 44106

Stanley Blaugrund, M.D.
Director, Department of Otolaryngology
Lenox Hill Hospital
47 East 77th Street
New York, New York 10021

Diane Bless, Ph.D.
University of Wisconsin
1975 Willow Drive
Department of Communicative Disorders
Madison, Wisconsin 53706

Louis Boves, Ph.D.
Institute of Phonetics
Nijmegen University
Erasmusplein 1
6525 HT Nijmegen
THE NETHERLANDS

Malcolm Brenner, Ph.D.
Speech Research Laboratory
VA Medical Center
4150 Clement Street
San Francisco, California 94121

Donald G. Childers, Ph.D.
College of Engineering
Department of Electrical Engineering
University of Florida
Gainesville, Florida 32611

Rene Collier, Ph.D.
University of Antwerp
UFSIA
Prinsstraat 13
Antwerp, Belgium

William A. Conrad, M.S.
Biophysics Consultant
30 West 71st Street
New York, New York 10023

Donald S. Cooper, Ph.D.
Research Division
Department of Otolaryngology
903 Hoffman
2025 Zonal Avenue
Los Angeles, California 90033

Bert Cranen, Ph.D.
Institute of Phonetics
Nijmegen University
Erasmusplein 1
6525 HT Nijmegen
THE NETHERLANDS

Gunnar Fant, D.Sc.
Head, Department of Speech Communication and Music Acoustics
Royal Institute of Technology
S-100 44 Stockholm
SWEDEN

Robert Feder, M.D.
Assistant Clinical Professor
Head and Neck Surgery at U.C.L.A.
Adjunct Professor of Vocal Arts at U.S.C.
8635 West 3rd Street, Suite 465W
Los Angeles, California 90048

Dom V. Finocchio, Ph.D.
Department of Physiology and Biophysics
University of Washington
Seattle, Washington 98195

Osamu Fujimura, Ph.D.
Head, Department of Linguistics and Artificial Intelligence Research
AT&T Bell Laboratories
600 Mountain Avenue, #2D-545
Murray Hill, New Jersey 07974

Jan Gauffin, D.Sc.
Department of Speech Communication and Music Acoustics
Royal Institute of Technology
KTH S-100 44 Stockholm
SWEDEN

Carole E. Gelfer, M.A.
Haskins Laboratories
270 Crown Street
New Haven, Connecticut 06510

Wilbur J. Gould, M.D.
Department of Otolaryngology
Lenox Hill Hospital
47 East 77th Street
New York, New York 10021

Bernard Guérin, Ph.D.
Institut National Polytechnique de Grenoble
Institut de la Communication Parlée
U.A. CNRS 368
23, Rue des Martyrs
38031 Grenoble-Cedex
FRANCE

Katherine S. Harris, Ph.D.
Haskins Laboratories
270 Crown Street
New Haven, Connecticut 06510

Malcolm H. Hast, Ph.D.
Director of Research
Department of Otolaryngology - Head and Neck Surgery
Northwestern University
Searle Building 12-535
303 East Chicago Avenue
Chicago, Illinois 60611

Minoru Hirano, M.D.
Chairman, Department of Otolaryngology
Kurume University Medical School
67 Asahi-Machi
Kurume-Shi 830
JAPAN

Hajime Hirose, M.D.
Research Institute of Logopedics and Phoniatrics
Faculty of Medicine
University of Tokyo
7-3-1 Hongo Bunkyo-ku
Tokyo 113
JAPAN

Harry Hollien, Ph.D.
Director, Communication Science
Institute for Advanced Study of the Communication Processes
63 ASB
University of Florida
Gainesville, Florida 32611

Kyoshi Honda, M.D.
Research Institute of Logopedics and Phoniatrics
Faculty of Medicine
University of Tokyo
7-3-1 Hongo Bunkyo-ku
Tokyo 113
JAPAN

Kenzo Ishizaka, Ph.D.
Communication Engineering Department
University of Electro-Communications
1-5-1 Chofugaoka
Chofu-shi, Tokyo, 182
JAPAN

Nobuhiko Isshiki, M.D.
Director, Division of Plastic and
Reconstructive Surgery
School of Medicine
Kyoto University
Kyoto, 606
JAPAN

Joel C. Kahane, Ph.D.
Department of Audiology and
Speech Pathology
Memphis Speech and Hearing Center
Memphis State University
807 Jefferson Avenue
Memphis, Tennessee 38105

James F. Kaiser, Sc.D.
Bell Communications Research, Inc.
Room 2E-354
435 South Street
Morristown, New Jersey 07960

Yuki Kakita, Ph.D., D.M.Sc.
Department of Electronics
Kanazawa Institute of Technology
7-1 Oogigaoka, Nonoichi-machi
Kanazawa-Minami 921
JAPAN

Toshio Kaneko, M.D.
Professor and Head, Department
of Otorhinolaryngology
Chiba University School of Medicine
1-8-1 Inohana, Chiba, 280
JAPAN

Gail B. Kempster, Ph.D.
Department of Communicative Disorders
Northwestern University
Evanston, Illinois 60201

Peter Kitzing, M.D.
University of Lund
Department of Phoniatrics
Malmö General Hospital
S-214 01 Malmö,
SWEDEN

Kensuke Komatsu, M.D.
Department of Otorhinolaryngology
Chiba University School of Medicine
1-8-1 Inohana, Chiba, 280
JAPAN

Charles R. Larson, Ph.D.
School of Speech
Northwestern University
Frances Searle Building
2299 Sheridan Road
Evanston, Illinois 60201

Erich S. Luschei, Ph.D.
Department of Speech Pathology
and Audiology
University of Iowa
Iowa City, Iowa 52242

Leslie Malmgren, Ph.D.
Department of Otolaryngology
College of Medicine, Research Laboratories
Upstate Medical Center
750 East Adams Street
Syracuse, New York 12310

Paul Milenkovic, Ph.D.
Department of Electrical Engineering
and Speech Motor Control Laboratories
University of Wisconsin
Madison, Wisconsin 53706

Kevin G. Munhall, Ph.D.
Department of Psychology
Stewart Biological Science Building
McGill University
Montreal P.Q.
Quebec, CANADA H3A 1B1

David J. Ostry, Ph.D.
Department of Psychology
McGill University
Montreal, Quebec
H3A, 1B1

Adrienne Perlman, M.A.
Department of Audiology and Speech
Pathology
Department 126
VA Hospital
Iowa City, Iowa 52242

Martin Rothenberg, Ph.D.
Department of Electrical and
Computer Engineering
Syracuse University
111 Link Hall
Syracuse, New York 13210

Shigeji Saito, M.D.
Chairman, Department of Otolaryngology
School of Medicine
Keio University
35 Shinanomachi
Shinjuku-Ku
Tokyo 160
JAPAN

Ronald C. Scherer, Ph.D.
Recording and Research Center
The Denver Center for the Performing Arts
1245 Champa Street
Denver, Colorado 80204

Prof. Carol Schoenhard
Associate Professor of Music
Westminster College
New Wilmington, PA 16142

Herbert M. Teager, Sc.D.
Chief, Bio-Medical Engineering Section
Boston University Medical Center
75 East Newton Street
Boston, Massachusetts 02118

Shushan M. Teager
Bio-Medical Engineering Section
Boston University Medical Center
75 East Newton Street
Boston, Massachusetts 02118

Shinzo Tanaka, M.D.
Department of Otolaryngology
Kochi Medical School
Kohasu Oko Nankoku
Kochi
JAPAN 781-58

Ingo R. Titze, Ph.D.
Department of Speech Pathology
and Audiology
University of Iowa
Iowa City, Iowa 52242

Barry Wyke, M.D., B.S.
The Royal College of Surgeons of England
35-43 Lincoln's Inn Fields
London WC2A 3PN
UNITED KINGDOM

Yoshikazu Yoshida, M.D.
Clinical Professor of Otorhinolaryngology
Kurume University
School of Medicine
1-12-8 Akasaka, Chuo-Ku
Fukuoka 810
JAPAN

James A. Zagzebski, Ph.D.
Department of Medical Physics
1530 Medical Sciences Center
University of Wisconsin
Madison, Wisconsin 53706

David L. Zealear, Ph.D.
Department of Otolaryngology - Head
and Neck Surgery
Northwestern University Medical School
320 East Superior
Chicago, Illinois 60611

FOREWORD

The material in this volume was delivered at the International Conference on the Physiology and Biophysics of Voice in Iowa City, Iowa, U.S.A., May 7-11, 1983. It was the third in a series of conferences that was planned in the initial stages by several members of the Scientific Advisory Board of the Voice Foundation. These individuals were Drs. Hirano, Fujimura, Titze, and Gould. It was felt that a collection of in-depth papers on vocal fold physiology would require the joint work of representative researchers from around the world, each having different disciplines and working environments. Dr. Stevens joined the group of advisors and assumed the important role of becoming one of the co-editors of the first volume, together with Dr. Hirano. This first volume consisted of a collection of papers given in Kurume, Japan in 1980, and was published by Tokyo University Press. The second volume, edited by Drs. Bless and Abbs and published by College-Hill Press, describes the proceedings of the first U.S. conference held in Madison, Wisconsin, in 1981. The fourth conference is now scheduled for New Haven, Connecticut, and will be under the direction of Drs. Katherine Harris and Thomas Baer.

Although termination after five conferences was contemplated in the original plan, the scope has become broader, the quality of the contributions has been sustained and the demand for the material has been sufficiently high that indefinite continuation of this series is conceivable. In order to allow for marked advancement in the various areas of research, a two-year interval between meetings has become more practical than the original one-year projection. No major deviation, however, is expected with regard to the approximate number of participants. The organizers have consistantly stressed the importance of having a small workshop atmosphere with lots of discussion and adequate time for personal exchange of information. Furthermore, the main theme of Vocal Fold Physiology has been preserved, although variations of that theme occur as a continuous reservoir of new topics is being tapped.

The Voice Foundation has helped - and will continue to help - as much as possible. We have maintained our goal of having the meetings in different geographical areas, and this has been quite successful in involving different research groups that would not otherwise be able to attend.

W.J. Gould, M.D.
Chairman of the Board
The Voice Foundation

PREFACE

As the series of vocal fold physiology conferences unfolds, it is becoming evident that each conference is assuming its own identity and particular emphasis. Although the title VOCAL FOLD PHYSIOLOGY is sufficiently broad to include all areas of basic and clinical research on the larynx, the editors are finding it advantageous to highlight the new volumes with a subtitle. This subtitle tends to reflect a small bias toward the areas of investigation of the host institution, which conducts tours of its laboratory facilities and involves much of its staff in technical presentations and discussion.

The first volume, edited by K. Stevens and M. Hirano, carried no subtitle. It served as an introductory volume to the general field of laryngeal physiology. The series of papers on morphological and histological descriptions of vocal fold tissue, along with vibratory patterns observed by various optical techniques, reflected not only the focal point of contemporary research, but also the strength of the Kurume laboratories. The second volume, edited by D. Bless and J. Abbs, carried the subtitle: Contemporary Research and Clinical Issues. It not only provided the reader with a thorough update of many of the topics covered in the first volume, but also included major contributions in the area of clinical voice. Emphasis was on a better merger between the voice laboratory and the voice clinic.

The present issue has been given the subtitle: Biomechanics, Acoustics, and Phonatory Control. This is not to exclude clinical or pedagogical topics, but rather to focus intensely on the mechanism of voice production with a biophysical orientation. In particular, an attempt was made to bring together a group of experts who could address the problem of acoustic interaction of a source of sound with a resonator when independence between the two is not a valid assumption. Thus, while the traditional source-filter theory of speech production was beginning to be challenged in the previous (Madison) Conference, an effort was made at Iowa to have adequate representation to make the discussion a focal point of the conference.

Interactions between the source and the resonator are not limited to the acoustic domain. The entire soft-walled vocal tract, including many structures of the head, neck and shoulders, are set in motion during phonation. The greatest of these tissue-air interactions occur at the vocal folds. This oscillating structure, with its rich sensory and motor nerve supply, is one of the most intriguing and unique control systems to analyze. In no other organ of the body does one encounter tissue being accelerated and decelerated at such rapid rates, with continual collision being a requirement, with the balance of fluid within and on the surface of the tissue being critical for sustained oscillation, and with the formation of jet streams and air turbulence being key factors in the energy conversion process from the pulmonary airstream to acoustic waves radiated at the mouth.

In a number of ways the system is analogous to a hydroelectric power plant with its multifaceted controls. The available power from the reservoir (or lungs) depends upon the product of the pressure developed at the bottom of the dam (or below the vocal folds) and the average flow produced at exit. The efficiency with which the high-energy jet drives the turbine and the generator (or the vocal tract) depends upon the pressure impulses imparted to the vanes of the turbine (or to the supraglottal air column). Maintaining constant power output and constant frequency of the rotating shaft of the turbine and generator is a delicate control problem. Similarily, regulating intensity and fundamental frequency of the voice as the acoustic load of the vocal tract or the viscoelastic properties of the vocal folds change, is a delicate task for the sensory and motor systems of the body.

In order to fully understand the operation of such a complex physical (or physiologic) plant and make predictions about its behavior under yet unobserved conditions, one needs to study the structural components separately, one needs to have methods for observing and monitoring the integrated dynamic behavior, and one needs to understand the physical laws that govern the interactions between the components.

To this end, we have divided this volume into three major parts,

I. Anatomy and Neurophysiology of the Larynx.
II. Vocal Fold Kinematics and Techniques for Clinical and Pedogogical Observation
III. Biomechanic, Aerodynamic, and Acoustic Interactions

Part I begins with a comparative anatomy of the larynx across species (Chapter 1) followed by structural details of intrinsic laryngeal muscles (Chapters 2-4). The neuroanatomy of the motor system is then investigated (Chapters 5 and 6), and four chapters (7-10) on control of fundamental frequency and vocal fold adduction demonstrate the important role that electromyography has played in relating specific motor commands to phonatory gestures. Part I is concluded with a chapter on the reflexogenic (sensory) system that is responsible for controlling the vocal output (Chapter 11).

Part II provides a sampling of various techniques for observation of the vocal folds in action. Chapters 12 and 13 describe ultrasonic techniques for tracking the motion of tissue-air and tissue-tissue interfaces, while Chapters 14 and 15 describe procedures for mapping the two-dimensional trajectories (vertical and horizontal) of tissue particles either experimentally (by use of implanted lead pellets and an x-ray tracking beam) or computationally (by use of the finite-element method). Optical methods of quantifying vocal fold movement are then discussed (Chapter 16) and the combination of optical techniques with glottographic techniques (Chapter 17) leads into a series of clinical and pedagogical topics (Chapters 18-23) that combine glottographic, acoustic, and fiberscopic observation with subjective assessment of vocal function.

Part III deals exclusively with the biophysics of voice production. It is essentially a discussion of the physical laws that govern the motion of tissue and air as sound is produced. These laws, derived from established theories in the fields of aerodynamics, acoustics, and biomechanics, are believed to contain sufficient predictive power to model the mechanics of the vocal folds in a quantitive fashion. Chapters 24 through 26 describe the deformation of vocal fold tissue when external and internal (muscular) forces are applied. This leads to a direct measure of the viscoelastic properties of the vocal folds. Indirect ways of determining these properties, by measuring heat production (Chapter 28), offer the advantage of keeping the system more intact. The self-oscillatory mechanism resulting from the interaction between the airstream and vocal fold tissue is then discussed in Chapters 29-30, which is followed by a new look at glottal and vocal tract aerodynamics (Chapters 31-34), and finally by a series of contributions to the problem of acoustic interactions between the glottal source and the vocal tract resonator (Chapters 35-40).

The organization of the book turned out to be somewhat different from the organization of the conference. Greater effort was made to group the papers according to content rather than equal distribution per session. We are very grateful to the section editors, Dr. L. Malmgren, Dr. C. Larson, Dr. K. Munhall, Dr. D. Childers, and Dr. P. Milenkovic for their unselfish and unheralded contributions to the editing process, particularly with regard to the discussion material at the end of each chapter.

The major credit for bringing about this volume, as well as the previous ones, has to go to Dr. W.J. Gould and the Voice Foundation. In the field of voice, Dr. Gould has become the primary motivation force in bringing to action what many of us only dream about, and the Voice Foundation has become the initial source of support for many projects that might otherwise have been discarded as being unproven or unconventional. The Voice Foundation was the primary sponsor of this conference.

Special thanks must also be given to the University of Iowa for co-sponsorship. Dr. D.C. Spriesterbach, Vice President for Educational Development and Research, made available all the local university facilities, Dr. K. Moll presented a stimulating welcome address, and Robert Davis provided the resources of his Conference Center staff.

The entire conference would have been doomed to failure, however, without the untiring services of our conference secretary, Mary Lee, who took care not only of the immediate crises, but worked well in advance to foresee the needs of the conference participants. Also, Kim Sheplor and Debby Pittard have gained our deepest affection and gratitude for typing the manuscripts and maintaining order in the publication process. We would also like to thank Vicki Grabowski, Melanie Simonet, Grant Kuhn, Kerry Kovac, and Janice Lampert from The Denver Center for the Performing Arts' Publication Services Department for their untiring help and dedication in making this book become a reality. Finally, we would like to thank Vivian Heggie and Andy Cleary from Network Graphics, Denver, for their diligent help in the typesetting of this book.

Ingo R. Titze, Ph.D.
The University of Iowa Iowa City, Iowa 52242
Ronald C. Scherer, Ph.D.
The Recording and Research Center, The Denver Center for the Performing Arts,
Denver, Colorado 80204

I

Anatomy and Neurophysiology of the Larynx

Section Editors

Leslie T. Malmgren
Charles R. Larson

1 • COMPARATIVE ANATOMY OF THE LARYNX: EVOLUTION AND FUNCTION

Malcolm H. Hast

When the first vertebrates migrated from an aquatic to a terrestrial environment in the early Carboniferous period of the Paleozoic era some 350 million years ago (Romer, 1966; Romer and Parsons, 1977), they were preceded by the vertebrate ancestor of the mammalian larynx, the bichir Polypterus. This African ray-finned fish possesses both external gills and paired ventral lungs, as opposed to other ray-finned fishes which have an air or swim bladder, the function of which is that of a hydrostatic organ (Romer and Parsons, 1977). The "larynx" of Polypterus is composed of a simple muscle sphincter to protect the lung from water (Figure 1-1) (Negus, 1929, 1949). In the later-staged African lung fish (Protopterus) and the Australian lung fish (Neoceratodus), a mechanism is developed that not only protects the airway by a sphincteric musculature but allows for active dilatation by a group of separate muscle fibers that pull the valvular margins apart. The primitive larynx of the lung-fish offers a great advantage to this vertebrate. The muscular sphincter protects the lung when this river fish is in water, but during a drought, the dilator muscle fibers actively open the sphincter to allow for air to be taken into the lung by swallowing (Romer and Parsons, 1977; Negus, 1949). Development of dilator fibers by Protopterus, a structural and functional change in this early vertebrate, is an excellent example of an adaptive modification to a particular environment and mode of life.

The next modification of the larynx is seen in the evolution of the vertebrate from aquatic to terrestrial being (Negus 1949; Romer and Parsons, 1977). With the development of tetrapod amphibia, the salamander for example, the larynx truly becomes an individual organ, with

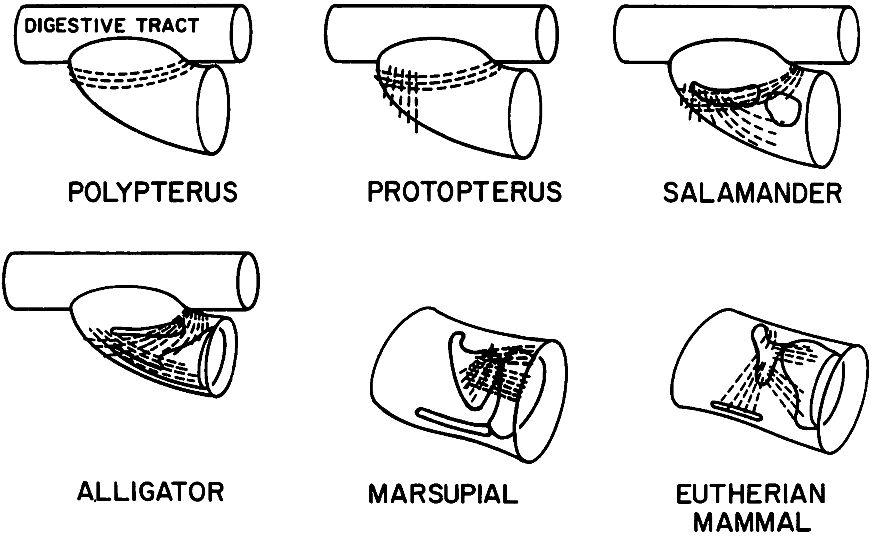


FIGURE 1-1
Evolution of the larynx (partly after Negus) illustrating development of cartilages and muscles.

lateral cartilaginous plates (representing the future arytenoid cartilages and part of the future cricoid cartilage) into which are inserted the dilator muscle fibers for active abduction of the glottis to meet enhanced ventilatory flow requirements of a land animal. In the frog (a tailless amphibian) the larynx becomes even more sophisticated for a more terrestrial life.

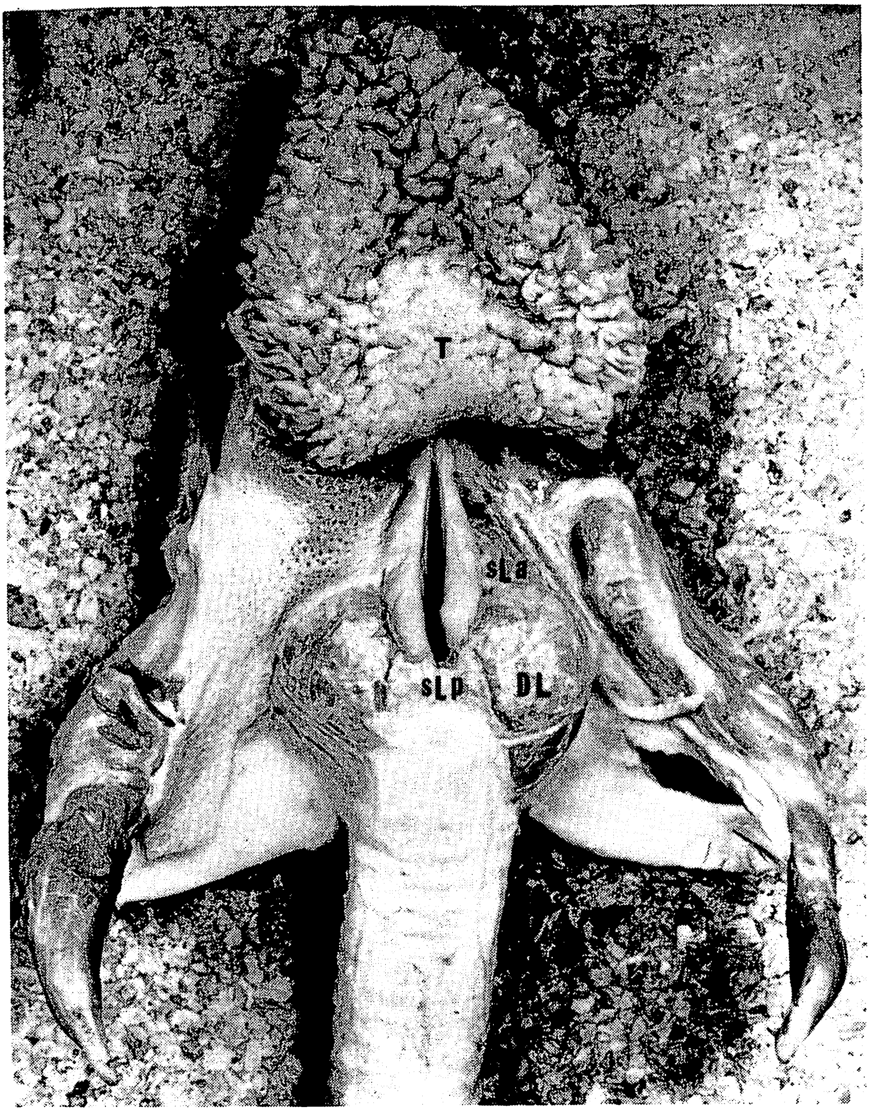


FIGURE 1-2
Dorsal view of the larynx of the Aldabran giant tortoise (Geochelone gigantea gigantea); (DL) m. dilator laryngis, (SLa) m. sphincter laryngis anterior, (SLp) m. sphincter laryngis posterior (t) Tongue.

The succeeding stage in the development of the larynx is seen in the modification of the laryngeal skeleton of the reptile. A fused cricothyroid plate is developed, with both the musculus dilator laryngis and musculus sphincter laryngis taking origin from the cricothyroid cartilage and inserting into the arytenoid cartilage. The tortoise and alligator are representative examples of this stage of animal which has become better adapted for a terrestrial habitat than the amphibian (Negus, 1949; Romer and Parsons, 1977).

Figure 1-2 shows the larynx of the Aldabran giant tortoise. Although well adapted to the functions of protection of the airway from fluids and for respiration, this larynx appears to be poorly adapted for phonation, possessing long and rigid arytenoid cartilaginous vocal folds. It is interesting to note that this tortoise larynx is at least as "advanced" hylogenetically, if not more so, than the frog; yet the infrequent grunts or roars of this reptile are hardly

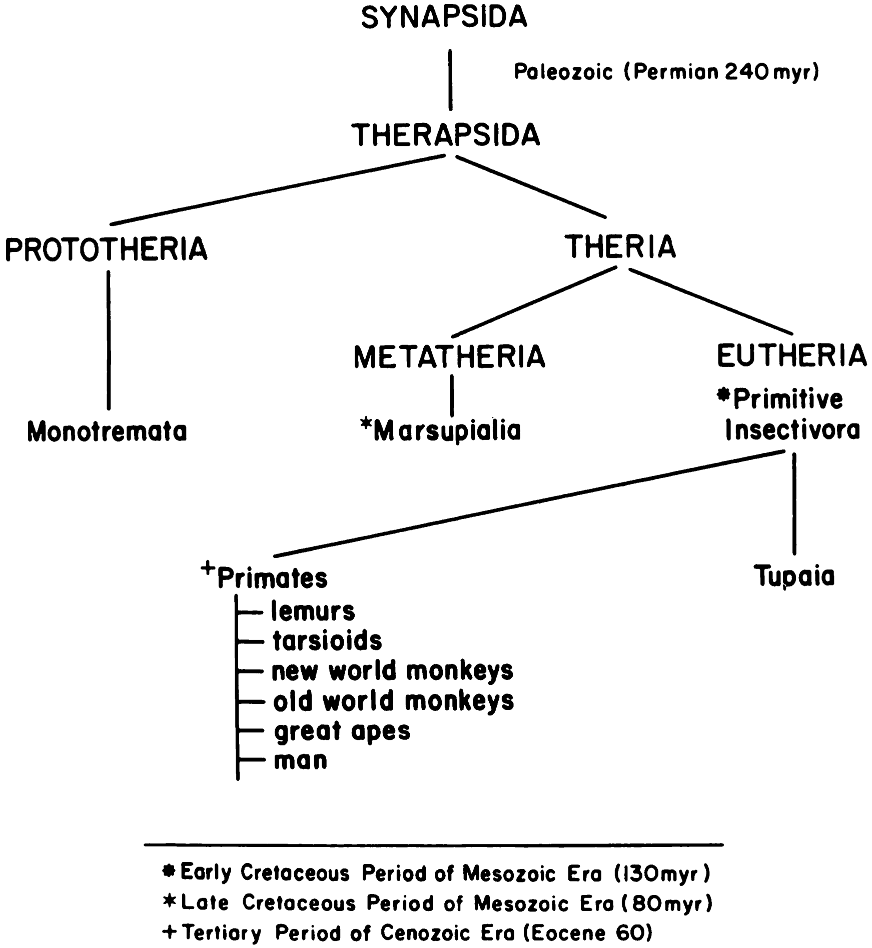


FIGURE 1-3
Simplified genealogy of hypothetical lines of descent of several mammalian orders with geological time scale.

noticeable when compared to the frequent and noisy croaks of the frog. One could attribute the louder voice to the frog's buccal air sac employed as a vocal resonator. But there is no morphological explanation for the difference in the use of the vocal mechanism. The needs of an animal probably determine its phonatory activity, not only the anatomy of its larynx, for the vocal repertoire has many functions: recognition, mating, social organization, territoriality, etc.

The above observation demonstrates the difficulty in adopting an evolutionary philosophy or hypothesis based on the scala naturae, a pseudo-evolutionary interpretation of the "phylogenetic scale." In fact, the theory of evolution is antithetical to a scala naturae that places Hominidae at the top and amoeba at the bottom; the "tree of life" is not vertical (simple) but many branched. What is meant by a "higher animal"? It should also be noted, and this is a profound weakness in trying to reconstruct a phylogeny (evolutionary history) based on any organ of the body, that we must reason from the evidence at hand, describing characteristics of contemporary animals which are themselves the end-points. Very similar anatomical structures can be found in superficially dissimilar animals.

With the appearance of the mammal (Figure 1-3) (Romer, 1966; Romer and Parsons, 1977) a true but immovable thyroid cartilage is found articulated to the cricoid cartilage (Romer and Parsons, 1977). This skeletal arrangement is seen in the reptile-like modern mammal Monotremata, and in Marsupialia which probably arose 80 million years ago in the Crestaceous period of the Mesozoic era (Vaughan, 1978). The New Guinea spotted cuscus (Phalanger maculatus), a typical marsupial (Figure 1-4), demonstrates the larynx of a phylogenetically early mammal. No cricothyroid muscle is found to stretch the vocal cords, and the dilator muscle is composed of two parts, cricoarytenoideus dorsalis and a cricoarytenoideus latero-dorsalis. The only other muscle is a bilateral sphincter whose

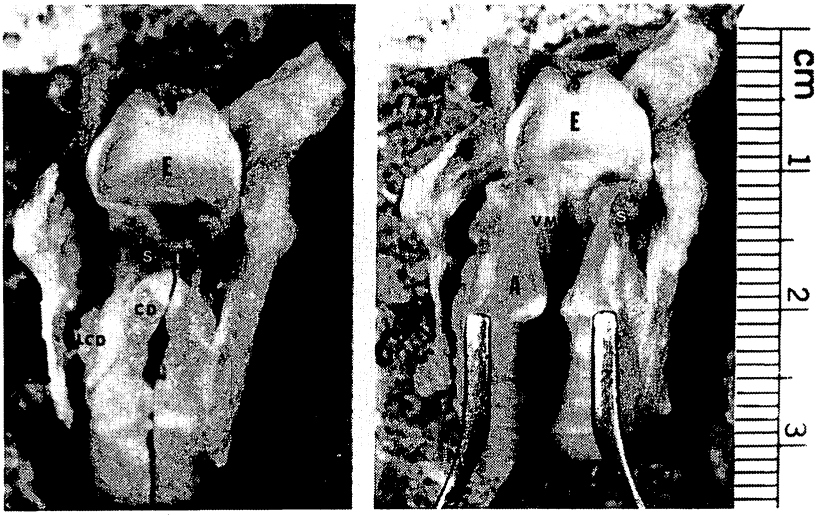


FIGURE 1-4

Dorsal and internal dorsal views of the larynx of the marsupial spotted cuscus (Phalanger maculatus). (E) epiglottis, (C) corniculate cartilage, (A) arytenoid cartilage, (S) sphincter muscle, (t) transverse sphincter ligament, (LCD) m. cricoarytenoideus lateral-dorsalis, (CD) m. cricoarytenoideus dorsalis, (VM) vocal membrane.

origins are from the lateral and posterior portions of the very large arytenoids, to insert at a common transverse arytenoid tendon. More than three quarters of the vocal fold is cartilaginous, unfavorable to versatile voice production but advantageous to increased respiratory demands in running.[1] A typical mammalian structure, the epiglottis is large and well developed in this keen scented (macrosomatic) animal; it lies dorsal to and in contact with the palate. It is to be noted that the primitive, nonvalvular type of larynx of the cuscus is atypical of an aboreal mammal with prehensile tail (Negus, 1929, 1949; Walker, 1974).

Beginning with the insectivore order, there is a skeletal change in the larynx that is unique to all modern mammals (except Monotremata and Marsupialia); a true cricothyroid joint develops with movements of the cricoid and thyroid cartilages adjusted by a muscle of the same name (Negus, 1929, 1949). Additional modifications include:

1. the division of the thyroarytenoid fold into inferior and superior folds, plica vocalis and plica vestibularis, found in insectivores (Figure 1-5), endentates (Figure 1-6), some ungulates, and primates for example;
2. development of the laryngeal sphincter into a lateral thyroarytenoid muscle and medial vocalis muscle;
3. a reduction in size of the arytenoid cartilage, notably in primates (Figures 1-7, 1-8);
4. the evolution of a small epiglottis in microsomatic animals like the California sea lion, many arboreal mammals and cetaceans;
5. aryepiglottic folds;
6. the increased development of intralaryngeal (ventricular) and extralaryngeal air sacs in many placentals and in all primates; and
7. the descent of the larynx to its lowest position in the great apes.

Examples of the above developments or modifications in the evolution of the larynx can be demonstrated by a study of its morphology in a number of species from several orders of mammal. These include the following: African hedgehog (Erinaceus albiventris), an

FIGURE 1-5
Ventral and dorsal views of the larynx of the insectivore African hedgehog (Erinaceus albiventris); (CT) cricothyroid muscle, (E) epiglottis, (T) thyroid cartilage, (CD) dorsal cricoarytenoid muscle.

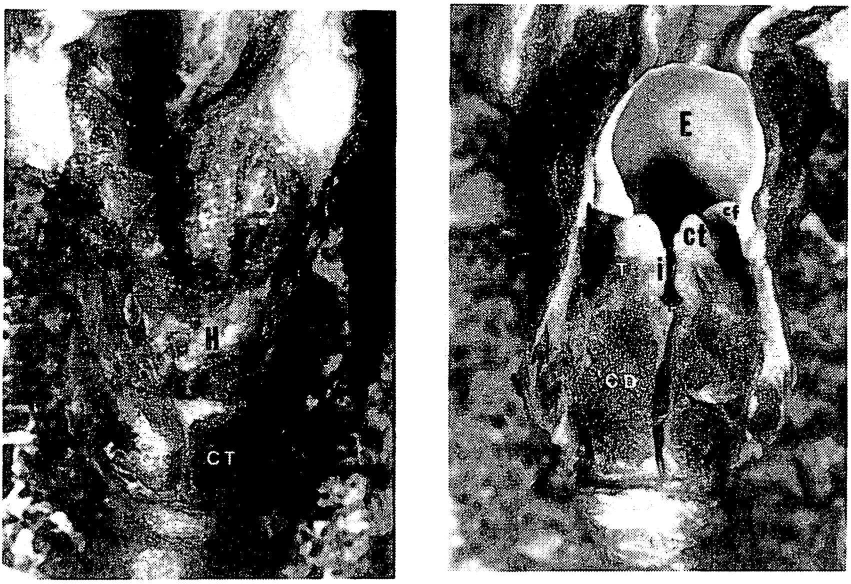


FIGURE 1-6
Ventral and dorsal views of the larynx of the edentate collared anteater (Tamandua tetradactyla); (H) Hyoid, (CT) cricothyroid muscle, (E) epiglottis, (T) transverse arytenoid muscle, (i) interarytenoid ligament, (ct) corniculate cartilage, (cf) cuneiform cartilage, (CD) dorsal cricoarytenoid muscle.

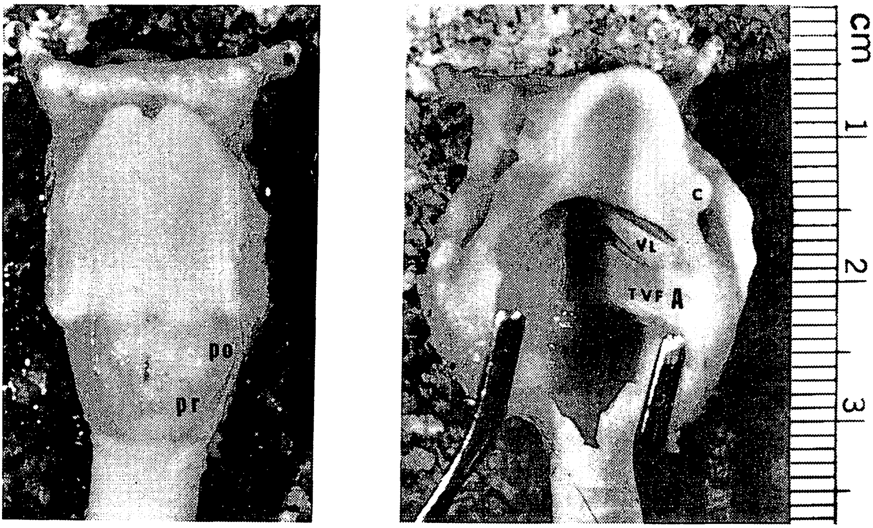


FIGURE 1-7
Ventral and internal dorsal views of the larynx of the prosimian primate brown lemur (Lemur fulvus). (po) pars obliqua and (pr) pars recta of cricothyroid muscle, (c) cuneiform cartilage, (VL) vestibular ligament, (TVF) true vocal fold, (A) arytenoid cartilage.

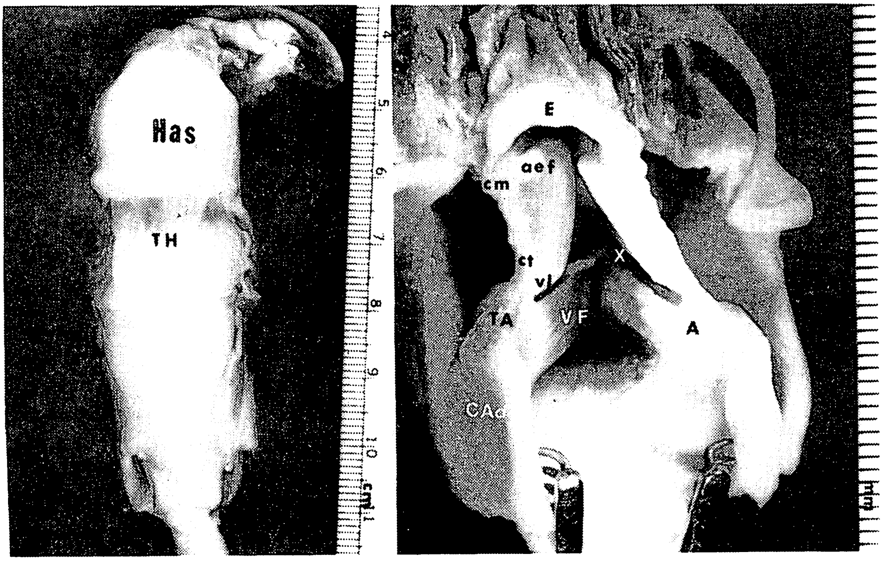


FIGURE 1-8
Ventral and internal dorsal views of the larynx of the primate black howler monkey (Alouatta carayo); (Has) hyoid air sac, (TH) thyrohyoid membrane, (E) epiglottis, (aef) aryepiglottic fold, (cm) cuneiform cartilage, (ct) corniculate cartilage, (vl) vestibular ligament, (TA) transverse arytenoid muscle, (VF) vocal fold, (A) arytenoid cartilage, (X) laryngeal ventricle leading to hyoid air sac, (CAd) dorsal cricoarytenoid muscle.

---

insectivore; collared anteater (Tamandua tetradactyla), an edentate; sheep (Ovis aries), an artiodactylan; brown lemur (Lemur fulvus) and black howler monkey (Alouatta carayo), primates; Amazon river dolphin (Inia geoffrensis), a cetacean; hammer-headed fruit bat (Hypsignathus monstrosus), a chiropteran.

In both the African hedgehog (Figure 1-5) and Old World monkeys and apes, the arytenoids are short and the inferior folds (vocal) are long and sharp. With the prosimian brown lemur (Figure 1-7) and the New World prehensile black howler monkey (Figure 1-8), the arytenoids are very short and the vocal folds are very long and sharp. But unlike the nocturnal and macrosomatic insectivore with a big intranarial epiglottis, the microsomatic ape has a small epiglottis that just touches the palate. The lemur has a large larynx and intranarial epiglottis, while the black howler monkey's large larynx and small epiglottis just touches the palate. Although not affording as big and protective an airway as the sheep (Figure 1-9) (Hast, 1979), with its large arytenoids and high vaulted flat vocal folds, the separate long and sharp inferior vocal folds of the primate provide both for an efficient valve which fixes the thorax for climbing and lifting a heavy weight, and for excellent adaptation for superior vocalization.

It is interesting to note that the macrosomatic anteater, Tamandua, shown in Figure 1-6, with a large larynx similar in structure to many ungulates, can only open its mouth about 6 mm (Walker, 1974), and is usually silent. In the concepts of natural selection and adaptation for survival, the morphology of this anteater's laryngeal organ probably can be found and explained only by a detailed study of the orders ancestral to Edentata. This approach is called by Darwin "unity of descent," which explains the old law of "unity of type." In Darwin's words: "By unity of type is meant that fundamental agreement in structure, which

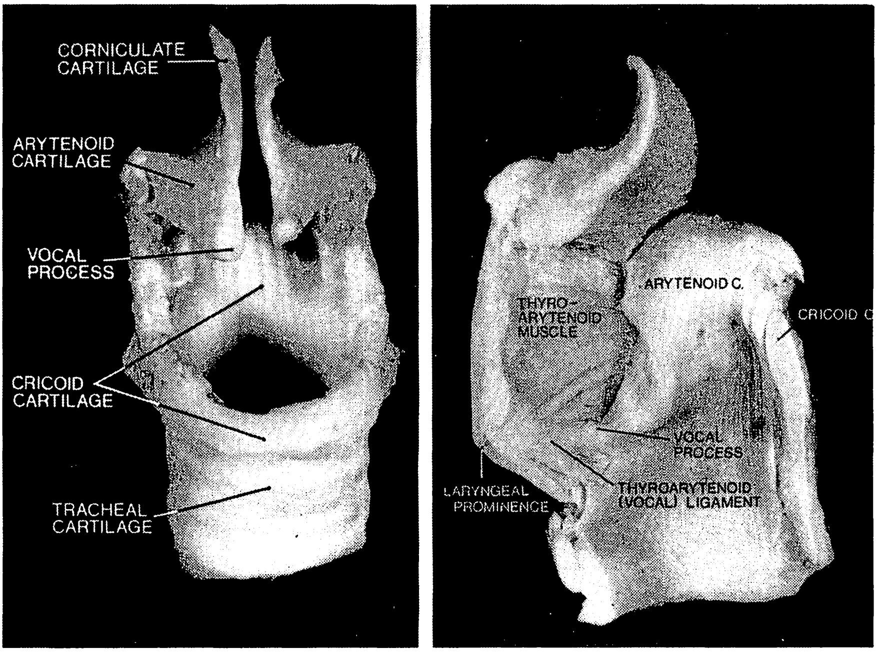


FIGURE 1-9
Cartilaginous skeleton and midsagital views of the larynx of the artiodactylan sheep (Ovis aries).

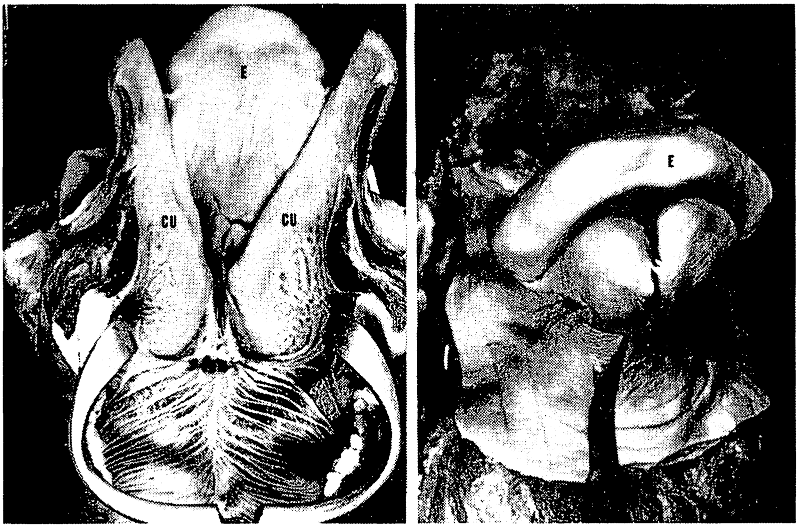


FIGURE 1-10
Internal dorsal views and entrance of the larynx of the cetacean Amazon river dolphin (Inia geoffrensis); (E) epiglottis, (CU) cuneiform cartilage, (TA) thyroarytenoid muscle.

we see in organic beings of the same class, and which is quite independent of their habits of life" (Darwin, 1859).

Before proceeding to a description of two very different larynges in this study of comparative anatomy, an important modifier of vocalization, the air sac, must be considered. Both ventricular and non-ventricular air sacs are particularly well developed in the primate (Negus, 1929, 1949). A notable example of a non-ventricular air sac is seen in the hyoid air sac of the howler monkey. In addition to the tracheobronchial tree and pharynx, the bony hyolaryngeal air sac of the howler monkey is an important resonator of the voice of this primate. Since the howler monkey is primarily herbivorous, it has the typically large epiglottis which is in opposition to the soft palate, to block off the oral cavity. With the loss of the oral cavity as a resonating chamber, the hyolaryngeal air sac can compensate for this loss and becomes the principal resonator of the voice (Hilloowala, 1976). The loud howl or roar of the howler monkey is essential to this mammal, since vocalizations must reach up to two miles through tropical rain forests to its mate or members of its troop (Vaughan, 1978). With the ultimate decensus of the larynx in Hominidae, the oral cavity, teeth, and lips become active resonators of the voice.

Figure 1-10 shows the larynx of the Amazon river dolphin, a cetacean. This intranarial

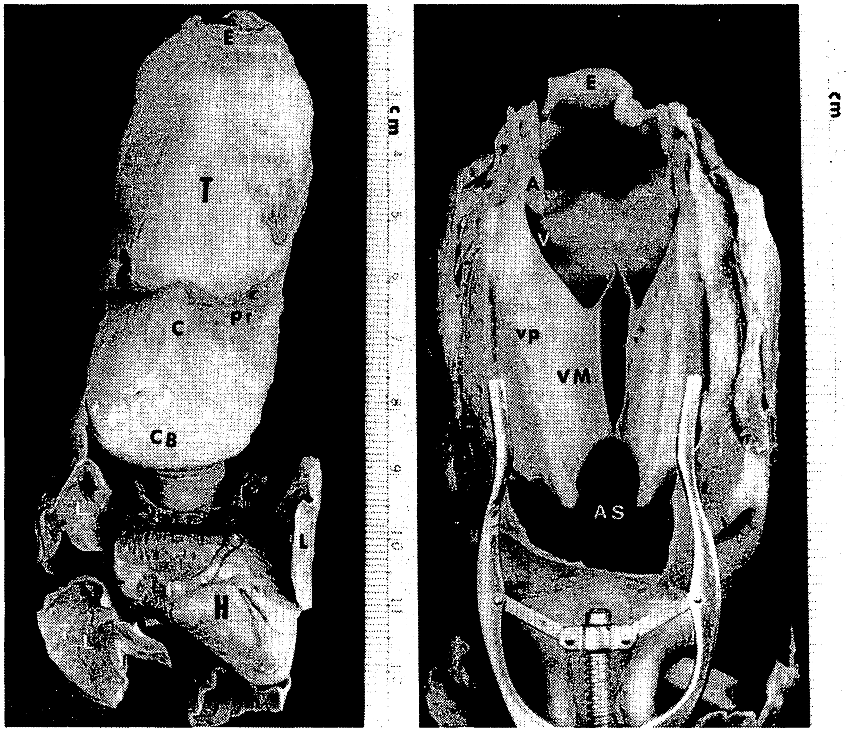


FIGURE 1-11

Ventral and internal dorsal views of the larynx of the megachiropteran hammer-headed bat (Hypsignathus monstrosus); (E) epiglottis, (T) thyroid cartilage, (C) cricoid cartilage, (pr) pars recta of cricothyroid muscle, (CB) cricoid bulla (subcricoid air sac), (L) Lungs, (H) Heart, (A) arytenoid cartilage, (V) ventricle, (vp) vocal process of arytenoid, (VM) vocal membrane, (AS) sub-cricoid air sac.

larynx is composed of the same cartilages and muscles as most mammals. Unlike the typical mammal, though, this larynx possesses unusually big and lengthy cuneiform cartilages, forming with the epiglottis a spout-like structure which produces whistles and click sounds (Blevins and Parkins, 1973). Two other interesting structures are: (1) the entrance to the larynx that resembles the mouth of the dolphin and (2) prominent fibers that radiate from the anterior cricothyroid membrane to encircle the mucosa at the level of the cricoid, which resemble the musculi pectinati of the right atrium of the heart. Could these distinctive fibrous ridges function as acoustic lenses?

Finally, we come to the largest larynx, proportionate to body size, of any living mammal—the male larynx of the hammer-headed fruit bat (Figure 1-11) (Vaughan, 1978). In contrast to the larynx of the nocturnal feeding echolocating Microchiroptera, which emits ultrasonic pulses for orientation, the Megachiropteran hammer-headed bat relies on vision and olfaction to search for its principal food, fruit. The male larynx of this mammal measures 68 mm in length, which, in proportion to its crownrump length of 260 mm, is about one-quarter of total body length. This larynx is so large that it occupies most of the space usually filled by the heart and lungs of most mammals. Using its guttural, explosive, and blaring call to attract females, this bat is unique among mammals in its unusual development of the laryngeal organ, for, "the remarkable specialization of the vocal apparatus clearly evolved in association with the importance of loud vocalizations during breeding displays" (Vaughan, 1978). This flying "loud speaker" also has developed very large lips which can be formed into a kind of "megaphone," and a subcricoid cartilaginous air sac that has not been found in other mammals. Another feature of this larynx, and of other bat larynges, is the structure of the cricothyroid muscle, which is composed of both superior and inferior oblique parts in addition to a straight part.[2] Research has shown that the cricothyroid muscle has a more important function in the mechanism of vocalization in chiropteras than in other mammals (Suthers and Fattu, 1973).

A study of the comparative anatomy of the larynx has been made and the results presented in the context of a general theory of vertebrate evolution. From a primitive muscle sphincter protecting the airway of the bichir fish to the simple cartilaginous structure of the salamander larynx, and finally to the advanced laryngeal organ of the hammer-headed bat, we have seen the larynx change and develop into a very specialized organ of communication that is essential to the daily existence and survival of mammals.

ACKNOWLEDGEMENT

The author wishes to thank the Hunterian Institute of the Royal College of Surgeons of England for providing facilities to carry out this study; the interest shown and the assistance given by the staff of the Department of Anatomy deserve special mention. I am grateful also to the Lincoln Park Zoological Gardens, Chicago, for supplying most of the specimens.

Footnotes:

[1] The large arytenoids, internal sphincter muscle (transverse arytenoid), and posterior cricoarytenoid muscle are the most prominent structures observed in the early human embryo (42 days) (Hast, 1972).

[2] For a more detailed description of this larynx, see Schneider et al, 1967.

REFERENCES

Blevens, C.E. and Parkins, B.J. (1973). "Functional anatomy of the porpoise larynx," *Am. J. Anat.* 138, 151-164.

Darwin, C. (1859). *The Origin of Species*. John Murray, London.

Hast, M.H. (1972). "Early development of the human laryngeal muscles," *Ann. Otol., Rhinol. and Laryngol.* 81(4), 524-531.

Hast, M.H. (1979). "Anatomy of the sheep larynx," *J. Anat.* 129(4), 824-855.

Hilloowala, R.A. (1976). "The primate hyolaryngeal apparatus and herbivorous modifications," *Acta Anat.* 95, 260-278.

Negus, V.E. (1929). *The Mechanism of the Larynx*. Wm. Heinemann Medical Books Ltd., London.

Negus, V.E. (1949). *The Comparative Anatomy and Physiology of the Larynx*. Wm. Heinemann Medical Books Ltd., London.

Romer, A.S. (1966). *Vertebrate Paleontology*, 3rd Ed. University of Chicago Press, Chicago.

Romer, A.S. and Parsons, T.S. (1977). *The Vertebrate Body*, 5th Ed. Saunders College, Philadelphia.

Schneider, R., Kuhn, H. and Kelemen, G. (1967). "Der Larynx mannlichen" *Hypsignathus monstrosus*, Allen, 1981 (Pteropodidae, Megachiroptera, Mammalia). *Z. Wiss. Zool.* 175(1-2), 1-53.

Suthers, R.A. and Fattu, J.M. (1973). "Mechanisms of sound production by echolocating bats," *Amer. Zool.* 13, 1215-1226.

Vaughan, T.A. (1978). *Mammalogy*, 2nd Ed. Saunders College, Philadelphia.

Walker, E.P. (1974). *Mammals of the World*, 3rd. Ed. 2 Vols. John Hopkins Press, Baltimore.

DISCUSSION

N. ISSHIKI: I am much impressed by your presentation, which shows there are so many varieties of larynges among different animals. I have long been wondering why some animals can phonate and some others cannot. What are the structural features of the larynx which separate the vocal larynx from the non-vocal larynx?

M. HAST: I've been asked this question before. First of all, if there are 4,700 species of mammals, you could say there are 4,700 different larynges. Secondly, why do some animals with very sophisticated larynges have very little to say, so to speak, and others with simpler larynges have a great deal to say; for example, the sea lion and the frog? I think it's the way animals live. It's what is needed for them to survive. Voice is part of their pattern of behavior, if they need it. Or do they need olfaction more: Which particular sense of communication do they use? I think this is part of it. When people ask me, "Well, if the monkeys, for example, have larynges that are just as good as ours," (and they are just as sophisticated) "why don't they communicate in a language like humans?" I think the answer will be found in studying the development of the brain. It may sound simplistic, but very frankly, if we with our brain had the primitive larynx of a marsupial, we would be able to communicate in a human language, even if we didn't have a movable cricothyroid joint. I don't even know how much a wallaby would benefit possessing a human larynx. I think the amount of vocalization is probably governed more by the mode of life, behavior, and intelligence than by structure.

P. KITZING: Is there any other animal which has a descended larynx like humans?

M. HAST: That is far away from the palate, that is descended a great distance from the palate?

P. KITZING: I mean in newborns, the larynx is high, very high.

M. HAST: The human neonate is very similar to these animals, yes, and then it descends.

No, not even in animals that are more bipedal, like the anthropoid ape, is the larynx as high in relation to the cervical vertebrae. The more natural stance for most of the apes is still bent over or on all fours. Among the primates and anthropoid apes, the larynx is just like you talk about in the newborn; it is very high up, about cervical vertebra C-1 or C-2, unlike the mature humanoid larynx which has descended to between C-3 or C-4. Because we are bipedal all the time, (with gravity) all parts seem to be descending.

D. BLESS: You describe characteristics of the larynx peculiar to specific species (e.g. bifed epiglottis). How much variation in laryngeal structures do you see among animals of the same species?

M. HAST: I have found that, among animals of the same species, there is no obvious variation; among animals of the same genera, there are slight or few differences; and among animals of the same order, there are significant differences. There are, however, obvious differences in size between sexes, as in the human.

W. CONRAD: In which species do we start to find false vocal folds? Is there a gradual development of the ability to close or adduct the false vocal folds? Is there any correlation between the sophistication of sounds produced and the ability to adduct the false vocal folds?

M. HAST: In answer to your first question, we find false folds in the tapir, pig, porcupine, raccoon, horse, dog, bear, hedgehog, badger, bat, monkey, ape, and human, for example. In answer to your other questions, as a scientist (and physiologist) I cannot answer these two questions based solely on my observations of structure. These are interesting questions; I just don't know the answer.

T. BAER: I am interested in the relationship you note between the presence of an air sac and the intensity of the species' voice. In some cases, at least, the main purpose of the sac appears to be for rebreathing in species that need to close the respiratory tract for extended durations (Negus). To determine whether the sac would be useful as a resonator, it would be useful to know both the dimensions of the sac and of the connecting duct. Also, the location of the connecting duct. Are these data available?

M. HAST: I have some unpublished data in my notes; but for a better discussion of this problem, I refer you to Hilloowala's article (see text) or to a book by P. Herskowitz on "Living New World Monkeys".

2 ▪ BOUNDARIES OF THE CRICOTHYROID, THYROARYTENOID AND LATERAL CRICOARYTENOID MUSCLES FORMED BY ARTERIAL PEDICLES FROM THE SUPERIOR AND INFERIOR LARYNGEAL ARTERIES

Joel C. Kahane

Thyroarytenoid (TA), lateral cricoarytenoid (LCA) and cricothyroid muscles (CT) make up 69 percent of the total weight of the intrinsic laryngeal muscles (Bowden and Scheure, 1960). Despite their anatomical importance and functional significance, information is lacking regarding the anatomical boundaries between these muscles in different parts of the larynx. Such information is needed to provide complete descriptions of these muscles which may be particularly useful to morphologists, electromyographers and those developing models of the larynx.

Determining the boundaries between the TA, LCA and CT muscles is difficult for several reasons. These include, their small size and the close proximity of the muscles to each other, the similarity of fiber direction of different muscles in a given region of the larynx and the lack of well developed fascial sheaths.

Although these muscles are grossly distinguishable from each other, their histologic boundaries are not nearly so distinct. It would be useful to find a means to determine the boundaries between them. It appeared to us that work from Terracol and Guerrier (1951) on the arterial supply of the larynx might provide a framework to study the aforementioned intrinsic laryngeal muscle relationships.

These investigators showed that the interior of the larynx is supplied by three arterial pathways called "arcades," which are derived from the superior and inferior laryngeal arteries. The three arcades form a triangular network in the interior of the larynx which consists of anterior, posterior and transverse (anteroposterior) routes. These vessels are accompanied by branches from the superior and inferior laryngeal veins and are also closely associated with the superior and inferior laryngeal nerves. Terracol and Guerrier unfortunately did not describe the specific pattern(s) of the blood supply to the intrinsic laryngeal muscles.

Such information might be useful in defining morphologic relationships among the intrinsic laryngeal muscles. It has been noted by Warwick and Williams (1973) that in many skeletal muscles "... branches of the principal artery and nerve... enter [the muscle] together along a strip, often fairly constant in position ..." (p. 483). This suggests that closer examination of the relationship between the intrinsic laryngeal muscles and branches of the arterial arcades described by Terracol and Guerrier (1951) may be useful in establishing an anatomical basis for defining the boundaries between these muscles.

The purpose of the present study was to (1) investigate the morphology of the vascular pedicles derived from the superior and inferior laryngeal arteries and (2) describe their relationship to the TA, LCA and CT muscles.

METHODS

Laryngeal Specimens

Twenty human larynges (10 male and 10 female) were obtained from autopsy. Each group contained a one month infant and 9 adult specimens. The male adult specimens ranged in age from 16-71 years, Md age = 49 years; the female adult specimens ranged in age from 20-92 years, Md age = 56 years. All larynges came from persons who were free of laryngeal trauma or from diseases known or suspected of adversely affecting laryngeal function.

Histologic Preparation

Specimens were preserved in 10% buffered formalin. They were decalcified in EDTA and prepared for routine histologic processing and embedding in paraffin. Each larynx was sectioned in the coronal plane at 6μm thickness. Sections were taken at 125μm with either Safranin-Phloxine-Hematoxylin or Iron Gallien Elastic stains.

Microscopic Analysis

Serial sections were studied via light microscopy and stereo macroscopy. Histologic reconstructions were made using a Bauch and Lamb Tri Simplex projection microscope. Sections were examined to determine the origin, termination, course, muscular relationships and composition of the principal arterial arcades and their branches associated with the TA, LCA and CT muscles.

Tracings were made of vascular compartmentation in the aforementioned muscles in the anterior commissure, anterior, middle and posterior regions of the vocal folds and the mid-arytenoid cartilage and cricothyroid joint regions. These tracings were used to make composite reconstructions of the typical pattern(s) of the vascular pedicles in these regions, respectively. Light microscopy was used to resolve fine details of structure and to confirm the identity of contiguous muscle fibers from two or more different muscles which share common attachments (e.g., medial fibers of pars oblique and pars rectus of the CT muscle from LCA, LCA from TA in the arytenoid cartilage area and LCA from PCA at the muscular process).

RESULTS

Typical Morphology

The typical topography of a vascular pedicle is shown in coronal sections from the anterior and middle thirds of the vocal fold (Figure 2-1). Each of the pedicles are derived from a neighboring arcade. Note how they separate contiguous muscles (discussed more fully in a subsequent section). A more detailed microscopic view of the morphology of a vascular pedicle is seen in Figure 2-2, taken from the middle third of the vocal fold. Note that this section illustrates a portion of a pedicle that contains both arterial and venous branches. The vessels are supported by a connective tissue stroma principally composed of collagenous fibers which pass between two intrinsic laryngeal muscles (LCA and TA). Branches from pedicles such as this one penetrate the surrounding muscles to form complex networks as shown in Figure 2-3.

Sex Differences

Blood vessels in female arterial pedicles tended to be composed of smaller caliber vessels than in male counterparts. The connective tissue stroma also appeared to be less dense than in males.

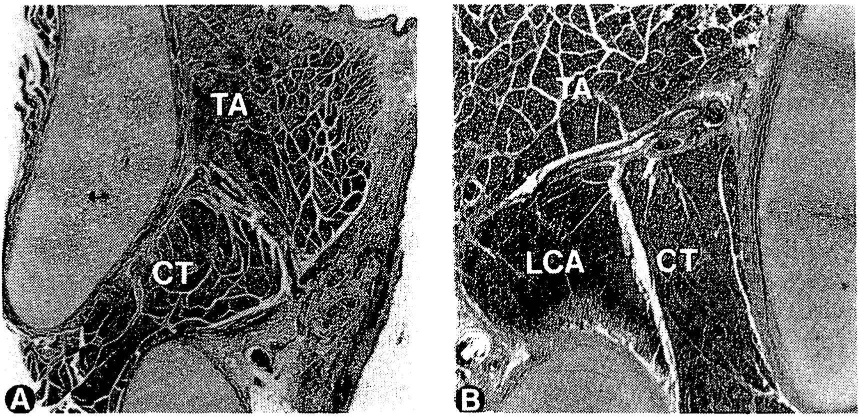


FIGURE 2-1
Micrographs of vascular pedicles from the (A) anterior and (B) middle thirds of the vocal folds. (A) is taken from the right side of the larynx and shows a pedicle that separates the cricothyroid muscle (CT) from the thyroarytenoid muscle (TA). (B) is taken from the left side of the larynx and shows a pedicle that separates CT, TA and lateral cricoarytenoid (LCA) muscles. The space between LCA and CT is due to preparation artifact. See text for discussion. (12X, original magnification. Iron Gallien Elastic Stain).

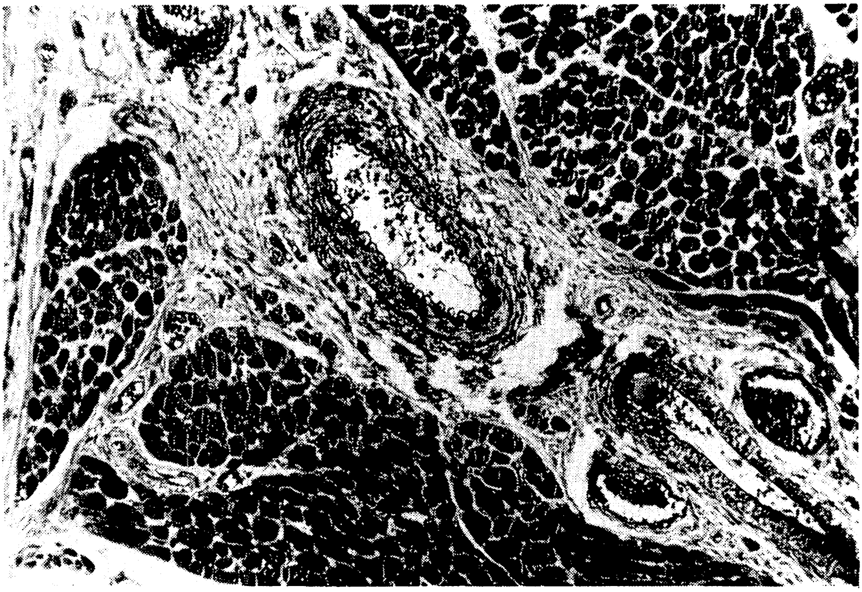


FIGURE 2-2
Micrograph illustrating the anatomy of a typical vascular pedicle from the middle third of the vocal fold. This pedicle separates the thyroarytenoid (above) from lateral cricoarytenoid (below) muscles. One large and two smaller arterial branches are shown, along with a branch of vein which is sectioned longitudinally. The vessels are supported by connective tissue. Note that spaces in the connective tissue surrounding the largest blood vessel are preparation artifacts. (33X, original magnificaiton, Iron Gallien Elastic Stain).

Regional Differences

Distinctive differences were found in the origin and distribution of the arterial pedicles in different regions of the larynx. Though readily identifiable in adult histologic sections, these relationships were especially well defined in the infant larynx. The small size and excellent condition of the infant histologic sections made them particularly well suited to illustrate the typical morphology of the arterial arcades and their relationships to the intrinsic laryngeal muscles.

Anterior Commissure. In the anterior commisure (Figure 2-4), the CT muscle is separated from the TA muscle by the main branch of the cricothyroid artery which penetrates the muscle and passes medially beneath the lamina of the thyroid cartilage. It is joined by a medial branch of the artery which perforates the cricothyroid ligament. Both branches of the artery along with supporting connecting tissue intermingle with the perimysium of the respective muscles and serves to separate CT from TA. The lateral border of TA is bounded by vessels from the superior laryngeal artery and vein.

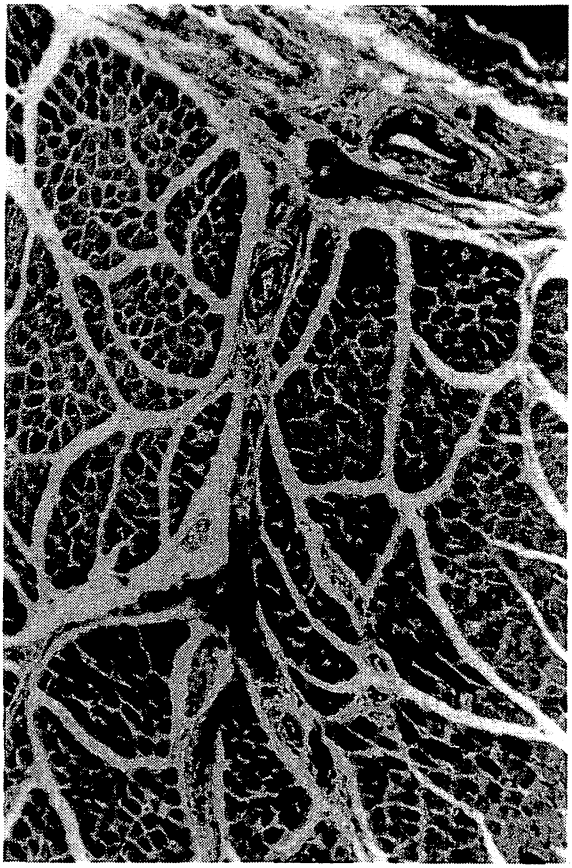


FIGURE 2-3
Micrograph illustrating the distribution of a branch from an arterial arcade into the cricothyroid muscle. Fibers from the external branch of the superior laryngeal nerve which accompanied this vessel can be seen within the pedicle and towards the bottom of the micrograph. (33X, original magnification, Safranin-Phloxine-Hematoxylin Stain).

Anterior Region of Vocal Fold. In the anterior region of the vocal fold (Figure 2-5), the CT and TA muscles appear to be separated by a vascular pedicle formed through a joining of the cricothyroid artery and a longitudinally coursing branch of the inferior laryngeal

FIGURE 2-4

Coronal section from a human infant larynx (7x magnification) from the anterior commissure illustrating the typical relationship of the arterial arcades and pedicles to the intrinsic laryngeal muscles. 1 = main branch of cricothyroid artery, 2 = medial branch of cricothyroid artery (enters lateral to cricothyroid membrane), 3 = conus elasticus, connective tissue and glands in anterior commissure, 4 = vessels from superior arterial arcade, 5 = nodular portion of vocal ligament, 6 = ventricle of Morgagni, T = thyroid cartilage, C = cricoid cartilage, TA = thyroarytenoid muscle, CT = cricothyroid muscle, Tr = trachea.

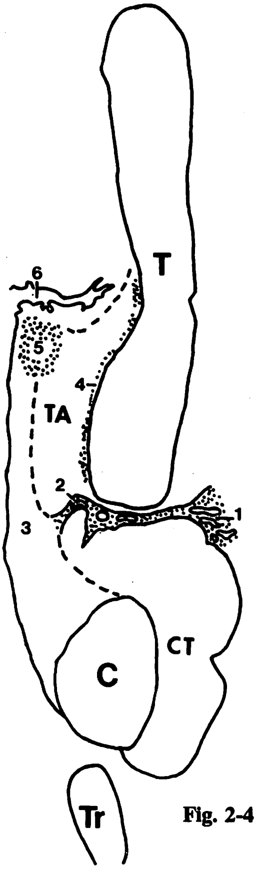


Fig. 2-4

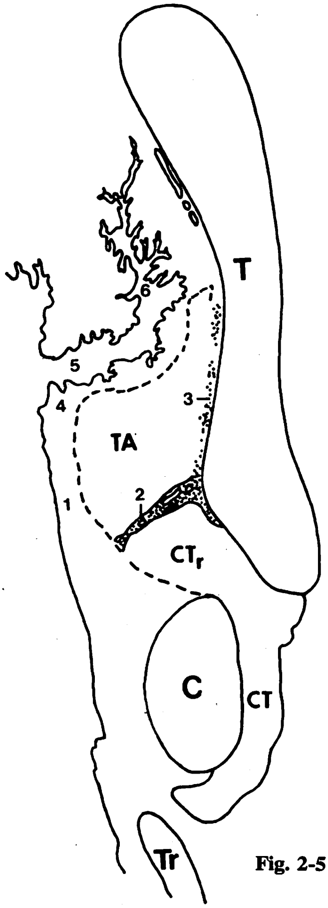


Fig. 2-5

FIGURE 2-5

Coronal section of the human infant larynx (7x magnification) from the anterior third of the vocal fold illustrating the typical relationship of the arterial arcades and pedicles to the intrinsic laryngeal muscles. 1 = conus elasticus, 2 = branches from transverse and cricothyroid arteries, 3 = branches of vessels from superior laryngeal artery and vein (anterior arcade), 4 = lamina propria of vocal fold, 5 = ventricle of Morgagni, 6 = sacculus, T = thyroid cartilage, C = cricoid cartilage, TA = thyroarytenoid muscle, CT_r = pars rectus portion of cricothyroid muscle, CT = cricothyroid muscle, Tr = trachea.

FIGURE 2-6
Coronal section of the human infant larynx from the middle third of the vocal fold (7x magnification) illustrating the typical relationship of the arterial arcades and pedicles to the intrinsic laryngeal muscles. The configuration of the arterial pedicle in the middle and posterior regions of the vocal folds is similar. 1 - vessels from posterior arcade, 2 = transverse arcade, 3 = medial limb of transverse arcade, 4 = inferior limb of transverse arcade, 5 = conus elasticus, 6 = lamina propria of vocal folds, 7 = ventricle of Morgagni, 8 = vestibular folds, T = thyroid cartilage, C = cricoid cartilage, TA = thyroarytenoid muscle, CT = cricothyroid muscle, Tr = trachea.

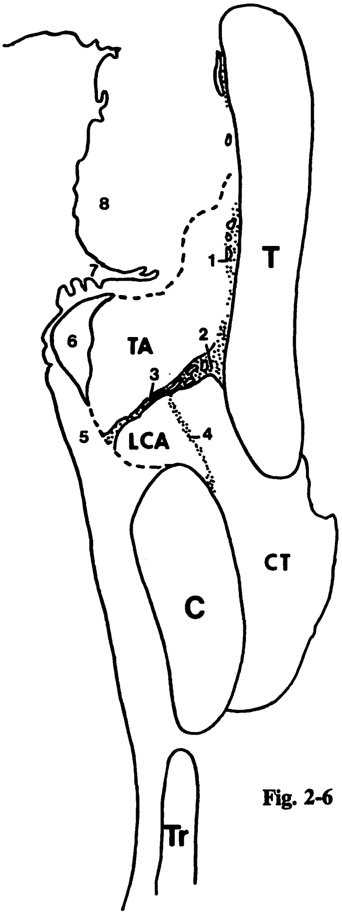


Fig. 2-6

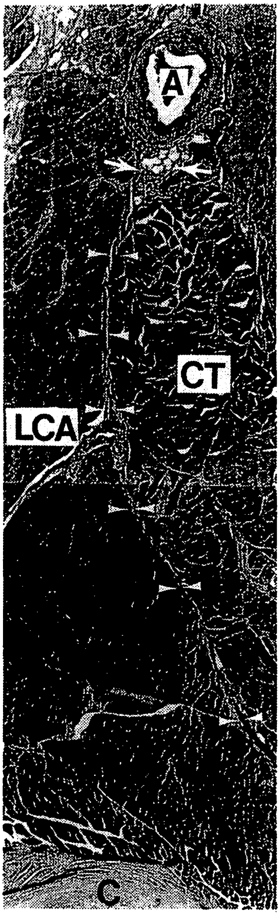


Fig. 2-7

FIGURE 2-7
Micrograph of the inferior limb of the transverse arcade (A) which is composed of a thin connective tissue sheath (arrowheads) separating the lateral cricoarytenoid muscle (LCA) from pars oblique portion of the cricothyroid muscle (CT). Note attachments of LCA muscle fibers to the cricoid cartilage (C). (13x, original magnificaiton, Iron Gallien Elastic Stain).

artery (called the transverse arcade by Terracol and Guerrier). The latter courses along the lateral border of the LCA muscle giving off medially directed branches along its anteroposterior course. The lateral border of the TA and CT muscles continue to be supplied by superior laryngeal blood vessels.

Middle and Posterior Regions of Vocal Fold. In the middle and posterior regions of the vocal folds the vascular pedicles separate the CT, LCA and TA muscles from each other and from the thyroid lamina (Figure 2-6). A thick connective tissue sheath supports the transverse and posterior arcades. This sheath separates the CT from the TA, LCA muscle complex.

The longitudinally coursing branch of the inferior laryngeal artery (transverse arcade) is well defined in these regions of the vocal folds and is accompanied by substantial amounts of supporting connective tissue. The pedicle consists of medial and inferior limbs. The medial limb is the most prominent separating the LCA from TA muscles. It courses toward the conus elasticus, occasionally anastomosing with a more medialy coursing branch of the inferior laryngeal artery which runs anteroposteriorly in an areolar tissue space between the cricoid cartilage and conus elasticus. The medial limb contains small calibre blood vessels and varying amounts of connective tissue. The inferior limb is composed of connective tissue which is thin and delicate (Figure 2-7). It appeared to be routinely damaged in the histologic sectioning of our paraffin embedded specimens (Figure 2-1B) resulting in a plane of cleavage between the CT and LCA muscles. Occasionally, a small arterial branch from the transverse arcade was found in the inferior limb, although this was not typical.

Arytenoid Cartilage. In the arytenoid cartilage (mid portion used for illustration), the transverse arcade, a branch of the inferior laryngeal artery separates LCA from TA muscles (Figure 2-8). The pars oblique portion of CT muscle is separated from LCA by a connective tissue partition which becomes thickened at the upper border of the CT muscle. Blood vessels and branches of the recurrent laryngeal nerve are contained in parts of this septum (Figure 2-9). Thickening of this septum results from a convergence of blood vessels and supporting connective tissue from the transverse and posterior arterial arcades (Figures 2-8 and Figure 2-9).

The lateral border of TA is bounded by vessels from the posterior arcade. It is formed by joining of branches of the superior and inferior laryngeal arteries and veins.

It appears that the upper regions of the larynx derive most of their arterial support from the superior laryngeal arteries and veins. The exception is the interarytenoid area which appears to receive substantial vascular supply from both superior and inferior arteries and veins. The lower region of the larynx and most of the middle of the larynx receive arterial supply from the inferior laryngeal arteries.

Cricothyroid Joint. The inferior laryngeal artery is closely associated with the cricothyroid joint (Figure 2-10). In the posterior aspect of the joint, the inferior laryngeal artery courses deep to PCA muscle fibers which converge on to the muscular process of the arytenoid cartilage. The inferior laryngeal artery passes over the posterior cricothyroid ligaments, giving off a a small branch to the pars oblique portion of CT muscle and then dividing into posterior and anterior branches. The posterior branch is the largest and it courses superiorly and somewhat anteriorly to join the posterior branch of the superior laryngeal artery to form the posterior arcade. The anterior branch, passes over the horizontal fibers of the posterior cricothyroid ligament and courses longitudinally and lateral to the LCA muscle. This forms the transverse arterial arcade, branches of which separate LCA from TA muscles. Vessels in the posterior and transverse arcades are surrounded by areolar tissue and collagenous fibers which form distinct boundaries between the musculature and the thyroid laminae.

FIGURE 2-8
Coronal section from the human infant larynx through the mid-region of the arytenoid cartilage (7x magnification) illustrating the typical relationship of the arterial arcades and pedicles to the intrinsic laryngeal muscles. 1 = posterior arcade, 2 = convergence of posterior and transverse arcade, 3 = medial limb of transverse arcade, 4 = inferior limb of transverse arcade, A = arytenoid cartilage with associated glandular tissue, T = thyroid cartilage, C = cricoid cartilage, TA = thyroarytenoid muscle, LCA = lateral cricoarytenoid muscle, CT_0 = pars oblique portion of cricothyroid muscle.

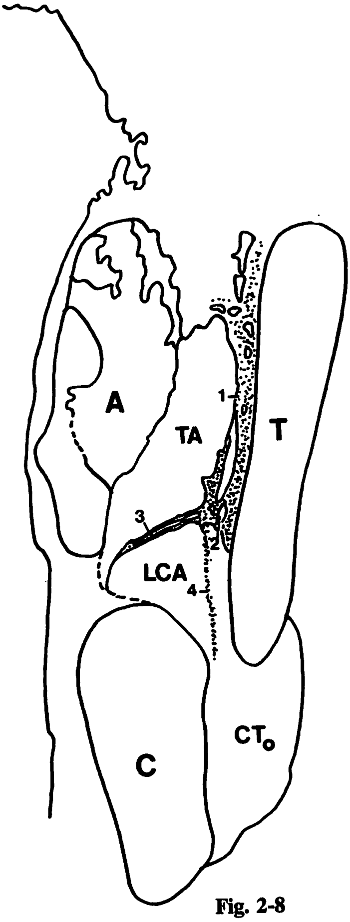


Fig. 2-8

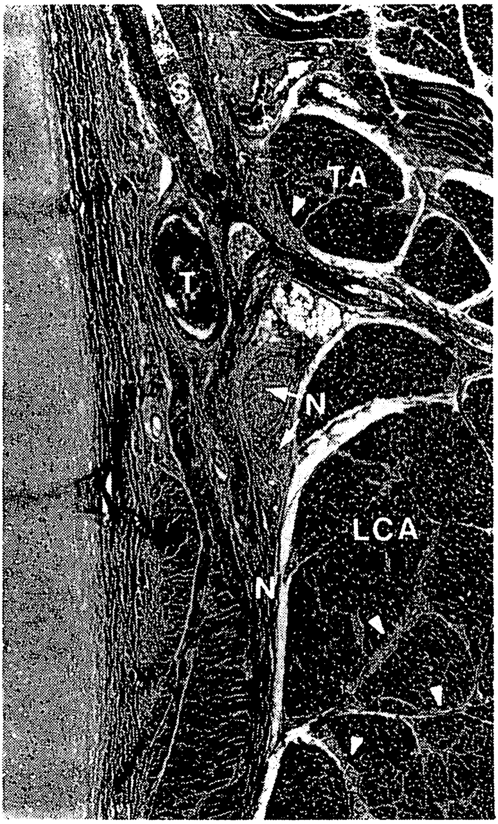


Fig. 2-9

FIGURE 2-9
Micrograph from the mid arytenoid cartilage region which shows the relationship between a portion of the recurrent laryngeal nerve (N) and the superior (S) and transverse (T) arterial arcades. One of the pedicles from the transverse arcade is separating thyroarytenoid (TA) from lateral cricoarytenoid (LCA) muscles. Note branches (arrowheads) of the nerve into LCA and TA. (33x, original magnification, Iron Gallien Elastic Stain).

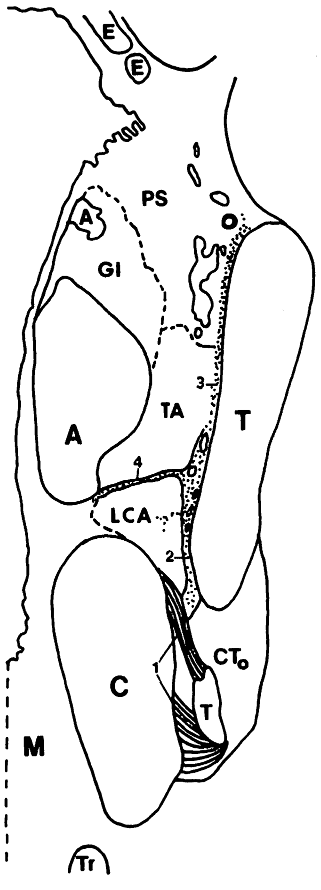


FIGURE 2-10
Coronal section from a human infant larynx from the region of the cricothyroid joint (7x magnification) illustrating the typical relationship of the arterial arcades and pedicles to the intrinsic laryngeal muscles. 1 = ligaments of cricothyroid joint, 2 = inferior laryngeal artery, 3 = posterior arcade, 4 = branch from transverse arcade, A = arytenoid cartilage, Gl = glands in mucosa of arytenoid cartilage, PS = paraglottal space, E = epiglottis, T = thyroid cartilage, C = cricoid cartilage, M = mucosal lining of lumen of cricoid cartilage, TA = thyroarytenoid muscle, LCA = lateral criocoarytenoid muscle, CT$_0$ = pars oblique portion of cricothyroid muscle, Tr = trachea.

DISCUSSION

Data from the present study substantiate the arterial pattern within the larynx described as "arcades" by Terracol and Guerrier (1951). Results from this study extend these authors findings to show how branches of the arcades relate to the cricothyroid (CT), thyroarytenoid (TA) and lateral cricoarytenoid (LCA) muscles.

The use of the arterial arcades and their major branches (referred to as pedicles in this paper) appears to be a feasible method for determining histologic boundaries between the CT, TA and LCA muscles. These relationships are clearly visible using low power objectives (20x and lower). It should be noted that in determining the arterial boundaries between muscles, it is essential to verify the origins of the muscle fibers being separated or compartmented. This is particularly the case when the fibers from two or more muscles are closely approximated (e.g., lateral fibers of LCA from pars oblique fibers of CT). Such verification was found to be reliably made at magnifications of 20-40x.

The extra- and intramuscular courses of the arterial arcades are quite consistent, while those of the accompanying veins exhibit considerably more variability. The arterial pedicles were thickest in the middle and posterior portions of the larynx. This appears to result from larger calibre vessels and increases in areolar and dense connective tissues which accompany them.

Branches of the inferior (recurrent) laryngeal nerve closely follow the distribution of the posterior and transverse arterial arcades. They appear to exhibit greater complexity within the muscle than the blood vessels.

The CT muscle receives extensive blood supply from the cricothyroid artery (derived from the superior laryngeal artery) and also from the inferior laryngeal artery through branches from the posterior and transverse arcades. This arrangement appears to insure an adequate blood supply to this large, physiologically important and active muscle.

More information is needed about the distribution of blood vessels within the intrinsic laryngeal muscles. Such data will enhance our understanding about the complex morphology of these muscles and also provide a detailed "map" of the vascular plan within them. Such data may be useful in future research. It might become feasible to use radioisotopes to track and localize specific regions of activity within the musculature of the larynx during phonation as has recently been done in the studies of cortical activity (Lassen *et al*, 1978). Detailed information about the vascular supply to the intrinsic laryngeal muscles may also be clinically useful in helping to understand the spread of disease within the larynx.

ACKNOWLEDGEMENTS

Appreciation is expressed to Alice Kahn for her carefully drawn tracings of histologic sections. This research was supported in part by a Research Grant #049 from The Voice Foundation.

REFERENCES

Bowden, R. and Scheure, J.L. (1960). "Weights of abductor and adductor muscles of the human larynx," *J. Laryngol. and Otol.* 74, 971-980.

Lassen, N.A., Ingvar, D.H. and Skinhoj, (1978). "Brain function and blood flow," *Scientific Am.* 239/4, 62-71.

Terracol, J. and Guerrier, Y. (1951). "Les systeme arteriel du larynx. Etude anatomique," *Bull. des la societe-d'Oto-rhino-laryngologie des Lyon et de la Region* 2/4, 57-81.

Warwick, R. and Williams, P.L. (1973). *Gray's Anatomy* (35th British edition). W.B. Saunders, Philidelphia.

DISCUSSION

K. HARRIS: Could you comment again on the relationship between vascular arcades and nerve fibers?

J. KAHANE: The nerves (superior and inferior laryngeal nerves) travel along with respective arteries and veins, usually posterior and medial to them. We have observed that the major branches of the nerves follow the main trunks of the arterial arcades but exhibit more diffuse and complex branches within the muscles they innervate than the artery.

M. HIRANO: Some of your slides frighten me because we have conducted several hundred EMG studies of VOC in which we insert a needle through the junction between CT and VOC where those vessels shown in your slides exist. Fortunately, we have had problems very rarely. Could you comment on the possibility of injuring the vessels during the EMG procedure?

J. KAHANE: This is an interesting and relevant point. It is an area which I have not given a great deal of thought. It seems to me that there is natural protection afforded to the vessels in several ways. They are small caliber vessels and are surrounded by areolar tissues and in some places also by significant amounts of fibrous connective tissues. Although the major features of the the morphology of the arcades appear to fit into definable subtypes, the representation of each aspect of this vascular plan within a single muscle or region of the larynx is likely to vary. Thus the inherent variation in the basic plan of the arcade affords some insurance against injury. Finally, if memory serves me correctly, the placement of your electrodes (Hirano and Ohala, *J. Speech Hear. Res.* 1969, 12/2, 362-373) is more likely to be in the neighborhood of the smaller branches supplying the muscle than the larger branch of the arcade. They are probably more frequently passing through a network of small arerioles and capillaries than approaching the main trunk of the arcade. Your question deserves serious study in the future.

3 ▪ VIBRATORY BEHAVIOR VERSUS THE STRUCTURE OF THE VOCAL FOLD

Minoru Hirano, Koichi Matsuo, Yuki Kakita, Hiroshi Kawasaki and Shigejiro Kurita

INTRODUCTION

Over the past four decades, analyses of ultra high speed motion pictures of the vocal folds have greatly increased the information of the vibratory behavior of the vocal folds and the mechanism of vocal fold vibration. In many studies employing ultra high speed photography, trajectories of two selected points facing each other on the edges of the bilateral vocal folds were investigated. These points were usually located at the place where the amplitude of lateral excursion was the greatest. However, the exact location of the point of the greatest lateral excursion has not been determined. It has been presumed to be located at the anteroposterior midpoint of the membranous portion of the vocal fold. For example, it is written in many publications that vocal fold nodules and polyps usually occur at the middle of the membranous portion of the vocal fold because the mechanical stress during vibration is the greatest at this location.

Moore and von Leden (1958) determined the lateral excursion patterns of four selected points along the edge of the vocal fold during inspiratory phonation and demonstrated significant differences in vibratory behavior among these four points. In addition, Tanabe *et al* (1975) illustrated lateral excursion patterns of five selected points on the edge of the vocal fold during normal phonation at varying fundamental frequencies and intensities, demonstrating marked differences in phase and amplitude between these points.

Moreover, Hirano and his co-workers reported variations in the layer structure of the human vocal folds along the anteroposterior dimension (Hirano, 1975 and 1981; Hirano, Kurita and Nakashima, 1981; Kurita, 1980).

The purpose of this paper is to investigate how the variations in the layer structure along the length of the vocal fold are related to the vibratory behavior.

MORPHOLOGICAL STRUCTURE

Before going into the present investigation, an explanation of the morphological structure of the vocal fold will be given for the sake of a better understanding of the present study.

Figure 3-1 presents a histological picture of a horizontal section of a human vocal fold. As shown in this figure, the structure of the vocal fold is not uniform along the length. Since the details have been reported elsewhere (Hirano, 1981; Hirano, Kurita & Nakashima, 1981; Kurita, 1980), only a brief explanation will be given here.

At the anterior commissure of the vocal folds, there is a mass of collagenous fibers which is connected anteriorly to the inner perichondrium of the thyroid cartilage and posteriorly to the lateral portion of the vocal ligament. This mass is called the anterior commissure tendon. Posterior to this mass is the anterior macula flava, a mass of elastic fibers. This is a continuation of the medial portion of the vocal ligament. Thus, there are gradual changes in stiffness of the structure from the hard thyroid cartilage to the pliable vocal fold mucosa. At the posterior end of the membranous vocal fold, there is a mass of elastic fibers called the posterior macula flava. It is a continuation of the medial portion of the vocal ligament and is connected to the vocal process of the arytenoid cartilage posteriorly. Again, there are gradual changes in stiffness from the firm arytenoid cartilage to the pliant vocal fold mucosa.

The length of the anterior and the posterior macula flava is approximately 1.5 mm.

Figure 3-2 presents frontal sections of an adult male vocal fold at the five places shown in Figure 3-1. Figure 3-3 shows thickness measures of three layers, i.e. the cover (I), the medial portion of the vocal ligament (II), and the lateral portion of the ligament (III). The data were obtained from the larynges of 5 normal male adults in their thirties. Each line in Figure 3-3 shows the values of each individual. Important findings in these figures are the following:

1. The loose and pliable cover is the thickest at the middle (C) of the membranous portion, becoming thinner towards the anterior and the posterior ends. It tends to be thinner in the anterior portion (A, B) than in the posterior portion (E, D).
2. The intermediate layer of the lamina propria, i.e. the medial portion of the vocal

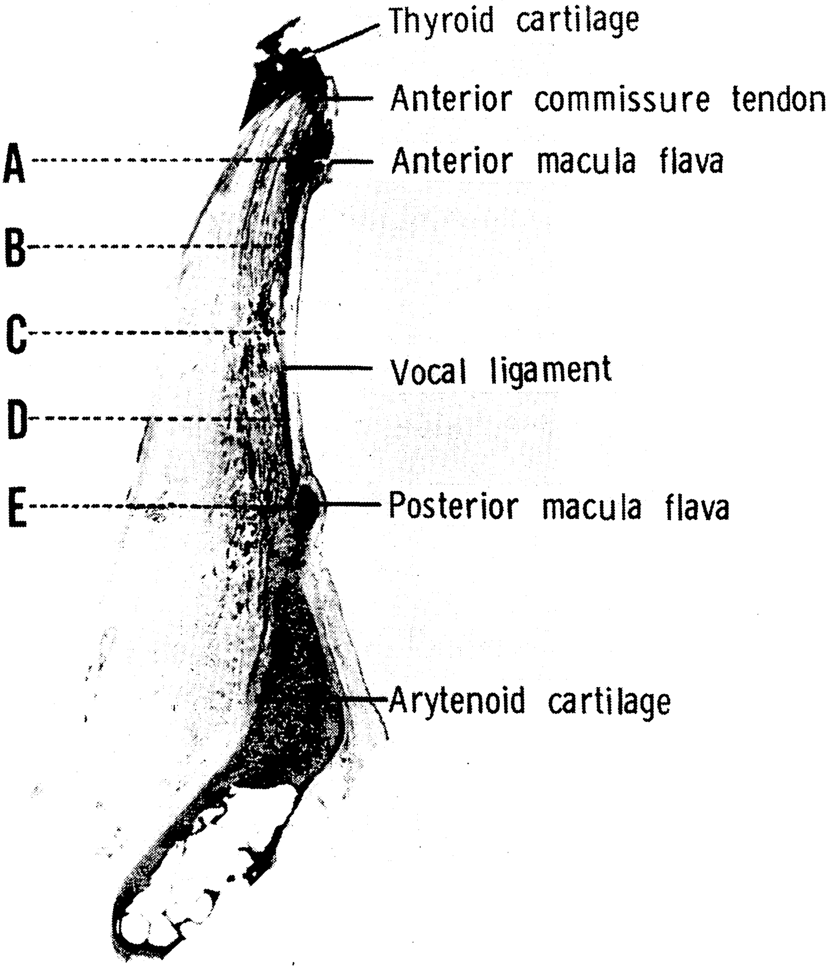


FIGURE 3-1
A histological picture of a horizontal section of a human vocal fold. (from Hirano, 1981).

ligament, consisting primarily of elastic fibers, is thinnest at the midpoint (C), becoming thicker towards both ends. It tends to be thicker in the anterior portion (A, B) than in the posterior portion (E, D).

3. The deep layer of the lamina propria, i.e. the lateral part of the vocal ligament, consisting chiefly of collagenous fibers, is thickest at the posterior end, becoming thinner towards the anterior portion.
4. The location of the vocal ligament relative to the upper surface of the vocal fold becomes lower as the posterior end is approached.

METHOD

Subject and Photography

A 39-year-old normal male, who had no voice training, served as the subject. His vocal folds were photographed under indirect laryngoscopy at speeds of approximately 3500 frames

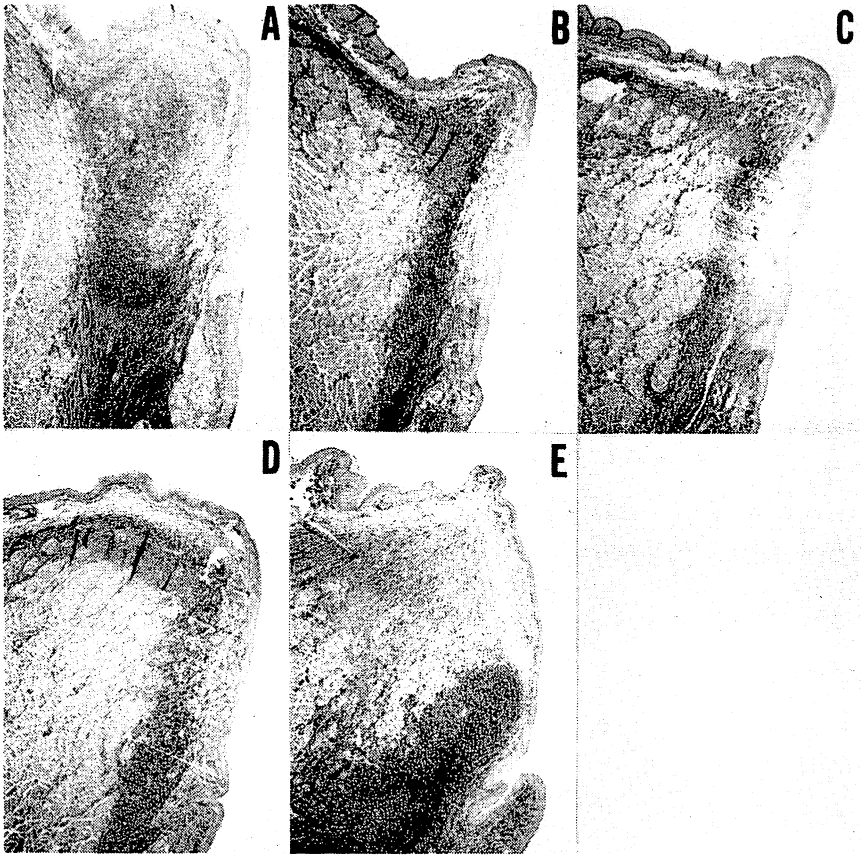


FIGURE 3-2
Frontal sections of the vocal fold of a 22-year-old male at five different places along the longitudinal axis which are shown in Figure 3-1. A: At the anterior macula flava. B: At the middle of A and C. C: At the middle of the membranous portion. D: At the middle of C and E. E: At the posterior macula flava. (from Kurita, 1980).

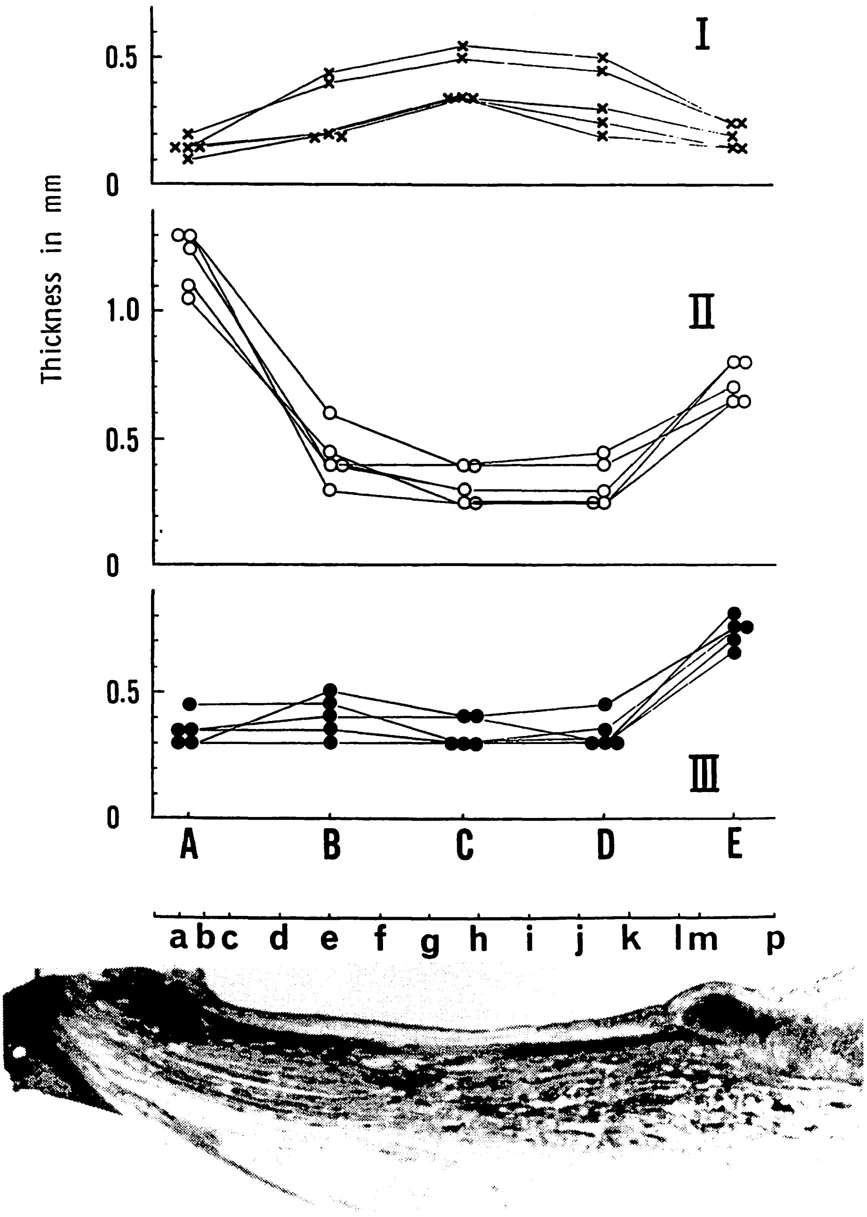


FIGURE 3-3
Thickness of the cover (I), the medial portion of the vocal ligament (II) and the lateral portion of the vocal ligament (III). The values of five males in their thirties are presented. The marks a-p indicate the locations equivalent to those whose vibratory behavior was investigated in this study.

per second. Since the entire length of the vocal folds were not exposed on one film, we photographed the anterior and the posterior portion of the vocal folds separately on two films. The subject phonated at comfortable pitch and loudness level during filming. The apparatus we used for photography has been described elsewhere (Hirano *et al*, 1974).

Reference Scale for Absolute Value Determination

In order to obtain absolute values in measuring various parameters, the following procedures were performed. The subject's vocal folds were photographed via a laryngoscope while phonating at the same pitch and loudness as those during the ultra high speed photography. A piece of graph paper was then photographed while placed at the same distance from the laryngoscope as that of the vocal folds during the filming. Four photographs were therefore obtained: two ultra high speed photographs showing the anterior and the posterior portion of the vocal folds, a laryngoscopic photograph of the entire vocal folds, and a laryngoscopic photograph of a portion of a sheet of graph paper. The photograph of the graph paper was used as the reference scale. Two blood vessels having two bends, one on the right vocal fold and the other on the left, were clearly observed on the ultra high speed film of the posterior portion of the vocal folds. The vessel on the right vocal fold was clearly visualized also on the laryngoscopic photograph and the vessel on the left was found on the ultra high speed film of the anterior portion of the vocal folds (Figure 3-4). Therefore, the absolute value of the distance between the bends of the two vessels could be determined, and this measure was used as the reference scale for the anteroposterior dimension on the ultra high speed films.

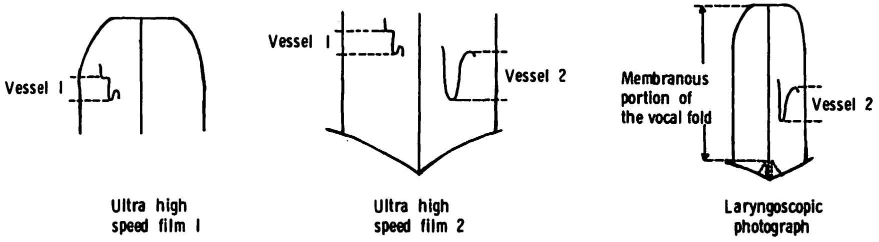


FIGURE 3-4
Schematic presentations of two vessels used as the reference scale.

In order to apply the reference scale correctly for the mediolateral dimension, we measured the angle of the light to the upper surface of the vocal folds while the subject was positioned in the same posture used during ultra high speed photography. The angle was determined by means of X-ray. It was presumed to be a right angle in the frontal plane whereas it measured 60.5° in the sagittal plane. Therefore, there exists the following relation:

$$\begin{aligned} Y &= X \cos (90^\circ - 60.5^\circ) \\ &= 0.87X \end{aligned}$$

where Y is the anteroposterior dimension and X is the mediolateral dimension on the film for a given actual length.

This relationship was applied in determining absolute values of length in the mediolateral direction.

Selection of Points for Frame-by-Frame Analysis

Thirteen points on the edge of each vocal fold were selected for frame-by-frame analysis as shown in Figure 3-5. Points a and b were located on the anterior macula flava. The average length of the anterior macula flava obtained from our separate histological investigation was approximately 1.5 mm as described earlier. We applied this value to the present investigation. Point c was located at the posterior end of the anterior macula flava. Points d-l were placed at millimeter intervals from Point c. Point m was located at the anterior end of the posterior macula flava. The posterior portion of the posterior macula flava and the entire cartilaginous portion of the vocal folds were not visualized on the ultra high speed film.

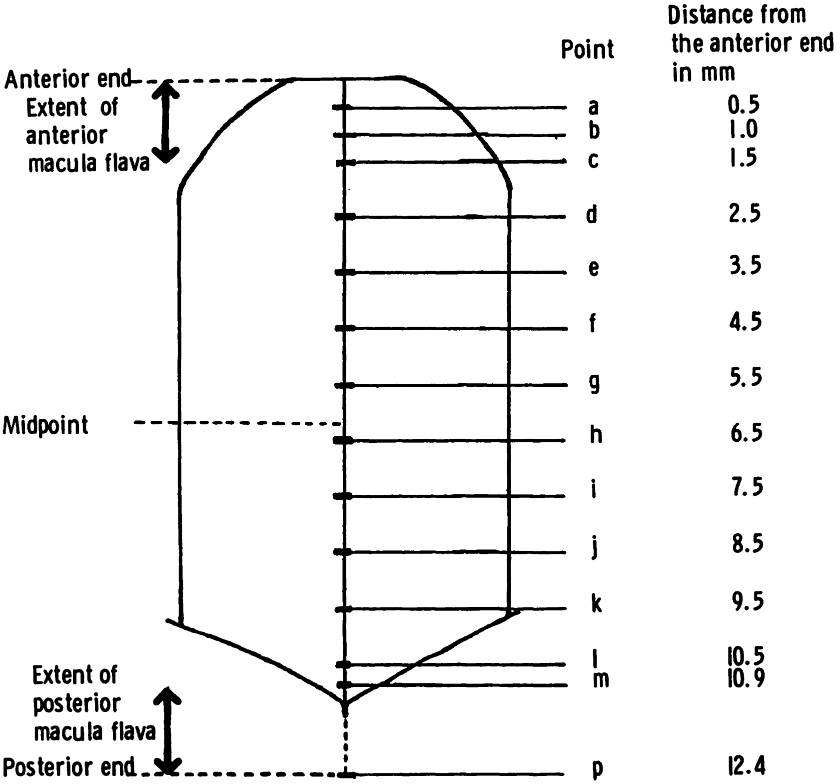


FIGURE 3-5

A schematic presentation demonstrating the location of the thirteen points selected for frame-by-frame analyses of vibratory behavior.

Points a-f were included in the film (Film 1) for the anterior portion and Points f-m in the film (Film 2) for the posterior portion of the vocal folds. Thus we had Point f on both films.

One vibratory cycle was selected for frame-by-frame analysis. For frame-by-frame analysis, we used a film motion analyzer (Vanguard Model 16S).

RESULTS

Comparison of the Two Films

The subject attempted to phonate in the same manner for the two filmings. In order to examine if the vibratory pattern of the vocal folds was exactly the same for the two filmings, we compared the vibratory pattern of Point f on the two films. For the sake of differentiation, we call this point Point f in Film 1 and Point f' in Film 2. As shown in Figure 3-6, the vibratory pattern of Point f and that of Point f' were similar but not exactly the same. Therefore, the manner of phonation should be regarded to be slightly different for the two filmings.

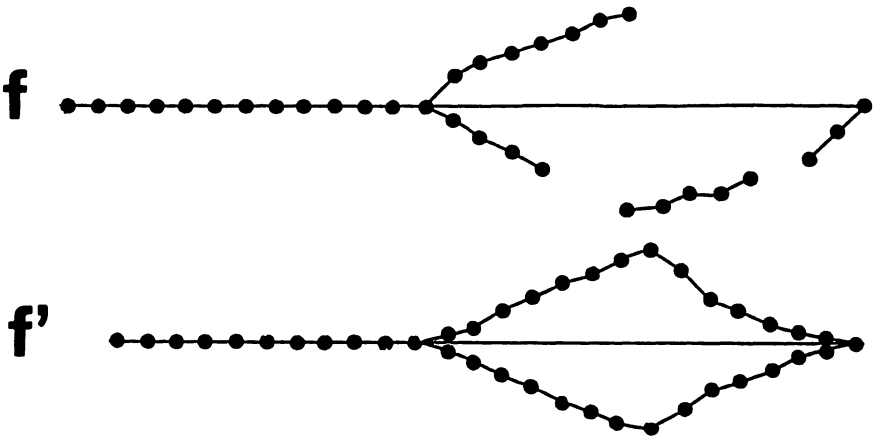


FIGURE 3-6
Lateral excursion pattern of the vocal fold edges at Points f and f'.

General Vibratory Pattern

Figure 3-7 presents patterns of lateral excursion of the 13 points and Figure 3-8 shows glottal width curves at each point.

In Film 1 the acoustic signal was superimposed on the image of the vocal folds. The location of some of the points to be investigated was difficult to determine in some frames when the acoustic signal obscured them. The blanks in Figure 3-7A represent these points. When it was not possible to determine the point on one vocal fold, the glottal width for that point was obtained by doubling the distance from the midsagittal plane to the point on the other vocal fold.

Table 3-1 presents maximum amplitude, duration of the closed, opening and closing phases, speed of opening and closing, open quotient, speed quotient and speed index of the points investigated. The maximum amplitude and speed shown in this table are the average of those for the points on the bilateral vocal folds.

The fundamental period of phonation was 7.70 msec (Fo = 130Hz) for Film 1 and 7.41 msec (Fo = 135Hz) for Film 2.

Phase Differences in Vibration Among the Points Along the Length

As shown in Figures 3-7 and 3-8, the glottis started opening at its posterior half (Points h-m). The opening movement proceeded anteriorly up to Point c, the posterior end of the anterior macula flava. The glottis did not open at all over a large portion of the anterior macula flava (Points a and b).

The maximum opening occurred first at Point c in Film 1 and at Point m in Film 2. It took place lastly at Point f in both films. Point f is not the midpoint of the membranous portion of the vocal fold but is located near the junction between the anterior and middle thirds of the membranous portion.

The glottis was closed first at Point c. At all the other points, the glottis was closed simultaneously.

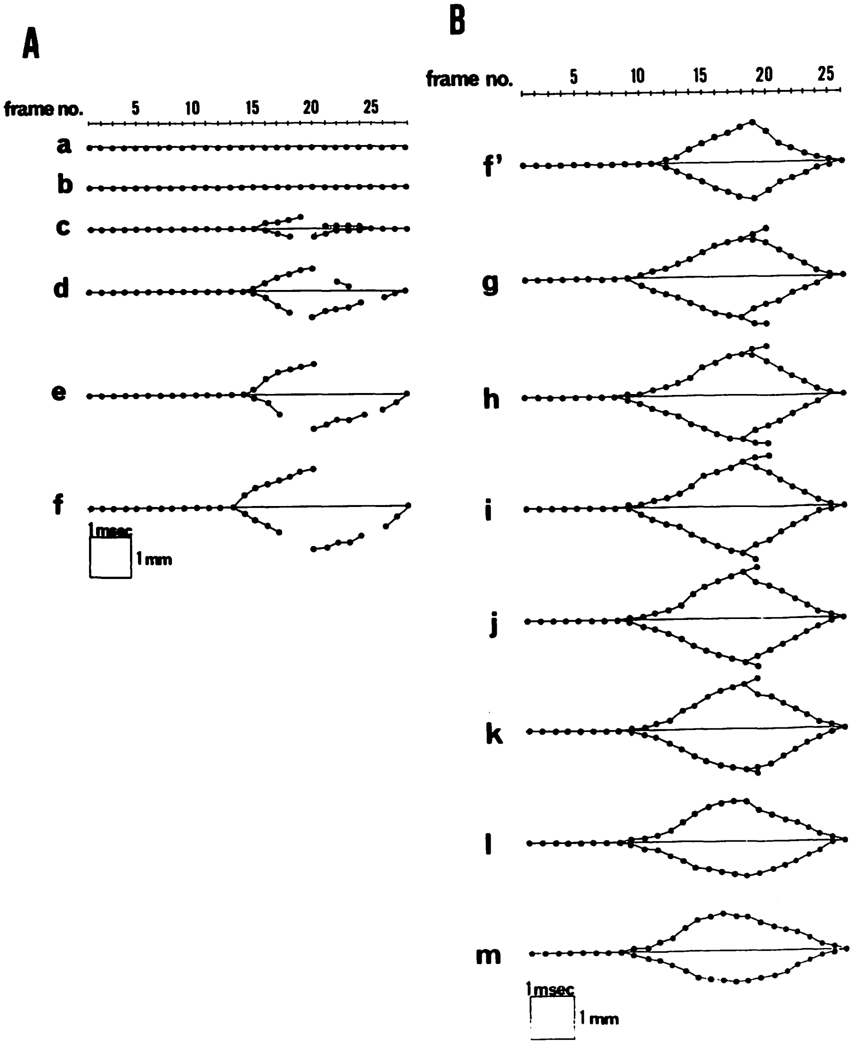


FIGURE 3-7
Lateral excursion pattern of the vocal fold edges at the thirteen points investigated.

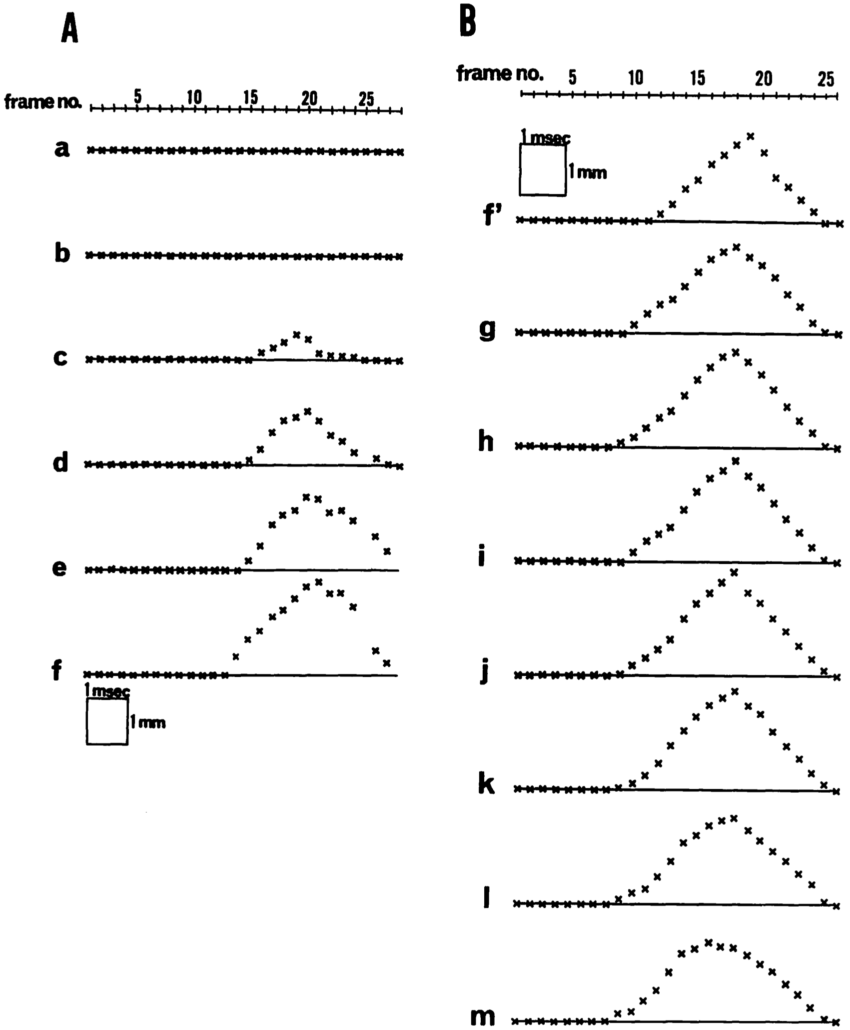


FIGURE 3-8
Glottal width curves at the thirteen points investigated.

Maximum Amplitude

Figure 3-9 depicts the maximum amplitude for each point shown as a function of the distance from the anterior end of the vocal fold. In order to compare the amplitude between the points photographed on Film 1 (a-f) and those on Film 2 (f-m), values of the amplitude measured on Film 1 were normalized by the ratio of the amplitude for Point f′ to that for Point f. The normalized values are connected with the dotted line in Figure 3-9.

The maximum amplitude was the largest at Point j and the second largest at Point i. The difference in amplitude between these two points was very small. These points are not located

TABLE 3-1

Maximum amplitude, duration of the closed, opening and closing phases, speed of opening and closing, open quotient, speed quotient, and speed index of the points investigated.

Point	Maximum amplitude in mm	Duration in m/sec			Speed in m/sec		Open quotient	Speed quotient	Speed index
		Closed phase	Opening phase	Closing phase	Opening	Closing			
a	0	7.70	0	0	0	0	0.00	—	—
b	0	7.70	0	0	0	0	0.00	—	—
c	0.28	4.85	1.14	1.71	0.25	0.16	0.37	0.67	-0.20
d	0.59	3.71	1.71	2.28	0.35	0.26	0.52	0.75	-0.14
e	0.87	3.71	1.71	2.28	0.51	0.38	0.52	0.75	-0.14
f	0.95	3.42	2.00	2.28	0.48	0.42	0.56	0.88	-0.07
f'	0.87	3.13	2.28	2.00	0.38	0.44	0.58	1.14	0.07
g	0.89	2.57	2.57	2.28	0.35	0.39	0.65	1.13	0.06
h	0.96	2.28	2.85	2.28	0.34	0.42	0.69	1.25	0.11
i	1.01	2.28	2.85	2.28	0.35	0.44	0.69	1.25	0.11
j	1.02	2.28	2.85	2.28	0.36	0.45	0.69	1.25	0.11
k	0.97	2.28	2.85	2.28	0.34	0.43	0.69	1.25	0.11
l	0.86	2.28	2.85	2.28	0.30	0.38	0.69	1.25	0.11
m	0.77	2.28	2.57	2.57	0.30	0.30	0.69	1.00	0.00

near the midpoint but rather at approximately the junction between the middle and posterior thirds of the membranous portion.

Speed of Lateral Excursion

Figure 3-10 presents the average speed of lateral excursion for each point as a function of the distance from the anterior end of the vocal fold. In Film 1, the opening speed was faster than the closing speed at all the points investigated. In Film 2, however, the closing speed was faster than the opening speed at all the points except for Point m for which the speed was the same for opening and closing. This indicates that the manner of phonation was not the same for the two filmings as described earlier. Therefore, we are not able to compare the speed between the two films directly. Nevertheless, we can presume that the speed was faster for a large portion of the middle region (Points e-k) than for the anterior (Points c-d) and posterior (Points l-m) portions and that the speed was faster at the posterior portion than at the anterior portion. The speed appeared to be slightly slower near the midpoint than near places anterior or posterior to it.

Open Quotient, Speed Quotient and Speed Index

Figure 3-11 presents the open quotient (OQ) and the speed index (SI) for each point as a function of the distance from the anterior end of the vocal fold. As shown in this figure and Table 3-1, OQ was greater in the posterior half (Points h-m) of the membranous portion than in the anterior half (Points a-g). In the posterior half, there was no difference in OQ between the varying points. In the anterior half, OQ became smaller as the anterior end was approached.

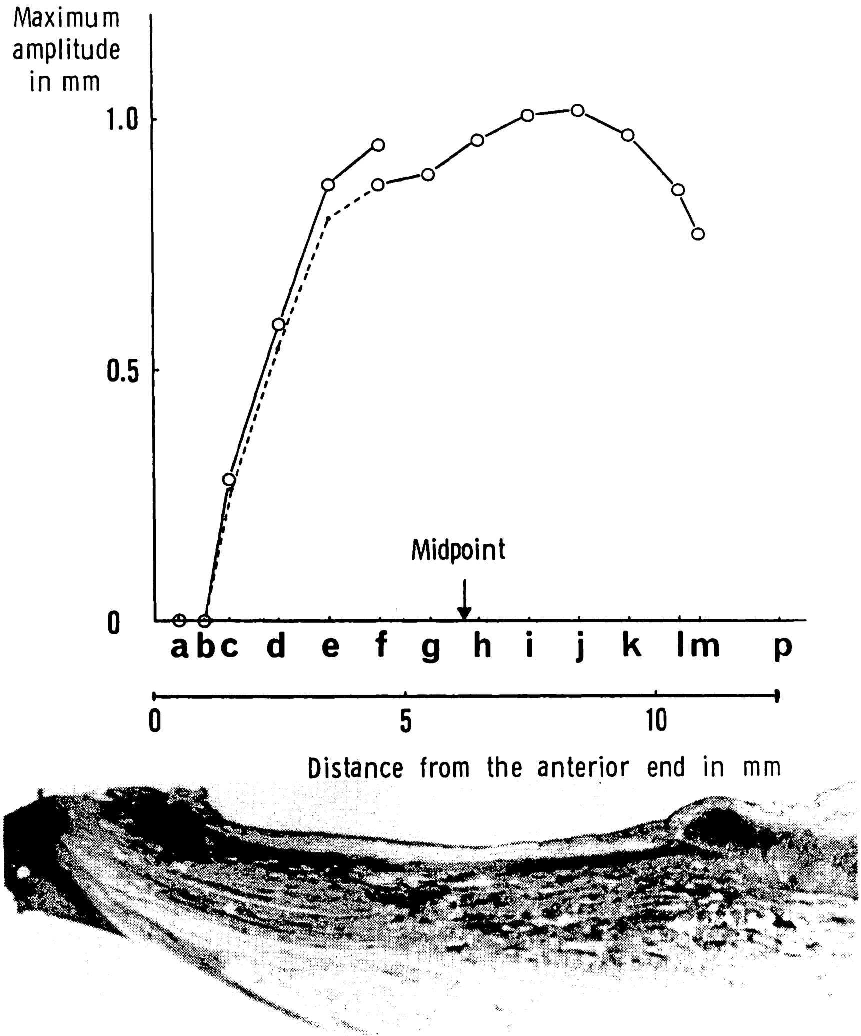


FIGURE 3-9
Maximum amplitude for the thirteen points investigated shown as a function of the distance from the anterior end of the vocal fold.

The speed quotient (SQ) was less than 1 and the speed index (SI) had negative values in Film 1, indicating that the opening speed was faster than the closing speed. SQ and SI became smaller as the anterior end was approached. In Film 2, SQ was greater than 1 and SI showed positive values except for Point m, implying that the opening speed was slower than the closing speed. SQ and SI did not vary throughout Points h-l, i.e. in most of the portion of the posterior half of the membranous vocal fold.

COMMENTS

Histologically, the vocal fold appears to be the most pliable at the area around the midpoint of the membranous portion. The posterior portion of the membranous vocal fold appears to be slightly more pliant than the anterior portion. At the anterior and the posterior macula flava, the tissue appear to be firm. The greatest amplitude of lateral excursion, however, occurred not at the midpoint but at a place posterior to the midpoint. This can be accounted for, at least in part, by the fact that the anterior end of the vocal fold is a definitely fixed point while the posterior end can be movable. The amplitude for the anterior end of the posterior macula flava (Point m) was much larger than that for the posterior end of the anterior macula flava (Point c). This suggests that the entire posterior macula flava and presumably the tip of the vocal process also were moving during phonation.

Tanabe *et al* (1975) showed figures demonstrating the vibratory pattern of 5 points along the vocal fold edge for 12 different phonations. The tip of the vocal process vibrated in 10 of the 12 phonations. One of the two exceptional phonations in which the tip of the vocal

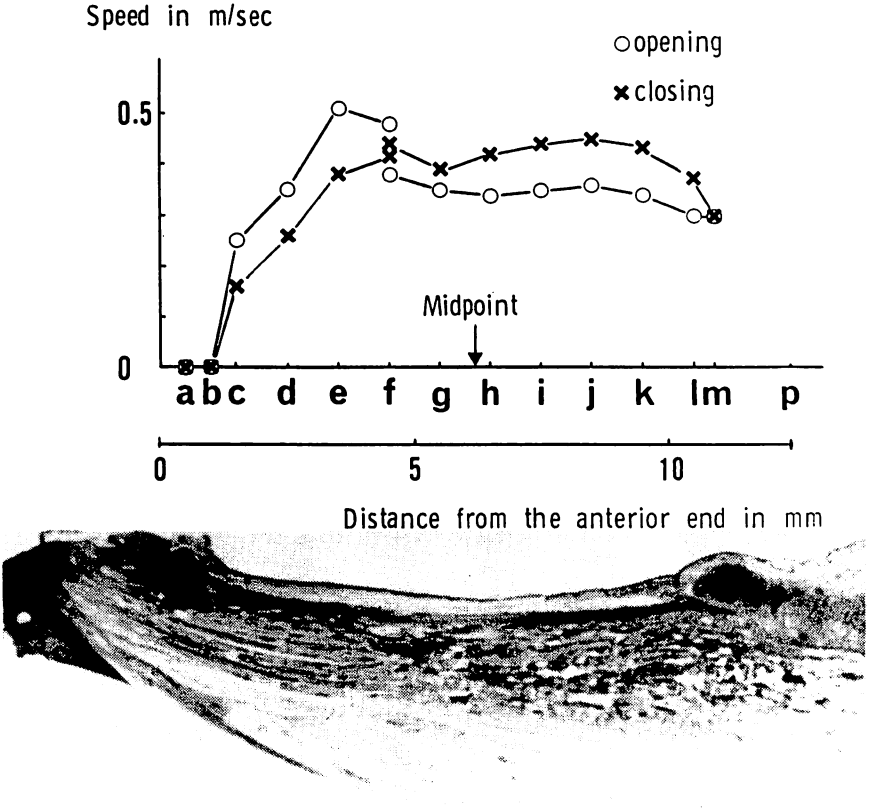


FIGURE 3-10
Average speed of opening and closing for the thirteen points investigated shown as a function of the distance from the anterior end of the vocal fold.

process was fixed was made at the fundamental frequency of 400Hz, which was very high for a male. (The authors did not specify the sex of the subjects. We suppose, however, that the subject was a male because the phonations were at 116Hz, 180Hz, 250Hz and 400Hz in their article). The other exceptional phonation seems to have been an unusual one, because the maximum glottal width was measured to be approximately 0.7 mm and the glottis was not completely closed (Tanabe *et al*, Figure w-12, right). On the basis of the illustrations by Tanabe *et al* and the results of the present investigation, one can say that the tip of the vocal process vibrates in ordinary phonation.

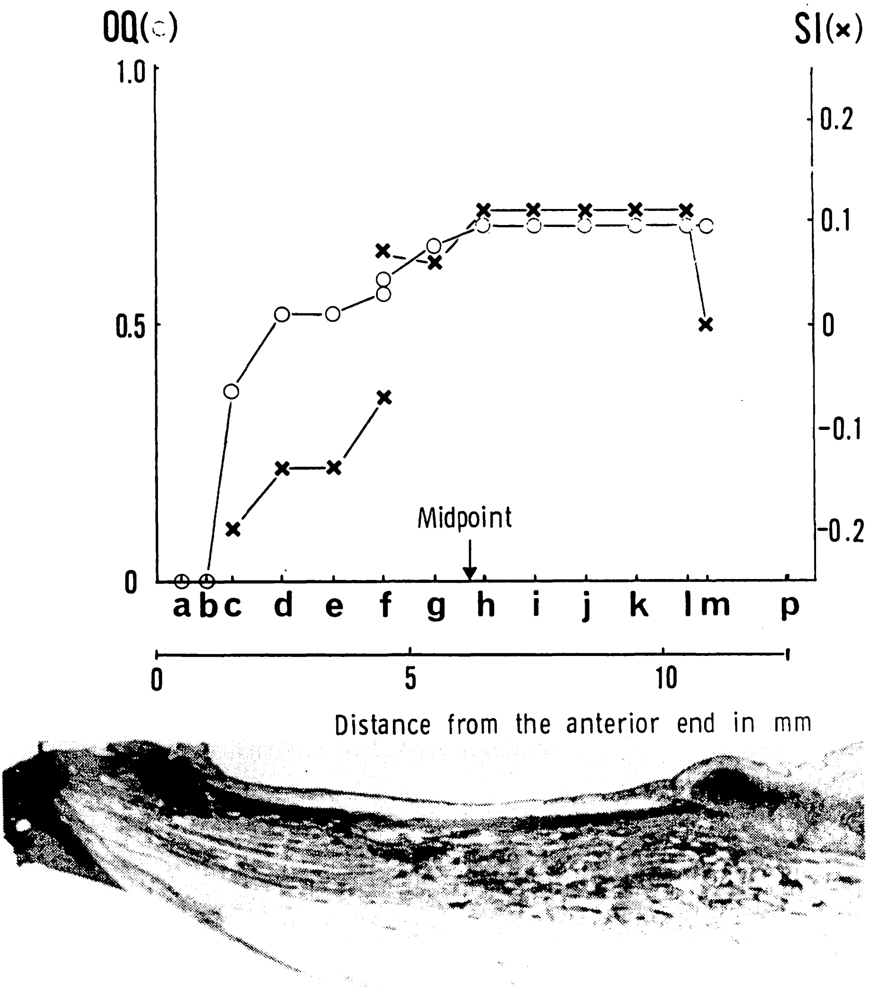


FIGURE 3-11
Open quotient (OQ) and speed index (SI) for the thirteen points investigated shown as a function of the distance from the anterior end of the vocal fold.

The present investigation indicated that the glottis was not opened over a large portion of the anterior macula flava. This does not necessarily imply that the anterior macula flava does not move at all. We rather think that the anterior macula flava is moving in a different manner from the main portion of the membranous vocal fold. The macula flava presumably moves in the manner like that of a head rotating back and forth.

In the present investigation, the glottis started opening at the posterior half of the membranous portion. The opening movement then proceeded forward. This indicates, we think, that the bilateral posterior maculae flavae and the tips of the bilateral vocal processes were not tightly approximated. The maximum opening first took place at the end of the anterior and posterior maculae flavae and was observed lastly near the junction between the anterior and middle thirds of the membranous portion. It is an interesting fact that different portions along the edge of the vocal fold move in inverse directions at the same time. At present, we are not able to explain the mechanism of this phenomenon.

The phase differences in vibratory movements between different points along the vocal fold edge were demonstrated also in the article by Tanabe *et al.* No systematic rules were observed with respect to the occurrence of the phase differences in their study.

It has already been well documented that there are phase differences along the longitudinal dimension. This complex vibratory behavior seems to be mandatory to protect the vocal fold tissue from mechanical shock or damage possibly caused by vibration. Further analysis and some theoretical interpretation of the relationship between the structure of the vocal fold and the vibratory behavior are required.

REFERENCES

Hirano, M. (1975). "Phonosurgery: Basic and clinical investigations," *Otologia* (Fukuoka) 21, 239-440.

Hirano, M. (1981). "Structure of the vocal fold in normal and disease states. Anatomical and physical studies," *ASHA Reports* 11,11-30.

Hirano, M., Kurita, S. and Nakashima, T. (1981). "The structure of the vocal folds," in K.N. Stevens and M. Hirano (Eds.) *Vocal Fold Physiol.*, Tokyo: Univ. of Tokyo Press, 33-41.

Hirano, M., Yoshida, Y., Matsushita, H. and Nakashima, T. (1974). "An apparatus for ultra-high-speed cinematography of the vocal cords," *Ann. Otol. Rhinol. Laryngol.* 83, 12-18.

Kurita, S. (1980). "Layer structure of the human vocal fold. Morphological investigation," *Otologia* (Fukuoka) 26, 973-997.

Moore, P. and von Leden, H. (1958). "Dynamic variations of the vibratory pattern in the normal larynx," *Folia phoniat.* 10, 205-238.

Tanabe, M., Kitajima, K., Gould, W.J. and Lambiase, A. (1975). "Analysis of high-speed motion pictures of the vocal folds," *Folia phoniat.* 27, 77-87.

DISCUSSION

C. LARSON: Is there any difference in the elastic stiffness versus the viscous stiffness as you progress along the length of the vocal folds?

M. HIRANO: We have no actual data on that point. However, Yuki Kakita, one of my colleagues, estimated on the basis of some experimental data, that the roll of viscous stiffness is much smaller than the roll of elastic stiffness. (See *Vocal Fold Physiology.* University of Tokyo Press, 1981).

R. SCHERER: Does the amplitude of the vocal fold motion of the medial surface increase as subglottic pressure increases?

M. HIRANO: Yes. Although we haven't made any quantitative measurements on that, we often see the phenomenon subjectively. Your question is really worthy of some research.

T. BAER: Considering that the most posterior point you measured showed significant vibratory amplitude, could you comment on how much of the arytenoid cartilage participates in vibration?

M. HIRANO: Since the cartilages and the posterior macula flava are stiff and will not flex significantly, you can probably draw a straight line between point "m" and the posterior end of the cartilage, which is fixed to the posterior cricoarytenoid ligament. Considering the length of the arytenoid cartilage and the location of the fixed point at the posterior cricoarytenoid ligament, you can estimate the dimension of the movement of the cartilagenous portion.

G. FANT: Can you describe the phonatory mode of this particular subject? Was it a medium intensity?

M. HIRANO: Medium intensity and pitch, comfortable phonation.

H. HIROSE: Not infrequently, we observe that there is a phase difference in the longitudinal direction in pathological cases under stroboscopy. Since this type of phase difference is observed in the normal vibration patterns that you presented, I am wondering if it is reasonable to interpret this finding as "pathological". What effect do you think the longitudinal phase difference would give to the glottal wave form?

M. HIRANO: Your observation concerning pathological cases is quite true. We have already reported some results on that using ultra high speed photography, but I do not think that the findings shown in our present paper are pathological. I rather think that almost all normal vocal folds present some phase differences between different points along the longitudinal direction, which, however, are often overlooked unless you make a very close observation. I have no idea about the relationship between the longitudinal phase difference and the glottal wave form. Is the glottal wave form the same if the glottal area curve is the same no matter how the longitudinal phase difference is? I don't know.

N. ISSHIKI: I found your presentation extremely interesting from the clinical viewpoint. If I remember correctly, the report by Dr. Tanabe and by Dr. Gould that the vocal process vibrates during phonation was based on male subjects. You suggested that this cartilage vibration may have something to do with a protective mechanism of the vocal cords. Do you suggest that this supposed difference between the male and female may predispose a difference in the frequency of incidence of vocal cord nodules?

M. HIRANO: I have no definite answer. But I found in Tanabe's paper that one of the two exceptions which did not show any movements at the tip of the vocal process was a phonation at extremely high Fo (400 Hz). Another exception was also an unusual phonation because the maximum glottal width was small (0.7mm) and there was no complete glottal closure. We have not analysed any ultra high speed film for female subjects, but I think they also probably show some movements of the cartilage. Anyway, your idea seems to me to be a productive area for further research.

4 ▪ MUSCLE FIBER TYPES IN THE HUMAN POSTERIOR CRICOARYTENOID MUSCLE: A CORRELATED HISTOCHEMICAL AND ULTRASTRUCTURAL MORPHOMETRIC STUDY

Leslie T. Malmgren, Richard Gacek and Carol A. Etzler

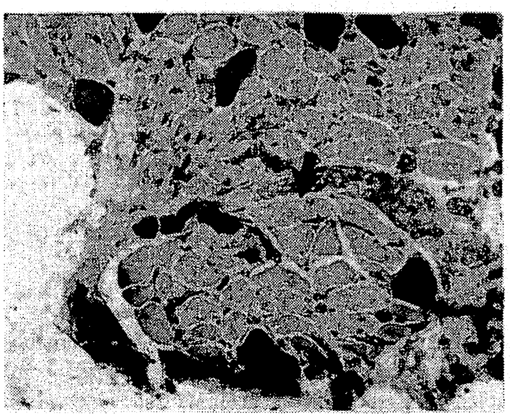
INTRODUCTION

The posterior cricoarytenoid muscle (PCA) differs from most other muscles for which physiological, biochemical and ultrastructural data are available in that it is composed of muscle fibers that have physiological characteristics and enzyme profiles that are not entirely consistent with commonly designated fiber types (Hirose *et al*, 1969; Edström, Lindquist and Martensson, 1974; Malmgren and Gacek, 1981; Hinrichsen and Dulhunty, 1982). Motor unit activity patterns are considered to be a major determinant of muscle fiber type differentiation (Pette *et al*, 1975; Salmons, Gale and Sréter, 1978; Salmons, 1980). Therefore it is likely that the specific activity patterns characteristic of motor units in the PCA lead to the observed differences between the fibers in this muscle as compared to fiber types that have been studied in most other muscles. Although such differences in the biochemical and physiological differentiation of fiber types in the PCA suggest corresponding adaptations in the ultrastructure of these fiber types, there is little data available concerning the ultrastructural characteristics of muscle fiber types in the human PCA. In the present investigation we have determined the ultrastructural characteristics of histochemically-identified muscle fiber types in the human PCA using morphometric techniques.

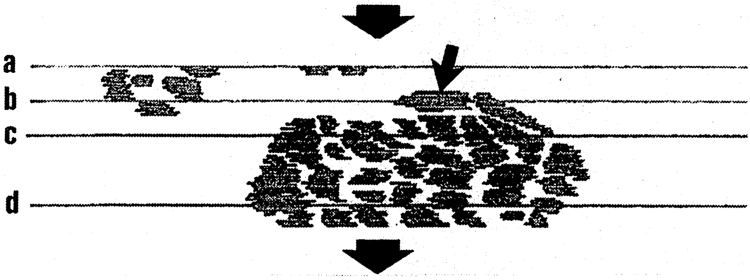
METHODS

Ultrastructural morphometric data was obtained from 29 muscle fibers that had been directly identified as being either type 1 or type 2 on the basis of their relative alkaline actomyosin ATPase histochemical staining densities. This tissue was sampled from the PCA muscle of a 51-year old, non-irradiated, male volunteer who was undergoing laryngectomy for malignancy. The muscle was removed at the time of surgery and immediately frozen by immersion in N-methylbutane cooled to its freezing point with liquid nitrogen. The tissue was then sectioned in a cryostat with some serial sections processed for the histochemical determination of fiber type (ATPase staining according to Guth and Samaha, 1970) and other sections processed essentially according to Eisenberg and Kuda (1977) for correlated electron microscopy. The sections taken for electron microscopy were fixed in cacodylate buffer (0.1M, pH 7.2) 5% glutaraldehyde containing 20mm EGTA [Ethyleneglycol-bis-(B-amino-ethyl ether) N,N'-tetracetic acid]. These sections were then post-fixed with 1% osmium tetroxide in 0.1M cacodylate buffer (pH 7.4), dehydrated in an ethanol series, washed in prophylene oxide and embedded in Spurr's (1969) embedding medium.

The morphometric data were collected prior to the identification of muscle fiber types through the use of the computer reconstruction so that the individuals involved in data collection were blind with respect to fiber type. Point counting techniques (Weibel and Elias, 1a



1b



1c

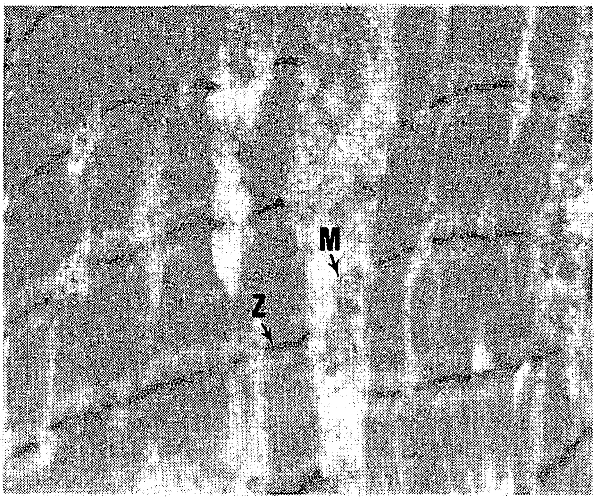


FIGURE 4-1

Example of technique used to permit the direct correlation of the histochemically determined muscle fiber types with data for corresponding ultrastructural morphometric characteristics. Figure 4-1a shows a cryostat section from a human PCA muscle that was processed for the histochemical localization of ATPase. The majority of muscle fibers in this field have little staining for this enzyme (arrow) and can therefore be classified as type 1 (slowly contracting). Some darkly stained type 2 (rapidly contracting) fibers are also present in this field. In order to correlate the histochemically-determined fiber type with the corresponding ultrastructural morphometric data, an adjacent cryostat section was processed for electron microscopy using the techniques of Eisenberg and Kuda (1977). In order to permit measurements of Z-line width, this section was oriented such that the muscle fibers were cut longitudinally rather than transversely as with the cryostat section used for fiber type (ATPase) staining. Serial thick (4μ) and thin sections were then cut on an ultramicrotome, and computer reconstruction (Figure 4-1b) was used to identify the fiber sectioned for electron microscopy. In the computer reconstruction (1b) the numberous short lines indicate the shapes and positions of the muscle fibers which have been reconstructed from points obtained from many serial thick plastic sections. The long lines (a, b, c, d) indicate the planes at which thin sections were taken. The arrow in Figure 4-1b indicates a fiber that is also identified in the histochemical series (arrow, Figure 4-1a). This section "b" (Figure 4-1b) includes a section through the center of the same fiber that was identified in both the histochemical series and in the computer reconstruction (arrows, Figures 4-1a and 4-1b). An electron micrograph of a field in this same fiber is seen in Figure 4-1c. Note z-lines (z) and mitochondria (M).

TABLE 4-1 DIFFERENCES ON SINGLE VARIABLES

The ultrastructural characteristics of type 1 (low alkaline actomyosin ATPase) muscle fibers are compared to those of type 2 (high alkaline actomyosin ATPase) muscle fibers. All data except ZLINE and DIAMETER values are given as volume fractions. A Hotelling's T square indicated a significant difference between type 1 and type 2 fibers for one or more of the variables.

	Type 1		Type 2	
	Mean	S.D.	Mean	S.D.
DIAMETER	28.6μ	10.1	41.5μ	9.1
ZLINE	1145 A°	110	1275 A°	126
MCENTER	0.189	0.085	0.179	0.050
MOUTER	0.184	0.115	0.381	0.173
HOMOG	0.067	0.081	0.033	0.040
LIPID	0.009	0.010	0.002	0.003

Hotelling's T Square = 64.185 (P = 0.002)

1967; Underwood, 1970; Elias, Hennig and Schwarts, 1971) were used to determine mitochondrial volume fraction values for fibers identified by a code unknown to the observer.

Due to the small sample size and the relatively low lipid droplet volume fraction, the lipid droplet volume fraction was measured directly using a compensating polar planimeter rather than through the use of point counting techniques. Mean values for Z-line width (5 measurements/fiber) were also recorded for each fiber.

In order to permit Z-line width measurements, these sections were reoriented so that the muscle fibers were sectioned longitudinally on the ultramicrotome rather than transversely as with the sections used for the correlated fiber type (ATPase) determinations. A computer reconstruction (Figure 4-1) was then used to establish the identity of each fiber examined with respect to the fiber type data for that fiber as indicated by the section processed for ATPase histochemistry. Micrographs used for determinations of mitochondrial and lipid droplet volume fraction values were taken as continuous montages on a JEOL 100S transmission electron microscope at a primary magnification of 5K (photographically enlarged to approximately 12K). Micrographs were also taken of each fiber at 10K primary magnification (enlarged to 24K) for use in Z-line width determinations.

RESULTS

Ultrastructural morphometric data (Table 4-1) was obtained from 29 muscle fibers that had been directly typed (using histochemistry) as type 1 or type 2 on the basis of their actomyosin ATPase activities. The fibers included in this sample were from various regions of the muscle.

Muscle Fiber Size: Our results indicate that the lesser fiber diameters of type 1 fibers in the human PCA were, in general, somewhat smaller than those of type 2 fibers in this muscle (Table 4-1).

Mitochondria: In the central compartment of these muscle fibers (outermost 3 microns excluded) the mitochondrial volume fractions of type 1 fibers were only slightly greater than those of type 2 fibers. However, the mitochondrial volume fractions obtained for the muscle fiber surface compartment (outermost 3 microns only) were approximately twice as high for type 2 fibers as for type 1 fibers. In this outermost compartment both the type 1 and the type 2 fiber samples frequently had volume fraction values above 0.27 which was very uncommon for areas sampled from the central compartment of either cell type.

The outermost annulus was sampled at two separate sites along the perimeter of the fiber. The absolute value of the difference between the mitochondrial volume fractions sampled on the two sides of each fiber (variable HOMOG, Table 4-1) provided an indication of the homogeneity of surface mitochondrial distributions characteristic of each fiber type. A value of 0 would indicate a completely homogeneous distribution of mitochondria between the two surface sites sampled, while larger values for this variable would indicate a more clustered arrangement of mitochondria in the surface compartment. The mean values calculated for type 2 fibers (HOMOG, Table 4-1) indicated that this fiber type had a somewhat more homogeneous distribution of subsarcolemmal mitochondria than was characteristic of type 1 fibers in the PCA.

Lipid Droplets: In the central compartment of fibers in the PCA the lipid droplet volume fraction in type 1 fibers was approximately four times as high as in type 2 fibers.

Z-Line Widths: In general the Z-line widths of type 2 fibers were larger than those of type 1 fibers. When the relationship between the Z-line width and the mitochondrial volume fraction in the central core of the fiber was examined (Figure 4-2), there was only a very

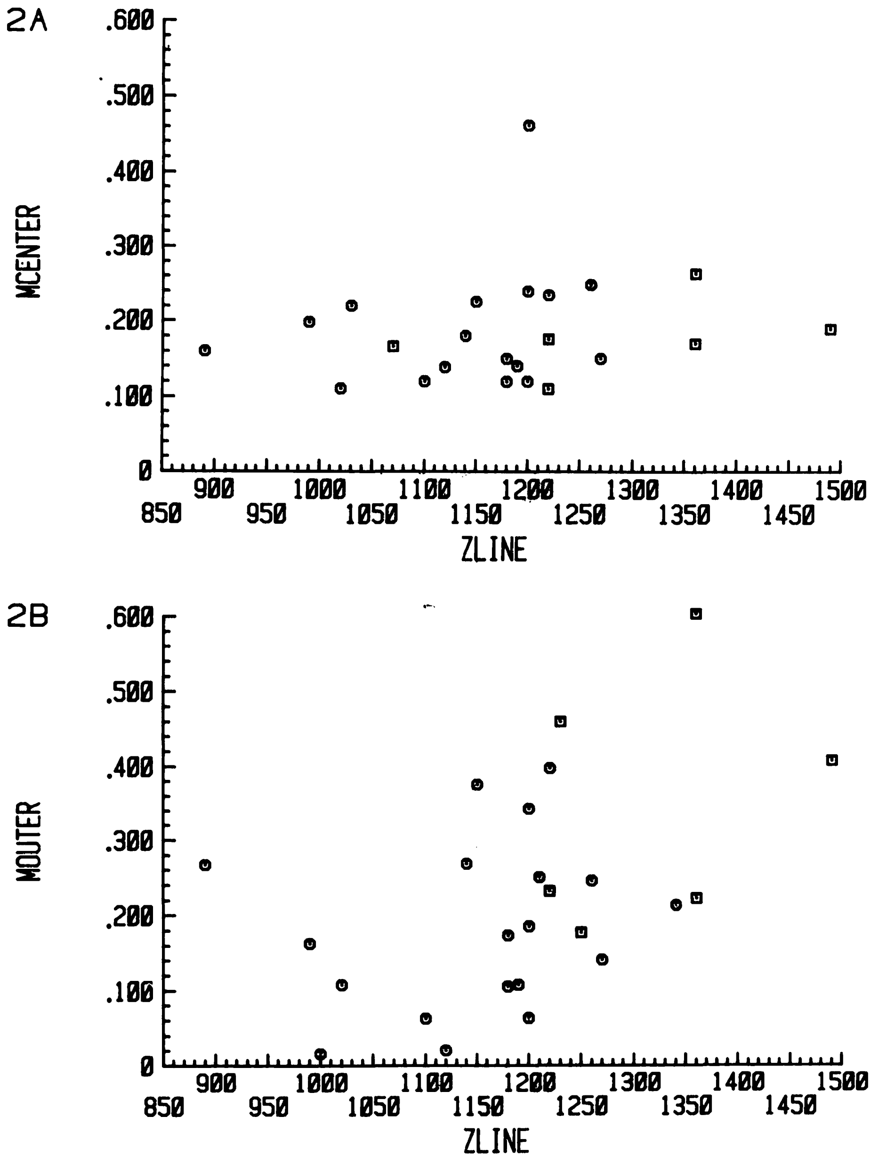


FIGURE 4-2

a. Scatterplot for relationship between the Z-line width (ZLINE) in A^0 and the mitochondrial volume fraction of the central compartment (MCENTER) of type 1 (circles) versus type 2 (squares) muscle fibers in the human PCA.

b. Scatterplot for relationship between the Z-line width (ZLINE) in A^0 and the mitochondrial volume fraction in the surface compartment (outermost 3 microns only) MOUTER of type 1 (circles) versus type 2 (squares) muscle fibers in the same muscle fibers as above.

TABLE 4-2. Single Linkage Cluster Analysis

The following single linkage cluster analysis tree was generated for the variables, MCENTER, ZLINE, DIAMETER, LIPID, MOUTER and HOMOG, using the BMDP1M cluster analysis program. The tree is printed over an absolute correlation matrix, and clustering is defined by the minimum distance method. The MCENTER variable forms a cluster with the variable HOMOG below it is indicated on the tree by the intersection of the dashes beginning above MCENTER with the slashes starting next to HOMOG. The MCENTER-HOMOG cluster then joins with the variable MOUTER as indicated by the intersection of the dashes beginning above MCENTER with the slashes starting next to MOUTER. The MCENTER-MOUTER cluster formed then joins with a cluster below it consisting of the variables ZLINE and LIPID. This new cluster is indicated on the tree by the intersection of the dashes beginning above MCENTER with the slashes starting next to LIPID. Finally the variable DIAMETER joins the cluster formed by all of the other variables as indicated by the intersection of the dashes starting above MCENTER with the slashes starting next to DIAMETER.

```
                    ------------------------------/
MCENTER             77 / 59 / 22  32 / 14 /
                      /     /           /      /
                      /     /           /      /
HOMOG               / 43 / 20   6 /  7 /
                         /            /     /
                         /            /     /
MOUTER                / 43  29 / 18 /
                                 /      /
                        ----    /      /
ZLINE                       45 / 6 /
                                /    /
                              /    /
LIPID                          / 8 /
                                  /
                                 /
DIAMETER                        /
```

weak positive correlation between these two variables (r = .1798). However, in the outermost three micron compartment of the same fibers there was a fairly clear and graded increase of mitochondrial volume fraction with increasing Z-line width (Figure 4-2B; r = .4592). The finding that these muscle fiber type characteristics vary in a graded fashion suggests that muscle fiber "type" differentiation occurs as a graded continuum rather than as discrete quantal fiber types as might be implied by the more traditional grouping of fibers into discontinuous typ designations such as type 1, 2A and 2B.

A cluster analysis was carried out for the MCENTER, ZLINE, DIAMETER, LIPID, MOUTER and HOMOG variables using the BMDP program (Hartigan, 1981) for single linkage cluster analysis of variables (Table 4-2). The finding that the mitochondrial variables, MCENTER, MOUTER and HOMOG form a cluster is consistent with the possibility that the density distributions of mitochondria are controlled by a common regulatory mechanism(s) which is (are) presumably influenced by the activity pattern of the motor unit. The Z-line width (ZLINE), the variable most closely correlated with the muscle fiber type (as defined by alkaline actomyosin ATPase staining density), forms a separate cluster with LIPID, the lipid droplet volume fraction.

This study was designed to directly examine the relationship between the ultrastructural characteristics of the muscle fiber types in the human PCA and the corresponding fiber type as determined by the relative pH dependent actomyosin ATPase activities. The technique used to permit this direct correlation necessitated the use of tissue that had been frozen and then thawed. Consequently, certain ultrastructural components such as the sarcoplasmic reticulum were not sufficiently well-preserved to permit morphometric analysis. It would therefore be desirable in future studies to employ some means of identifying the muscle fiber type on the ultrastructural level in conventionally fixed PCA muscles. In an effort to achieve this goal, a stepwise discriminant analysis was run for the variables from this study in order to obtain classification functions that maximize the separation of type 1 from type 2 fibers. Of the available variables, the combined use of Z-line with diameter values was indicated to maximize the separation of these two fiber types.

The use of the classification functions derived from this discriminant analysis yielded a percent of correct fiber type identification of 93.7% for type 1 fibers and 100% for type 2 fibers with an overall percent correct of 95% (Table 4-3). However, this estimate of the power of this classification procedure is probably somewhat optimistic, since in order to relate the fiber type (ATPase) to the corresponding ultrastructural characteristics, it was necessary to use a very time consuming technical procedure which required a small sample size (29 muscle fibers total). Due to the relatively small sample size it was not possible to use split-sample validation to assess the extent to which the equations used idiosyncratic sampling error to create the classification functions. Consequently these functions may be to some

TABLE 4-3. Stepwise Discriminant Analysis
The following stepwise discriminant analysis was carried out using the BMDP program P7M to calculate functions to discriminate between type 1 and type 2 muscle fibers in the human PCA under MCENTER, ZLINE, DIAMETER, LIPID, MOUTER and HOMOG as the available variables.

Classification Function* for Most Useful Discriminating Variables

Fiber Type	One	Two	
ZLINE	0.14431	0.17530	
DIAMETER	0.95633	1.24109	
Constant	−97.34498	−146.45479	
Percent Correct	93.7%	100.0%	Total =95%

**In order to use the classification functions on new data, compute the classifiction score for each group by multiplying the new data by the above coefficients and adding the values to the constant term. The case is then assigned to the group for which the classification score is highest.*

extent more accurate for the sample of muscle fibers examined than for the entire population. In addition since we were only able to examine a single human case using these techniques, it is likely that individual variation in other PCA muscles examined in future studies will result in a somewhat lower percentage of correct fiber type identifications. However, in spite of the limitations imposed on the assessment of the precise power of the classification functions, the data provided by the 29 fibers examined are probably sufficiently representative to provide classification functions that can be used to provide better fiber type identifications than would be possible on the basis of a single variable such as Z-line width alone. These classification functions may be of value in future studies on conventionally fixed PCA muscles, and they may therefore make it possible to obtain data concerning ultrastructural components that were inadequately preserved using the techniques employed in the present study.

As indicated stepwise discriminant analysis yielded a classification function based on ZLINE and DIAMETER values for each fiber type. Since the ZLINE value must be obtained from muscle fibers that are sectioned longitudinally, while the DIAMETER value must be taken from the same fiber sectioned transversely, a computer reconstruction was used to obtain these two values from the same fiber. In future studies it would be easier to use a discriminant function based only on variables that could be obtained from a single sectioning plane. An additional stepwise discriminant analysis was accordingly run using the same variables as in the first analysis with the exception of DIAMETER, the only variable unavailable from longitudinally sectioned fibers. The analysis indicated a discriminant function based solely on the ZLINE variable (Type 1 fiber: coefficient = 0.10736, constant = -62.12236; Type 2 fiber: coefficient = 0.12735, constant = -87.12279) with a percent correct of 87.5% for type 1 fibers and 75.0% for type 2 fibers and an overall percent correct of 85%. Although this discriminant function is not as powerful as the function based on both ZLINE and DIAMETER values for the same fiber, it may provide some indication of fiber type in studies where both values are not available.

DISCUSSION

The posterior cricoarytenoid muscle has an important role in phonation as well as in respiration (Green and Neil, 1955; Hast, 1967; Murakami and Kirchner, 1972; Fukuda, Sasaki and Kirchner, 1973; Hirose, 1976; Hirose and Sawashima, 1981). In man, where capabilities for phonation and speech are highly developed, it might be anticipated that this muscle would be specialized in its differentiation as an adaptation to these functions. Histochemical studies have in fact provided evidence that the muscle fiber types in the PCA differ from those in most other muscles (Edström *et al*, 1974; Malmgren and Gacek, 1981). However, in spite of the physiological importance of the PCA and this evidence that muscle fiber type differentiation in this muscle is atypical of that occuring in most other muscles, there is very little data available concerning the ultrastructural specializations characteristic of fiber types in the PCA. The primary finding of the first ultrastructural study on the intrinsic laryngeal muscles (Berendes and Vogell, 1960) was that the cricothyroid muscle appeared to have fewer mitochondria and a proportionally greater amount of sarcoplasm than the vocalis muscle. This study, however, did not report any findings concerning the ultrastructural characteristics of muscle fibers in the PCA. In a more recent investigation Hirano and Ito (1981) used scanning electron microscopy to examine the ultrastructure of the intrinsic laryngeal muscles of the dog. Although scanning electron microscopy does not offer the resolution possible with transmission electron microscopy, using this tool they were able to detect unusually long, L-shaped or U-shaped mitochondria in the medial portion of the vocalis muscle, in contrast to oval or short rod-shaped mitochondria in the PCA and the other intrinsic laryngeal muscles. They suggested that these two varieties of mitochondrial shape may be related to some functional aspects of the muscle fibers. The present investigation differs from previous ultrastructural studies on the laryngeal muscles in that we have used histochemical and computer reconstruction techniques to relate the muscle fiber type to the corresponding ultrastructural characteristics. Since the muscle fiber type reflects the physiological capacities of the fiber, (Burke *et al*, 1971, 1973; Garnett *et al*, 1971) this technical approach has made it easier to relate the specific ultrastructural characteristics of a fiber type to its probable role in the overall function of the PCA. In addition the use of quantitative techniques has made it possible to make absolute comparisons between the ultrastructural characteristics of specific muscle fiber types in the PCA and data collected by other investigators for muscles serving other functions.

Although there is little comparable quantitative data available in the literature concerning mitochondrial volume fractions in the surface compartment of other human muscles, there have been a number of studies concerning mitochondrial volume fractions in the central core of fibers in other muscles. A comparison of our data for tne central core with values for the same compartment in the other human muscles (gastrocnemius, quadriceps, deltoid, triceps, vastus lateralis: Cullen and Weightman, 1975, Hoppeler *et al*, 1973) indicates that the mitochondrial content of the human PCA muscle is extremely high, with mitochondrial volume fractions that are from three to six fold those reported for the other human muscles examined. Similar findings have recently been reported for the mitochondrial volume fractions of rat PCA muscle fibers, (Hinrichsen and Dulhunty, 1982) which have mitochondrial volume fractions that are approximately two to eight fold those of fibers in other muscles in this animal model (Davey *et al*, 1981). The relatively high mitochondrial volume fractions characteristic of muscle fibers in the human PCA indicate a high capacity for oxidative metabolism and help to explain the high fatigue resistance displayed by this muscle.

Our technical approach allowed us to collect ultrastructural morphometric data for histochemically identified muscle fiber types. We found that the mitochondrial volume fractions characteristic of the central core (outer 3 microns excluded) of type 1 in the human posterior cricoarytenoid muscle are only slightly higher than those in the corresponding compartment of type 2 fibers. These results differ somewhat from the reported mitochondrial values for type 1 and type 2 fibers in human vastus lateralis, deltoid and biceps muscles, (Saltis and Mendel, 1974) where these was a significant ($p < .05$) difference between type 1 (number of mitochondria per micron^2 A-band area = 0.839 ± 0.298 S.D.) and type 2 fibers (number of mitochondria per micron^2 A-band area = 0.549 ± 0.266 S.D.). Their data, however, are for the number of mitochondria per area of A-band rather than for the mitochondrial volume fraction with respect to a cell volume reference. Since our data are expressed as the mitochondrial volume fraction, a direct comparison of our findings with their data for other human muscles is not possible.

The results obtained for the outermost annulus (outermost 3 microns) or human PCA muscle fibers differed from those for the central core in that the mitochondrial volume fractions in this compartment were approximately twice as high in type 2 fibers (mean = 0.381 ± 0.173 S.D.) as in type 1 fibers (mean = 0.184 ± 0.115 S.D.). Saltis and Mendel (1974) found that in other human muscles occasional subsarcolemmal aggregates of mitochondria were found in type 1 fibers but not in type 2 fibers. This observation differs with the above results as well as with our finding that both type 1 and type 2 fibers in the human PCA had higher mean volume fractions in the subsarcolemmal compartment (outermost 3 microns) than in the central core of the fiber. These differences between the fiber types in the human PCA and those in other human muscles probably reflect the activity patterns of motor units in the PCA and the effects of these activity patterns on the differentiation of type 1 and type 2 fibers in this muscle.

Our results for lipid droplet volume fractions in the central core of PCA muscle fibers differed from those for the mitochondrial volume fraction in this compartment, in that type 1 fibers had lipid droplet volume fractions that were approximately four times higher than those of type 2 fibers. Although most of the results that we have obtained concerning the ultrastructural characteristics of type 1 and type 2 fibers in the PCA differ from those obtained by Saltis and Mendel (1974) for other human muscles, the results for the lipid droplet volume fraction suggest a similarity between PCA muscle fiber types and those in other muscles. It should be noted, however, that the seven-fold higher lipid droplet content reported by Saltis and Mendel (1974) for type 1 fibers as compared to type 2 fibers cannot be compared quantitatively to our finding of a fourfold difference between type 1 and type 2 fibers in the PCA. These authors measured the number of lipid droplets per micron^2 A-band rather than the absolute volume fraction of lipid droplets in this compartment. Since it is technically difficult to determine the three-dimensional size of lipid droplets from electron micrographs, it is hard to calculate absolute volume fractions from data concerning the number of lipid droplets in a two-dimensional micrograph sample.

Our results for Z-line widths in the human PCA can be best compared to those of Sjöström and coworkers (1982) who used an almost identical technique to directly relate muscle fiber ATPase histochemistry to Z-line width in the human tibialis anterior muscle. In contrast to their results for the tibialis anterior we have found that in the human PCA type 2A fibers (see Malmgren and Gacek, 1981) have wider Z-line widths than type 1 fibers. Furthermore, our finding is in contrast with many other reports that have compared Z-line widths of fiber types in muscles other than the PCA (see for example: Saltis and Mendel, 1974; Eisenberg and Kuda, 1977).

Although our results for Z-line widths in the PCA are somewhat surprising, it should be noted that the muscle fiber types in the cat (Edström, Lindquist and Martensson, 1974) and human (Malmgren and Gacek, 1981) PCA have previously been reported to have combinations of relative enzyme concentrations that differ from the enzyme profile characteristics of fiber types found in most other muscles. The motor unit activity patterns in the PCA also differ from those in most other muscles in that they are active during each inspiratory cycle (Green and Neil, 1955; Hast, 1967; Suzuki and Kirchner, 1969; Murakami and Kirchner, 1972; Fukuda, Sasaki and Kirchner, 1973). There is much evidence that the physiological, biochemical and ultrastructural characteristics of muscle fibers reflect the pattern of activity of their motor unit (Pette *et al*, 1975; Salmons, Gale and Sreter, 1978; Salmons, 1980). Therefore, the relatively unique activity patterns in PCA motor units might be the basis for our finding of wider Z-lines in type 2A fibers than in type 1 fibers in this muscle. Since our results differ from those obtained for the relationship between Z-line width and fiber type in other muscles, it appears that the regulatory mechanisms controlling these two variables are to some extent separate, and that the relationship between Z-line width and fiber type can be reversed from that seen in most muscles under certain circumstances such as those prevailing in the PCA. Such multiple regulatory mechanisms might be separated in the PCA as opposed to most other muscles on the basis of different activation response times and/or response decay times with respect to motor unit activity patterns in this muscle. Another possible explanation might involve a differential effect of an additional determinant of either fiber type or Z-line width such as a neurotrophic interaction. There is in fact some possibility of an influence of neurotrophic substances on the differentiation of muscle fibers (Gallego *et al*, 1978; Goldring *et al*, 1981).

The diameter of type 1 (mean = 28.6μ ± 10.1) as well as those of type 2 fibers (mean

= 41.5μ ± 9.1) in the human larynx were smaller than values reported by Polgar and coworkers (1973) for the vast majority of many other human muscles examined (mean of mean fiber diameters for various muscles: type 1 = 53.2μ ± 10.3; type 2 = 57.1μ ± 13.2). The relatively small size of muscle fibers in the human PCA would result in lesser oxygen diffusion distances in these fibers as compared to those in most other human muscles. This characteristic seems consistent with the extremely high mitochondrial volume fractions observed in the PCA, and it would suggest that these fibers are highly specialized for oxidative metabolism and resistance to fatigue. Similarly, the finding that type 1 fibers in the PCA have smaller fibers than type 2 fibers in this muscle indicates a lesser oxygen diffusion distance in type 1 fibers as compared to type 2 fibers. However, the mitochondrial volume fractions for the central core compartment were only slightly higher in type 1 fibers than in type 2 fibers, and in the surface compartment this relationship was, in fact, reversed with mitochondrial volume fractions in type 2 fibers being approximately twice as high as type 1 fibers. It may be that the high surface concentrations of mitochondria found in type 2 fibers in the PCA help to compensate for limitations imposed on the diffusion of oxygen by the larger size of these fibers.

The maximum tension force output for a single motor unit can be calculated as the product of the innervation ratio, the average area of the unit's muscle fibers and the specific tension output for the fiber type expressed as force per unit of cross-sectional area (Close, 1972). Another consequence therefore of the relatively small size of muscle fibers in the PCA would be, that at a constant innervation ratio the total cross sectional area of muscle fibers included within a single motor unit would be smaller than in a motor unit composed of larger muscle fibers. This would mean that the tensions developed by the recruitment of each motor unit would be smaller than with motor units composed of the same number of larger fibers of the same type. In addition the number of muscle fibers included within each motor unit is probably low in the PCA as compared to most other muscles (Faaborg-Andersen, 1957; Hinrichsen and Ryan, 1982). This finding together with the finding that the PCA muscle fibers are unusually small would suggest that relatively small increments of tension would be associated with the progressive recruitment of individual motor units in this muscle. These characteristics would therefore form a basis for very fine control of tension generation in the PCA.

SUMMARY

We have used a technique based on computer reconstructions and histochemistry to directly relate quantitative ultrastructural data to fiber type in 29 muscle fibers from a single human posterior cricoarytenoid muscle obtained from a nonirradiated male volunteer undergoing laryngectomy for malignancy. Since the technique used was extremely time consuming, it was not feasible to sample additional human cases to determine the range of individual variation with respect to the variables considered. However, if it can be assumed that the case examined was not extremely unusual, our findings suggest that the ultrastructural characteristics of the human PCA differ greatly from those of most other human muscles and that this muscle is highly specialized toward its physiological functions.

Our results suggest that the muscle fibers of the human PCA differ from those in most other muscles in that they have a substantially higher proportion of the cell volume occupied by mitochondria and lipid droplets. Both of these findings are indications of a specialization toward oxidative metabolism and resistance to fatigue. The fiber types in the PCA are also much smaller than fibers found in most other muscles which may represent a further specialization for fatigue resistance by improving the logistics of oxygen diffusion.

Type 2A (high alkaline ATPase) fibers in the PCA differ from type 1 fibers in this muscle (low alkaline ATPase) in that they are larger and have wider Z-line widths than type 1 fibers. Their larger size may be an adaptation to their position in the recruitment sequence (presumably later than type 1 fibers), since limitations on oxygen diffusion distances would be of lesser importance in fibers that generate tension with an activity pattern more intermittent than that of type 1 fibers. The spacial distribution of mitochondrial volume fractions in these two fiber types is also consistent with this interpretation. In type 2A fibers there are relatively dense concentrations of mitochondria near the cell surface which would help to decrease the oxygen diffusion distance in these larger type 2A fibers. This should help the 2A fibers to approach the fatigue resistance of type 1 fibers, yet still offer the greater muscle fiber cross sectional areas (greater maximum tension force output per fiber) needed to generate the relatively larger tensions typical of motor units recruited late in the sequence.

ACKNOWLEDGEMENTS

We wish to thank Mrs. Norma Slepecky for her technical assistance and Mrs. Pat Kane for her help in the organization of the manuscript. This project was supported by NIH Grant NS 1826502.

REFERENCES

Berendes, J. and Vogell, W. (1960). "Kehlkopfmuskeln im electronenmikroskopischen Bild," *Archiv. Fur Ohr. Nas. -und Kehlkopf Heilkunde* 176, 730-741.

Burke, R.E., Levine, D.N., Tsairis, P., and Zajac, F.E. (1973). "Physiological types and histochemical profiles in motor units of the cat gastrocnemius," *J. Physiol.* 234, 723-748.

Burke, R.E., Levine, D.H., Zajac, F.E., Tsairis, P. and Engel, W.K. (1971). "Mammalian motor units: physiological-histochemical correlation in three types of the cat gastrocnemius," *Science* 174, 709-712.

Close, T. (1972). "Dynamic properties of mammalian skeletal muscles," *Physiol. Rev.* 52, 129-197.

Cullen, M.J. and Weightman, D. (1975). "The ultrastructure of normal human muscle in relation to fiber type," *J. Neurol. Sci.* 25, 43-56.

Davey, D.F., Dunlop, C., Hoh, J.Y. and Wong, S.Y.P. (1981). "Contractile properties and ultrastructure of extensor digitorum longus and soleus muscles in spinal cord transected rats," *Aust. J. Exp. Biol. Med. Sci.* 59, 393-404.

Edström, L., Linquist, C., and Martensoon, A. (1974). "Histochemical properties of intrinsic laryngeal muscle," in Wyke, B. (Ed.) *Ventilatory and Phonatory Control Systems*. London: Oxford University Press, 392-407.

Eisenberg, B.R. and Kuda, A.M. (1977). "Retrieval of cryostat sections for comparison of histochemistry and quantitative electron microscopy in a muscle fiber," *J. Histochem. Cytochem.* 25, 1169-1177.

Elias, H., Hennig, A. and Schwarts, D.E. (1971). "Stereology: applications to biomedical research," *Physiol. Rev.* 51, 158-165.

Engel, W.K. (1970). "Selective and nonselective susceptibility of muscle fiber types," *Arch. Neurol.* 22, 97-117.

Engel, W.K., Bishop, D. and Cunningham, G.G. (1970). "Tubular aggregates in type II muscle fibers: ultrastructural and histochemical localization," *J. Ultrastruc. Res.* 31, 507-525.

Engel, W.K., Gold, G.N. and Karpati, G. (1968). "Type I fiber hypotrophy and central nuclei: A rare congenital muscle abnormality with possible experimental model," *Arch.*

Neurol. 18, 435-444.

Faaborg-Anderson, K. (1957). "Electromyographic investigation of intrinsic laryngeal muscles in humans," *Acta Physiol. Scand. Suppl.* 140, 1-147.

Fenichal, G.M. and Engel, W.K. (1963). "Histochemistry of muscle in infantile spinal muscular atrophy," *Neurology* 13, 1059-1066.

Fukuda, H., Sasaki, C.T. and Kirchner, J.A. (1973). "Vagel afferent influences of the phasic activity of the posterior cricoarytenoid muscle," *Acta Otolaryngol.* 75, 112-118.

Gallego, R., Huizar, P., Kuno, N. and Kuno, M. (1978). "Disparity of motorneurone and muscle differentiation following spinal transection in the kitten," *J. Physiol.* 281, 253-265.

Garnett, R.A., O'Donovan, M.J., Stephans, J.A. and Taylor, A. (1979). "Motor unit organization of human medial gastrocnemius," *J. Physiol.* 287, 33-43.

Goldring, J.M., Kuno, M., Nunez, R. and Weakly, J.N. (1981). "Do identical activity patterns in fast and slow motor axons exert the same influence on the twitch time of cat skeletal muscles?" *J. Physiol.* 321, 211-223.

Green, J.H. and Neil, E. (1955). "The respiratory function of the laryngeal muscle," *J. Physiol.* 129, 134-141.

Guth, L. and Samaha, F.J. (1970). "Procedure for the histochemical demonstration of actomyosin ATPase," *Exp. Neurol.* 28, 365-367.

Hartigan, J. (1981). "P1M-Cluster analysis of variables," in Dixon, W.J. (Ed.) *BMDP Statistical Software, 1981 Edition.* Berkeley: University of California Press, 447-458.

Hast, M.H. (1967). "The respiratory muscle of the larynx," *Ann. Otol. Rhinol. and Laryngol.* 234, 489-497.

Hinrichsen, C. and Dulhunty, A. (1982). "The contractile properties histochemistry, ultrastructure and electrophysiology of the cricothyroid and posterior cricoarytenoid muscles in the rat," *J. Mus. Res. & Cell Motil.* 3, 169-190.

Hinrichsen, C.F.L. and Ryan, A. (1982). "The size of motor units in the laryngeal muscles of the rat," *Experimentia* 38, 360-361.

Hirano, M. and Ito, T. (1981). "Ultrastructure of laryngeal muscles," *Biomed. Res.* 2, 289-291.

Hirose, H. (1976). "Posterior cricoarytenoid as a speech muscle," *Ann. Otol. Rhinol. & Laryngol.* 85, 334-443.

Hirose, H. and Sawashima, M. (1981). "Functions of the laryngeal muscles in speech," in K.N. Stevens and M. Hirano (Eds.) *Vocal Fold Physiology.* Tokyo: Univ. of Tokyo Press, 137-154.

Hirose, H., Ushijima, T., Kobayahsi, T. and Sawashima, M. (1969). "An experimental study of the contraction properties of the laryngeal muscles in the cat," *Ann. Otol. Rhinol. & Laryngol.* 78, 297-307.

Hoppeler, H., Luthi, P., Claassen, H., Weibel, E.R. and Howald, H. (1973). "The ultrastructure of the normal human skeletal muscle: A morphometric analysis on untrained men, women, and well-trained orienteers," *Pflugers Arch.* 344, 217-232.

Malmgren, L.T. and Gacek, R.R. (1981). "Histochemical characteristics of muscle fiber types in the posterior cricoarytenoid muscle," *Ann. Otol. Rhinol. & Laryngol.* 90, 423-429.

Mendel, J.R. and Engel, W.K. (1971). "The fine structure of type II muscle fiber atrophy," *Neurology* 21, 358-365.

Murakami, Y. and Kirchner, J.A. (1972). "Respiratory movements of the vocal cords. An electromyographic study in the cat," *Laryngoscope* 82, 454-467.

Pette, D., Ramirez, B.U., Muller, W., Simon, R., Exner, G.V. and Hildebrand, R. (1975). "Influence of intermittent long-term stimulation on contractile, histochemical and metabolic properties of fibre populations in fast and slow rabbit muscles," *Pfuegers Arch.* 361, 107.

Polgar, J., Johnson, M.A., Weightman, D. and Appleton, D. (1973). "Data on fibre size in thirty-six human muscles: An autopsy study," *J. Neurol. Sci.* 19, 307-318.

Salmons, S. (1980). "The response of skeletal muscle to different patterns of use - some new developments and concepts," in Pette, D. (Ed.) *Plasticity of Muscle.* Berlin: de Gruyter, 387-399.

Salmons. S., Gale, D.R. and Sreter, F.A. (1978). "Ultrastructural aspects of the transformation of muscle fibre type by long term stimulation: changes in Z discs and mitochondria," *J. Anat.* 127, 17-31.

Saltis, L.M. and Mendell, J.R. (1974). "The fine structural differences in human muscle fiber types based on peroxidatic activity," *J. Neuropath. & Exp. Neurol.* 34, 632-639.

Sjöström, J., Kidman, S., Larsen, K.H., and Angquist, K.-A. (1972). "Z- and M-band appearance in different histochemically defined types of human skeletal muscle fibers," *J. Histo. Cytochem.*

Spurr, A.R. (1969). A low-viscosity epoxy resin embedding madium for electron microscopy," *J. Ultrastruct. Res.* 26, 31-38.

Underwood, E.E. (1970). *Quantitive Stereology.* Addison-Wesley Publishing Co., Reading, Massachusetts.

Weibel, E.R. and Elias, H. (1967). *Quantitative Methods in Morphology.* (First International Stereology Conference), Springer-Verlag, New York, Inc. New York.

DISCUSSION

D. COOPER: Several investigators have observed fiber type grouping in apparently healthy PCA muscles. Are there specific reasons why this might happen?

L. MALMGREN: This was first reported by Teig and coworkers (*Acta Otolaryngol.* 85, 272) and later also observed in a study carried out in my lab (*Ann. Otol. Rhinol. & Laryngol.* 90, 423). I should first explain that muscle fiber type grouping is an indication of partial denervation followed by reinnervation by collateral sprouting; so the fact that this phenomenon indicates partial dennervation implies that there is something such as compression of a nerve, leading to partial denervation, followed by reinnervation. Teig's group suggested that this may be the result of compression on the recurrent laryngeal nerve caused by swallowing large pieces of food. Another possible explanation that I've suggested is that this phenomenon may be in some way related to the specific activity patterns characteristic of the PCA motor units. In my presentation, I've indicated that the histochemical and ultrastructural characteristics of the PCA muscle fibers differ greatly from the type of differentiation characteristic of muscle fiber types found in most other muscles in the body. The unusual characteristics of fiber types in the PCA are presumably a consequence of the specific activity patterns typical of PCA motor units which would differ greatly from those in, for example, a limb muscle. It is therefore conceivable that the specific activity patterns characteristic of the PCA motor units also lead in some way to an earlier cell death in some of the motor neurons in this system. This would ultimately result in the observed muscle fiber type grouping. These are only some thoughts. However at this point, I don't think that we really have any clear explanation for these observations of muscle fiber type grouping in the PCA.

D. ZEALEAR: Are there any sources of human PCA muscles other than from laryngectomies, where the effect of the cancer may be a consideration?

L. MALMGREN: We used tissues from laryngectomy cases since the originally intended focus of our study was the histochemistry of the muscle fiber types in the human PCA, and since there are rapid postmortem changes in some enzymes we limited our study to fresh tissues obtained in the operating room. There was a recent study based on the use of tissues taken from autopsy. They didn't report muscle fiber type grouping so there is certainly a possibility that the muscle fiber type grouping that has been observed in PCA muscles obtained from laryngectomies is in some way associated with the tumor.

D. ZEALEAR: I assume that these were unilaterally involved tumors. Did you look at the tumor side as well in comparison with regard to the extent of type grouping?

L. MALMGREN: In the cases examined the PCA muscles were not involved in the tumor so we examined both PCA muscles in some cases. There was no consistent relationship between the side on which the tumor was found and the extent of type grouping in the right as compared to the left PCA muscle. Muscle fiber type grouping is also known to occur frequently in the PCA muscles of horses (*Equine Vet. J.* 4, 144) where the extent of type grouping is generally greater on the left side than on the right side. We therefore also compared the extent of muscle fiber type grouping in the right human PCA muscles to that found in the left PCA muscle and failed to see any consistent difference in the extent of type grouping between left and right human PCA muscles (*Ann. Otol. Rhinol. Laryngol.* 90, 423).

B. WYKE: It might be appropriate to point out that muscle fiber type grouping has also been commonly observed in the small muscles of the foot in older people, and there is a suspicion now among neuropathologists that this is in fact a consequence of aging. This may happen in a number of other striated muscles. There is a study in progress in Stockholm right now which suggests that this also occurs in the diaphragm. So I think that we now need to start to pay particular attention to the ages of the cases studied and to specify the ages quite specifically in reports.

L. MALMGREN: I am particularly interested to know that this has also been observed in the diaphragm. I have been aware of studies indicating that this was found in some of the small muscles in the foot, and it has also been suggested that this might be due to the wearing of tight shoes, but the diaphragm is quite a different situation anatomically which would favor the hypothesis that muscle fiber type grouping, at least in some situations, may be related to aging.

Note added after meeting: The occurrence of muscle fiber type grouping in the diaphragm is also interesting in the sense that both the diaphragm and the PCA muscles are involved in respiration and the activity patterns of at least some of the motor units in each of these muscles would therefore be similar. This might be consistent with the idea that I mentioned earlier concerning the possibility of an earlier neuronal cell death in motor units having such activity patterns.

E. LUSCHEI: To push that very far, you have to arrive at some definition of what is meant by type grouping. For example, if I see two of the same type of fiber together, do I call that a group? How about three? In other words, I would say that in a big field even if you have a random distribution of fiber types you might observe some apparent grouping by chance alone so I think that you probably need to consider the probability of the occurence of type grouping as a random event.

L. MALMGREN: Yes, but the extent of type grouping observed in the human PCA was very extreme (see micrographs in Teig, Dahl and Thorkelsen, *Acta Otolarngol.* 95, 272; Malmgren and Gacek, *Ann. Otol. Rhinol. & Laryngol.* 90, 423) and such considerations would only be necessary in order to reliably detect lesser degrees of type grouping in the PCA. By "extreme" I mean muscles having many areas with, for example, 50 fibers all of the same type together in a group bordered by groups of equal size but another fiber type. This would clearly be highly improbable as a random event in a muscle such as the PCA which typically has approximately a 50/50 % composition of type 1 and type 2 fibers. However I certainly agree that in order to detect lesser degrees of type grouping reliably it would be necessary to consider the statistical probability that the fiber type distribution cannot be expected to occur in a random distribution.

E. LUSHEI: I've seen in the jaw muscles of young monkeys an enormous change across a fascicle from say all type 1 to all type 2. Would you consider that a group?

L. MALMGREN: That probably reflects a sharp gradient in the topographical distribution of motor unit territories. Such fiber type distributions occur in a number of muscles and are typically found in muscles that are capable of contracting in a geometrically complex way (i.e., spacially complex recruitment sequence of motor unit territories within the muscle). Muscle fiber type grouping is not generally as graded and occurs as sharply defined groups scattered about the muscle. The "groups" in muscle fiber type grouping are more sharply defined, and generally the groups themselves are more randomly distributed within the muscle.

5 ▪ FUNCTIONAL ORGANIZATION OF THE PRIMARY MOTOR CORTEX CONTROLLING THE FACE, TONGUE, JAW, AND LARYNX IN THE MONKEY

David L. Zealear, Malcolm H. Hast, and Zoya Kurago

INTRODUCTION

Vocal communication can be broken down into two components, verbal and nonverbal. The nonverbal component is the intonation or emotional context of a phonation. The verbal component is the extrapersonal information conveyed in vocalization and is dependent upon the nature of the specific sounds emitted and their syntax. Observations have been made during brain stimulation and following brain damage in humans suggesting that the cortex and the subcortex are differentially involved in the control of these two components (Penfield 1938, 1949; Jurgens 1979). Stimulation of particular regions within the diencephalon of the subcortex elicits phonations such as laughter or moaning which reflect particular emotional states. In contrast, stimulation of the lateral face area or the supplementary motor area of the motor cortex produces less primitive vocalizations (i.e., long drawn out vowels possibly representing individual phonemes). Bilateral damage involving these cortical areas severely compromises speech production but has little effect on emotionally coupled phonation (Larson *et al*, 1981).

Vocalization in non-human mammals (e.g., cats, monkeys, apes) also appears to be controlled, to some extent, by both the cortex and subcortex. As in the human, stimulation of particular subcortical regions produces emotionally coupled phonations which are stereotypic for a given species but akin to those produced in the human. The degree of cortical influence on oral communication in non-human mammals, on the other hand, is less certain and appears to vary with the level of cortical evolvement of a given species. Stimulation of the lateral face area has only successfully produced vocalization in the ape, although laryngeal muscle contractions have been reported in both monkey and cat. The laryngeal control is more extensive in monkey than cat, suggesting a more significant role of this area in monkey vocalization. Another line of observations indicating the phonatory importance of the lateral face area in mammals with more developed cortication comes from neuroanatomical tracing experiments. Direct cortical connections from this area to the nucleus ambiguous, the source of laryngeal motoneurons, have been demonstrated to be strong in man, significant in the ape, possibly absent in monkey, and absent in cat (Jurgens, 1979).

Given the importance of the lateral face area in speech production, we set out to study this region of the primary motor cortex using state of the art neurophysiological techniques, such as microstimulation and single unit recording in the awake behaving animal. Our choice of experimental animal was the rhesus monkey, since these invasive techniques can not be used ethically in experiments with either apes (i.e., all species are endangered) or man. Although brain surface macrostimulation, evoked potential recording, and ablation (or damage) studies in humans and other mammals have provided us with a gross approximation as to the localization of brain function, these newer techniques can give us detailed information, on a cell to cell basis, regarding the internal spatial and functional organization of an individual brain region. More specifically, by microstimulating at numerous sites within the lateral face area and recording the peripheral response obtained, it has been possible to gradually establish a topographical map of the area with respect to its control of particular parts of the orofacial musculature (i.e., muscles of the face, tongue, jaw, and larynx). By then recording from single neurons in an awake behaving monkey at sites first stimulated for identification, it has also been possible to determine the functional roles of neurons in communication or other orofacial activity. A basic principle involved in the organization of this cortical area became apparent during the course of these studies as will be described. This principle may also be common to the organization of other regions of the primary motor cortex and may provide a key to our understanding the manner in which the lateral face region directs orofacial activity during speech production in man.

METHODS

Chronic studies were conducted on seven (7) monkeys. A general anesthetic was administered to an animal before any surgical procedure was performed. Following removal of a bone flap overlying the lateral aspect of the left primary motor cortex, a microelectrode chamber was affixed to the skull of each animal. A restraining bolt and EMG plug were also affixed to the skulls. Bipolar EMG electrodes were introduced into eleven (11) different laryngeal muscles and their lead wires brought subcutaneously to the skull plug. For all animals except the last, sessions were conducted under light ketamine anesthesia three (3) times a week for as long as three (3) months. During a session, an animal was placed in a monkey chair with head immobilized by the restraining bolt. A microelectrode was introduced into the lateral face area. Recordings were first taken with the microelectrode to pickup neuron activity, indicating that the microelectrode had pierced the dura and entered the surface of the brain. Microstimulation (25 μamps) was then performed at sites separated by 50-500 microns, the degree of separation depending upon whether there was a tendency for changes to occur in the peripheral response. Stimulus trains of 100 ms duration, 333/second, and 200 microseconds pulse duration were used. Laryngeal muscle responses were monitored by observing changes in EMG activity. Face, tongue, and jaw responses were observed visually and documented by videotaping for later analysis. If a response was obtained at a site, the current level was lowered and the threshold for the response and its latency determined if possible. Changes in the threshold level for a response with depth, the threshold profile, provided a basis for determining the boundaries of the elemental zones in cortex, termed "efferent zones". That is, if the threshold for a response rose with electrode advancement and the nature of the response then changed, we knew we had crossed the boundary of a zone. Increases in the threshold for a response, along with disappearance of recorded neuron activity, also indicated when we had left gray matter and entered white matter. The nature of responses within zones and their stereotaxic coordinates were recorded. Following the terminal session on an animal, zones were referenced to surface photography of the cortex and to histological sections of the brain containing marking lesions of zone sites.

The last animal (S24) was implanted similarly to the others, but then trained to vocalize for food rewards before physiological sessions commenced. Sessions were conducted in the awake state, every day for six (6) months. Although microstimulation was performed to map this area and characterize efferent zones, considerable time was spent recording from neurons to determine their roles in orofacial activity such as oral communication (e.g. barking, cooing, shrieking), facial communication (lip smacking), and feeding.

RESULTS AND DISSCUSSION

The region of the primary motor cortex studied is circumscribed in Figure 5-1. Since more extensive information was obtained from experiments on our last two animals, only the results from these experiments will be presented. The map of laryngeal muscle representation obtained from one of these monkeys (1709) is shown in Figure 5-2. Each symbol or group of symbols at a spot indicates the particular laryngeal muscle or muscles which responded to microstimulation within a single efferent zone. If more than one zone was encountered during a microelectrode penetration, only the most superficial zone was plotted in the figure. The intrinsic adductor miscles (solid symbols excluding stars) and extrinsic miscles (open symbols) were represented within an ovoid region extending from the subcentral dimple inferiorly, towards the Sylvian fissure. The posterior cricoarytenoid (the vocal fold abductor muscle) representation extended outside this region (stars). Although not depicted in the figure, it also completely covered the ovoid region, giving it a much larger representation than any other muscle.

Although maintenance of continuity in the neural representation of the body, so called "somatotopy", is a rule which applies to the primary motor cortex as a whole (Woolsey *et al*, 1952), there was no obvious display of somatotopy in the representation of laryngeal muscles. As shown in Figure 5-2, none of the muscles dominated a particular field within the ovoid laryngeal region. On the other hand, it is possible that somototopy with respect to individual muscles existed but was obscured because of the extensive overlap among individual muscle fields.

At the level of the efferent zone, the overlap in muscle fields could have been either a reflection of the mixing or interdigitation of zones controlling single but different muscles or simply a reflection that zones themselves controlled more than one muscle. Although there is support for the general notion that zones in motor cortex only control single muscles (Asanuma, 1975; McGuiness *et al*, 1980), all the zones encountered in the laryngeal region were multimuscle zones. In particular, every zone controlled the posterior cricoarytenoid and one or more adductor or extrinsic muscle. For example, the arrow in Figure 5-2 indicates a zone encountered during one penetration controlling the PCA, TA, and CT muscles. The EMG responses and the threshold current levels required for responses from this zone are displayed as a function of microelectrode depth in Figures 5-3 and 5-4. Referring to Figure

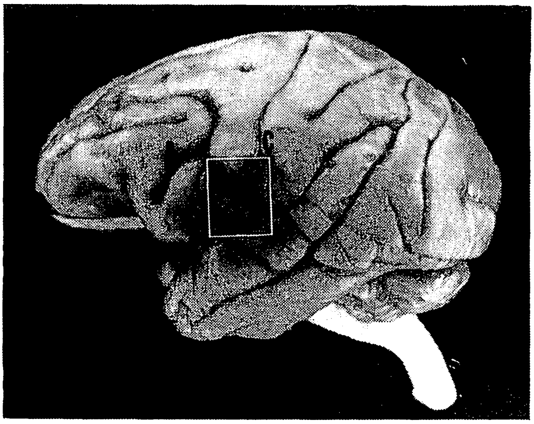


FIGURE 5-1
Left hemisphere of monkey S24 showing region studied.
A: Arcuate sulcus;
C: Central sulcus;
S: Sylvian fissure.
Arrow indicates subcentral dimple.

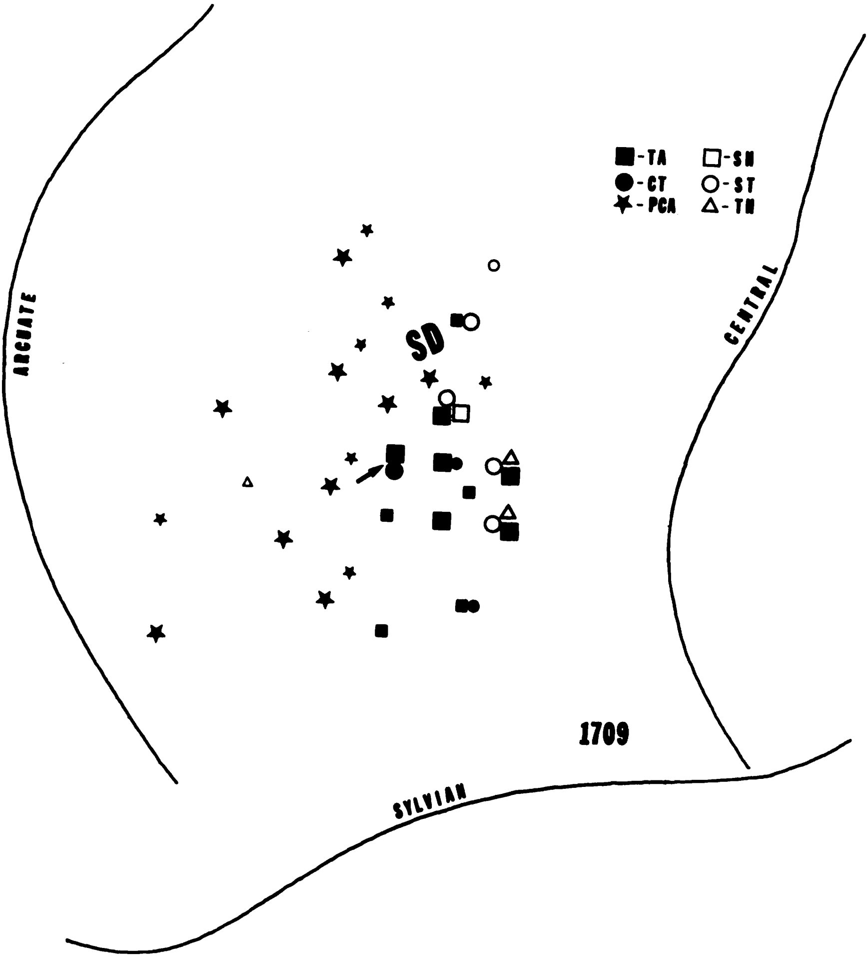


FIGURE 5-2
Map of laryngeal zones. Large symbols indicate responses with thresholds of 5 μamps or less and small symbols thresholds of 6-10 μamps. TA: Thyroartenoid; CT: Cricothyroid; PCA: Posterior cricoarytenoid; SH: Sternohyoid; ST: Sternothyroid; TH: Thyrohyoid. All ipsilateral muscles. SD: Subcentral dimple. Although not indicated, posterior cricoarytenoid inhibitions were associated with excitatory responses of all other muscles.

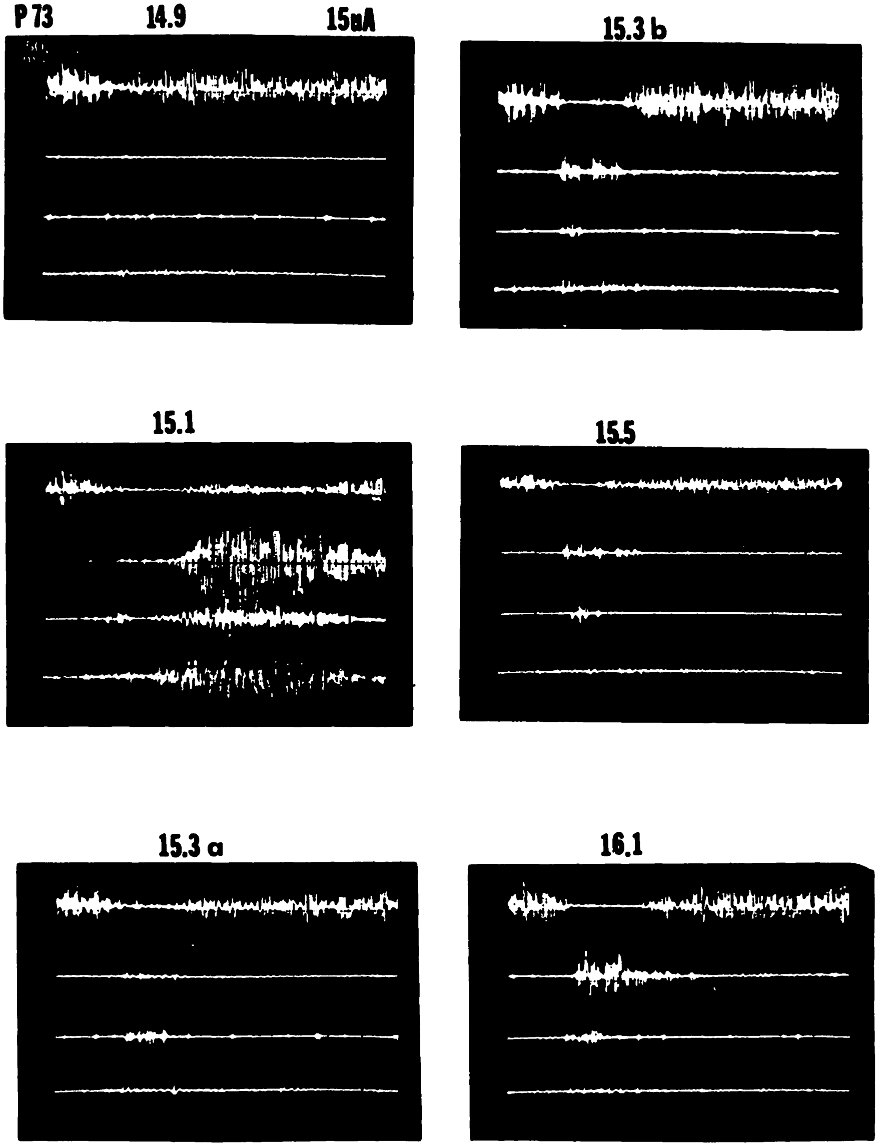


FIGURE 5-3
EMG responses elicited by mircostimulation as a function of microelectrode depth. In the first frame 14.9 mm corresponded to a depth of .4 mm below the cortical surface. The traces in each frame are recordings from the PCA, TA, CT, and ST muscles (top to bottom). Stimulus current was 15 μamps.

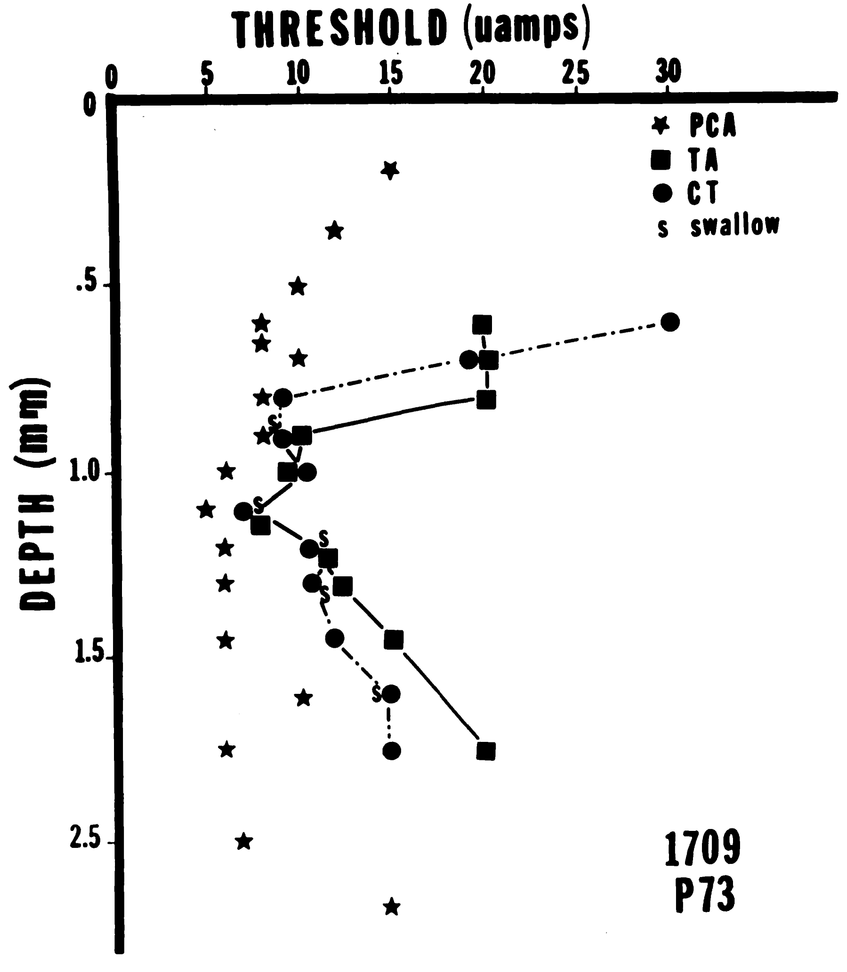


FIGURE 5-4
Threshold levels as a function of depth for the responses of Figure 5-3. The threshold levels decreased with depth as the microelectrode more closely approached and directly activated an output neuron or neurons of the cortex. The same neuron or neurons must have been responsible for all the responses, since the response thresholds changed similarily (i.e., a single efferent zone was involved).

5-4, the first response observed following penetration of the cortex with the microelectrode was an inhibition of the PCA muscle. An example of the inhibitory response is shown in the top trace of Figure 5-3 at a depth of .4 mm (corresponding to 14.9 mm). Notice that the latency for the response was quite long, approximately 60 msec from the onset of the stimulus. With further advance of the electrode, microstimulation evoked an excitatory response of the CT muscle (Figure 5-3, 15.3a, third trace) concomitant with the PCA inhibition. The threshold

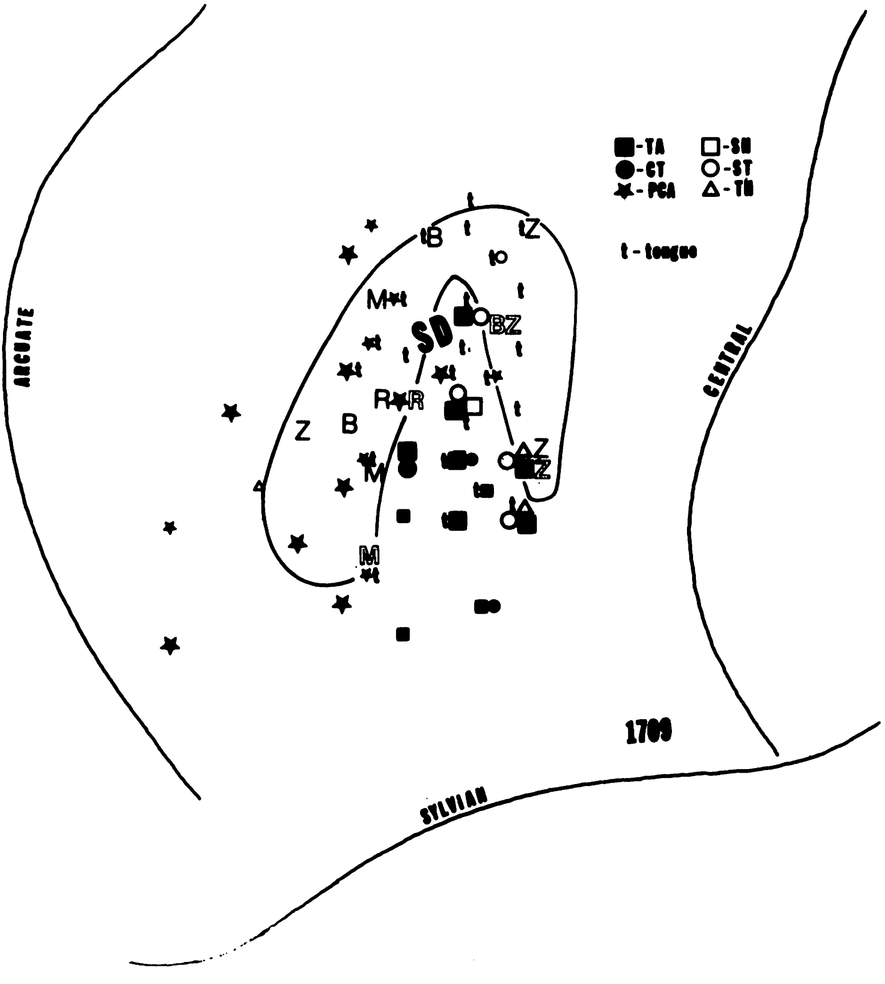


FIGURE 5-5

Face and tongue responses obtained with thresholders of less than 10 μamps added to the laryngeal map of Figure 5-2. Open letters indicate ipsilateral face muscles, solid letters indicate contralateral face muscles. M: Mentalis; R: Orbicularis oris; B: Buccinator; Z: Zygomaticus, Threshold profiles of face and tongue responses not studied.

for the response at this depth (Figure 5-4, .8 mm) was identical to that for the PCA. With slight advance of the microelectrode, an excitatory response of another adductor muscle, the TA, was recruited with microstimulation (Figure 5-3, 15.3b, second trace), and this excitatory response became the dominant one (e.g., compare the TA and CT responses in right hand frames of Figure 5-3). The thresholds for all three of the responses decreased with further advance of the microelectrode (Figure 5-4), reaching a minimum at 1.2 mm below the cortical surface, corresponding to layer 5 of the gray matter, as determined histologically. The thresholds then rose, as shown, when microstimulating past layer 5 and within the white matter. In addition to these muscle responses, a patterned swallow like movement of the pharynx was observed during microstimulation within this and other efferent zones. The "swallow" was similar to the muscle responses in its threshold profile (Figure 5-4) but was distinguished in that it could not be evoked consistently with every stimulation, regardless of the current level. The muscle EMG responses associated with a "swallow" are shown in Figure 5-3 (15.1) and were demarcated by their longer latency and longer duration.

It might be helpful at this time to discuss some of the features involved in cortical laryngeal control. First, although the laryngeal map defined by microstimulation was less extensive than that determined previously by Hast, *et al.* (1974) using macrostimulation [macrostimulation has limitations as a mapping tool because of excessive current spread] the laryngeal representation was, nevertheless, relatively large. The PCA representation was particularly widespread, indicating that the primary cortical influence on the larynx was directed at the PCA to "not open the glottis". On the other hand, the ovoid region was also significant in size so that instructions from this region to the other laryngeal muscles dictating specific glottal configurations would be important for particular voluntary laryngeal activities. Second, an unexpected feature of the laryngeal representation was that the cortex did not exhibit direct control over individual laryngeal muscles. The efferent zones, the building blocks of the motor cortex, controlled more than one muscle. That the control was indirect was suggested by the long EMG latencies following cortical microstimulation. Certainly latencies of less than 10 - 20 msec would have been observed if direct cortical connections to laryngeal motoneurons existed. In this regard, it should be reiterated that anatomical tracing studies have found "little if any" direct projection from the motor cortex to the nucleus ambiguous in the monkey. The reciprocal relationship demonstrated in our studies between the abductor and adductor responses during cortical microstimulation lend support for the idea that the cortical influence is mediated through reticular inspiratory and expiratory interneurons. A final point of discussion concerns the vertical organization of the laryngeal region. In penetrations orthogonal to the cortical surface, such as the one described in Figure 5-4, only one efferent zone was encountered. That is, although the thresholds for responses changed with depth as exemplified in Figure 5-4, the nature of the responses did not change within an orthogonal column of gray matter. Microstimulation was effective in eliciting a response when the microelectrode was deep, in layer 5, because Betz cells, the cortical output neurons, could be activated directly. Superficial microstimulation was less effective because Betz cells were only affected indirectly, via activation of superficial neurons within the column. The point is, activation of neurons within an orthogonal column gave the same response. A columnar vertical organization has been found to exist in the motor cortex and in many other regions of the cerebral cortex in general (Mountcastle, 1957; Hubel *et al* (1962; Asanuma, 1967).

During the course of study of the cortical laryngeal region, we often elicited movements of the face, tongue, and jaw with microstimulation. In our last two animals we documented these responses and undertook a formal investigation of the cortical representation of these three other structures. In Figure 5-5 a map of the face and tongue obtained from monkey 1709 was added to the map shown in Figure 5-2 of the laryngeal region (the jaw representation was not studied in this animal). The tongue representation occupied the center of the precentral gyrus and extended from above inferiorly into the ovoid region of the laryngeal representation. The tongue and laryngeal regions were flanked by an anterior and posterior face field which were bridged in the middle, forming a horseshoe shaped face representation as outlined in the figure. The face representation seemed to be somatotopically organized. The more exterior or lateral aspect of each face field represented the more lateralized muscles of the

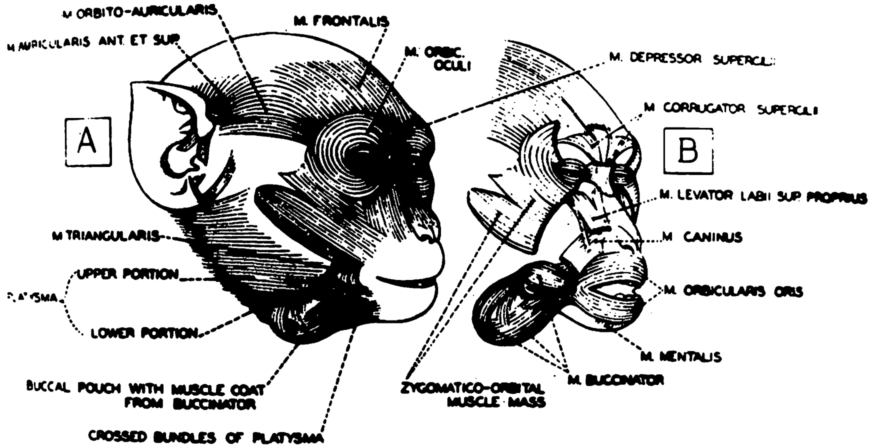


FIGURE 5-6
Drawing of muscles of facial expression in the rhesus monkey. Taken from Huber (1933).

face such as the zygomaticus, and the central bridge area represented the more midline face muscles, such as the orbicularis oris and mentalis (see Figure 5-6).

In our last animal (S24) we performed a more detailed study of the face, tongue, and jaw representations including an investigation of the composition of their efferent zones. In view of the unexpected finding regarding the nature of laryngeal efferent zones, we were particularly interested in determining if face, tongue, and jaw zones controlled single muscles as claimed for other regions of the motor cortex. In Figure 5-7 the most superficial zone encountered at each penetration is described by symbols representing the structure or structures which responded. Only responses which had thresholds of 10 μamps or less were included, so that we could be confident that they resulted from activation of neural elements within single zones. [This assumes that neural elements a distance of less than 80-90 microns from the electrode tip belong to one zone (see Stoney, *et al*, 1968).] It is quite obvious from the map that the majority of the zones controlled more than one muscle, because the majority controlled a muscle in more than one structure. However, it should be emphasized that the responses produced within a zone were not necessarily equal in strength. (Large symbols indicate strong responses, small symbols weak responses in Figure 5-7). It could be argued that when multiple responses were obtained at a microstimulation site that additional responses to the dominant one resulted from spread of current and activation of adjacent zones. This is unlikely first because there was not a dominant response in approximately half of the multimuscle zones, and second the minimum threshold levels and profile curves for responses were not significantly different. One would expect higher threshold levels and possibly different threshold curves during activation of adjacent or nearby zones. This second point is illustrated in Figure 5-8 for the penetration indicated by the arrow in Figure 5-7. The penetration was somewhat oblique to the cortical surface, because the penetration was made down the superior bank of the sulcus depicted in Figure 5-7, extending from the arcuate sulcus to the subcentral dimple. For this reason the microelectrode cut across four (4) different zones as described in the legend of Figure 5-8. Note that in each zone the minimum thresholds for responses were indistinguishable and their threshold profiles similar.

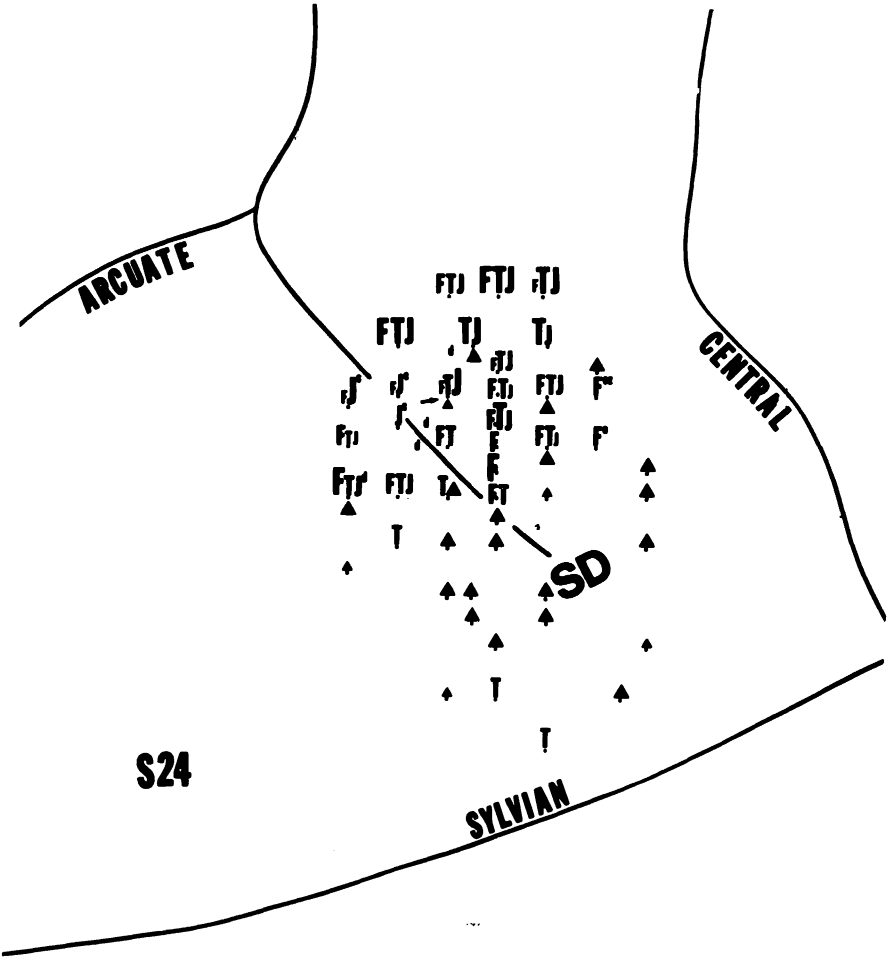


FIGURE 5-7
Map of face, tongue, jaw (and larynx) zones for monkey S24. Only the most superficial zone encountered in each penetration was plotted (except for "d"). Triangles indicate laryngeal zones (large triangles thresholds less than 10 μamps, small triangles thresholds 6-10 μamps.) All zones with face, tongue, or jaw responses had thresholds of 10 μamps or less (large symbols indicate strong responses, medium sized symbols moderate responses, and small symbols weak responses). "J^{c}" indicates zone with jaw closing. "J^{d}" indicates zone with jaw deviation. "d" indicates approximate position of a deeper zone with jaw deviation. "F^{o}" indicates orbicularis oculi zone. "F^{oz}" indicates orbicularis oculi and zygomaticus zone.

In viewing a complete map of the representations of the face, tongue, jaw, and larynx as shown in Figure 5-7, a strict adherence to somatotopy was not obvious. There was considerable overlap in their representations, and this overlap occurred partially at the level of the efferent zone as demonstrated in Figure 5-8. A somatotopic organization was more apparent in viewing a map of the muscles of the face alone as shown in Figure 5-9. Some features

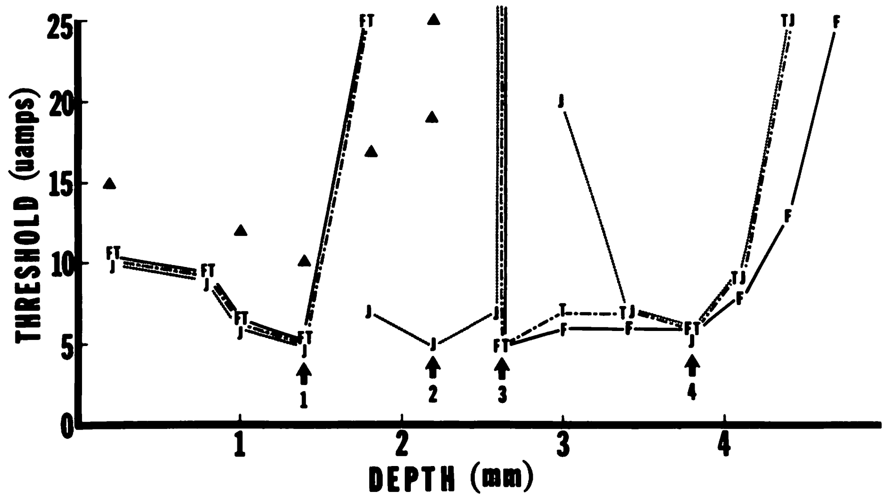


FIGURE 5-8

Threshold profiles for responses obtained during the penetration indicated by the arrow in Figure 5-7. Four zones were encountered as indicated by the arrows in this figure. The first was a face, tongue, jaw (and possibly laryngeal) zone and was the one plotted in Figure 5-7.. The second was a jaw deviation zone; the third a face, tongue zone; and the fourth a face, tongue, jaw zone.

in this organization were mentioned previously (see Figure 5-5). There were two representations of the face, one in an anterior field and one in a posterior field. In Figure 5-9 the boundary between the fields was just superior to the sulcus which terminated in the subcentral dimple. Within each field, lateral muscles of the face (e.g. zygomaticus) were represented more laterally or externally along the gyrus, while medial muscles were represented more medially, near the middle of the gyrus. A second feature was that individual face muscles or parts of face muscles were represented repeatedly within each field. Finally, a very obvious feature was that the two fields were distinguished in the frequency with which particular face muscles were represented. The posterior field allotted disproportionate space for representation of ipsilateral (open symbols) lateral muscles, while the anterior field allotted disproportionate space for representation of contralateral (solid symbols) medial muscles.

It was not certain whether muscles of the tongue also exhibited a somatotopic spatial organization, since it was not possible to observe contractions of individual tongue muscles in these studies. However, if tongue actions elicited by microstimulation were plotted, a meaningful organization became apparent, particularly if the actions were viewed in relation to associated face responses. In figure 5-10, tongue actions were plotted for each zone where there were also associated face responses. Within the posterior face field, zones controlling lateral face muscles of the ipsilateral side (see Figure 5-9) tended to produce deviation of the tongue (Figure 5-10, symbol "D") towards the same ipsilateral side (open symbols). In contrast, within the anterior face field, zones controlling midline face muscles on the contralateral side (e.g. M,R) tended to produce protrusion (P) or flipping (F) of the tongue towards the same side (solid symbols). Thus, a zone tended to produce or control tongue movement "towards" the particular face muscle also controlled by the zone. An ap-

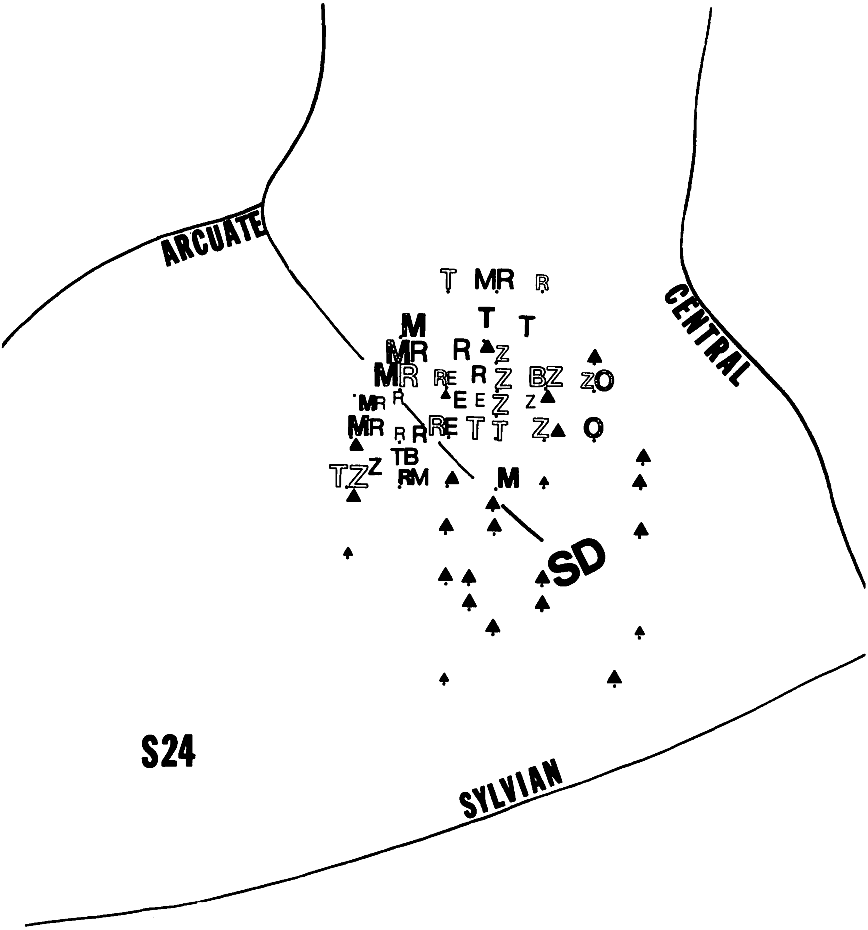


FIGURE 5-9
Face zones plotted alone with the particular muscle responding indicated. Z: Zygomaticus; T: Triangularis; B: Buccinator; M: Mentalis; R: Orbicularis oris; E: Levator labii; O: Orbicularis oculi. Point of penetration indicated by a dot for superficial zone. Position of deeper zones approximated from histology sections. Open letters: Ipsilateral face muscles responded; Solid letters: Contralateral face muscles responded; Stipled letters: Bilateral face muscles responded. Three sizes of symbols indicate relative response strengths.

parent exception to this relation is signified by the two tongue retraction zones (R) where the ipsilateral zygomaticus was also controlled. On the other hand, these zones were distinguished from other zygomaticus zones in that laryngeal muscles were also controlled. Possibly the zones produced tongue retraction "towards" the larynx?

It became clear to us in analyzing the responses produced by microstimulation that the particular combinations of responses controlled by single zones was not arbitrary but had

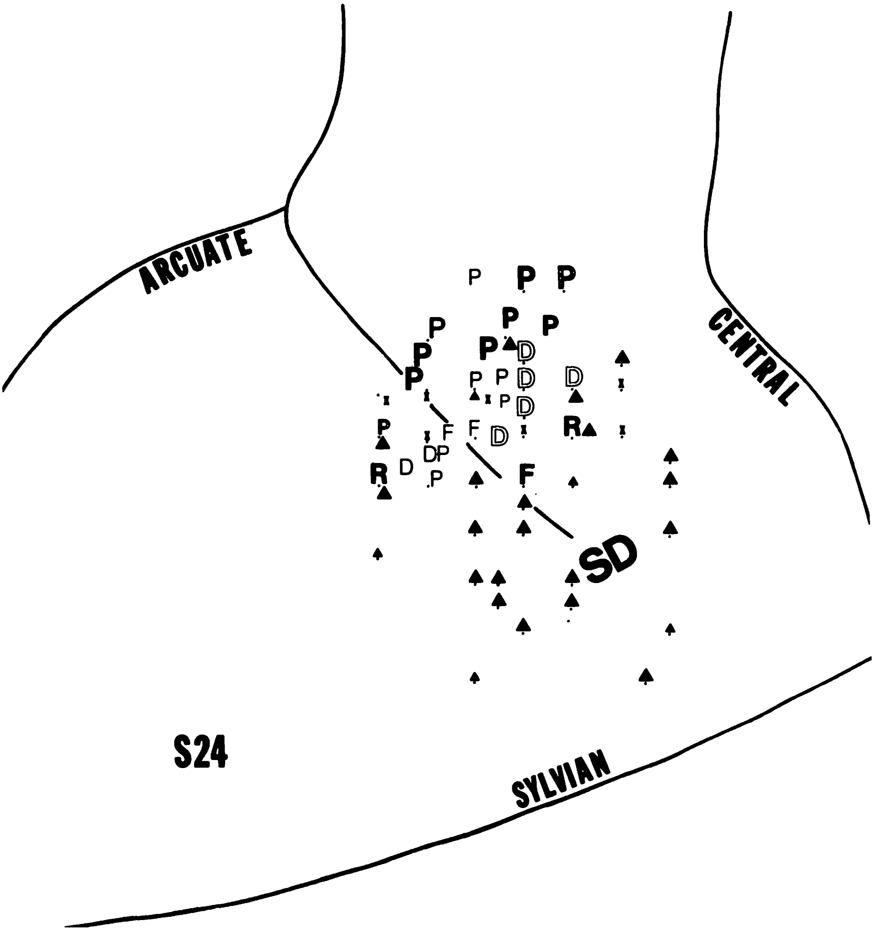


FIGURE 5-10
Tongue actions which had associated face responses within zones. D: Deviation; P: Protrusion; R: Retraction: F: Tip flick. Open letters indicate that the action occurred towards the ipsilateral side, solid letters that the action occurred towards the contralateral side, and stipled letters that the action did not occur to either side. "x" no tongue response.

functional significance. This point became clearer in sessions conducted with the last animal in which single unit recordings were made in identified zones while the animal was feeding or vocalizing. For example, many of the units recorded from ipsilateral tongue deviation-zygomaticus zones in the posterior face field had sensory input from both the ipsilateral side of the tongue and skin overlying the zygomaticus. These units also responded vigorously when the animal rolled his tongue to the ipsilateral side and contracted that side of the face to clear food from the vestibule into the oral cavity. In contrast, units recorded from tongue protrusion-contralateral orbicularis oris zones in the anterior face field had sensory input from more midline aspects of the face, tongue, and/or teeth. Furthermore, these units responded during the initial contacting of food with the lips and tongue. Thus, the responses produced by microstimulation at a site appeared to have functional significance and reflected the kinds of orofacial movements governed by neurons at that site during food manipulation or communication. Although further analysis of single unit activity is required, neurons in the lateral face area appeared to be more involved in controlling orofacial movements during food manipulation than during communication.

In summary, although there are regions of efferent zones controlling single muscles, there is not a strict adherence to somatotopy in the representation of muscles of the face, tongue, jaw, and larynx in the lateral face area. A basic principle underlying the functional organization of this area is that muscles are repeatedly represented across the cortex in fields which overlap one another in a specific manner and related to the kinds of muscle synergies required for particular movements. The data from these studies further suggest that the overlap partially occurs at the elemental level of the efferent zone.

REFERENCES

Asanuma, H. and Sakata, H. (1967). "Functional organization of a cortical efferent system examined with focal depth stimulation in cats," *J. Neurophysiol.* 30, 35-54.

Asanuma, H. (1975). "Recent developments in the study of the columnar arrangement of neurons within the motor cortex," *Physiol. Rev.* 55, 143-156.

Hast, M.H., Fischer, J., Wetzel, A. and Thompson, V. (1974). "Cortical motor representation of the laryngeal muscles in macaca mulatta," *Brain Res.* 73, 229-240.

Hubel, D. and Wiesel, T. (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol. Lond.* 160, 106-154.

Huber, E. (1933). "The facial musculature and its innervation," in *The Anatomy of the Rhesus Monkey,* C. Hartman and W.L. Strauss (Eds.), William and Wilkins, Baltimore.

Jurgens, U. (1979). "Neural control of vocalization in nonhuman primates," in *Neurobiology of Social Communication in Primates,* Academic Press, New York, 11-44.

Larson, C. (1983). "Preliminary observations on cortical and brainstem mechanisms of laryngeal control," in *Vocal Fold Physiology,* D. Bless and J. Abbs (Eds.), College-Hill Press, San Diego, CA, 82-95.

McGuiness, E. (1980). "Organization of the face representation in macaque motor cortex," *J. Comparative Neurol.* 193, 591-608.

Mountcastle, V. (1957). "Modality and topographic properties of single neurons of cat's somatic sensory cortex," *J. Neurophysiol.* 20, 408-434.

Penfield, W. (1938). "The cerebral cortex in man," *Arch. Neurol. Psych.* 40, 417-442.

Penfield, W. and Rasmussen, T. (1949). "Vocalization and arrest of speech," *Arch. Neurol. Psych.* 61, 21-27.

Stoney, S., Thompson, W. and Asanuma, H. (1968). "Excitation of pyramidal tract cells by intracortical microstimulation: Effective extent of stimulating current," *J. Neurophysiol.* 31, 659-669.

Woolsey, C., Settlage, P., Meyer, D., Sencer, W., Hamuy, T., and Travis, A. (1952). "Patterns of localization in precentral and 'supplementary' motor areas and their relation to the concept of a premotor area," *Res. Publ. Ass. Nerv. Ment. Dis.* 30, 238-263.

DISCUSSION

B. WYKE: One thing I wasn't clear about: Are the particular topographic distribution maps that you have shown true for a single set of stimulus parameters or are they consequences of the results you've got from varying the stimulus input at any given level of penetration? Because as you well know the neurons in the radial columns in this particular region of the cortex have nerve cell bodies that are of different sizes and with different thresholds and the number of motor neuron pools to which they project through their descending branches varies considerably also. So the question is, as in all cortical exploration experiments, does the map represent the picture at any particular set of stimulus parameters or is it a composite of the results obtained by a range of stimulus parameters?

D. ZEALEAR: We're really looking for the threshold for a motor response, and the threshold for that motor response will change. As you're within the superficial gray layers, you're activating more sensory cells that do feed on to those fifth layer cells that produce the motor act.

B. WYKE: If I can interrupt, the thing that I'm really worried about, as I'm always worried with these cortical electrical stimulation experiments, is do I now understand you to say that the distributions that you've shown are the distributions of the most excitable cells in the system at the time that you did the stimulation?

D. ZEALEAR: I think this question is of fundamental importance. Let me rephrase the question if I may. What factors determine how many and which neurons are activated near the electrode tip and would one get a different map if any of these factors, stimulus parameters, are changed? Although excitability may have some bearing on whether a cell is activated or not, it is probably the proximity of the cell to the electrode tip and the degree of current spread which has primary importance in activation. Current spread is, in turn, dependent upon the size of the electrode tip and the current level used. In microelectrode stimulation studies, the current spread is quite limited, since the electrode tips are quite small. In particular, our microelectrodes had 20 micron exposed tips which is smaller than an average sized cell. The current spread was also limited by using low level stimulation, just adequate for a threshold response. Furthermore, only those responses obtained with currents less than 10 μamps were plotted on the map. These responses were obtained at the minimum of a depth-threshold profile curve, most of them in layer 5, so that direct activation of pyramidal cells probably occurred. Stoney et al. (see references) has estimated that the spread of current with microstimulation below 10 μamps is sufficient to activate only a single pyramidal cell. Even if we were activating more than one cell, it seems likely that the spread of current was limited to an efferent zone. These zones are the indivisible building blocks controlling the peripheral musculature.

Would one get a different map using higher currents or bigger tipped electrodes, approaching the sizes of those used during macrostimulation? As you have suggested, we noticed that the peripheral response changed when the current level was increased sufficiently to activate neurons in more than one zone. In particular, the response area enlarged about the original focus or foci of activity. It seems likely that a map like ours would be obtained using bigger electrodes or higher currents, but the map would probably be less precise with borders between topographical regions somewhat fuzzy. However, macrostimulation serves a purpose in that it provides a quick and easy method of obtaining a first approximation of the topographical layout of the region being investigated. Microstimulation, on the other hand, can determine not only the topographical layout of the region, but also the fundamental principles upon which it is organized, that is, a description of what comprises single zones. Probably the most important finding from our study is that a significant area of this region of cortex includes zones which represent more than one muscle, and that the combination of muscles represented within a zone has functional significance.

H. HIROSE: When you are monitoring the effect of stimuli only by visual observation, you may have difficulty in determining the threshold of stimuli for jaw movement since the closer and the the opener of the jaw may be excited simultaneously. Could you comment on this point?

D. ZEALEAR: It was more difficult to determine the thresholds for jaw movement, as you have suggested, not only because the opener and closer may be simultaneously activated but also because jaw muscles cannot be inspected visually for contraction like tongue or mimetic face muscles. We might consider implanting EMG electrodes in jaw muscles in future animals, since not only can thresholds be precisely estimated but also one can obtain response latency data. In these animals we had eleven laryngeal muscles implanted with bipolar pairs.

E. LUSCHEI: Have you any evidence concerning whether or not the cortical input to the larynx, face and jaw (from the far-lateral precentral cortex) is bilaterally equivalent in the monkey? That is, are the maps of this area in the same animal bilaterally symmetrical and of roughly equivalent thresholds?

D. ZEALEAR: We don't know if the left and right far-lateral precentral cortices are identical topographically and functionally, since all of our experiments were conducted on the left cortex.

N. ISSHIKI: Penfield, a long time ago, reported that he elicited sustained phonation by electrical stimulation given on a certain portion of the cortex in awake human subjects during operations. Could you comment concerning this apparent difference between the human and the monkey?

D. ZEALEAR: We have parallel evolution going on here, humans and monkeys. It could be in part due to differences in neuronal machinery. The great apes might be a lot more like man in terms of the basic neuronal substrate at the cortical level for producing calls while the monkey is not wired that way. I'm doing this study because I think that there is some substrate similarity at the level of the cortex between the monkey and man, but the monkey has developed for different behavior really. Every animal has it's nervous system wired for what it does, and the monkey is not a vocalizer, not like we are, so that may also be an explanation for why we are seeing the kinds of responsiveness that we have noted in the monkey.

C. LARSON: I have a comment on Dr. Isshiki's question. There is anatomical confirmation by Kuypers of different projections from the cortex in humans, apes and monkeys. Humans have direct cortical projections to motoneuron pools. Chimpanzees have a few, but not many. Monkeys, cats and dogs don't appear to have any direct cortical projections to laryngeal motoneurons.

D. ZEALEAR: Just to add another comment to that. In some studies a number of years ago by Sutton and Larson, lesions in this area of cortex in the monkey produced little if any noticeable vocal deficit.

M. HIRANO: I think that cortical stimulation experiments such as this invite a lot of questions because there are so many neuronal networks involved in this type of reflex or movement. Still I think this is very relevant work. I am especially interested in the swallowing in your figure. Do you mean by swallowing, the pharyngeal stage of the swallowing movement? In other words do you mean a series of reflexes occurring in the pharynx, oral cavity and the larynx?

D. ZEALEAR: It's hard to say. You can describe it as a swallow, but it may be a patterned movement that may be just part of a swallow. In the videotape there is a demonstration of another patterned movement like pouch elevation following stimulation of white matter below a zone where we were getting contractions of the triangularis muscle, a muscle used in pouch elevation. Both of these patterned movements may be just part of a whole temporal sequence of movements involved in a particular stereotypic behavior.

M. HIRANO: It can be derived from a single stimulus?

D. ZEALEAR: Right. It's a train. I'm sorry I didn't mention my stimulus parameters. We used a 100 Hz stimulus of 200 microsecond squarewaves.

M. HIRANO: Is there a specific frequency range that is effective in the production of these "swallowing" movements? I am asking this because frequencies of 30-40 Hz are the most adequate in producing a swallowing reflex through the stimulation of the internal branch of the superior laryngeal nerve.

D. ZEALEAR: We haven't looked at the effectiveness of different frequencies in producing a "swallow". However, if there is a limited frequency range which elicits this response, it may not necessarily be the same as that found during stimulation of sensory axons in the peripheral nerve.

6 - LOCALIZATION OF THE LARYNGEAL MOTONEURONS IN THE BRAIN STEM AND MYOTOPICAL REPRESENTATION OF THE MOTONEURONS IN THE NUCLEUS AMBIGUUS OF CATS. —AN HRP STUDY

Yoshikazu Yoshida, Toshimi Miyazaki, Minoru Hirano,
Takeski Kanaseki.

INTRODUCTION

It has been accepted that the motoneurons innervating the intrinsic laryngeal muscles are located in the nucleus ambiguus of the brain stem. This fact was first demonstrated with the use of the retrograde degeneration technique by Bunzel-Federn (1899), Kosaka (1909) and Molhant (1911-1912) in the beginning of the nineteenth century. Szentágothai (1943), Koyama (1951) and Furstenberg (1955) investigated the localization of the motoneurons for the cricothyroid (CT), the abductor and the adductor muscle. Lawn (1966) studied anatomical organization of the nucleus ambiguus and determined the cytoarchitectural map. By means of the horseradish peroxidase (HRP) technique, Gacek (1975) investigated the central location of each laryngeal motoneuron for the first time.

We have determined the precise localization of not only the laryngeal motoneurons but also the other motoneurons located in the nucleus ambiguus (Yoshida *et al* 1980, 1981, 1982). In addition, we have studied the myotopical representation of the motoneurons in the nucleus of cats utilizing the HRP technique. This paper will describe results of our recent studies.

MATERIALS AND METHODS

Adult cats (2.7-3.5kg) were anesthetized with Ketamine (30-40mg/kg i.m.) and Xylazine hydrochloride (0.5-1.5mg/kg). After exposing the muscle to be investigated, 1-5μl of 30% HRP (Toyobo Grade I-C) dissolved in sterile 0.9% saline was injected manually into the belly of the muscle with a microsyringe under an operation microscope. In some cats, HRP was injected in the recurrent laryngeal nerve or in the inferior ganglion of the vagus nerve (Table 6-1). After survival periods of 24-72 hours, the cats were sacrificed for the determination of the location of motoneurons. Details pertaining to the procedures of HRP injection, perfusion, tissue fixations, section cutting and mounting have been described elsewhere (Yoshida *et al* 1980). 3.3'-dimethoxybenzidine dihydrochloride or Benzidine dihydrochloride was adopted for the reaction for identification of peroxidase activity. In the case of the former, cresyl violet and in the case of the latter, neutral red were used for the counterstaining.

RESULTS AND DISCUSSION

Summary of results after injection of HRP into each muscle is shown in Table 6-2. Figure 6-1 schematically demonstrates the location of the cell column for each muscle, the nodose ganglion and the recurrent laryngeal nerve with respect to the level in the rostrocaudal direc-

TABLE 6-1. Subjects for an HRP stydy

Subjects				Location HRP Injected	volume HRP injected (μl)
Muscle	Glosso-pharyngeal nerve	Pharyngeal muscle	Stylopharyngeal muscle (STP)	muscle belly	1-3
	Vagus nerve	Pharyngeal constrictor muscles	Cephalopharyngeal muscle (Cep)	rostral portion of Passavant ridge	2-5
			Hyopharyngeal muscle (HP)	muscle belly	2-5
			Thyropharyngeal muscle (TP)		
			Cricopharyngeal muscle (CP)		
		Esophageal muscles	Cervical esophagus muscle (CE)	level of Thyroid gland	2-10
			Middle part of thoracic esophagus muscle (ME)	level of caudal portion of Aortic arch	
			Lower part of thoracic esophagus muscle (LE)	level of immediate rostral to Diaphragm	
		Soft palate	Levator muscle of velum palatinum (LVP)	soft palate	1-4
		Intrinsic laryngeal muscles	Cricothyroid muscle (CT)	muscle belly	1-5
			Posterior cricoarytenoid muscle (PCA)		
			Thyroarytenoid muscle (TA)		
			Lateral cricoarytenoid muscle (LCA)		
			Interarytenoid muscle (IA)		
Nerve	Nodose ganglion (NG)			ganglion	10-15
	Recurrent laryngeal nerve (RLN)			level of caudal portion of Thyroid gland	5

tion. Figure 6-2 schematically shows the outline of the somatotopic arrangement of the labeled motoneurons in the transverse plane of the nucleus ambiguus at the three levels which are indicated in Figure 6-1.

Laryngeal Muscles

Following HRP injection into the cricothyroid (CT), posterior cricoarytenoid (PCA), thyroarytenoid (TA), lateral cricoarytenoid (LCA) and interarytenoid muscles (IA), the labeled motoneurons were observed in the ipsilateral nucleus ambiguus exclusively. In the present paper, the term "nucleus ambiguus" includes the retrofacial nucleus. The motoneurons of CT were recognized in the rostral part of the nucleus ambiguus whereas those of the other laryngeal muscles were identified in the caudal half.

CT: The motoneurons of CT showed compact form and were located in the ventral part of the nucleus ambiguus (Figure 6-3). Their location extended from the level immediate caudal to the facial nucleus to the level of the rostral end of the hypoglossal nucleus.

PCA: The motoneurons of PCA were aggregated and occupied the middle part of the nucleus (Figure 6-3). The labeled cell column was found between the level of the caudal end of the CT motoneurons and the level where the central canal develops. The levels of the labeled cell column of PCA, TA, LCA and IA differ slightly from each other and become more caudal in the order of PCA, TA, LCA and IA.

TA: The motoneurons of TA were scattered and were seen in the dorsal part of nucleus (Figure 6-3).

LCA, IA: The motoneurons of LCA and IA were sparse and were scattered throughout in the nucleus (Figure 6-3). The caudal end of the IA cell column was recognized at the level of the caudal end of the inferior olivary nucleus.

According to Bunzel-Federn (in rabbits, 1899), Kosaka (in dogs, 1909) and Molhant (in rabbits, 1911-1912), the motoneurons supplying the superior laryngeal nerve were located in the compact group of nerve cells of the rostral part of the nucleus ambiguus while those innervating the inferior laryngeal nerve were situated in the loose group of nerve cells of the caudal part. Szentágothai (in cats and dogs, 1943), Koyama (in cats, 1951) and Furstenberg (in a monkey, 1955) showed that the location of the motoneurons became more caudal in the order of CT, PCA, TA, LCA and IA in the rostrocaudal pattern of innervation in the nucleus ambiguus. Lawn (1966) stated that the medial column innervates CT and the cells of the diffuse formation of the caudal part of the principal column supply the remainder of the intrinsic laryngeal muscles in rabbits. All these studies were made chiefly by the retrograde degeneration method. The first investigation by means of HRP technique for the laryngeal muscles in kittens was reported by Gacek (1975). He described that the adductor neurons were located in the dorsal division, on the other hand the abductor neurons occupied the ventral division. In addition, he also stated that CT and PCA derived innervation from the retrofacial nucleus. Sugimoto *et al* (in cats, 1979), Kalia and Mesulam (in cats, 1980) and Lobera *et al* (in rats and cats, 1981) also used HRP in investigating motoneurons for the laryngeal muscles. However, they did not investigate each laryngeal muscle separately. The results of our study generally agree with previous investigations. However, our findings in regard to the location of motoneurons for PCA were different from those reported by Gacek (1975). He found the PCA motoneurons in the retrofacial and ambiguus nuclei. In our study no labeled cell was seen in such a rostral position as that in Gacek's report. Lobera *et al* (1981) stated that a possible explanation for this difference is enzyme spread in the kitten to pharyngeal muscles, therefore labeling their motoneurons located in the rostral part of the nucleus ambiguus. We determined the geographic relationship between the motoneurons innervating the pharyngeal, esophageal and laryngeal muscles. Among the laryngeal muscles, the motoneurons of CT are located much more rostral than those of the others. They are closer to the motoneurons

muscle HRP injected		HRP labeled neurons			
				location in the nucleus	
group	muscle	nucleus	side	rostrocaudal direction	transverse plane
intrinsic laryngeal muscles	CT	AM	ipsilat.	caudal 2/3 in rostral half	ventral
	PCA			caudal half	middle
	TA				dorsal
	LCA				widely
	IA				
pharyngeal constrictor muscles	CeP	AM	ipsilat.	rostral half	dorsomedial
	HP				
	TP			caudal 2/3 in rostral half	dorsomedial and ventrolateral partical
	CP				
esophageal muscles	CE	AM	ipsilat.	caudal 2/3 in rostral half	ventrolateral
	ME	AM	bilat.	caudal 2/3 in rostral half	lateral
		DMV		rostral most	ventral
	LE	AM		caudal 2/3 in rostral half	dorsolateral
		DMV		rostral most	ventral
stylopharyngeal muscle	STP	AM	ipsilat.	rostral 1/3 in rostral half	dorsolateral part of FN
soft palate	LVP	AM	ipsilat.	middle 1/3	middle

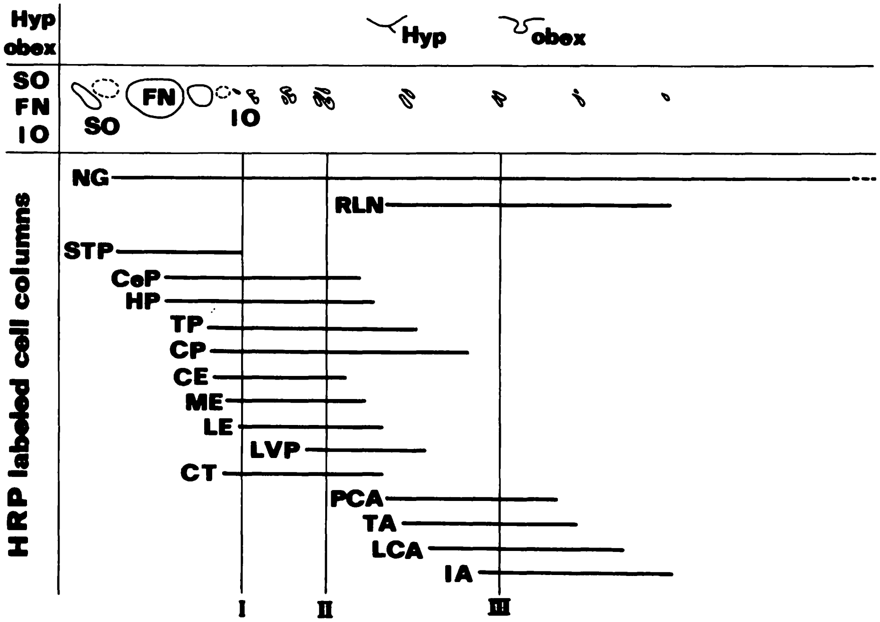


FIGURE 6-1
A diagram which draws schematically the level of the labeled cell column for the nodose ganglion, the recurrent laryngeal nerve and the muscles subjected in the rostrocaudal direction.
The level in the brain stem is indicated with the shape of the hypoglossal nucleus (Hyp), obex, the superior olivary nucleus (SO) and the inferior olivary nucleus (IO).
NG: the nodose ganglion, RLN: the recurrent laryngeal nerve, STP: the stylopharyngeal muscle, CeP: the cephalopharyngeal muscle, HP: the hyopharyngeal muscle, TP: the thyropharyngeal muscle, CP: the cricopharyngeal muscle, CE: the cervical esophagus muscle, ME: the middle part of thoracic esophagus muscle, LE: the lower part of thoracic esophagus muscle, LVP: the levator muscle of velum palatinum, CT: the cricothyroid muscle, PCA: the posterior cricoarytenoid muscle, TA: the thyroarytenoid muscle, LCA: the lateral cricoarytenoid muscle, IA: the interarytenoid muscle.

TABLE 6-2
Summary of Results. The term "nucleus ambiguus" includes the retrofacial nucleus.

AM: the nucleus ambiguus, DMV: the dorsal motor vagal nucleus, FN: the facial nucleus, CT: the cricothyroid muscle, PCA: the posterior cricoarytenoid muscle, TA: the thyroarytenoid muscle, LCA: the lateral cricoarytenoid muscle, IA: the interarytenoid muscle, CeP: the cephalopharyngeal muscle, HP: the hyopharyngeal muscle, TP: the thyropharyngeal muscle, CP: the cricopharyngeal muscle, CE: the cervical esophagus muscle, ME: the middle part of thoracic esophagus muscle, LE: the lower part of thoracic esophagus muscle, STP: the stylopharyngeal muscle, LVP: the levator muscle of velum palatinum.

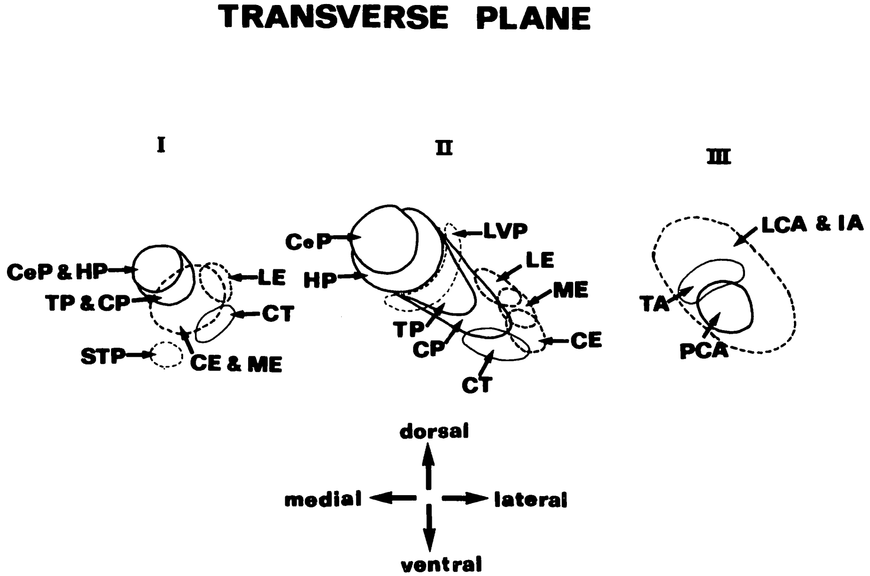


FIGURE 6-2

Outline of the somatotopic arrangement of the labeled motoneurons for each muscle in transverse plane of the nucleus ambiguus at three levels (I, II and III) which are indicated in Fig. 6-1. I: at the level of the rostral portion of the inferior olivary nucleus. II: at the level where the principal nucleus of the inferior olivary nucleus develops well. III: at the level of obex. STP: the stylopharyngeal muscle, CeP: the cephalopharyngeal muscle, HP: the hyopharyngeal muscle, TP: the thyropharyngeal muscle, CP: the cricopharyngeal muscle, CE: the cervical esophagus muscle, ME: the middle part of thoracic esophagus muscle, LE: the lower part of thoracic esophagus muscle, LVP: the levator muscle of velum palatinum, CT: the cricothyroid muscle, PCA: the posterior cricoarytenoid muscle, TA: the thyroarytenoid muscle, LCA: the lateral cricoarytenoid muscle, IA: the interarytenoid muscle.

of the pharyngeal muscles than to those of the other laryngeal muscles (see Figure 6-1). This is understandable because CT as well as the inferior pharyngeal constrictor is phylogenetically derived from the fourth branchial arch.

Pharyngeal Constrictor Muscles

After HRP injection into the cephalopharyngeal (CeP), hyopharyngeal (HP), thyropharyngeal (TH) and cricopharyngeal muscles (CP), the labeled neurons were seen exclusively in the rostral part of the ipsilateral nucleus ambiguus.

CeP and HP: The motoneurons of CeP and HP were scattered in the rostral part of their cell column and located in the lateral part of the nucleus ambiguus whereas they were aggregated and occupied the dorsomedial part of the nucleus in the caudal part. The rostral end of the labeled cell column for CeP and HP was located at the level of and lateral to the middle of the facial nucleus.

TP and CP: The motoneurons of TP and CP occupied a larger part of the nucleus than those of CeP and HP. The rostral end of the labeled cell column for both muscles was

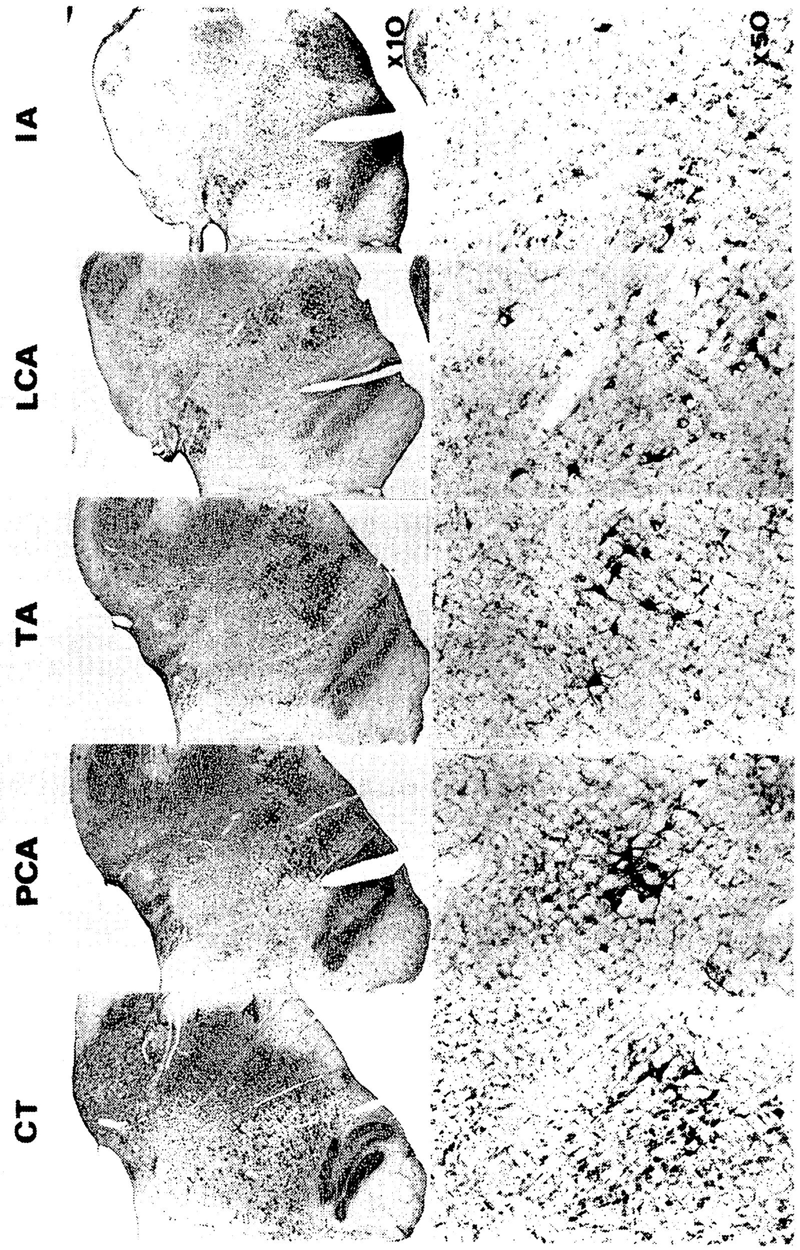


FIGURE 6-3

Photomicrographs of cross sections of the brain stem of cats showing the location of the HRP labeled cells innervating the cricothyroid (CT), the posterior cricoarytenoid (PCA), the thyroarytenoid (TA), the lateral cricoarytenoid (LCA) and the interarytenoid (IA) muscles. Upper line, x10; lower line, x50. (By Y. Yoshida, *Acta Otol.* 1982.)

found at the level immediately caudal to the facial nucleus. The level of the caudal end of the labeled cell column became more caudal in the order of CeP, HP, TP and CP.

Kosaka (1909) described for the first time that the pharynx of dogs is innervated by the motoneurons of the rostral part of the nucleus ambiguus. Molhant (1911-1912) demonstrated that the nucleus ambiguus of rabbits consisted of three segments: the compact, the semicompact and the loose formation. He also stated that the pharyngeal branch of the vagus nerve is supplied by both the compact and the semi-compact formation. According to Lawn (1966), the motoneurons of the pharynx of rabbits was located in the principal column of the nucleus ambiguus. Our results basically agree with these previous reports. However they did not differentiate the motoneurons of each of the pharyngeal constrictor muscles.

Esophageal Muscles

Following HRP injection into the cervical esophagus muscle (CE), labeled neurons were located in the rostral part of the ipsilateral nucleus ambiguus exclusively. After HRP injection into the middle part (ME) and lower part of the thoracic esophagus muscle (LE), labeled neurons were found in the rostral part of the nucleus ambiguus and the dorsal motor vagal nucleus (DMV) bilaterally with an ipsilateral predominance.

CE: The motoneurons of CE occupied a small ventrolateral portion of the nucleus ambiguus. The rostral end of the labeled cell column for CE was found at the same level as that of TP and CP, whereas the caudal end was located at the level slightly caudal to the well developed portion of the principal nucleus of the inferior olivary nucleus.

ME and LE: The motoneurons of ME and LE were identified in the ventrolateral part of the nucleus ambiguus and the ventral portion of DMV. Within the ventrolateral part of the nucleus ambiguus, the motoneurons were found more ventral in the order of LE, ME and CE. In the ventral portion of DMV, there was no difference in the location between ME and LE motoneurons. The labeled cell column of ME was located more rostrally than that of LE and they were situated more caudally than that of CE. The rostral end of the labeled cell column of ME was found at the level of the rostral end of the inferior olivary nucleus. The caudal end of the cell column of LE was observed at the level of the rostral portion of the hypoglossal nucleus.

There are two muscle layers in the esophagus of cats: the inner circular and outer longitudinal layers. Both layers consist of striated muscles in their rostral portion whereas they are composed of smooth muscles in their caudal portion. The transition from striated to smooth muscle takes place in the caudal third for the inner circular layer and at the level of the diaphragm for the outer longitudinal layer (Yanohara 1955). Therefore, the motoneurons labeled in the nucleus ambiguus in the present investigation are presumably those for the striated muscles whereas the motoneurons recognized in DMV are those for the smooth muscles. Kosaka (in dogs, 1909), Molhant (in rabbits, 1911-1912) and Lawn (in rabbits, 1966) reported that the esophageal motoneurons were situated in the rostral part of the nucleus ambiguus. Kosaka (1909) and Molhant (1911-1912) also demonstrated the existence of differences in the location of the motoneurons for different segments of the esophagus. They also described that some motoneurons of the esophagus muscle were located in DMV. Our results agree with theirs. Getz and Sernes (1949) stated in their investigation of DMV that the most caudal portion of DMV in rabbits innervates thoracic organs such as the trachea, the bronchi and the esophagus. However, no labeled cells were found in such a caudal location of DMV in our investigation.

Stylopharyngeal Muscle (STP)

STP motoneurons labeled with HRP were only found in the most rostral part of the nucleus ambiguus ipsilaterally. Their location extended from the level immediately caudal to the superior olivary nucleus to the level of the rostral portion of the inferior olivary nucleus. The motoneurons in the rostral part of the labeled cell column were scattered and occupied the dorsolateral part of the facial nucleus but were not within the facial nucleus itself while those in the caudal part were clustered and occupied the ventral part of the nucleus ambiguus.

The motoneurons of the glossopharyngeal nerve were studied by Molhant (in rabbits, 1911-1912), Szabo and Dussardier (in sheep, 1964), Lawn (in rabbits, 1966) and Satomi *et al* (in cats, 1979). Molhant (1911-1912) divided the nucleus ambiguus into three segments as mentioned before. He reported that the glossopharyngeal motoneurons were situated in the rostral part of the compact formation of the nucleus ambiguus. On the investigation of the cells of origin of the vagus nerve, Szabo and Dussardier (1964) stated that the nucleus ambiguus was separated into three groups: *group principal*, *group accessorie* and *group disperse*. In their description, the motoneurons supplying the glossopharyngeal nerve were located in the ventral part of the rostral of *group principal*. Lawn (1966) described that the rostral part of the medial column of the nucleus ambiguus innervates the glossopharyngeal nerve. According to Satomi *et al* (1979), following HRP bathings of the transected glossopharyngeal nerve root, the labeled cells were identified in the nucleus ambiguus. He guessed that the HRP labeled neurons were STP motoneurons. Our results agree with the previous investigations.

Levator Muscle of Velum Palatinum (LVP)

Following HRP injection into the soft palate, labeled neurons were recognized ipsilaterally in the rostral part of the nucleus ambiguus exclusively. According to Crouch (1969), the musculature of the soft palate of cats consists of levator and tensor veli palatini muscles. M. uvulae and M. palatopharyngeus are not described in his text book. M. tensor veli palatini is innervated by a branch of the trigeminal nerve. Therefore, HRP positive cells should be the motoneurons of LVP. LVP motoneurons extended between the level slightly rostral to the well developed portion of the principal nucleus of the inferior olivary nucleus and the level slightly caudal to the rostral part of the hypoglossal nucleus. They were situated in the central part of the nucleus.

Kosaka (1909) reported that the soft palate of dogs is innervated by the middle part of the nucleus ambiguus. In Lawn's (1966) cytoarchitectural map, the motoneurons of the soft palate of rabbits were located in the principal column of the nucleus ambiguus. And they were situated more caudally than that of the esophageal motoneurons. These previous investigations agreed with our results in general. The present investigation demonstrates the precise location of the motoneurons of LVP.

Recurrent Laryngeal Nerve (RLN)

After applying HRP to RLN, most of the labeled neurons were found in the caudal part of the ipsilateral nucleus ambiguus. They extended from the level of the rostral end of the hypoglossal nucleus, that is, the same level as the rostral end of PCA motoneurons, to the level of the caudal end of the inferior olivary nucleus, which is identical to the level of the caudal end of IA motoneurons. Some labeled neurons were found in the ipsilateral dorsal motor vagal nucleus (DMV) at the same level. The RLN motoneurons greatly increased in

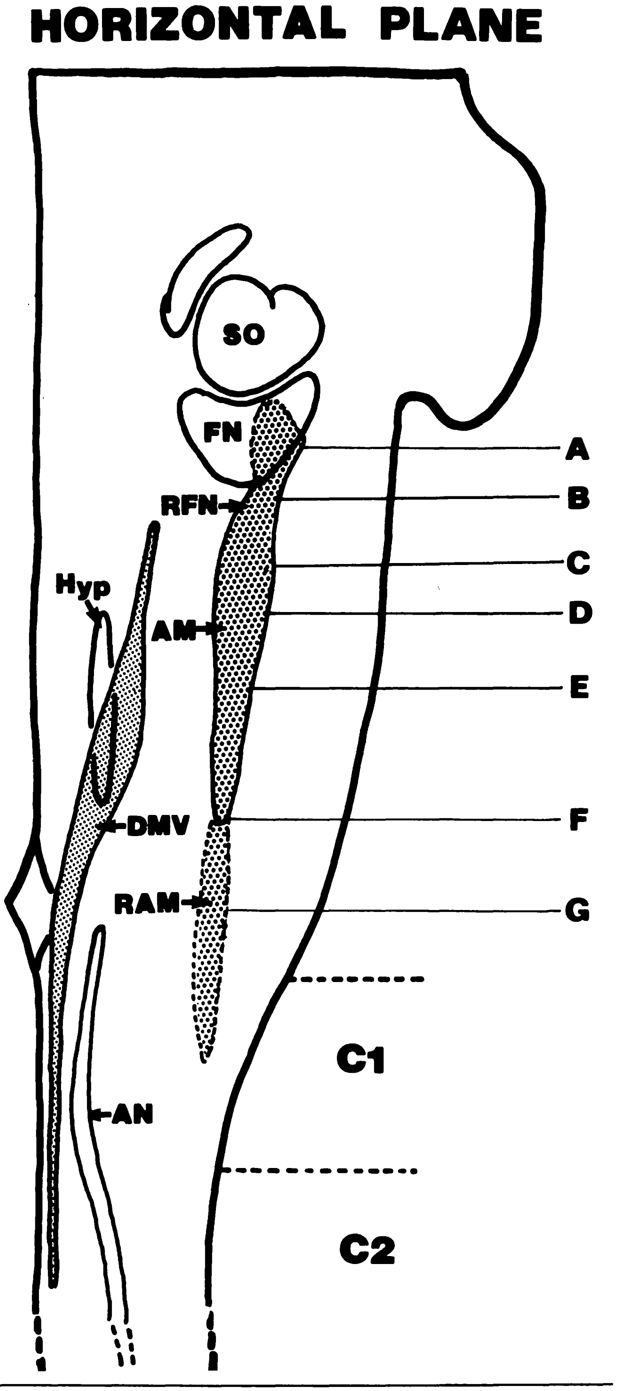


FIGURE 6-4
Outline drawing of the location of HRP labeled cells in the caudal brain stem and the cervical cord of the cats following HRP injection of the nodose ganglion. A-G indicate the level of the photomicrographs in Figure 6-5. SO: the superior olivary nucleus, FN: the facial nucleus, RFN: the retrofacial nucleus, AM: the nucleus ambiguus, Hyp: the hypoglossal nucleus, DMV: the dorsal motor vagal nucleus, RAM: the retroambigualis nucleus, AN: the accessory nucleus. (by T. Miyazaki, *Otologia* Fukuoka, 1982.)

number at the level slightly rostral to the obex and showed aggregated form in the ventromedial and the central portions and sparse form in the other portion of the nucleus.

Nodose Ganglion (NG)

When HRP was injected into NG, two cell columns extending from the lower brain stem to the upper cervical spinal cord were identified on the ipsilateral side (Figure 6-4). One was the nuclei ambiguus et retroambigualis located laterally and the other the dorsal motor vagal nucleus located medially. The former extended from the level of the caudal end of the superior olivary nucleus to Cl and the latter from the level of the rostral end of the inferior olivary nucleus to C2 (Figure 6-4). The rostral end of the cell column of the nucleus ambiguus was located at the same level as the rostral end of the STP neurons. The arrangements of the motoneurons in the nucleus at each level of A-G which is indicated in Figure 6-4 is as follows.

A. At the level of the rostral portion of the nucleus ambiguus and also at the level of the middle part of the facial nucleus.
The motoneurons of the cell column of the nucleus ambiguus were scattered in the outer area dorsal and lateral to the facial nucleus, but never observed within the facial nucleus (Figure 6-5A).

B. At the level of the rostral part of the inferior olivary nucleus.
The labeled motoneurons of the nucleus ambiguus were slightly aggregated in the central part and scattered in the other part of the nucleus. A small number of labeled motoneurons of DMV also were recognized in the most rostral part of DMV (Figure 6-5B).

C. At the level of the well-developed portion of the principal nucleus of the inferior olivary nucleus.
The greatest number of motoneurons of the nucleus ambiguus were at this level and could be divided into two groups: the small compact group of small cells occupying the dorsomedial part and the large scattered group of larger cells located in the ventrolateral part of the nucleus. The motoneurons of DMV showed an aggregated cluster in the large region of the nucleus (Figure 6-5C).

D. At the level of the rostral portion of the hypoglossal nucleus.
The motoneurons of the nucleus ambiguus were scattered throughout the nucleus and they appeared as large multipolar cells. The motoneurons of DMV at this level increased in number and were situated near the fourth ventricle (Figure 6-5D).

E. At the level of the rostral portion of the obex.
The labeled motoneurons of the nucleus ambiguus were aggregated in the ventromedial and the central portion and sparse in the other portion of the nucleus. The labeled motoneurons of DMV were located close to the fourth ventricle and formed an oval shape (Figure 6-5E).

F. At the level of the caudal end of the inferior olivary nucleus.
The motoneurons of the nucleus ambiguus were sparse but located throughout the nucleus. The DMV motoneurons were observed close to the central canal (Figure 6-5F). The transition from the nucleus ambiguus to the nucleus retroambigualis takes place at this level.

G. At the level of the pyramidal decussation.
The retroambigual motoneurons were situated in the dorsolateral region of the ventral horn and were small in comparison with the ambigual motoneurons. The DMV motoneurons appeared in a compact shape of the nucleus (Figure 6-5G).

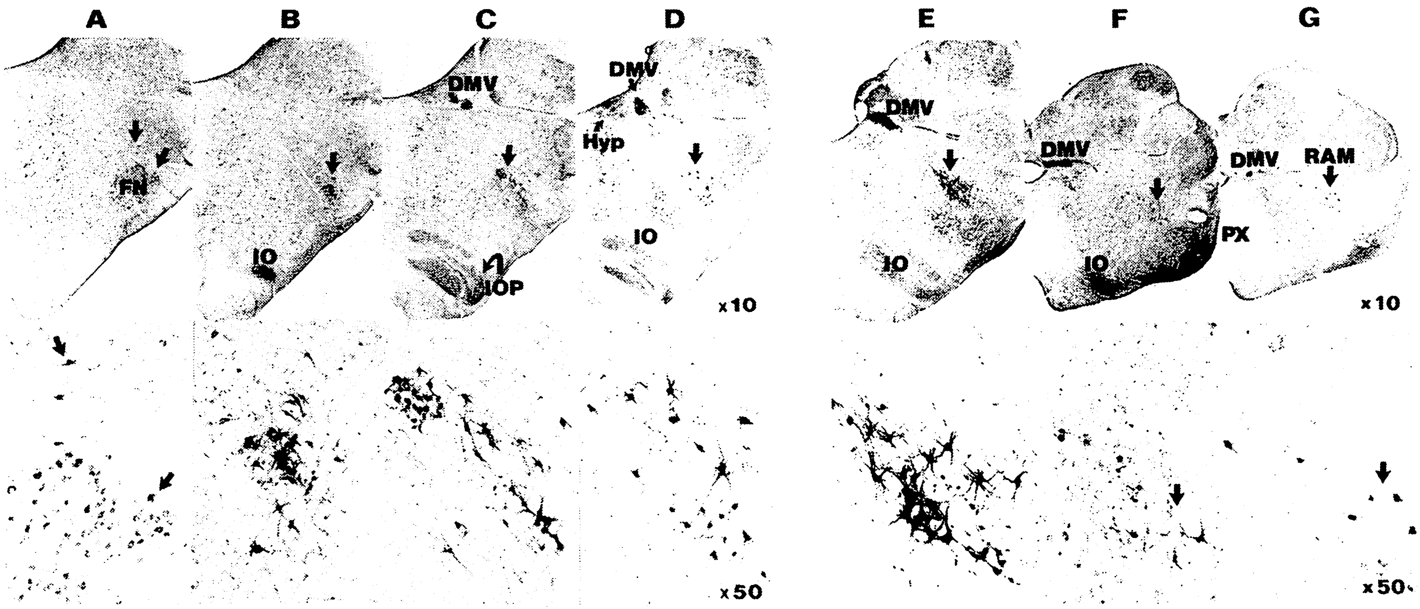


FIGURE 6-5
Photomicrographs of cross sections of the brain stem of cats demonstrating the location of labeled cells following HRP injection of the nodose ganglion. A: the level of the rostral portion of the nucleus ambiguus and also at the level of the middle part of the facial nucleus. B: the level of the rostral part of the inferior olivary nucleus. C: the level of the well-developed portion of the principal nucleus of the inferior olivary nucleus. D: the level of the rostral portion of the hypoglossal nucleus. E: the level of the rostral portion of the obex. F: the level of the caudal end of the inferior olivary nucleus. G: the level of the pyramidal decussation. FN: the facial nucleus, IO: the inferior olivary nucleus, IOP: the principal nucleus of the inferior olivary nucleus, DMV: the dorsal motor vagal nucleus, Hyp: the hypoglossal nucleus, PX: the pyramidal decussation. (By T. Miyazaki, *Otologia*. Fukuoka, 1982.)

TABLE 6-3
Somatotopic arrangement of the nucleus ambiguus in cats.

rostrocaudal direction	density of cells	arrangement of cells	muscle innervated
rostral 1/6	loose	loose, dorsolateral of FN	most of STP
		loose, lateral of FN	a part of CeP and HP
middle 2/6	compact	large cell group (scattered, occupying ventrolateral large region of the nucleus)	CT, CE, ME, LE a part of TP and CP
		compact, small cell group (compact, occupying dorsomedial small region of the nucleus)	CeP, HP, LVP, TP, CP
caudel 3/6	loose	aggregated, middle part of the nucleus	PCA, TA
		sparse, widely in the nucleus	LCA, IA

FN: the facial nucleus, STP: the stylopharyngeal muscle, CeP: the cephalopharyngeal muscle, HP: the hyopharyngeal muscle, CT: the cricothyroid muscle, CE: the cervical esophagus muscle, ME: the middle part of thoracic esophagus muscle, LE: the lower part of thoracic esopahgus muscle, TP: the thyropharyngeal muscle, CP: the cricopharyngeal muscle, LVP: the levator muscle of velum palatinum, PCA: the posterior cricoarytenoid muscle, TA: the thyroarytenoid muscle, LCA: the lateral cricoarytenoid muscle, IA: the interarytenoid muscle.

Myotopical Representation of the Motoneurons in the Nucleus Ambiguus of Cats

On the basis of the present results, we defined the myotopical representation of the motoneurons of the nucleus ambiguus. The exact extent of this longitudinal cell column was from the level immediately caudal to the superior olivary nucleus to the caudal end of the inferior olivary nucleus. The cell column was approximately 5.5-6 mm long and was divided into three parts cytoarchitecturally; i.e. the rostral loose formation, the middle compact formation and the caudal loose formation. The rostral loose formation, which was the rostral one-sixth, contained most of the STP motoneurons and a part of CeP and HP motoneurons. The middle compact formation which was the next two-sixths, was composed of two groups of cells; i.e. a large cell group and a compact small cell group. The former, occupying a large region of the nucleus, was located ventrolaterally and innervated the esophageal muscle and CT. The latter was situated in the dorsomedial region and contributed CeP, HP and LVP. Both of them supplied TP and CP. The caudal loose formation, which was the caudal half, connected the groups of the middle compact formation and contained PCA, TA, LCA and IA motoneurons (Table 6-3).

Regarding the extent of the nucleus ambiguus in the rostrocaudal direction, Yoda (in cats, 1940), Olszewski *et al* (in humans, 1954) and Taber (in cats, 1961) reported that the rostral end of the cell column was the caudal end of the facial nucleus whereas the caudal end was the level of the caudal end of the inferior olivary nucleus. On the other hand Szabo and Dussardier (in sheep, 1964) described that the cell column of the nucleus ambiguus was found between the level of the facial nucleus and the level of the caudal end of the inferior olivary nucleus. And Lawn (in rabbits, 1966) stated that the location of the nucleus ambiguus extended from the caudal end of the facial nucleus to the level of a border between the medulla oblongata and the cervical cord. In respect to the location of the caudal end of the nucleus ambiguus, our results agree with Yoda (1940), Olszewski *et al* (1954), Taber (1961), Szabo and Dussardier's (1964), description. However, in our investigation, the rostral end of the nucleus ambiguus was identified at the level of the caudal end of the superior olivary nucleus and also at the level of the middle part of the facial nucleus. Lawn (1966) studied the localization of the motoneurons in the nucleus ambiguus of rabbits by means of retrograde degeneration. According to Lawn, the nucleus ambiguus of rabbits contained about 1500 cells and was approximately 4-5 mm long. It was divided into a rostral compact formation and a caudal diffuse formation. He further subdivided the compact formation into a lateral part, the principal column, and a medial part, the medial column. And the principal column of the compact formation contained the motoneurons for the esophagus, pharynx and soft palate. The rostral part of the medial column formed a separate group of cells which supply STP. The remainder of the medial column innervated CT. The cells of the caudal diffuse formation supplied the remainder of the intrinsic muscles of the larynx. Our results concerning the somatotopical arrangement of the motoneurons of the nucleus ambiguus basically agree with Lawn's description.

CONCLUSION

Following HRP injection into the laryngeal, pharyngeal constrictor, esophageal, stylopharyngeal muscles and the soft palate, all of the motoneurons labeled with HRP were found in the nucleus ambiguus ipsilaterally except for some labeled neurons of ME and LE. Only ME and LE motoneurons were recognized in both the dorsal motor vagal and the ambigual nuclei bilaterally with ipsilateral predominance. The motoneurons of CT were situated in the ventral part of the caudal 2/3 in the rostral half of the nucleus. The HRP labeled neurons of PCA, TA, LCA and IA were observed in the caudal half of the nucleus. Within this caudal half of the nucleus PCA were located centrally, TA motoneurons dorsally and LCA and IA motoneurons widely. The motoneurons of CeP and HP were situated in the dorsomedial part of the rostral half of the nucleus, TP and CP neurons in the dorsomedial and ventrolateral part of the caudal 2/3 in the rostral half, CE motoneurons in the ventrolateral part of the caudal 2/3 in the rostral half, STP motoneurons in the outer dorsolateral part of the retrofacial nucleus of the rostral 1/3 in the rostral half and LVP motoneurons in the middle part of the middle 1/3 of the nucleus. The HRP labeled neurons of ME and LE were identified both in the caudal 2/3 in the rostral half of the nucleus ambiguus and in the ventral part of the most rostral part of the dorsal motor nucleus of the vagus. Within the caudal 2/3 in the rostral half of the nucleus ambiguus ME were located laterally, LE motoneurons dorsolaterally. For the purposes of the determination of the exact extent of the ambigual cell column, both the recurrent laryngeal nerve and the nodose ganglion were examined by means of the HRP technique.

On the basis of these investigations, we attempted to define the myotopical representation of the ambigual motoneurons. The cell column which was located between the end of the superior olivary nucleus and the end of the inferior olivary nucleus, was divided into three segments; the rostral loose formation, the middle compact formation and the caudal loose formation. The rostral loose formation, which was the rostral one-sixth, contained most of

STP motoneurons and a part of CeP and HP motoneurons. The middle compact formation which was the next two-sixths, was composed of two groups; i.e. the large cell group and the compact small cell group. The former, occupying the large region of the nucleus, was located ventrolaterally and innervated the esophageal muscle and CT. The latter was situated in the dorsomedial part and contributed CeP, HP and LVP. Both of them supplied TP and CP. The caudal loose formation, which was the caudal half, connected the groups of the middle compact formation and contained PCA, TA, LCA and IA motoneurons.

REFERENCES

Bunzel-Federn, E. (1899). "Der centrale Ursprung des N. vagus," *Monatsschr. F. Psychiat. Neurol.* 5, 1-22.

Crouch, J.E. (1969). *Text-Atlas of Cat Anatomy*. Lea and Febiger. Philadelphia.

Furstenberg, A.C. and Magielski, J.E. (1955). "A motor pattern in the nucleus ambiguus. Its clinical significance," *Ann. Otol. Rhinol. Laryngol.* 64, 788-793.

Gacek, R.R. (1975). "Localization of laryngeal motor neurons in the kitten," *Laryngoscope* 85, 1841-1861.

Getz, B. and Sernes, T. (1949). "The localization within the dorsal motor vagal nucleus," *J. Comp. Neurol.* 90, 95-110.

Kalia, M. and Mesulam, M.M. (1980). "Brain stem projections of sensory and motor components of the vagus complex in the cat: II. Laryngeal, tracheobronchial, pulmonary, cardiac and gastrointestinal branches," *J. Comp. Neurol.* 193, 467-508.

Kosaka, K. (1909). "Uber die Vaguskerne des Hundes," *Neurol. Centralbl.* 28, 406-410.

Koyama, H. (1951). "The morphological studies on the central innervation of the cat intrinsic laryngeal muscles in the nucleus ambiguus," *J. Jpn. Bronchoesophagol. Soc.* 2, 31-34.

Lawn, A.M. (1966). "The localization, in the nucleus ambiguus of the rabbit, of the cells of origin of motor nerve fibers in the glossopharyngeal nerve and various branches of the vagus nerve by means of retrograde degeneration," *J. Comp. Neurol.* 127, 293-306.

Lawn, A.M. (1966). "The nucleus ambiguus of the rabbit," *J. Comp. Neurol.* 127, 307-320.

Lobera, B., Pasaro, R., Gonzolez-Baron, S. and Delgado-Garcia, J.M. (1981). "A morphological study of ambiguus nucleus motoneurons innervating the laryngeal muscles in the rat and cat," *Neurosci. Lett.* 23, 125-130.

Miyazaki, T. (1982). "Somatotopical organization of the motoneurons in the nucleus ambiguus of cats - the horseradish perxidase method," *Otologia* (Fukuoka) 28, 649-679.

Molhant, M. (1911-12). "Localizations nucleaires," *Le Névraxe* 12, 277-316.

Olszewski, J. and Baxter, D. (1954). *Cytoarchitecture of the human brain stem*. S. Karger. Basel, New York.

Satomi, H., Yamamoto, T., Ise, H. and Takahashi, K. (1979). "Identification of the inferior salivatory nucleus in the cat as studied by HRP bathings of the transected glossopharyngeal nerve root," *Neurosci. Lett.* 11, 259-263.

Sugimoto, T., Itoh, K., Mizuno, N., Nomura, S. and Konishi, A. (1979). "The site of origin of cardiac preganglionic fibers of the vagus nerve: An HRP study in the cat," *Neurosci. Lett.* 12, 53-58.

Szabo, T. and Dussardier, M. (1964). "Les noyaux, d'origine de nerf vague ches le mouton," *Zeitschr. F. Zellforsch.* 63, 247-276.

Szentágothai, J. (1943). "Die lokalisation der kehlkopfmuskulatur in den vaguskernen," *Zeitschr. F. Anat.* 112, 704-710.

Taber, E. (1961). "The cytoarchitecture of the brain stem of the cat. 1. Brain stem nuclei of cat," *J. Comp. Neurol.* 116, 27-52.

Yanohara, T. (1955). "Anatomical studies of esophagus in clinical point of view," *Pract. Otol.* (Kyoto) 48, 686-723.

Yoda, S. (1940). "Über die kerne der Medulla oblongata der Katze," *Zeitschr. F. Mikr. -anat. Forsch.* 48, 529-582.

Yoshida, Y., Miyazaki, T., Hirano, M., Shin, T., Totoki, T. and Kanaseki, T. (1980). "Location of motoneurons supplying the cricopharyngeal muscle in the cat studied by means of the horseradish peroxidase method," *Neurosci. Lett.* 18, 1-4.

Yoshida, Y., Miyazaki, T., Hirano, M., Shin, T., Totoki, T. and Kanaseki, T. (1981). "Localization of efferent neurons innervating the pharyngeal constrictor muscles and the cervical esophagus muscle in the cat by means of the horseradish peroxidase method," *Neurosci. Lett.* 22, 91-95.

Yoshida, Y., Miyazaki, T., Hirano, M., Shin, T. and Kanaseki, T. (1982). "Arrangement of motoneurons innervating the intrinsic laryngeal muscles of cats as demonstrated by horseradish peroxidase," *Acta Otolaryngol.* 94, 329-334.

DISCUSSION

C. LARSON: What is the size of the motoneuron cell bodies for the intrinsic laryngeal muscles, and is there a difference in the size of the motoneuron cell bodies for the different muscles you studied?

Y. YOSHIDA: The motoneuron cell sizes were less than 60 microns and we did not observe any difference between the different muscles that we studied with respect to motoneuron cell size.

D. ZEALEAR: Why is it that the CT motoneurons are located more rostrally in the brainstem than those of other intrinsic laryngeal muscles? Is there a functional reason, or does the location simply reflect topography (i.e., they are located in association with motoneurons of muscles with similar embryological origin - the constrictors).

Y. YOSHIDA: On the basis of our investigation we are not able to draw any interpretation to answer your question.

7 ▪ VOICE FUNDAMENTAL FREQUENCY CHANGES FOLLOWING DISCHARGE OF LARYNGEAL MOTOR UNITS

Charles R. Larson and Gail B. Kempster

INTRODUCTION

The present study focuses on the manner in which different intrinsic laryngeal muscles affect voice fundamental frequency (Fo). Several studies have described the role of various laryngeal muscles in pitch control (Baer, Gay & Niimi, 1976; Faaborg-Andersen, 1965; Gay, Hirose, Strome, & Sawashima, 1972; Hirano, Ohala, & Vennard; 1969; Hirano, Vennard & Ohala, 1970; Shipp & McGlone, 1971; Sutton, Larson & Farrell, 1972). Despite the considerable number of these studies, there are still unresolved questions on just how the activity of certain laryngeal muscles affect Fo. Implicit in such discussions are the biomechanical effects of muscular adjustments. Based on anatomical and mechanical studies of vocal fold tissue, Hirano (1977) suggested that the vocal folds consist of two functionally different elements: a cover of epithelium overlying a body which is comprised of muscular tissue. Contractions of either cricothyroid (CT) or thyroarytenoid (TA) muscle may affect the length and stiffness of the cover and body differentially (see discussions of Hirano, 1977 and Fujimura, 1981). According to this model, contraction of the CT would increase the length of the vocal folds thus increasing the stiffness of the cover and resulting in a Fo increase. In support of this notion, numerous studies note CT EMG activity increases with Fo (Faaborg-Andersen, 1965; Gay *et al*, 1972; Hirano *et al*, 1970; Sutton *et al*, 1972).

Effects of TA contraction are less clear. Hirano (1977) suggested that under some conditions, contraction of the vocalis muscle (VOC) reduces stiffness of the vocal folds whereas under other conditions, VOC activity increases vocal fold stiffness (Hirano, 1977). One problem that confronts our understanding of the function of the TA muscle stems from its morphological character. The TA is still considered by many investigators to be one homogeneous muscle whose anterior attachment is the angle of the thyroid cartilage and whose posterior attachment is the arytenoid cartilage extending from the apex to the muscular process on the lateral side. Other opinions vary. Some researchers claim that the TA consists of two muscles: the vocalis, which is the most medial part lying adjacent to the vocal cords, and the TA proper, which lies lateral to the vocalis and comprises the main bulk of the muscle (for a general review see Zemlin, 1981; Dickson & Maue-Dickson, 1982; and Kahane, 1982). In addition, some anatomists have said that parts of the vocalis muscle insert on the vocal cords themselves, whereas other anatomists deny such an insertion. Another report suggests that the TA and lateral cricoarytenoid (LCA) muscle are functionally all one muscle, with the more lateral part originating from the cricoid cartilage (Dickson & Maue-Dickson, 1982). According to this view, the function of the more lateral parts of the TA is thought to be that of adduction of the vocal folds just as is the function of the classically defined LCA. The TA and LCA may accordingly be thought of as one muscle whose function changes depending on the medial-lateral position of the muscle fibers. A recent report (Rosenfield, Miller, Sessions, & Patten, 1982) described the morphology of the TA muscle in which the fibers comprising the TA were not found to be all aligned in the same direction, but appeared to course in different directions, much like that of the tongue. Suffice it to say, the TA is a complex muscle, and different subcomponents may have different functions. For example, it is possible that the medial portion may increase the stiffness of the medial parts of the vocal folds with concomitant Fo increase, whereas more lateral portions may cause shortening and relaxation of the vocal folds with decreases in Fo (Kahane, 1982).

In the study of the laryngeal muscles, many investigators have used wide-field EMG techniques. A problem with such techniques, however, is that the interference pattern reflects activity of many different motor units, and possible functions of different parts of the muscle being studied are obscured. Similarily, most studies that have examined the mechanical properties of the laryngeal muscles have been done by observing the contraction due to whole muscle stimulation. With this technique, it is not possible to study the mechanical properties of individual motor units or different parts of a given muscle. For instance, animal studies have shown that the contraction time of the CT is about 35-44 msec, whereas those for the vocalis and LCA are more on the order of 15-19 msec (Hast, 1967; Hast & Golbus, 1971; Hirose, Ushijima, Kobayashi and Sawashima, 1969; Martensson & Skoglund, 1964). The only report of single motor unit (SMU) contraction properties is from Zealear's study (1981) of cat TA in which time to peak contractions ranged from 4 to 43 msec. It is important to learn the contraction properties of SMUs in human laryngeal muscles in order to understand how the nervous system may be operating to control them. A study of the contraction properties of single motor units (SMU) might also provide insight into possible functions of different parts of such muscles as the TA. Since a fundamental property of muscle is its speed of contraction, which determines to a large extent its viscoelastic, biomechanical properpties (for discussion see Cooker, Larson & Luschei, 1980), such information also has implications for attempting to model these muscles.

In 1981, Baer reported the results of a study in which he adapted a technique utilized by Milner-Brown, Stein and Yemm (1973) for the study of the contraction properties of single motor units (SMU) in awake human subjects. Baer (1981) studied the CT muscle and an unidentified strap muscle. In brief, he triggered a computer on the discharge of a SMU and averaged voice Fo following the unit. He reported that the CT SMU caused a Fo increase with a peak at 80 msec latency. In support of this finding, Sapir, McClean and Luschei (1982) electrically stimulated the CT muscle in humans and observed peak changes in Fo at latencies of about 80 msec. The peak contraction time of the CT muscle appears, however, to be an order of magnitude longer than that reported for nonhuman CT muscles (Hirose *et al*, 1969; Martensson & Skoglund, 1964). It is unclear from Baer's study whether other units were firing in synchrony with the triggering unit and thus might be affecting the apparent contraction time of the unit in question. Similarily, in the study of Sapir *et al* (1982), it is uncertain if one or more than one unit was stimulated.

If the technique utilized by Baer is reliable, it would be a useful way to increase our understanding of how each of the laryngeal muscles contributes to Fo control as well as of the contracting properties of SMUs. However, in order for this technique to reveal information about SMUs, one must be certain that the SMU in question is not firing in synchrony with other units. Milner-Brown *et al* (1973) reported synchronous activity in one of their subjects, and other investigators have reported synchronous SMU activity in other muscles (Dietz, Bischofberger, Wita & Freund, 1976; Mori, 1975). Therefore, it is not unreasonable to suspect that laryngeal SMU's may also be activated synchronously.

In the present study, we have investigated the TA as well as the CT muscle using Baer's approach. In so doing, we have attempted to deal with the problem of synchronous discharge of SMUs from different muscles. Our results indicate some differences in the contraction times of CT compared to Baer's findings, and also suggest that different SMU's in the TA muscle may have different functions. The latter finding has implications for understanding the biomechanics of the TA muscle as well as its role in Fo control.

METHODS

Subjects. Two adult male and one female, all with normal voices and no evidence of vocal pathology, served as subjects. None of the subjects had received voice training.

Apparatus. SMUs were recorded with bipolar, hooked wire electrodes. The electrodes were .001″ teflon-coated, stainless steel wires. Usually about .25 mm of insulation was removed from the tip. Both electrode wires were inserted into a muscle with a single 30 ga hypodermic needle. Potentials from the electrodes were amplified with Grass P511 AC Preamplifiers.

The subject's voice was transduced with an Electro-Voice, omni directional microphone, amplified differentially, and recorded on magnetic tape along with the SMU potentials. A locally-fabricated instantaneous frequency meter was used to extract voice Fo from the voice signal.

Procedures. Attempts were made to insert electrodes bilaterally into each CT and TA muscle, but we were not always successful. Various tests (suggested by Hirano & Ohala, 1969) were used to verify correct electrode placement. Frequently electrodes were determined to be in one of the strap muscles, and those data will not be presented in this report.

Following electrode insertion, the subject was instructed to vocalize with a steady modal pitch and constant intensity for several seconds, take a breath, and repeat the procedure. While the subject was vocalizing, the experimenter triggered an oscilloscope from the fundamental of the voice signal and the gate output pulse from the oscilloscope was recorded on tape. Additionally, the experimenter displayed one of the 4 SMU channels on another oscilloscope as well as played the SMU signal through an audio monitor. While a subject was phonating attempts were made to keep the voice pitch such that a SMU was firing at a relatively low rate. After sufficient data were recorded, the subject was instructed to vocalize at a slightly different pitch, and the procedures replicated.

Following collection of data, one of the electrode pairs was connected to the output of a stimulus isolation unit, which was controlled by a Grass S-88 stimulator. The particular muscle was stimulated with square wave pulses of 1 msec duration and a threshold was determined at which the subject could just barely feel the stimulus. The stimulus intensity was then reduced to just below threshold and the subject was instructed to begin phonating. While he was phonating, the muscle was stimulated at a variable frequency from 1 to 10 Hz. The voice signal and a synchronizing pulse from the stimulator were recorded on magnetic tape.

Data Analysis. Data recorded on magnetic tape were later processed through an averaging program on a PDP 11/23 computer. This program allowed the averaging of multiple channels of data simultaneously. The gate pulse from the oscilloscope, which represented the subject's Fo, was passed from the tape playback to the frequency meter. The output of the frequency meter was essentially a DC signal whose magnitude was proportional to Fo (see Figure 7-1). There were slight variations in the DC Fo signal that represented small fluctuations in the Fo. The presumption of this study was that these fluctuations might be related to the discharge of SMUs. The signal representing voice Fo was then passed to the A/D convertors of the computer system. The SMU channels were full-wave rectified prior to entering the A/D convertors of the computer system. The averaging program was initiated by the discharge of a SMU from one of the 4 SMU channels.

The following procedures were used to "isolate" a non-rectified SMU potential in order to trigger the computer. In some cases a given SMU was well isolated and we were able to trigger an oscilloscope off its waveform. The gate pulse from the scope then triggered the computer. In other cases, we passed the SMU potential through a high pass filter in order to accentuate the amplitude of one SMU potential over that of another to thereby trigger the oscilloscope and computer. In still other cases, we used a locally fabricated dual-time and amplitude window comparator to isolate one SMU from another. This device then triggered the computer. The computer was programmed to display a certain time window prior to the discharge of the SMU as well as a longer window following the SMU. For electrical stimulation of muscle, the Fo signal was averaged with respect to the onset of the stimulus to the particular muscle.

Averaged data were subsequently displayed on a graphics computer terminal and measured. Since the averaged Fo signal was frequently noisy, we employed a computer smoothing process to eliminate the noise. This procedure does not alter the time characteristics of the signal, but merely eliminates noise transients of the signal (See Figure 7-2 for a comparison of the filtered and non-filtered Fo averaged signal). This process not only made the figures easier to read, but it also enhanced our ability to make accurate measures of time and magnitude of the averaged signal.

RESULTS

When an electrode was placed in the CT muscle, the EMG or SMU activity displayed remarkable sensitivity to elevations in pitch with rather little changes due to jaw depression or pushing the head forward. We found electrodes located in the TA to be very interesting from the standpoint that a wide variety of EMG or SMU patterns of activity were observed. For example, some TA SMUs showed clear increases in firing rate with pitch elevations. On occasion, some of the SMUs would decrease their activity when a certain pitch level was achieved. Units were also recorded that either increased their activity with respiratory effort on prolonged exhalation or units decreased their activity with increasing phonation time. Some units also were more active than others with adduction of the vocal folds. These different SMU patterns may reflect the fact that the electrode was located in different parts of the TA, but we were unable to make accurate judgements as to the precise location within the muscle. A detailed analysis of these SMU properties will form the basis of a subsequent report.

Although we attempted to isolate SMUs from all of our electrode insertions, on many occasions several SMUs were active and it was not possible to isolate a single SMU. The activity from such electrodes was nevertheless averaged with respect to a SMU from another electrode, since it has been argued that averaging EMG or multiple unit activity is a valid technique for establishing correlated activity in other muscles (Milner-Brown *et al*, 1973; Kirkwood, 1979). From the three subjects in this study, data were recorded and analyzed from 23 CT and 16 TA SMUs. In some cases, synchronous electromyographic (EMG) activity was present in one or more muscles, which made interpretation difficult.

Among the problems we encountered in analyzing the SMUs was the frequent inability to observe an identifiable Fo change following the SMU despite numerous averaging attempts. Frequently, if data were averaged from a SMU for hundreds or thousands of discharges, an averaged change in Fo would appear, disappear and perhaps reappear again. It is our impression that this resulted from subtle differences in the way a subject would control Fo from breath to breath, or even during the course of a single breath. To deal with this problem, we averaged data after relatively few firings of an SMU and compared several averages. We then selected from this group one average that best represented the group. Out of the total number of SMUs, we were able to describe contraction properties of only 12 CT and 6 TA

SMUs that were not obviously contaminated by synchronous activity in other muscles.

Figure 7-1 illustrates the nature of our unprocessed data. The top trace shows the audio signal played back from the tape recorder. Directly beneath is the DC analog of the Fo that was obtained by processing the audio signal through our frequency meter. Careful inspection reveals small changes in this DC signal. Also shown on this figure are simultaneous records of SMU activity from the right and left CT muscles. Note that there appears to be no systematic relationship between the units in the right and left CT muscles. In this particular case, we were able to trigger the computer off units in the R CT and obtain an averaged Fo change. Subsequently, we then triggered off units in the L CT and also obtained an Fo change.

Figure 7-2 illustrates the results from averaging the Fo off the right CT SMU shown in Figure 7-1. In this case the amount of Fo change was .4 Hz and the time to peak Fo change was 23 msec. From all the CT SMUs analyzed, the time to peak Fo varied from 9 to 57 msec with most having times less than 25 msec. In the lower part of this figure are the averages from the left CT and right TA, with very little indication of synchronous EMG activity from them.

Figure 7-3 illustrates the results from triggering the computer off a right TA SMU and averaging voice Fo, and left CT and TA muscles. In this example, Fo increased with a peak latency of approximately 9 msec following the onset of firing of the right TA SMU. The amount of Fo change in this particular case was .1 Hz. Analysis of the 6 TA SMUs revealed latencies from 9 to 20 msec. In this figure, the other muscles show no obvious correlated activity near the time of the triggering SMU.

Although the averaging technique yielded several examples of voice Fo like those illustrated above, we were quite surprised to observe Fo changes like those illustrated in Figure 7-4. In these examples from a TA and a CT unit, Fo exhibited oscillations that began after the triggering SMU. In all cases in which we observed such oscillations, the mean Fo maintained by the subject was in the 200 and 280 Hz range, and the time from a peak to a trough corresponded approximately to the reciprocal of voice Fo. These oscillations then represent periodic increases and decreases in Fo following the SMU. We are relatively certain that these oscillations did not result from chance synchronization of the sampling frequency of the computer system with voice Fo since we varied the sampling frequency, filtered the Fo signal at 100 Hz (low pass), reprocessed the data with the tape played back at one half the recording speed, and still were unable to alter this effect.

Our analyses of the relationship of SMU contractions and Fo were frequently confounded by synchronous activity in one or more muscles. In Figure 7-5, a slight increase in L CT and L TA activity (see arrows) immediately follows the triggering R TA SMU. We do not believe it is possible to relate the Fo change following a SMU in cases such as this since it could be caused in part by activity of the other muscles. Data such as these have not been included in our reports of SMU contraction effects on Fo.

The effects of direct muscle stimulation provided results that were very similar to those of triggering from a SMU potential, and help confirm the general properpties of the CT and TA SMUs. Figure 7-6 illustrates averaged Fo changes following stimulation of a CT and a TA muscle. In all cases, such as that in Figure 6A, CT stimulation resulted in an Fo increase, and the latencies varied from 11 to 60 msec.

In these examples of direct muscle stimulation we have no guarantee only one SMU was stimulated. The latencies of Fo peaks, therefore, may not accurately reflect contraction times of SMUs. In fact, we believe the above example of muscle stimulation may probably reflect the activity of several SMUs. To illustrate this point, Figure 7-6B shows the results from

FIGURE 7-1
Display of unprocessed data. Top trace is voice signal, second trace is output of frequency calculator, with calibration mark at the left. The next two traces are SMU data from right and left cricothyroid muscles.

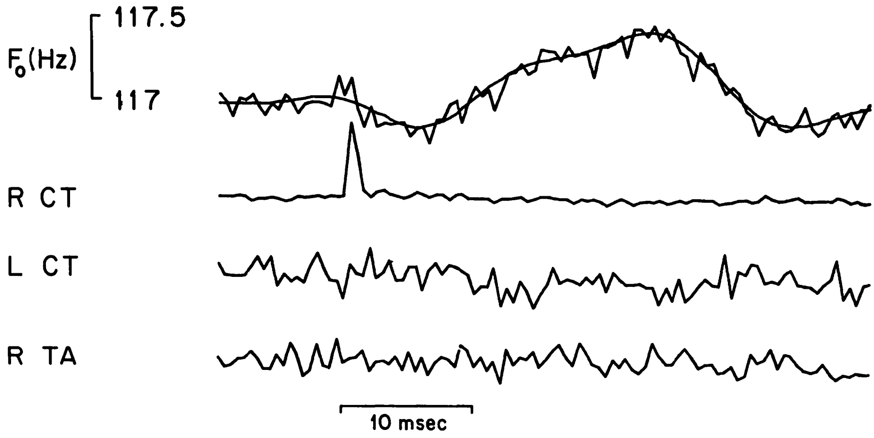


FIGURE 7-2
Illustration of averaged data. The top trace shows the averaged Fo from the frequency calculator. Superimposed on this trace is a digitally smoothed function that represents Fo (for further discussion see text). The second trace represents the SMU from which the computer was triggered. The location of the SMU potential also represents the time from which data on the other channels have been averaged. The next two traces represent averaged SMU activity from left cricothyroid and right thyroarytenoid muscles.

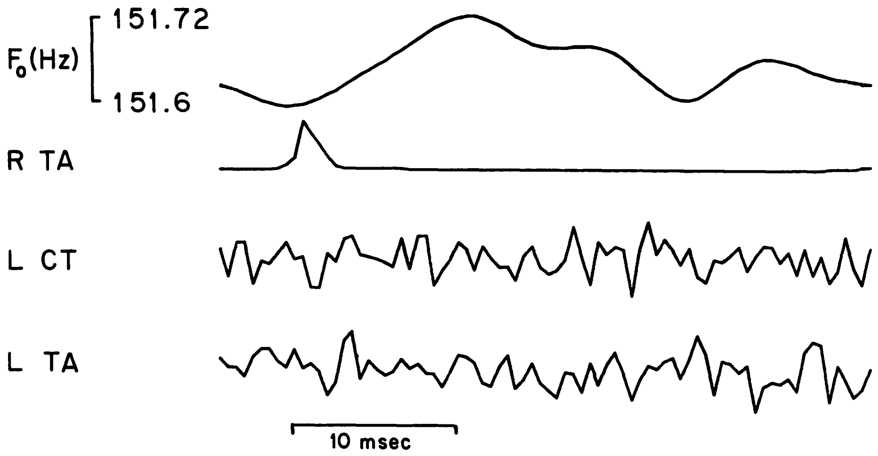


FIGURE 7-3
Averaged data similar to Figure 7-2, but with data from the discharge of a right thyroarytenoid SMU.

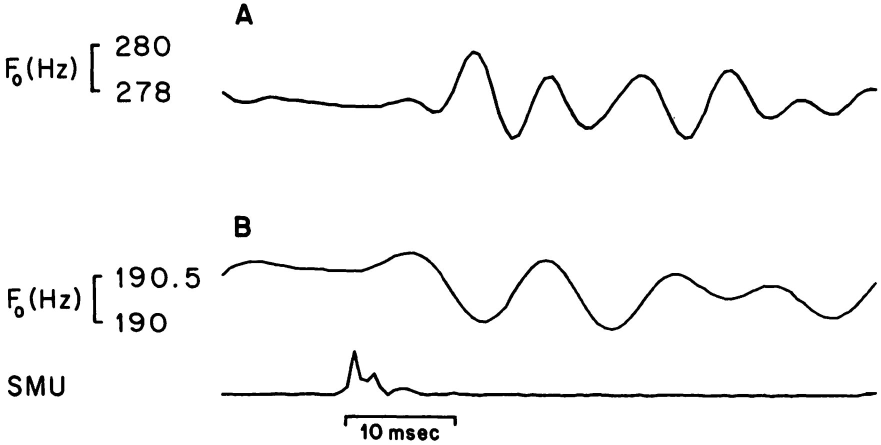


FIGURE 7-4
Illustration of oscillatory changes in Fo from one subject in the thyroarytenoid and another subject in the cricothyroid muscle. The SMU potential trace at the bottom is displayed to indicate the time from which data were averaged.

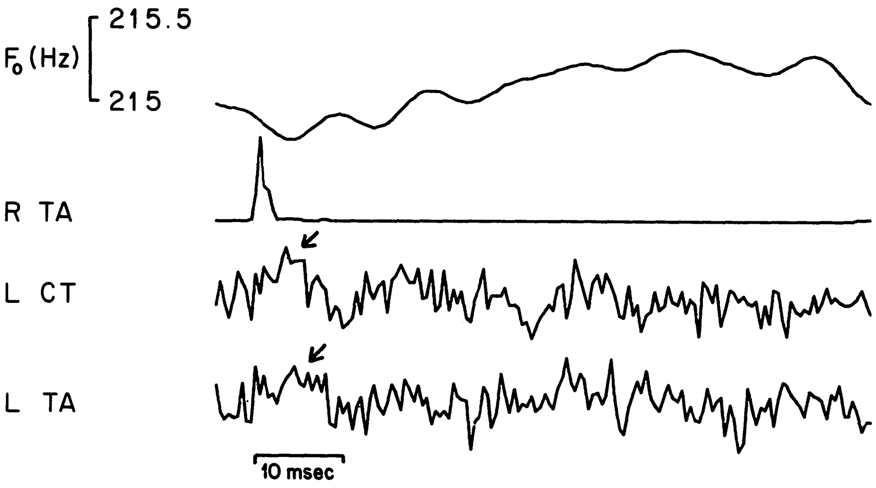


FIGURE 7-5
Averaged data similar to that shown in previous figures. The arrows shown on the bottom traces indicate periods of synchronous EMG activity in those two muscles.

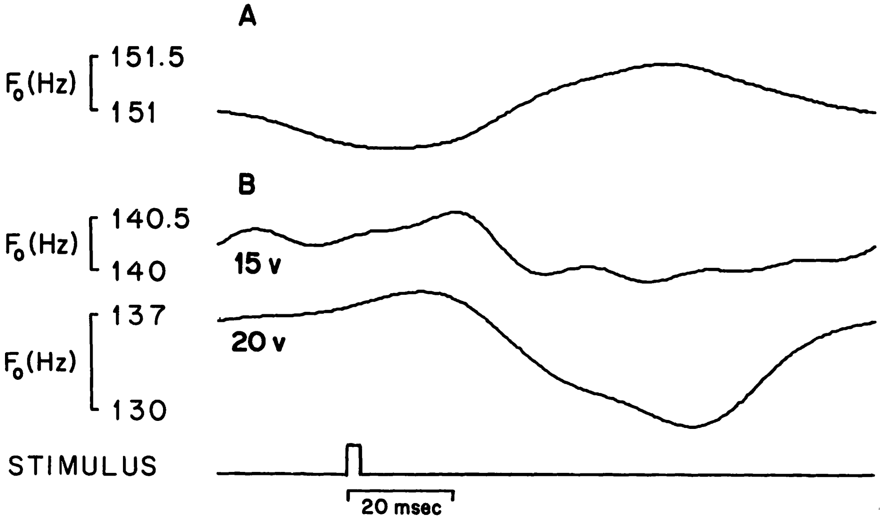


FIGURE 7-6
Averaged Fo data following stimulation of the cricothyroid (top trace) and thyroarytenoid (bottom two traces) muscles. The bottom trace illustrates the stimulus pulse from which data were averaged. The thyroarytenoid muscle was stimulated at 15 and 20 volts as indicated; cricothyroid was stimulated at 10 volts.

stimulating a L TA muscle at two different intensities. In the top trace, a very slight increase in Fo is seen with a peak latency of 20 msec when TA was stimulated at 15 volts. When the voltage was increased to 20, the bottom trace reveals a peak increase at about 20 msec but then a much more pronounced decrease in Fo with a latency of 65 msec. In this case, we believe that with relatively low stimulus intensity, one or more units were active which had the primary effect of increasing Fo. However, with higher intensity, more units were excited, some of which significantly reduced Fo. This is the only example for which we observed a decrease in Fo; in all other cases Fo increased similarily to that shown in the top trace of Figure 7-6B. The latencies of Fo change with TA stimulation ranged from 20 to 65 msec.

DISCUSSION

The results of the present study extend and confirm Baer's (1981) findings that it is possible through averaging techniques to detect a voice Fo change following the discharge of an intrinsic laryngeal muscle SMU. From studies in other somatic musculature, it has been found that the force produced by a SMU can be determined through such averaging techniques (Clark, Luschei, & Hoffman, 1978; Milner-Brown *et al*, 1973). Accordingly, it seems reasonable to suggest that minute changes in stiffness or length of the vibrating parts of the vocal folds, caused by SMU contractions, could affect Fo. There are likely to be slight time delays in the conversion of mechanical perturbations to acoustical energy and its transmission to the recording site in front of the lips, but these delays are probably less than 1 msec. It therefore seems that the averaged Fo changes represent the time course of the contraction of the SMUs.

The contraction times we observed for CT are somewhat less than that measured by Baer (1981). Our longest contraction time was about 60 msec, whereas Baer's unit was about 80 msec. Given the sampling problem, we do not regard this as a fundamental difference, but we would stress that the majority of our SMUs in the CT had contraction times in the 20 to 30 msec range and some were as short as 11 msec. These values are slightly less than whole muscle contraction times of the CT muscle in cats and dogs (Hirose *et al*, 1969; Martensson & Skoglund, 1964). The CT SMUs in every case produced increases in Fo which supports previous evidence that the CT muscle's function is that of pitch elevation (Baer *et al*, 1976; Faaborg-Andersen, 1965; Gay *et al*, 1972; Hirano *et al*, 1969; Hirano *et al*, 1970; Sutton *et al*, 1972;).

Averaging on TA SMUs also produced Fo increases, but the contraction times were faster than 30 msec, with most being less than 15 msec. This finding agrees with studies of whole muscle contraction times that have reported the TA to be a faster muscle than the CT (Hirose, *et al*, 1969; Martensson & Skoglund 1964), and with Zealear's study (1981) of cat TA SMUs which had contraction times as low as 4 msec. The voice Fo increases also support previous claims that the function of the TA muscle is to increase pitch (Baer *et al*, 1976; Hirano *et al*, 1969; Hirano *et al*, 1970; Shipp & McGlone, 1972).

The recording of Fo oscillations following either a CT or TA SMU was unexpected and remains difficult to interpret. In a study of vocal perturbation, Lieberman (1961) noted a similar pattern of long-short variations in glottal periods. He attributed this effect to the glottal excitation waveform. In the present study the Fo oscillations were synchronized with respect to SMU contractions. Therefore, other explanations on the origin of such oscillations should be considered. This effect was unchanged even by altering the sampling frequency of the computer, filtering the DC Fo signal to be averaged and playing the data back at one half speed, which strongly indicates that the oscillations are not the result of artifact due to our instrumentation. It should be noted that in all cases where oscillations were observed, the voice Fo was usually in the 200 to 280 Hz range, and this may provide a clue as to the cause of such oscillations. We suggest that vibration of the vocal folds excites sensory receptors in the laryngeal system, and these receptors then excite laryngeal motor neurons. Sasaki and Suzuki (1976) determined by electrical stimulation of the SLN, that the laryngeal-laryngeal reflex in humans has a latency of about 25 msec. Thus activation of laryngeal receptors should cause reflex activation of the muscles with a latency of 25 msec. When a person is phonating in the pitch range of 200 to 280 Hz, the interval between glottal cycles is 4 to 5 msec. Since the 25 msec latent period of the reflex is a whole number multiple of the glottal inter-cycle interval, the reflex activation of the receptors would tend to cause laryngeal muscle excitation in time with the glottal period. Considering a SMU, if it too were influenced by laryngeal reflexes, it would tend to discharge in synchrony with the glottal cycle because of the timing interaction of the laryngeal reflex with glottal periods. Such a result in the averaging process would cause periodic fluctuations in Fo following the SMU. Such oscillataions in all probability do not reflect any mechanical effects of the SMU, only the tendency to synchronize glottal periods on the SMU discharge. In keeping with this argument, it is important to realize that a given SMU would not be reflexively activated for each glottal cycle. Indeed because such SMUs fire at intervals usually of about 40-60 msec, they would only be activated on every eighth to twelfth glottal cycle. But when they are activated, they would tend to fire in synchrony with the glottal cycles because of the interaction of the laryngeal reflex latency and the glottal inter-cycle interval. Another point to be made is why one would not expect to see such oscillations when a person is vocalizing with an Fo of 100 Hz, which has a glottal inter-cycle interval of 10 msec. In this case, the 10 msec time period is probably sufficiently long that the tendency for the SMU to discharge in synchrony with the glottal period is reduced, simply because of the greater time interval of the glottal cycle, i.e., more time smearing is possible with a longer glottal period.

If we are correct in interpreting the cause of the oscillations we have reported, our explanation has implications for voice control. If indeed reflexes enhance the tendency for muscles to contract in synchrony with the glottal cycle, then at some pitches, active muscle contraction would coincide with the movements caused by the interaction of the aerodynamic and myoelastic factors contributing to vocal fold vibration. In such a case, vocal fold vibrations might be enhanced or could occur with less effort than at other pitches.

The other method we used to study the functions of the CT and TA muscles was through direct electrical stimulation. With CT stimulation, the latency of the peak Fo increase was about the same as we found with SMU averaging. We can thus say that in some of these cases we were probably stimulating one or at most a few SMUs. The only other report utilizing this technique is that of Sapir *et al* (1982) who reported a Fo increase with CT stimulation at a latency of 80 msec. Considering that Baer also reported a SMU averaged Fo peak response of 80 msec latency, it is possible that Sapir *et al* (1982) stimulated one SMU. However, our results suggest that they may have stimulated more than one SMU.

The effects of stimulation of the TA muscle were in most cases similar to the effects of averaging Fo off a SMU. There was only one case in which TA stimulation did not cause an increase in Fo, which helps confirm previous reports of the TA muscle being a pitch elevator (Baer *et al*, 1976; Hirano *et al*, 1969, Hirano *et al*, 1970; Shipp & McGlone, 1971). The one case in which we observed a fall in Fo after stimulation illustrates two points. First, this observation supports previous suggestions that part of the TA muscle decreases pitch (see Zemlin, 1981; Kahane, 1982). Secondly, in our figure illustrating this effect (Figure 7-6B), we also demonstrated how with electrical stimulation it is easy to excite multiple SMUs, some with possible different functions. In our example, the lower intensity perhaps had a greater tendency to excite SMUs whose effect was to increase pitch. With higher intensity, other SMUs were activated which caused Fo to decrease.

With many of the SMUs isolated, we were unable to detect a significant Fo change despite numerous averaging attempts. Perhaps the most obvious explanation for this failure is that if a given SMU is firing at a relatively high level, the average tension it develops will become fused to a relatively steady-state level of muscle force. In such a case, the Fo would show very little if any modulation. Another possibility is that a SMU might produce such a small amount of force compared to other units that the effects of triggering off a SMU might be obscured by the other units. A third possible explanation is that some TA SMUs have little if any effect on vocal fold stiffness and hence would not contribute to Fo. A fourth factor was that with higher pitches, Fo oscillations occurred which obscured any possible Fo changes related to mechanical changes in the vocal folds.

The occasional problem we encountered with averaged synchronous EMG activity in other muscles indicates that in future studies of this type multiple muscles should be sampled. With synchronous muscle activity, it is difficult, if not impossible to ascribe a perturbation in Fo to a particular SMU.

CONCLUSION

The results of the present study indicate that the contraction speed of human CT and TA SMUs are relatively fast, with the TA units being somewhat faster than the CT units. Both TA and CT units caused Fo increases. At higher pitches, we frequently observed Fo oscillations following a SMU discharge. Our tentative explanation of the Fo oscillations is that they result from synchronization of the glottal cycle and the SMU discharge. This synchronization may result from excitation of SMUs by laryngeal afferents, which may be activated by the glottal vibrations themselves. These results confirm previous evidence that the primary function of the cricothyroid muscle is to increase pitch. Our results also suggest that the primary function of the thyroarytenoid muscle is to increase Fo, but our stimulation data suggest some parts of this muscle may decrease pitch as well. Electrical stimulation of the muscles supported the findings of the contraction properties of individual motor units.

REFERENCES

Baer, T. (1981). "Investigation of the phonatory mechanism," Proc. of the Conf. on the Assessment of Vocal Pathol., *ASHA Reports* 11, 38-47.

Baer, T., Gay, T. and Niimi, S. (1976). "Control of fundamental frequency, intensity and register of phonation," *Haskins Labs: Stat. Rep. on Speech Res.*, SR-45/46

Clark, W.R., Luschei, E.S. and Hoffman, D.S. (1978). "Recruitment order, contractile characteristics and firing patterns of motor units in the temporalis muscle of monkeys," *Exp. Neurol.* 61, 31-52.

Cooker, H.S., Larson, C.R. and Luschei, E.S. (1980). "Evidence that the human jaw stretch reflex increases the resistance of the mandible to small displacments," *J. Physiol.* (London) 308, 61-78.

Dickson, D.R. and Maue-Dickson, W. (1982). *Anatomical and Physiological Bases of Speech.* Boston: Little, Brown and Co.

Dietz, V., Bischofberger, E., Wita, C. and Freund, H.J. (1976). "Correlation between the discharges of two simultaneously recorded motor units and physiological tremor," *EEG and Clin. Neurophysiol.* 40, 97-105.

Faaborg-Andersen, K. (1965). "Electromyography of laryngeal muscles in humans: technics and results," in *Current Problems in Phoniatrics and Logopedics*, Basel. S. Karger.

Fujimura, O. (1981). "Body-cover theory of the vocal fold and its phoenetic implications," in K.N. Stevens and M. Hirano (Eds.) *Vocal Fold Physiology*, Tokyo: Univ. of Tokyo Press, 271-290.

Gacek, R.R. (1975). "Localization of laryngeal motor neurons in the kitten," *Laryngoscope* 85, 1841-1861.

Gay, T., Hirose, H., Strome, M. and Sawashima, M. (1972). "Electromyography of the intrinsic laryngeal muscles during phonation," *Ann. Otol. Rhinol. Laryngol.* 81, 401-409.

Hast, M.H. (1967). "Mechanical properties of the vocal fold muscle," *Prac. Otol. Rhinol. Laryngol.* 29, 53-56.

Hast, M.H. and Golbus, S. (1971). "Physiology of the lateral cricoarytenoid muscle," *Prac. Otol. Rhinol. Laryngol.* 33, 209-214.

Hirano, M. (1977). "Structure and vibratory behavior of the vocal folds," in M. Sawashima and F.S. Cooper (Eds.) *Dynamic Aspects of Speech Production*, Tokyo: Univ. of Tokyo Press.

Hirano, M., Vennard, W. and Ohala, J. (1970). "Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal msucles," *Folia phoniat.* 22, 1-20.

Hirano,M. and Ohala, J. (1969). "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles," *J. Speech Hear. Res.* 12, 362-373.

Hirano, M., Ohala, J. and Vennard, W. (1969). "The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation," *J. Speech Hear. Res.* 12, 616-628.

Hirose, H., Ushijima, T., Kobayashi, T. and Sawashima, M. (1969). "An experimental study of the contraction properties of the laryngeal muscles in the cat," *Ann. Otol.* 78, 297-306.

Kahane, J.C. (1982). "Anatomy and physiology of the organs of the peripheral speech mechanism," in N.J. Lass, L.V. McReynolds, J.L. Northern and D.E. Yoder, (Eds.) *Speech, Language, and Hearing 1: Normal Prosesses*. Philadelphia: W.B. Saunders Co.

Kirkwood, P.A. (1979). "On the use and interpretation of cross-correlation measurements in the mammalian central nervous system," *J. Neurosci. Methods* 1, 107-132.

Lieberman, P. (1961). "Perturbations in vocal pitch," *J. Acoust. Soc. Am.* 33/5, 597-602.

Martensson, A. and Skoglund, C.R. (1964). "Contraction properties of intrinsic laryngeal muscles," *Acta Physiol. Scand.* 60, 318-336.

Milner-Brown, H.S., Stein, R.B. and Yemm, R. (1973). "The contractile properties of human motor units during voluntary isometric contractions," *J. Physiol.* (London) 228, 285-306.

Mori, S. (1975). "Entrainment of motor-unit discharges as a neuronal mechanism of synchronization," *J. Neurophysiol.* 38, 859-869.

Rosenfield, D.B., Miller, R.H., Sessions, R.B. and Patten, B.M. (1982). "Morphologic and histochemical characteristics of laryngeal muscles," *Arch. Otolaryngol.* 108, 662-666.

Sapir, S., McClean, M. and Luschei, E. (1982). "Time relations between cricothryoid muscle activation and Fo changes," paper, Ann. Convention ASHA, Toronto, Canada.

Sasaki, C.T. and Suzuki, M. (1976). "Laryngeal reflexes in cat, dog, and man," *Arch. Otolaryngol.* 102, 400-402.

Shipp, T. and McGlone, R.E. (1971). "Laryngeal dynamics associated with voice frequency change," *J. Speech Hear. Res.* 14, 761-768.

Sutton, D., Larson, C. and Farrell, D. (1972). "Cricothyroid motor units," *Acta Oto-laryngol.* 74, 145-151.

Zealear, D. (1981). "The contractile properties and functions of single motor units of thyroarytenoid muscle in the cat," paper, Conf. on Vocal Fold Physiol., Madison, WI.

Zemlin, W.R. (1981). *Speech and Hearing Science*. Englewood Cliffs, N.J.; Prentice-Hall, Inc.

DISCUSSION

I. TITZE: How do you control for tremor and vibrato in your interpretation of small frequency changes?

C. LARSON: As for tremor and vibrato, we ask the subjects to maintain as steady a pitch and intensity as possible. We realize, of course, that there still will be some tremor and vibrato, but we think that these aspects of voice may be due in part to SMU contractions.

T. BAER: I noticed that to the left of the line-up point in some of your figures, there was not a flat curve which suggests that when you were stimulating, the interval between the stimuli must have been of the same order of magnitude as the duration of the twitch, and for the single motor unit work you probably were not rejecting small interspike intervals.

C. LARSON: For the single motor unit analysis we first examined their firing frequency and attempted to set up the window on the computer to match that of the intervals of the single motor units. A lot of the laryngeal motor units we've looked at fired at frequencies of around twenty or thirty Hz and then set the computer window at a value to match the fastest stimulus frequency. Now as far as the decline in pitch that you see prior to the stimulation point, that's a little bit difficult for us to explain, but it might be that there were slight periodic fluctuations in the voice or an occasional stimulus induced change of longer duration than most. I should also say, the reason we had to vary the stimulus frequencies as much as we did was to avoid chance synchronization of the stimulus pulse and voice fundamental frequency because then we would get huge oscillations across the entire records.

T. BAER: That would be the equivalent of random firing of the single motor units.

C. LARSON: Exactly. With the single motor units, we assume there's some randomness there, but we think some of our results suggest there may be some non-random things going on.

T. BAER: Just to pursue a thought. When you were recording the spike activity, did you in any way attempt to reject the short firing intervals where the twitches would superimpose?

C. LARSON: Yes, the computer automatically rejected short intervals.

G. ZIMMERMAN: When you stimulated one muscle, was there any noticeable action in any other muscle?

C. LARSON: We disconnected the other muscles during stimulation because you just see large stimulus artifacts on those channels. So it was not possible to see effects in other muscles.

M. HIRANO: I have a vague guess that the effect of TA on Fo may depend on the overall level of CT activity. Did you find any evidences demonstrating some relationship between the effect of TA upon Fo and CT activity level?

C. LARSON: Your suggestion is a good one. We have not looked for such a relationship yet, but it would probably be worthwhile to do so.

J. KAHANE: I would like to say a few words regarding your comments on the heterogeneity in the organization of muscle fibers in the thyroarytenoid. The question is, what's the nature of that heterogeneity? I think part of the answer has to do with the way the fibers which localize themselves close to the vocal process are oriented, versus those fibers which spread out in the fovea oblonga and ultimately upwards in the arm of muscularis. Those fibers that tend to localize themselves closest to the vocal process and to the arytenoid contiguous to that area, tend to come from the highest points of attachment on to the deep surface of the thyroid cartilage, and they for the most part tend to run longitudinally from the deep surface of the thyroid cartilage to the arytenoid. So that in cross-section, they tend to look more ovoid.

They're smaller, and they're more densely packed. The fibers that tend to spread outwards into the most lateral portion of the thyroarytenoid are coming from successively lower points of attachment on the thyroid cartilage. As they course towards the arytenoid cartilage, they're spreading outwards. So in coronal section, they're more obliquely oriented. So I think in the section that you saw in the histochemical study the field is relatively small, and I think they probably all were sampled pretty much close to the arytenoid, so they all look oval, or relatively oval in shape. But if you look at the total cross-section of the thyroarytenoid, you will see those regional differences. And they have to do largely with the geometry of the muscle fibers as they course in the anterior and posterior direction.

M. ROTHENBERG: You mentioned the fundamental frequency extractor used. Did it average over time?

C. LARSON: No, it computed the time interval between successive pulses, and its output was a DC signal proportional to the time between pulses. We took the voice signal, low-pass filtered it, triggered it on an oscilloscope, took the gate pulse out of the oscilloscope and fed that into the device that computed the time interval between successive pulses.

M. ROTHENBERG: And that value is held to the next pulse?

C. LARSON: Right, its held so if you're holding at a steady pitch, the fundamental frequency extractor just looks like a DC level that's proportional to fundamental frequency.

M. ROTHENBERG: In describing that, I guess you should specify the filtering you used because that would basically add an averaging time constant that would affect your final results.

C. LARSON: Yes, the filtering contributed about a .5 msec delay, and so that is a constant error in our estimates of the contraction properties of the units.

D. COOPER: It may be difficult to compare time relations for single MU's to time relations for whole muscle, as in your and Zealear's work. Richard Stein, in his work on frequency response of soleus indicated that damping of the 2nd order model he fitted to the frequency response decreased as more units were recruited, i.e., a stimulated single MU's mechanical activity is highly damped by the other inactive MU's, while a larger number of active MU's is less damped.

C. LARSON: I think you raise a good point and that may explain why we were unable to detect Fo changes following discharge of some of the SMU's we recorded. Whether or not the damping you speak of would alter the time relations of SMU contractions or their stiffness values is another matter. At this time, I would have to say that further work is needed to assess to what extent damping by non-active muscle would alter the time characteristics of active SMU's.

D. BLESS: You mentioned that you have been somewhat disheartened by the variations in activity observed for the same pitch levels produced immediately following breath intake. Do you think any of these differences could be due to interactions between lung volume level and glottal valving? And if so, do you control for lung volume level?

C. LARSON: I think you misunderstood me. We did not notice effects immediately following breath intake. We do, however, think that some of our results are affected by different strategies in Fo control that the subjects used from one breath to the next. Nevertheless, I think your suggestion about the interaction between lung volume level and glottal valving is good, and perhaps if we had controlled for this variable our results would have been more consistent.

8 ▪ ACTIVITY OF THE THYROARYTENOID MUSCLE IN THE PRODUCTION OF KOREAN STOPS AND FRICATIVES

Hajime Hirose, Hea Suk Park and Masayuki Sawashima

INTRODUCTION

In Korean, stop consonants having the same place of articulation can be classified into three different types with respect to manner of articulation. The three types are generally referred to as "forced", "lax" and "aspirated". All stop types may occur in the syllable-initial position to be realized as voiceless, while in the medial position, the lax stops are usually manifested by voiced allophones. In syllable-final position, the three stop types are phonetically realized as voiceless "applosives", being characterized by the absence of oral release.

Similar manner classifications are also applied to affricates and fricatives in Korean, with affricates classified into the same three types. Fricatives are classified into two types, i.e. "forced" and "lax", although the latter type is referred to as "aspirated" by some investigators.

The present study aimed at examining the laryngeal adjustments for Korean stops both in the syllable-initial and syllable-final positions with special reference to the activity of the thyroarytenoid muscle. Observations on the laryngeal adjustments for the fricatives are also performed.

PROCEDURES

Electromyographic (EMG) signals from the thyroarytenoid (VOC) muscle were recorded in three native Korean speakers of Seoul dialect using hooked-wire electrodes while the subjects repeatedly read meaningful test words containing different types of stops and fricatives placed in different phonological environments. The EMG signals were subsequently reproduced and computer-processed to obtain the average time curve of the EMG activity relative to the acoustic signals which were recorded simultaneously.

Test words were prepared so as to place the stops and fricatives in different phonological conditions. They were all meaningful words with one exception. Those test words which appear in the figures of the present report are presented as examples below in phonetic orthography, with the "forced" type written as geminates.

1. Syllable-initial stops and fricatives
   "k^{h}ul" (nonsense word) "sirɰm" ("anxiety")
   "kul" ("oyster") "ssirɯm" ("wrestling")
   "kkul" ("honey")
   These words were uttered in the frame "ike ____ ita (that is ________)".
2. Syllable-final stops (applosives)
   "hyət^{h}" ("side")
   In this series, a similar word ending in a vowel was added for reference.
   "hyə" ("bran")
   These were uttered in the frame "ike________".
3. Syllable-final applosive followed by the syllable-initial stops. In this case, initial "lax" stops are said to become "forced" stops.

"pɛk kaci" ("a hundred kinds") ‘tɛk k^{h}i" ("your key")
"pɛk kkaci" ("up to a hundred") "kyɔt^{h} kaci" ("nearby branch")
Here, test words with forced syllable-initial stops were added for reference and were uttered in frame "ike______ta".
"pɛ kkaci" ("to the boat")
Sound spectrographic analysis was carried out on the same speech samples used for the EMG data assessment.

RESULTS

In general, VOC activity patterns in the three subjects appeared to be quite similar to each other for each token type. Therefore, the averaged EMG patterns for one of the three subjects are shown in this report as a representative sample.

EMG findings on syllable-initial stops. It was revealed that VOC activity was suppressed for each type of stop consonant examined, the degree of which was slightest for the forced type and most marked for the aspirated type, while it was moderate for the lax type. After suppression, VOC activity increased again toward the voice onset of the postconsonantal vowel. The timing of the reactivation after suppression was earliest for the forced type, followed in order by the aspirated and the lax types. The peak of the reactivation was higher for the forced and the aspirated types than for the lax type (Figure 8-1). Incidentally, the vowel preceding the lax stop tended to be laryngealized in this subject.

EMG findings on syllable-initial fricatives. EMG patterns of VOC appeared to be very similar to each other with the lax and forced types (Figure 8-2).

EMG findings on syllable-final applosives. When compared to the EMG pattern for the test word ending with a final vowel, VOC activity appeared to decline abruptly after peaking for the vowel preceding the applosive stop (Figure 8-3).

EMG findings on syllable-final stops followed by syllable-initial stops. It was found that the pattern of VOC activity for those utterance samples containing final stop followed by a syllable-initial lax or forced stop was very similar to that for an initial forced stop in terms of the degree and timing of VOC suppression for the consonant segment and of the reactivation of VOC for the postconsonantal vowel. In contrast, when the final stop was followed by an aspirated stop in the syllable-initial position, the pattern resembled that for the single syllable-initial aspirated type (Figure 8-4).

DISCUSSION

The present study revealed that the three types of Korean stops in the syllable-initial position are characterized by different patterns of VOC activity. For the production of the aspirated stop, VOC is markedly suppressed and then reactivated for the following vowel. Such activity patterns seem to correspond to a marked abduction followed by a quick adduction gesture of the vocal folds, which has been observed in fiberscopic studies.

EMG patterns for the lax stop type can be considered to correspond to the glottal abduction gesture for this type, in which a moderate degree of glottal opening is always observed in fiberscopic studies. A relatively less marked reactivation of VOC after suppression in this case may be due to the fact that the glottal opening is moderate and that the F_0 of the following vowel is relatively low. The apparent increase in VOC activity preceding the consonantal suppression in the case of the initial lax stop shown in Figure 8-1 is considered to be due to the laryngealization of the preceding vowel.

It has been observed by fiberoptic study that the glottal opening is smallest for the forced

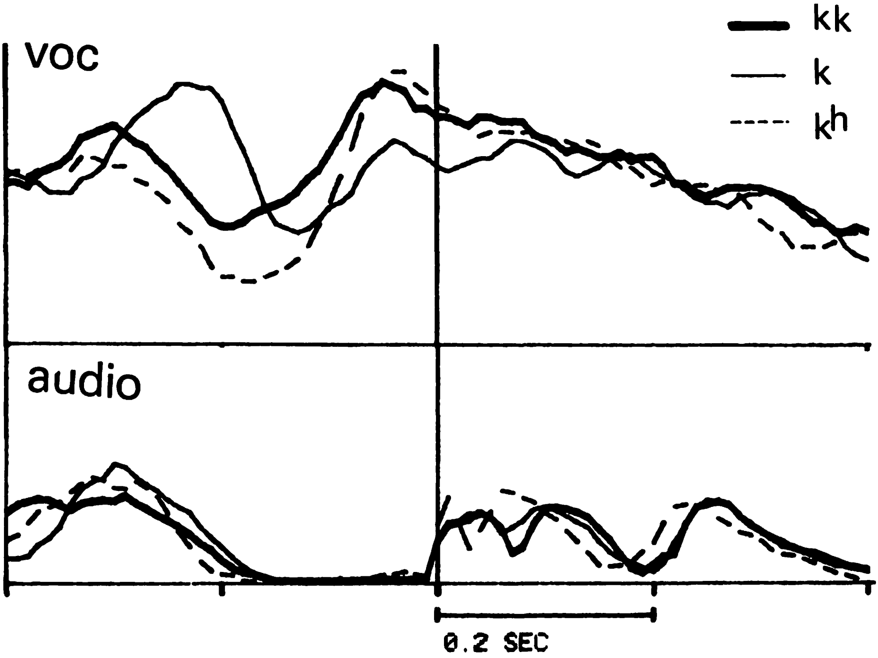


FIGURE 8-1
Averaged EMG curves of the thyroarytenoid (VOC) for the utterance "ike Culita" — where "C" stands for a forced (thick line), a lax (thin line) and an aspirated (dashed line) stop, respectively. The lower curves are averaged audio-envelopes. The line-up for the averaging was taken at the voice onset of the following vowel.

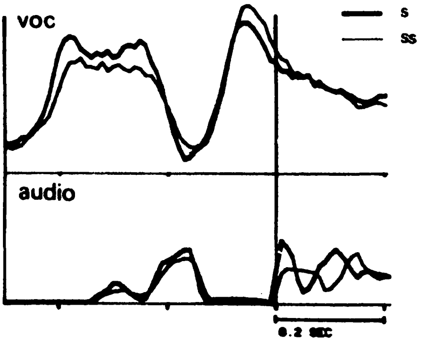


FIGURE 8-2
Averaged EMG curves of VOC for two types of syllable-initial fricatives. The line up point was at the implosion of the consonant.

FIGURE 8-3
Averaged EMG curve of VOC for a syllable-final applosive stop (thin line) compared to that for a similar word ending with a vowel (thick line). The line up was taken at the onset of the vowel of the syllable nucleus.

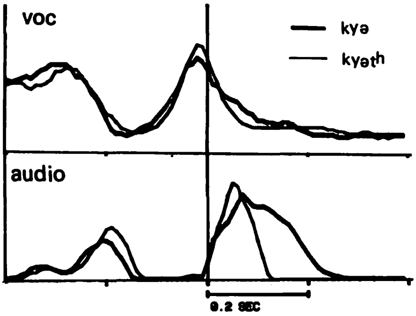
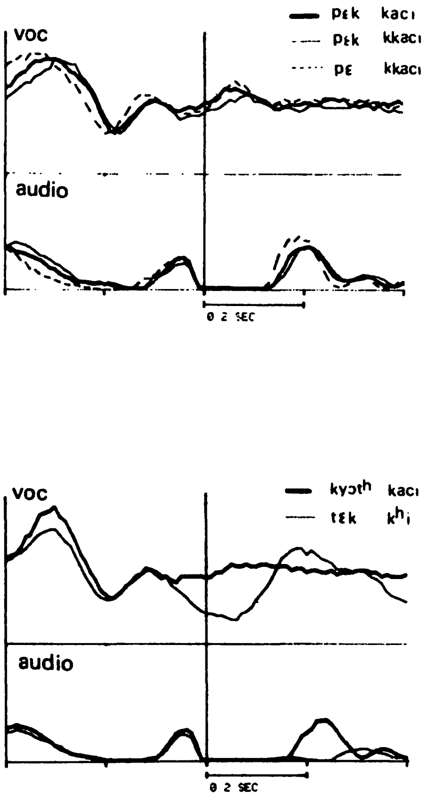


FIGURE 8-4
Above, averaged EMG curves of VOC for the utterances “ike pɛk kacita” (thick line), “ike pɛk kkacita” (thin line) and “ike pɛ kkacita” (dashed line). Below, “ike kyɔt^{h} kacita” (thick line) and “ike tɛk k^{h}ita” (thin line) are included below for comparison.



type among the three stop types, and that the glottis tends to close earlier relative to the voice onset of the following vowel in this case. Apparently, the minimum suppression and early reactivation of VOC activity found in the present study for the syllable-initial forced stop seem to correspond at least to the temporal features of the glottal dynamics for the forced type stop. However, the increase in VOC activity preceding the voice onset of the vowel following the forced type stop, the degree of which is relatively high, cannot be explained solely by the simple dimension of glottal abduction-adduction. Rather, as already suggested by Hirose *et al* (1974), the relatively steep increase in VOC activity for the forced stop type must be taken as a characteristic feature of this type of Korean stop. This activity pattern may correspond to the acoustic feature of "laryngealization" as described by Abramson and Lisker (1972) and Ladefoged (1973), and could be a physiological correlate of the rapid build-up in intensity after the stop release which was found to be characteristic for the forced stop type by Han and Weitzman (1970).

Even so, the degree of increase in VOC activity observed in the three subjects in the present study was less marked than that found in our previous study reported in 1974. The difference may be due to a dialectal difference among the experimental subjects. In our previous study, a subject of the Taegu dialect was examined, in whom the distinction between the forced and lax stops could not be made on the basis of a difference in VOT value. On the other hand, the distinction was actually possible on the basis of the VOT value in the present subjects who were speakers of the Seoul dialect.

Our fiberscopic study revealed that there was no significant difference in the time curves of the glottal gestures between the syllable-initial forced and lax fricatives (Sawashima *et al*, 1980). The present study also failed to disclose any conceivable difference in VOC activity between the two types.

Kagaya (1974), reported a considerable difference in the time course of the glottal closing gesture in syllable-initial forced and lax types of fricatives, the latter being called "aspirated" in his categorization. The gestures of the two types of fricatives in syllable-medial position were similar to each other. In his experiments using a fiberscope, Kagaya used isolated CV syllables without any preceding carrier so that he observed the glottal gestures for the production of fricatives in absolute initial position. Since the syllable-initial fricatives were uttered with a preceding carrier in the present study, the apparent discrepancy between his results and ours regarding the glottal closing gesture may be explained, at least in part, by the different phonetic conditions in the two studies.

As for the syllable-final applosives, VOC activity showed an abrupt decline which may correspond to the transient glottal abduction confirmed by our fiberoptic observation on this type of stop (Sawashima *et al*, 1980). Iwata *et al* (1979) reported that in Fukienese, a southern Chinese dialect, syllable-final applosive stops were associated with a marked adduction of the false vocal folds. Most probably, this type of glottal constriction is produced with a marked increase in VOC activity. In fact, a significant increase in VOC activity is often observed in Japanese and in American English immediately after the cessation of a string of utterances where a tight glottal constriction is confirmed by fiberoptic observations. In any case, the mechanism for the production of applosive stops appears to be completely different in Korean and Fukienese.

The results of the present study indicate that the pattern of VOC activity for Korean clusters consisting of a syllable-final applosive and a syllable-initial forced or lax stop is quite similar to that of the syllable-initial forced stop. It was also shown that the pattern of VOC activity for the cluster type consisting of a syllable-final applosive and a syllable-initial aspirated stop resembles that of the syllable-initial aspirated type. These results are compatible with our fiberoptic observations (Sawashima *et al*, 1980). Thus, the laryngeal feature of the final stop seems to be assimilated by the following syllable-initial stop, regardless of the difference in place of articulation.

The role of the VOC in speech production has not been completely clarified as yet. It has been claimed that this muscle contributed to vocal fold adduction, glottal constriction, laryngealization and creaky voice production. In contrast, its activity is suppressed in vocal fold abduction and — at least for Japanese and American English — for consonantal segments irrespective of the voiced-voiceless distinction (Hirose and Gay, 1972; Hirose and Ushijima, 1978). In singing, it has been reported that the VOC plays an important role in producing "heavy" register (Hirano *et al*, 1970).

The increase in VOC activity observed for the Korean forced stop seems to indicate an "extra strong tensing" (Fischer-Jørgensen, 1977) of the VOC, producing a creaky voice character. In general, creaky voice — or laryngealization — gives a harsh voice quality with a comparatively low pitch. However, in the case of the Korean forced stops, the postconsonantal vowel usually starts with a high F_O. Our preliminary EMG study on the activity of the cricothyroid (CT) muscle revealed that CT activity always increases together with VOC activation for the forced stop type. Thus, the state of the vocal folds for forced stop production may be characterized by a stiffening of its "body" and a strong medial compression associated with relatively high longitudinal tension, resulting in creaky voice quality with a high F_O. Further study is needed to elucidate the nature of the synergetic actions of the laryngeal muscles in the articulatory adjustments for different types of speech sounds, including Korean stops. Possible dialectal differences should also be taken into consideration.

SUMMARY

1. Thyroarytenoid (VOC) activity patterns appeared capable of characterizing the three different types of Korean stops known as "forced", "lax" and "aspirated".
2. In particular, for the production of the forced stop, the VOC shows a marked increase in activity with relatively early timing before the onset of the following vowel. Based on these findings, the state of the vocal folds for the production of this type of stop was discussed.
3. The syllable-final applosive stop is characterized by a steep decline in VOC activity.
4. There is no difference in VOC activity for the two types of Korean fricatives.
5. When a final stop was followed by a syllable-initial stop, the EMG pattern for the cluster resembled that of the syllable-initial consonant.

REFERENCES

Abramson, A.S. and Lisker, L. (1972). "Voice timing in Korean stops," *Proc. of the 7th International Congress of Phonetic Sciences,* Montreal, 1971, the Hague: Mouton, 439-446.

Fischer-Jørgensen, E.(1977). Discussion. In M. Sawashima and F.S. Cooper (Eds.) *Dynamic Aspects of Speech Production*. Tokyo: Univ. of Tokyo Press, 70.

Han, M.S. and Weitzman, R.S. (1970). "Acoustic features of Korean /P,T,K/, /p,t,k/ and /pʰ, tʰ,kʰ/," *Phonetica* 22, 112-128.

Hirano, M., Vennard, W. and Ohala, J. (1970). "Regulation of register, pitch and intensity of voice - an electromyographic investigation of intrinsic laryngeal muscles," *Folia phoniat.* 22, 1-20.

Hirose, H. and Gay, T. (1972). "The activity of the intrinsic laryngeal muscles in voicing control - An electromyographic study," *Phonetica* 25, 140-164.

Hirose, H., Lee, C.Y. and Ushijima, T. (1974). "Laryngeal control in Korean stop production," *J. Phonetics* 2, 145-152.

Hirose, H. and Ushijima, T. (1978). "Laryngeal control for the voicing distinction in Japanese consonants," *Phonetica* 35, 1-10.

Iwata, R., Sawashima, M., Hirose, H. and Niimi, S. (1979). "Laryngeal adjustment of Fukienese stops - initial plosives and final applosives," *Ann. Bull. RILP* 13, 61-81.

Kagaya, R. (1974). "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives," *J. Phonetics* 2,161-180.

Ladefoged, P. (1973). "The feature of the larynx," *J. Phonetics* 1, 73-83.

Lisker, L. and Abramson, A.S. (1964). "A cross-language study of voicing in initial stops: acoustical measurements," *Word* 20, 384-422.

Sawashima, M., Park, H.S., Honda, K. and Hirose, H. (1980). "Fiberscopic study on laryngeal adjustments for syllable-final applosives in Korean," *Ann. Bull. RILP* 14, 125-138.

DISCUSSION

M. ROTHENBERG: The measurements presented by Dr. Hirose support the viewpoint proposed in my doctoral dissertation (Rothenberg, 1968) that there must be a differentiation between the categories "voiced" and "tightly-voiced" in a universal phonetic theory for stop consonant production. In that work it was established from measurements of intraoral pressure and from acoustic pressure recordings made with the bilabial closure bypassed by a short tube aimed at the microphone (thus giving a qualitative measure of tube air-flow) that the Korean "tense" unvoiced stop was produced with a closing or adductory gesture timed roughly 50 msec or less before the articulatory release (fig. 6.5.3 of the dissertation), whether in a prevocalic or intervocalic postion. That the glottal adjustment was of the "tightly-voiced" category (defined on p. 69 of the dissertation) was indicated by the fact that the spectrum of the voicing present in the bypassed air showed a uniformly strong, harmonic-rich voicing, with no sign of breathing-voice transition, and the fact that the time course of the intraoral pressure recordings with no bypass showed a consistent downturn of intraoral pressure before the release. A marked reduction of intraoral (supraglottal) pressure just before the release would be expected when the glottis was closed (a condition of a tightly-voiced adjustment when the transglottal pressure was not sufficient to initiate voicing), and the beginning of the articulatory opening movement caused the supraglottal air volume to increase (as represented by $I_{j,t}$ in figure 2.4.1 of the dissertation).

The theoretical question proposed by this analysis was then to determine how a consistently unvoiced stop was produced when the voiced or tightly-voiced adjustment was attained as long as 50 msec before the release and the transglottal pressure was increased during that 50 msec by the drop in supraglottal pressure caused by the articulatory opening movement. At that time I proposed that to accomplish this, the adductory movement would be expected to be as close to the release as allowed by the physiological constraints on laryngeal-articulatory coordination, and that a tightly-voiced adjustment might be used instead of a voiced adjustment, as had been proposed by others from subjective descriptions of the production of this type of stop. The conjecture here is that with a tightly-voiced adjustment, the vocal folds would stay closed for the smaller transglottal pressure values that occur before the articulatory release, and begin oscillating only when the release brought the transglottal pressure suddenly up to the full subglottal pressure (since the supraglottal pressure drops to near zero at that instant). For this mechanism to be effective (no voicing before the release when the transglottal

pressure is small, but immediate full voicing following the release) we might also expect a somewhat higher than normal subglottal pressure to be used with these stops, to ensure immediate voicing at the release.
The study just reported is significant in that it may have better defined the laryngeal mechanisms that can be used to meet the proposed aerodynamic imperatives in this category of stop.

M. Rothenberg, "The Breath-Stream Dynamics of Simple-Released-Plosive Production," *Bibliotheca Phonetica,* Vol 6, Karger, Basel, New York, 1968.

H.HIROSE: I completely agree with your comment suggesting the contribution of aerodynamic factors to the classification of Korean stops. Earlier research by Smith and Lee certainly indicated the high subglottal pressure, for example.

9 ▪ IS DECLINATION ACTIVELY CONTROLLED?

Carole E. Gelfer, Katherine S. Harris, Rene Collier, and Thomas Baer

INTRODUCTION

It is generally assumed that, for read speech at least, the fundamental frequency of the voice declines over the course of major syntactic constituents. These units correspond to what has previously been termed the "breath group" (Lieberman, 1967; Lieberman, Sawashima, Harris, & Gay, 1970) or "intonation group" (Breckenridge, 1977), being marked on either end by a pause and/or inspiration. The general downdrift of F_O is exclusive of local perturbations secondary to syllable prominence and segmental effects, and is probably best characterized by a steadily declining baseline upon which these local movements are superimposed (Cohen, Collier, & t'Hart, 1982; Fujisaki & Hirose, 1982).

Variations in subglottal pressure (P_s) and cricothyroid (CT) muscle activity are thought to bear most directly on F_O variation, although it has been difficult to separate the CT's contribution to the global prosodic structure of an utterance from its involvement in ongoing local adjustments. However, despite these methodological problems, there has been little evidence to suggest a gradual decline in CT activity corresponding to that in F_O.[1] Rather, the CT's most active involvement in intonation appears to be confined to instances of local emphasis (e.g., Collier, 1975; Maeda, 1976). Subglottal pressure, on the other hand, does exhibit a declination of its own which at least grossly mirrors the F_O contour (Atkinson, 1973; Collier, 1975; Lieberman, 1967; Maeda, 1976), thus suggesting that F_O declination might be a passive phenomenon. However, despite the apparent relationship between P_s and F_O, attempts to establish a direct correlation between the two (Atkinson, 1978; Maeda, 1976) have been unsuccessful in that the drop in F_O exceeds the 3-7 Hz/cm-H_2O that a purely passive model would predict (Baer, 1979; Hixon, Klatt, & Mead, 1971; Ladefoged, 1963).

Some researchers have proposed that declination, and the physiological processes underlying it, is under active speaker control. This assumption derives from observations of variations in some aspects of F_O as a function of utterance length. Cooper and Sorenson (1981), for example, found significant, if not robust, increases in initial peak F_O for progressively longer utterances, while Breckenridge (1977) and Maeda (1976) observed relatively constant values under the same conditions. However, there is a large amount of data to suggest that final F_O values are invariant despite changes in the length of utterances (Boyce & Menn, 1979; Cooper & Sorenson, 1981; Kutik, Cooper, & Boyce, 1983; Maeda, 1976), initial starting frequency (Liberman & Pierrehumbert, 1982), or the insertion of dependent clauses such as parentheticals (Kutik, Cooper, & Boyce, 1983). What the seemingly discrepant results regarding F_O starting frequency would suggest, then, is that, as an utterance increases in length, either the total amount of declination increases or the rate of decline decreases. However, it is not entirely clear whether 1) these are mutally exclusive aspects of F_O declina-

---

Footnote:

[1] This should not be interpreted as meaning that there is not a gradual relaxation of the muscle, but that the EMG activity associated with cricothyroid contraction does not appear to lessen gradually over the course of an utterance.

tion and 2) length-dependent variations in F_O necessarily refute the predictions of a passive model of declination and favor theories involving elaborate speaker pre-planning.

The present study examined the F_O declination, and some physiological variables presumed to underly it, under various linguistic conditions. Our purpose was to elucidate further the relationships among these variables and to speculate as to whether speakers exercise significant control over any or all of them.

METHODS

The subject was a native speaker of Dutch who produced five repetitions of Dutch utterances of three lengths; six, thirteen, and twenty syllables. Mean utterance durations were 1.35, 2.065, and 3.02 seconds, respectively. All three lengths had the first four syllables in common; for the longer utterances, the first eight syllables were identical (see appendix). Each utterance type was also produced in reiterant form, using either the syllable /ma/ or /fa/.[2] The purpose of employing reiterant speech was to neutralize segmental effects while preserving overall intonation and syllable timing (Larkey, 1983; Liberman & Streeter, 1978). In addition, by using syllables with expected differences in airflow requirements, the effect of these differences on subglottal pressure and, possibly, F_O, could be assessed.

For each length condition, emphatic stress was placed either on the first syllable receiving lexical stress (the second syllable in the utterance), the last syllable receiving lexical stress (the penultimate syllable), or both. We will refer to these as early, late and double stress conditions, respectively. In all, there were twenty-seven utterance types (3 phonetic conditions x 3 stress conditions x 3 length conditions).

The results were analyzed with respect to the effects of utterance length and syllable emphasis on initial F_O, P_s, CT, and respiratory activity, as well as the magnitude and rate of decline in each of these variables over entire utterances.

Subglottal pressure was recorded by means of a pressure transducer inserted through the cricothyroid membrane into the trachea. Standard EMG techniques were used to record from the cricothyroid muscle (Harris, 1981). Lung volume was inferred from the calibrated sum of thoracic and abdominal signals from a Respitrace inductive plethysmograph, and F_O was derived from the output of an accelerometer attached to the pretracheal skin surface. A cepstral technique was used to extract F_O from the signal.

RESULTS

Figure 9-1 shows Respitrace comparisons for each phonetic condition across stress types for Length 2 utterances. Within utterances of a given phonetic composition (i.e., Dutch, /ma/ or /fa/), the rate of air expenditure appears to remain constant within a stress condition, as is obvious from the parallel tracings. However, the peak inspiration varies inconsistently across parallel sets.[3] Thus, it appears that, on the respiratory level, local variables such as the degree or place of emphasis were not reflected in the air flow management of this speaker's utterances.

Footnote:

[2] The durations of the reiterant utterances are, on average, somewhat longer than the corresponding Dutch due to the intrinsic duration of /a/ and/or the inadvertent addition of extra syllables. Those tokens for which the latter was evident were still included in all analyses on the assumption that the speaker's intention was to produce an utterance of a given duration, so that any length-dependent adjustments would be identical.

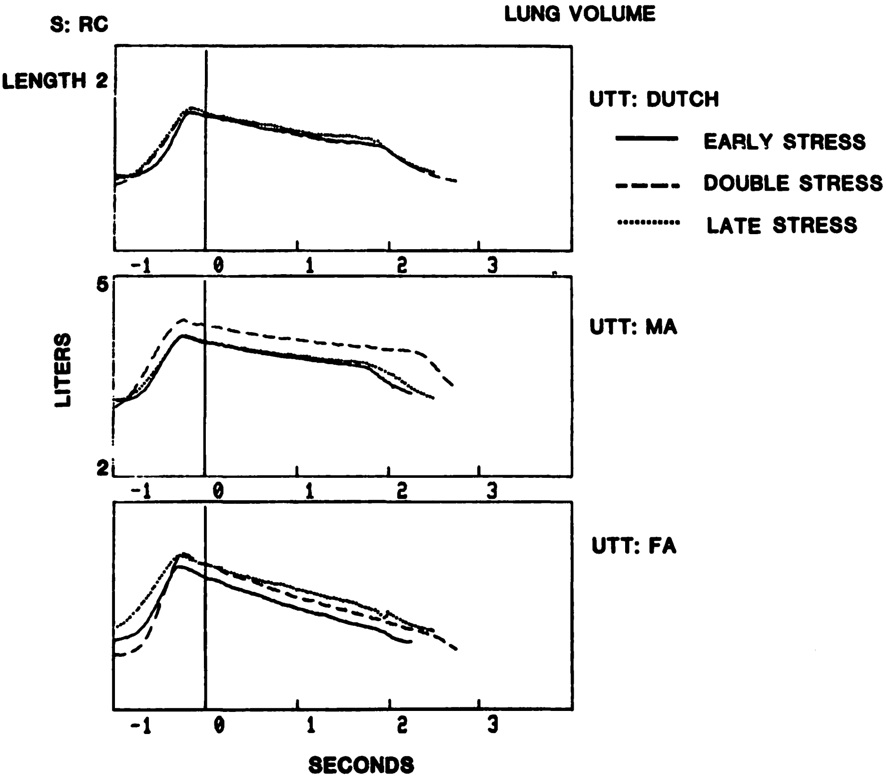


FIGURE 9-1
Comparison of Respitrace curves for Length 2 utterances across all stress types shown for each of the three phonetic conditions.

Across phonetic conditions, however, airflow rates did differ, as is evidenced by the apparent differences in the rate at which these curves decline. The left-hand section of Figure 9-2 shows a comparison of the Respitrace curves for each phonetic condition for early stress across three utterance lengths. It appears that airflow rate for the /fa/ condition always exceeds that for the /ma/ condition, while, for the Dutch, air expenditure is more variable. The obvious question is whether these changes in airflow are reflected in the pressure. From the corresponding subglottal pressure tracings on the right of Figure 9-2, it can be seen that they were not. Furthermore, while local segmental effects are apparent in the P_s curves, par-

---

Footnote:

[3] As used here, peak inspiration corresponds to the maximum amplitude of the output signal of the Respitrace immediately preceding the onset of speech. Because of the built-in filter characteristics of the Respitrace, however, this point may not represent actual peak inspiration. Furthermore, baseline drift and positional changes may introduce artifact into the signal as well. Therefore, peak inspiration measures should be interpreted with caution.

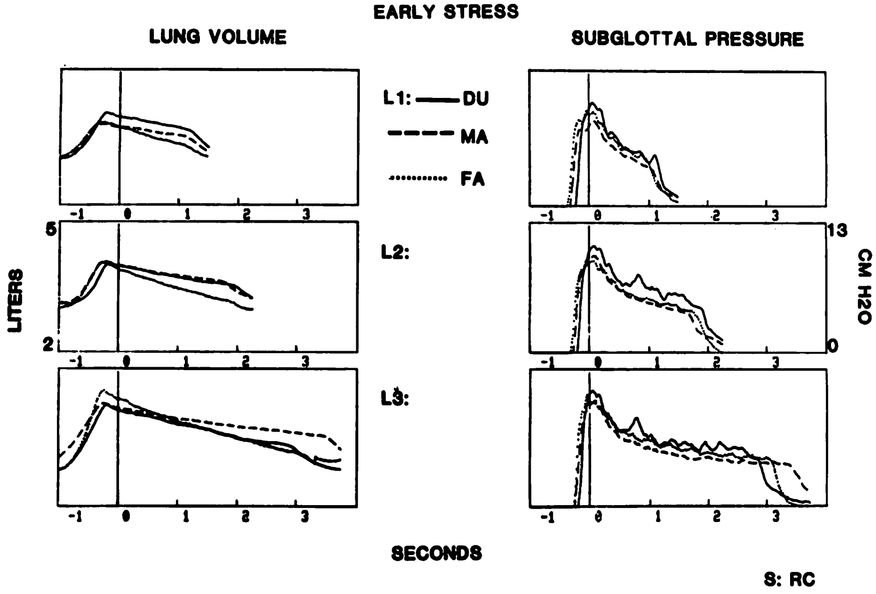


FIGURE 9-2

Corresponding Respitrace and subglottal pressure curves for the early stress condition across the three phonetic conditions. Comparisons are shown for each utterance length.

ticularly for the Dutch utterances, it is also apparent that, for the three comparisons made at each length, a single function could most likely characterize the decline of subglottal pressure, despite the variations in phonetic composition and concomitant airflow characteristics.

Because of this demonstrated uniformity of P_S across phonetic conditions, the remainder of this paper will focus on the analysis of the reiterant /ma/ utterances on the assumption that they are at least generally representative of normal speech.

Figure 9-3 again shows Respitrace curves for each stress type across the three utterance lengths. It can be seen that there are no visually significant differences in the rate of expiration nor evidence of systematic adjustments in peak inspiration as a function of anticipated length. It is the case, though, that the depth of inspiration (with the exception of one utterance type) appears to be adjusted according to utterance length. This is evident from the values in Table 9-1, which shows both the point of peak inspiration and the amount of inspiration (calculated by subtracting the preceding valley from the peak) for each utterance. However, because the experiment was designed in such a way that all tokens of a particular stress type were produced in blocks of utterances of increasing length, it is impossible to determine the significance of this finding. In other words, because Length 3 tokens were always preceded by tokens of the same length or, in one instance, by the last token of the shorter length utterance, inspiration necessarily began at a point lower in this speaker's vital capacity than, for example, Length 2 tokens, which could have only been preceded by tokens of the same length or shorter. Thus, we are unable to determine whether the increase in the depth of inspiration as a function of length represents an artifact of experimental design or evidence of anticipated

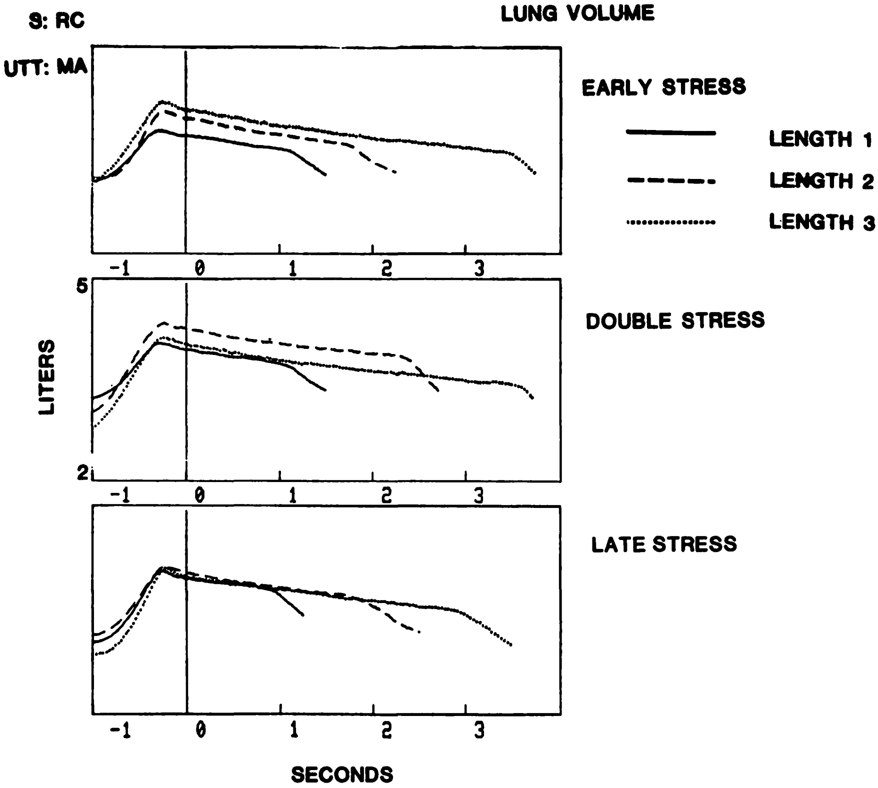


FIGURE 9-3
Respitrace curves for reiterant /ma/ utterances across lengths. Comparisons are shown for each stress condition.

pulmonary requirements.

Overall, the Respitrace data fail to demonstrate conclusively the manner or extent to which this speaker makes prephonatory adjustments of this kind under the various conditions. However, in light of the otherwise uniform nature of these Respitrace curves, and the absence of any obvious relationship with the subglottal pressure, their influence on the ultimate trajectory of fundamental frequency declination appears questionable.

Figure 9-4 depicts the F_0 contours for the three stress conditions for each utterance length. These contours probably represent what has been termed "baseline declination" in as pure a form as possible in that significant segmental effects are absent. For the early and double stress conditions, there is an obvious peak associated with every emphatic syllable, and a consistent initial peak height difference as a function of utterance length. However, F_0 does not decline steadily from these peaks. Rather, there is a rapid drop in frequency to a point from which F_0 then begins a steady decline. While the time course of this initial plunge is identical across lengths, despite differences in peak height, the points from which the slow decline begins for each length are not, bearing instead the same relationship as the initial peaks. This relationship appears to be maintained throughout the course of at least

TABLE 9-1
Peak inspiration (left) and total inspiration (right), in liters, for the three length conditions across all stress types.

	Peak Inspiration (liters)				Amount Inspiration (liters)			
	Early	Double	Late	Mean	Early	Double	Late	Mean
Length 1	3.83	4.06	4.05	3.98	.85	.93	1.17	.98
Length 2	4.1	4.35	4.12	4.19	.99	1.36	1.11	1.15
Length 3	4.23	4.16	4.09	4.16	1.41	1.61	1.26	1.43
Mean	4.05	4.19	4.09		1.08	1.3	1.18	

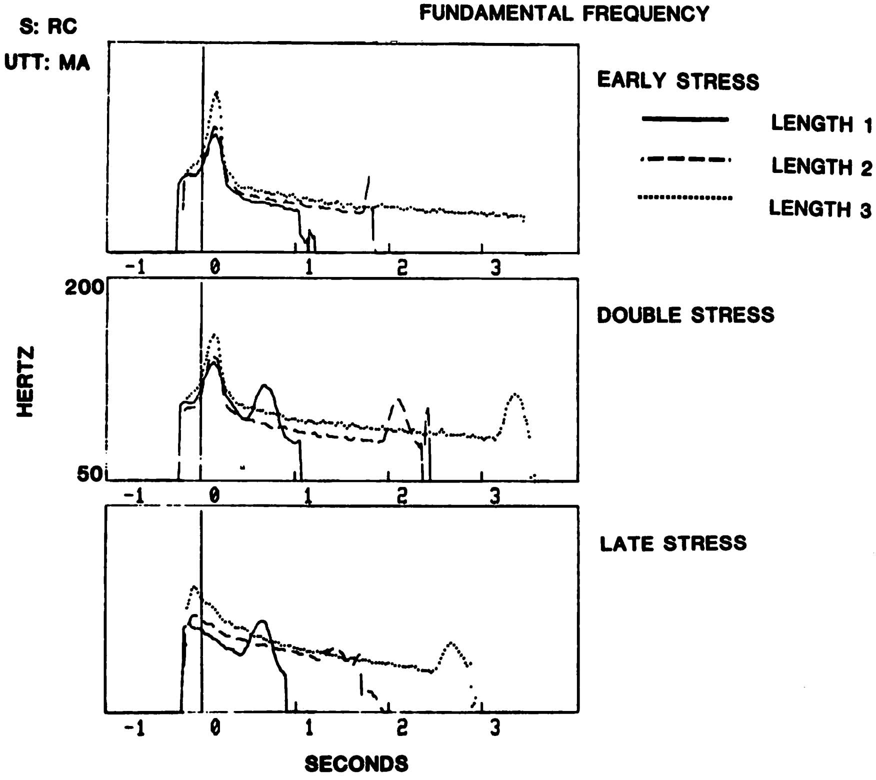


FIGURE 9-4
Fundamental frequency curves for reiterant /ma/ utterances across lengths. Comparisons are shown for each stress condition.

the longer utterances, although they appear to decline in parallel. In the absence of early emphasis in the late stress condition, the F_O peaks occur upon initiation of the utterance and are thus displaced in time relative to the second syllable peaks in the former two conditions. Furthermore, the decline of F_O from these peaks is far more gradual and less strikingly parallel. However, it is of some interest to note that the relationship of these nonemphatic initial peaks across lengths is the same as for their emphatic counterparts.

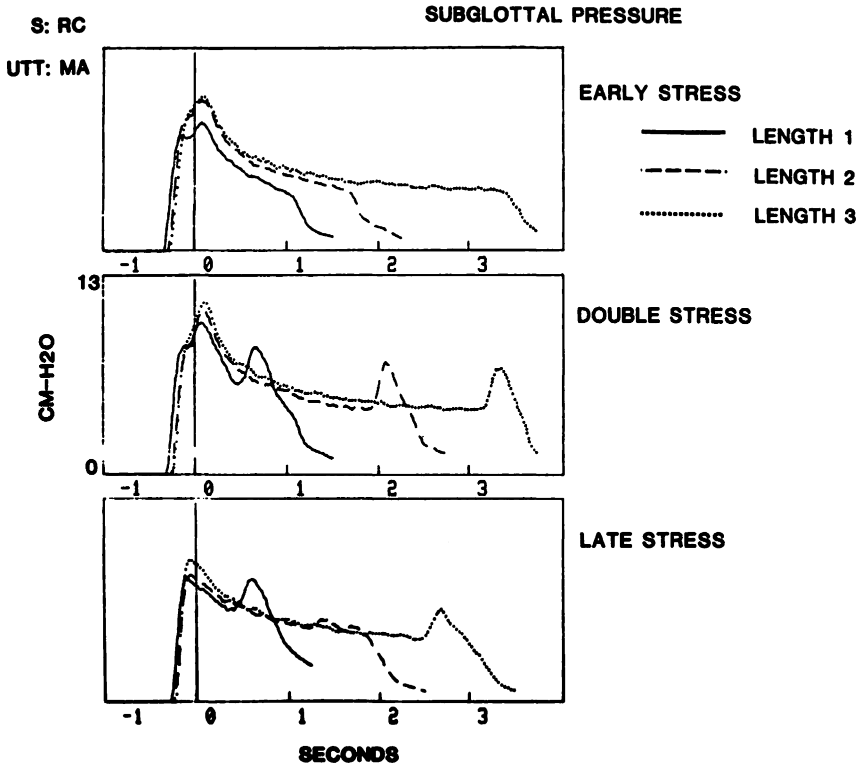


FIGURE 9-5
Subglottal pressure curves for reiterant /ma/ utterances across lengths. Comparisons are shown for each stress condition.

Figure 9-5 shows the corresponding subglottal pressure tracings. It can be seen that the same general tendencies prevail. That is, there is an effect of utterance length on the initial peak pressure and a relatively rapid initial pressure drop into a more-or-less parallel and steadily declining function for the longer utterances. Again, the peaks occur earlier in the late stress condition and the initial pressure drop is less rapid.

If the F_O and P_s tracings are examined in parallel, it becomes apparent that there is a point in time, following the initial peaks, after which the decline in F_O almost mirrors that of P_s. However, the parallelism is less obvious for the earlier portions of early and double stress utterances, since the most rapid drop in P_s is far more gradual than that for F_O.

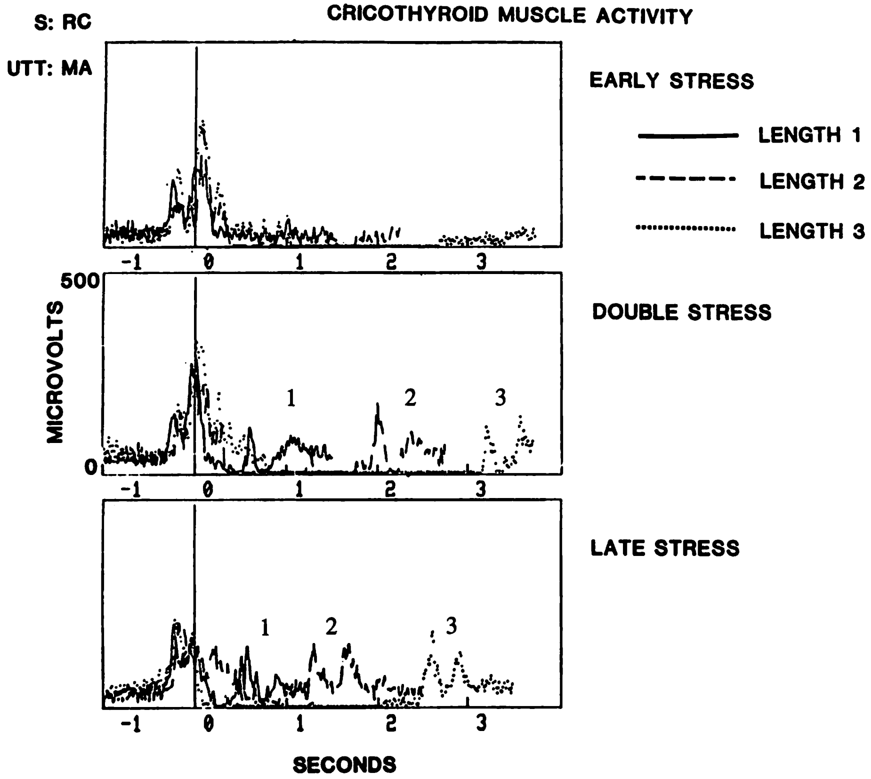


FIGURE 9-6

Averaged cricothyroid muscle activity for reiterant /ma/ utterances across lengths for each stress condition. Final peaks for each utterance length are denoted by the numbers above these peaks.

Figure 9-6 shows CT activity for the three stress conditions. The overall CT pattern differs from those of P_s and F_O in that, although inherently noisy, CT activity appears to be relatively binary in utterances of this form. There are significant increases in CT occurring for initial syllables, whether stressed or not, and for all final stressed syllables, which in this figure are marked in the double and late stress conditions according to their respective lengths. CT activity during the early portion of these utterances is characterized by double peaks whose timing is identical across stress types, but whose relative magnitudes differ with stress type, corresponding to the placement of the P_s and F_O peaks. The double peaks associated with the final stressed syllables of Lengths 2 and 3, however, are the result of averaging events that are distant from the line-up point in tokens of slightly unequal lengths, and are not characteristic of CT activity for final stress peaks.

In order to examine the effect of anticipated length on the initial portions of utterances, we compared initial peak values of CT, P_s, and F_O for each stress type across lengths. It should be recalled that the initial utterance peaks for P_s and F_O in the late stress condition were displaced relative to those with early and double stress, while the timing of the CT peaks

TABLE 9-2

Initial peak measurements of cricothyroid activity, subglottal pressure, and fundamental frequency for the three length conditions across stress types. P_s and F_0 values for the late stress condition do not represent absolute peak values (see text).

		Early	Double	Late	Mean	
	Length 1	202	273	159	211	
CT	Length 2	296	277	169	247	Cricothyroid
	Length 3	310	331	189	277	(μV)
	Mean	269	294	172		
	Length 1	8.3	9.9	7.1	8.4	Subglottal
P_s	Length 2	9.7	10.7	7.4	9.3	Pressure
	Length 3	9.9	11.3	8.2	9.8	(cm-H_2O)
	Mean	9.3	10.6	7.6		
	Length 1	135	137	102	125	Fundamental
F_0	Length 2	141	143	111	132	Frequency
	Length 3	166	158	124	149	(Hz)
	Mean	147	146	112		

remained constant irrespective of stress type. In the interest of consistency, then, the values reported here for P_s and F_O in the late stress condition are those that correspond in time to the peaks for the other two stress conditions and, thus, actually represent values on the declining portion of these curves. The results are shown in Table 9-2. It can be seen that a consistent effect of sentence length obtains for every stress condition for all physiological and acoustic measures.[4]

If the corresponding F_O and CT curves are examined in parallel, there appears to be a close correspondence beteween the time course of the CT suppression and the point at which F_O begins its steadiest decline. We would thus hypothesize that the combined activity of CT and P_s accounts for the behavior of F_O near peaks, but not during the period of F_O slow decline. We acknowledge, of course, that the activity of a number of muscles, not monitored in this study, may also have causal effects on F_O.

Assuming, then, that CT plays little or no active role in F_O declination, we examined the relationship between P_s and F_O in two different ways. First, the amount of drop in F_O and P_s was calculated between the point at which the CT activity ceased and the end of the utterance in the early stress condition, and between the minimum values just preceding the last peak in the double and late stress conditions. In the second analysis, we used the average duration of Length 1 of the early stress utterances as a fixed endpoint and determined the amount and rate of F_O and P_s decline between the offset of CT activity and this fixed endpoint for all utterances.

The offset of CT activity was defined as the time at which the EMG output (measured

---

Footnote:

[4] The actual peak values for P_s and F_O for the late stress condition evidence a similar length effect. They are, in order of increasing length: P_s: 7.9, 8.1, 9.0; F_O: 113, 119, 139.

TABLE 9-3
Analyses of rate of decline in F_0 and P_s across lengths for each stress condition, calculated for (1) the interval from the point of CT offset to P_s minima (variable interval) and (2) the interval from the point of CT offset to a fixed endpoint corresponding to the average duration of Length 1 of the Early stress condition (constant interval). The frequency-to-pressure ratios are also shown for each analysis.

		ANALYSIS 1			ANALYSIS 2		
		FO	Ps	FO/Ps	FO	Ps	FO/Ps
Early	Length 1	22.21	3.94	5.64	22.21	3.94	5.64
	Length 2	14.39	2.47	5.83	19.7	3.73	5.28
	Length 3	7.03	1.07	6.57	17.42	3.95	5.2
Double	Length 1	—	—	—	—	—	—
	Length 2	15.37	2.38	6.46	22.52	4.12	5.47
	Length 3	10.76	1.36	7.91	19.75	3.49	5.66
Late	Length 1	—	—	—	—	—	—
	Length 2	20.79	2.57	8.09	17.32	2.61	6.64
	Length 3	16.85	1.56	10.8	35.22	3.52	10.01

in microvolts) dropped to, and remained below, a level equivalent to the baseline plus 10% of the peak level. These analyses were not performed on Length 1 of the double and late stress conditions. In the former condition, the interval between CT offset for the first peak and CT onset for the second was too short. In the latter condition, CT activity was never consistently suppressed, so that an offset time could not be obtained. Furthermore, in both cases, the designated interval for the second analysis extended into the final stress peak. These analyses were performed on a token-by-taken basis in order to accommodate variability in the timing of CT activity.

Table 9-3 shows the results of both analyses in terms of F_O slope (Hz/sec), P_s slope (cm-H_2O/sec) and the frequency-to-pressure ratios (Hz/cm-H_2O). Looking first at the ratios from both analyses, it should be noted that six of the seven values from Analysis 2 fall within the acceptable range of 3-7 Hz/cm-H_2O, while only four of the seven values from Analysis 1 fall within this range. However, even those values that fall outside the range are considerably lower than those reported when the effects of CT and possibly other muscle activity are not neutralized (see Maeda, 1976). Thus, a passive mechanism whereby F_O declination is determined by a steadily falling subglottal pressure should be reconsidered.

DISCUSSION

As for the influence of utterance length on the slope of F_O and P_s change, the results of Analysis 1 show a substantial decrease in the rate of change with increasing length. This effect has been observed in previous studies and assumed to represent high level preplanning whereby certain physical aspects are represented in a speaker's utterance plan. However, when slope is calculated over fixed portions of these same utterances, as in Analysis 2, this effect is substantially lessened, demonstrating a more constant rate of decline across lengths. (For

Length 3 of the late stress condition, there is probably some peculiarity in the data, particularly for F_0.) The results of the latter analysis further suggest that neither F_0 nor P_s decline at a constant rate across an entire utterance. If they did, we would expect the slopes to be identical over any portion of a given utterance, despite its length. However, the results of Analysis 1 demonstrate that this is not the case. It appears that, with the obvious exception of the late stress utterances, the rate of decline in P_s and F_0 is greatest earlier in an utterance, as is indicated by the steeper slopes in the second analysis, and that these curves would be best characterized by an exponential function. Thus, the apparent "length effect" that we and others observe when slope is calculated over an entire utterance can probably be attributed to the nonlinear nature of F_0 declination and not to elaborate precalculations or ongoing reorganization on the basis of utterance length. Our data substantiate the claims of Liberman and Pierrehumbert (1982) that the F_0 contour gradually approaches an asymptotic value as well as Maeda's finding that the latter portion of some utterances may not show any evidence of declination.

The systematic adjustments in peak F_0 suggest that, on some level, this speaker does take sentence length into account. However, these peaks do not appear to influence the trajectory of the total declination contour. Rather, their influence appears to be limited to their immediate vicinity, probably including the frequency from which declination actually begins. However, the latter is probably a function of temporal constraints whereby, in a fixed amount of time, the frequency to which F_0 falls is a function of the frequency from which it starts. Thus, whatever its purpose, manipulating peak height does not appear to be essential to the realization of declination, per se.

In summary, we have found that, for reiterant utterances composed of voiced continuants, where normal segmental adjustments were presumed to be neutralized, CT activity was prominent in instances of emphatic syllable stress, and relatively inactive elsewhere. Subglottal pressure, on the other hand, showed a gradual decline before and/or after stress peaks and was paralleled by a falling fundamental frequency. Thus, while we cannot rule out effects such as vocal fold relaxation on F_0 and P_s during these intervals, the data do suggest that, where CT activity is negligible, F_0 declination can be accounted for on the basis of a falling P_s alone.

Our conclusions at this point must be tentative for two reasons: First, because we have analyzed the data of only one subject and second, because there are inconsistencies between the late stress utterances and the other two stress conditions. However, we believe there are strong indications that certain aspects of F_0 are the result, not of high-level (i.e. cognitively-generated) planning processes, but of the intrinsic behavioral properties of underlying physiological systems.

ACKNOWLEDGEMENT

This work was supported by NINCDS Grants NS-13870 and NS-13617 to Haskins Laboratories.

APPENDIX

Early Stress

Length 1: Je weet dat jan nadenkt.
Length 2: Je weet dat jan erover nadenkt te betalen.
Length 3: Je weet dat jan erover nadenkt ons daarvoor met genoegen te betalen.

Double Stress

Length 1: Je weet dat jan nadenkt.
Length 2: Je weet dat jan erover nadenkt te betalen.
Length 3: Je weet dat jan erover nadenkt ons daarvoor met genoegen te betalen.

Late Stress

Length 1: Je weet dat jan nadenkt.
Length 2: Je weet dat jan erover nadenkt te betalen.
Length 3: Je weet dat jan erover nadenkt ons daarvoor met genoegen te betalen.

REFERENCES

Atkinson, J.E. (1973). "Aspects of intonation in speech: Implications from an experimental study of fundamental frequency," unpublished doctoral diss., Univ. of CT.

Baer, T. (1979). "Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes," *J. Acoust. Soc. Am.* 65, 1271-1275.

Boyce, S. and Menn, L. (1979). "Peaks vary, endpoints don't: Implications for linguistsic theory," Proc. from the 5th Ann. Meeting of the Berkeley Linguistics Soc.

Breckenridge, J. (1977). "Declination as a phonological process," Murray Hill, N.J. *Bell Sys. Tech. Memo.*

Cohen, A., Collier, R. and 'tHart, J. (1982). "Declination: Construct or intrinsic feature of pitch?" *Phonetica* 39, 254-273.

Collier, R. (1975). "Physiological correlates of intonation patterns," *J. Acoust. Soc. Am.* 58, 249-255.

Cooper, W.E. and Sorenson, J.M. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.

Fujisaki, H. and Hirose, K. (1982). "Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation," Preprints of papers, Working Group on Intonation, The 13th International Congress of Linguistics, Tokyo, 57-70.

Harris, K.S. (1981). "Electromyography as a technique for laryngeal investigation," Proc. of the Conf. on the Assessment of Vocal Pathol. *ASHA Reports* 11, 70-86.

Hixon, T.J., Klatt, D.H. and Mead, J. (1971). "Influence of forced transglottal pressure changes on vocal fundamental frequency," *J. Acoust. Soc. Am.* 49, 105.

Kutik, E.J., Cooper, W.E. and Boyce, S. (1983). "Declination of fundamental frequency in speakers' production of parenthetical and main clauses," *J. Acoust. Soc. Am.* 73, 1723-1730.

Ladefoged, P. (1963). "Some physiological parameters in speech," *Lang. and Speech* 6, 109-119.

Larkey, L.S. (1983). "Reiterant speech: an acoustic and perceptual evaluation," *J. Acoust. Soc. Am.* 73, 1337-1345.

Liberman, M. and Pierrehumbert, J. (1982). "Intonational invariance under changes in pitch range and length," *Bell Labs Internal. Tech. Memo.*

Liberman, M. and Streeter, L.A. (1978). "Use of nonsense-syllable mimicry in the study of prosodic phenomena," *J. Acoust. Soc. Am.* 63, 231-233.

Lieberman, P. (1967). *Intonation, Perception and Language*. Res. Monograph #38. Cambridge: MIT Press.

Lieberman, P., Sawashima, M., Harris, K.S. and Gay, T. (1970). "The articulatory implementation of the breath-group and prominence: Crico-thyroid muscular activity in intonation," *Language* 46, 312-327.

Maeda, S. (1976). "A characterization of American English intonation," unpublished doctoral diss., Massachusetts Institute of Technology

DISCUSSION

M. HIRANO: The fact that Fo is related to Ps after the early stress is very interesting, because it suggests that there are some factors in Ps which somehow regulate Fo. One thing we must consider is that Ps is a parameter determined by various multiple muscles. Therefore, your result may be a manifestation that there are some parameters which cause a decrease in Ps as well as a decrease in Fo.

K. HARRIS: This study suffers from the limitations of all essentially correlational studies—that it does not show cause directly.

I. TITZE: I would like to propose a biomechanical explanation for declinatin in Fo. Our stress relaxation measurements on vocal fold tissue show between 10% and 50% decrease (after step elongation) in longitudinal tension in the time periods you show. It appears that Fo declination may be directly dependent on tissue relaxation. As I have suggested previously, subglottal pressure seems to affect Fo indirectly through nonlinear, amplitude-dependent stiffness. A sudden length increase, however, would explain the higher initial Fo with increased initial tension, which then relaxes.

K. HARRIS: As we said in response to Dr. Hirano, it may be that the correlations we show are the surface reflections of some deeper underlying phenomenon.

T. BAER: Well, I don't agree that it's been established that subglottal pressure shouldn't increase pitch. There are, of course, a number of experiments, including my own, that show that when subglottal pressure is changed as an isolated variable, in the absence of other changes, pitch goes up.

I. TITZE: Yes, but have you ever decoupled subglottal pressure from vibrational amplitude? If you could show that the vibrational amplitude stays constant when you change subglottal pressure and then you get a change in Fo, I'll buy the argument. But as long as you increase the amplitude of vibration, and you drive the tissue non-linearly, you will then always get an effective change in stiffness of the tissue.

T. BAER: Well, I don't disagree that there are going to be nonlinear effects, and perhaps non-linear effects can account for the increase in pitch. But I'm not quite sure how you get from there to relaxation over the period of a couple of seconds.

I. TITZE: Well, I'm saying that there may be two mechanisms for declination of Fo. One would be tissue relaxation, which does not require an explanation directly on the basis of subglottal pressure change. It is still related indirectly to a stiffness change, on the basis of what we know. I agree with you that phenomenologically, there is a correlation betwen Ps and Fo, but I believe that's always through a changing stiffness.

T. BAER: Well, I think, that depends on how you define stiffness.

K. HARRIS: Could I make one other comment on this particular issue? We would have gotten a somewhat different picture of what was going on at the larynx if we had also looked at strap muscles. That is, what we're looking at is the muscles that again phenomenologically appear to have a lot to do with Fo control at the top of the range, but probably not so much as Fo itself decreases. That is, in lower ranges it may well be that the relationship between Fo and subglottal pressure or Fo and strap muscle activity is different from what it is higher in the range.

M. ROTHENBERG: A claim that Fo varies with subglottal pressure can have at least two meanings. The first is simply that a variation in Fo is the direct result of a change in subglottal pressure, with the laryngeal adjustment held constant. This claim can be easily tested. The second, more complex argument is that when the subglottal pressure changes there is some relatively automatic change of laryngeal adjustment to one that is somehow more natural, efficient or otherwise suitable for the new pressure, and that this change in laryngeal adjustment, when combined with the change in subglottal pressure, causes the change in Fo.

C. LARSON: I'd like to make a comment, Ingo, on your nonlinearity suggestion on relaxation of tissue. The cricothryroid EMG activity dropped off markedly, suggesting a nonlinearity. Doesn't that EMG change go in just the opposite direction of your suggestion? That is, the EMG dropped off. Whereas, if it were tissue relaxation causing the declination in Fo, wouldn't you expect to see almost an increase in cricothyroid activity during the course of that declination as an active means of compensating for the tissue relaxation?

I. TITZE: But we don't know about vocalis, if it drops off too, then those effects could compensate somehow.

B. WYKE: I'm afraid to get into this argument. I'd like to suggest that maybe there is a solution in that, as I hope to show later this afternoon, there is a physiological mechanism that couples the providing activity of muscles like the cricothyroid and the vocalis to the prevailing subglottal pressure as long as that pressure is greater than 3cm H_20. There are some very sensitive mechanoreceptors in the subglottal mucosa, the frequency of which is constantly proportional to variations in subglottal pressure. These, in turn, reflexly affect the tension in the vocal fold musculature and the stiffness of the vocal fold. And I think this might be relevant.

10 ▪ VARIABILITY ANALYSIS OF LARYNGEAL MUSCLE ACTIVITIES

Kiyoshi Honda

INTRODUCTION

A variety of voiced sounds are produced by the coordinated activity of the larynx and the supra-laryngeal articulators, using the respiratory air flow as a source energy. This process has been modelled as a simplified form in the source-filter concept, where the source generator simulates the vocal fold vibration and a couple of filters represent the acoustic characteristics of the vocal tract. In this theory, these two subunits of the speech production mechanism are supposed to be controlled independently. While the larynx regulates the source waveforms, the articulators such as the tongue and the jaw determine the filter characteristics. However, in production of human speech, such an independent control is not likely to occur, since the articulatory and phonatory organs interact with each other by means of biomechanical interactions between them. An inspection of the anatomical configuration of the vocal tract organs reveals that all the structures are networked by muscles and ligaments, suggesting that movements of the organs have an inter-dependency with each other. The tongue articulation, for example, may apply a force to the larynx, modifying its phonatory function. Contrarily, muscular control for changing vibratory mode of the vocal folds can produce a substantial change in the vocal tract shape as a result of the external frame function by the supra-laryngeal articulators. Thus, natural speech produced by humans appears to be realized by a complex control across organs.

The previous studies (Honda, 1983; Honda *et al*, 1982) have described a biomechanical interaction between vowel articulation and pitch control. The experimental results suggest that the tension of the vocal folds is affected by horizontal, rather than vertical, movements of the hyoid bone produced by tongue articulation. In high vowels, the posterior fibers of the genioglossus muscle produce forward movements of the tongue root and the hyoid bone, which can rotate the thyroid cartilage and increase the longitudinal tension of the vocal folds. This mechanism exemplifies a type of the larynx-vocal tract interaction and proposes an implication on the physiological mechanism for the intrinsic pitch of vowels (the tendency of vowel-to-vowel variation in FO). Furthermore, it has been suggested that the vocal tract shape may systematically change with fundamental frequency due to the activity of tongue muscles to enhance pitch control. In the vowel /i/ in high FO, the articulation tends to be extreme due to increased activity of the genioglossus posterior to pull the hyoid bone further forward. Thus, the variability in acoustic characteristics of speech sound reflects "the physiological constraints of speech organs", providing an aspect of the naturalness of human speech.

This study presented here is concerned with the biomechanical influences between tongue articulation and pitch control. The variability of physiological signals during vowel production was statistically analyzed using a correlation technique. First, the mechanism of tongue articulation and its effect on laryngeal function were examined using electromyographic and acoustic signals. Second, the analysis was performed on the intrinsic laryngeal muscles and a few vibratory paramenters in order to predict an evidence of laryngeal adjustments to the articulatory influences.

METHODS

This study includes two separate experiments on a subject of an American English speaker. Table 10-1 shows the muscles and the utterances used in the experiments. In the first experiment, electromyographic (EMG) signals from six tongue muscles and the cricothyroid muscle were recorded during production of the 11 English vowels in isolated /əpVp/ words. Jaw movements were monitored simultaneously by a photo-electronic device using the position sensing detecter (PSD). The data for the vowels /i/ and /ɑ/ were analyzed in this study. In the second experiment, EMG data from the intrinsic laryngeal muscles were measured for the same vowels during the utterances /pip/ and /pɑp/ with a frame sentence. The electroglottogram was used to monitor the glottal vibration. All the EMG measurements were taken by inserting bipolar hooked-wire electrodes which were newly devised (Honda *et al*, 1983), and the integrated signals were used for the processing and the analysis. Audio signals were also collected for extracting acoustic parameters. The pre-processings of the collected data were performed by the Physiological Signal Processing (PSP) system of Haskins Laboratories, and finally, the correlation analysis was made by use of a microcomputer.

TABLE 10-1
List of Muscles and Utterances

1st Experiment	2nd Experiment
Genioglossus Post. (GGP)	Lateral Cricoarytenoid (LCA)
Genioglossus Ant. (GGA)	Cricothyroid Anterior (CTa)
Geniohyoid (GH)	Cricothyroid Posterior (CTp)
Mylohyoid (MH)	Inter-arytenoid (INT)
Hyoglossus (HG)	
Styloglossus (SG)	
Cricothyroid (CT)	
/əpip/, /əpɑp/ (in isolation)	/pip/, /pɑp/ ("Guess ______ again.")

Correlation analysis was performed on the token-to-token variability of the measured data. The level of EMG activity associated with each token for each of the muscles was represented by a segment average, obtained by summing the activity within an appropriate window. The window for the first experiment was 200 msec in length centered at 20 msec prior to the voice initiation and 100 msec for the second experiment centered at the point of the voice initiation. The other measures (such as FO, F1, F2, and jaw movement) were analyzed on the same basis with a suitable window position adjusted to the vowel segment. The left side of Figure 10-1, for example, shows the segments for the analysis during production of the ten tokens from the genioglossus posterior muscle. The distribution of segment averages for each muscle was standardized and compared across measures in the joint histogram as shown in the right side of the figure. Here, for example, the relation between the genioglossus (GGP) and the cricothyroid (CT) are demonstrated, showing that these two measures were positively correlated.

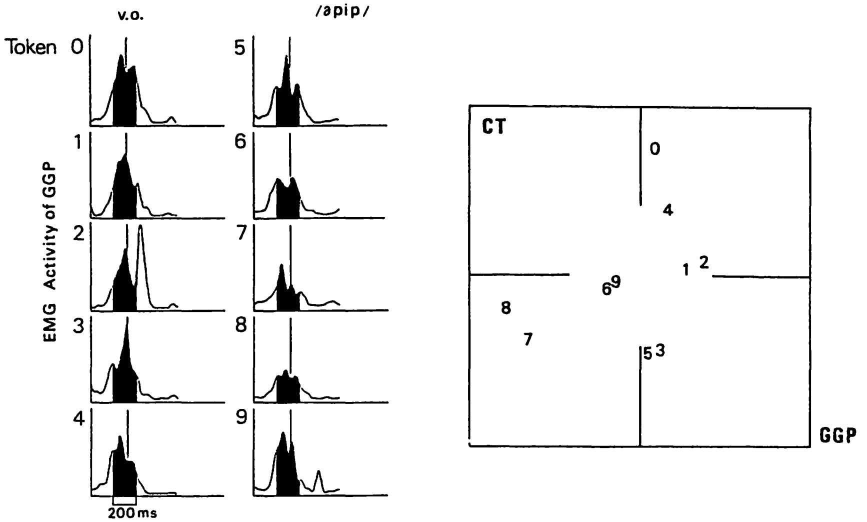


FIGURE 10-1

The segments used for the analysis and an example of the joint histogram. The left side of the figure indicates the temporal patterns for the 10 tokens of the genioglossus posterior muscle during production of /əpip/. The segments used for the analysis on the token-to-token variability are shown by the dark areas. The right side of the figure demonstrates an example of the joint histogram for the relationship between the genioglossus posterior (GGP) and the cricothyroid (CT) muscles. The numeric characters on the figure represent the standardized distributions of the segment averages for the two muscles.

RESULTS AND DISCUSSION

1st Experiment

The EMG and formant frequency data from the first experiment were analyzed for each vowel separately, and compared with FO variation. In the following joint histograms, each data point of the segment average for a pair of measures is marked by the number from 0 to 9, representing the ten tokens of the utterance, so that a multi-dimensional consideration on the data distribution could be done. Table 10-2 shows the statistical values of the acoustic data.

Production of the vowel /i/

Figure 10-2 gives the variability of the tongue EMG activity and the two formant frequencies in the ten repetitions of /əpip/ utterance, compared with the FO variation. In all the pairs of the measures in the figure, some of them show evidence of correlation. Among the tongue muscles, the anterior and the posterior genioglossus muscles showed a higher positive correlation with FO. These muscles are most relevant for producing the tongue shape for /i/. GH and MH do not have a clear relation with FO. HG and SG have some correlation, but their influence on the tongue shape seems to be small, because they are less active in

TABLE 10-2
Statistical Values of Acoustic Data

Vowel		Mean	Maximum	Minimum	S.D.
/i/					
	FO	161.8	180.6	152.5	8.3
	F1	300.6	313.2	281.8	9.1
	F2	1833.7	1858.7	1788.9	20.5
/ɑ/					
	FO	143.0	152.8	135.9	4.6
	F1	634.5	649.2	626.1	6.9
	F2	1004.1	1017.5	981.9	11.1

this vowel. The acoustic parameters indicate that the articulatory configuration was not identical across tokens. F1 showed a negative correlation, and F2 showed a positive correlation with FO. These EMG and acoustic data indicate that the articulation of the vowel /i/ is more extreme when FO is higher. It appears that the variation in the formant frequency data results from the variability of the genioglossus activity to enhance pitch control for this vowel.

Production of the vowel /ɑ/

Figure 10-3 shows the data for the utterance type /əpɑp). In this vowel, correlations between each muscle and FO are less clear. However, the formant frequency data indicate that F2 alone showed an obvious correlation with FO. It is interesting that F2 shows a positive correlation with FO, because the data indicate that the articulatory configuration is more extreme when FO is lower. This is probably because the EMG data for both HG and SG show a weak negative correlation with FO. Since these muscles show a significant activity in the vowel /ɑ/, and they have mechanical connections with the hyoid bone, the hyoid bone may be pulled backward due to increased activity of these muscles for producing the vowel /ɑ/. Although the effect of HG and SG on hyoid bone position does not seem to be so significant as observed in GGP activity for the vowel /i/, this result suggests that low FO tends to be observed in an extreme articulation of the vowel /ɑ/.

Summary for the 1st Experiment

The original data for the tongue EMG used in this study is already reported by Alfonso *et al* (1982). According to their study, the difference in the EMG activity is more remarkable in the front vs. back comparison in tongue position rather than high vs. low comparison. GGP alone contributes to a high vs. low distinction. In the vowel /i/, the two genioglossus muscles (GGA and GGP), the mylohyoid (MH), and the genioglossus (GH) are active. It is noted that GGP raises the tongue dorsum synergistically with GH and MH, whereas GGA seems to pull the midline of the dorsum down to maintain a constricted space in the front part of the vocal tract. In the vowel /ɑ/, the hyoglossus (HG) and the styloglossus (SG) show a relevant activity to draw the tongue down and back. The analysis in this study and in the previous report (Honda, 1983) revealed that the variability of each muscle activity was com-

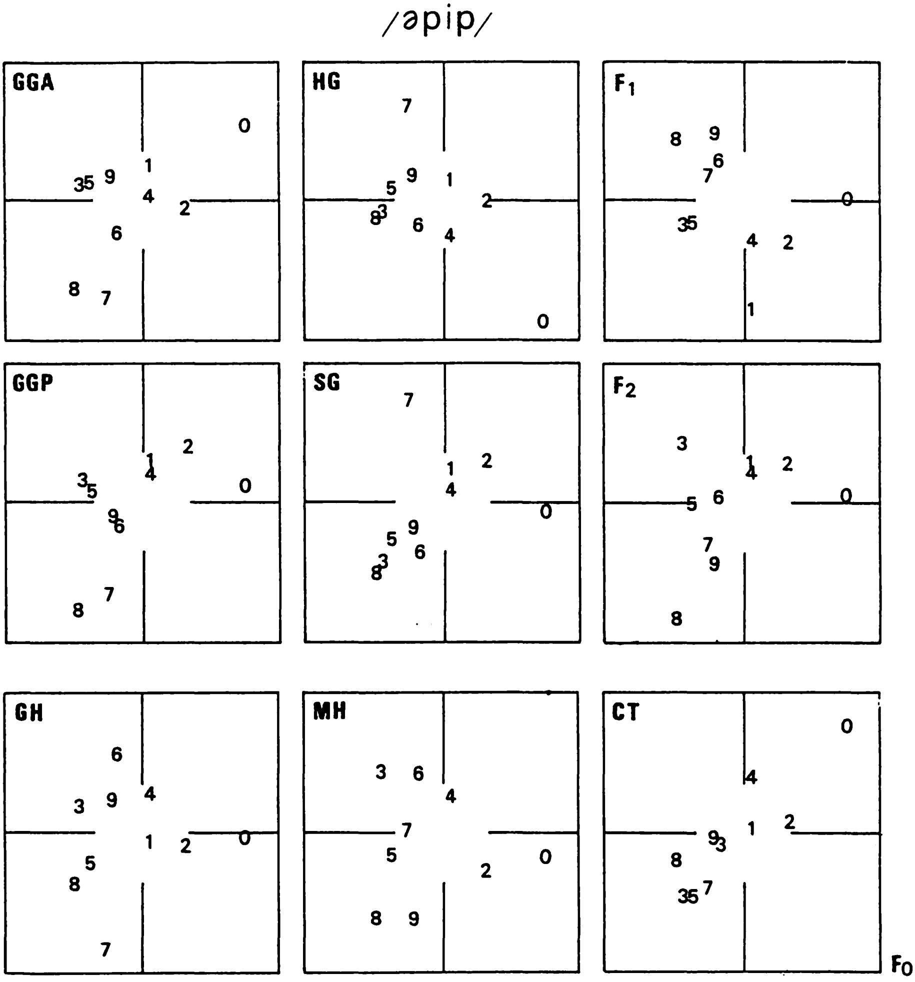


FIGURE 10-2
The token-to-token varability in the EMG data and the formant frequencies (F1 and F2) in /əpip/, in comparison with the fundamental frequency (FO in abscissa).

pensated by each other, finally resulting in a smaller variation in the vocal tract shape. However, a slight difference in the vocal tract shape still remains, corresponding to the activity variation of the most significant muscle(s) such as GGP for /i/, and HG and SG for /ɑ/. This notion is also supported by the variability of F1 and F2.

Figure 10-4 gives a tongue model representation for the compensatory effect and the remaining variation in the vocal tract shape. This simple model is built in order to visualize the effect of each muscle activity on the vocal tract configuration for the vowel /i/. The shape of the tongue is determined by the segment averages of all the EMG and the jaw movement data obtained in this experiment. The left side of the figure demonstrates the effect of GGP activity alone on the vocal tract shape. The remaining parameters for the other measures were

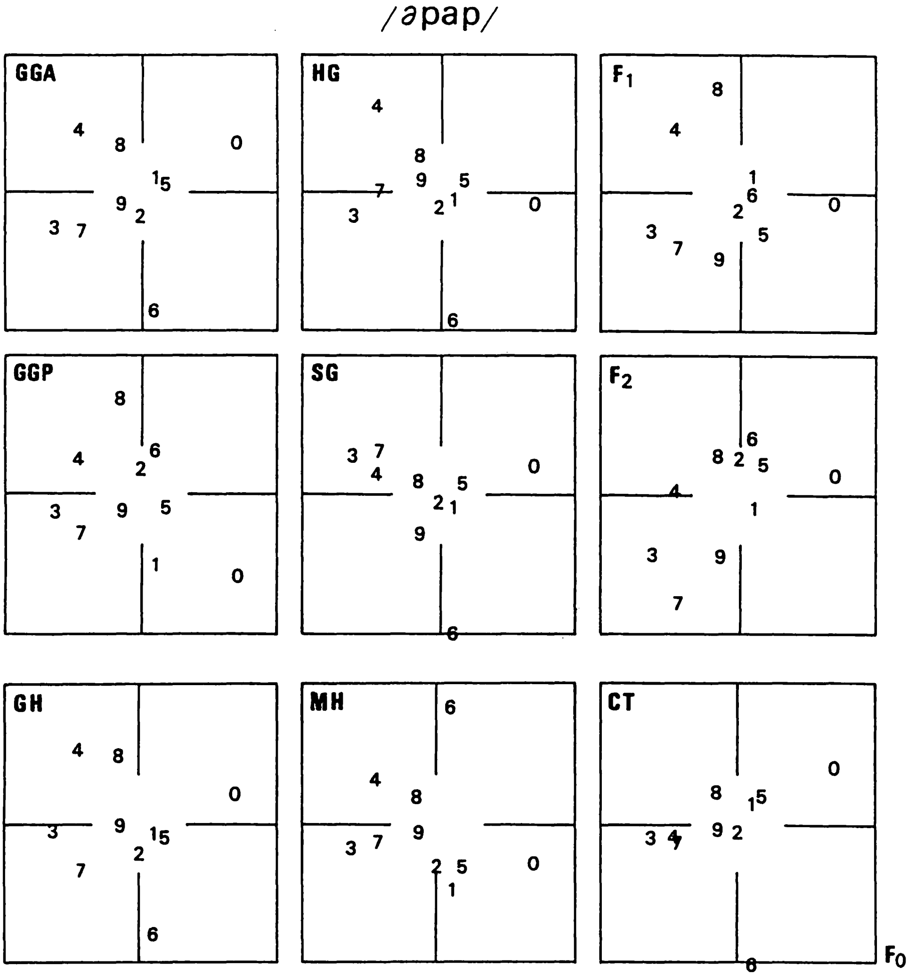


FIGURE 10-3
The same as Figure 2 in /əpɑp/.

kept in the mean values for the ten tokens. The difference in the tongue shape shown in the figure is considerably large for the same vowel quality. In low FO, the activity of GGP was almost equal to the mean of GGP activity in the ten tokens of the lax vowel /I/. The right side of the figure indicates that the effect of GGP variation is compensated by the other parameters, effectively reducing the variability in the vocal tract shape. However, the constriction of the vocal tract is still greater in the token with higher FO than with lower FO. Note that the size of the pharynx is larger and that the hyoid bone position is more forward in higher FO. These differences in the tongue shape indicate the predominant effect of the genioglossus posterior muscle. From a view point of the external frame function, it is reasonable to consider that the effort to produce a high pitch phonation of the vowel /i/ in-

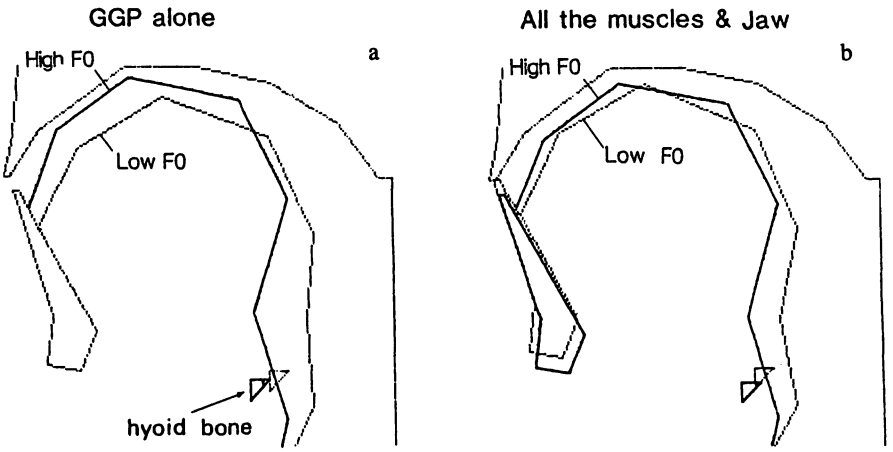


FIGURE 10-4

Tongue model representation of the effect of GGP activity on the vocal tract shape and compensations by the other factors. Figure 10-4(a) is drawn by the data for the EMG of GGP alone in the tokens with high and low FO. Figure 10-4(b) is determined by all the measures in the same tokens.

tion of the extrinsic laryngeal muscles and results in the change in the tongue shape associated with FO variation.

However, for the vowel /ɑ/, HG and SG do not have such a suitable anatomical relation with the hyoid bone as to help raising FO. Contrarily, they may lower FO by drawing the hyoid bone backward when their activity is increased. Therefore, a lower FO is predicted by an extreme articulation of /ɑ/. The result of the analysis for the vowel /ɑ/ supports a possibility of this negative correlation between FO and the articulatory effort. It is suggested that a phonation of /ɑ/ in high FO may require a reduced activity of HG and SG. It may be also possible that GH is involved for raising FO intensionally.

2nd Experiment

The variability in the tongue muscle activity described above may require some adjustments by the activity of the intrinsic laryngeal muscles in order to maintain a consistent quality of phonation. The EMG data from the 2nd experiment were analyzed for both vowels together, and also compared with FO. In the following joint histograms, the 16 data points for each utterance type are represented by the character markings, 'i' and 'a', obvious to the vowel types.

Figure 10-5 shows the variability in the cricothyroid (CT) activity for both vowels. CTa and CTp represent the anterior and the posterior part of the cricothyoid muscle, respectively, as shown in Figure 10-5(a) and Figure 10-5(b). They are supposed to correspond to the anatomical subdivisions of the pars recta and the pars obliqua of this muscle. The FO value obtained from the experiment varied from 150.0 to 183.7 Hz in /i/, and from 133.0 to 149.4 Hz in /ɑ/, in agreement with the tendency of the intrinsic pitch. Although the cricothyroid is the primary pitch raising muscle, the correlation analysis in this experiment did not entirely indicate a good linear relation between CT activity and FO. In comparison of the two

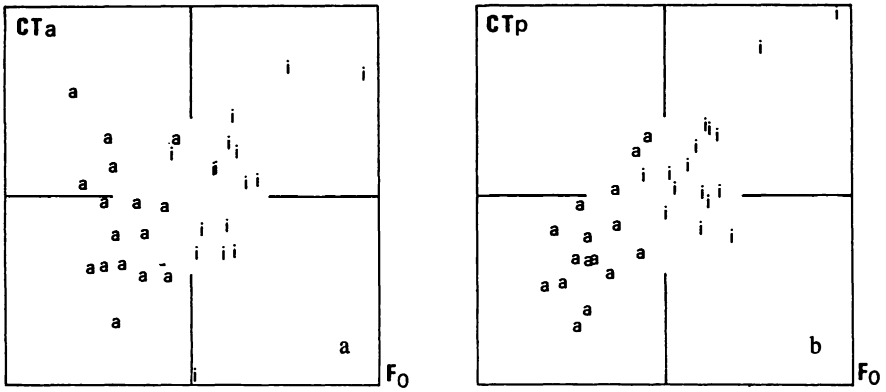


FIGURE 10-5
The joint histograms for the two cricothyroid muscles (CTa and CTp) compared with FO (abscissa) in /pip/ and /pɑp/.

CTs in both vowels together, CTp showed a relatively good correlation with FO, whereas CTa did not exhibit a significant relation. Observing each vowel type separately, there appears to be some qualitative difference between these two muscles. CTa showed a positive correlation with FO in /i/, but rather a very weak negative correlation in /ɑ/. CTp, in both vowels, showed relatively good correlations with FO. The differences observed in the two muscles are interesting from many points. Note that HG and SG in /ɑ/ showed an inverse effect of the articulatory activity on FO in the first experiment. It could be assumed that CTa functioned to reduce the effect of back-vowel articulation on FO, showing a higher activity in an extreme articulation. On the other hand, the linear correlation between CTp and FO across vowels contradicts the theories for the mechanism of the intrinsic pitch, rather implicating an synergistic control between the tongue and the larynx.

Figure 10-6 shows the data for the two laryngeal adductors, the lateral cricoarytenoid (LCA) and the inter-arytenoid (INT). LCA, in Figure 10-6(a), had a fairly good correlation with FO, however, this muscle showed a conspicuous difference across vowels. The correlation with FO is very high in /i/, whereas no correlation in /ɑ/. Furthermore, the variation in EMG activity is much greater in /i/ than in /ɑ/. Contrary to LCA, INT did not show a significant difference across vowels. In each vowel, INT had a weak positive correlation with FO, however, the average activity level is almost identical for both vowels. Although both muscles are the adductor of the larynx, there seems to be other functional differences. LCA is sometimes called as the medial compressor of the glottis. By this function, LCA can directly contribute to pitch raising by decreasing the length of vocal fold vibration. Figure 10-7 demonstrates this function by schematic drawings. When the force of the medial compression is larger, the vibratory length along the vocal folds becomes shorter, as noted by Broad (1973). Conversely, when the medial compression is weak, vocal fold vibration may involve the arytenoid cartilages, effectively increasing the length of vibration.

Finally, a preliminary analysis was performed on the variability of the vocal fold vibration in comparison with EMG and FO variations. The open quotient (OQ) was used as a parameter of glottal vibration, and it was calculated from the derivative of the electro-

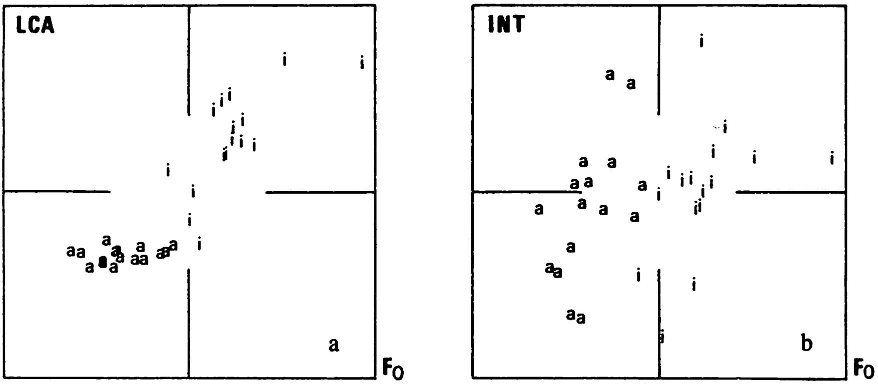


FIGURE 10-6
The joint histograms for the lateral cricoarytenoid (LCA) and the inter-arytenoid compared with FO (abscissa) in /pip/ and /pɑp/.

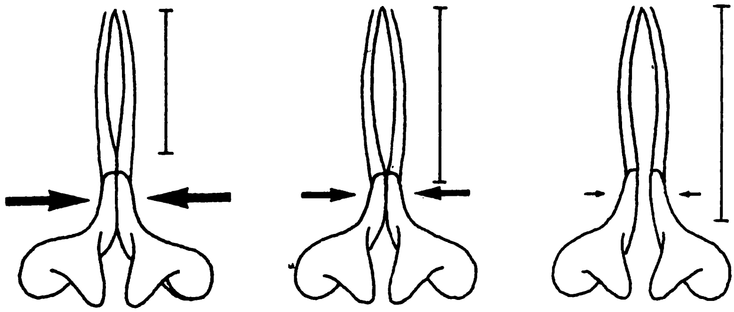


FIGURE 10-7
The effect of the medial compression on the length of vocal fold vibration. The three different conditions are shown as the constricted, normal and spread configurations of the glottis. The vertical lines indicate the effective length of vibration. (Modified from Broad, 1973).

glottographic (EGG) signals. EGG measures a change in the trans-glottal conductance during vibration, and it has been used as a simple, non-invasive technique to monitor the glottal vibration. The significance of EGG waveform has been examined (Baer, 1983), confirming a consistent correspondency between EGG waveform and the vocal fold vibration in terms of timing. In Figure 10-8, OQ variation was compared with LCA, INT and FO. The distribution of the OQ data for each vowel overlaps with each other, but the variation and the mean value of the data are larger in /i/ than in /ɑ/, showing that the data with high OQ appear more frequently in /i/. In Figure 10-8(a), LCA exhibited an inverse correlation with OQ in /i/, but almost no correlation in /ɑ/. The difference agrees with the relation between LCA and FO. In /i/, LCA may help to constrict the glottis as FO increases, overcoming the effect of CT activity to produce a slight abduction. INT showed a similar tendency in both vowels, in-

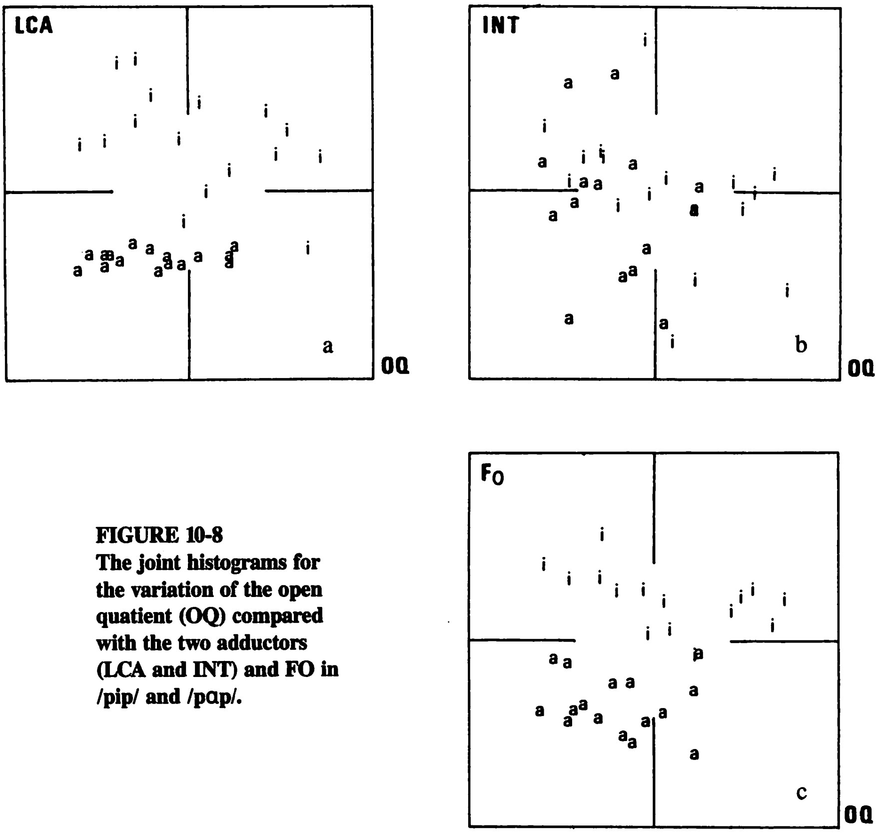


FIGURE 10-8
The joint histograms for the variation of the open quatient (OQ) compared with the two adductors (LCA and INT) and FO in /pip/ and /pɑp/.

versely correlated with OQ. It is reasonable that the adduction force by INT activity may enhance a glottal constriction. In general, the glottal vibration tends to be more sinusoidal in higher frequency. However, Figure 10-8(c) shows a complicated relation between FO and OQ. A different distribution of the data was observed across vowel types. In the tokens with relatively low OQ, a simple relation may be seen as an inverse correlation with FO in each vowel. It is suggested that the glottis tends to be constricted as FO increases. In the higher OQ range, however, a phonation in low FO appears to be prohibited.

Summary for the 2nd experiment.

The data collected in the 2nd experiment do not satisfy the complete set of measures required for the purpose of this study, since the other intrinsic muscles are left and subglottal driving factors are not considered. The audio waveform showed some other phonetic differences between two vowels such as the duration of the pre-voiced aspiration, longer in /i/ than in /ɑ/. A considerable difference in phonatory mechanism is suggested to be associated with the vowel-to-vowel variation. However, the results obtained here may be interpreted to summarize that a coordination of the intrinsic muscles varies with vowel quality, probably in anticipation of the articulatory effect on the larynx. This observation implies that the physical property of the vocal folds and the configuration of the glottis may vary with vowels, even though the variation in the coordination among the muscles results from a compensatory adjustment to the external forces to keep the vibratory mode maintained.

GENERAL REMARKS

It is well known that vowel-to-vowel variations are observed in the acoustic characteristics in vowel sounds. Other than their resonant characteristics, FO, intensity, and duration of vowels tend to vary uniquely to their quality. All these acoustic parameters during dynamic speech utterances exhibit evidence of correlation resulting from the physiological restraints of speech organs, realizing a natural impression of speech sounds. Since some of these variations reflect a consequence of larynx-vocal tract interactions, source parameters and vocal tract shapes provide a tendency of correlation.

The present study was attempted to reveal the physiological origins of the correlation between the fundamental and formant frequencies by analyzing the token-to-token variability of physiological and acoustic signals. In addition, the variability in muscular control of vocal fold vibration was examined with respect to the vowel difference. It is indicated that the relationship between fundamental and formant frequencies can be predicted by the forces on the hyoid bone resulting from the phonatory and articulatory activities. It is also suggested that the coordination of the intrinsic muscles seems to vary with vowels in anticipation of the articulatory influences.

REFERENCES

Alfonso, P.J., Honda, K., Baer, T. and Harris, K.S. (1982). "Multi-channel study of tongue EMG during vowel production," *J. Acoust. Soc. Am.* S1, 71, S54.

Broad, D.J. (1973). "Phonation," in Minifie, D.F., Hixon, T.J. and Williams, F. (Eds.) *Normal Aspects of Speech, Hearing, and Language*. New Jersey: Prentice Hall, 127-167.

Honda, K. (1983). "Relationship between pitch control and vowel articulation," in Bless, D.M. and Abbs, J.H. (Eds.) *Vocal Fold Physiology: Contemporary Research and Clinical Issues.* San Diego: College-Hill Press, 286-297.

Honda, K., Baer, T. and Alfonso, P.J. (1982). "Variability of tongue muscle activities and its implications," *J. Acoust. Soc. Am.*, S1, 71, S103

DISCUSSION

M. HIRANO: The data you presented are all very interesting in many aspects. But I must point out that some figures can be very misleading unless you describe other parameters in phonation. For example, the inverse relationship between Fo and OQ you presented is mysterious to everybody. But if the increase in Fo is associated with an increase in vocal intensity, the result will be quite understandable.

K. HONDA: Yes, I agree. Other factors such as subglottal pressure, air flow and the amplitude of vibration are necessary to interpret these data.

11 ▪ REFLEXOGENIC CONTRIBUTIONS TO VOCAL FOLD CONTROL SYSTEMS

Barry D. Wyke

At the last Vocal Fold Physiology Conference (in Madison, Wisconsin) I presented a general outline of our schema of the laryngeal neuromuscular control sequence of events in phonation. This Communication attempts a more detailed analysis of the reflexogenic contributions to that sequence, in the light of currently available animal exerimental and human observations.

In general, the reflex systems that modulate the phonatory activity of the vocal fold musculature may be divided into two categories - intrinsic and extrinsic.

Intrinsic Modulatory Reflexogenic Systems

These are operated by the co-ordinated activity of three sets of low-threshold mechanoreceptors embedded in the tissues (subglottic mucosa, muscles and joint capsules) of the larynx itself, whose afferent discharges are relayed through the laryngeal nerves to the brain stem motoneurone pools of the laryngeal muscles and upon which they exert reciprocally co-ordinated facilitatory and inhibitory influences.

1. Subglottic mucosal reflexogenic systems. A mixture of rapidly and slowly adapting corpuscular mechanoreceptors is distributed through the subglottic mucosa of the larynx, with population densities that vary in different regions of the mucosa. These sensitive receptors are stimulated by subglottic air pressures in excess of about 3 cm H_2O: so that during phonatory expiration (when, depending on the phonatory circumstances, subglottic pressures may range between 5 and 70 cm H_2O) their discharge frequency rises and falls in response to the surges of subglottic pressure that then occur. This fluctuating receptor activity exerts graduated facilitatory effects on the motoneurone pools of the vocal fold adductor muscular (with coincident inhibitory effects on posterior crico-arytenoid motoneurones) that contribute to the regulation of the vocal fold adduction that has been initially established during the preceeding phase of prephonatory tuning of the vocal folds.

2. Myotatic reflexogenic systems. Each of the laryngeal muscles contains stretch-sensitive mechano-receptors (consisting of a few muscle spindles, and many spiral nerve endings coiled around individual laryngeal muscle fibers) that are activated by stretching forces in excess of 3 gm applied to the individual muscles. Since during phonation such stretching forces are applied from below to the already adducted vocal folds by the advancing expiratory air stream, afferent discharges are provoked from the myotatic mechanoreceptors at frequencies related to the fluctuations in the stretching forces thus applied to the vocal folds during phonantory expiration. These discharges exert reciprocally co-ordinated facilitatory and inhibitory reflexogenic influences on the laryngeal adductor and abductor motoneurone pools that help to maintain and adjust the degree of vocal fold adduction during each phonatory episode that has been initiated during prephonatory tuning of the vocal folds.

3. Arthrokinetic reflexogenic systems. The capsules of the intercartilaginous joints of the larynx (and especially those of the crico-arytenoid and cricothyroid joints) are provided with an array of corpuscular mechanoreceptors - most of which are low threshold, rapidly adapting receptors. These receptors are stimulated by small increments of tension in the region of joint capsule in which they are embedded, the pattern and magnitude of which is deter-

mined by the direction and amplitude of the joint movements. Since rapidly fluctuating movements of the laryngeal joints occur during phonation, phasic changes in the discharge rate of the articular mechanoreceptors are thereby evoked during each phonatory episode. The resulting afferent discharges are then relayed to the motoneurone pools of the adductor and abductor musculature of the vocal folds, upon which they exert reciprocally co-ordinated facilitatory and inhibitory influences additional to those being exerted simultaneously from the mucosal and myotatic laryngeal mechanoreceptors.

The coordinated interaction of this triad of intrinsic laryngeal mechanoreceptor systems thus plays a major (and subconscious) part in the process of phonatory modulation of the activity of the vocal fold musculature once phonatory expiration is in train.

Extrinsic Modulatory Reflexogenic Systems

These are of two orders - pulmonary and cochlear.

1. Pulmonary reflexogenic systems. The lungs contain three varieties of receptor systems whose activity is transmitted via the vagus nerves to (inter alia) laryngeal motoneurone pools, upon which they exert varying effects. But their influence upon the phonatory activity of the vocal fold musculature appears to be negligible in normal circumstances, although their activity may be involved in the changes in voice quality that occur in patients with a variety of pulmonary (e.g. emphysema, chronic asthma, chronic bronchitis) and cardiovascular (e.g. pulmonary hypertension) disorders.

a. Stretch-sensitive nerve endings are distributed throughout the interalveolar connective tissues of the lungs, and are stimulated during lung inflation. But while their stimulation during normal inspiratory lung inflation appears to have no significant effect on the laryngeal musculature, their excessive stimulation results in reflex inhibition simultaneously of both vocal fold adductor and abductor muscles, thereby decreasing laryngeal resistance to the expiratory air flow.

b. J-receptors are unmyelinated nerve endings embedded in the walls of the pulmonary alveoli. Normally they are inactive; but they are stimulated by the development of pulmonary congestion and oedema. When so stimulated their afferent discharges produce simultaneous facilitation of vocal fold adductor and abductor motoneurones, thereby increasing laryngeal resistance to the expiratory air flow.

c. Irritant receptors are unmyelinated nerve endings ramifying through the epithelial lining of the respiratory tract from the trachea to the terminal bronchioles. Normally they are inactive: but they are stimulated by the presence of irritant gases (such as cigarette smoke) in the inspired air and by bronchoconstriction (as occurs in asthma, for example). Their stimulation results in the same effects on the laryngeal musculature as with the J-receptors; namely, simultaneous facilitation of laryngeal adductor and abductor motoneurone activity, with an increase in laryngeal resistance.

2. Cochlear reflexogenic systems. In addition to their other central projections, neurones in the cochlear nuclei give off projections to the laryngeal motoneurone pools within the lower brain stem that are activated by acoustic stimulation. This cochleo-laryngeal reflexogenic system has not yet been adequately studied, but its stimulation certainly results in facilitation of laryngeal adductor motoneurones. It may be, then, that this system contributes a reflexogenic (and thus subconscious) component to the process of acoustic automonitoring adjustment of vocal fold status that occurs during phonation.

REFERENCES

Adzaku, F.K. and Wyke, B.D. (1979). *Folia phoniat.* 31, 271-283.

Adzaku, F.K. and Wyke, B.D. (1982). *Folia phoniat.* 34, 57-64.

Brodal, A. (1981). *Neurological Anatomy in Relation to Clinical Medicine.* 3rd Ed. Oxford: Oxford Univ. Press.

Widdecombe, J.G. and Glogowska, M. (1973). *Acta Neurobiol. Exp.* 33,21-31.

Wyke, B.D. (Ed.) (1973). *Ventilatory and Phonatory Control Systems.* Oxford: Oxford Univ. Press.

Wyke, B.D. (1978). In. *Proc. VIIth Cong. Union of European Phoniatricians.* 10-26. Jyväskaliä, Finland.

Wyke, B.D. (1982). In *Tiefenstruktur der Musik. Festschrift fur Fritz Winckel.* C. Dahlhaus and M. Krause (Eds.) Universitätsbibliothek der Technischer Universität, Berlin, 129-156.

Wyke, B.D. and Kirchner, J.A. (1976). In *Scientific Foundations of Otolaryngology.* R. Hinchcliffe and D. Harrison (Eds.) Heinemann; London, 546-574.

DISCUSSION

G. FANT: What is the relative role of mechanical and chemical receptors in voice regulation?

B. WYKE: Assuming the question refers to the receptor systems in the subglottic laryngeal mucosa, expiratory stimulation of the mechanoreceptors therein (to degrees proportional to the subglottic pressure) results in reciprocally coordinated reflexogenic facilitation of motor unit activity in the vocal fold adductor musculature with inhibition of posterior crico-arytenoid activity. The subglottic mucosal chemoreceptor system has not yet been adequately studied, but at present it appears that its stimulation by CO_2 concentrations in excess of 5% in the subglottic air (or by cigarette smoke) results in reflexogenic adduction of the vocal folds. Thus both reflexogenic systems influence the length, tension and mass of the vocal folds during phonation by reinforcing adductor muscle activity.

R. SCHERER: Would you please elaborate on the effects of laryngeal surface anesthesia on declamatory speech and singing?

B. WYKE: Topical anesthesia of the subglottic laryngeal mucosa (by injection through the cricothryoid membrane) has little effect on the conversational speaking voice, but produces perturbations of the declamatory speaking voice (involving higher subglottic pressures) and renders singing (which requires still higher subglottal pressures) almost impossible, because of lack of accurate pitch and loudness control.

J. CURTIS: Is it the control of the pitch that seems to be destroyed so they cannot sing or carry a tune?

B. WYKE: That's right, they get a pitch wobble. And what is more, they cannot project their own pitch. In other words, what comes out in terms of pitch from moment to moment during singing is something over which they have no control.

G. ZIMMERMAN: The pathways involved in gate and mastication are modulable by further inputs. In voice production, what could modulate the pathways you're talking about during a production of pitch? What are the possibilities?

B. WYKE: You mean we can modulate what's traversing these reflexogenic pathways through the synaptic relays? Well, the main modulator as far as we're able to determine are some of the neurons in the brain stem reticular system and some in the cerebellum.

G. ZIMMERMAN: So lots of inputs could alter those loops that you're talking about.

B. WYKE: Well, I could go on explaining this ad infinitum, but the activity going on in the loops, as you call them, for example, represents a primary substratum of inputs.

G. ZIMMERMAN: Would you call them loops? What role do these reflex pathways play in the control of pitch? It seems to me, if I understand you correctly, you're indicating once vibration begins that it's controlled mainly by these reflexogenic sets.

B. WYKE: Mainly, but not exclusively.

G. ZIMMERMAN: Ok. So there are lots of other modulators.

B. WYKE: Sure. And what I'm suggesting is that these are sufficiently potent to have practical importance particularly in relation to disorders of phonation. But, I'd be the last person to say that these were the only modulators. But they're the ones I wanted to talk about this afternoon.

J. CURTIS: I must have misunderstood what you were saying. Let me see if I can articulate logic to that. In anesthetizing certain subglottal mucosa, the disrupting effect became greater as you went from simple conversational speech to what you call declamatory speech to singing.

B. WYKE: That's right.

J. CURTIS: It would seem to me that voluntary control of what one is doing is greater as you progress from conversational speech to declamatory speech to singing. You're having to control pitch, for example, voluntarily more in that progression. And yet, if this is an involuntary reflexive system, then it ought not to disrupt in that progression that you suggest. Am I wrong somewhere?

B. WYKE: I have to say, yes you are. Because what is often referred to as voluntary control of muscular activity, is really nothing of the sort. Any neurologist will tell you that there is no muscular activity and indeed no muscle in the body, the activity of which can be regulated exclusively by perceptually organized control which is what is meant by the shorthand phrase, voluntary control. All striated muscle activity, whether it's being operated from a prestored program which is very often the case in singing, or whether it is true voluntary control, that is, being controlled by continual perceptual attention to the process, is determined in its final output by the coincident input to the motor neuron pools of a series of reflexogenic systems. Now if you pull one or more of the reflexogenic systems out of the entire system, then what is left of the system cannot generate a normal performance. This is a very familiar experience in clinical neurology and neurosurgery, and it's as true of muscles in the larynx as it is anywhere else. So I hope you will forgive me and say, that when I said that you were wrong, I said so with the greatest respect.

II

Vocal Fold Kinematics and Techniques for Clinical and Pedagogical Observation

Section Editors

Kevin G. Munhall
Donald G. Childers

12 ▪ ULTRASONIC MEASUREMENT OF LARYNGEAL KINEMATICS

Kevin G. Munhall and David J. Ostry

Recent evidence from the study of laryngeal activity has been successful in demonstrating the overall character of abduction and adduction gestures in speech. This includes the identification of the muscular substrate (Hirose & Sawashima, 1981) as well as the demonstration of suprasegmental adjustments (e.g., Sawashima, 1970). In the present paper we have attempted to extend these observations on vocal fold dynamics by examining a large sample of laryngeal gestures obtained with pulsed ultrasound. Our data show that abduction and adduction gestures have different temporal characteristics, but are based on similar control mechanisms. Specifically, the data suggest that differences in the duration of the abduction and adduction movements are accompanied by systematic changes in articulator stiffness.

The findings to be presented are based on a kinematic analysis of the timing of vocal fold gestures. While kinematic information can be equivocal without accompanying physiological and biomechanical data (Müller, Abbs, & Kennedy, 1981), the kinematic phenomena may constrain the form that control can take and thus provide a basis for a detailed account of laryngeal function. The work described here involves manipulations of speech rate and stress for gestures in different consonant and vowel environments. The presentation of these findings will be preceded by a description of our ultrasound recording and data analysis procedures.

METHOD

The data were collected with a computerized pulsed ultrasound recording and analysis system. (See Hamlet (1981) for a review of ultrasound technology and its application to

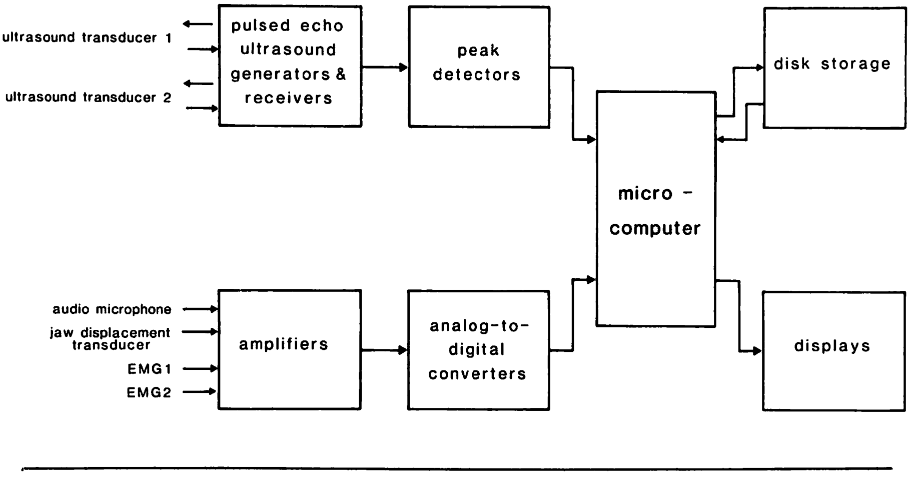


FIGURE 12-1
Schematic diagram of computerized pulsed-ultrasound system for the measurement of speech movements.

laryngeal research.) An earlier version of this system is described in Keller and Ostry (1983). The system consists of separate ultrasound units for laryngeal and tongue movements (Picker models 103 and 104 A-scan pulsed ultrasound units) and a 12 bit analog to digital channel for voice. There are several additional analog channels that can be used to monitor physiological signals and movements of the jaw. All input channels are presently sampled at a 1 kHz rate. A Cromemco CS2 microcomputer is used for data collection, display, and analysis (Figure 12-1).

The ultrasound signals are emitted from transducers placed beneath the chin for tongue movements and against the thyroid lamina for laryngeal movements. The ultrasound pulses thus travel through soft tissue to the articulator surface. Ultrasound has the property that part of its energy is reflected at changes in acoustic impedance. In diagnostic and experimental work with ultrasound, reflections occur at changes in tissue density with a large percentage of the acoustic energy being reflected at tissue-air boundaries. The large amplitude reflection from the articulator surface allows us to track laryngeal activity during speech. This is accomplished by timing the interval between the emission of the source pulse and the reception of the large amplitude echo (Figure 12-2). These intervals, timed with a 2 MHz clock, are sampled once a millisecond and transferred to the microprocessor.

The time between the emission of the ultrasound source pulse and the reception of the large amplitude reflection from the free margin of the vocal fold is converted to a distance estimate by assuming an average speed of ultrasound in soft tissue of 1540 m/sec (Goss, Johnston, & Dunn, 1978). The speed of ultrasound in soft tissue is not altered significantly

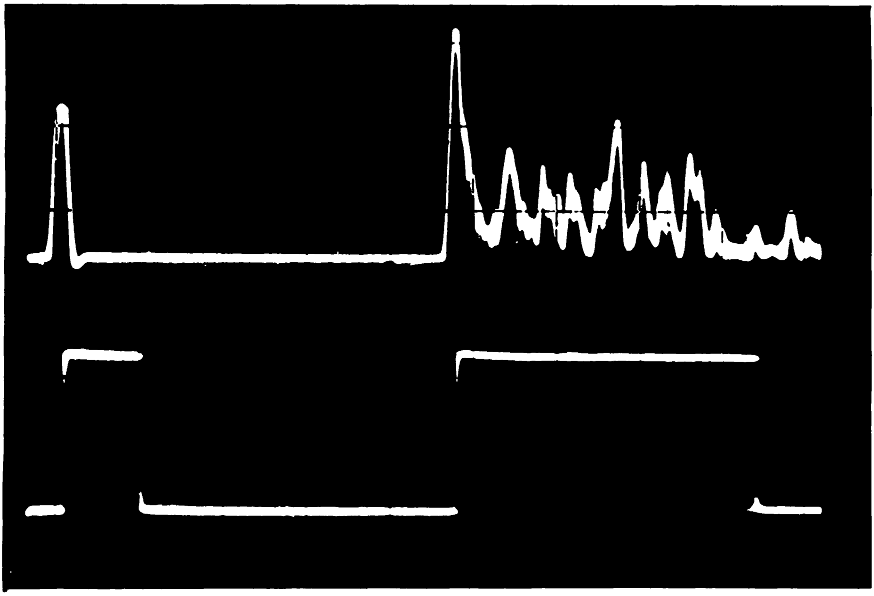


FIGURE 12-2
Oscilloscope record of an emitted ultrasound pulse and the reflections from the tongue dorsum (upper trace). Corresponding signals from the peak detection circuitry are shown in the lower trace. The distance of the tongue dorsum from the crystal of the ultrasound transducer is indicated by the interval between the leading edges of the signals in the lower record.

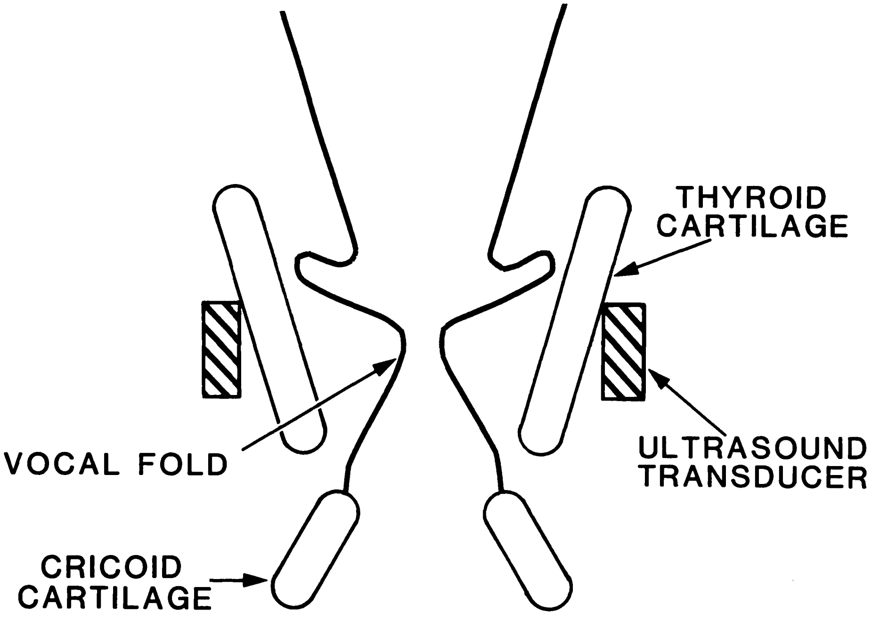


FIGURE 12-3
Schematic drawing of the position of the ultrasound transducers in relation to the anatomical structures of the larynx.

by the level of muscular contraction or by the orientation of the ultrasound beam relative to the muscle. In the case of the monitoring of the vocal folds, passage of the ultrasound signal through the hyaline cartilage of the thyroid lamina results in a small but constant error in the estimated distance from the transducer to the margin of the vocal fold.

The transducer placement for laryngeal recording is determined using a procedure reported previously by Hamlet (1981), Holmer and Rundqvist (1975), and Kaneko, Uchida, Suzuki, Komatsu, Kanesaka, Kobayashi, and Naito (1981). The subject is seated in front of a stand which holds a pair of matched transducers. A transducer is placed on each side of the thyroid lamina below the thyroid notch and the ultrasound unit's through-transmission mode is used to locate the folds (Figure 12-3). Maximum through-transmission at this level will occur when the folds are in contact. Fold location is identified when a discontinuous signal is observed during a sustained vowel and no signal is observed during non-contact laryngeal maneuvers such as respiration. Next, the amplitude of the through-transmitted signal is maximized during repetitive syllable production at the pitch and amplitude required for testing. During the actual test conditions the system is switched to a pulsed-echo mode and unilateral measures with a single transducer are taken of the distance from the transducer to the fold's surface.

The height of the laryngeal cartilages and hence the height of the folds themselves is known to vary during natural speech. These height changes are primarily associated with differences in pitch. While such variation in height could have consequences for an ultrasound measurement system, as it does for measurements with transillumination and fibre

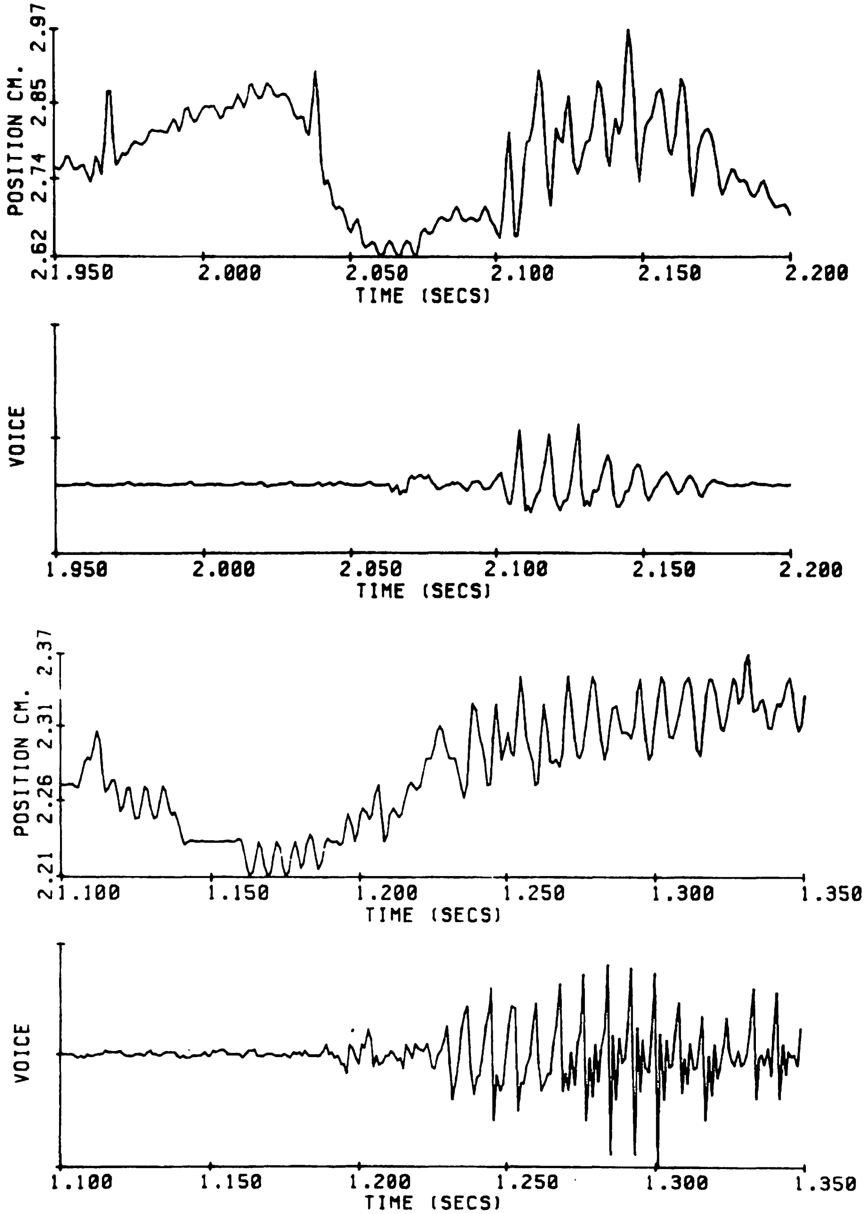


FIGURE 12-4

Raw ultrasound records obtained at a 1 kHz sampling rate using two different ultrasound transducers. The upper panels show a recording obtained using the 1.8 MHz elliptical transducer described in Hamlet (1980). The lower panels show a recording obtained with a 2 MHz circular transducer. Acoustic records are shown below each of the ultrasound records.

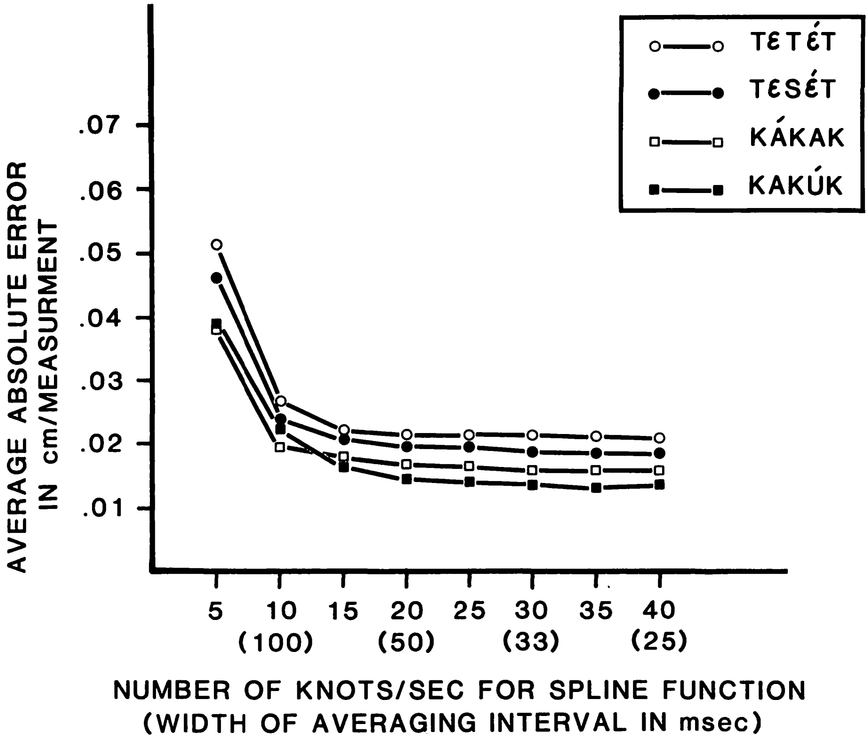


FIGURE 12-5
Average absolute error per measurement as a function of the number of knots used to calculate the spline function. Corresponding widths of averaging intervals in msec are given in parentheses.

optics, we feel this does not affect the interpretation of the present data. Specifically, subjects produce repetitive monotone trials of stimuli that eliminate the major intonation contour height changes. While height changes specific to stop consonants have been observed (e.g., differences between voiced and voiceless stops; Riordan, 1980, Perkell, 1969) these height changes tend to occur after the release of an intervocalic consonant and therefore do not affect the gestures being measured. Further, the magnitude of these height changes, for example, in Riordan's data, is no more than a millimetre or two and well within the beam width of our transducer.

One should note that the actual measurements of abduction and adduction gestures depend on the anterior/posterior placement of the transducer on the thyroid cartilage. Our current placement criteria typically result in mid to posterior transducer positions. The requirement of a discontinuous signal during through-transmission eliminates the possibility of far anterior placements as continuous transmission of ultrasound through the anterior wall of the larynx would presumably be observed. Further, our measurements are from the membraneous portion of the glottis as opposed to the cartilaginous portion. Placement along the membrane is indicated by the presence of a discontinuous signal in through-transmission and by periodic pulsing in the raw ultrasound records (Figure 12-4). As a control for anterior/posterior placement variation, a particular transducer position on the lamina is maintained for at least one measurement of each condition in a given experiment.

We have tested a variety of ultrasound transducers for laryngeal measurement. These have included the transducers used by Hamlet (1980). The results reported below are based primarily on data collected with a 2 MHz unfocused circular transducer with a crystal diameter of 13 mm. The transducers used previously by Hamlet have a 1.8 MHz frequency and an elliptical beam profile (see Hamlet, 1980 for a detailed description). Raw records obtained with both transducers are displayed in Figure 12-4.

The data analysis involves the fitting of natural cubic spline functions to the raw data (Johnson & Riess, 1977). For purposes of analysis the data are divided into intervals of equal width and the average values in the intervals serve as knots for a spline fitting program. Cubic splines are piecewise polynomial functions that approximate a set of data points. These functions were chosen for the present application as their piecewise form makes no *a priori* assumptions about the overall shape of the function and enables the approximation to follow trends in the data closely. Further, the functions are differentiable numerically, thus allowing the user to obtain values for velocity and acceleration as well as position.

The averaging intervals for the spline approximation were selected on the basis of a numerical procedure in which the average absolute difference between the spline function and the raw data was calculated for a variety of averaging interval widths. Figure 12-5 shows the average absolute error between the raw laryngeal measurements and the fitted spline function for interval widths ranging from 25 msec to 165 msec. It can be seen that for a variety of different speech gestures the error due to the approximating solution reaches an essentially constant value for interval widths of 45 msec or less. This averaging interval width was used for all the data presented in this paper. Examples of spline approximated data are shown in Figure 12-6.

In order to characterize the abduction and adduction movements of the vocal folds, the spline function must sacrifice information about higher frequency deformations associated with glottal vibration. This is a consequence of the averaging operation that provides knots to the spline fitting program. It should be noted that the 1 kHz sampling rate used in our system is sufficient to demonstrate glottal vibration (Figure 12-4) but it seems inadequate for the detailed measurement of individual vibration kinematics. Researchers studying such phenomena have generally used higher pulse rates (e.g., Kaneko *et al*, 1981).

The data were collected by recording a number of 3.5 sec trials of vocal fold movements. For a given trial or block of trials the ultrasound transducer placement was verified as described above and the subjects produced the speech tokens repetitively. Experimental conditions were randomly ordered across trials.

The stimuli were of the form CVCVC with the primary data being measurements on the intervocalic consonant. The stimuli included /tɛtɛt/ and /tɛsɛt/ with stress on either the first or second syllable and /kakak/ with stress on either the first of second syllable. Rate was also manipulated.

RESULTS

The data were partitioned with respect to the following variables: duration of the movement, displacement, maximum velocity, and the time from the initiation of movement to the point of maximum velocity. In Figure 12-7, the top panel shows the distance from the crystal of the ultrasound transducer to the vocal fold surface. The folds are fully adducted at the top

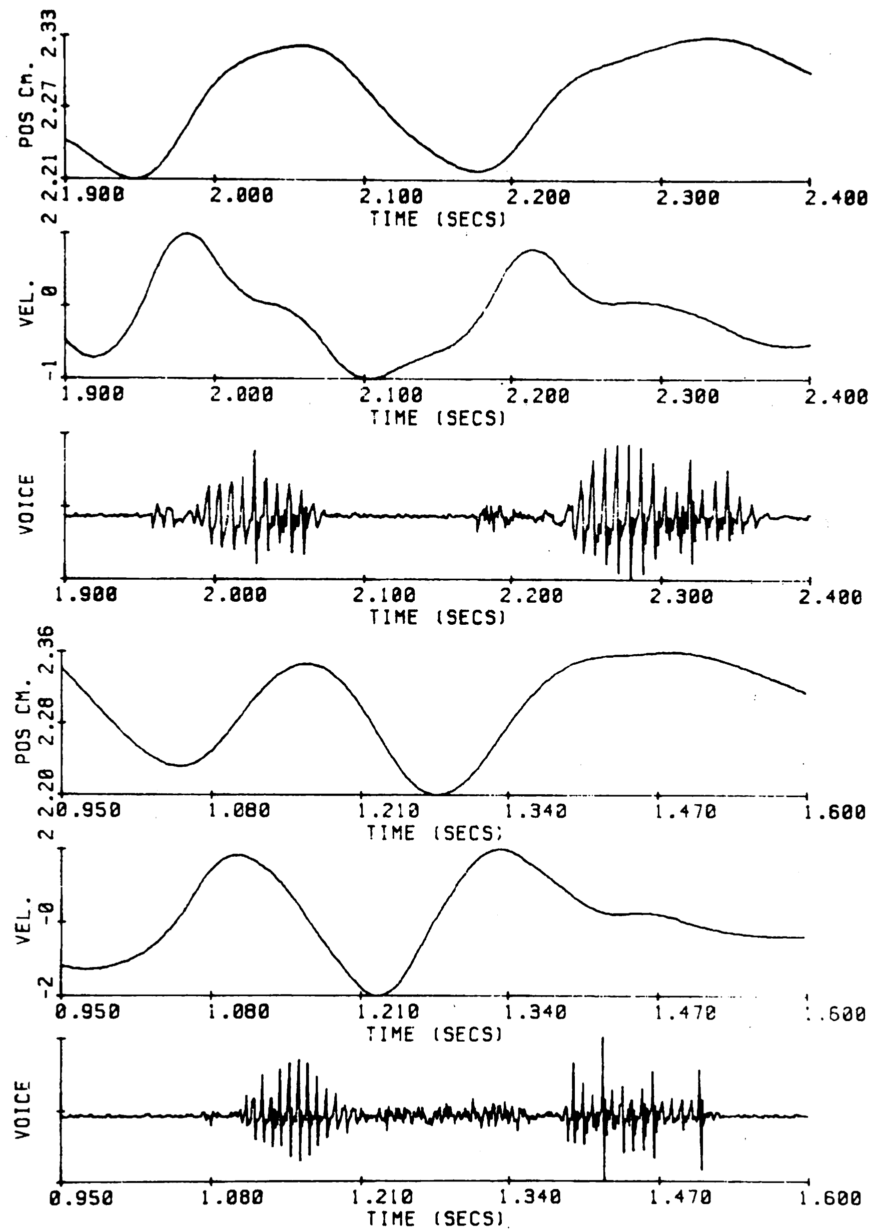


FIGURE 12-6

Ultrasound records of single repetitions of /tɛtɛt/ (upper panels) and /tɛsɛt/ (lower panels), with corresponding velocity and raw acoustic records. In both panels the folds are adducted at the top of the record and abducted at the bottom. The values shown on the ordinate correspond to the distance in cm from the ultrasound crystal to the free margin of the vocal fold. Adduction velocities are shown as positive values; abduction velocities are negative values.

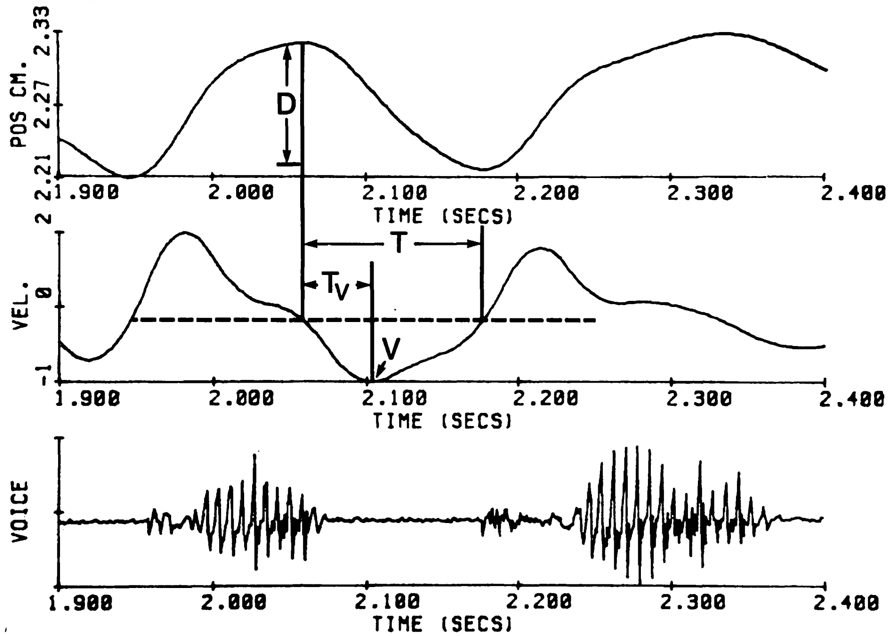


FIGURE 12-7

Ultrasound displacement record (top panel), instantaneous velocity (middle panel), and accompanying acoustic signal (lower panel) showing standard measurement variables. D is the displacement or extent of the movement in cm. T is the duration of the movement in msec. V is the maximum velocity in cm/sec. Tv is the duration of the interval between the initiation of movement and the point of maximum velocity in msec.

of the record and abducted at the bottom. The middle panel displays the instantaneous velocity of adduction and abduction gestures, while the bottom panel shows the accompanying acoustic signal.

To date we have observed good correspondence between our data and previously reported findings using transillumination and fibre optics. For example, as reported by Löfqvist and Yoshioka (1981) the time from vowel offset to peak glottal opening varies as a function of oral closure or approximation duration. Figure 12-8 shows, for a single subject, a scattergram of raw data values for /tεtεt/ with stress and rate manipulated. As can be seen the closure duration varies with the interval from vowel offset to peak glottal opening. Further, average glottal abduction/adduction displacements for fricatives and stops differ (Table 12-1). The glottal displacements are significantly larger for fricatives than for stops. This parallels results reported by a number of researchers (e.g., Yoshioka, Löfqvist, & Hirose, 1979).

In our current work we have been studying the form of the trajectory of single gestures. The studies have focused on finding patterns among kinematic variables that are specific to manipulations of speech rate, stress, vowel, or consonant, or that are independent of such divisions. One of the kinematic relationships we have examined is the relationship between the maximum velocity of a movement and the amplitude or extent of that movement. This relationship has been shown to vary depending on the manner in which movements are controlled (Kelso, Holt, Rubin, & Kugler, 1981; Nelson. 1983).

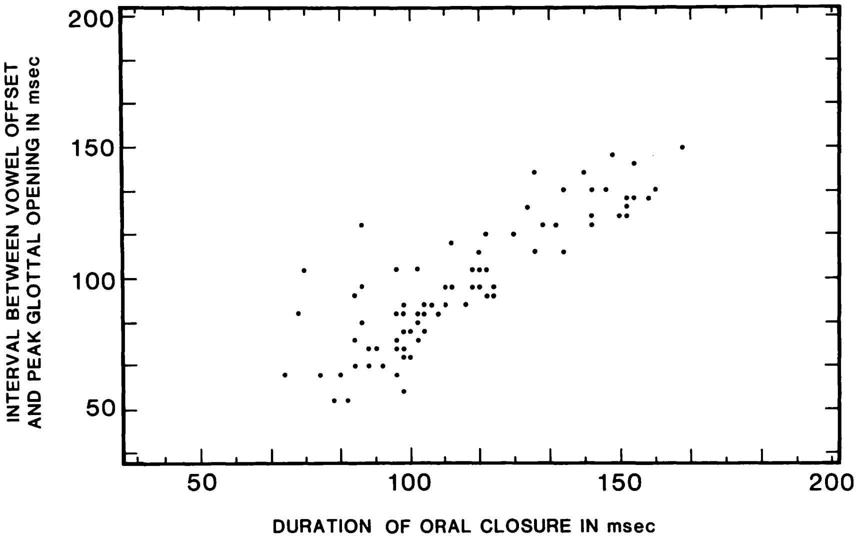


FIGURE 12-8
The relation between the duration of oral closure and the interval between vowel offset and peak glottal opening for the intervocalic consonant in a single subject's productions of the nonsense syllable /tɛtɛt/. Each point represents the value for a single production.

TABLE 12-1
Average abduction/adduction displacement in mm during intervocalic stop and fricative production in a /tɛCɛt/ context.

	/S/	/T/
Subject KM	1.2	1.0
Subject DO	1.7	1.4

The relationship between maximum velocity and movement extent has been examined previously in studies of the physiology of limb movement. Cooke (1980, 1982) has demonstrated that in limb movements changes in the maximum velocity/displacement relationship are a kinematic concomitant of changes in the overall stiffness of the limb. Specifically, the slope of the regression line relating these two variables (with maximum velocity presented as a function of displacement) has been shown to vary with limb stiffness. Relatively flat regression lines are observed for movements about joints in which the overall stiffness of the limb is low whereas steep regression lines correspond to greater stiffness. Particularly compelling evidence for this interpretation of the kinematic profiles comes from the study of patients with decreased muscle tone as a result of cerebellar dysfunction. As predicted these patients exhibit significantly flatter maximum velocity/displacement relationships than normal (Cooke, 1980).

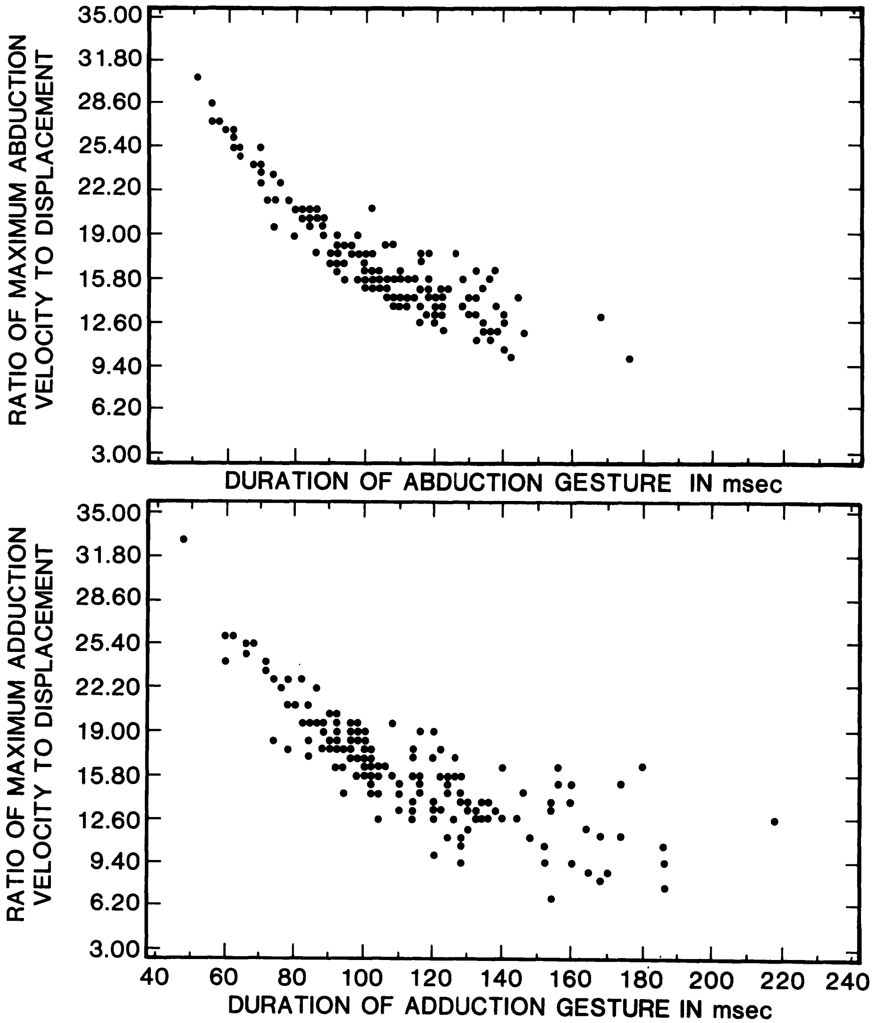


FIGURE 12-9

Scattergram showing the ratio of maximum velocity to displacement as a function of the duration of abduction (top panel) and adduction (lower panel) gestures for a single subject. The data represent individual productions for the intervocalic consonant in /tɛtɛt/and /tɛsɛt/ with speech rate and stressed vowel manipulated.

We have examined the relationship between maximum velocity and displacement for both abduction and adduction gestures in a number of phonetic and suprasegmental conditions. In all cases the correlations within conditions were highly reliable with an average r^2 typically above .65.

To examine changes in the slope of the maximum velocity/displacement relationship, we obtained point estimates of the slope for a given gesture (i.e. the ratio of maximum velocity

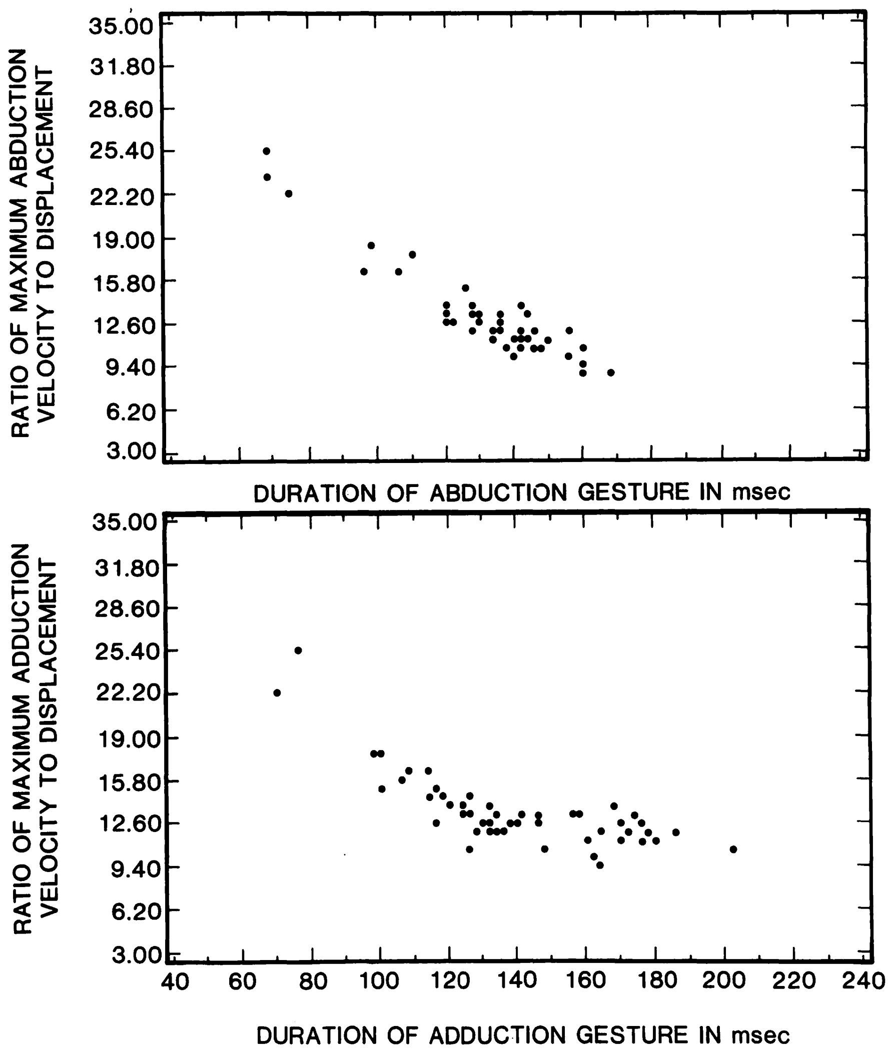


FIGURE 12-10

Scattergram showing the ratio of maximum velocity to displacement as a function of the duration of abduction (top panel) and adduction (lower panel) for a single subject. The data represent individual productions of the intervocalic consonant in /kakak/ with the stressed vowel manipulated.

to displacement). These estimates were examined relative to the duration of the corresponding movement. As can be seen in Figure 12-9, a scattergram of individual observations for a single subject, the ratio of maximum velocity to displacement varies systematically with the duration of the gesture. As gesture duration increases the ratio decreases. This is true for both abduction and adduction gestures. Figure 12-10 shows a similar relationship for a different speech token. As the maximum velocity-displacement slope has been shown to vary with stiffness, these findings suggest the possibility that decreases in the duration of glottal gestures are accompanied by increases in articulator stiffness. This is consistent with evidence that the dynamic stiffness of the limb increases as the movement decreases in duration (Feldman, 1980a; Cooke, 1982).

Figures 12-9 and 12-10 display data for the intervocalic consonant in CVCVC sequences in which speech rate, stress, and consonant are varied. The data appear to form single functions in spite of this variation. An examination of the average maximum velocity/displacement ratios was carried out to determine the effects of consonant, stress, and rate on this ratio. As can be seen in Table 12-2, rate, stress, and consonant manipulations produce average values in different regions of the function. The observed ratios were higher for stops, unstressed, and fast conditions.

TABLE 12-2

Average maximum velocity/displacement ratios by speech rate, stress, and consonant for the intervocalic consonant in a /tɛCɛt/ context.

Subject DO		
	/T/	/S/
Abduction	16.32	15.29
Adduction	16.31	14.99
	Fast	Slow
Abduction	16.65	15.12
Adduction	16.24	14.82
	Unstressed	Stressed
Abduction	16.43	15.53
Adduction	17.88	13.34
Subject KM		
	/T/	/S/
Abduction	17.95	16.30
Adduction	16.94	15.78
	Fast	Slow
Abduction	17.83	16.60
Adduction	16.62	16.21
	Unstressed	Stressed
Abduction	18.18	16.04
Adduction	16.97	15.75

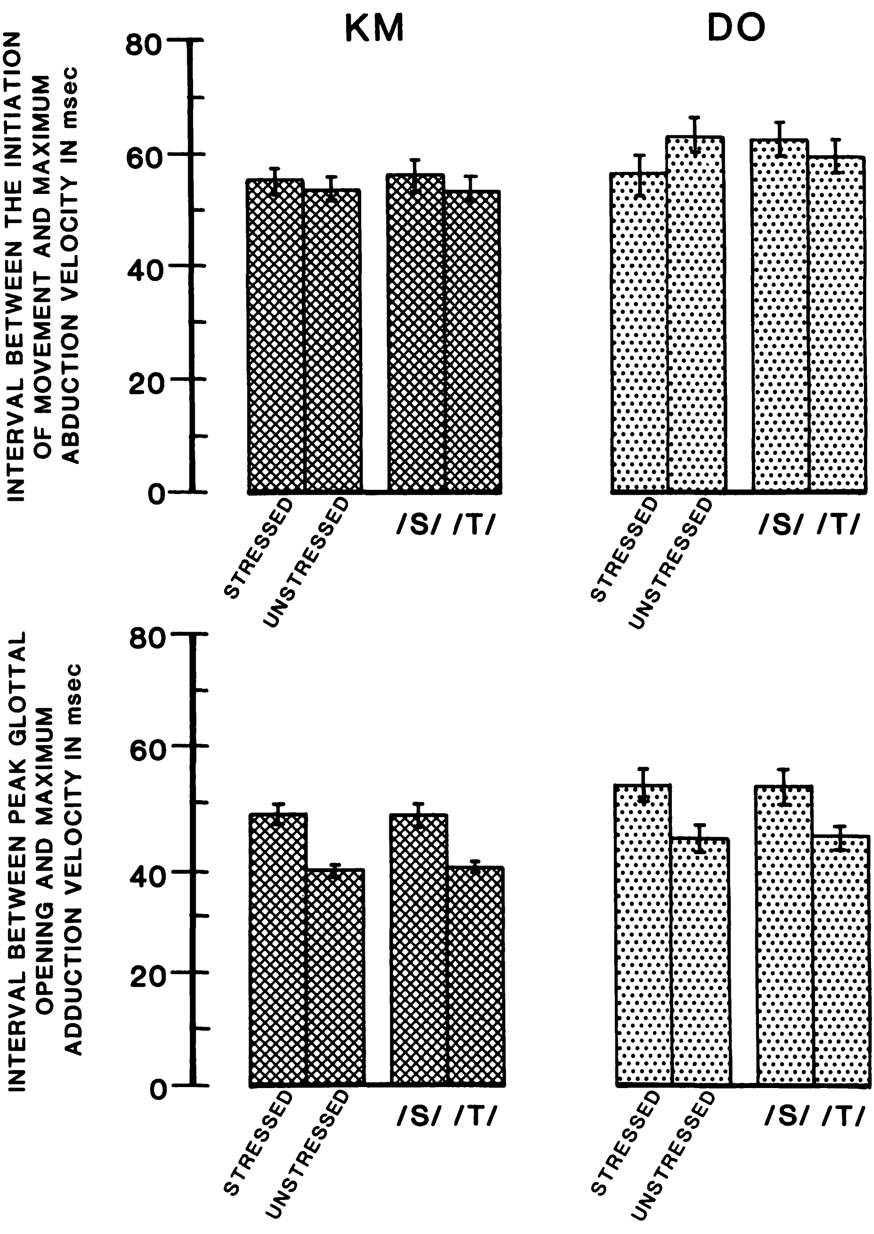


FIGURE 12-11

Duration of the interval from the initiation of movement to maximum abduction velocity (upper panel) and duration of the interval from the initiation of movement to maximum adduction velocity (lower panel) for the intervocalic consonant in /tɛtɛt/ and /tɛsɛt/ with either the first or second vowel stressed.

In a different analysis of the data we examined the relation between abduction and adduction. While the displacement or extent of abduction and adduction gestures is necessarily related except for cases in which complete laryngeal closure is not achieved, our data have indicated that the time courses of the two gestures are somewhat different.

In the case of abduction movements the time from the initiation of movement to the point of peak velocity was not found to vary significantly with rate or stress manipulations. This is consistent with Löfqvist and Yoshioka (1981) who found that the average interval from implosion to peak velocity did not vary across different obstruents for speakers of Swedish and Japanese. In an examination of adduction gestures, on the other hand, this interval was found to reliably differ with stress and consonant (Figure 12-11).

DISCUSSION

To date we are encouraged by the use of our system for laryngeal research. The system has the advantage of the automated collection of readily interpretable data. Our data so far have provided a replication of findings obtained with other technologies, which points to the validity of the measurement technique and the independence of the findings from the particular procedures used.

The data presented here demonstrate a number of characteristics of laryngeal movements. Within individual gestures the ratio of maximum velocity to displacement increases systematically with decreases in the duration of the movement. While these relationships are observed for both abduction and adduction gestures, the durational characteristics of the two movements differ.

The systematic increase in the maximum velocity/displacement ratio with decreases in gesture duration occurs with changes in speech rate, stress, and consonant. To the extent that this ratio is an adequate index of articulator stiffness, the data indicate that stiffness differs for various speech manipulations (Ostry, Keller, & Parush, 1983). The presence of a single function which accounts for a high proportion of the variance suggests that durational changes, associated with differences in speech rate, stress, and vowel, may be all produced by increasing the overall stiffness of the glottal articulators (cf. Tuller, Harris, & Kelso, 1982). Feldman (1980a,b) has shown that articulator stiffness can be controlled by changing the zero-lengths, that is, by regulating the muscle lengths at which the tonic activity of motor units in agonist and antagonist muscles begins. The regulation of zero-lengths would enable overall changes in articulator stiffness without directly controlling the tension-extension characteristics of the individual abducting and adducting muscles.

A function similar to the maximum velocity/displacement functions shown in Figures 12-9 and 12-10 is predicted by a mass spring system with negligible damping (Nelson, 1983). Under these conditions, the maximum velocity/displacement ratio varies as a function of $\pi/2T$, where T is the duration of the movement. Changes in this function accompany changes in stiffness in the mass spring model. As can be seen in Figures 12-12 and 12-13 the predicted functions provide a good first approximation to the laryngeal data. (See also Kelso & Holt, 1980.)

In contrast to the similar changes observed for both adduction and abduction movements in the point estimates of articulator stiffness, the opening and closing movements themselves showed different kinematic patterns. Specifically, for abduction, the time from the initiation of the movement to the point of maximum velocity showed no systematic pattern over differences in stress, speech rate, or consonant. On the other hand, the time to peak velocity for adduction gestures varied with consonant and stress but not rate. Time to peak velocity

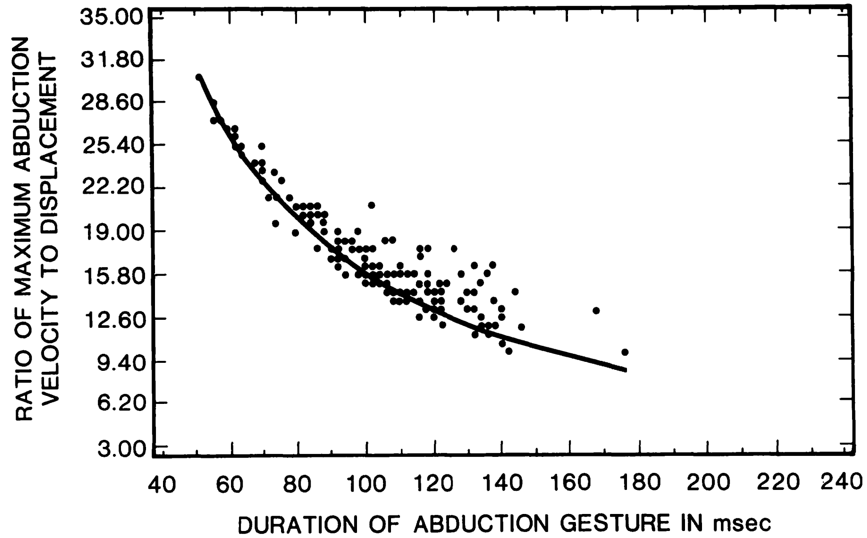


FIGURE 12-12

Scattergram showing the ratio of maximum velocity to displacement as a function of the duration of abduction for a single subject and the function predicted by a mass spring system with negligible damping. The data represent individual productions of the intervocalic consonant in /tɛtɛt/ and /tɛsɛt/ with speech rate and stressed vowel manipulated.

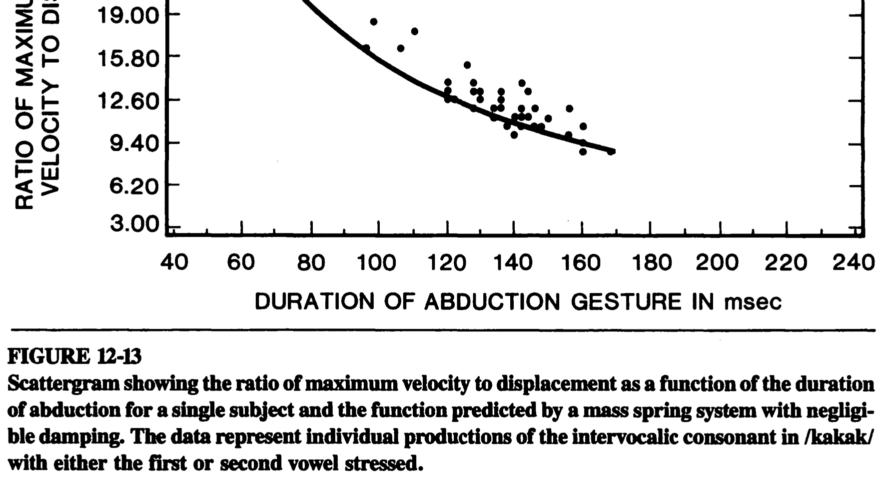


FIGURE 12-13

Scattergram showing the ratio of maximum velocity to displacement as a function of the duration of abduction for a single subject and the function predicted by a mass spring system with negligible damping. The data represent individual productions of the intervocalic consonant in /kakak/ with either the first or second vowel stressed.

was greater for stressed than for unstressed vowels and for fricatives than for stops.

The differences in the kinematic patterns of abduction and adduction perhaps result from differing functional roles of the opening and closing gestures. The timing of the abduction gesture appears to be related to the oral closure timing. The abduction movement must ensure that the glottal state can produce the appropriate aerodynamic conditions for consonant production. The adduction movement on the other hand is associated with bringing the laryngeal apparatus to a vowel production posture. The timing demands of these two roles are presumably different.

In summary, the principle finding with respect to the control of laryngeal gestures is that increases in gesture duration for both abduction and adduction are accompanied by decreases in the estimated stiffness of the glottal articulators. Durational changes in speech rate, stress, and consonant all seem to be accommodated by this single function.

ACKNOWLEDGEMENTS

This research has been supported by grants to the second author from the Natural Sciences and Engineering Research Council of Canada and the FCAC program of the Quebec Department of Education. The authors would like to express appreciation to Sandra Hamlet for the generous loan of a number of ultrasound transducers. Avraham Parush and George Hargrave made helpful comments on earlier versions of this work.

REFERENCES

Cooke, J.D. (1980). "The organization of simple, skilled movements," in G. Stelmach and J. Requin (Eds.) *Tutorials in Motor Behavior*. Amsterdam: North Holland.

Cooke, J.D. (1982). "Position-velocity-torque relations during human arm movement," *Soc. for Neuroscience Abstracts* 8, 731.

Feldman, A.G. (1980a). "Superposition of motor programs - 1. Rhythmic forearm movements in man," *Neuroscience* 5, 81-90.

Feldman, A.G. (1980b). "Superposition of motor programs - 2. Rapid forearm flexion in man," *Neuroscience* 5, 91-95.

Goss, S.A., Johnston, R.L. and Dunn, F. (1978). "Comprehensive compilation of empirical ultrasonic properties of mammalian tissues," *J. Acoust. Soc. Am.* 64, 423-457.

Hamlet, S.L. (1980). "Ultrasonic measurement of larynx height and vocal fold vibratory pattern," *J. Acoust. Soc. Am.* 68, 121-126.

Hamlet, S.L. (1981). "Ultrasound assessment of phonatory function," Proc. of the Conf. on the Assessment of Vocal Pathol., *ASHA Reports* 11, 128-140.

Hirose, H. and Sawashima, M. (1981). "Functions of the laryngeal muscles in speech," in K. N. Stevens and M. Hirano (Eds.) *Vocal Fold Physiol.* Tokyo: Univ. of Tokyo Press.

Holmer, N.G. and Rundqvist, H.E. (1975). "Ultrasonic registration of the fundamental frequency of a voice during normal speech," *J. Acoust. Soc. Am.* 58, 1073-1077.

Johnson, L.W. and Riess, R.D. (1977). *Numerical Analysis.* Reading, Mass.: Addison - Wesley, Inc.

Kaneko, T., Uchida, K., Suzuki, H., Komatsu, K., Kanesaka, N., Kobayashi, and Naito, J. (1981). "Ultrasonic observations of vocal fold vibration," in K.N. Stevens and M. Hirano (Eds.) *Vocal Fold Physiol.* Tokyo: Univ. of Tokyo Press.

Keller, E. and Ostry, D.J. (1983). "Computerized measurement of tongue dorsum movements with pulsed echo ultrasound," *J. Acoust. Soc. Am.* 73, 1309-1315.

Kelso, J.A.S. and Holt, K.G. (1980). "Exploring a vibratory systems account of human movement production," *J. of Neurophysiol.* 43, 1183-1196.

Kelso, J.A.S., Holt, K.G., Rubin, P. and Kugler, P.N. (1981). "Patterns of interlimb movement emerge from the properties of non-linear limit cycle oscillatory processes: theory and data," *J. Motor Behav.* 13, 226-261.

Löfqvist, A. and Yoshioka, H. (1981). "Interarticulator programming in obstruent production," *Phonetica* 38, 21-34.

Müller, E.M., Abbs, J.H. and Kennedy, J.G. (1981). "Some systems physiology considerations for vocal control," in K.N. Stevens and M. Hirano (Eds.) *Vocal Fold Physiol.* Tokyo: Univ. of Tokyo Press.

Nelson, W.L. (1983). "Physical principles for economies of skilled movements," *Biol. Cybernetics* 46, 135-147.

Ostry, D.J., Keller, E. and Parush, A. (1983). "Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movement in speech," *J. Exp. Psychol.: Human Perception and Performance*, 9, 622-636.

Perkell, J.S. (1969). *Physiology of Speech Production*. Cambridge Mass.: MIT Press.

Riordan, C.J. (1980). "Larynx height during English stop consonants," *J. of Phonetics* 8, 353-360.

Sawashima, M. (1970). "Glottal adjustments for English obstruents," *Haskins Labs Stat. Rep. on Speech Res.* SR-21/22, 187-200.

Tuller, B., Harris, K.S. and Kelso, J.A.S. (1982). "Stress and rate: differential transformations of articuation," *J. Acoust. Soc. Am.* 71, 1534-1543.

Yoshioka, H., Löfqvist, A. and Hirose, H. (1979). "Laryngeal adjustments in the production of consonant clusters and geminates in American English," *Haskins Labs Stat. Rep. on Speech Res.* SR-59/60, 127-151.

DISCUSSION

H. HIROSE: I am particularly interested in your data showing the difference between abduction and adduction in terms of the time from the onset of displacement to the point of maximum velocity with respect to differences in stress or manner of articulation. Would you give us an interpretation for this finding?

K. MUNHALL: I'd like to be cautious about this for the moment. If I look at the distributions of the raw data for this interval for abduction and adduction, they are very similar. The abduction data just do not show any systematic variation. I presume that the differences we've observed are based on the different roles the two gestures play in sound production or perhaps on different roles in interarticulator coordination. However, we can't tell without further research.

M. ROTHENBERG: Two questions. Do you have any estimate or measurement of the rate limitation imposed by your processing scheme?

K. MUNHALL: The processing system is quite flexible and the temporal resolution can be set at any value that is appropriate for the measured behavior. In the present data the bandwidth is approximately 23 Hz.

M. ROTHENBERG: Now as far as the explanation of the abduction/adduction difference goes, I've noticed the same thing in airflow data, that the rate in adduction is faster than the abduction. I've given that an interpretation in my mind. In producing a stop you have a maximally fast movement. You have some dynamic limitations for the opening and closing. If you pay attention to timing the adduction, then you lose some control over the abduction and vice versa. So it's a matter of where you want to put your money. In terms of the aerodynamics of producing the stop, it's the adduction that's more important. That determines the duration of aspiration, which is more critical, so that you don't lose too much air. The abduction is not as critical.

K. HARRIS: By this interpretation you would expect the stress difference for all languages, but you would expect the absolute values to be different depending upon the aspiration characteristics.

M. ROTHENBERG: Yes, I think you would expect to see some differences depending on aspiration characteristics.

13 ▪ CORRESPONDENCE OF ULTRASONIC AND STROBOSCOPIC VISUALIZATION OF VOCAL FOLDS

James A. Zagzebski and Diane M. Bless

In previous presentations (Bless, 1981; Zagzebski, 1982) it was shown that ultrasound scanning of the larynx could provide detail on the vibrating vocal folds during speech production. Whereas the spatial resolution and the imaged field cannot compare with that of other procedures for studying the larynx, such as stroboscopic video recording and high speed cinematography, ultrasound does offer some unique advantages in this application. For example, studies can be done using external probes, permitting movement patterns to be studied without the tongue or epiglottis obstructing the view. Ultrasound also lends itself quite naturally for carrying out studies related to precise timing of the movement patterns of articulators or different regions of the same articulator.

The present phase of our research is directed towards exploring the capabilities and limitations of ultrasound techniques for studying the larynx. Questions that are being addressed include: What structures can be seen reliably with ultrasound scanning? How can these structures best be delineated with ultrasound? And, what studies can be carried out reliably with present instrumentation? In this presentation we will describe studies showing simultaneous viewing of the larynx with B-mode ultrasound and stroboscopic video laryngoscopy.

ULTRASOUND SCANNER

The ultrasound system we are currently using for imaging the larynx is a rapidly scanning, "mechanical sector scanner" pulse echo instrument. (Biodynamics, Biosound, Indianapolis, Indiana.) A 2.2 cm by 1.5 cm elliptically shaped piezoelectric transducer array is the sound source and the echo detector. A pulse of sound from the arrray is transmitted through a liquid path in the transducer assembly and deflected by an acoustic mirror into the larynx. Interfaces in the sound beam path, characterized by discontinuities in the tissue density and/or compressibility, reflect and scatter the incident beam. Part of the reflected energy travels back to the transducer, where it is detected as an echo signal. Images are produced using a B-mode echo display. Here echo signals are converted to brightness modulated dots on a display screen and the position of each dot corresponds to the corresponding reflector position.

The scanner employed has an 8 MHz transducer, yielding spatial resolution on the order of less than 1 mm. Scans are produced at a nominal rate of 60 sec^{-1} by oscillating the sound beam deflecting mirror. The image produced represents a 3 cm x 4 cm tomographic section, with the exact plane imaged depending on the transducer assembly position. The system used has no image memory or "scan converter". The video signals from the scanner are sent directly to television monitors for viewing and photography and for application to a video tape recorder.

SYNCHRONIZATION

The magnitude of vocal fold displacement and the precise timing of displacement patterns recorded with ultrasound were compared with vocal fold movements recorded simultaneously with stroboscopic video glottography. Precise timing of the ultrasound B-mode image and the stroboscopic image was achieved by synchronizing the light source in the strobe unit to the ultrasound scanner. This arrangement was necessary because the mechanical scanner requires a fixed scanning rate that is difficult to alter with externally applied signals.

The present version of the stroboscopic unit (Wolf Stroboscopic Recording System) appears to synchronize quite well to a voice signal. However, it could not be synchronized directly to an external signal source, and we were not permitted to modify the apparatus to do so. Therefore, it was necessary to use the arrangement shown in Figure 13-1 to achieve synchronization.

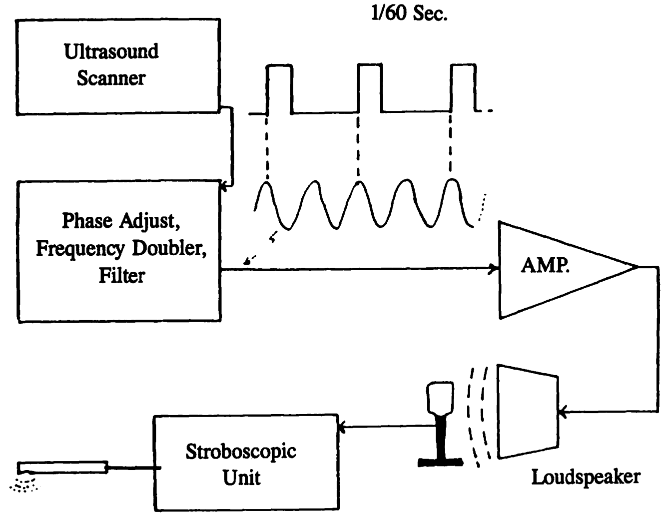


FIGURE 13-1
Apparatus for noninvasive synchronization of the light flashes of a stroboscopic video glottography apparatus to the image frame of a rapidly scanning ultrasound instrument.

Image synchronization pulses produced by the ultrasound scanner (60 per second) were applied to a frequency doubling and signal conditioning circuit, producing a 120 hz sine wave. (120 hz was necessary because 60 hz was outside the passband of the strobe circuitry.) The sine wave was used to drive a loudspeaker, the output being picked up by the strobe's input microphone. This locked the light source frequency to exactly twice that of the scanner. The phase of the sine wave relative to the input trigger from the scanner could also be varied. This allowed the light flashings of the strobe unit to be synchronized to any scan line of the B-mode ultrasound image.

For these studies the phase of the sine wave was adjusted so that the light strobed coincidentally with a position cursor on the B-mode image. A cadmium sulfide detector (Figure 13-2) was used to verify that the strobe and the cursor remained time locked to within 300 nanoseconds for repeat activations of the system.

COMPARISON STUDIES

We have carried out frame-by-frame analysis of split screen video images of the stroboscope and ultrasound images. When the two imaging modalities are synchronized pro-

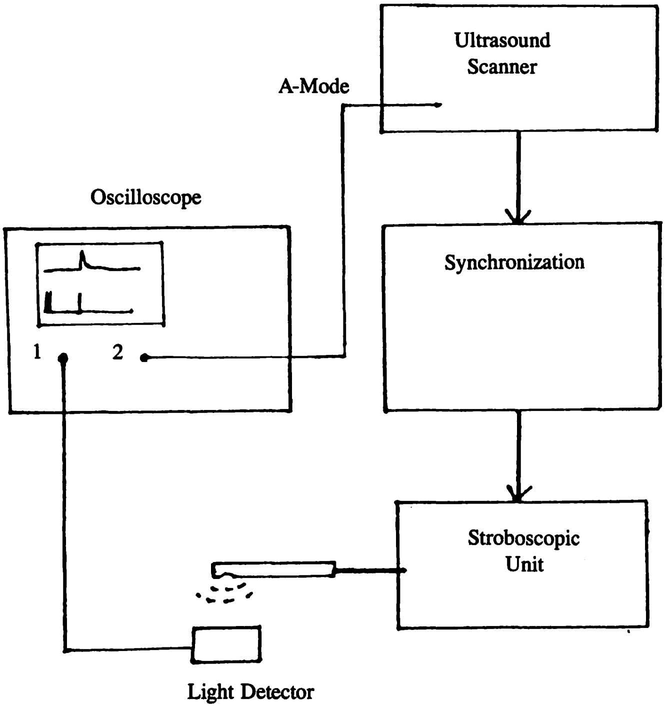


FIGURE 13-2
System used to test the reproducibility of the synchronization system shown in Figure 13-1.

perly the mesio-lateral excursions of the vocal fold as viewed from the strobe are matched by corresponding movements on the ultrasound image. The superior aspects of the B-mode image of the vocal fold correlate best with the optical image. Thus, it would appear that the mechanical sector scanner is indeed providing reliable images of the vocal folds.

DISCUSSION

In its present stage of development B-mode ultrasound imaging could be used for several types of studies of the larynx during speech production. Laryngeal heights can be determined during various speech activities and non speech maneuvers. This may be done in the presence of relative motion between the thyroid cartilage and vocal ligaments. Thus, effects of breath level, pitch and vowel type on the exact position of the vocal cords may be determined. The amount of mesio-lateral motion of the vocal folds can also be determined.

Although the present ultrasound system has been found useful for some investigations of vocal fold activity, it does exhibit several limitations relative to this application. These include too restrictive an image plane for visualizing extreme vertical excursions of the larynx and the fixed, 60 Hz scanning rate.

The above limitations of the present system are by no means fundamental to the modality, and an optimal instrument for studying and assessing vocal fold motion with B-mode ultrasound can be constructed. One proposed system is illustrated in Figure 13-3. The scanning assembly shown is a linear transducer array, operating at a 7 MHz ultrasound frequency. The

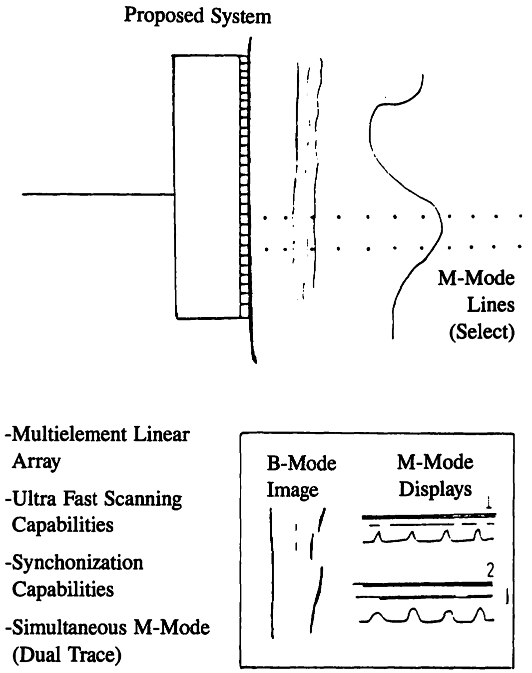


FIGURE 13-3
Proposed scheme for scanning the vocal folds with ultrasound. The linear array system could be synchronized to a voice signal, or a "real-time" scan could be carried out over a limited region.

array extends over a vertical segment of 5 cm, which would permit viewing the vocal folds during extreme excursions of the thyroid cartilage. The scanning speed could be improved upon, permitting speeds as high as 500 images/sec for images consisting of 40 separate acoustic lines and imaged depths of 4 cm, and scan speeds as high as 1,000 images/sec for a more limited, 20 line image. Alternatively, the scan could be synchronized to a voice signal or other source, since the electronically switched array is not subject to inertial effects as is the mechanical scanner. It would also be feasible to provide one or more M-mode lines on the image, each having a separate M-mode display indicating the movement patterns of structures along specific lines on the image. The lines involved could be switch selectable by the operator. Thus the precise timing of movements of specific structures or different regions on the same structure may be studied quite readily.

REFERENCES

Bless, D., Zagzebski, J., and Ewanowski, S. (1981). "Ultrasonic observations of the larynx," Paper, Julliard Conference on Care of the Professional Voice, New York.

Zagzebski, J., Bless, D., and Ewanowski, S. (1982). "Pulse echo imaging of the larynx using rapid ultrasonic scanners," in Bless, D.M. and Abbs, J.H. (Eds.) *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. San Diego: College-Hill Press, 210-222.

DISCUSSION

K. HARRIS: What is the research goal of this use of the stroboscope, or is it to be used clinically?

D. BLESS: We have two objectives in our work with ultrasound. The first is to be able to describe vocal fold vibration and vertical height changes in voice-disordered speakers without having to insert a probe in the oral cavity or having to expose the patient to unnecessary radiation. Although we have used this system with patients, we do not feel we have sufficient understanding of the images to begin applying it routinely to the clinical population. We are currently working to describe the structures (e.g., cartilages and muscles) visualized with ultrasound in normal speakers. Once this has been accomplished we will begin to make more rigorous application to the clinical population. Our second objective is to describe the influence of lung volume level and tongue position on the vibratory characteristics and vertical height of the larynx. To this end we are making simultaneous measurements of ribcage and abdomen movements, lung volume, laryngeal movements (ultrasonic and stroboscopic recordings), and frequency and intensity recordings made during a variety of phonatory conditions.

P. KITZING: As you know, we have worked with ultrasound in Malmö, Sweden and we used the M-mode. It was so difficult, we stopped as we didn't think we could progress any more at that time. One of the problems we ran into was that the ultrasound pulse was transmitted into the fold on the other side when there was contact between the vocal folds. The other problem we had was with height changes when different pitches were produced. We tried to measure ongoing speech and that was very difficult. What I want to ask now is, could your frame rate, which seems to be 60 per second, be synchronized with the Fo to get a stroboscopic effect?

J. ZAGZEBSKI: Yes, and it could be done more efficiently using a transducer array scanner than a mechanical scanner as we used in the present study.

P. KITZING: Do you have any comments on the signal going over to the other side?

J. ZAGZEBSKI: In our imaging, that would reduce the amplitude of the echo signal from the tissue-air interface of the vocal fold. We may be losing part of our echo signal from the vocal folds. This may, for example, result in our underestimating the actual mesiolateral displacements of the vocal folds. We haven't really looked into this. Since it could lead to artifacts, this is certainly one of the things that requires looking at with these combined images.

O. FUJIMURA: I think the most important artifact that has to be corrected in this kind of stroboscopic technique is that in this case, there is a phase progression from frame to frame. But at the same time there is a very similar progression in phase of the same order and magnitude from position to position. As you sweep there is a delay, and the duration of the interval from position to position is comparable to the difference in terms of frame. So you have to correct for that in order to get the correct image.

J. ZAGZEBSKI: Even though a single scan takes a sixtieth of a second to carry out, the actual time spent with the sound beam traversing the region of the vocal fold is about 2 milliseconds. So that limits the details of the motion patterns that we can pick up. We can probably improve on it (with an optimized system) by perhaps a factor of two or three, but certainly scanning speed is always going to be a limitation.

O. FUJIMURA: The point is, that could be corrected if you store the information you obtain in a computer and make adjustments in terms of the interpretaion of the phase. You can do that, except that we have to assume that the vocal fold vibration is regular. But that's necessary in any case for any stroboscopic technique.

J. ZAGZEBSKI: Well, one technique that might be utilized, if you were trying to do a precise reconstruction of the time events, is creating M-mode or time motion mode tracings on a time versus distance scale at different locations in the larynx. You could easily incorporate the correction that you referred to.

14 ▪ PELLET TRACKING IN THE VOCAL FOLD WHILE PHONATING — EXPERIMENTAL STUDY USING CANINE LARYNGES WITH MUSCLE ACTIVITY

Shigeji Saito, Hiroyuki Fukuda, Satoshi Kitahara, Yutaka Isogai, Tohru Tsuzuki, Hiroshi Muta, Etsuyo Takayama, Tadashi Fujioka, Nobuyuki Kokawa, and Katsumi Makino

Generally speaking, a great deal of information about vocal fold vibrations can be gained with a laryngeal mirror, used with the convenient stroboscopic observation or the ultrahigh speed cinematographic observation. Based on this information, we can draw a three dimensional vibratory pattern. However, such a procedure is only a hypothetical depiction. In order to confirm this characterization, it is necessary to observe the vibration from the frontal plane view.

We have already reported the results of our experiments with the newly designed X-ray stroboscope system with which the vibratory pattern from the frontal plane view can be observed.

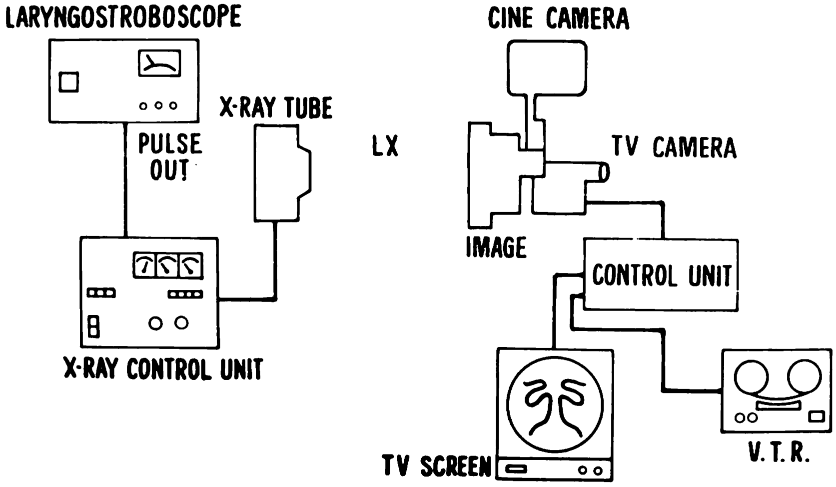


FIGURE 14-1
Block diagram of the system.

Voice-synchronized X-rays can be obtained by the use of a pulse signal from a laryngostroboscope as a trigger to pulse generated X-ray equipment. The block diagram of this system is shown in Figure 14-1.

At previous conferences of vocal fold physiology held in Kurume and Madison, we demonstrated the superficial movement of the vocal fold.

The movements of the tiny lead pellets injected into the various parts of the vocal fold are demonstrated in Figure 14-2. The phases of movement of each lead pellet are delayed one behind the other. The movement of the lead pellets is greatest at the free edge. The

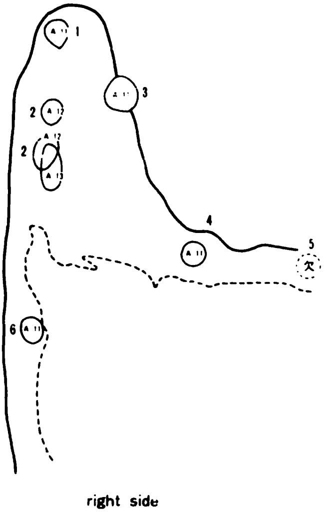


FIGURE 14-2
Schematic figure showing location of pellets.

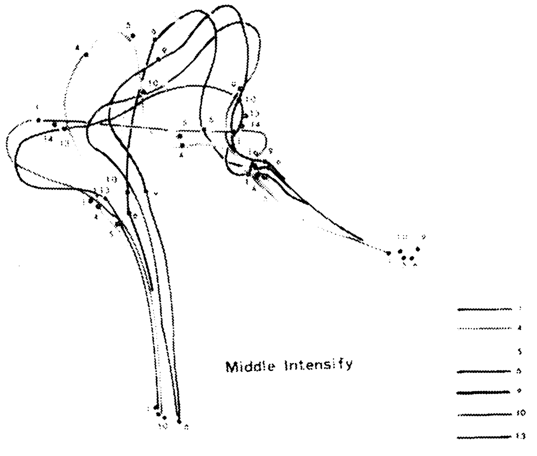


FIGURE 14-3
Imaginary drawing of the vibration.

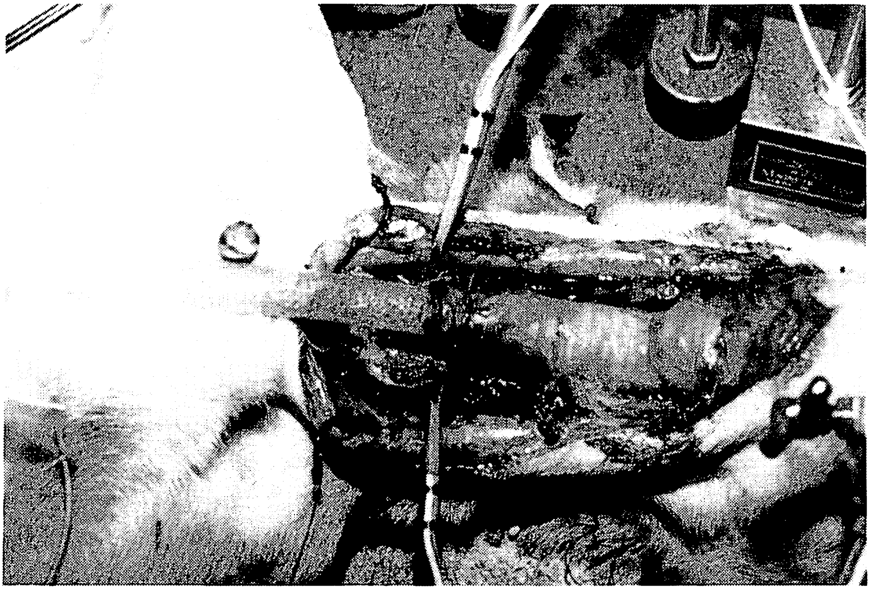


FIGURE 14-4
Exposed larynx with intact RLNs prepared for electrical stimulation.

vibratory pattern of the vocal fold is obtained by analyzing movement of each lead pellet as shown in Figure 14-3. This imaginary figure was obtained by drawing lines between particles at the same phase.

Basic studies including the one mentioned above let us conclude that the mode of vibration of the vocal fold is considered to be a traveling wave. This wave is made by a rotatory movement with differentially shifted phases along the vibratory line of the vocal fold.

For further studies, it is extremely necessary to observe the motion of the inner layer of the vocal fold, namely the muscle layer, and discuss the difference between the inner layer movement and the superficial one.

METHOD

Adult dogs weighing 10 - 15 kg were used for this study. Dogs were anesthetized with 25 - 35 mg/kg body weight pentobarbital sodium (Nembutal Sodium®) using the intravenous route. Additional pentobarbital was later added to maintain the uniform level of anesthesia.

A vertical midline incision was made on the neck from below the mandible to just above the sternum. The trachea was cannulated. Then the larynx was exposed with intact bilateral recurrent laryngeal nerves (RLN).

The bilateral vocal folds were forced to adduct to obtain glottal closure by suture of the posterior commissure using nylon thread. After this procedure, the RLNs were electronically stimulates to obtain muscle activity. The high frequency electrical stimulation just above the threshold for the RLNs is thought to mainly activate the thyroarytenoid muscles. A polyethylene tube was inserted into the cut-end trachea for introducing the air flow needed to obtain experimental sound. (Figure 14-4).

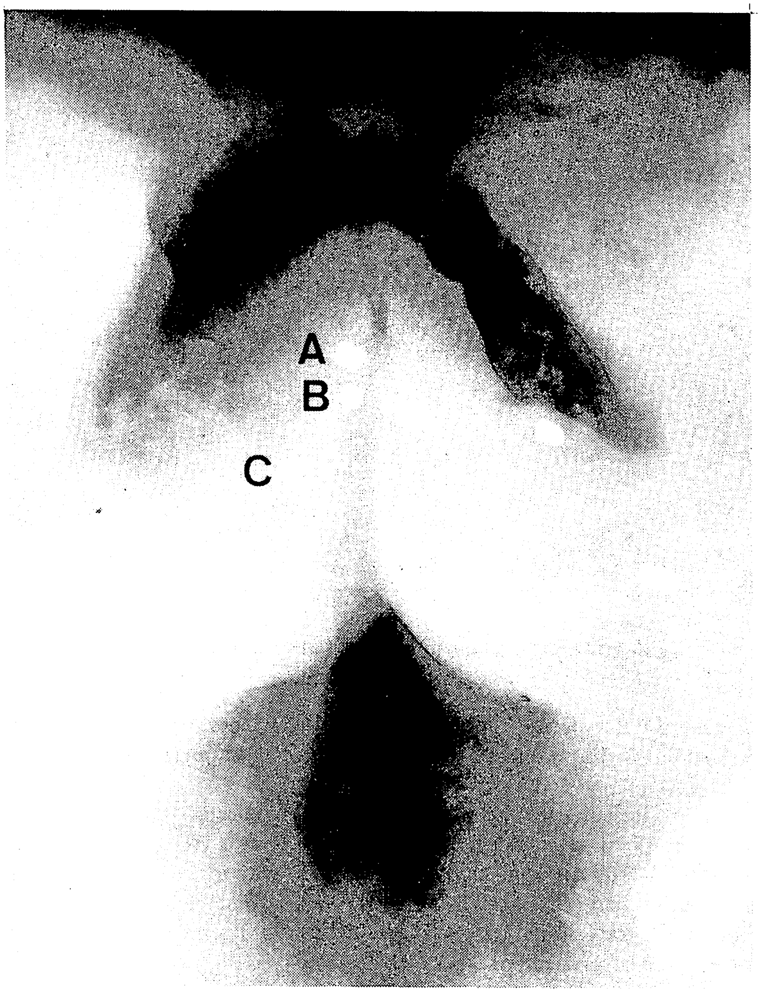


FIGURE 14-5
X-ray film showing location of pellets.

Tiny lead pellets were inserted into the muscle layer and into the superficial layer. Thus, we can see movement of the lead pellets inserted into both the superficial layer and inner layer while the laryngeal muscles are activated.

RESULTS

The location of the pellets are shown in Figure 14-5. Dynamic analysis was performed with three pellets inserted into the left vocal fold. Two pellets (A, B) among them were inserted into the so-called free edge (superficial layer). The histological section is shown in Figure 14-6. The other pellet (C) was inserted into the muscle layer, not so far from the so-called free edge (Figure 14-7).

FIGURE 14-6
Histological figure showing location of pellets (A, B).

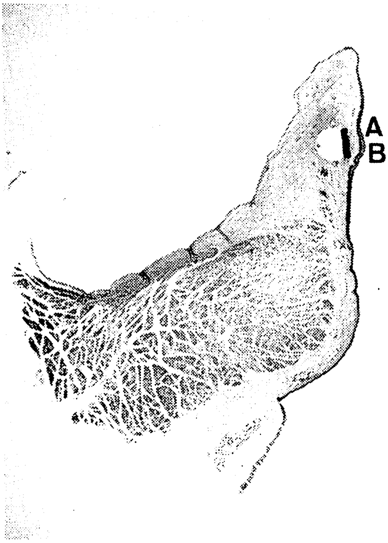
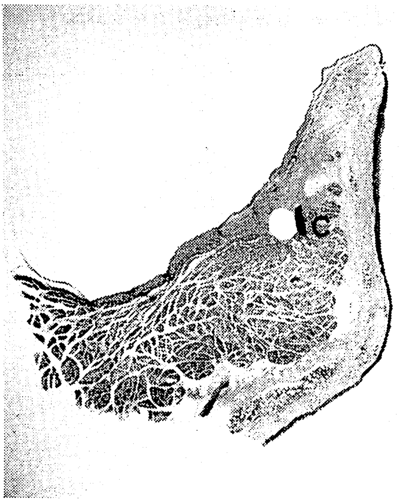


FIGURE 14-7
Histological figure showing location of pellet (C).



The phasic motions of pellets A and B along the X and Y axes are shown in Figure 14-8 and 14-9. The components of movement in both the X and Y axes are very phasic. This fact reveals that there should be vibratory movements in this area during phonation. On the other hand, pellet (C) inserted into the muscle layer did not vibrate regularly (Figure 14-10).

Their Lissajous' figure is shown in Figure 14-11. Pellet (C) of the same figure presents a very different pattern. Random spots show a different pattern.

Another representative case is demonstrated in the following figures. The pellets inserted into the very superficial part and the middle depth of the so-called free edge (Figure 14-12, 14-13), vibrate well (Figure 14-14, 14-15).

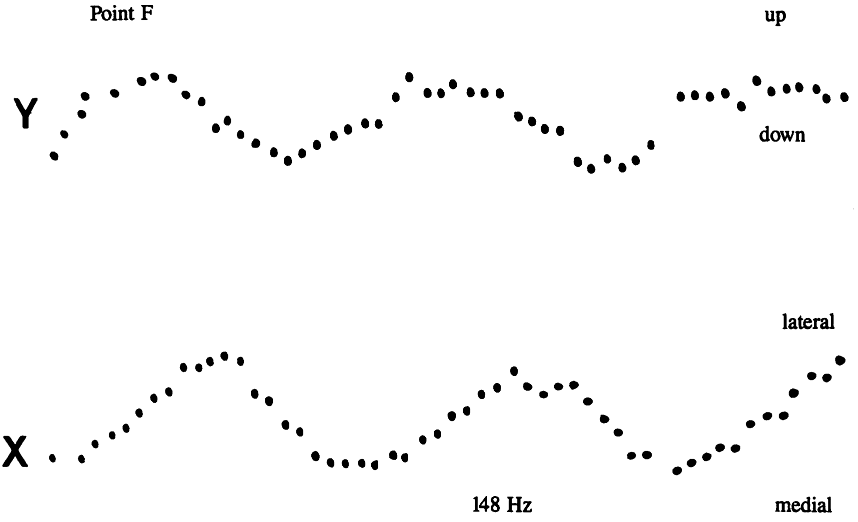


FIGURE 14-8
Pellet (A) tracking along X and Y axis.

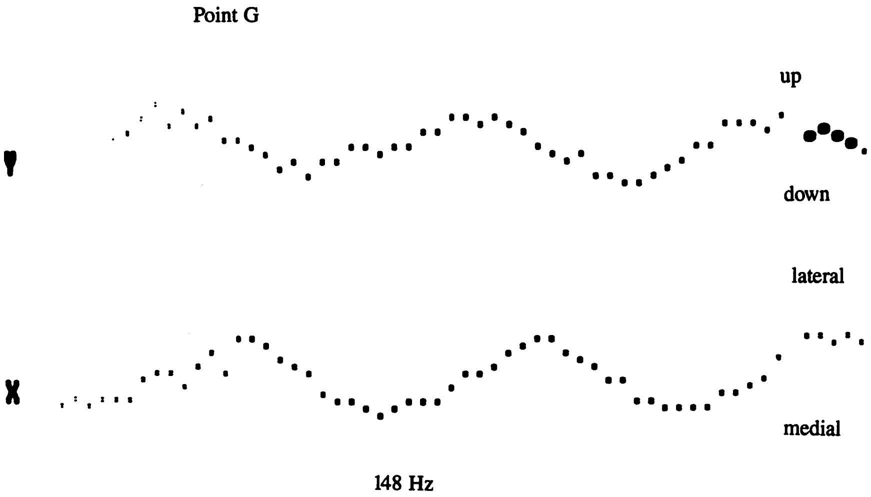


FIGURE 14-9
Pellet (B) tracking along X and Y axis.

However, we could not obtain any regular movement in the third pellet (Figure 14-16) as shown in Figure 14-17. This pellet was inserted into the border of the muscle and superficial layers. Their Lissajous' figures are shown in Figure 14-18. The spot group indicated by an arrow was from the movement of the latter pellet.

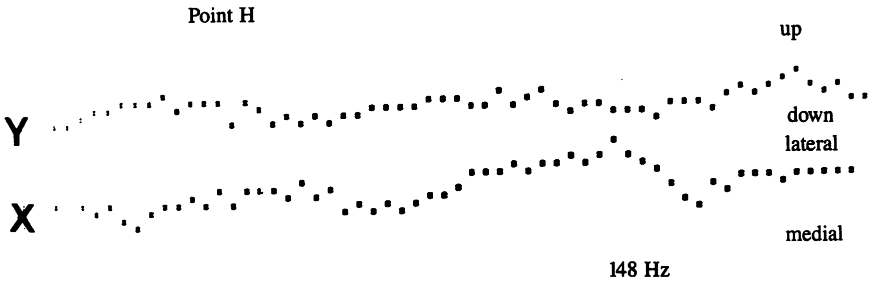


FIGURE 14-10
Pellet (C) tracking along X and Y axis.

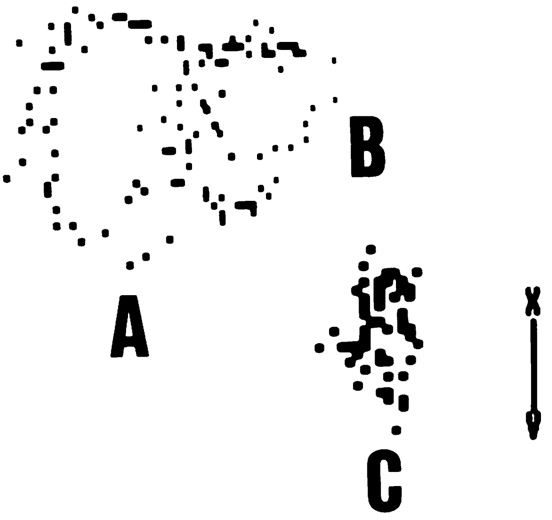


FIGURE 14-11
Lissajous' figure obtained in trackings of pellets A, B and C.

DISCUSSION

It is very beneficial to study the vibratory mode of the vocal fold using the larynx with activated laryngeal muscles instead of excised larynges, especially for clarifying whether or not vibration is in the inner layer (muscle layer).

From our experiments, it is very clear that no meaningful movement is obtained in the pellet inserted into the muscle. The fact that the spots were at random in Lissajous' figures shows that the movement of the inner layer does not play an important role in regard to "VIBRATION" which is essential during phonation.

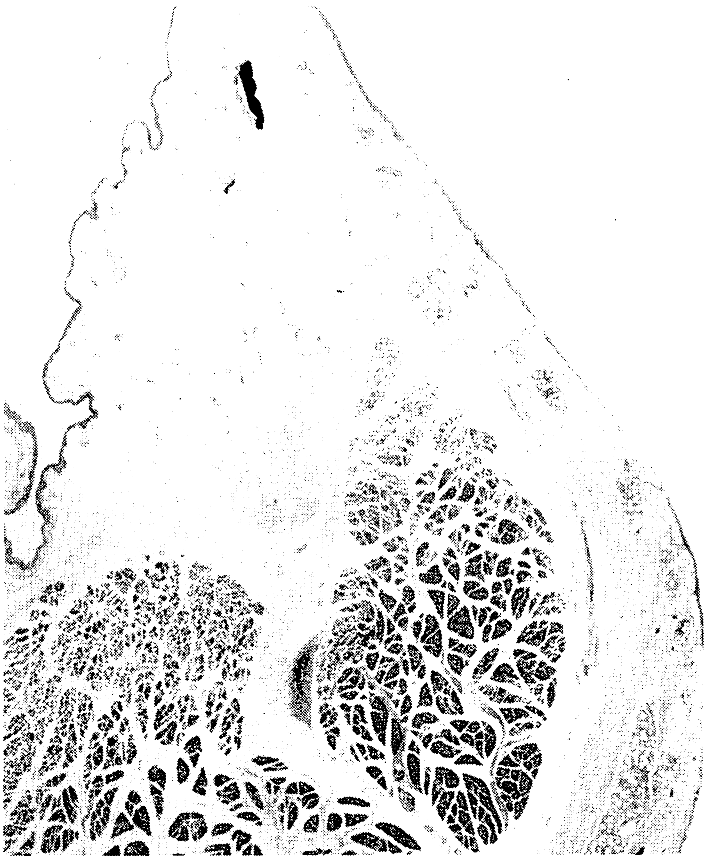


FIGURE 14-12
Histological figure showing a pellet in the superficial layer.

Therefore, we think that the role of the muscle layer is to regulate the shape of the vocal fold, based on the physiological characteristics of the phonation. By this action, the condition and character of the membranous layer should be altered. Thus, the speed and amplitude of the traveling wave might be precisely regulated.

CONCLUSION

1. Vibratory movements were obtained in the pellets inserted into the so-called free edge.
2. On the other hand, no meaningful movement was obtained in the pellets inserted into the muscle layers.
3. It is clear from the facts mentioned above and reported formerly that the essential principle of the vibration is the traveling wave in the superficial layer under any condition of phonation.
4. The muscle layer might regulate the mode of traveling waves, depending on phonatory conditions.

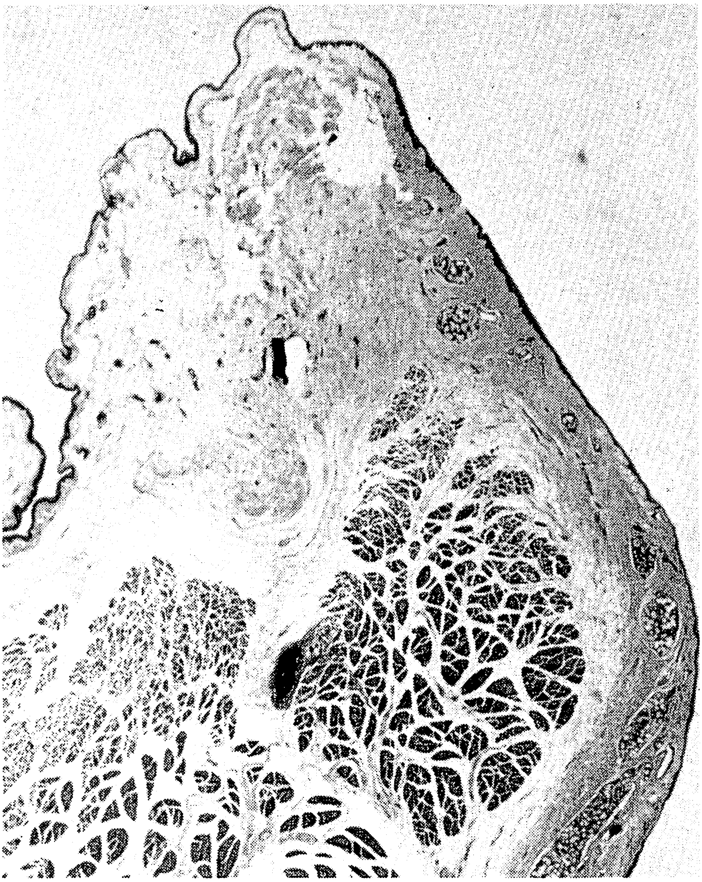


FIGURE 14-13
Histological figure showing a pellet in the middle depth of the superficial layer.

REFERENCES

Baer, T. (1981). "Observations of vocal fold vibration: Measurement of excised larynges," in Stevens, K.N. and Hirano, M. (Eds.) *Vocal Fold Physiology*. Tokyo: Univ. of Tokyo Press.

Hirano, M. (1975). "Phonosurgery," *Otologia (Fukuoka)* 21, 239-440.

Kakita, Y., *et al* (1976). "Schematical presentation of vibration of the vocal cords as a layer-structured vibrator," *J. Otolaryngol. Jpn.* 79, 1333-1340.

Kaneko, T., Uchida, K., Suzuki, H., Komatsu, K., Kanesaka, T., Kobayashi, N. and Naito, J. (1981). "Ultrasonic observations of vocal fold vibration," in Stevens, K.N. and Hirano, M. (Eds.) *Vocal Fold Physiology*. Tokyo: Univ. of Tokyo Press.

Saito, S. (1977). "Phonosurgery," *Otologia (Fukuoka)* 23, 171-384.

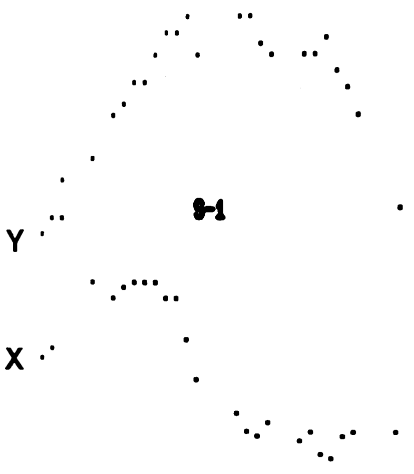


FIGURE 14-14
Tracking of the pellet in Fig. 14-12.

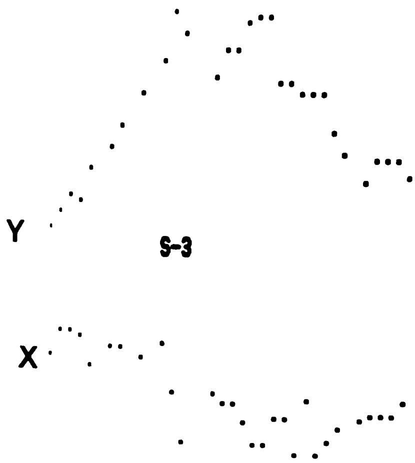


FIGURE 14-15
Tracking of the pellet in Fig. 14-13.

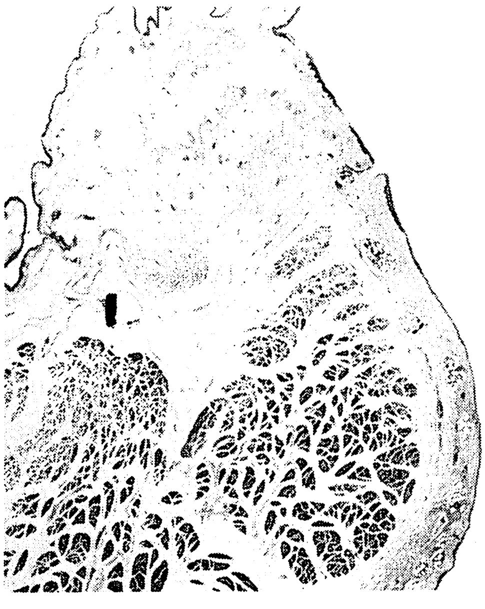


FIGURE 14-16
Histological figure showing a pellet at the border of the muscle and superficial layer.

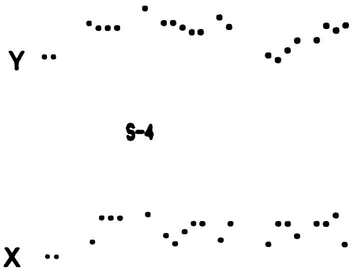


FIGURE 14-17
Tracking of the pellet in Fig. 14-16.

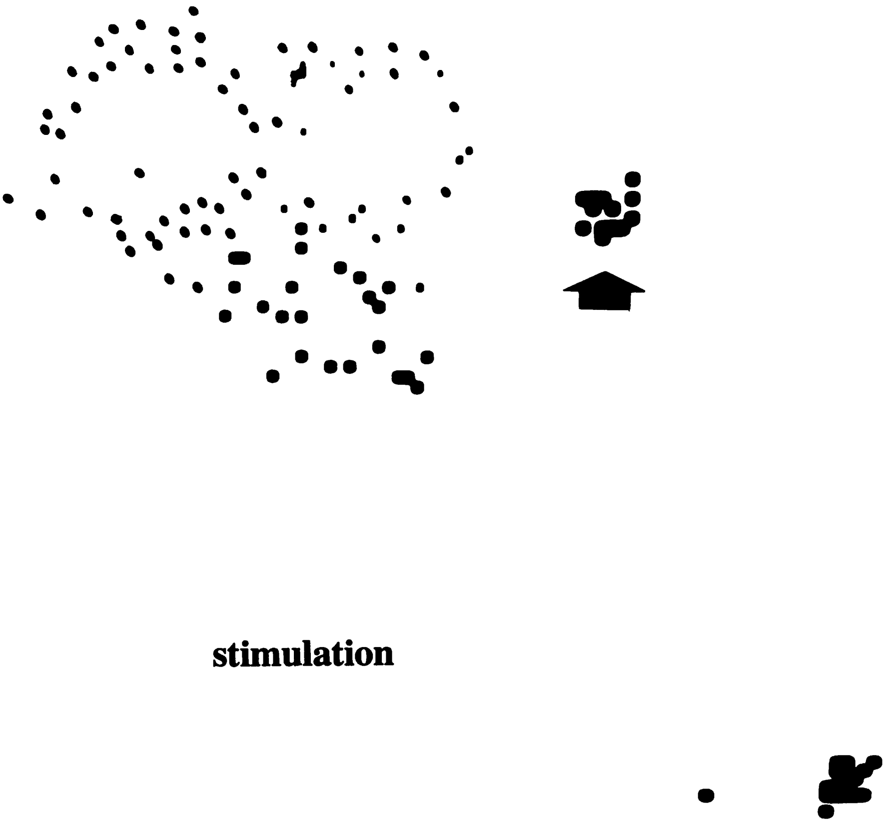


FIGURE 14-18
Lissajous' figure obtained in pellets tracking of Fig. 14-14, 14-15 and 14-17.

DISCUSSION

R. SCHERER: Were you able to determine pellet movement differences with respect to longitudinal placement?

S. SAITO: We are now doing research that will examine differences in pellet movement due to differences in their longitudinal placement, but we don't have any final results yet.

P. KITZING: Could you please comment on the voice quality, register and fundamental during your experiments?

S. SAITO: The vibratory pattern of the vocal folds was steady. The fundamental frequency was about 150 Hz.

P. KITZING: Is there any difference in the vibratory pattern between specimens and in vivo larynges?

S. SAITO: From my observation, they were almost the same. There was no difference observed between the two.

M. HAST: Please describe the parameters of your stimulation artifact; i.e., nerves that were stimulated, at what voltage or amperes, length of pulse, number of pulses per second in train?

S. SAITO: The conditions of stimulation were from 0.5 to 0.7 volts from 30 to 60 Hz with 1 msec duration.

M. HAST: What effect did the parameters of the stimulus artifact have on the vibration (Fo) of the vocal folds and on the movement of the pellets?

S. SAITO: We have not performed quantitative measurements of the relationship between the nerve stimulation and the movement of the pellets. When a good experimental sound is obtained, then we observe the movement of the pellets.

J. KAHANE: This is a beautiful demonstration of complex laryngeal physiology. You and your colleagues are to be congratulated. Where along the superior surface of the vocal fold did you find greatest movement and displacement?

S. SAITO: The movement of the lead pellets is greatest at the free edge. The pellet at the lateral part of the surface of the vocal fold moves at a smaller amplitude. The phases in the movement of each pellet are delayed one behind the other.

W.J. GOULD: I just wanted to make a comment, and then also a short list of questions. First, I think this is a landmark paper, and I want to congratulate you. Number two is that I would like to know if you have any thoughts of doing an experiment where you strip the vocal fold of the canine, let it heal, and then try the lead pellet experiments, because I think that could lead to tremendous clinical implications for means of handling endoscopic work.

S. SAITO: Thank you for your advice. I would like to try that. I have performed microsurgery of the larynx about five thousand times since 1965 and we have only been able to observe the scars and the superficial layers with steroscopic examination, not the vibration.

W.J. GOULD: That leads into the implication that the experimental work leads right into the clinical experience, which I happen to agree with you about.

S. SAITO: I would like to examine the larynx not only under normal conditions but also under pathological conditions.

M. HIRANO: In conjunction with Dr. Kitzing's question, I think I'd like to point out one thing about the vibration mode of the muscles and the mucosa. In 1974 I presented four hypothetical patterns of vibration (*Folia phoniat*. 1974, 26, 89-94) and I think Dr. Saito's experiment today very beautifully demonstrates one of those patterns, that is the case where the vocalis muscle is in full contraction and the cricothyroid muscle is not activated. In that case I postulated that the vibratory movement will be contained chiefly in the mucus membrane. In this kind of experiment, as Dr. Saito mentioned, it is very difficult to regulate the degree of contraction of the muscle. But, if you were able to adjust the degree of contraction in varying degrees, maybe you would observe different patterns.

B. WYKE: You yourself quite correctly said that the activity of the vocal fold musculature influences the parameters of the traveling wave in what you call the superficial layers of the vocal fold, and I think we'd all go along with that. But in the type of experiment you are doing, you are producing tonic motor unit activity in the vocal fold musculature throughout the course of your observations. Now that is not the way motor units behave in the vocal fold musculature during natural phonation as every electromyographer knows. And secondly, it completely ignores, as do all isolated laryngeal experiments, the role of the multiplicity of feedback systems that are involved in modulating the activity of the musculature during phonation. And this is why we are very worried if the implication is that these experiments or the observations in these experiments should be extrapolated to the understanding of natural phonation. We are very doubtful that they have anything to do with it.

C. VAUGHAN: I have problems with this because we don't normally go around with a mass of lead pellets in our mucous membrane, and I wonder how you've accounted for that mass? Can you find some other radiopaque substance?

S. SAITO: A number of people, many of whom are here, have given me advice on this. It is possible, for example, to use small gold pellets.

C. VAUGHAN: The mass must have some influence. How have you accounted for it?

S. SAITO: At first I thought the same as you because in human clinical cases if there is a small polyp or small nodule in the free edge of the vocal fold, we can observe the disturbed vibration. So at first, we only inserted one or two pellets, but we could get vibration. So day by day we added pellets. Of course, the difficulty is that the vibration is slightly different, I think. But I haven't made quantitative meaurements of this observation.

O. FUJIMURA: Congratulations on the success. I consider this a great step forward. I wish very much that you would be able to do this in vivo. Also, I would like to congratulate Dr. Hirano for the experimental verification of what he has been saying.

The size of the pellet possibly can be reduced if you can get an x-ray source wtih a smaller source spot, because the detectability of the pellet is probably limited by the image blurring. Given the physical limitation that you have to use a certain amount of mass to make a radiopaque pellet, the approach one can take is to obtain results using different masses deliberately and extrapolate the comparison toward the no mass condition.

C. VAUGHAN: But until you have understood the influence of that mass, you cannot draw conclusions about the vibration pattern and I think that's the next thing you have to do.

O. FUJIMURA: That's what I'm saying.

I. TITZE: Just to reiterate the problems that might be associated with the size of the pellets. I can't remember what the specific gravity of lead is in relation to water, but its more than 10, so we have to imagine the pellets to be about twenty times their size if we want to relate them to the specific gravity or density of tissue. So it's not like having a nodule of the same size, which does perhaps affect the vibratory pattern a little bit, but it would have a much greater effect on it. Maybe the thing to do is not to implant too many. I know the principle of superposition doesn't apply in some of these complex movements, but perhaps you could look at the effect of several by looking at one at a time rather than loading the folds down with maybe five or six. I also think it's a marvelous piece of work. We look forward to using the data in our simulations.

T. BAER: In support of the methodology and perhaps one piece of evidence about the effects of the pellets on the vibratory pattern is that you have different placements on both sides. You often have pellets in both left and right folds in different positions, and it appeared to me from the films that the vibrations are still quite symmetric which would suggest that the pellets are not having a very great effect on the vibratory pattern. That would be one more piece of evidence to answer Dr. Vaughan's question.

L. MALMGREN: I'd like to ask whether the pellets were solid lead or whether they were lead coating on something. I wonder how thick the lead would have to be before it became opaque?

S. SAITO: It is solid lead, but it is flat, not round.

15 ▪ SIMULATION OF PARTICLE TRAJECTORIES OF VOCAL FOLD TISSUE DURING PHONATION

Fariborz Alipour-Haghighi and Ingo R. Titze

INTRODUCTION

The purpose of this study was to investigate the possibility of getting self oscillation of the vocal folds by using a finite element method, and to obtain tissue particle trajectories. The emphasis is on computational techniques rather than specific results.

The governing equations for vocal fold vibration are time dependent, and since the driving forces are air pressure, these equations are also coupled to nonlinear air flow equations. A closed form analytic solution is therefore not possible and a numerical method must be used. The finite element technique was chosen in this study because of its capability to handle not only complex geometry and boundary conditions, but also because it can handle inhomogeneous, anisotropic and multi-layered media. Unlike some other methods, such as finite difference methods in which field variables are defined only at discrete points, the finite element technique assumes a continuous distribution of field variables within the domain of each element.

I. FINITE ELEMENT MODELING OF TISSUE MECHANICS

Cinematography of vocal folds has shown that small portions of the vocal folds (hereafter called particles) typically exhibit planar motion, such that at any moment the displacement vector of any particle of the fold can be expressed by a two dimensional vector,

$$\Psi = u\underline{i}+w\underline{j} \qquad (15\text{-}1)$$

where u and w are lateral (x) and vertical (z) components of displacement, respectively, with corresponding unit vectors $\underline{i}$ and $\underline{j}$. In general, these displacements are functions of both time and space, and the medium is viscoelastic with complex boundary conditions.

An attempt was made to model the vibrations of the folds and to find the time-dependent displacement field $\Psi(r,t)$ by using a finite element method. To make modeling easier, the vocal fold was divided into plane layers parallel to vectors u and w so that two-dimensional motion could be applied to each layer. For sufficiently thin layers, the displacement field does not change very much across the thickness of each layer, so that simplifications are valid, except perhaps at the end layers where the boundary is fixed.

A. Variational Principle

Formulation of the problem with a finite element technique is based on the minimization of the total potential energy of the system via a variational principle. A variational principle for this problem can be obtained directly by identifying strain energy from elasticity relations, such that

$$\delta\pi = \delta(U - W_p) = 0 \qquad (15\text{-}2)$$

where π is potential energy, U is strain energy, and W_p is virtual work. The strain energy is further defined as

$$U = \iiint U_o \, dv \tag{15-3}$$

$$U_o = \tfrac{1}{2}(\sigma_x \varepsilon_x + \sigma_y \varepsilon_y + \sigma_z \varepsilon_z + \tau_{xy}\gamma_{xy} + \tau_{yz}\gamma_{yz} + \tau_{zx}\gamma_{zx}) \tag{15-4}$$

where U_o is the strain energy density of Hookian elastic material and σ, τ, ε and γ are normal stresses, shear stresss and strains. Then

$$W_p = \iiint_v (F_x u + F_y v + F_z w) \, dv + \iint_s (T_x u + T_y v + T_z w) \, dA \tag{15-5}$$

is virtual work done by the body force $\underline{F}$ and the surface force $\underline{T}$. In this case, $\underline{F} = -\rho \frac{\partial^2 \psi}{\partial t^2}$ is the inertial force and $\underline{T} = \underline{P}$ is the pressure force.

Using stress - strain displacement relations for a linear elastic system with small deformations, we have

$$\varepsilon_{ij} = \tfrac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$$

$$\{\sigma\} = [c]\{\varepsilon\} \tag{15-6}$$

where [c] is the material stiffness matrix containing 36 terms for general anisotropic materials. The potential energy of the system reduces to

$$\begin{aligned}
\pi = \tfrac{1}{2} \iiint & \Big\{ (\lambda + 2\mu)\Big[\Big(\frac{\partial u}{\partial x}\Big)^2 + \Big(\frac{\partial w}{\partial z}\Big)^2\Big] + 2\lambda \frac{\partial u}{\partial x}\frac{\partial w}{\partial z} \\
& + \tfrac{1}{2}\mu' \Big[\Big(\frac{\partial u}{\partial y}\Big)^2 + \Big(\frac{\partial w}{\partial y}\Big)^2\Big] + \tfrac{1}{2}\mu\Big[\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\Big]^2 \Big\} dv \\
& + \iiint \rho\Big(u \frac{\partial^2 u}{\partial t^2} + w \frac{\partial^2 w}{\partial t^2}\Big) dv \\
& - \iint (p_x u + p_z w) \, dA
\end{aligned} \tag{15-7}$$

Minimization of this energy provides the governing equations of the system.

If we divide the system into finite elements, this energy splits between the elements, and integrals can be carried out in the elemental domain. The same minimization principle holds true for each element. To simplify modeling and integration, we assume that the effects of y-variation can be given as a separate function such that

$$\begin{aligned} u &= U(x, z, t) \cdot f(y) \\ w &= W(x, z, t) \cdot g(y) \end{aligned} \tag{15-8}$$

where f and g satisfy the end boundary condition, i.e.,

$$f(0) = f(L) = 0$$

$$g(0) = g(L) = 0 \qquad (15\text{-}9)$$

Now in a thin layer Δy we assume the functions f and g are not changing rapidly and take on their average values for the whole layer. The 3-dimensional problem thereby transforms to a 2-dimensional problem. The functions f and g may be assumed of the form of $\mathrm{Sin}(\pi y/L)$ to satisfy the anterior and posterior boundary conditions and to provide the approximate variation in displacement observed along the length of the folds.

B. Element Type and Shape Function

Triangular finite elements are widely used in 2-dimensional problems of stress analysis. The description of any irregular boundary can be specified by the density of the mesh without creating stability problems (as is frequently encountered with the finite difference method). The difficult task in any finite element method is mesh generation, especially in cases involving complicated geometries. In this study a mesh pattern is used in all layers as shown in Figure 15-2. The displacement function within the element is approximated by a linear interpolation function N for both U and W as

$$U(x, z, t) = \sum_{i=1}^{3} N_i U_i$$

$$W(x, z, t) = \sum_{i=1}^{3} N_i W_i \qquad (15\text{-}10)$$

where

$$N_i\,(x,z) = \frac{1}{2A}\,(\alpha_i + \beta_i x + \gamma_i z) \qquad (15\text{-}11)$$

A is the area of the element and α, β, and γ are evaluated from nodal coordinates. Substituting these expressions into the potential energy integral, differentiating with respect to nodal field variables U_i and W_i, and carrying out integration over the element domain, gives six equations of equilibrium for each element in the form of

$$[M]^e \left\{ \begin{matrix} \ddot{U} \\ \ddot{W} \end{matrix} \right\}^e + [K]^e \left\{ \begin{matrix} U \\ W \end{matrix} \right\}^e = \left\{ \begin{matrix} F_x \\ F_z \end{matrix} \right\}^e \qquad (15\text{-}12)$$

Effects of viscosity could be taken into account by replacing μ with $\mu + \eta \dfrac{\partial}{\partial t}$ and μ' with $\mu' + \eta' \dfrac{\partial}{\partial t}$. This adds extra terms of the form

$$\frac{d}{dt} \iiint [\cdots]\, dv \qquad (15\text{-}12a)$$

to the equations, which eventually appear as damping terms. We thus obtain an equation of motion of the form

$$[M]^e\{\Psi\}^e + [D]^e\{\Psi\}^e + [K]^e\{\Psi\}^e = \{F\}^e \qquad (15\text{-}13)$$

where $\{\Psi\}^e$, $\{\dot{\Psi}\}^e$ and $\{\ddot{\Psi}\}^e$ and $\{F\}^e$ are 6x1 vectors and $[M]^e$, $[D]^e$ and $[K]^e$ are element mass, damping and stiffness matrices. Assembly of the equations for the elements yields the matrix differential equation

$$[M]\{\Psi\} + [D]\{\Psi\} + [K]\{\Psi\} = \{F\} \qquad (15\text{-}14)$$

which we write simply as

$$M\ddot{\Psi} + D\dot{\Psi} + K\Psi = F \qquad (15\text{-}15)$$

Here Ψ, $\dot{\Psi}$, $\ddot{\Psi}$ and F are global vectors for all nodal points and M, D, and K are global mass, damping and stiffness matrices which have contributions from all the elements.

C. Solution Technique

The finite element method used in the solution of the spacial problem gives a second order matrix differential equation in time. Solution of this equation can be obtained by a finite difference scheme by marching with specific time step Δt from a known initial condition. If at the n-th time step the displacement vector is denoted by Ψ_n and its time derivatives by $\dot{\Psi}_n$ and $\ddot{\Psi}_n$, we can write

$$\begin{aligned} \dot{\Psi}_n &= \frac{\Psi_{n+1} - \Psi_{n-1}}{2\Delta t} + 0(\Delta t)^2 \\ \ddot{\Psi}_n &= \frac{\Psi_{n+1} - 2\Psi_n + \Psi_{n-1}}{(\Delta t)^2} + 0(\Delta t)^2 \end{aligned} \qquad (15\text{-}16)$$

Application of the Crank-Nicholson technique to stiffness force (to give stability to the system) yields

$$K\Psi = K\frac{\Psi_{n+1} + \Psi_n}{2} \qquad (15\text{-}17)$$

We have the following difference equations:

$$\frac{M}{(\Delta t)^2}\{\Psi_{n+1} - 2\Psi_n + \Psi_{n-1}\} + \frac{D}{2\Delta t}\{\Psi_{n+1} - \Psi_{n-1}\} + \frac{K}{2}\{\Psi_{n+1} + \Psi_n\} = F_n \qquad (15\text{-}18)$$

Multiplying this equation by $(\Delta t)^2$ and rearranging terms give the displacement vector Ψ_{n+1} in terms of two previous time step vectors Ψ_n and Ψ_{n-1} as

$$\begin{aligned}&[M + \tfrac{1}{2}(\Delta t)D + \tfrac{1}{2}(\Delta t)^2K]\Psi_{n+1}\\&\quad = (\Delta t)^2F_n - [\tfrac{1}{2}(\Delta t)^2K - 2M]\Psi_n - [M - \tfrac{1}{2}(\Delta t)D]\Psi_{n-1} \qquad (15\text{-}19)\end{aligned}$$

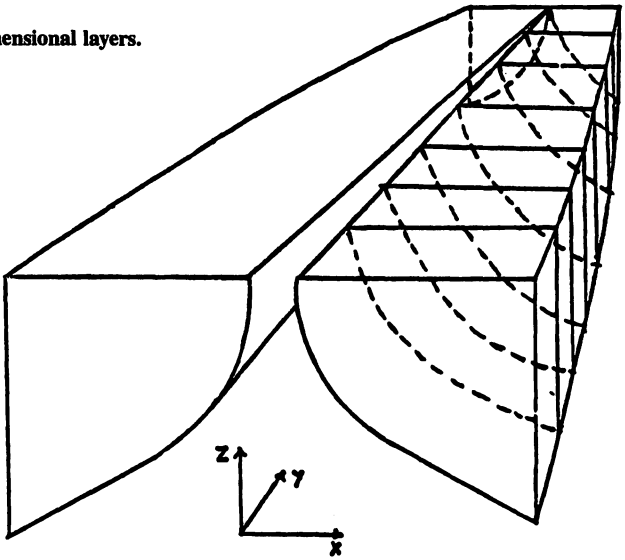
Having two initial conditions, namely ψ_0 and ψ_0, allows the determination of the displacement vector at any moment in time by using the above recursive relationship, which set-up linear algebraic equations through an elimination algorithm.

II. RESULTS OF THE SIMULATION

Preliminary results were obtained for an initially closed glottis. As shown on Figure 15-1, the folds were divided into 7 layers of equal thickness. Triangular elements of different size were selected such that smaller elements existed in the tip of the fold, where the largest movement is expected. Two different materials were considered to be within the fold: elements 1-8, soft tissue with elastic constant of 10 kPa; elements 9-17, a harder tissue with elastic constant of 25 kPa and Poison ratio of 0.33.

Figure 15-3 shows the movement of the vocal folds in one cycle with a period of approximately 10 ms. Figure 15-4 shows the trajectories of nodal points in the medial layer of the fold for a subglottal pressure of 8 cm H_2O. The trajectories start with a rising counterclockwise curve, turn into a clockwise rise, and then fall to complete a figure-eight shaped trajectory which varies from node to node. Also it was found that the trajectories are highly affected by elastic constants and pressure, especially the glottal pressure profile. In this study a simple parabolic pressure profile was used, without solving for the air flow in the glottis. Also, the viscoelastic properties are approximate values. Thus more realistic trajectories would

FIGURE 15-1
Vocal folds and 2-dimensional layers.



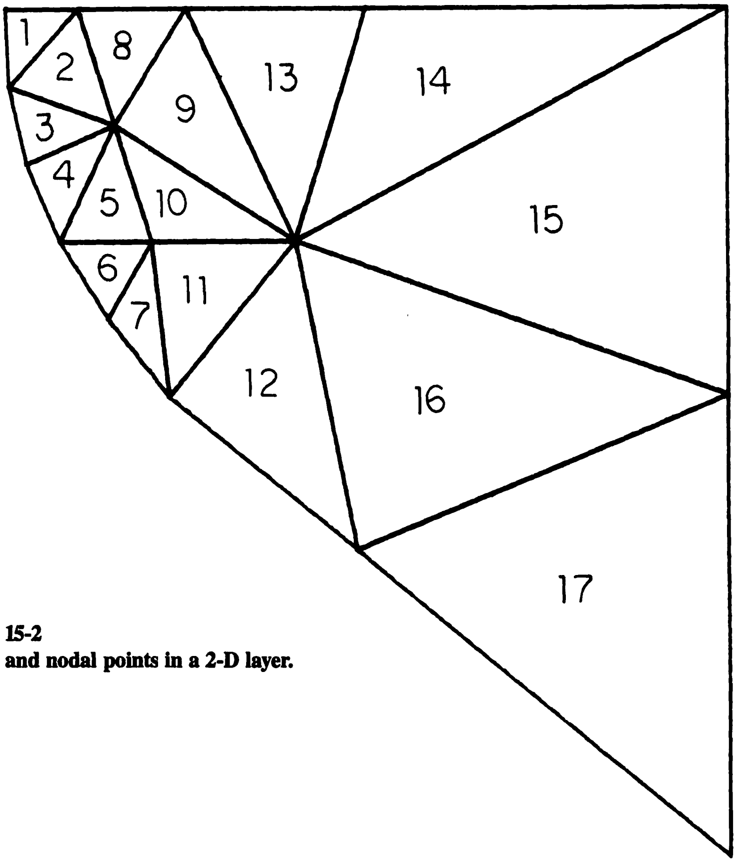


FIGURE 15-2
Elements and nodal points in a 2-D layer.

perhaps be obtained with a more accurate solution of the air flow and better knowledge or viscoelastic constants.

In summary, the finite-element technique seems to show promise toward providing a better solution for the spatial variation of the tissue displacement. In particular, it is better suited for accurate description of surface forces in the glottis, for quantifying the layered structure, and for maintaining computational stability. Several problems still remain, however. In order to keep the number of elements manageable, we have specified the longitudinal variations by an overly simplified formula. A computational approach for handling the coupling of adjacent layers is needed. Finally, more work is needed to quantify the pressure-flow relationships in the glottis.

REFERENCES

Archer, J.S. (1963). "Consistent mass matrix for distributed mass system," *J. of Struct. Div.*, Proc. of ASCE, ST4, 161-178.

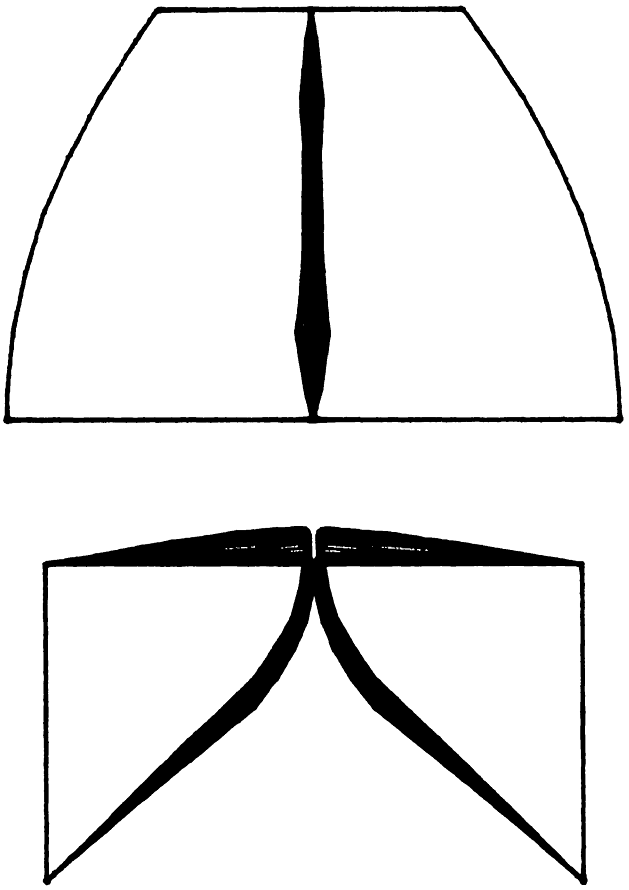


FIGURE 15-3
Folds opening in one cycle.

Desai, C.S. (1979). *Elementary Finite Element Method*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Jennings, A. (1977). *Matrix Computation for Engineers and Scientists* John Wiley & Sons.

Nickell, R.W. (1973). "Direct integration methods in structural dynamics," *J. of Engr. Mech. Div.*, Proc. of ASCE, EM2, 303-317.

Oestreicher, H.L. (1951). "Field and impedence of an oscillating sphere in a viscoelastic medium with an application to biophysics," *J. Acoust. Soc. Am.* 23/6, 707-714.

Smith, I.M. (1982). *Programming the Finite Element Method with Application to Geomechanics*, John Wiley & Sons.

Titze, I.R. (1981). "Biomechanics and distributed-mass models of vocal fold vibration," in Stevens, K.N. and Hirano, M. (Eds.) *Vocal Fold Physiology*. Tokyo: Univ. pf Tokyo

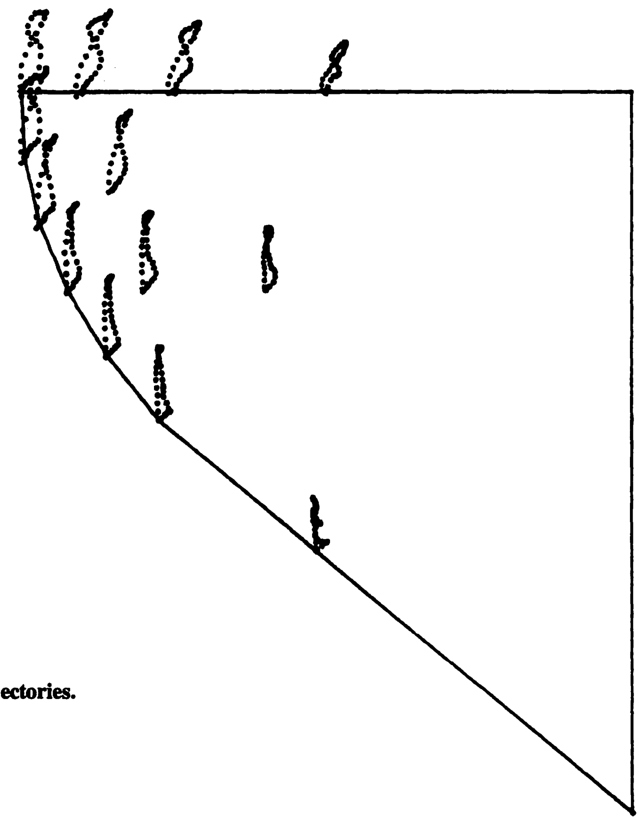


FIGURE 15-4
Nodal points trajectories.

Press, 245-264.

Titze, I.R. and Talkin, D.T. (1979). "A theoretical study of the effects of various laryngeal configuration on the acoustics of phonation," *J. Acoust. Soc. Am.* 66/1, 60-74.

Ugural, A.C. and Fenster, S.K. (1981). *Advanced Strength and Applied Elasticity*. New York: Elsevier.

Zienkiewicz, O.C. (1977). *The Finite Element Method*. McGraw-Hill.

16 ▪ STEREO-LARYNGOSCOPY: A NEW METHOD TO EXTRACT VERTICAL MOVEMENT OF THE VOCAL FOLD DURING VIBRATION

Yuki Kakita, Minoru Hirano, Hiroshi Kawasaki, and Koichi Matsuo

INTRODUCTION

This paper presents a new method of extracting the vertical component of the movement of a selected point on the vocal fold. Since this method allows a high speed filming in the order of 10^3 frames/s, the displacement in vertical movement can be obtained for one actual vibratory cycle during vocal fold vibration in living humans.

A stereoscopic measuring device for the larynx was first, to our knowledge, developed by Fujita (1938a, 1938b, 1943) for measuring the length of the human vocal fold in vivo. This method used a regular laryngeal mirror and a specially made binocular, which was just like today's operation microscope. Stereoscopic observation was not used for measuring the vertical position of the larynx, but for compensating for the vertical distance in order to obtain the accurate length of the vocal fold.

The stereo-fiberscope was devised by Sawashima and Miyazaki (1974) and later improved by Fujimura and his colleagues (Niimi & Fujimura, 1976; Fujimura, Baer & Niimi, 1979; Fujimura, 1979). This device was for measuring the larynx height during speech utterance.

Honda, Hibi, Kiritani, Niimi and Hirose (1980) and Sawashima, Hirose, Hibi, Yoshioka, Kawase and Yamada (1981) reported about a stereoendoscope for observing vocal folds. This method basically uses the technique of the stereofiberscope with respect to the analysis of stereoscopic images. The brightness of the image is better than the fiberscope. However, due to the light source and the transmission loss of the light intensity through the optical fiber, the highest filming rate was limited to 100-200 frames/s.

X-ray stroboscopy was developed by Saito, Fukuda, Isogai and Ono (1981) for a direct observation of the vibratory movement of the vocal fold in the frontal plane. By this method, we can observe the vocal fold in vivo without inserting an instrument through the mouth or nose. However, we have to put a metal pellet on the vocal fold to locate a specific point, or use a contrast medium to enhance the image. Also, this method requires consideration about the safety of x-rays. We can observe the movement for one vibratory cycle in the stroboscopic sense, but not continuously. Therefore, it is not possible to observe an irregular vibration for one actual vibratory cycle.

Our method presented in this paper was developed to observe a 3-dimensional movement, particularly the vertical, during one vibratory cycle (see Hirano, Kakita, Kawasaki & Matsuo, 1982 for an overview; see Kakita, 1982 for theory and data analysis). This method uses an ultra-high-speed filming system together with a specially designed mirror, which we call the "Stereo-laryngo-mirror" (SLM, hereafter).

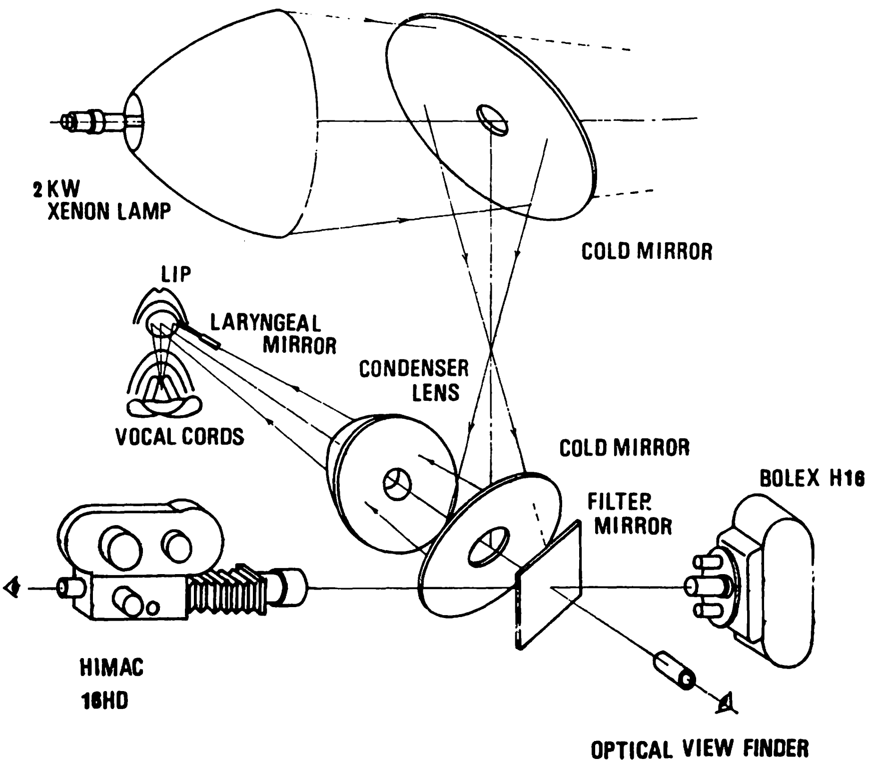


FIGURE 16-1
Schematic drawing of the ultra-high-speed filming system. (Yoshida *et al*, 1972; reprinted with permission).

2. METHOD

2.1 Instrumentation

Figure 16-1 shows a schematic drawing of the regular ultra-high-speech filming system. The details have been reported elsewhere. (Yoshida, Hirano, Matsushita & Nakashima, 1972; Hirano, Yoshida, Matsushita & Nakashima, 1974.) For stereoscopic filming the regular laryngeal mirror is replaced by the SLM.

Figure 16-2 shows three different views of the SLM. An SLM consists of two identical small mirrors combined to keep a specific angle. The height and the width of one mirror is about 2 cm and 1 cm, respectively. The four corners of the SLM are rounded so that the corners would not hurt the tissue of the subject's pharynx. The mirror portion is not larger than a laryngeal mirror with a diameter of 1 inch.

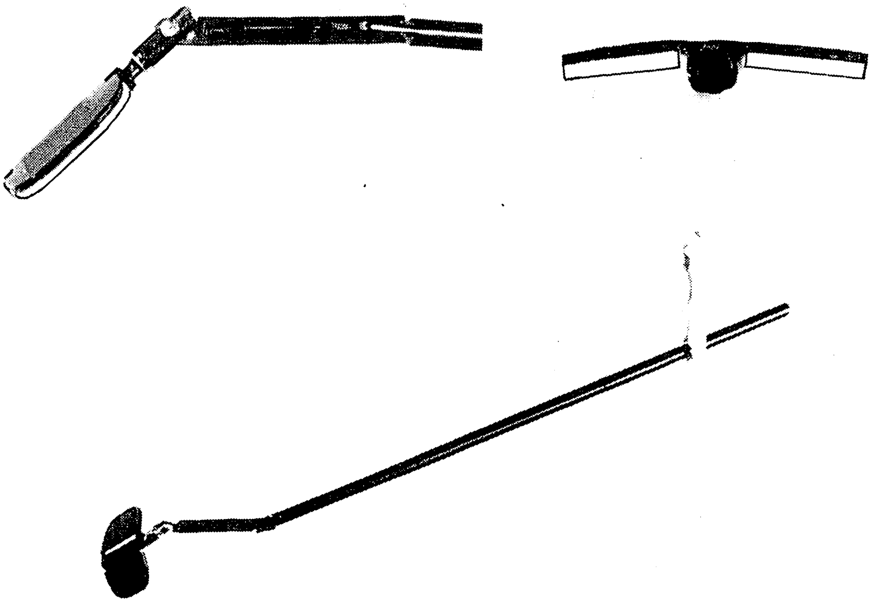


FIGURE 16-2
Three views of the stereo-laryngomirror. (Hirano *et al*, 1982; reprinted with permission).

2.2 Procedure

The examiner placed the SLM in the subject's pharynx. The position and tilt angle of the SLM were adjusted so that the image of the vocal fold could be seen through each mirror, and that the two images would look as symmetrical as possible. Filming was then done at 2600 frames/s for the example shown in this paper.

2.3 Theory

Figure 16-3 shows a schematic drawing of the stereoscopic imaging. M_L and M_R are the small mirrors constituting the SLM. Suppose that M_L and M_R are located symmetrically about the y-z plane. The virtual images of an object A made by M_L and M_R are I_L and I_R, respectively. When seen from the view point E ($y=l$), the two images look like those shown in Figure 16-4.

Let ϕ in Figure 16-3 be the tilting angle of the midline of SLM from the z-axis, θ, the tilting angle of each mirror toward the observer, and h, the z coordinate of a point P on the object A. When d in Figure 16-4 indicates the distance between the two images of P,

$$h \approx d/(4\theta\sin\Phi) \qquad (\theta \ll 1,\ l \gg h) \qquad (16\text{-}1)$$

(See Kakita, 1982 for the detail of the derivation of Equation (16-1)).

Equation (1) implies $h \propto d$ when θ, ϕ = constant. This means that the vertical distance of a point on the object from the mirror is obtained from the distance between the two images

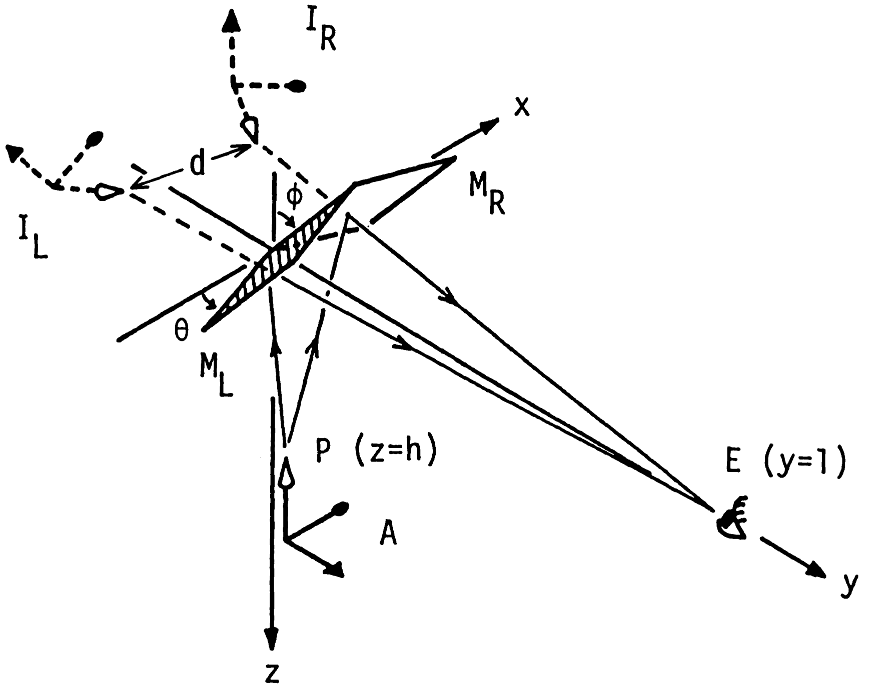


FIGURE 16-3
Schematic drawing of the stereoscopic imaging. (Kakita, 1982; reprinted with permission).

of that point. Since we are interested in the movement of the vocal fold during one vibratory cycle, i.e. about 10 ms for a male voice and 5 ms for female, assumption of θ, $\phi\approx$ const. during the interval would be valid.

To obtain d in actual scale, a scale in mm was filmed immediately after each shot of the larynx, keeping the focus of the lens unchanged. Since the depth of the focus is extremely small for the camera system, the scale can be located at the position of the vocal fold with good accuracy.

Let us talk about the error of h estimated from d. Suppose that $\theta\approx3.5°$ and $\phi\approx45°$, Equation (1) becomes $h\approx5.9d$ or $\Delta h\approx5.9\Delta d$. This indicates that the measurement error for d is magnified by about 6 times. To reduce the error, repeated measurement for the same point is recommended. (See Section 3.2.)

3. RESULTS

3.1. Preliminary data

A stereoscopic high-speed filming of the vocal fold vibration was done for a male adult

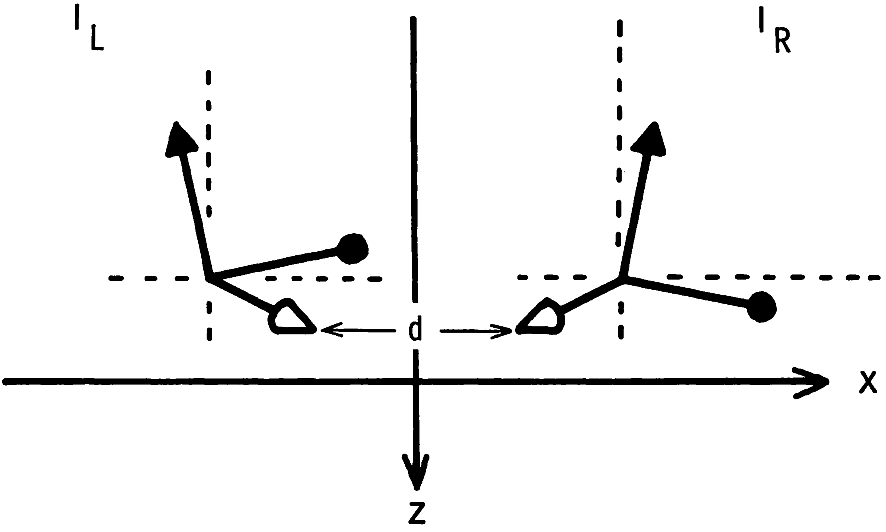


FIGURE 16-4
Schematic drawing of the two images of a 3-D scale photographed on the film. (Kakita, 1982).

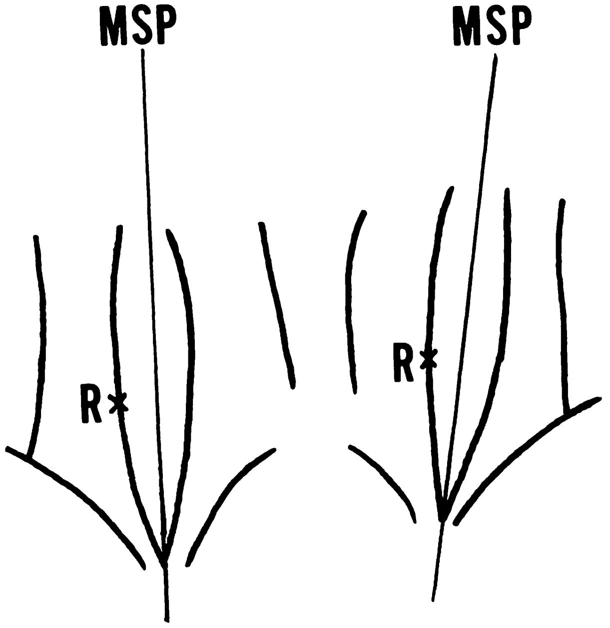


FIGURE 16-5
An example of the stereoscopic images of the vocal folds in one frame of the film. A point marked R on each image was selected for measuring. (Hirano *et al*, 1982).

with normal vocal folds. Filming rate was about 2600 frames/s.

Figure 16-5 shows a schematic drawing of a frame from the motion picture film. A point marked R was selected as a point for measurement, since it was markedly red and also it was close to the vibrating edge of the vocal fold. MSP indicates the "midsagittal plane" and was estimated from the configuration during the closed period of the vocal folds.

Using a film motion analyzer (Vanguard, U.S.A.), (ξ,η) coordinates for each R were measured. Let (ξ_L,η_L) and (ξ_R, η_R) be the coordinate values for the left and the right image, respectively. The horizontal distance d_a between the two images was computed by $d_a = \xi_R - \xi_L$. Strictly speaking, the true distance between the two images $d_a = \{(\xi_R - \xi_L)^2 + (\eta_R - \eta_L)^2\}^{1/2}$ should be obtained since the images show a slight tilting. However, since the tilting angle was 3°, the error between the two d_a's was about 1%. Therefore we chose $d_a = \xi_R - \xi_L$ for the ease of computation.

Actual distance d [mm] is obtained by

$$d = d_a/f \tag{16-2}$$

where f is a conversion constant obtained for the scale measurement mentioned in Section 2.3.

In this particular experiment, $f = 53$, which means the scale 53 on the analyzer corresponds to lmm on the vocal fold measured along the x, or y axis. By estimation, $\theta \approx 3.6°$ and $\phi \approx 51°$ (Details of this estimation can be found in Kakita, 1982.) Using equation (16-2),

$$h\ [\text{mm}] \approx d_a/10.35 \tag{16-3}$$

3.2. Data analysis

To reduce the measurement error of d_a, we repeated the measurement of ξ_L and ξ_R 4 times. By this d_a is expected to be ½ $(=1/\sqrt{4})$ due to the reason described in Section 2.3

Figure 16-6 shows the result. Supposing that ξ_{ij} is the value for *i*th measurement for *j*th frame, $\hat{\xi}_j = \text{median}_i(\xi_{ij})$ was first obtained. Then, for the time series $\hat{\xi}_j$, we obtained $\hat{\bar{\xi}}$ by applying Tukey's 3R smoothing (Tukey, 1977). The same was done for d_a.

Open circles indicate smoothed plots, and the vertical bars indicate errors in quartile deviation. The quartile deviation is the same as the probable error, which is an estimate to show the resolution of a single measurement. Notice that the error for $\hat{\bar{d}}_a$ is comparable to the change in $\hat{\bar{d}}_a$ sequence. Raw value plots of $\hat{d}_a$, before 3R smoothing, are also shown for reference. Actual scale value estimated by Equation (16-1) is shown at the right margin of Figure 16-6.

Smoothed curve of $\hat{\bar{d}}_a$ in Figure 16-6 still shows irregular humps and dents. We therefore used a more powerful smoother, 3RSSH3RSSH3 (See Tukey, 1977). The result is shown in Figure 16-7. We now see that the time course of $\hat{\bar{d}}_a$ roughly shows an upward movement as $\hat{\bar{\xi}}_R$ (or $\hat{\bar{\xi}}_L$) shows an inward movement, and vice versa.

Figure 16-8 shows two dimensional plots for $\hat{\bar{\xi}}_R$ and $\hat{\bar{d}}_a$, which indicate lateral (x) - vertical (z) trajectory of the point R within a frontal plane. The primary component of the trajectory is an elliptical pattern rotating counterclockwise. For the frames 5 through 14, however, the trajectory shows a clockwise pattern. This 8-shape pattern of the trajectory was observed by Baer (1975, 1981), too.

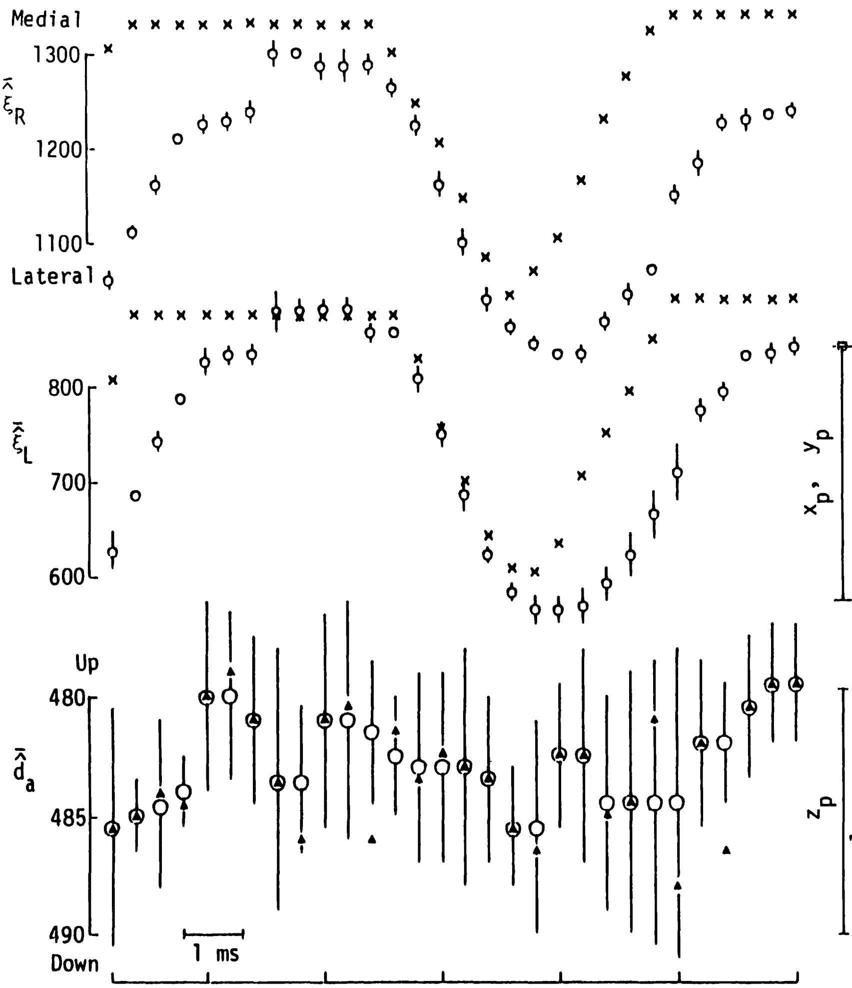


FIGURE 16-6

Three dimensional components of the movement of point R on the vocal fold during vibration. Open circles indicate smoothed plots by 3R (See Tukey, 1977) with vertical bars showing errors in quartile deviation. Crosses indicate the temporal change in the position of the edge of the vocal fold for reference. Filled triangles indicate the raw data of d_a. (Kakita *et al*, 1982).

4. DISCUSSION

4.1 Resolution

From the bar plots shown in Figure 16-6, the mean probable error for d_a was 4.5 accor-

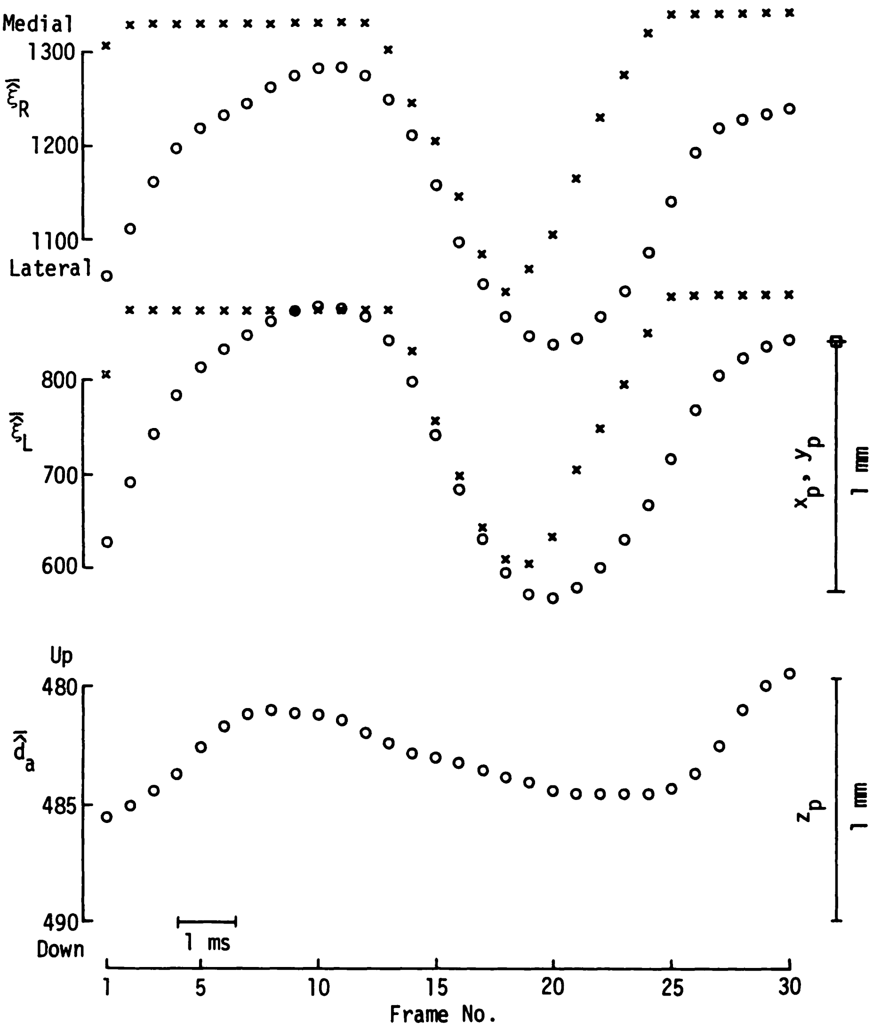


FIGURE 16-7
Plots similar to that in Figure 16-6. Plots are smoothed by 3RSSH3RSSH3.

ding to the analyzer's scale. This means that the resolution of d_a ($=\xi_R - \xi_L$) by a single measurement is estimated to be 4.5.

The blur of the image on a regular 16mm motion picture film is reported to be 20-50 μm (HITACHI, Operation manual) under optimal condition. When we compute the analyzer's scale values, 20 and 50 μm corresponds to 3 and 8, respectively. The resolution of 4.5, therefore, roughly equals the limit diameter of the finest image of a point.

To get a better resolution, repetition of the measurement is effective. This is to reduce the error of the most probable location, the median in this case, of a particular point. In the

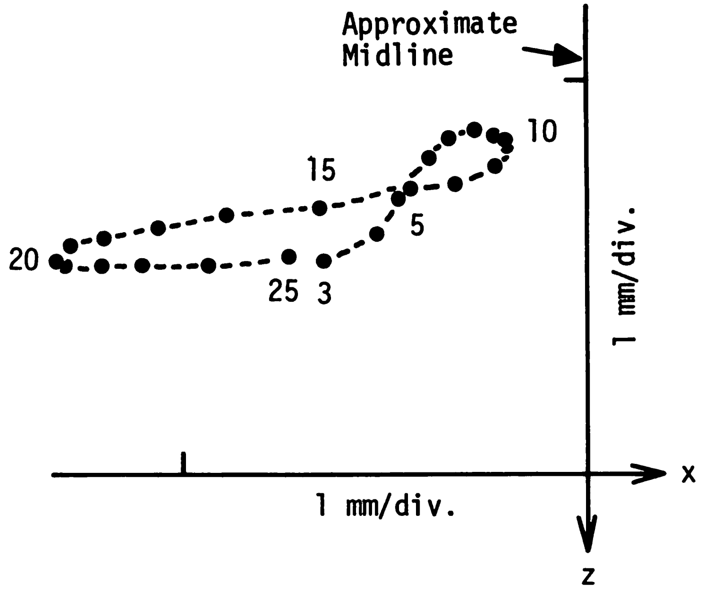


FIGURE 16-8
x-z (lateral-vertical) trajectory for the point R. $\bar{\xi}_R$ and $\bar{d}_a$ in Figure 16-7 corresponds to x and z, respectively.

sense of the least square estimation, $\underline{n}$ times repetition reduces the probable error of the location to $1/\sqrt{n}$. Since we measured 4 times, the error for $\bar{d}_a$ is 2 $(4.5 \times 1/\sqrt{4})$.

In the result shown in Figure 16-6, the ratio of amplitude and error was about 1, for a single measurement. When the improvement by 4 measurements is considered, the ratio of amplitude and the probable error of the median becomes 2.

Theoretically, the larger the number of the repetition, the better the resolution of the location of a point. However, repetition of less than 10 times is realistic. If repeated 9 times, the ratio becomes 3. Provided that other conditions are the same, this would be the best resolution.

4.2 Estimation of actual vertical scale

To estimate the vertical scale in mm, we used Equation (1): $h = d/(4\theta\sin\phi)$. During an experiment, particularly during one vibratory cycle, all the d, θ, ϕ are expected to be unchanged. However, ϕ is not always assured to be 45° in an experiment. We checked this by taking an x-ray photograph of the side-view of the subject separately from the motion picture filming. Examiner and the subject resumed their postures as close as possible to the actual filming condition. We obtained $\phi \approx (45 \pm 5)°$ from the x-ray film. This error for ϕ causes about 10% error for the estimated value of h.

Note that the error of h caused by the error of ϕ in this section should be treated separately from the error caused by the error of d_a discussed in Section 4.1. The former is related to the actual scale value, but the latter to the vertical location of a point relative to the entire

h. Therefore, if we are only interested in the pattern of the movement of a point and its measurement error, the discussion in this section can be skipped.

5. CONCLUSION

Using a specially designed mirror (stereo-laryngo-mirror), we succeeded in filming stereoscopic images of the vibrating vocal folds at a very high filming rate. The vertical location of a point on the vocal fold can be obtained from the distance between the left and the right images of that point in the stereoscopic image. Accuracy of the vertical location of the point was about 0.45 mm, and this is about the same as the amplitude of the vertical movement of that point, i.e. 0.5 mm. By repeating the measurement 4 times for the same images, the accuracy was improved from 0.45 mm to 0.2 mm.

REFERENCES

Baer, T. (1975). "Investigation of phonation using excised larynxes," doctoral diss. MIT.

Baer, T. (1981). "Observation of vocal fold vibration: Measurement of excised larynges," in Stevens, K.N. and Hirano, M. (Eds.) *Vocal Fold Physiology*. Tokyo: Univ. of Tokyo Press.

Fujimura, O., Baer, T. and Niimi, S. (1979). "A stereo-fiberscope with a magnetic interlens bridge for laryngeal observation," *J. Acoust. Soc. Am.* 65/2, 478-480.

Fujimura, O. (1979). "Fiberoptic observation and measurement of vocal fold movement," Proceedings of the Conference on the Assessment of Vocal Patholology *(ASHA Rep.* 11, 1981, 59-69).

Fujita, K. (1938a). "On a device for measuring the larynx in vivo," *J. Otolaryngol. Jpn.* 45A, 936-937. (in Japanese)

Fujita, K. (1938b). "Structure and characteristics of the larynx measuring device we devised," *J. Otolaryngol. Jpn.* 45A, 1524-1525. (in Japanese)

Fujita, K. (1943). "Study on the length of the vocal fold for Japanese," *J. Otolaryngol. Jpn.* 50, 470-492. (in Japanese)

Hirano, M., Yoshida, Y., Matsushita, H. and Nakashima,T. (1974). "An apparatus for ultra-high-speed cinematography of the vocal cords," *Ann. Otol. Rhinol. Laryngol.* 83, 12-18.

Hirano, M., Kakita, Y., Kawasaki, H. and Matsuo, K. (1982). "Stereoscopic ultra high speed photography of the larynx," *J. Otolaryngol. Jpn.* 85, 260-264. (in Japanese)

Himac high speed camera 16HD, Operation Manual, Hitachi Co. (in Japanese) (year unknown).

Honda, K., Hibi, S., Kiritani, S., Niimi, S. and Hirose, H. (1980). "Measurement of the laryngeal structures during phonation by use of a stereoendoscope," *Ann. Bull. RILP.* 14, 73-78.

Kakita, Y. (1982). "Stereo-laryngoscopy: Theory and practice," *Trans. Committee Speech Res.* S82-33, Acoust. Soc. *Jpn.* 257-264. (in Japanese)

Niimi, S. and Fujimura, O. (1976). "Stereo-fiberscopic investigation of the larynx - A new apparatus and preliminary results of testing," a written version of the talk given at the Symposium, "Care of the Professional Voice," Juilliard School of Music, New York.

Saito, S., Fukuda, H., Isogai, Y. and Ono, H. (1981). "X-ray stroboscopy," in Stevens, K.N. and Hirano, M. (Eds.) *Vocal Fold Physiology*. Tokyo: Univ. of Tokyo Press.

Sawashima, M. and Miyazaki, S. (1974). "Stereo-fiberscopic measurement of the larynx: A preliminary experiment by use of ordinary laryngeal fiberscopes," *Ann. Bull. RILP.* 8, 7-12.

Sawashima, M., Hirose, H., Hibi, S., Yoshioka, H., Kawase, N. and Yamada, M. (1981). "Measurements of the vocal fold length by use of stereoendoscope - a preliminary study," *Ann. Bull RILP.*

Tukey, J.W. (1977). *"Exploratory data analysis,"* Reading, Mass: Addison Wesley Publishing Co.

Yoshida, Y., Hirano, M., Matsushita, H. and Nakashima, T. (1972). "Ultra high larnygocine model KU-II," *J. Otolaryngol. Jpn.* 75, 32-38. (in Japanese)

DISCUSSION

O. FUJIMURA: Is the source of random variation of measurements primarily the lack of clear definition of measurement points in the two image fields?

Y. KAKITA: I believe we had a clear enough definition of the particular point which we measured. But generally speaking, your point is true. If there is no visually distinctive marker on the film, we cannot measure the location of any point. The primary reason for the random variation of the point location in the example I have shown, however, was that as I mentioned in my presentation, we are at the limit of the resolution of the 16 mm movie film. The repetitive measurement is one means to improve locating a centre of a point. We could use a high resolution film. But, unfortunately, a high resolution film usually has a low filming rate; so, we have to sacrifice the resolution to some extent.

17 ▪ ELECTROGLOTTOGRAPHY, SPEECH, AND ULTRA-HIGH SPEED CINEMATOGRAPHY

Donald G. Childers, Jayant M. Naik, Jerry N. Larar, Ashok K. Krishnamurthy and G. Paul Moore

ABSTRACT

One of our research goals is to evaluate the feasibility of electroglottography (EGG) for assessing laryngeal function in a research and clinical environment. To this end we have been conducting a series of experiments which simultaneously record EGGs and acoustic phonations while ultra-high speed films are made of the vibratory modes of the vocal cords of various subjects. These three data records are synchronized by a special timing signal. This paper documents our research protocol, data analyses, potential data errors, and provides illustrative results from various experiments.

INTRODUCTION

The electroglottograph has been in use for some time, as reviewed by Lecluse (1977). Fifteen years ago Fant and his colleagues (1966) concluded that the EGG, "when adjusted to have a long time constant, is a valuable instrument for indirect studies of the voice mechanism as well as of laryngeal articulations." However, there has been little work validating the hypotheses concerning the various features of the EGG waveform and their relationship to the vibratory modes of the vocal cords. One validating technique records on film the vibratory motion of the vocal cords while simultaneously and synchronously recording the EGG signal. The only work along these lines, that we are aware of, is that of Lecluse *et al* (1975) and Teaney and Fourcin (1980) who have used stroboscopic photography, which, of course, has the disadvantage that there are typically many vibratory cycles between successive strobe images. We overcome this disadvantage by using ultra-high speed laryngeal films to establish the mechanical vocal cord vibratory basis for the EGG waveform features and characteristics. Since we also simultaneously record the acoustic utterance, we are able to relate the EGG waveform features to features in the speech recording as well, e.g., the instant of glottal closure can be identified in the phonation.

Our group has recorded data from subjects with a normal larynx as well as patients with various laryngeal disorders (Childers *et al*, 1984; Childers *et al*, 1982). We believe our experiments will result in a reliable method for interpreting the EGG waveform in both research laboratories and a clinical setting.

METHODS

The EGG waveform and timing signal were recorded on one stereo tape recorder (Sony, model TC530), while the phonation and timing signal were recorded on another stereo tape recorder (Revox A77 or TEAC A-2060). The subject's utterance was obtained using a hearing aid microphone coupled directly to the tape recorder. The microphone was attached to the mirror handle at the point where the mirror frame joins with the handle. This location was approximately 11 cm from the vocal cords and 7 cm from the teeth, but this distance, of course, varies from subject to subject. The small hearing aid microphone was used at this particular location so that camera motor noise did not intrude into the recordings of the phonation. The audio bandwidth of the microphone has been measured to be approximately 6 kHz with a slight peak at 4 kHz.

Both tape recorders were run at 7½ ips so that a flat frequency response was obtained from approximately 50 Hz to 5 kHz. The EGG signal was provided by one of two electroglottographs (A. Fourcin or D. Teaney, Synchrovoice). Both the speech and EGG waveforms were compensated for phase distortion introduced by the audio tape recorders (Berouti *et al*, 1977). This compensation was necessitated since we only recently obtained access to an FM tape recorder. The Fourcin EGG waveform was similarly computer corrected using a square-wave calibrator provided by Fourcin. Similarly, the Synchrovoice EGG waveform can be computer corrected using a calibrator of our own design.

The photographic equipment and configuration for this study have been described elsewhere (Moore, 1975; Childers, 1977). A Fastax camera model WF-14 was used, which is capable of exposure rates up to 8000 frames per second. Recent film speeds were 5000 frames per second. Illumination was provided by an incandescent lamp having a color temperature of 3200° K. The light passes through two condenser lenses as well as a water cell to remove heat. A plano-convex lens focuses the light onto a plane mirror which turns the converging light beam by 90°, directing it to the laryngeal mirror. With the subject in place, the mirror was located at the back of the pharynx and directed the light 90° downward onto the vocal folds. The image of the folds was reflected by the laryngeal mirror through an opening in the center of the plane mirror and into the camera lens. We also photographed a grid placed in the focal plane of the vocal cords. This allows absolute measures of vocal cord length, displacement, and glottal area, while previously only relative measures could be made.

A second lens protruded from the side of the camera, specifically designed for photographing an oscilloscope face. Two timing signals generated by a time code generator were photographed through this lens. The two traces were positioned on one edge of the film. The EGG waveform was displayed on a third trace of the oscilloscope and was positioned on the other edge of the film. Because the two camera lenses are at a 90° angle the three oscilloscope traces appear on a film frame that is displaced five frames behind the film frame that is recording the position of the vocal cords.

The time code generator was specifically designed to provide a means for establishing temporal alignment between the three synchronous measures, i.e., the speech and EGG waveforms and the film. This generator provided a 5 kHz square wave for one oscilloscope trace and an 8 bit binary number for the other trace. This binary number indicated the number of successive occurrences of blocks of 100 cycles of the 5 kHz square wave. Synchronized with these two waveforms was a 10 kHz square wave which was recorded on the timing channel of the audio tape recorders. All waveforms were initiated simultaneously and synchronously. Thus, we were able to "count" the film frames to a desired location. This was usually done visually. Similarly we were able to count along the audio recordings to the same synchronous location. This was done with a computer program operating on the digitized data.

For comparison purposes the speech and EGG waveforms were sampled in synchrony with the 10 kHz timing signal. These data were aligned and compensation was made for the acoustic propagation delay of the speech waveform relative to the EGG signal. Similarly, we adjusted these data to align their position relative to the position of the vocal cords exposed on each film frame.

Since the EGG waveform was recorded on the film, we may digitize this signal as well.

This redundancy in our data recording allowed an additional verification of our data alignment procedures.

For our experiments the subject, whether possessing a normal voice or a vocal disorder, attempted to phonate the vowel /i/ so that the epiglottis was held out of the optical pathway of the vocal fold image but usually the sound produced approximated an /a/. The recorded phonation was sustained for about three seconds. The subjects were instructed to 1) phonate at a comfortable level and hold the fundamental frequency (Fo) constant, or they were instructed to 2) vary Fo or 3) vary intensity. Both male and female subjects participated in our experiments.

The vocal fold motion was measured from the ultra-high speed laryngeal films using an image processing system which projects a film frame via a television camera on to a high quality gray level computer image terminal. This terminal has graphics overlay capability as well as a joy stick cursor which can be moved under operator control. The position of the cursor is read by the computer. We measure the glottal area, the length of vocal cord contact from anterior to posterior, the width of the glottis at various locations, etc.

The three data waveforms, namely, glottal area (absolute measure), EGG, and speech were plotted in synchrony on the face of one of our graphics terminals. We may photograph or make copies of these waveforms. Since these data are synchronized with the ultra-high speed laryngeal films, we may then match a film frame with the corresponding EGG and speech data points.

Since the EGG and utterance were recorded on separate recorders, these signals must be digitized separately using the timing signal recorded on the second channel of each tape. There exists the possibility of a slight displacement in the sampling of these two waveforms since the external timing signal on one tape, which activates the analog-to-digital converter computer circuitry, may differ slightly in amplitude from the timing signal on the other tape. Further, this error may cause a displacement between these wave-forms and the glottal area measured from the laryngeal films. To partially compensate for this error we used the digitized EGG trace taken from the films to validate the proper alignment between the three waveforms, namely, glottal area, EGG, and speech. The EGG recorded on the films was in exact synchrony with the glottal area, except for the known aforementioned five film frame displacement. In the future, a four channel FM tape recorder will be used for data collection and many of the above problems will be eliminated or greatly reduced.

DATA PROCESSING

A number of special signal processing programs have been developed to smooth and filter the data. These include the following:

1. Non-linear median filtering (smoothing) of the glottal area curve (Tukey, 1974; Rabiner *et al*, 1975).
2. Trend removal and 60 Hz filtering of EGG and speech data.

The median filter smooths sharp, point-like jumps in the glottal area curve. These are caused by improper operation of the image digitization equipment, as well as by the limited resolution of the equipment.

Trend removal and 60 Hz filtering are performed to alleviate obvious noise and errors which arise in the data collection and digitization process.

The aforementionned tape distortion correction of the speech signal is necessary if proper inverse filtering results are to be obtained. Such a correction is also applied to the EGG signal to correct low frequency distortion.

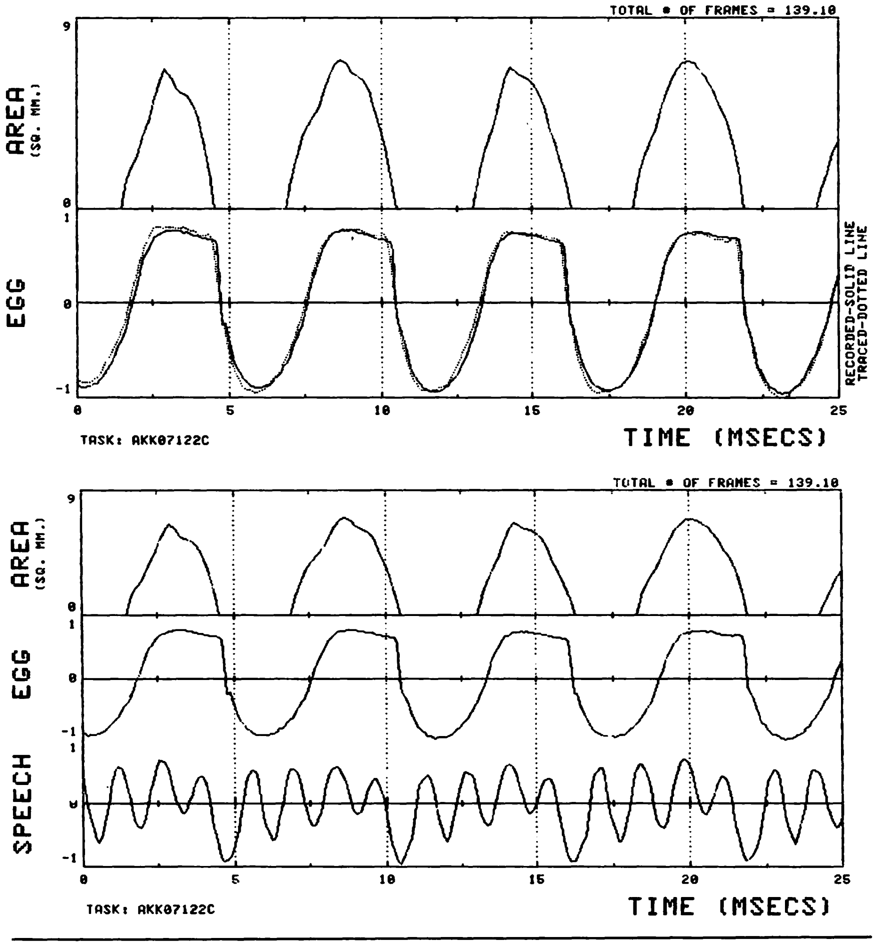


FIGURE 17-1
Upper graphs: glottal area, tape recorded EGG (solid line), EGG traced from film (dotted line). Lower graphs: glottal area, EGG, and speech. These data are synchronized and aligned. The speech waveform has been advanced 0.4 ms (4 data samples).

Figures 17-3 and 17-4 show the EGG, differentiated EGG, and glottal area for the same subjects and phonation tasks. The differentiated EGG waveform is useful for detecting the points of glottal opening and closure, although this waveform does not yield consistent results. Our observations have led us to conclude that glottal closure may be predicted from the differentiated EGG waveform as the instant at which this waveform has its absolute minimum during a pitch period. The instant of glottal opening is more variable, but seems to occur the instant at which the differentiated EGG waveform has its absolute maximum during a pitch period. We show the reasons for this conclusion in Figure 17-6, which we discuss after presenting Figure 17-5.

SYNCHRONIZED EGG, SPEECH AND ULTRA-HIGH SPEED FILM DATA AND OBSERVATIONS

Our synchronized data base now exceeds 80 films and recordings. One experiment in this series is a set of 36 films with associated synchronized EGG and speech recordings. This experiment called for four male subjects to perform nine tasks each. These tasks consisted of three variations in fundamental frequency, Fo, (125, 170, and 340 Hz) and three variations in intensity for each Fo. The intensity was measured by a sound dB meter while the subjects matched their Fo to a sinewave function generator they listened to via a set of headphones.

Figures 17-1 and 17-2 are representative of our preliminary data for four subjects (JMN, AKK, GPM, DMK) showing absolute area measurements superimposed with the synchronized EGG and speech waveforms.

The upper set of graphs in Figure 17-1 shows the glottal area function as measured from the ultra-high speed films. The film frame scale is 139.10 so that each tic mark corresponds to 13.9 frames. This odd scale arises because we have chosen to represent our data on a millisecond time basis. The lower two EGG graphs in the upper portion of Figure 17-1 show the corrected EGG obtained from the tape recording (solid line) and the EGG measured from the ultra-high speed laryngeal films (dotted line). These two curves are nearly identical. The lower set of three graphs in Figure 17-1 shows the glottal area, the EGG waveform, and the speech signal. These three curves are aligned and corrected. The alignment procedure includes compensation for the acoustic propagation delay as well.

On the films one can see that the initial point of vocal fold contact corresponds to the "knee" or "break" in the negative slope of the EGG waveform and the EGG waveform minimum corresponds to maximum glottal closure. Some of our films show the vocal folds making initial contact along the lower margin. Then vocal fold contact increases upward toward the upper margin as the cords "roll" upward. Vocal fold contact also increases along the midsagittal plane as glottal closure is occurring. Thus, vocal fold contact is increasing simultaneously in two dimensions, longitudinally and sagittally. As the cords open we assume that the vocal folds start to separate first along the lower margin and then finally at the upper margin. But this has not yet been visually confirmed. One experiment, already performed, has called for the subjects to phonate during inhalation, which shows the vocal cords making contact initially at the upper margin progressing toward the lower margin and then separating initially at the upper margin, i.e., the opposite to that observed during normal exhalation phonation.

From Figures 17-1 and 17-2 one can see that glottal closure (minimum glottal area) corresponds with the minimum in the EGG waveform. Because the minimum in the glottal area waveform is rather broad, the temporal alignment of the film frames and the EGG signal appears precise when, in fact, there could be a displacement of several film frames.

From these two figures one may conclude that the EGG characteristic that indicates glottal closure is the initiation of the sharp downward break in the EGG waveform or possibly the instant at which the EGG waveform crosses the zero axis on its positive to negative swing. This interpretation appears valid, as we will discuss later, except when glottal closure does not take place, as in the high Fo tasks. In these cases the minimum in the glottal area function corresponds with the EGG minimum.

For the microphone we used and its polarity output, the instant of glottal closure may be estimated from the speech waveform as the largest negative peak. If glottal closure does not take place then the minimum glottal area corresponds to the largest negative peak in the speech waveform.

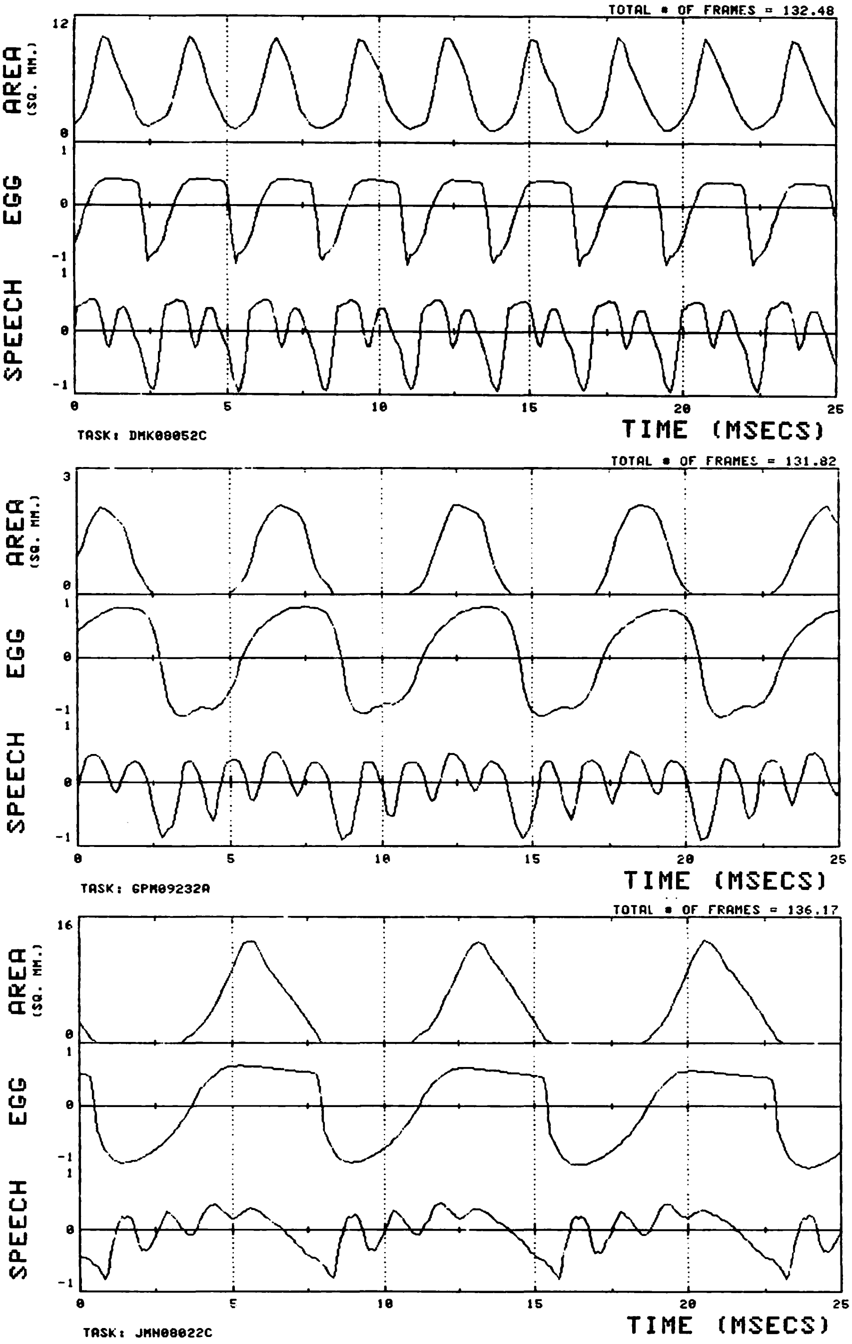


FIGURE 17-2
Glottal area, taped EGG, and speech for three additional subjects. Same corrections as in Figure 17-1.

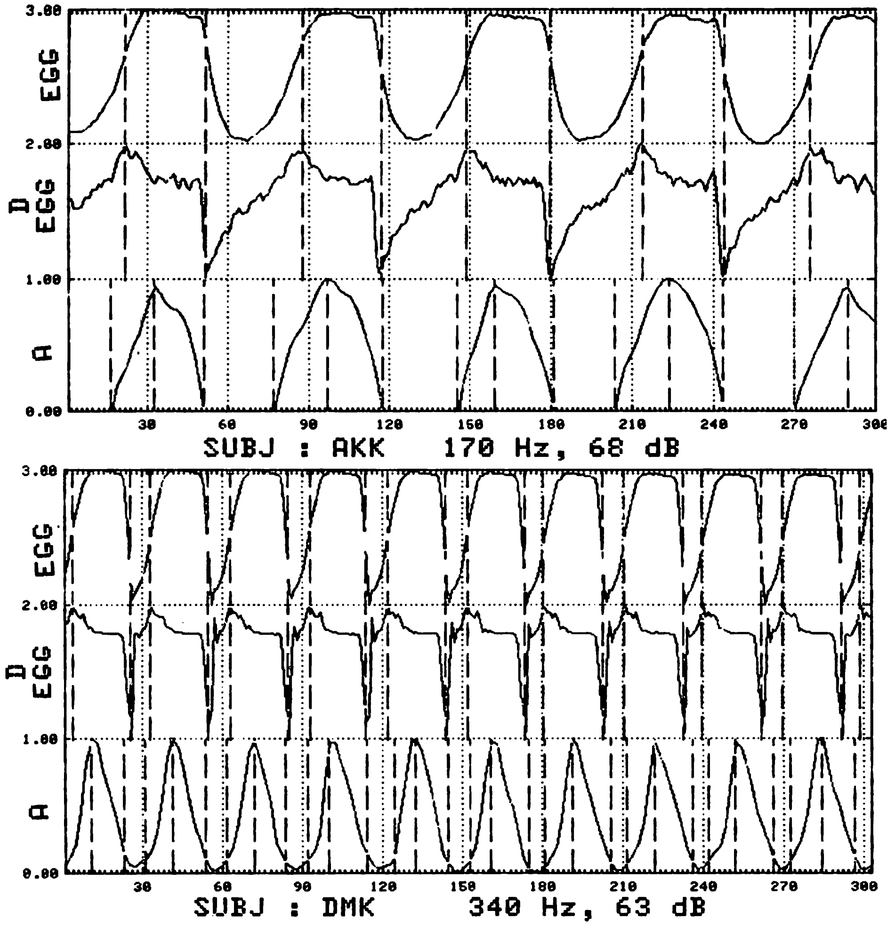


FIGURE 17-3
EGG, differentiated EGG (D EGG), and glottal area (A) for two of the subjects shown in Figures 17-1 and 17-2.

Figure 17-5 illustrates the EGG waveforms for one subject, with Fo fixed, but three different utterance intensities. We believe the EGG "rate of closure" is becoming steeper as the intensity is increased. We have tried to quantify this by determining the spectra for these three waveforms, which are shown. While there is a tendency for the spectra to show high frequency content as the intensity increases, the tendency is not as marked as we had expected.

We have also done some work on estimating the glottal volume-velocity waveform by inverse filtering the speech signal. Here we use a 12 pole linear prediction (LP) model. An illustrative result is depicted in Figure 17-6 along with the glottal area, the EGG, differentiated EGG, and differentiated volume-velocity.

In Figure 17-6 we also show a similar set of data obtained by simultaneously digitizing the EGG and speech while a subject phonated an /a/. These two signals were fed directly

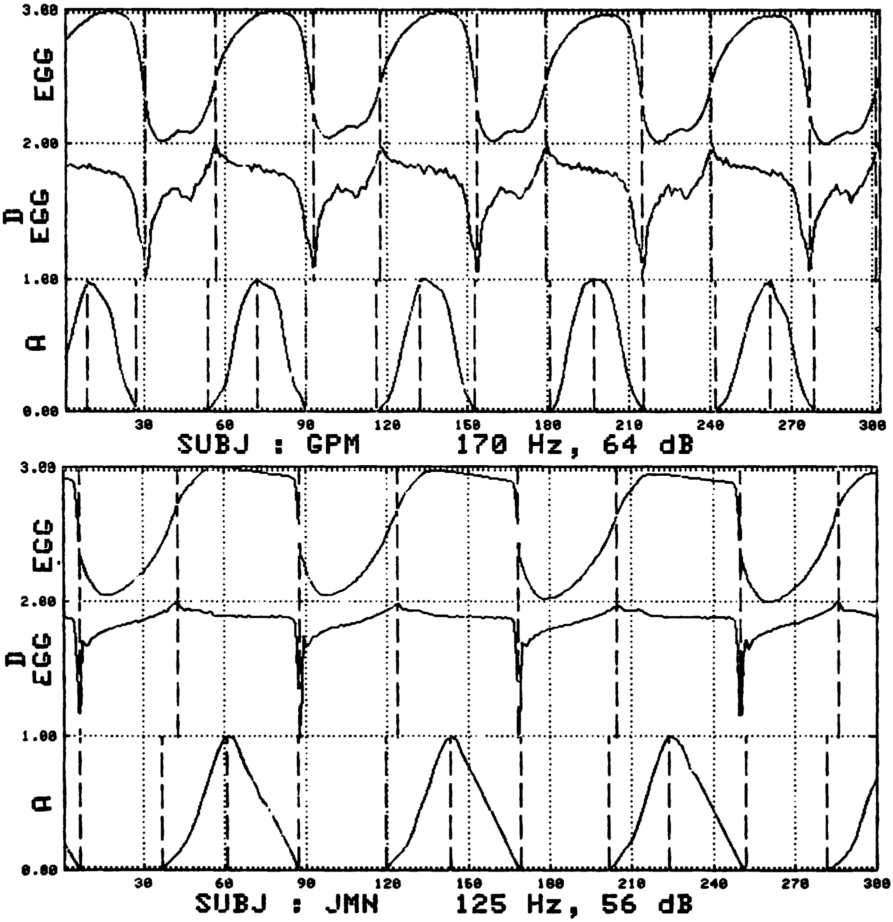


FIGURE 17-4
EGG, differentiated EGG (D EGG) and glottal area (A) for the two other subjects shown in Figure 17-2.

to the analog-to-digital converter, thus eliminating tape recorder distortion and any timing errors. We then inverse filtered the utterance to obtain the volume-velocity. A 12 pole LP model was used. We show the EGG, volume-velocity and their respective derivatives.

Note that in both sets of graphs shown in Figure 17-6 that for a given pitch period the absolute maximum of the differentiated EGG corresponds to the glottal opening, while the absolute minimum aligns with the instant of glottal closure. Further, glottal opening and closure are easily determined using waveform features drawn from the differentiated volume-velocity. The glottal area function was not available for the directly digitized experiment.

Some investigators have compared the spectra of various excitation waveforms, such as glottal area and volume-velocity. Figure 17-7 shows the spectra for the EGG, glottal area, and volume-velocity depicted in the upper graph in Figure 17-6. Note that the EGG spec-

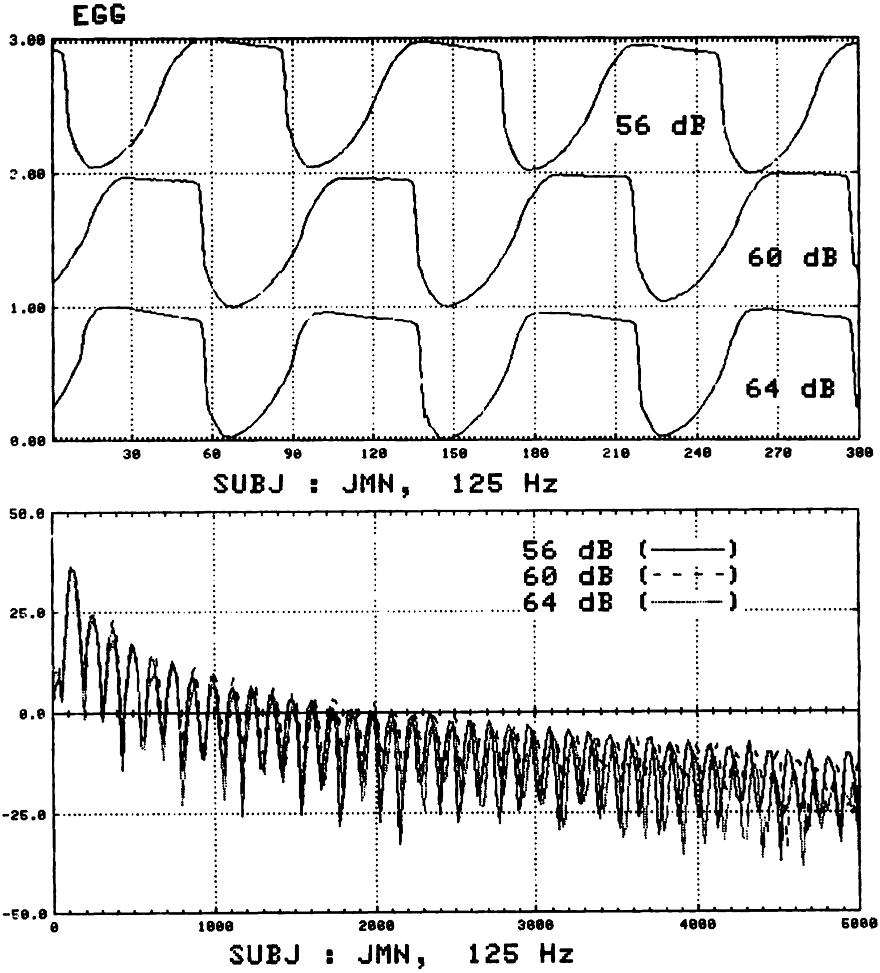


FIGURE 17-5
EGG and respective spectra for one subject phonating at three different intensities.

trum has the most high frequency content, followed by the spectrum of the volume-velocity, and finally by the glottal area spectrum.

Figure 17-8 is an example of using the EGG for voiced/silence-unvoiced phonation classification (Yea *et al*, 1983). This use of the EGG is compared to a similar speech based algorithm (Yea *et al*, 1983). The figure shows that the EGG waveform indicates laryngeal activity one pitch period before the speech signal reflects such motion. If this proves to be a consistent feature, then this will be useful for real-time algorithms.

In Figure 17-8 the pitch contour is estimated from the EGG waveform by measuring the interval between successive positive going zero crossings of the signal. These intervals are averaged for the analysis frame (30 ms) to give the average pitch period. The speech based pitch contour algorithm shown uses a forward auto-correlation of center clipped speech, but

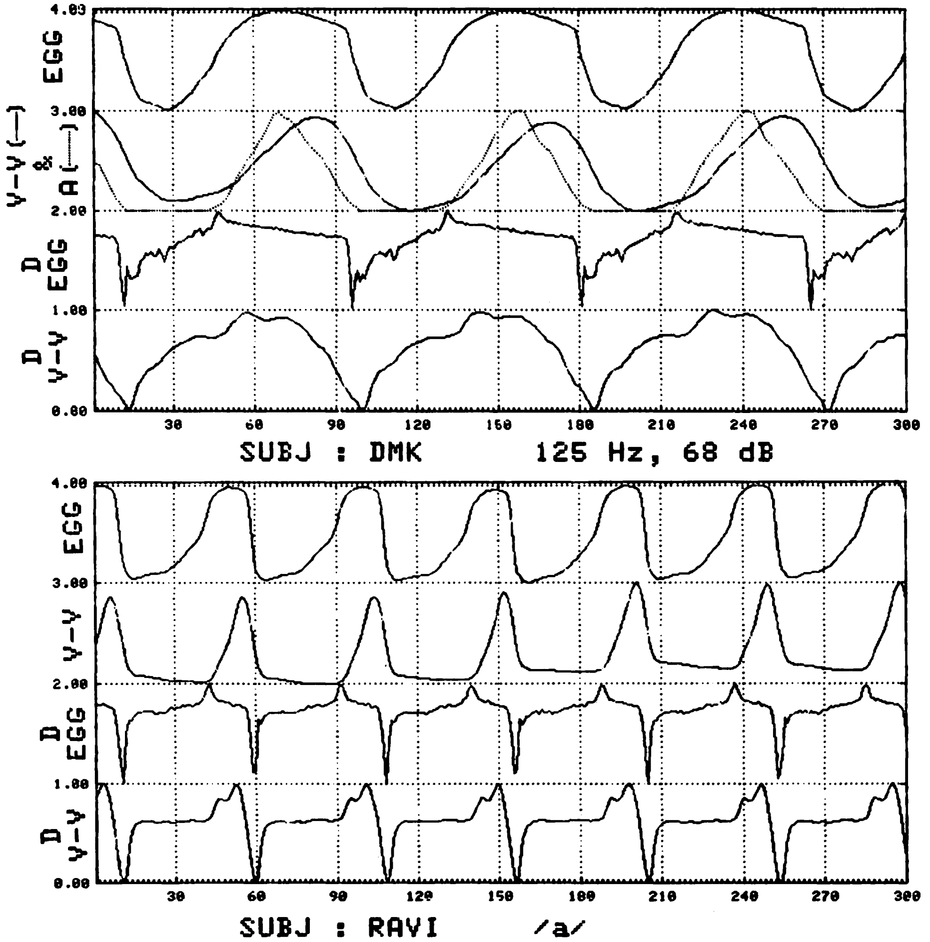


FIGURE 17-6

EGG, volume-velocity (V-V), glottal area (A), differentiated EGG (D Egg), and differentiated volume-velocity (D V-V). The upper graph is for one of our filmed subjects. The lower graph is data directly digitized from another subject. The D V-V minimum is aligned with the instant of glottal closure measured from the glottal area. The V-V is similarly aligned.

is supplemented by zero crossing calculations to improve the detection of mixed excitation intervals. A more sophisticated algorithm gives similar results (Yea *et al*, 1983).

Observations made from our films have convinced us that the presence of mucus on the vocal cords can lead to the misinterpretation of the vocal cord vibratory events. The mucus may be the last layer to separate as the cords are opening. Such strands of mucus introduce noticeable changes in the glottal length waveform, which is measured as the length of the glottal opening. One example is shown in Figure 17-9, where the glottal length has a "bend" in the rising portion of the waveform. The cords are open at the beginning of the bend, but

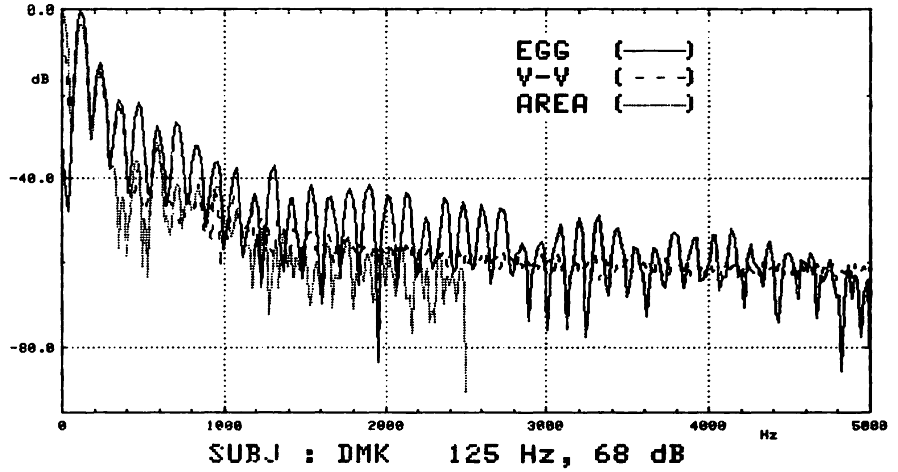


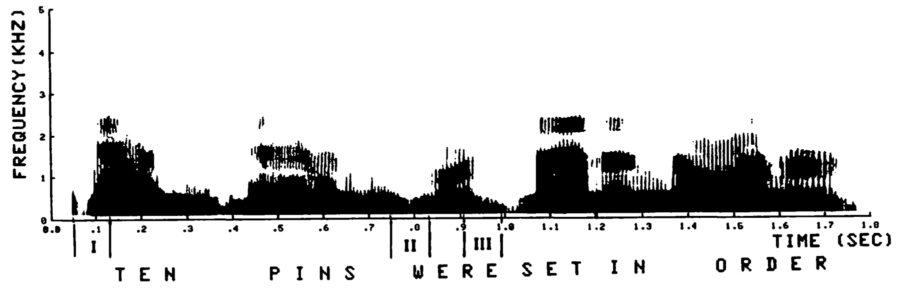
FIGURE 17-7
Spectra for the EGG, V-V, and glottal area shown for the subject in the upper graph of Figure 17-6.

a strand of mucus bridges the cords. When this strand breaks the length waveform rises sharply to its characteristic flat top. While the EGG waveform appears normal for this case, the differentiated EGG does not have its maximum at the instant of glottal opening, rather, the maximum occurs just as the mucus strand begins to break. This point corresponds to the "knee" in the opening phase of the glottal area waveform. The EGG waveform reaches its maximum just as the mucus strand has broken. Thus, it appears that the EGG waveform does not always represent glottal contact area, since a strand of mucus may provide a highly conductive current path for the EGG. This may lead to misinterpretation of the data. We illustrate this case with an ultra high speed laryngeal film.

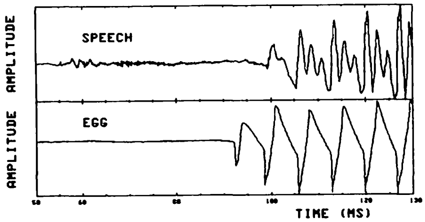
The film shows that the EGG indication of glottal opening is delayed due to the mucus bridging the vocal folds. As the mucus strand breaks the EGG rises sharply to its characteristic flat top. Actually the cords were open about seven frames earlier, but the mucus apparently provides a conductive path, which leads to a false indication of vocal fold opening.

CONCLUSIONS

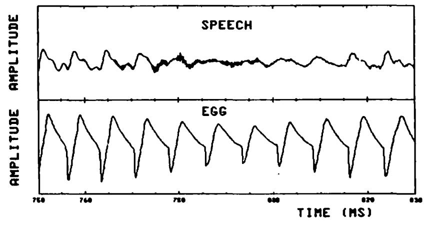
Our data tend to show that 1) the rising portion of the EGG waveform corresponds to the glottal opening phase, 2) the maximum of the EGG waveform coincides with the maximum glottal opening, 3) the declining portion of the EGG waveform corresponds to the glottal closing phase, 4) the "knee" or "break" in the negative slope of the EGG waveform corresponds to initial vocal fold contact on the closing phase, 5) glottal closure or the point of minimum glottal area coincides with the minimum of the EGG waveform, and 6) the positive going axis crossing of the EGG corresponds with initial glottal opening. These observations agree with a model proposed by Rothenberg (Moore, 1975). Figure 17-10 is our summary of this model.



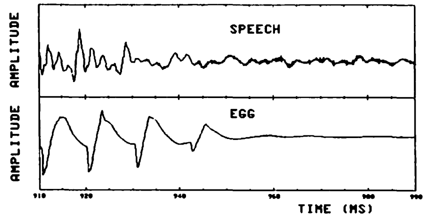
SPECTROGRAM OF SENTENCE 'TEN PINS WERE SET IN ORDER'.



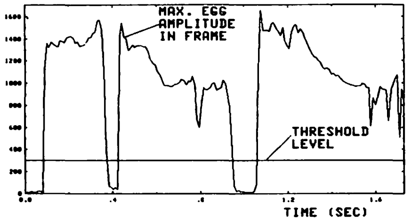
VOICING ONSET, SECTION I



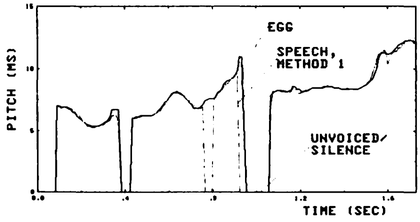
MIXED EXCITATION, SECTION II



VOICING DECAY, SECTION III



EGG AMPLITUDE THRESHOLDING FOR VOICED/UNVOICED-SILENCE DECISION.



MEASURED PITCH CONTOUR FROM EGG AND SPEECH(METHOD 1).

FIGURE 17-8

Spectrogram, speech and EGG waveforms. Voiced/unvoiced-silence detection using EGG amplitude thresholding. Pitch detection using EGG and the speech signal.

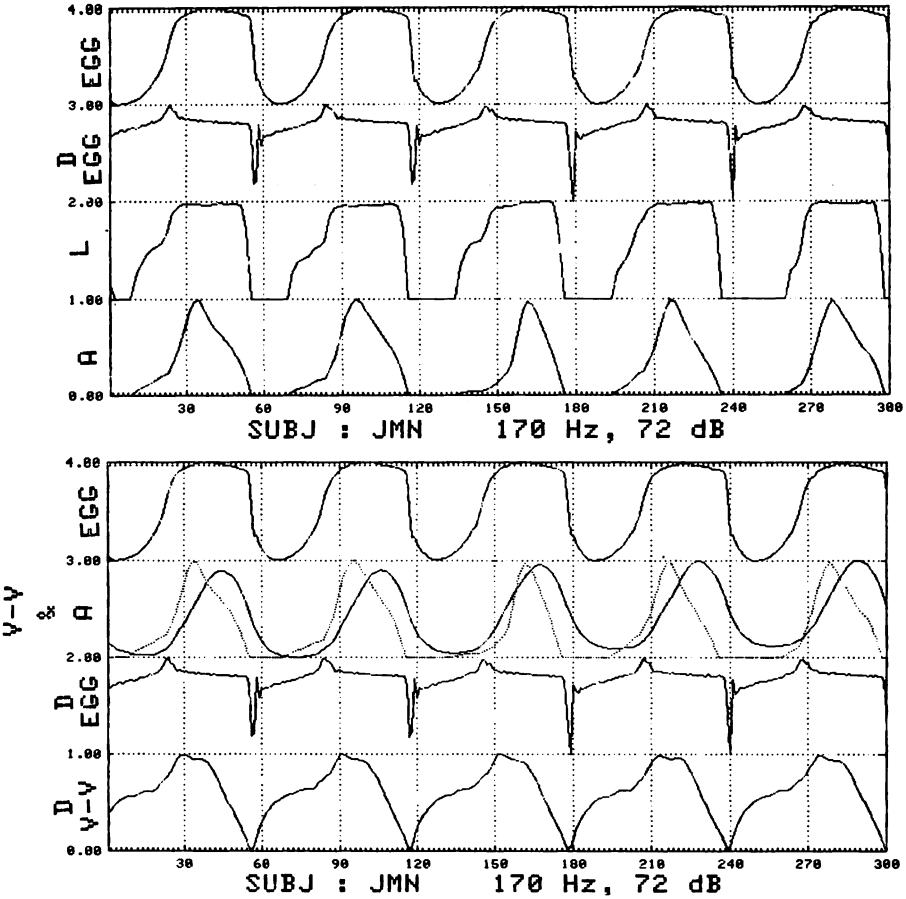


FIGURE 17-9
EGG, differentiated EGG, glottal opening length (L), glottal area, V-V and differentiated V-V for one subject.

ADDENDUM

As this paper was being typed Baer *et al* (1983) published their recent results. They compare information on laryngeal vibrations obtained from high-speed filming, photoglottography (PGG), and electroglottography (EGG). Analyzing data from two subjects, they report that PGG and film measurements give essentially the same information for peak glottal opening and glottal closure. They believe the EGG signal appears to reliabily indicate vocal fold contact. Taken together, these authors feel that the PGG and EGG may provide much of the information obtained from high-speed filming as well as potentially detect horizontal phase differences during opening and closing.

There appears to be general agreement amongst the various investigators, that there is more uncertainty about the moment of glottal opening than about closure. (See our earlier figures and Baer *et al*, 1983.)

A common assumption is that the EGG is an indicator of lateral glottal contact area (Baer *et al*, 1983). But there is scanty direct evidence for this conclusion. We have noted that the EGG is often strongly correlated with the length of glottal contact (or opening) as measured from the ultra-high speed films (Childers *et al*). But we are still uncertain as to whether the EGG reflects the lateral area of vocal fold contact. It may be that during vibration the vocal folds are dynamically coming into contact along the midsagittal plane such that the depth of contact is relatively constant once the folds have closed. That is, as the folds come into initial contact at the lower margin, their rolling action causes them to part along the lower margin as they roll into contact upward toward the upper margin. During this action, the folds are also closing lengthwise (often in a zipper-like fashion). The lengthwise contact may vary more between subjects than the depthwise contact.

If it is assumed that the EGG is a measure of electrical impedance across the glottis, then observations of vocal fold contact and EGG measurements would lead one to conclude that the impedance should be smaller at closure than at opening because it is commonly believed that more tissue is in contact at closure than during opening (Baer, 1981). Figure 1 in (Baer *et al*, 1983) supports this conclusion, while Figure 2 in (Baer *et al*, 1983) does not. Additional evidence against the above hypothesis is presented in our Figure 17-11, i.e., the impedance at closure is larger than at opening. One explanation for this phenomenon is the free mucus bridging the folds, as in the example explained earlier. Consequently, we recommend caution in interpreting the EGG as a measure of lateral glottal contact area.

We also must remind the reader that the EGG can give a waveform which appears to indicate that vocal fold contact is taking place, when in fact it is not, as in high intensity, high fundamental frequency phonations.

ACKNOWLEDGEMENT

This work was supported in part by Grant NIH NS17078 from the National Institute of Health, Grant NSF ECS-8116341 from the National Science Foundation and the University of Florida Center of Excellence Program for Information Transfer and Processing. We also have received support previously from NIH Biomedical Grant 5S07RR07021. We thank Yacob Alsaka for his help on a number of occasions.

REFERENCES

Baer,T. (1981). "Investigation of the phonatory mechanism," *Proc. of Conf. on the Assessment of Vocal Pathol.*, Chap. 4, C.L. Ludlow and M. Hart (Eds.) ASHA Rep. 11, 38-47.

Baer, T., Lofquist, A. and McGarr, N.S. (1983). "Laryngeal vibrations: a comparison between high-speed filming and glottographic techniques," *J. Acoust. Soc. Am.* 73/4, 1304-1308.

Berouti, M.G., Childers, D.G. and Paige, A. (1977). "A correction of tape recorder distortion," *Proc. IEEE on Acoust., Speech and Sig. Proc.*, 397-400.

Childers, D.G., Smith, A.M. and Moore, G.P. (1984). "Relationships between electroglottograph, speech, and vocal cord contact," *Folia phoniat.*, 36, 105-118.

Childers, D.G., Moore, G.P., Naik, J.M., Larar, J.N. and Krishnamurthy, A.K. (1982). "Assessment of laryngeal function by simultaneous, synchronized measurement of

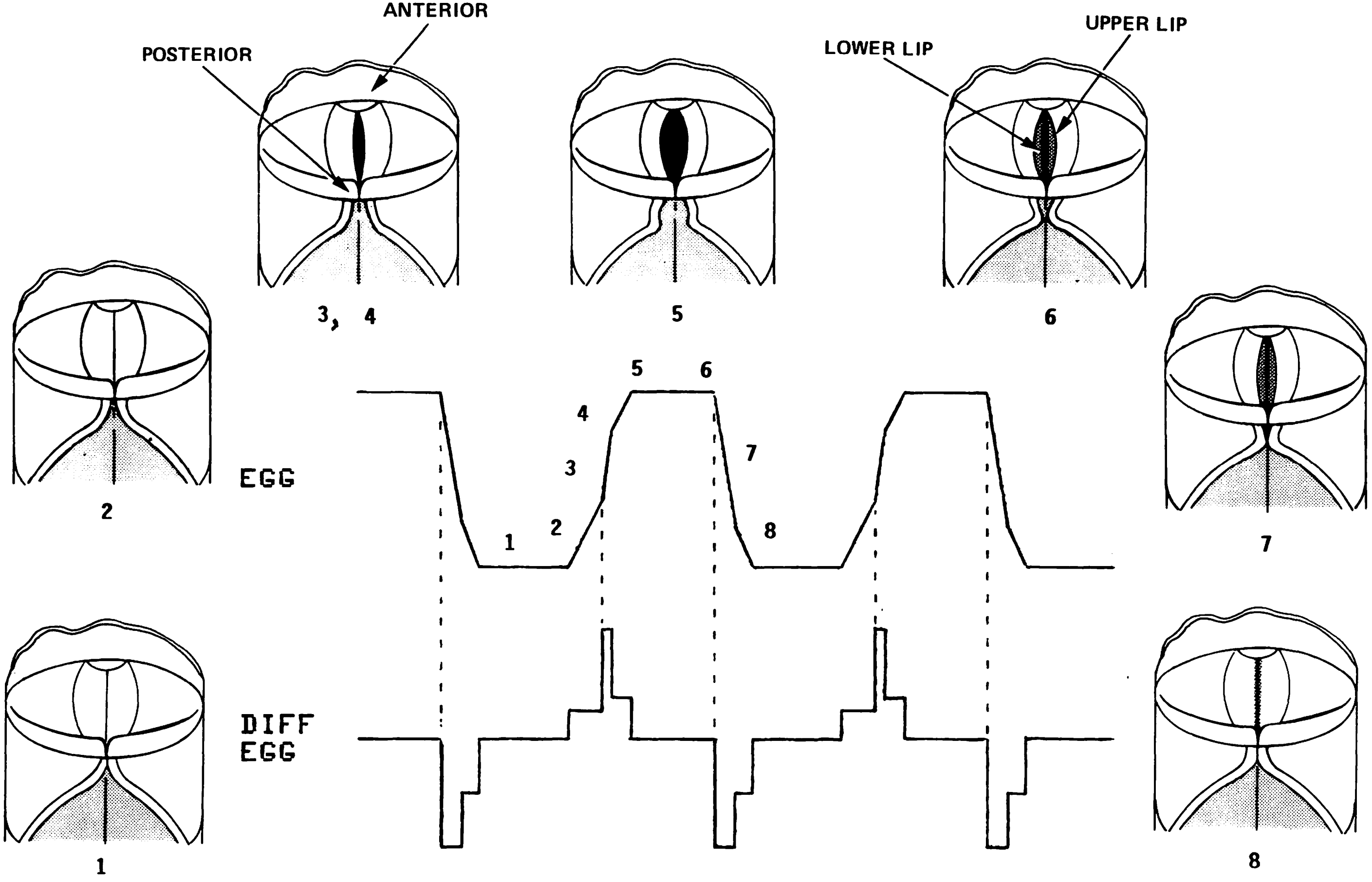

POSTERIOR
ANTERIOR
LOWER LIP
UPPER LIP
3, 4
5
6
2
7
1
8
EGG
DIFF EGG
1
2
3
4
5
6
7
8

FIGURE 17-10
1. Vocal folds maximally closed. Complete closure may not be obtained. Flat portion idealized. 2. Folds parting along lower margin (LIP) toward upper margin. 3. Folds opening along upper margin, usually opening is also from posterior to anterior. 4. Folds continuing to open. 5. Folds apart, no lateral contact. 6. Folds starting to close, lower margin first progressing anterior to posterior, but still no contact. 7. Folds making first lateral contact along lower margin and at anterior. 8. Folds completing closure with increasing lateral contact.

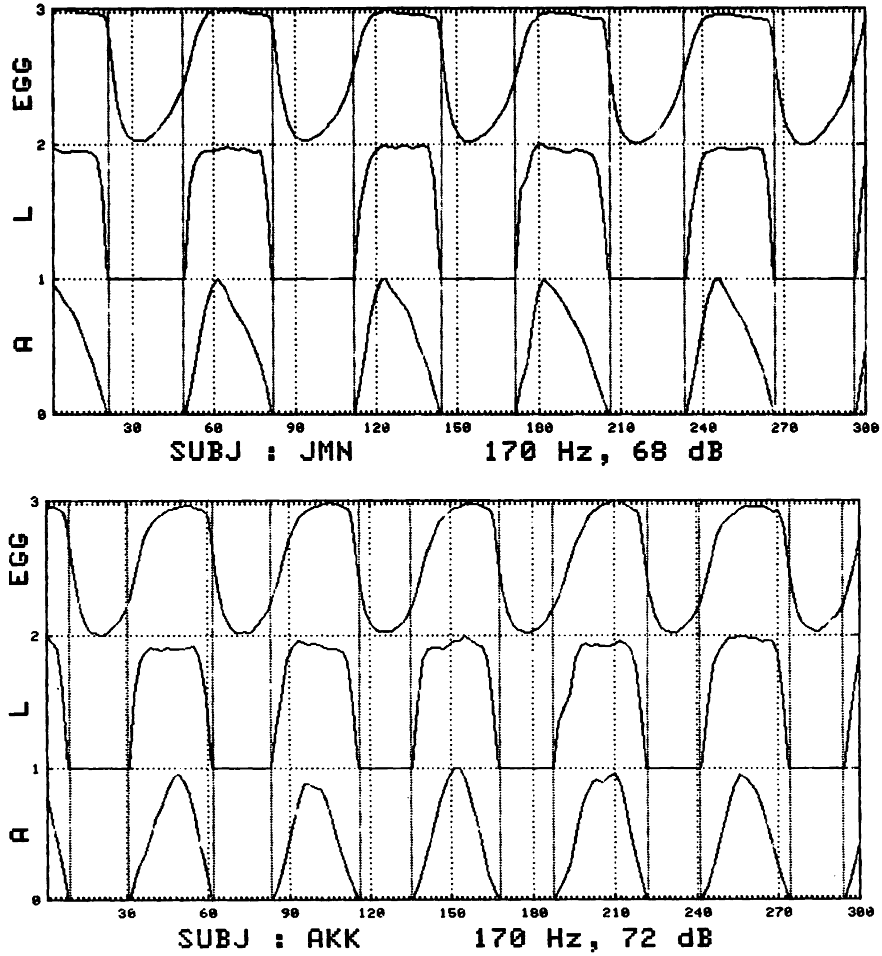


FIGURE 17-11
EGG, glottal opening length (L), and glottal area for two subjects.

speech, electroglottography, and ultra-high speed film," *Eleventh Symposium: Care of the Professional Voice*, the Juilliard School.

Childers, D.G. (1977). "Laryngeal pathology detection," *Critical Reviews in Bio-Engineering* 2, 375-426.

Fant, G., Ondrackova, J., Lindquist, J. and Sonesson, B. (1966). "Electrical glottography," *STL/QPSR-4*, 13-25.

Lecluse, F.L.E., Brocaar, M.P. and Verschuure, J. (1975). "The electroglottograph and its relation to glottal activity," *Folia phoniat.* 27, 215-224.

Lecluse, F.L.E. (1977). *Elektroglottografie. Drukkerij Elinkiwijk*, B.V., Utrecht, The Netherlands, 181.

Moore, G.P. (1975). "High speed photography in laryngeal research," *Can. J. of Otolaryngol.* 4, 793-799.

Rabiner, L.R., Sambur, M.R. and Schmidt, C.E. (1975). "Applications of a nonlinear smoothing algorithm to speech processing," *IEEE Trans. on ASSP*, ASSP-23/6, 552-557.

Rothenberg, M. (1981). "Some relationships between glottal air flow and vocal fold contact area," *Proc. of the Conf. on the Assessment of Vocal Pathol.*, Chap. 8, ASHA Rep. 11, 88-96.

Teaney, D. and Fourcin, A.J. (1980). "The electrolaryngograph as a clinical tool for the observation and analysis of vocal fold vibration," The Voice Foundation.

Tukey, J.W. (1974). "Nonlinear (nonsuperposable) methods for smoothing data," in *Conference Record*, EASCON, 673.

Yea, J.J., Krishnamurthy, A.K., Naik, J.M., Moore, G.P. and Childers, D.G. (1983). "Glottal sensing for speech analysis and synthesis," *Proc. International Conf. on Acoust. Speech and Sig. Proc.*, Boston, 3, 1332-1335.

DISCUSSION

R. SCHERER: Do mucus strands across the glottis have a great effect on the EGG waveform?

D. CHILDERS: I'd say it is influential when interpreting the EGG with regard to glottal opening or closing. I feel that it is a rather high conductive path for the current. The shape of the EGG waveform still looks normal, it's just that the rise time and fall time are changed. The slope is what is crucial.

R. SCHERER: If you have a relatively large amount of mucus over the glottis, does that affect the airflow?

D. CHILDERS: Yes. It may impede air flow or cause it to be re-routed. We do not see particles of mucus fly about. The mucus strands simply break as the vocal cords move apart. But the width of the mucus strand can vary from pitch cycle to pitch cycle as our Figure 9 (length) shows.

G. FANT: Congratulations on a beautiful study. I think this is rewarding to see all these different measures aligned at the same time. I remember Paul Moore's early publications on measuring the area functions of the waveform. Paul Moore measured just the distance across the cords. How did you measure the area function? Did you have proper integration? And if so, what kind of accuracy did you have?

D. CHILDERS: We measure the glottal area in two ways. The first method uses a computer algorithm to trace out the boundary of the glottis. This is a contour of many discrete points. The computer program measures the length of the straight line which joins the point on the edge of the left vocal cord to the corresponding point of the glottis on the edge of the right vocal cord. We add up the lengths of all such lines and multiply by the length of the glottal opening. The second method is fundamentally the same, but only measures the glottal width at five or six points, locates the anterior and posterior glottal end points and uses straight line interpolation to connect the points along each glottal edge. This gives two triangles (one at the anterior and one at the posterior end) and a series of trapezoids in between. Since we know the lengths of all edges on the triangles and the trapezoids we can calculate the glottal area.

The accuracy has been estimated by calculating the area of known geometric shapes which we have digitized and processed on the image digitization equipment. These measurements were very accurate (1% or less) but they were high contrast images with no "fuzzy" edges. The glottal area measurements have a larger error, but we cannot be sure what the exact value would be. My estimate is 5%.

We list here several publications which document our measurement techniques and the accuracy.

A. Krishnamurth and D.G. Childers (1981). "Vocal fold vibratory patterns, comparison of film and inverse filtering," Proc. Internat'l Conf. on Acoustics, Speech, and Signal Proc., 133-136.

J.R. Booth and D.G. Childers (1979). "Automated analysis of ultra high-speed laryngeal films," IEEE Trans. on Biomed. Eng. BME-26, 185-192.

T. BAER: I'd like to agree, it's beautiful data. I would just like to correct what might have been a slight misinterpretation of what we were saying in our paper. Its very difficult to speak of an instant at which you say opening or closing has occurred. I think what we were saying is that the interval between when the curve first deflects and when it reaches a point that corresponds with complete glottal closure, is more clearly defined than the change that corresponds with glottal opening. Opening occurs over a smaller vertical extent. So, just prior to opening, the vertical extent is much slower and it is probably associated with greater horizontal phase differences than closing. I think the closing is almost always much more nearly the same across the longitudinal extent, while that's rarely true during opening.

D. CHILDERS: I would agree. So if my comments appeared to conflict with that, it's just a matter of words.

T. BAER: One additional point that seems to me a little unclear about opening is concerned with the mucus that is bridging the gap and mucus that is considered to be part of the folds.

D. CHILDERS: The only point I wanted to make with that was about possible misinterpretation of the electroglottographic signal. Clearly you can define the measurement for glottal area to be one that ignores that the mucus is there, or you can include the mucus. It's a matter of how you want to define it.

T. BAER: Right, but I guess if the mucus is occluding the folds, that really should affect the flow and should be equivalent to vocal fold contact.

D. CHILDERS: Right, and just a slight comment back to yours. You'll notice in the plot that the knee in the length function varies as you go down, so that the mucus, as it's bridging the vocal folds, is dynamically varying during the film. At the beginning, it looks like just a small portion. Later there's more. And then again, it goes back to a small portion again. So you can see that in just one cycle there's quite a bit of variation.

M. ROTHENBERG: I also agree that this study is very carefully done and unambiguous. I have a few small comments. One is, we've been using the electroglottograph and air flow for clinical measurements, and I agree about misinterpreting it. For example, if you're looking at perturbations in the EGG waveform as evidence of a polyp or something like that, you can easily make a mistake, and think that there's a mass there when there is really just the mucus. That's important. I'm glad that you point that out. Also, you looked at large masses of mucus, but then that led me to think, what about the thin layers of mucus that are on the cords themselves. You never or rarely see a jump in the waveform at the instant of glottal opening. I was thinking that if there was mucus all along the vocal folds, you probably can't see that probably clearly on your films. But if the folds were separating, is there some effect where the liquids, because of their adhesiveness, stayed together longer than the cords, even a small extent, and sort of peeled away?

D. CHILDERS: That actually can be seen here, if we had a stop action projector. You could see it frame by frame on this particular example that we chose. That's just one of many that we've done. So that does happen.

M. ROTHENBERG: So that would cause that rounding?

D. CHILDERS: That's correct.

M. ROTHENBERG: That hadn't occurred to me before seeing your films. The third thing that I would like to bring up is that we have been looking at the electroglottograph as a means of measuring the glottal duty cycle, that is, the closed and open quotients. You use the peaks of the differentiated waveform for this. You could use that method, but we have been using the instants that the original, undifferentiated waveform crosses come specified level between the positive and negative extremes of the waveform. We presently use the level about 65% from the negative extreme of the inverse vocal fold contact area waveform (i.e., from the most-closed glottis extreme). Now I thought that the level crossings might result in a more robust measure of the duty cycle than do the peaks in the derivative. In your experience, do the peaks in the derivative waveform provide a robust measure? Do you get too much error when the noise is strong? How well can you use it in running speech?

D. CHILDERS: Probably you could not. It is certainly sensitive to noise. It was a simple procedure for clean signals that one might be able to use in algorithms for speech processing such as in real time. Now, what you've been proposing before, which I learned about, as you know, not too long ago, we simply haven't done yet.

W.J. GOULD: I think that we have to also maintain another thought in mind; that is, we can alter the viscosity of the mucus. The viscosity of the mucus in normal clinical tests has been shown to be tremendously different at different times. Therefore that has to enter into calculations and thoughts.

D. CHILDERS: That is a good point. We haven't seen many patients. They're hard to come by. We've seen some, but they don't always have mucus.

A. KEIDAR: What kind of instructions did you give your subjects when you did your measurements?

D. CHILDERS: Very simple instructions. First, if they can tolerate the laryngeal mirror, we simply use a set of headphones and a sine wave generator, so that it doesn't interfere too much with the instructions that are given to the subject, such as, cock their head a little bit to try to get the field of view more correct. So it's simple in terms of trying to match a frequency in terms of the fundamental frequency of phonation. For intensity, we take whatever we can get when the subject's phonating. They just know that in steps they have to go up, and you happen to see nice four dB steps there. They don't always come in nice four dB steps. Sometimes they're two and a half dB steps or something else like that. With the same instructions, the same subject can reproduce the "same" EGG waveform the next day or the next week.

18 ▪ SIMULTANEOUS PHOTO- AND ELECTROGLOTTOGRAPHIC MEASUREMENTS OF VOICE STRAIN

Peter Kitzing

INTRODUCTION

In phoniatric practice a very common patient is the one who must speak a great deal in his professional work and who complains of excessive vocal fatigue, feelings of irritation in his throat, and increasing hoarseness. The most prominent feature of the voice quality of these patients is usually their straining and tenseness, a voice quality dimension which sometimes also is called pressed or hyperfunctional. Rather often there can be heard hard attacks and some compressed vocal fry. Voice strain is closely correlated with vocal effort, but in perceptual evaluations of voice quality, the dimensions of intensity and strain should be distinguished.

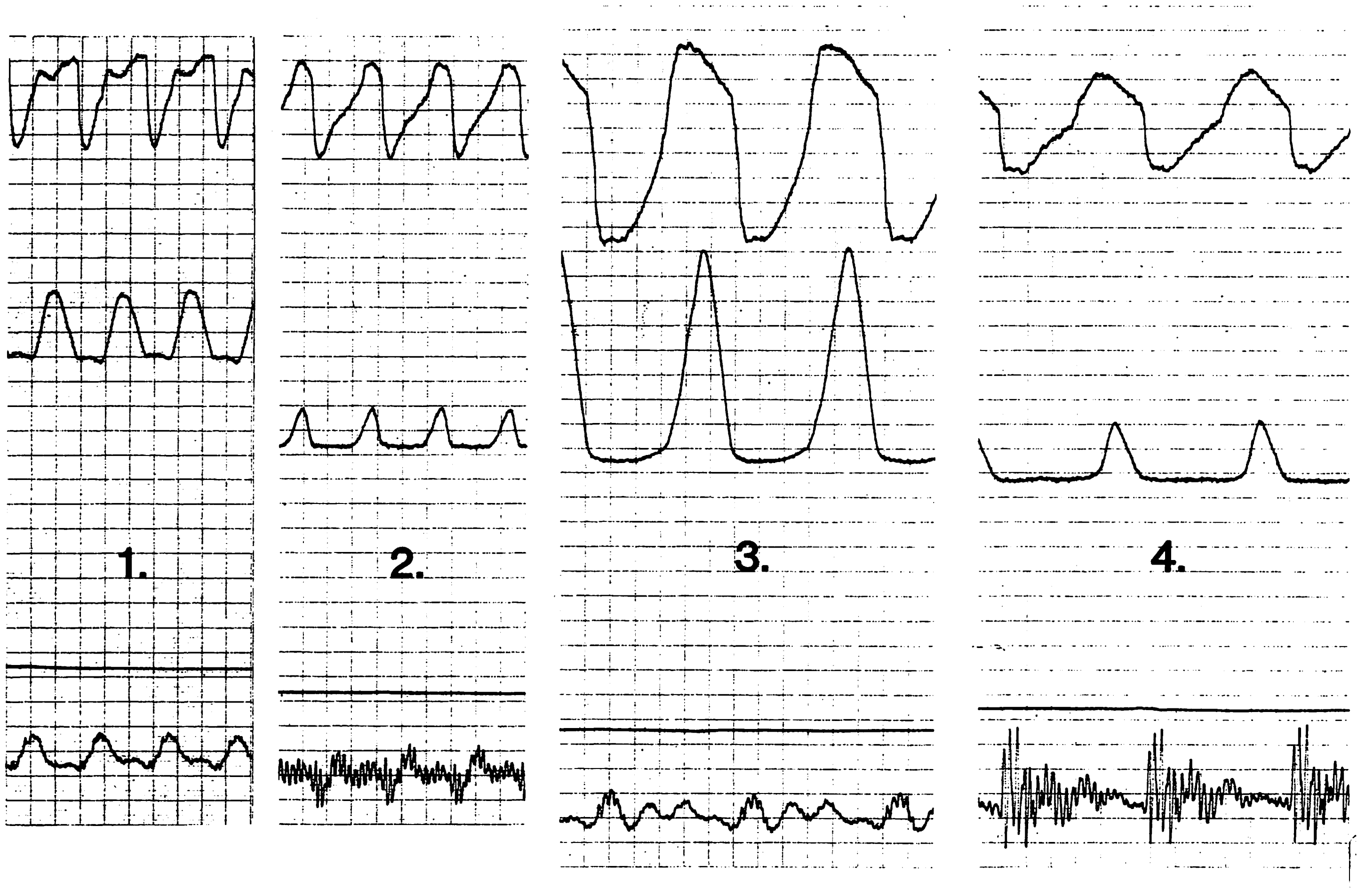
One important goal of voice therapy is to teach the patient to avoid vocal strain even when raising his intensity or else speaking in stressful situations, and the extent to which he can do so is a useful criterion of therapeutic success. Therefore a practical method for quantifying the degree of voice strain would be of benefit for phoniatrical diagnostics and for the evaluation of voice therapy.

So far, there seem to exist only indirect methods, which have not been generally accepted in clinical routine work. One would be to measure subglottal pressure during phonation (Schutte, 1980). Another one is to use the inadvertent rise of vocal pitch during effort (part of the so called Lombard sign) as a measure of vocal strain. This is done in a load test of vocal function developed in Malmö, where the mean voice fundamental frequency and range are measured on line (Glottal Frequency Analyzer) while reading aloud before, during and after 30 minutes of competing white noise at 70 dB SPL (Kitzing, 1979). One serious disadvantage of these indirect measures of vocal strain in clinical work is the difficulty to distinguish strain from effort.

The most appropriate way to investigate organic laryngeal pathology (nodules, polyps, cancer, paresis, etc.) seems to be the use of optical methods, especially indirect microscopy of the larynx and microstroboscopy (magnifying the image of the laryngeal mirror by means of an operating microscope equipped with a stroboscopic light source). On the other hand, pathologic function of the larynx can be recorded acoustically by means of tape recordings for qualitative evaluation or quantitative acoustic analysis. The present paper is an attempt to quantify a special kind of laryngeal dysfunction viz. vocal strain, by combining acoustic and glottographic recordings.

SUBJECTS AND METHOD

Two adults, one female and one male, without any known vocal pathology, produced the vowel [ɛ:] for periods of 5 to 10 seconds, trying, by aid of a tuning fork, to keep the pitch constant at 110 and 220 Hz, respectively, and changing their degree of vocal strain according to a preset schedule: normal, strained, very strained, normal, loose, very loose. The schedule was carried out three times, yielding six samples of normal phonation and three samples of each of the other qualities for each speaker. During loose phonation, the speakers


1.
2.
3.
4.

took care to avoid a breathy quality of phonation.

The following recordings were carried through:

1. Electroglottography. Registration of the electrical impedance changes during phonation in an AC-field in the soft tissues of the neck between two electrodes placed about 3 cm apart on the skin in front of the larynx. Apparatus: Electro-Glottograph, Type EG 830, FJ Electronics.
2. Photoglottography. Transillumination of the neck tissues immediately beneath the larynx by means of a fibre optic light source and monitoring the variations of the glottal aperture during phonation with a light sensitive transducer in the pharynx, inserted through the nose. Apparatus: Photoglottograph, LG 900, FJ Electronics.
3. Recording of the audio signal by means of a high quality microphone (Sennheiser MD 421), and a custom-built amplifier.

The three signals were recorded on an FM tape recorder, Lyrec TR 86 (frequency range 0-10 kHz) which was played back at 1/64 of its original speed to provide curves by means of an ink recorder, Elema Schönander Mingograph 800. Typical recordings appear in Figure 18-1.

A second microphone signal was stored by means of an ordinary, studio quality tape recorder, Revox A 700, and used for acoustic analyses and a rating experiment. The audiotape recordings were analysed with regard to the overall sound level of the voice samples compared to an electrical standard (integrating sound level meter, Bruel & Kjaer type 2218), and the intensity of the fundamental was measured by an FFT-frequency analyzer (Nicolet 444 A) calibrated to a different standard.

The tape-recorded voice samples were copied on a second tape in random order while keeping the sound level equal for all samples. Each sample on the slave tape was preceded by a spoken number for identification and a copy of a sample of normal phonation. This normal sample served for comparison in a forced choice rating test, carried out by twelve experienced voice therapists, using a scale of five degrees of vocal strain. The raters were told only to consider the dimension of vocal strain, trying to neglect other dimensions as for instance degree of breathiness or vocal fry. The standard deviation from the mean ratings can be said to represent a measure of the interrater reliability. On the average, it amounted to 16% (range 0-35%) of the mean rating scores.

MEASUREMENTS

In the glottograms from each voice sample, for seven approximately equally distanced places five consecutive periods were measured by hand and averaged. The resulting seven figures were then averaged to represent the numerical glottographical value for the sample which was correlated with the results from the rating test.

FIGURE 18-1

Typical glottographic registrations and microphone signals of normal and strained voices. From top to bottom: Electroglottograph, photoglottograph, microphone signal.

1. **Female speaker, normal voice 224 Hz.**
2. **Female speaker, strained voice, 227 Hz.**
3. **Male speaker, normal voice, 109 Hz.**
4. **Male speaker, strained voice, 108 Hz.**

In the glottograms glottal closure is indicated by downwards direction of the curves.

TABLE 18-1
Glottographic parameters (means and standard deviations).

Sex of speaker	Intended voice quality	Photoglottographic						
		Period	Closed time (mm)*) n	Open time (mm)*) =	Open quotient 42	Speed index	Amplitude (mm on standard registration)	Difference of Electroglottographic Closed time (mm)*)
female	hyper-function	14.1 0.21	7.6 1.68	6.5 1.83	0.46 0.13	0.13 0.05	7.8 2.49	−0.04/0.42 n = 39
	normal	14.1 0.15	5.1 0.58	8.9 0.59	0.63 0.04	0.02 0.05	10.6 4.71	0.24/1.35 n = 11
	hypo-function	13.9 0.20	4.8 0.46	9.2 0.60	0.66 0.03	−0.05 0.04	9.0 4.78	0.68/0.28 n = 17
male	hyper-function	29.6 0.28	18.7 1.28	10.9 1.22	0.39 0.04	−0.10 0.06	21.5 15.32	−2.9/1.58 n = 23
	normal	29.3 0.21	16.7 0.72	12.8 0.63	0.44 0.02	0 0.09	32.6 8.92	−0.8/0.55 n = 33
	hypo-function	30.2 0.31	15.2 1.62	15.0 1.49	0.50 0.05	0.01 0.06	21.9 7.04	−0.9/0.79 n = 28

*) 1 mm corresponds to 0.3125 ms

TABLE 18-2

Correlations between perceptual rating of vocal strain and different parameters of the voice samples.

Parameter	Female speaker	Male speaker
Speakers' intended degree of hyperfunction	.95***	.93***
PGG period	.34	−.57**
PGG closed time (c)	.81***	.78***
PGG open time	−.76***	−.86***
PGG open quotient	−.79***	−.75***
PGG speed index	.89***	−.21
PGG amplitude (a)	−.16	−.13
Overall sound level (I)	.60**	.89***
Level of fundamental frequency (I_{Fo})	.65**	−.36
c/a	.33	.52*
c/I_{Fo}	.86***	.82***
EGG closed time (c_E)	.62*	.56*
c_E/I_{Fo}	.56*	.71**

CALCULATIONS. RESULTS.

The photoglottographic open quotient (open time divided by entire period) and the speed index (difference between opening time and closing time, divided by open time; cf. Hirano, 1981) were calculated for each voice sample. To show the general tendencies of the results, the measurements from each of the three speaker intended main voice qualities were averaged: hyperfunction = 3 x 7 samples of very strained quality plus 3 x 7 samples of moderately strained quality; normal = 6 x 7 samples of normal quality, regarding degree of vocal strain; hypofunction = 3 x 7 samples of moderately loose quality plus 3 x 7 samples of very loose (but not breathy) quality. The results appear in Table 18-1.

Secondly, different numerical parameters of each of the eighteen voice samples from either speaker were correlated to the mean rating scores of vocal strain. The results appear in Table 18-2.

DISCUSSION

As the strained quality of a voice sometimes might be difficult to ascertain, the intention of the subjects to produce a certain degree of vocal strain did not seem sufficient to form a basis for further investigations. Therefore, a group of expert listeners which had demonstrated an adequate inter-rater reliability was used to evaluate the degree of straining. The correlation between the speakers' intended degree of vocal strain and the results from the rating test (Table 18-2) turned out to be very good.

From the measured period times it appears, that both speakers had some difficulty in producing the intended pitch of 220 Hz and 110 Hz, respectively. The actual pitches were about 227 Hz and 108 Hz, with the male speaker showing a tendency to raise his pitch systematically with increasing degree of vocal strain. This is documented in the weak but significant negative correlation between the degree of rated vocal strain and the period time for that speaker. On the whole, the stability of the pitch parameter in this investigation is regarded as acceptable. Too great variations of pitch could have influenced the value of L_{Fo}, - see below.

Photoglottography (PGG), as developed by Sonesson (1960, Kitzing & Sonesson, 1974) has been compared to measurements from high speed films of the larynx. Generally, there has been a good correspondence between measurements of glottal area variations from the two methods (Zemlin 1959, Harden 1975, Baer *et al* 1982). This seems especially to be the case with respect to the temporal aspects of the period, viz. the time of glottal closure and the moment of maximal glottal opening. On the other hand, the amplitude of the glottal vibratory period cannot be measured reliably by means of photoglottography.

Comparisons between electroglottography (EGG) and stroboscopical investigations of the vibrating vocal folds (Fourcin 1975, Lecluse 1977) or high speed films (Baer *et al* 1982) seem to substantiate the common view, that the EGG signal indicates variations of vocal fold contact area. The EGG has been shown to clearly identify the termination of the glottal pulse in comparisons with the PGG (Kitzing 1977, Baer *et al* 1981, Dejonckere 1981) and inverse filter glottograms of oral air flow (Rothenberg 1979). On the other hand, it seems impossible to define the moment of total closure from the EGG (Lecluse *et al* 1975), nor can the point of maximal opening of the glottis (and thereby the vibratory amplitude) be recorded by this method. The non-invasive EGG method has been used to study irregularities of vibration in voice disorders (Neil *et al* 1976, Wechsler 1977, Jentzsch *et al* 1981) and to measure changes of glottal closure after voice therapy (Thyme & Frokjaer Jensen 1982). Schutte (1982) was unable to observe specific differences in EGG recordings from speakers with pathological conditions of the larynx as compared to normal speakers.

Since vocal strain means a sphincteric activity of the larynx, a consequent increase of the closed time and a corresponding decrease of the open time in the glottal period might be expected, which means a decrease of the open quotient (open time divided by entire period). This was the case in both speakers, resulting in the high (positive or negative) correlation coefficients for the aforementioned parameters (Table 18-2). From a statistical point of view, therefore, a measure of glottal closure seems sufficient to describe the degree of strain.

However, physiologically, there seem to be two ways to accomplish an increase in the closed time. One is vocal strain, which should be avoided in healthy voice habits. The other one would be a glottal closure during phonation caused primarily by the aerodynamic forces due to the Bernoulli effect as a result of the upward propagating mucosal wave. Accomplishing phonatory closure in this manner is likely to be less harmful and tiring, and therefore more desirable from the point of voice hygiene.

One essential difference between the two modes to accomplish phonatory closure is the amount of the vibratory amplitude. When the tonus in the laryngeal muscles is comparatively low and the vocal folds are relaxed, the vibratory amplitude increases with a consequent enhancement of the pressure drop in the glottis due to the Bernoulli effect, meaning a more efficient vibratory closure. On the other hand, it is a common experience from clinical laryngostroboscopic examinations, that the vibratory amplitudes decrease the more the voice quality is strained.

This observation was substantiated by the inverse-filter experiments of Sundberg and Gauffin (1979), who found "pressed" phonation to be characterized by a small inverse filter glottogram amplitude as well as by a comparatively long closure time.

Unfortunately, amplitude measurements from glottograms are not reliable enough to be used, since they depend too much on irrelevant conditions, such as the thickness of the neck tissues, the position of the larynx in relation to the transducer, etc. As could be expected, there is only a weak correlation between degree of vocal strain and a quotient, c/a, which takes into account both the increasing closed time and the decreasing amplitude as measured on the photoglottograms (Table 18-2).

Here another of Sundberg's and Gauffin's findings (1979) seems to be of help. In their investigations of the spectrum of the glottal voice source they found a high correlation between the amplitude of the glottogram and that of the fundamental, the best straight line approximation of this relationship being

$$L_0 = 25 \log Ag + C_1$$

where L_0 is the level of the fundamental in dB, Ag is the glottogram amplitude and C_1 a constant.

The level of the fundamental might also be influenced by the vocal pitch and by the quality of the vowel. Keeping these paramenters under control, an attempt was made to use the level of the voice fundamental (L_{Fo}), from an FFT-frequency analysis as a measure of the glottogram amplitude. As the quotient between the closed time and L_{Fo} correlated very well with the rating scores of voice strain (cf. Table 18-2), the attempt seemed to be successful. This is felt to be the most important finding of the present investigation.

Some other interesting correspondences between rated vocal strain and physiological parameters appeared. The high correlation between the rated degree of strain and the overall sound level of the voice samples is in accordance with Sundberg's and Gauffin's findings (1979) of a "comparatively high" SPL in "pressed" phonation. The lower value for the female speaker ($r = 0.60$) is due to varying degrees of pharyngeal resonance, inconsistently used when straining her voice. When this could be heard on the tape recordings, there appeared a richness of high frequency harmonics in the FFT-analysis, a finding which is in accordance with Hirano's description of "hyperfunctional" voice quality (Hirano, 1981).

Strained voice being a result of laryngeal sphincteric activity, a hightened speed of phonatory glottal closure during voice strain would have been expected. As a measure of glottal closure speed, the glottographic speed quotient, Sq (opening time divided by closing time), has been in common use, and it has been possible to show a decrease in Sq as a result of successful voice therapy, meaning less straining (Kitzing & Löfqvist, 1977). However, as Hirano (1981) points out, the use of another quotient, the speed index, SI (difference of opening and closing time divided by the sum of both times), seems to be a more rational measure, in part because it is easier to visualize the waveform from SI values than from Sq values (for a detailed discussion, cf. Hirano p. 46). As was expected, in the present investigation there was found a high correlation between the rated degrees of strain and the SI in the female speaker. The lack of correlation of this parameter in the male speaker is so far unexplained.

The feasability of using the c/L_{Fo} quotient as a measure of voice strain in future clinical practice seems to depend on at least three factors. First, the use of an FFT-analyzer to measure the fundamental seems too expensive. Since the detection of the voice fundamental from the glottal vibratory period time is rather trivial, it is recommended to use this information to control a band-pass filter in a simplified device just to register the Fo level. Secondly, all measurements and calculations should be computerized by means of some desk computer, because they are far too time consuming to be done by hand.

Thirdly, the recording of the closed time of the vibratory period should be simplified. The photoglotography is a semi-invasive method, not well accepted by some subjects, and liable to fail when there is too much sphincteric activity in the throat. For this reason, the PGG recordings were combined with those from an electroglottograph. As could be expected, there was no difference between the PGG and EGG registrations of vibratory period length. However, the EGG equipment available for this investigation produced measurements of the closed time which differed substantially from the PGG measurements (Table 18-1), resulting in a poorer correspondence of the EGG parameters (c_E and c_E/L_{F_0}) with the rated degree of vocal strain (Table 18-2).

To conclude, the quotient between the vocal vibratory closed time and the level of the voice fundamental (as a measure of the vibratory amplitude) seems to be a useful measure of voice strain based on current concepts of voice physiology. The feasability of adopting this measure to routine clinical work with cases of voice disturbances depends on the development of more simplified methods. The aim of our future investigations is to evaluate the findings of this pilot study by the use of more subjects and to develop a technical apparatus to meet the aforementioned prerequisites.

ACKNOWLEDGEMENTS

The author is grateful to Johan Sundberg, Department of Speech Communication and Music Acoustics, KTH, Stockholm, for valuable comments on the discussion version; to Per Nilsson, M.D., Univ. E.N.T.-Clinic, Malmö, for putting the FFT-analyzer at our disposal; and to Derek E. Dunn, Ph.D., Exerimental Dept., Malmö General Hospital, for revising the English text.

REFERENCES

Baer, T., Titze, I.R. and Yoshioka, H. (1981). "Multiple simultaneous measures of vocal fold activity," Proceedings of the Vocal Fold Physiology Conference, Madison, WI.

Baer, T., Löfqvist, A. and McGarr, N.S. (1982). "Laryngeal vibrations: a comparison between high-speed filming and glottographic techniques," to be published in J. Acoust. Soc. Am.

Dejonckere, P.H. (1981). "Comparison of two methods of photoglottography in relation to electroglottography," *Folia phoniat.* 33, 338-347.

Fourcin, A.J. (1975). "Laryngographic examination of vocal fold vibration," in B. Wyke (Ed.) *Ventilatory and Phonatory Control Mechanisms*. Oxford Univ. Press, 315-333.

Harden, R.J. (1975). "Comparison of glottal area changes as measured from ultrahigh-speed photographs and photoelectric glottographs," *J. Speech and Hear. Res.* 18, 728-738.

Hirano, M. (1981). *Clinical Examination of Voice*. Wien, New York: Springer Verlag.

Jentzsch, H., Unger, E. and Sasama, R. (1981). "Elektroglottographische Verlaufskontrollen bei Patienten mit funktionellen Stimmstörungen," *Folia phoniat.* 33, 234-241.

Kitzing, P. and Sonesson, B. (1974). "A photoglottographic study of the female vocal folds during phonation," *Folia phoniat.* 26, 138-149.

Kitzing, P. (1977). "Methode zur kombinierten photo- und elektroglottographischen Registrierung von Stimmlippenschwingungen," *Folia phoniat.* 29, 249-260.

Kitzing, P. and Löfqvist, A. (1977). "Clinical application of combined electro- and photoglottography," *I.A.L.P. Cong. Proc.*, Copenhagen 1, 529-539.

Kitzing, P. (1979). "Glottografisk frekvensindikering," thesis, Malmö.

Lecluse, f., Brocaar, M.P. and Verschuure, J. (1975). "The electroglottography and its relation to glottal activity," *Folia phoniat.* 27, 215-224.

Lecluse, F. (1977). "Electroglottografie," thesis, Univ. of Rotterdam, 1-181.

Neil, W.F., Wechsler, E. and Robinson, J.M.P. (1977). "Electrolaryngography in laryngeal disorders," *Clinical Otolaryngol.* 2, 33-40.

Rothenberg, M. (1979). "Some relations between glottal air flow and vocal fold contact area," Proc. of the Conf. on the Assessment of Vocal Fold Pathol., Nat'l Inst. of Health, Bethesda, MD, *ASHA Rep.* 11, 88-96.

Schutte, H.K. (1980). "The efficiency of voice production," thesis, Groningen.

Schutte, H. (1982). "Interpretation of electroglottograms, based on zonal representation within the phonetogram," manuscript to be published.

Sonesson, B. (1960). "On the anatomy and vibratory pattern of the human vocal folds," thesis, *Acta Otolaryngol. Suppl.* 156, 35-80.

Sundberg, J.K. and Gauffin, J. (1979). "Waveform and spectrum of the glottal voice source," in Lindblom, B. and Öhman, S. (Eds.) *Frontiers of Speech Communication Research.* London, New York, San Fransisco: Academic Press, 301-320.

Thyme, K. and Frokjaer-Jensen, B. (1982). "Acoustic changes caused by speech training compared with changes in electro-glottographic registrations," paper, Congress of the Union of European Phoniatricians, Prague.

Wechsler, E. (1977). "A laryngographic study of voice disorders," *Brit. J. Dis. Com.* 12, 9-22.

Zemlin, W.R. (1959). "A comparison of high speed cinematography and a trans-illumination-photoconductive method in the study of the glottis during voice production," thesis, Univ. of Minnesota. 1-34.

DISCUSSION

M. ROTHENBERG: I've noticed sometimes in using and comparing the electroglottograph waveform with air flow, that there is a shape factor when you have a pressed or adducted voice. Now you mention, and I agree, that there may be different ways of having adducted voice, and I'm not sure which one applies here. The shape factor being that the waveform loses its sharp rise during the adduction and starts to look more like a sawtooth. In that case, it's hard to separate the open and closed portions. Now, have you noticed such a shape factor in your cases of the strained voice or the pressed voice?

P. KITZING: That might also come about, as we heard this morning, because of mucus; I just don't know why it is sometimes easy to measure and other times not. (Slide is projected.) Now, this is normal voice, and was obtained using my electroglottograph. This is what the waveforms would look like. (Pointing to slide.) This would be a smooth voice, and this is the cracked voice. This has the longer closed time, and I usually try to measure from here to over there, if there is not a notch. Sometimes you can see this kind of notch; which we could see also this morning in the electroglottograph. But this notch isn't always seen.

M. ROTHENBERG: That's not the type of pattern that I'm talking about. Maybe we could talk about this later.

P. KITZING: What I want to do is give you a warning against the electroglottograph.

19 ▪ CLINICAL SIGNIFICANCE OF A VOCAL EFFICIENCY INDEX

Nobuhiko Isshiki

At the first Vocal Fold Conference in Kurume, the idea of vocal efficiency was presented, with its technical problems and clinical results. Since then, a variety of proposals have been made by many investigators to assess some aspect of the efficiency of voice production (Figure 19-1). Here in this paper discussed are the theoretical background for the vocal efficiency index we employ as a clinical test, the associated technical problems and the tentative solution for them. The efficiency of voice production was first defined by van den Berg as a ratio of acoustic power of voice to the subglottic power. The acoustic power of voice was roughly calculated from the measured sound intensity. Due to the difficulties or inconveniencies involved in measuring the subglottal pressure, the efficiency of voice production defined like this has not come into wide clinical use.

In the Kurume Conference, the author proposed AC/DC ratio as a vocal efficiency index, avoiding confusion with the true efficiency of voice production. To be more specific, it is a ratio of effective value of a vocal sound wave detected at the sensor of a flow-meter to the DC component or mean flow rate of the expiratory air flow rate during phonation. Since then, many other proposals followed. Citing just a couple of examples, the decibel value of sound intensity devided by the air flow rate was proposed as a simple indicator of efficiency. At first sight, this value may appear rational to express the vocal efficiency in rough estimate, because a louder voice produced with less air flow means higher efficiency. Supposing with the use of air flow of 150 cc/sec, one produced voice of 60 dB SPL and the other 80dB. Then the latter can be said to have a higher efficiency than the former. Conversely, the voice 80dB SPL produced with 100cc/sec air flow is said to be more efficient than the same intensity voice produced with 200c/sec air flow. It seems not contradictory to our common sense. But the story is not so simple. According to this formula of expressing the vocal efficiency, the efficiency of voice of 60dBSPL produced with 100cc/sec air flow should be equal to that of 90dBSPL voice produced by 150cc/sec air flow. Now, we will notice that this formula may be inadequate for expressing the vocal efficiency, because we all feel that the voice as loud as 90dB produced with 150cc/sec has certainly a higher efficiency than the voice as soft as 60dB produced with 100cc/sec air flow.

Let's be more theoretical now, and first let's begin with the definition of efficiency. According to the American Standard Acoustical Terminology, approved by American Standard Association, it is defined as follows. The efficiency of a device with respect to a physical quantity which may be stored, transferred or transformed by the device is the ratio of the useful output of the quantity to its total input. What is important here is that it must be the ratio, which implies that the numerator and denominator must have the same dimension. It follows that we can not divide the decibel value by flowrate in order to obtain the efficiency. The vocal efficiency index, or AC/DC ratio as we defined is consistent with the above American Standard Definition, since both AC and DC are expressed by volume velocity. In using the vocal efficiency index as a clinical diagnostic tool, we must be well aware of the two facts. First, it has nothing to do with the quality of voice. What is meant by a useful output in the previous definition is an acoustic amplitude expressed in volume velocity, no mat-

1) Efficiency of voice production

$$= \frac{\text{Acoustic Power}}{\text{Subglottal Power}}$$

2) Vocal Efficiency Index

$$= \frac{\text{AC(rms of AC component)}}{\text{DC(mean flow rate)}}$$

3) Vocal Intensity(dB)/mean flow rate

4) Vocal Intensity(dB) - $\log(\text{Flow rate})^2$

FIGURE 19-1
A number of proposals were made to express some aspect of vocal efficiency.

ter whether it be harmonics or noise. No consideration is given to the quality of voice, whether normal or hoarse.

Second, a high vocal efficiency does not always mean well-functioning superior larynx. It may be likened to a device which is functioning at a very high efficiency but will soon be in trouble due to overwork. What we regard as an ideal larynx is not necessarily a high-efficient one, but the one which produces high quality voice, without any trouble forever. I am wondering whether or not there is any critical vocal efficiency, over which the voice tends to deteriorate, leading to hyperfunctional voice disorder. To my knowledge, there has been no objective test available to detect or to assess the hyperfunctional factor in functional dysphonias. Now, I begin to feel that the vocal efficiency index can be a breakthrough against this difficult problem, in other words, a potential tool for the diagnosis of hyperfunctional dysphonia. In the previous paper, we mentioned that the vocal efficiency below 50 implies imperfect closure of the glottis. The value indicates what the extent of trouble is and helps judge the effect of therapy, when the larynx is already in dysfunction. What seems equally or more important is to detect the precipitating factor for hyperfunctional dysphonia, that is too much strain. There may be a very high overloading efficiency at the initial stage of the disease.

From the technical viewpoint of measuring the AC/DC ratio, there is a problem when AC/DC ratio is high, say above 50, as in normal phonation without glottal gap. In the research process, in cooperation with Dr. Ishizaka, of how to get rid of the effect of resonance of the vocal tract and the tube, we became more and more aware of and concerned about the possible non-linear function of the hot-wire output at a very low flow rate. In Figure 19-2 shown is the frequency response of the hot-wire flow-meter we obtained, when a constant flow of N2 gas at a ratio of 200cc/sec was superimposed on the sound waves. However, the frequency

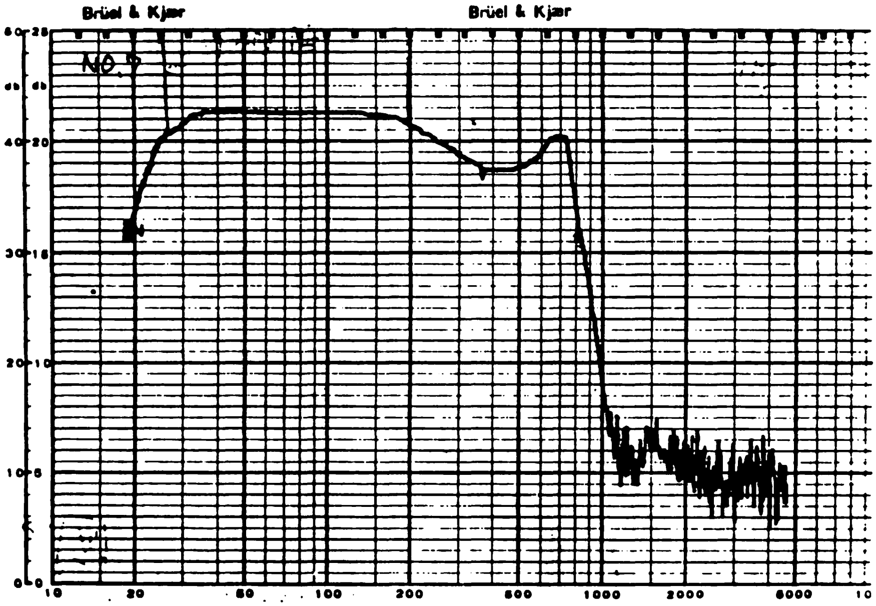


FIGURE 19-2
The frequency-response of the hot-wire flow-meter we obtained, when a constant flow of nitrogen gas at a ratio of 200cc/sec was superimposed on the sound wave.

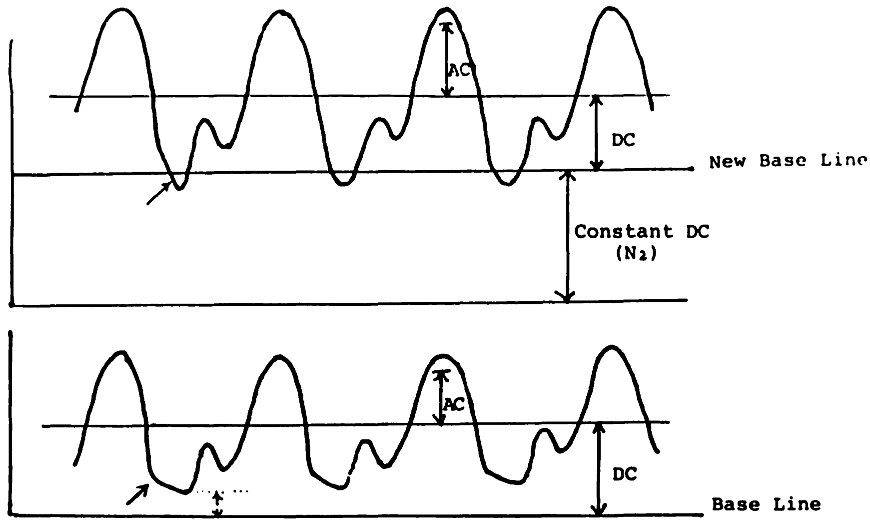


FIGURE 19-3
Difference in the wave-form is noted between the two curves: the above wave obtained with constant DC flow superimposition has much less distortion than the below obtained without the flow superimposition.

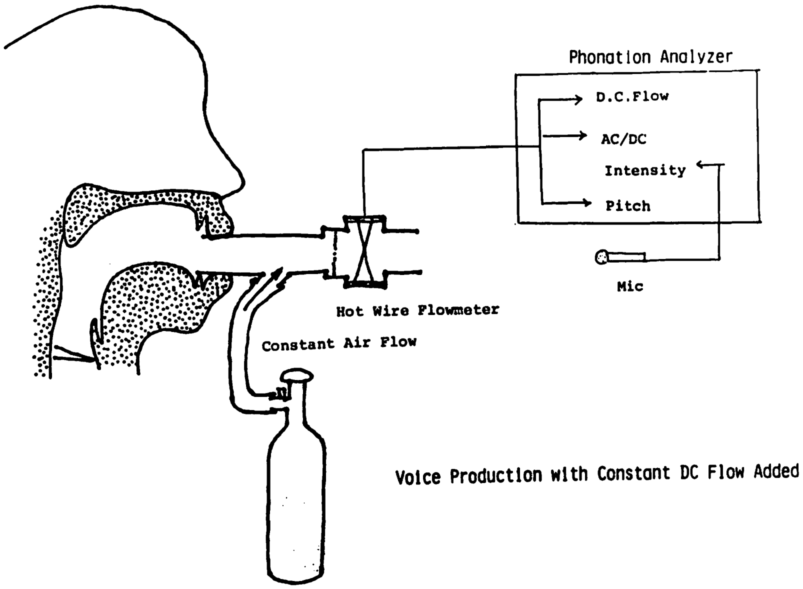


FIGURE 19-4
Experimental set-up for voice production with constant flow of gas superimposed. A hot-wire flow meter has a low frequency-response at low DC flow rate. In order to avoid this low frequency response region, DC constant flow was added.

response at a very low flow rate could not be measured, because a precisely constant low flow is very difficult to produce. At a very low flow rate, a heated wire takes time to be cooled down, because of a slight wind. If a constant flow is superimposed, there would be no or much less problem. This low frequency response at low flow rate is almost an inevitable draw-back of a hot-wire flow meter, although much improved by the use of a feed-back system to maintain a constant temperature of the wire. This non-linearity causes some distortion in the wave form, especially near the base-line, as you see in Figure 19-3. This distortion in turn results in false decrease in measured AC component, increase in DC component, and consequently erronously low AC/DC ratio.

There will be no such problem, when the flow is high in such cases as vocal cord paralysis with a large glottal gap. This is really a complicated problem, and in an attempt to find some tentative solution, we made the following experiments as shown by the Figure 19-4. A constant nitrogen gas flow at a rate of 200cc/sec was provided, joining the expiratory air flow during phonation. This gas addition made the difference in the wave form and also the relation between the wave and the base line. Our new model of hot wire flow-meter is capable of displaying the mean flow rate, AC/DC ratio, intensity and pitch at any chosen moment in digital form (Figure 19-5). During phonation with superimposed nitrogen gas flow, these parameters were measured, and the real mean expiratory flow and AC/DC ratio without N2 gas were deduced (Figure 19-6). Some of the results are shown in Figure 19-7. In general,

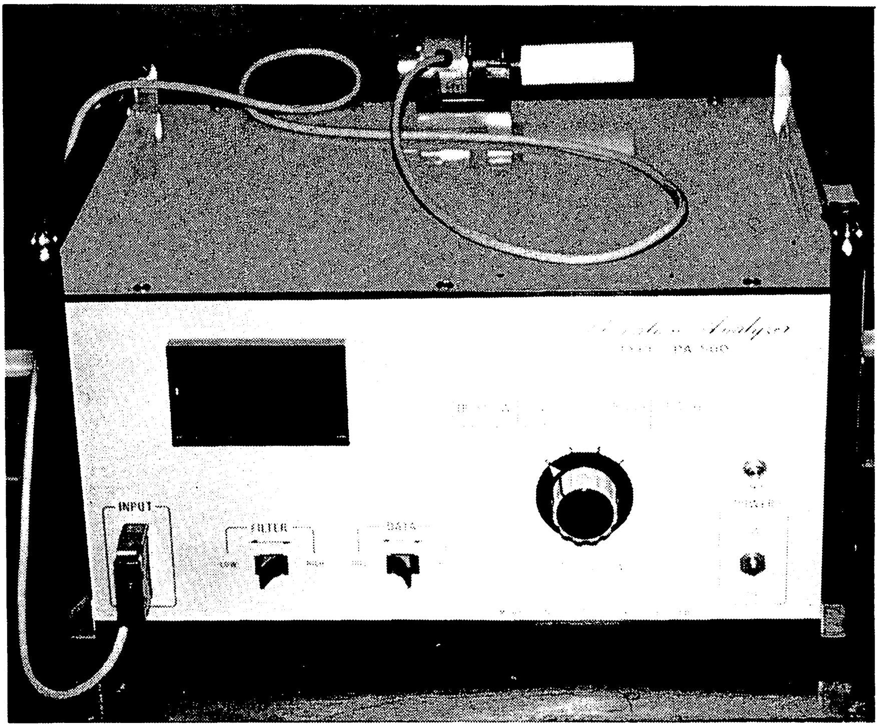


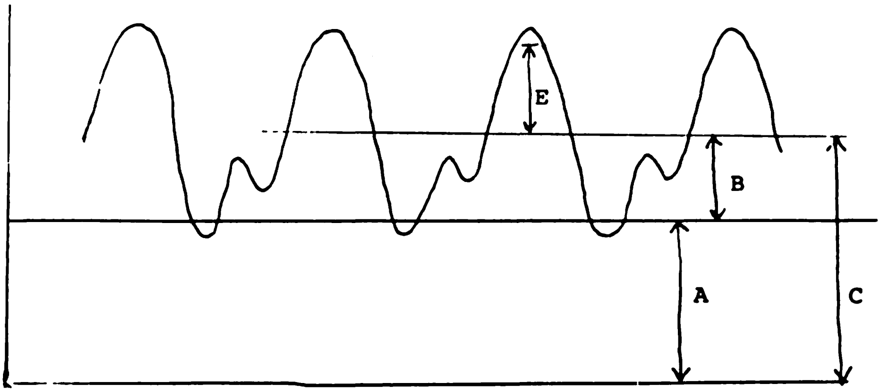
FIGURE 19-5
Phonation analyzer permits digital display of the mean flow rate, AC/DC, intensity, and pitch at a selected instant.

the calculated AC/DC ratio was substantially higher than the AC/DC ratio obtained without nitrogen gas supply, say by over 10%. However, when AC/DC ratio is low, as low as 30, or when imperfect closure of the glottis is expected, there was no much difference noticed in AC/DC ratio, whether with N2 gas superimposed or not.

These findings were quite conceivable and just what we expected. Although we feel that this method is a step forward toward the ideal instrumentation for evaluating vocal efficiency index, it is inconvenient and may create another problem. Although we have been engaged in this for over 6 years, we are sorry but we are not yet at the stage to report the many clinical data which you may think interesting. It is a very interesting and potential instrument but also very difficult to handle. Now I would like to end this interim story of our struggle with the instrument, and where we are now.

In summing up,

1. The vocal efficiency index as we proposed is consistent with the definition of the standard terminology.
2. The vocal efficiency index appears to have clinical meanings or application in (1) objective judgement of degree of dysphonia and consequent assessment of any therapeutic effect, and (2) potential detection of hyperfunctional or overloading way of voice production.



DC Bias (N_2Gas) : A

Mean Flow Rate : B

Total Mean Flow Rate : C

AC/DC : D

AC : E

Corrected AC/DC : F

$$\text{Corrected AC/DC} = \frac{C \times D}{C - A}$$

FIGURE 19-6
Calculation of real AC/DC from the obtained values on the instrument, when constant DC gas is superimposed.

3. The technical problems are associated with (1) poor frequency response at low flow rate, and (2) resonance effect.
4. As a tentative solution, an experiment of nitrogen gas superimposition to expiratory air was attempted with some apparent improvement.

DISCUSSION

I. TITZE: There is a fundamental difficulty with a definition of an efficiency ratio that does not involve power or energy. Two equally efficient hydroelectric plants, for example, would be judged very differently in terms of efficiency if only flow were considered. A plant on a large river has small pressure, but large flow. A plant on a small mountain stream reservoir has small flow, but large pressure. The product of pressure and flow measures the available (input) power. As another example, a transformer can produce a ratio of output voltage to input voltage of 10,000 or more, but the efficiency is always less than 100%. It seems that we need to consider the energy distribution in tissue, airstream, and acoustic radiation to define vocal efficiency.

DC Bias N_2Gas	Total Mean Fl.R.	AC/DC	Int.	Pitch	Cor.AC/DC	Mean Fl.R
96	207	35	76	134	65	111
0	109	53	77	136		

96	237	48	82	185	81	141
0	141	65	82	186		

94	221	62	83	211	108	127
0	143	73	82	211		

FIGURE 19-7
Samples of corrected AC/DC as against non-corrected AC/DC.

N. ISSHIKI: Thank you for your comment. According to the subnote to the definition of efficiency in the American Standard Acoustical Terminology, the input and output usually mean the power, unless otherwise stated. There is no problem, if we use power for both the input and output quantities. But this subnote also implies that these terms need not be necessarily the power, if so specified. The key words in the definition are "ratio" and "useful". The input and output quantities should have the same dimension in order for them to be applicable to the idea of efficiency.

If we take the flow rate as a physical quantity of the input in the hydroelectric power plant, then the ouput quantity should also be the flow rate. The useful output in this case is zero, because the outflow of water is not useful from the viewpoint of an hydroelectric power plant. In other words, the flow rate cannot be used as an input and ouput quantity, because it is not useful in this case. Therefore, it is inadequate to apply the flow-efficiency to the hydroelectric power plant.

But, in applying the flow-efficiency to glottal function, the input and output have the same dimension, cc/sec, and the root mean square value of the AC component of the outflow is useful for vocal intensity. I think that the vocal efficiency index or flow-efficiency as defined in my paper seems compatible with the definition in the American Standard Terminology. The value can exceed 100%, which certainly appears strange to our common usage of "efficiency", but not irrational, I think, on the basis of the definition.

As for the example of a transformer, I am not sure whether the concept of voltage-efficiency can be applied to a transformer or not, as it is not clear to me what the useful output is. I would like to have the opinion of a specialist on this point.

I. TITZE: If we could find a better way to measure subglottal pressure, that would take care of it.

N. ISSHIKI: Yes, but it's actually impractical from the clinical viewpoint.

I. TITZE: I realize that.

G. FANT: It is apparently not enough to have a single measure of efficiency. We could use some kind of physiological efficiency parameter related to the type of voicing and its ability to save the voice. You could even use the ratio of AC over DC, which would be a good indication of efficiency. In a weak voice this ratio is small. One simplified measure of the AC to DC ratio could be obtained from the mouth output and its time integral. This index, however, would lack the leakage component of a breathy voice.

M. ROTHENBERG: Perhaps the problem is just one of terminology. Maybe it should be called a breathiness index instead of using the word efficiency index. If performed correctly such a measure could have some value, because it does relate to certain types of voice disorders. Perhaps just changing the name a little may satisfy the uneasiness expressed by some here. As I said, perhaps call it "breathiness index" and save the word efficiency for some other type of index.

N. ISSHIKI: I'd like to make a comment on that. If we call it a breathy component, we must be aware that normal voices may have different AC-DC ratios, and these voices may not be breathy. Some voices have a long closed-phase, and the AC-DC ratio is very high. And other voices may have a short closed-phase, and will not be breathy. These voices have different AC-DC ratios, but both sound normal. Thus, in terms of breathiness there is no breathiness at all. But still, the AC-DC ratio is different. So a breathy component is not indicated in this case.

M. ROTHENBERG: Maybe we need another word, that we can't think of now.

O. FUJIMURA: I have to agree with Ingo Titze. To be general, it's true that we have to consider power rather than just flow. But, if we take as constructive an attitude as possible, we must pin down what the problem is . We have to specify the impedance level for the two quantities that we compare, in order to be able to use just one quantity to specify efficiency. And then the question boils down to what are the impedance levels (approximately) for the AC component and DC component. This, obviously, depends very crucially upon the particular vocal tract condition. That is what we have to consider in defining what efficiency is. If, for example, you specify what particular vocal tract configuration is required for the measurement, this would make the measurement better. We could, for example, specify the vowel gesture. It doesn't solve the entire issue. We might have to think about it from this point of view, and try to make the approximate measure in the clinic as useful as possible.

N. ISSHIKI: Thank you for your comment. And I think that what you may mean by impedance may include the resonance effect or the size of the mouth opening. So we are trying to make it as constant as possible. The size of the mouth is actually held constant. We use the same size of pipe, and try to reduce the effect of resonance. With regard to the question raised by Dr. Titze, one thing I don't understand is that we need not restrict the meaning of efficiency to power because the terminology brochure stated that, unless stated otherwise. So I think it is unnecessary to restrict the efficiency terminology to power.

T. BAER: They probably don't say it in the brochure, but normally you would require, in order to use the term efficiency, that the measure be considered one that is conserved. Power is a physical quantity that needs to be conserved while AC flow does not. I don't know whether this adds anything to the argument.

N. ISSHIKI: The terminology said "store, transform, transfer". Three variations were included.

I. TITZE: If you consider an index that does not have to be in the range of 0-100%, then any operational definition will do for your purpose. But if you think of efficiency as varying between 0 and 100%, then, like Tom said, you have to use a quantity that physically is conserved, otherwise you'd get all sorts of inconsistencies.

N. ISSHIKI: That's right. What I found was that the index exceeded 100%. I thought I should change the type of index, because it did go over 100%. I agree that an index over 100% is strange.

P. KITZING: If the efficiency means the quotient between energy output and energy input, or work output to work input, then from a voice clinician's viewpoint, I would like very much to have an index of the work done in the larynx. I'm interested in the work of the larynx, not the work of the aerodynamic forces. Usually you have plenty of fuel (compressed air), but if the organ is worn out, then you would like to see if the organ is efficient. If one is very relaxed, and one doesn't work very much and one has a good voice, then from a clinician's viewpoint the organ has good efficiency.

N. ISSHIKI: The efficiency you mentioned may be called the vocal efficiency in a physiological sense, as separated from an aerodynamic one. The vocal efficiency in a physiological sense is of great clinical significance but almost inaccessible with the use of the instruments now available.

In order to discriminate the efficiency in my paper from the power efficiency, I add the word "index", which however apparently caused another confusion. I am looking for more appropriate words to express what I mean. The glottal flow efficiency may be the one.

G. FANT: I agree that we need a physiological functional index in addition to physical parameters for different aspects of the phonation. AC/DC quotients may be studied from the output of a Rothenberg mask. Production models provide alternative means of defining vocal efficiency.

N. ISSHIKI: Thank you for your comment. The reason why I use a tube instead of a mask is to keep the mouth-opening constant.

20 ▪ VOICE MEASURES OF PSYCHOLOGICAL STRESS — LABORATORY AND FIELD DATA

Malcolm Brenner, Thomas Shipp, E. Thomas Doherty, and Philip Morrissey

INTRODUCTION

Since Lieberman's (1961) initial report on perturbations in vocal pitch, many studies have reported cycle-to-cycle variations (pitch perturbations or "jitter") that occur in the voice fundamental frequency (F_O) as they relate to subjects' real and feigned emotional states (Williams and Stevens, 1981), to their vocal training (Murry, 1979), their physical health (Ramig and Ringel, 1983), and to their age, sex and smoking habits (Ludlow, Coulter, and Bassich, 1982), among others. One aspect of the voice signal that has received particular attention is the difference in magnitude of pitch perturbations between a control or normal emotional state and one in which the individual is under psychological stress. This issue has gained prominence recently with the development of a series of commercial devices, such as the Psychological Stress Evaluator, which claim to determine whether or not a speaker is lying based on analyses of recorded speech using a factor resembling jitter (Bell, Ford, and McQuiston, 1976; Dektor Inc., undated; McNeice and Cota, 1980). These devices have been largely discredited for purposes of lie detection (Brenner and Branscomb, 1979; Browne, 1979; Hollien, 1979; Shipp and Izdebski, 1981; Horvath, 1982). The evidence generated by the studies refuting the devices' lie detection capability, however, provides some support for the suggestion that the vocal signal may have acoustic properties that relate to psychological stress (Brenner, Branscomb, and Schwartz, 1979; Van Dercar, Greaner, Hibler, Spielberger, and Block, 1980).

A major difficulty encountered in laboratory stress studies is the need to determine if the subject is truly experiencing stress. Based on previous literature, several measures appear to be prominent indicators. Beatty (1982) summarizes evidence that eye pupil diameter can serve as a sensitive indicator of stress. Williams and Stevens (1981) reviewed evidence that elevation in voice frequency may accompany various emotional states.

An alternative method to a highly-controlled laboratory study is to perform acoustic analysis of taped speech samples recorded while individuals were undergoing extreme stress levels. Recordings such as the classic description of the Hindenburg disaster (Williams and Stevens, 1981) provide excellent research materials where high stress can be inferred from the situation and speech samples compared with those recorded immediately prior to stress onset.

This paper will describe our studies of acoustic correlates of phychological stress using both laboratory and field studies.

Experiment I

The first study employed a standard laboratory task varying workload demands on the subject. Several acoustic and physiologic measures previously associated with stress were samples simultaneously: F_O, frequency "jitter," pupil size, and EMG from the cricothyroid muscle (CT).

METHOD

Subjects were seven males, aged 22 to 33 years, free of any history of respiratory or laryngeal dysfunction. Hooked-wire electrodes were inserted to the left cricothyroid muscle of each subject using verification techniques previously described (Shipp, Fishman, Morrissey and McGlone, 1970). To establish a metric of EMG activity, subjects were instructed to relax the muscles of the shoulder, neck and face following which baseline CT activity was recorded. Next, subjects produced phonation near the top of their pitch range (falsetto). EMG output from the CT was integrated and averaged over three 300-msc windows at both the lowest (baseline resting) and highest (phonation) muscle activity. The lowest and highest EMG values from the three windows was used to establish the 0-100 metric by which muscle activity was quantified. All experimental EMG values were expressed as a percentage of this range.

Pupil size was tracked by a Gulf and Western (Whittaker) TV pupilometer, Model 831 that utilizes a camera with closeup lens to transmit an image of the eye to the instrument, which then performs a calculation of the pupil area and outputs a voltage proportional to pupil size. The pupilometer was calibrated over a one-volt range using targets of known diameter up to 4 mm. The pupilometer CRT image was continuously monitored to ensure the machine was tracking pupil size.

For the experimental task the subject sat at a table with his head stabilized against an individually-fitted nosebridge device so that the pupilometer TV camera could be focused on his right eye. The subject fixated his gaze on a low-level light blinking at 1 Hz approximately fifteen feet ahead. Subjects performed the number-repeating task of Kahneman and Beatty (1966). They listened to a free field presentation of a recorded talker saying the digits "nine" and "one" in strings of 2, 4, and 7 numbers. After each stimulus string presentation there was a two-second pause and then the recorded voice said, "Repeat." The subject then repeated the stimulus string at a rate of one/sec using the blinking light for timing. A series consisting successively of a 2, 4, and 7 number problems was presented for repetition five times. It was expected that the two-number string would represent the minimum anxiety state, and the subsequent four-number and seven-number task would represent increasing degrees of anxiety and stress associated with increased memory load. Previous literature reports that pupil size increases in direct proportion to the length of the required string (Beatty, 1982).

Output from the pupilometer, laryngeal contact microphone, EMG biological amplifier, and crystal-controlled timing signal generator were recorded simultaneously on a multichannel FM recorder. For analysis, the four data channels from each subject's procedure were written out on an optical oscillograph at 25 mm/sec to identify the two-second period between the stimulus tape's instruction to repeat and the subject's repetition of the first digit in the string (regardless of stimulus length). It was felt that this time period would incorporate the highest stress. Pupilometer values were obtained by a hand scoring system of averaging the DC signal during the critical two seconds and comparing these values to a matched baseline period of two seconds prior to stimulus presentation. Next, the subject's voice signal and accompanying timing signal were displayed on the oscillograph at 2000 mm/sec. Temporal measures of each F_O cycle and timing signal wave were obtained using a hand measuring technique accurate to the nearest 0.05 millimeter. Measures from the crystal-controlled timing signal showed variability in the tape recorder and/or oscillograph transport of up to 0.05 millimeters for each 100 Hz wave; therefore, measurements were made of each vocal period to the nearest 0.1 millimeter (0.05 msec).

An EMG value for each trial was obtained by integrating and averaging the muscle signal

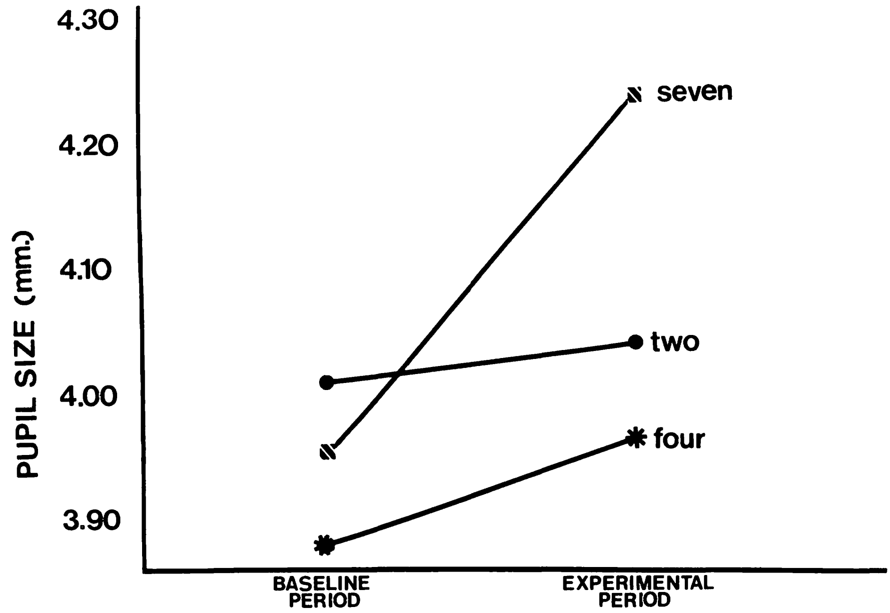


FIGURE 20-1

Mean eye pupil diameter averaged over all subjects in all trials. The baseline period is the 2-second interval prior to the first stimulus presentation, and the experimental period is the 2-second interval between stimulus and response.

during the subject's repetition of the first digit in the string since this utterance was temporally adjacent to the "critical" 2-second pause between stimulus and response.

RESULTS

Figure 20-1 shows a plot of eye pupil diameter size for the baseline period and during the critical period between the instruction to "repeat" and the start of subject voicing. These data are pooled for all subjects. The mean change in pupil size was smallest for the two-number strings (.04 mm), larger for the four-number strings (.08 mm), and largest for the seven-number strings (.29 mm) ($F_{(2,12)} = 17.1$, $p < .001$). There was no significant difference among the baseline scores alone. Pupil size data, then, suggest that the task did provide a condition consistent with some degree of increased psychological stress. This conclusion is supported by evidence of subjects' mean error rates, which were 0%, 3%, and 57% respectively for two, four, and seven-number strings. Based on this initial data analysis, detailed analysis was performed on only the first word from each subject's two-digit and seven-digit tasks, which were labeled "Easy" and "Difficult" conditions respectively. Items were selected for analysis that had unambiguous pupil measures and identical word in each condition.

Frequency jitter was calculated according to the formula:

$$\text{Jitter (\%)} = \frac{|P_i - P_{i+1}|}{\frac{P_i + P_{i+1}}{2}} \times 100$$

TABLE 20-1
Mean scores for seven subjects in response to an experimental manipulation of stress (task difficulty). JITTER respresents the proportion of jitter scores falling in the 0.00%-0.75% range.

	PUPIL DIAMETER (mm)	CT (%)	F_0 (Hz)	JITTER (%)
Easy Task	0.06	32	109	32.7%
Difficult Task	0.29	41	113	36.7%

where P_i and P_{i+1} are successive vocal periods. The denominator of this formula, which adjusts the score according to the average period of the two cycles being evaluated, normalizes the measurement on a cycle-by-cycle basis, which differs from the other reported jitter normalizing strategies. It was felt that the current formula, a more conservative strategy using short-timeframe adjustment, would make the jitter scores more immune to sudden intonation changes within the utterance. Inspection of the jitter data showed that approximately 95% of the reference vocal periods (P_i) were followed by periods that differ by less than 5%. Cycle-to-cycle temporal variations that exceeded the 10% value did not appear to be relevant to stress differentiation. These outliers, however, were valid perturbation and not instrumental or measurement artifacts. In the 0% to 5% range the responses to the "Difficult" task were associated with lower mean jitter scores than responses to the "Easy" task ($F_{(1,6)} = 3.8$, $p < .10$).

Table 20-1 summarizes the group mean scores in the "Easy" and "Difficult" condition. The statistic for the jitter score is the proportion of scores in the 0.00%-0.75% range, reflecting a cutoff point near the largest portion of the jitter distribution. When compared to the "Easy" condition, the "Difficult" condition was characterized by greater eye pupil diameter, higher F_0, increased CT activity, and lower jitter scores. An examination of individual subject means showed that the group trends applied to scores from six out of seven subjects for CT and Jitter, and five out of seven for F_0. Other measures approached, but failed, to reach significance at the .05 level.

Experiment II

The second study examined a methodological issue suggested by the results of Experiment I. To what extent is the accuracy of jitter measurement affected by timing variability in both the playback and display equipment used? Although Horii (1979) identified accuracy variables in jitter measurement, it seems possible that equipment can introduce random time variation that could well influence jitter scores significantly. For example, tape recorder flutter and wow characteristics can introduce jitter, which may be aggravated further if the recorded signal is played back for measurement on an oscillograph. It is possible that some of the disagreement reported on jitter measures results from variable amounts of instrumental jitter introduced during the extraction process. Experiment II investigated this issue.

METHOD

To evaluate the portion of the observed jitter attributable to the record/reproduce system, it was necessary to know the duration of every cycle in the original wavetrain. Rather than generate signals with a carefully specified amount of perturbation, a jitterless, steady-state test signal was produced on a Hewlett-Packard Function Generator, Model 3312A. Signal stability was verified by using a positive axis crossing of the test signal to trigger an oscilloscope beam (Tektronix 502A). The scope sweep speed was adjusted to display a single cycle of the test waveform (both sine and square waves) on the screen. To determine the period variability of the test signal, the time base was magnified by factor of 20. With the scope trigger anchoring the beginning of the sweep, the end of the trace could be examined for any observable movement along the X-axis. Estimates of the cycle boundry were easily made to the nearest millimeter; therefore, jitter values as small as 0.05% could be observed. The test signal, as far as could be determined, was without jitter; therefore, any jitter measured from oscillograms was introduced at one or more points in the record/reproduce system.

Measurements of each test signal cycle were made from oscillograms produced on a Honeywell Visicorder, Model 1108 at a chart speed of 2000 mm/sec. Two 2-second segments of the test signal were measured. The first oscillogram was produced with the signal generator feeding the test signal directly to the oscillograph; the second was the signal played back from a recording on the FM tape recorder (Honeywell 7600). Measurements were made to the nearest 0.1 mm, which provided a measurement accuracy of 0.5% per cycle. An average jitter value was calculated and expressed as a percent according to the jitter formula of Experiment I.

RESULTS

Calculations indicated that the oscillographic reproduction alone led to an average jitter value of 1.2%. When oscillograms were made from recorded signals, the measured jitter value increased to an average of 1.6%.

Since the amount of jitter in the test signal was determined to be less than 0.05% based on observing a magnified signal on an oscilloscope, the recorded signal was observed in the same manner and showed changes in duration of approximtely 0.5% (1.0 cm). Exact quantification of jitter in the recorded signal was not possible from the oscilloscope trace, but the observed "average" jitter is remarkably close to the 0.4% value derived from subtracting the difference between the tape playback jitter measures and the one directly from the function generator. This percentage figure is also remarkably close to the flutter specifications for the Honeywell 7600 FM tape recorder: 0.45% @ 7.5 IPS. It appears, then, that instrument error is of a magnitude that can add random noise to jitter measurement. As a result the absolute jitter scores from Experiment I become suspect, but the relative difference of scores between the two conditions was judged as valid.

In Experiment III it was possible to explore a mechanism for reducing this error by using a timing signal recorded on a parallel channel to the voice signal. This timing signal was used to correct the jitter scores and reduce any time distortion introduced by the instrumentation.

Experiment III

The speech material used in the third experiment was from recordings of crew conversation prior to aircraft crashes. There were three such recordings, with each providing speech samples from a crewmember before and during the emergency situation.

One advantage of examining aircrash tapes is the possibility of studying responses under conditions where the presence and nature of stress seems beyond question. Any trends developed in the laboratory for identifying acoustic correlates of stress should almost certainly be strengthened in this augmented stress situation. The advantage of aircraft recordings over, say, emergency calls to police or fire departments is that in aircraft communications there are usually exemplars of an individual's "usual" voice production before the stressful incident occurred; therefore, valid comparisons can be made between conditions using each voice at its own normal control.

Practical difficulties in studying aircrash material involve the recording quality and availability of appropriate material to analyze. When tapes are recorded on the ground (control towers or flight service centers), recording quality is usually acceptable, but may contain little or no speech material during an emergency. During an emergency, the cockpit personnel are usually too busy dealing with the situation to transmit anything but minimal information by radio. Further, cockpit tapes, recorded on board commercial passenger aircraft, typically have poor signal-to-noise characteristics because of high level background noise and the location of the omnidirectional cockpit microphone well above the heads of the crew.

METHOD

We obtained three tape recordings that appeared appropriate for establishing a data base. Analysis was performed on only identical words or phonetic segments produced by the same talker in the two conditions.

Tape 1, recorded in a control tower, contained routine communications and then an emergency announcement from a commercial passenger aircraft involved in a midair collision during a landing sequence. The pilot alerted the tower to the obvious crash situation and identified his aircraft. This identification, three syllables long, was the basis for comparison with two prior routine communications.

Tape 2 was a cockpit voice recording covering the final thirty minutes of a commercial passenger aircraft involved in an unsuccessful landing under poor visibility conditions. The aircraft first impacted the ground in a landing attitude and crashed 30-seconds later. The word "Oh" spoken by a crewmember after impact was compared with his previous air-to-ground transmissions in which the phoneme /o/ was used to identify the integer, "zero" in the aircraft's call number. These four /o/ transmissions occurred 20, 19, 14, and 1.5 minutes before impact. The final /o/ was spoken 11 seconds before the crash.

Tape 3, recorded from a flight service center, contained routine and emergency communications involving a helicopter flight. The pilot initially called requesting routine flight information, and, in the process, identified the aircraft's name, number, destination and location. Twenty-two seconds later, a sudden rotor failure (failure of the principal blade) caused an unavoidable crash situation. The pilot radioed an immediate emergency alert, repeating the previous identification and location information. This replication allowed us to compare 14 identical syllables in each condition. Total time from first emergency transmission to crash was 38 seconds. Recording quality was good, although normal helicoter mechanical vibration introduced audible "tremor" in the pilot's voice during all transmissions in the two conditions. Tape 3 was especially appropriate for analysis because of the substantial amount of common spoken material before and after the emergency, and the short time duration between the routine and emergency transmissions. An additional advantage of Tape 3 was the presence of a timing signal recorded simultaneously with the pilot's transmissions. (The flight service center routinely recorded a timing signal on a parallel recording channel.) The results of Experiment II suggested that even excellent playback and display equipment might produce random timing errors which could be on the order of magnitude of frequency jitter analysis. To reduce these possible sources of random noise, the 100 Hz timing tone available

TABLE 20-2
Experimental III: Proportion of jitter scores in the 0.00%-0.75% range for the routine and emergency aircraft transmissions (number of observations shown in parentheses).

	ROUTINE	EMERGENCY
Tape 1	18.9%	39.0%
	(37)	(41)
Tape 2	33.3%	46.2%
	(36)	(13)
Tape 3	20.7%	27.6%
	(402)	(508)

in Tape 3 was scored in parallel with the speech signal. Any observed jitter in the speech signal was corrected proportionately to timing signal jitter measured at that point.

PROCEDURE

Voice samples were analyzed for jitter using the procedures described in Experiment I.

RESULTS

For all three tapes, average F_0 increased significantly in the emergency situation: Tape 1, from 95 to 148 Hz; Tape 2, from 101 to 123 Hz; Tape 3, from 149 to 264 Hz. Similar increases appeared for standard deviation of F_0: Tape 1, from 12.9 to 23.7 Hz; Tape 2, from 12.6 to 63.8 Hz; Tape 3, from 30.1 to 66.0 Hz.

An examination of the jitter results indicated that trends present in Experiment I also appeared at Experiment III. As was found in Experiment I, the jitter distribution in all aircraft speech samples appeared to concentrate in the 0% to 7% range. Jitter scores also tended to be lower for emergency communications. As shown in Table 2, the proportion of jitter scores in the 0.00-0.75% range was larger for the emergency than for the routine communications (Tape 3: $X^2 = 5.80$, $p < .05$).

DISCUSSION

The trends noted in Experiment I tended to be consistent with those in the field setting of Experiment III. That is, situations in which increased stress can be measured or implied are characterized by increased eye pupil size (Beatty, 1982), increased F_0 (Williams and Stevens, 1981; Streeter, Macdonald, Apple, Krauss and Gallotti, 1983), and decreased vocal jitter (Lieberman, 1961). Experiment I used all of these measures simultaneously and found, as have previous researchers, a tendency towards identifying acoustic correlates of stress, but with sufficient variability in the experimental data to prohibit establishing statistical validation.

Another aspect of the findings from the first study that warrants attention is the correspondence between increased CT activity and decreased frequency jitter. There may be no causal relationship, yet it is tempting to attribute the known mechanical consequences of CT contraction to a change in the regularity of vocal fold vibration. It is possible that the inherent anatomic dissimilarity (no matter how small) between the paired vocal folds during speech contributes to the often-observed quasi periodicity in successive vibratory cycles; whereas, with CT contraction stiffening and thinning the tissues, the folds become more similar in configuration and vibrate in a more periodic manner. Since there is not an absolute one-to-one relationship between CT activity and fundamental frequency, a slight increase in CT activity during low frequency speech may not necessarily be reflected in upward F_O shifts, but rather, may align the folds in a more symmetrical fashion thereby decreasing the "normal" jitter percentage for that individual.

The relatively small cycle-to-cycle differences found in conditions of "normal" phonation suggest that jitter percentage greater than 10% may reflect another mechanism altogether, perhaps including transients in the afferent signal along the recurrent and superior laryngeal nerves. Such isolated "bursts" of CT activity were found by Baer (1979) in his study showing momentary voice frequency perturbations within 50 milliseconds following such EMG bursts.

In Experiment III, it should be noted that the results of Tape 3 did not appear to be affected by an audible tremor in the pilot's voice caused by the helicopter vibration. The normalizing procedure for jitter measures employed in the study was sufficiently precise to extract short-time jitter information despite potentially confounding background effects on voice such as vibratory-induced vocal tremor. It might also be noted that the pilot's voice in Tape 1, though slightly elevated in F_O, was not markedly different perceptually in the two conditions. This ability to "control" the more obvious speech features (F_O and word rate) during emergencies is similar to findings reported by Kuroda, Fujiwara, Okamura and Utsuki (1976).

An interesting trend appears in the speech analysis from Tape 2. Jitter percentage scores tended to decrease in direct relation to what we presume to be three levels of stress. The first, or lowest, stress level was during normal flight conditions. Speech samples analyzed during this period were from routine radio transmissions 20, 19, and 14 minutes pre-impact. Jitter percentage scores for these segments were 1.84, 1.90, and 1.86 respectively. The second stress level began three minutes before impact when the cockpit recordings indicated that the crew identified, and then became increasingly concerned about some equipment malfunctioning during the landing approach. During this three-minute period the talker radioed an acknowledgment of tower communications that incorporated the aircraft number. Jitter percentage for this utterance was 1.53. The third stress level began after initial ground contact when it was apparent to the crew that a fatal crash was imminent. The utterance analyzed was spoken by a crew member 19 seconds after ground impact. Jitter percentage for this segment was 1.47.

The results of Experiment II indicate a direction to improve the precision of jitter measurement and reduce unwanted background noise. If uncorrected, the effects of instrument noise are likely to be on the order of magnitude of jitter scores and would tend to inflate the absolute value of jitter and obscure any experimental effects present. The results of Experiment II argue for the use of a reference timing signal recorded at the time of the original speech, which can be used to indicate and mathematically correct timing variability introduced by laboratory instrumentation. Such an approach was tested satisfactorily in Experiment III. An even more exact solution, under consideration by our laboratory, is to use the timing signal to pace an analog-to-digital conversion of the speech signal prior to scoring. Such a procedure would automatically convert the speech into the correct timing sequence at which it was spoken removing all unwanted variance introduced by intervening instrumentation. Such procedures would add significantly to the accuracy of jitter measurement and to its value as a research measure.

Together, then, the two experiments provide additional support for frequency "jitter" as a measure of psychological stress. Detailed analysis of truly stressed speech using automated procedures on a voluminous speech sample may resolve the ambiguity of stress correlates in the speech signal. It seems possible that frequency "jitter" might eventually develop into a measure of stress with practical properties that would add greatly to a measurement battery in conjunction with previous voice measures.

ACKNOWLEDGEMENTS

This research was supported by the VA Research and Development Service. We thank the Federal Aviation Administration for its cooperation in providing several recording employed in Experiment III. For the loan of the pupilometer, we gratefully acknowledge Dr. Gary Pooch, Director, Man-Machine Systems Design Laboratory, Naval Postgraduate School, Monterey, California.

REFERENCES

Anon. "Psychological stress evaluator," (Sales Brochure), Dektor Counterintelligence and Security, Inc., Savannah, GA.

Baer, T. (1979). "Vocal jitter: a neuromuscular explanation," in V. Lawrence and B. Weinberg (Eds.) *Trans. of the 6th Symp. on the Care of the Prof. Voice, Part I.* The Voice Foundation, New York, 19-22.

Beatty, J. (1982). "Task evoked pupillary responses, processing load, and the structure of processing resources," *Psych. Bull.* 91, 276-292.

Bell, A.D., Ford, W.H. and McQuiston, R. (1976). U.S. Patent 3,971,034.

Brenner, M. and Branscomb, H.H. (1979). "Psychological stress evaluator: technical limitations affecting lie detection," *Polygraph* 8, 127-132.

Brenner, M., Branscomb, H.H. and Schwartz, G.E. (1979). "Psychological stress evaluator - two tests of a vocal measure," *Psychophysiol.* 16, 351-357.

Browne, N.W. (1979). "Voice 'Lie Detector' Decried as Unscientific," *New York Times*, Nov. 13, C1.

Feder, B. (1982). "Court test for voice analyzer," *New York Times*, Jan. 5, D1.

Hollien, H. (1979). "The case against stress evaluators and voice lie detection," Paper, Conf. of the Acad. for the Forensic Appl. of the Comm. Sciences, Gainesville, Fl.

Horii, Y. (1979). "Fundamental frequency perturbation observed in sustained phonation," *J. Speech Hear. Res.* 22, 5-19.

Horvath, F. (1982). "Detecting deception: the promise and the reality of voice stress analysis," *J. Forens. Sci.* 27, 340-351.

Kahneman, D. and Beatty, J. (1966). "Pupil diameter and load on memory," *Science* 154, 1583-1585.

Kuroda, I., Fujiwara, O., Okamura, N. and Utsuki, N. (1976). "Method for determining pilot stress through analysis of voice communication," *Aviation, Space and Environ. Med.* 47, 528-533.

Lieberman, P. (1961). "Perturbations in vocal pitch," *J. Acoust. Soc. Am.* 33, 597-603.

Ludlow, C.L., Coulter, D.C. and Bassich, C.J. (1982). "Relationships between vocal jitter, age, sex and smoking," *J. Acoust. Soc. Am. Suppl.* 1 71, Z1

McNeice, C. and Cota, R. (1980). "Build a vocal 'truth' analyzer," *Pop. Electron.* April, 66-71.

Murry, T. (1979). "Vocal jitter in sung and spoken vowels," in V. Lawrence and B. Weinberg (Eds.) *Trans. of the 6th Symp. on Care of the Prof. Voice, Part I.* The Voice Foundation, New York, 11-18.

Ramig, S.L. and Ringel, R.L. (1983). "Effects of physiological aging on selected acoustic characteristics of the voice," *J. Speech Hear. Res.* 26, 22-30.

Scherer, K.R. (1981). "Vocal indicators of stress," in J.K. Darby (Ed.) *Speech Eval. in Psychiatry*. Grune and Stratton, New York, 66-71.

Shipp, T., Fishman, B.V., Morrissey, P. and McGlone, R.E. (1970). "Method and control of laryngeal EMG electrode placement in man," *J. Acoust. Soc. Am.* 48, 429-430.

Shipp, T. and Izdebski, K. (1981). "Current evidence for the existence of laryngeal macrotremor and microtremor," *J. Forens. Sci.* 26, 501-505.

Streeter, L.A., Macdonald, N.H., Apple, W., Krauss, R.M. and Gallotti, K.M. (1983). "Acoustic and perceptual indicators of emotional stress," *J. Acoust. Soc. Am.* *73*, 1354-1361.

VanDercar, D.H., Greaner, M.A., Hibler, N.S., Speilberger, C.D. and Block, S. (1980). "A description and analysis of the operation and validity of the PSE," *J. Forens. Sci.* *25*, 174-188.

Williams, C.E. and Stevens, K.N. (1981). "Vocal correlates of emotional states," in *Speech Eval. in Psychiatry*, J. Darby (Ed.) Grune and Stratton, New York, 221-240.

DISCUSSION

C. LARSON: I have a couple of comments. First of all, I'd like to make some suggestions as to some possible neurological correlates of the work you reported on the correlation of pupil dilation. There is an area of the brain, namely the anterior cingulate gyrus, that has been known for years to be involved in pupil dilation. The same area of the brain is important in vocalization. If this area of the brain is eliminated, then sometimes vocalization is affected. This area of the brain projects to the periaqueductal gray. We are getting evidence that indicates that it projects to the larynx and controls muscles of the larynx. And so this could be a possible neurological pathway involved in this voice stress and jitter. I think your relating the cricothyroid activity to jitter is appropriate because as you increase the level of activity in the cricothyroid, more motor units would be reaching fused tetanus, and this should, if what I suggested yesterday concerning the relation between voice oscillations and single motor units is valid, increase the level of muscle activity and should cause a reduction in jitter.

M. BRENNER: Interesting. There is a saying in psychophysiology that all measures tend to follow this inverse relationship. That is, for heart rate and pupil size and I believe G.S.R., the data are more random when you're relaxed. As you become tense they tend to get more regular. So if there is some physiological correspondence, it's very interesting.

21 ▪ THE FIBERSCOPE: FLEXIBLE AND RIGID FOR LARYNGEAL FUNCTION EVALUATION

Wilbur James Gould

The past years have seen the development of new methods of vocal fold and laryngeal assessment.

The flexible fiberscope offers the advantage of lesser distortion of the hypopharyngeal and laryngeal area during examination. There are still considerable problems that have not been solved. Firstly, there is the distortion dependant upon which nostril was used. The distortion is either from right or left angulation. In turn, the angle used allows one to see either side more efficiently in a large number of cases. The further the instrument is from the vocal fold, the greater the distortion. In some instances the false vocal fold could appear larger than it truly is. Even motion can be distorted. Also, if anesthesia is used, the reflexogenic character is altered. If no anesthesia is used there is reflex stimulation up to the point of spasm caused by the apparatus itself. The other factor that must be borne in mind is the distance size distortion. It is very hard or rather impossible to maintain a fixed point to judge vertical motion during change of vocal function. The farther away you are from the vocal fold, the distance imaged has an exaggerated view. Therefore, the distance may be far less than is actually the case. The distinct advantage, however, is that it is possible to visualize almost all cases.

Stereoscopic examination of laryngeal areas offers the opportunity to give more accurate measure of structure size and distance. The flexible fiberscope can be used in both nostrils simultaneously in a limited number of cases, but still is a worthwhile form of examination in those where it is possible. The major problem has been in the handling of the distal ends which must be coupled to allow the stereoscopic effect. General usage demands an easily uncoupled mechanism rather than a method that would create difficulty upon attempt of separation. This is necessary in the event of any problem of discomfort or distress on the part of the individual being examined. More recently magnetic coupling has allowed easy separation and I look forward to this being a further tool that we will have readily available for clinical examination. The stereoscopic view will offer us a new milestone in the ability to evaluate the glottic apparatus. Dr. Kakita of Kurume, Japan, in an earlier paper, had demonstrated his stereoscopic laryngeal mirror which allows us to measure vertical aspects of motion as well. This is a means of evaluation by use of ultraspeed motion pictures. However, it does not allow more function analysis in view of the fact that the tongue is pulled and there is distortion by the oral route of instrumentation requiring depression of tongue and pulling of tongue as well.

The rigid fiberscope, such as the Wolf or Ward-Berci, offer beautiful images which can be very useful for close examination of detail. You could, as we show in one demonstration, see minor motion changes of sections of the vocal folds. However, here again as we all know, the distortion of a large object in the mouth with tongue depression or pull precludes a normal type of vocal function evaluation. It is also troubled by the same reflex problems we discussed for the flexible instrument.

It is, in my experience, better to use both methods for picture taking. The method selected depends upon the purpose of the examination. The flexible fiberscope serves as a more general instrument and allows more function analysis and the rigid fiberscope orally introduced offers more detail of the vocal folds themselves.

The microscope, when used in conjunction with a laryngeal mirror, is a useful instrument for detail examination in a more limited population. It is possible to do it comfortably in about 60-70% of patients when honestly assayed. The use of the Wolf rigid fiberscope allows closer examination of the larger percentage of patients although not at the magnification of the microscope in its ideal state.

The stroboscopic examination using the rigid fiberscope is satisfactory visually. It still does not offer enough light to take pictures of the motion by any practical television apparatus and, particularly, this is true when there is inflammatory reaction and light absorption. The light is just not sufficient. The microscope, again with the mirror, can be coupled to stroboscope and in the limited percent this examination can be done in quite excellent detail.

In view of the fact that the purpose of this paper is to stress visualization of function aspects of vocal production rather than inspection alone, we demonstrate first vocal function as observed by the flexible fiberscope. Dr. Robert Feder's television film of the multi adjustments of the supraglottic tissues in a mimic amply demonstrate this ability to see what other devices cannot demonstrate. The comparison of the visualization of the larynx by the two methods, the flexible and the rigid, showed well in a television film of a professional singer. Here the supraglottic tissues are well demonstrated as is the vertical position of the larynx. The rigid fiberscope shows the entire length of the vibratory fold and phonation. The mirror view rarely demonstrates the anterior area adequately. Any motion abnormality can be reasonably assayed by the use of the two methods as we demonstrated.

I would like to gratefully acknowledge Jo Estill, voice teacher, New York City; George Ramos, Vocal Dynamics Lab., Lenox Hill Hospital, New York City; and Dr. Robert Feder, Cedars of Lebanon, Los Angeles for their assistance in this effort.

DISCUSSION

R. FEDER: Jim, I've noticed that same asymmetry of the arytenoids with the crossover. What do you make of that?

W.J. GOULD: I don't know. I'm presenting it to the foremost voice scientists in the world, and I want them to tell me.

P. KITZING: Well, what I'm going to do now is very difficult for me. I feel like David meeting Goliath, only there can't be a question of a fight between friends. But I think there can be a question of a second-best method fighting the best method. What I feel in this connection is that the method of using an operating microscope to view the larynx with an indirect laryngeal mirror really gives you much better resolution, and that you really have better information if you do such an examination. Also, it is better to such a degree that if you have a concern about vocal fold vibration and blurring of the image, then you want to have your stroboscope in place so you can obtain a real sharp picture of the vocal folds. I think there can never be a question of this with the endoscope. I also work with both the endoscope and with the fiberscope. But I think if you want to have a real appropriate examination of the larynx, you should bring in the indirect microscope.

W.J. GOULD: I do have the indirect microscope. I find that in my hands, it's successful in about 60% of the cases, and I can't be satisfied with that. The majority of the cases that I get, where there really is pathology that I'm interested in looking at, I don't get a good laryngeal picture, at least in what I see. Now I know Dr. Van Lawrence has been using the microscope, and he has come to the same conclusion, i.e., it's about 60 to 70% successful. Now in the case of the flexible scope, it's close to 100% successful. And in the case of the Wolf, it's 90 to 95%, except for the one I mentioned yesterday. I think that's the reason I have to go to an instrumentation that allows me a higher and more uniform success rate. The microscope is handy. I do use it when indicated. I don't want to give that up, and I would not give up a regular laryngeal mirror. The flexible fiberscope, I think, offers us a medium that is close to perfection in the sense of being able to see everyone. There is no other source that I know of that can offer anything equal to that.

P. KITZING: Could there be a difference between European patients and American patients? (Laughter)

W.J. GOULD: As a matter of fact, no, I think that there may be a difference in the amount of anesthesia one would use and so forth. Now of course, if I were to completely anesthetize, and use a stronger agent and some Valium, then I could bring the laryngeal mirror up to a much higher percentage, but that's not the way I prefer to do it. I think that's the answer to it.

R. SCHERER: Dr. Gould, would you make a comment, please, on the advisability of using the stroboscope during laryngeal examination?

W.J. GOULD: Yes, I believe that the stroboscope is a very important instrument to use. Right now, friends of mine are trying to make stroboscopes to bring them within the practical realm of the average practitioner. Right now, the average practitioner would have to pay $15,000 to $30,000 for stroboscopes. That is the present going rate. Hopefully, there will soon be one or two that will bring the price down to around $2,000. If that becomes the case, then that would be a practical instrument. Today I was talking about flexible and rigid fiberscopy. But that doesn't mean that I don't believe in or do not want to use the other instrumentations.

22 ▪ THE FIBERSCOPE: ANALYSIS AND FUNCTION OF LARYNGEAL RECONSTRUCTION

Stanley M. Blaugrund, Wilbur J. Gould, Shinzo Tanaka and Kazutomo Kitajima

ABSTRACT

Points to be made in this presentation primarily concern themselves with clinical observations of the voicing larynx as viewed through the flexible fiberscope. Larynges which were ablated by several techniques of partial laryngectomy are presented and pertinent observations made as they appear on videotape. The operation performed was a three dimensional one designed to obtain tissue margins in accordance with adequate cancer surgery. In most instances, this involved the technique of vertical hemilaryngectomy, an operative procedure commonly performed in patients with T1 and T2 glottic cancers (Som and Arnold, 1960). Various types of laryngeal reconstructive procedures were utilized depending upon the nature of the defect created by the surgery (Ogura and Biller, 1969; Bailey, 1966; Blaugrund and Kurland, 1975). In instances where the arytenoid cartilage could be saved, pedicled flaps of local strap muscle were used in reconstruction of the vocal cord. Where the arytenoid cartilage could not be spared, an arytenoid replacement procedure utilizing a composite cartilage-muscle flap, was employed. Clinical observations of voicing during the postoperative period revealed that compensatory supraglottic sphincterization was the dominant method of achieving voice.

INTRODUCTION

Flexible fiberoptic instrumentation provides the laryngologist and voice scientist with a unique opportunity for examining the voicing larynx in its natural state unencumbered by techniques that produce distortion, such as anterior displacement of the tongue and epiglottis during indirect laryngoscopy or immobilization and anterior displacement, as in indirect laryngoscopy. In subjecting a group of patients having undergone vertical partial hemilaryngectomy to a critical analysis of their voicing mechanisms as viewed on videotape using the flexible fiberscope, several interesting findings are noted.

These include:

1. Hypertrophy of the true and false vocal cords, arytenoids and aryepiglottic folds of the contralateral side.
2. Sphincteric closure of the supraglottis sufficient to achieve phonation through the action of the remaining arytenoid cartilages, aryepiglottic folds bilaterally, and the epiglottis (Figure 22-1).
3. The degree of sphincteric action is proportional to the size of the initial surgical defect and the time elapsed after surgery.
4. Reconstructive procedures designed to provide bulk to enhance voicing may not always achieve the desired result due to atrophy of tissues used in the reconstruction.

The significance of these findings are conjectural, though worthy of consideration. One item of interest might involve a rethinking of post-operative voice therapy in these patients. In this context, more emphasis than is now being given should be directed toward supraglot-

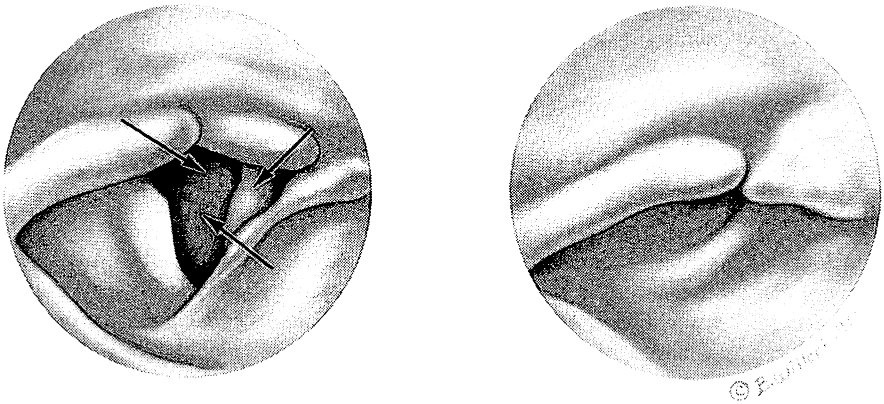


Figure 22-1. Sphinteric closure of supraglottis after hemilaryngectomy.

tal structures in the attainment of voice. In addition, these findings may offer some clue with respect to the elusive fundamental factors concerned with laryngeal vibrations.

MATERIAL AND METHODS

The operative procedure of vertical hemilaryngectomy involves subtotal excision of the ipsilateral thyroid cartilage along with contiguous intralaryngeal structures including the subglottis immediately above the cricoid cartilage, the true vocal cord, ventrical, false cord and arytenoid cartilage (when the vocal process is involved by disease). Where possible, in smaller lesions, the arytenoid cartilage can be preserved, in which case pedicled skin flaps or bipedicled strap muscle flaps are interposed. The defect created by vertical hemilaryngectomy thus constitutes a loss of intralaryngeal structures encompassing the true and false cords vertically and may extend in the horizontal plane to include the entire glottis leaving only one arytenoid. Reconstructive techniques attempt to introduce bulk through the use of local mesodermal flaps or free tissues and to provide epithelial resurfacing using adjacent pedicled mucous membrane or skin. Readers are referred to the original descriptions of these operations for details not pertinent to this paper.

Seven patients comprise this series, all of whom underwent vertical partial laryngectomy. Two patients received standard vertical partial procedures which spared the arytenoid cartilages and were reconstructed using local pedicled muscle flaps (Ogura and Biller). One of these patients for whom vocal dynamics data is available, phonated a sustained vowel with 335 cc/sec. of mean flow rate and exhibited 15 sec. of maximum phonation time. (Normal MFR = 46 to 222 cc/sec. and MPT = 15 to 62 sec.) He could not change pitch, and the noise component was more prominent than the harmonics in his sonogram.

Another patient, underwent total vertical partial hemilaryngectomy i.e., including the ipsilateral arytenoid. This individual was reconstructed using the cartilage-inferior constrictor muscle pedical flap (Blaugrund). The results of aerodynamic studies for this patient revealed a mean flow rate of 262 cc/sec. and maximum phonation time of 14 secs. The data thus shows that although there is documented evidence of mild to moderate airloss during phonation in these patients, their ability to engage in casual conversation is not significantly impaired. Another patient underwent vertical partial hemilaryngectomy with preservation of the ipsilateral thyroid ala and reconstruction using a bipedical strap muscle flap (Bailey). Two additional patients underwent similar vertical partials of varying extent and were reconstructed using local pedicled skin flaps (Conley). The seventh patient underwent simple cordectomy via thyrotomy 35 years previously. This case is included since it demonstrates how little from the standpoint of compensation is necessary when relatively small amounts of tissue are ablated.

A videotape demonstration and analysis of these seven patients follows.

CONCLUSION

Videotape analysis of postoperative patients having undergone vertical hemilaryngectomy, graphically demonstrates that there is considerable activity involving the supraglottic sphincters. For it would seem that not only do these mechanisms come into play from the standpoint of providing protection for the lower airway from aspiration, but they also contribute in a significant way to voice production as well.

The arytenoid cartilage/aryepiglottic fold/epiglottis complex, appears from these observations, to have the capacity to compensate for the production of speech from whatever defect created by surgery. This aspect of voicing postoperatively is not readily detectable without the use of the flexible fiberscope for reasons previously mentioned.

Hopefully these observations will contribute to a better understanding of voicing mechanisms in the hemilaryngectomized patient.

The authors gratefully acknowledge the efforts of Mr. George Ramos in compiling the data and in editing the video tapes for this presentation.

REFERENCES

Bailey, B.J. (1966). "Laryngectomy and laryngoplasty," *Trans. A.A.O.O.* 70, 559.

Blaugrund, S.M. and Kurland, S.R. (1975). "Replacement of the arytenoid following vertical hemilaryngectomy," *Laryngoscope* 85/6, 935-941.

Ogura J.H. and Biller, H.F. (1969). "Glottic reconstruction," *Laryngoscope* 79, 2181.

Som, M.L. and Arnold, L.M. (1960). "Hemilaryngectomy-surgery for cordal carcinoma with posterior extension," in *Otolayryngology*, Vol. 5, G. Coates, H.P. Schenk and M.V. Miller (Eds.), Prior, Hagerstown, Md. Chap. 7, 31-58.

DISCUSSION

R. FEDER: In that first case, where we heard that breathy voice and saw that gap, how long post-op was that?

S. BLAUGRUND: About four months, I think.

R. FEDER: Would you consider putting some teflon in there after another 4 to 6 months?

S. BLAUGRUND: I think I'd rather wait a while, at least based on the findings that we've seen. I think the longer you wait, the more chance there is of the patient having some compensatory improvements and augmentation of the voice. And I think that putting teflon in might alter this.

R. FEDER: If she were the same in a year, would you wait 18 months?

S. BLAUGRUND: Yes.

QUESTION: How long after surgery do you think rehabilitation should be begun? And is there any period after which you think you might not be able to give some voice rehabilitation? For example, if you don't do any sort of voice rehabilitation for a couple of months, say three or four months after surgery, is there a time after which you would never be able to use therapy?

S. BLAUGRUND: Yes, I think so. We try to begin therapy as soon as possible. As a matter of fact, we have our speech pathologist see the patient before the operation. We have a team at Lenox Hill and we meet every Thursday afternoon. We discuss the patients that are coming in. The patients that are going to have this type of surgery are seen beforehand and then followed as soon after surgery as possible.

QUESTION: How long is that?

S. BLAUGRUND: That's a week to ten days before surgery and for several months and years after surgery, depending upon the patient's response and compliance.

C. LARSON: You said that at least in one of your cases it looked like the false vocal folds had enlarged?

S. BLAUGRUND: There was one case that had shown that, yes.

C. LARSON: There are, apparently, a few muscle fibers in the false vocal folds. Are you suggesting that those muscle fibers became more functional through this operation?

S. BLAUGRUND: Yes, I think they undergo enlargement, and I think that applies not only in the larynx, but also to patients who have had a large ablated procedure like in a supraglottic laryngectomy. These patients have great difficulty swallowing in the first few weeks after surgery. I think it's because of a compensatory muscular hypertrophy that they are finally able to swallow. This is conjectural, but I feel that it is definitely the case.

23 ▪ THE RIDDLE OF THE “MIDDLE” REGISTER

Harry Hollien and Carol Schoenhard

INTRODUCTION

Vocal Register Definitions

It is now reasonably well established that operations appropriately identified as vocal registers exist in the human voice. Generalized (if not operational) definitions can be seen in the literature; one of the earliest being that of Garcia (1840) who indicated that a voice register is ... “a series of succeeding sounds of equal quality, a scale from low to high produced by the application of the same mechanical principal, the nature of which differs basically from another series of succeeding sounds of equal quality produced by another mechanical principal.” In the years to follow, many vocal pedagogists, laryngologists and phoneticians offered like definitions; virtually all described voice registers in much the same terms as did Garcia (see for example, Appelman, 1967; Fields, 1970; Hollien, 1974; Large, 1972; Nadoleczny and Zimmerman, 1938; Preissler, 1939; Ruth, 1963; Vennard, 1967). It would be of little value to list any portion of these hundreds of definitions or labels (see Morner *et al*, 1964, for examples of terms) as they add little to the rather simplistic concept articulated by Garcia. Our use of the term simplistic is not intended to be judgmental or negative. Rather, it indicates that Garcia’s definition, while a good one, does little but scratch the surface of the issue. That is, it must be asked: what are the scales to which Garcia refers? And...what series, what qualities, what mechanisms are involved? Admittedly, Garcia implies that the fundamental frequency level of the sung tone is one of the controlling elements. However, until all of the questions are answered, his definition must necessarily be viewed as superficial or, at least, incomplete.

On the other hand, there are a number of vocal pedagogists who have taken the position that voice registers do not exist within the singing voice; a number of references/arguments could be cited in this regard (see Fields, 1970, for example). However, Johnson (1982) articulates the argument succinctly when she suggests: 1) that it is only the untrained singer who distorts productions in such a manner that unrelated (register) sounds are produced, 2) that many great singers developed their voices without even being aware of the concept of registers, 3) that “smoothness of scale and tone” more functionally relate to good development of voice rather than does training “pieces” of the voice and 4) that the so called registers could be the result of illusions based on singers feeling vibrations in their chests (or head) when they sing certain frequencies.

There is little doubt but that successful singers — at least those trained in the classical, “western”, opera or concert mode — are able to conceal register differences when they sing. Indeed, while Sundberg (1982) agrees with Johnson to some extent (as do we), he points out (as do others), that some forms of singing depend on a singer’s ability to covary “articulation, subglottic pressure and formant frequencies” with phonatory frequency. Specifically, Sundberg stresses that register usage can be an important part of technique in certain types of singing. That such instances exist is conceded by Johnson. However, she argues that “distortions” of this type can be, and probably are, dangerous to the singer.

The cited controversies — occurring amongst scientists, scholars, teachers and practitioners in the field of vocal music — led Hollien and his associates (1966, 1968, 1974, 1976, 1982) to attempt to provide new perspectives for the study of voice registers. First, a series of experiments on voice were carried out; nearly all made contributions to the issue of voice registers even though this purpose often was only a secondary one. Further, Hollien noticed that few phoneticians experienced any real difficulty in their conceptualization of voice registers. Indeed, except for a controversy which concerned the nature (pathology/non-pathology) of the vocal fry register [this controversy was soon resolved by Hollien, Moore, Wendahl and Michel, 1966], few phoneticians disagreed in any major way as to the nature, boundaries and/or functioning of vocal registers. Moreover, it was noted then — and it should be noted now — that many singers could produce register-related sung tones which were perceptually identifiable whenever they were requested to do so.

These relationships led to a first-order attempt to refine or modify Garcia's definition and it was suggested that a voice register is a "series or range of consecutively phonated frequencies which can be produced with nearly identical voice quality...that there should be relatively little overlap between adjacent registers and that, to be a vocal register, the mechanism should be laryngeal in nature" (Hollien, 1974; Hollien, Gould and Johnson, 1976). Since, this definition was as skimpy as was Garcia's, its authors set out to identify and operationally establish those registers they could on the basis of the available (and relevant) perceptual, acoustic, physiological, aerodynamic and neurological evidence. Attempts to provide such operational definitions are considered important for, as Titze (1980) points out, little will be accomplished unless these entities are described with respect to as many levels and dimensions as possible "starting...(with) the neuomuscular level... to the bio-mechanical level, to the kinematic level (where you observe motion) to the aerodynamic and acoustic level, and finally, to the perceptual level." But...how far have we come in the realization of Titze's guidelines?

Speaking vs. Singing Registers

Recent attempts by the CoMeT (Collegium Medicorum Theatri) Committee on Vocal Registers[1] to develop appropriate models on the basis of the approaches cited above, have led them to take a conceptual position that appears to have been long overdue (it is reviewed in their 1982/1983 reports, which are edited by Hollien). While this postulate is so simple that it seems not to be very profound, its absence has resulted in a substantial amount of confusion relative to vocal registers. Simply stated, the concept suggests that singing registers and speaking registers are different entities. Of course, it must be conceded that they may overlap in function, that they may (in part anyway) have similar physiological roots. Nevertheless, it is recognized that while voice (laryngeal?) registers exist and are sometimes used in speech, no attempt is made to "train" them out of the productive repertoire of the speaker. Registers may exist in singing also. However, they do so primarily in the voices of beginn-

---

Footnote:

[1] The CoMet Committee on Vocal Registers includes G. Bellussi (Italy), F.S. Brodnitz (USA), W.S. Brown, Jr. (USA), B. Johnson (USA), J.-P. Köster (West Germany), V.L. Lawrence (USA), R. Miller (USA), A. Sonninen (Finland), J. Sundberg (Sweden), C. Timberlake (USA), I. Titze (USA), J.B. van Deinse (Netherlands), J. Wendler (East Germany), F. Winkel (West Germany) and H. Hollien (USA), Chair. In addition H. von Leden (USA), W.J. Gould (USA) and C.A. Schoenhard (USA) are consulting members.

ing, untrained or "not-yet-accomplished" singers; the trained singer works to remove or conceal their influence on tone, line or timbre — that is, this approach is followed by those singers who do not use register differences for dramatic effect or sing in styles that require their use. Please note, that by this very definition, we recognize the existence of voice registers in singing and the continued potential for the singer to use them if necessary or desirable.

But, what are some of these singing and/or speaking voice registers; what are their boundaries and/or dimensions? As was cited above, a rather substantial number have been proposed and labeled (Morner, *et al*, 1964); indeed, Vennard (1967) reports different scholars to have suggested that there are as few as one or as many as nine. Figure 23-1 should provide some insight as to the number, classification and extent of vocal registers; data here are drawn from the writings of four representative scholars. It should be noted, however, that the cited data were not necessarily obtained from a single reference. Rather, they are compilations of the "best" information each of the authors provide in their writings. It should be noted also that dashed lines extend certain of the register ranges. As may be obvious, these dashed lines suggest areas of uncertainty (either on the part of the author or on our part when interpreting his data). Moreover, the boundaries, as given, are not those of an individual or even the means of a subject group. Rather, they are the maximum extent of the register as portrayed by the most extreme individual within a sex.

Further, observation of Figure 23-1 will reveal that Vennard, Garcia and Appelman all suggest the presence of three registers whereas Hollien suggests only two. This difference is easily resolved. Hollien's data are based upon the registers encountered in speaking voice (two of the three he suggests have been established — out of a possible five) whereas the registers proposed by the other three authors relate specifically to singers and the singing voice.[2] Further, as will be observed, the terms used by Hollien tend to be somewhat generic; "modal" certainly is, as it refers to that register which provides the mode (i.e., it is used as the basis for most communicative behaviors). Of course, loft is a little artificial but it was proposed since it: 1) suggests a "higher" (frequency) register, 2) avoids the confusion and multiple use of the term "falsetto" and 3) had been proposed previously to identify just such an entity. The three other authors utilize "chest" to identify the lower register (it presumably is parallel to modal) as well as head (Vennard uses falsetto) to identify an upper register (presumably a register parallel to loft — or that which most phonetician's would classify as falsetto). However, these three authors identify a third — or middle register — as lying on the frequency continuum more or less in the upper region of "chest" and the lower region of "falsetto/head". Garcia and Vennard label this register as "head" and "falsetto" respectively; Appelman uses a somewhat more generic (and accurate) term: middle. It should be noted, however, that this register tends to be a rather foreign entity to most phoneticians — indeed, very few would recognize its existence, even in the singing voice. Thus, even though a given phonetician might employ a register system (or terms) somewhat different than those of Hollien, they would not tend to include a middle register in their approach/system; that is, for the speaking voice at least. Nonetheless, it is this, the middle register, that is the focus of the present paper.

Footnote:

[2] However, Wendler and Seidner (1982) argue, that only two registers can be adequately defended even for singers.

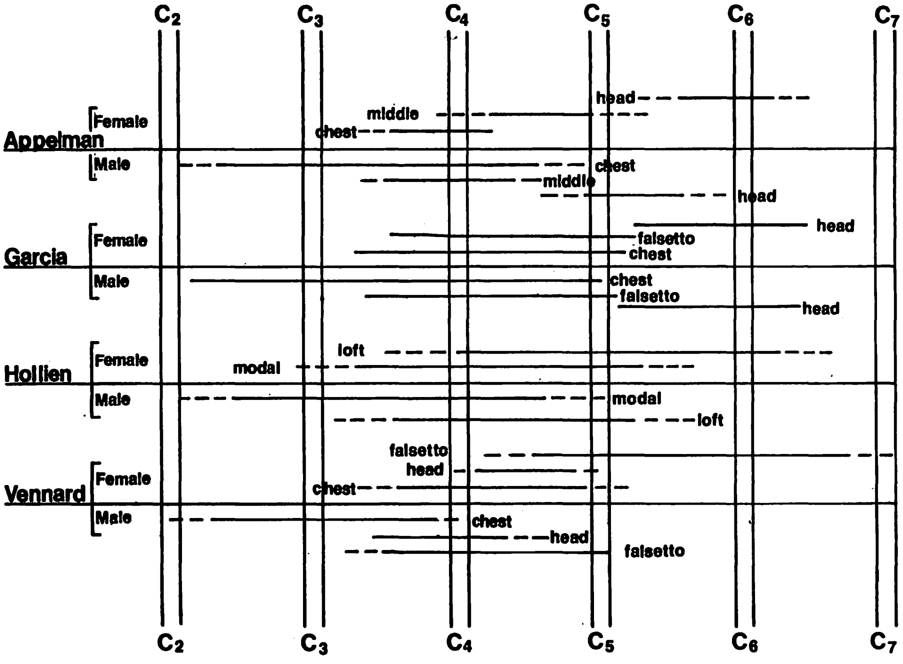


FIGURE 23-1 Number and extent of vocal registers as proposed by four representative authors.

A Misleading Concept

Prior to proceeding, however, another preliminary issue must be considered. This problem involves an unfortunate occurrance vis a vis, vocal registers; one that should be placed in perspective. As with many fields, traditions have developed in vocal music; sometimes they are helpful to the discipline, sometimes they hamper its proper development. The case in point belongs to the latter category; it involves the utilization of sensations felt by singers (when they phonate) to describe/label voice registers. The impact of doing so is both gross and subtle. It interferes with the proper development of appropriate concepts/structures (with respect to vocal registers) because it distracts the scholar/scientist or practitioner/clinician from establishing the accurate perceptual, acoustic, physiologic, aerodynamic and neurological correlates of the behavior. It is subtly detrimental because its use suggests that levels of sophistication and understanding exist which, indeed, do not. We concede that singers certainly do experience sensations — i.e., those resulting from sympathetic vibrations in the chest cavity when they sing lower frequencies and similar ones in the head, "mask" or sinuses when they sing those frequencies that are naturally resonant to these structures. But, what have these sensations to do with vocal registers...to those voice qualities that can be associated with the laryngeal tone and which are produced by the mechanical operation of the vocal folds? Probably very little. The two events (singers sensations, vocal registers) may relate in some minor way but there is little to no evidence that they are fundamentally linked. Indeed, it may be the fact that both are frequency dependent that has led to the error that one is causal to the other. In any case, it is the scientist and the practitioner — not the singer — who can (and should) establish and explain these behaviors. In the interim, simplistic and/or trivial relationships such as this one should be avoided as they can be dangerous. Here is one instance where convention must be modified in order that a better understanding of a complex mechanism may be accomplished. Since the convention does violence to the relevant concepts, it is necessary to coax traditional thought into a more reasonable and useful perspective. It matters little if the labels ultimately assigned to the various register categories are generic or artificial, just so long as they are not misleading (as are "chest" and "head").

The Middle Register — Its Perception

Just as there is a subgroup of voice teachers who argue that voice registers do not exist in the singer, there is a group which questions the validity of a "middle" register concept. Indeed, it rarely (if ever) appears in the phonetics literature nor does it appear to exist in the voices of the artistic level or premier singer (again, the cited exceptions should be noted). Where then is the domicile of this register? Apparently it appears primarily in the voices of student or developing singers. It is possible too that even accomplished singers can produce it volitionally. In any case, a rather substantial proportion of all singers, voice teachers, pedagogists and laryngologists appear to believe that this middle register exists; many consider it a problem to the aspiring singer. At this juncture, some scientists would argue that such is the stuff of legend. But, just as the legend of Troy proved to be more than simply fiction, so too have scientists long learned to trust the judgment/opinion of practitioners and clinicians relative to issues with which they are very familiar. Thusly, it is impossible to deny the potential existence of a middle register system in the face of such adamant opinion.

But, to repeat, what is the middle register — and where is it? Is there evidence that it can be perceived when heard? Obviously if it cannot be perceived as a distinct entity, it is rather difficult to argue that it exists (if, indeed it does) on the basis of other evaluations. Of course, a great many authors have provided data supporting a two register concept — both for speakers and singers — and perhaps it would be well to establish their nature as a point of reference to which the middle register may be compared. The perceptual literature related to the two major registers is extensive. Much of it can be found summarized in Hollien (1974, 1982, 1983); the acoustic and physiologic evidence is just as compelling. Moreover, there are two perceptual studies that are particularly relevant to this discussion. Better yet, they also attempt to contrast the middle register to the lower and upper registers. In the first of the two, Beard (1980) studied reasonably accomplished male singers and a second investigation was carried out on a similar group of eight female singers by Schoenhard and Hollien (1982). When the Beard data are recalculated to focus on just the two major registers (i.e., the lower or chest or modal and the upper or head or falsetto or loft), rather good correct identification scores were observed. That is, when only the low-high contrasts were analyzed, Beard's listeners correctly identified the low register 83% of the time and the high register 94% of the time. On the other hand, his correct identification levels for a three register sort averaged less than 58%. Comparable data for the female singers in the Schoenhard/Hollien study were 75% correct for the low register and 92% for the high when the auditors (listeners) were singers or singing teachers — and 72% and 85% respectively for all groups; the three-way mean for all groups/registers was almost identical to Beard's. Thus, it appears that it is relatively easy to perceptually establish the two main registers in speakers (see also previously cited research) — and in singers when they wish to produce them (see the studies cited above).

If the middle register is to be established as an entity for study, it too must be robust enough to be perceived easily. About the only experiments that have attempted to perceptually establish the middle register are the two cited (Beard, 1980 and Schoenhard and Hollien, 1982) plus Russo and Large (1978). The data here are not overly encouraging. As stated, Beard found that his 17 auditors were able to correctly identify the middle register productions of his four tenors only a little over half of the time (58.7%) with about half of the errors misassigned to the upper and the other half to the lower register. The data for the eight Schoenhard/Hollien female singers are similar if not quite as good. Here the 29 auditors (three groups of about 10 each: singing teachers, phoneticians and controls) correctly identified the middle register only a little over 47% of the time. However, in this case, nearly twice as many errors were made by a response shift to the upper register as by one downward to the lower register. The Russo-Large levels of correct identification are higher (69% correct) but they resulted from a binary choice between what appears to be the middle and lower register and it would be expected that such (two-way) scores would be higher than those for a three-way choice. In any case, it must be conceded that all the cited performances were rather weak — especially when it is remembered that most of the listeners were singing teachers, phoneticians or similar professionals. Accordingly, it must be concluded that the existence of the middle register is not easy to establish on a perceptual basis and an uncritical attitude that the register exists and its nature simply needs to be researched, may not be appropriate. In short, we find we must concede that perceptual data demonstrating the existence/nature/boundaries of a middle register are problematical in nature.

Conversely, it must be stressed that the correct identification scores (related to the middle register) were substantially above chance — and consistently so. Such relationships, coupled with the anecdotal evidence provided by pedagogists, make it difficult to simply jettison the concept. Although they tend to be weak and somewhat unstable, identification patterns persist for this, the middle, register.

Acoustic/Physiological Basis of the Middle Register

It is possible that acoustic/physiological information is available which can assist in the understanding of the middle register. Again, experiments that provide three-way contrasts will be given priority as it is hazardous to utilize data from studies where only two registers are considered — i.e., where the middle register is contrasted to either the upper or the lower register. That is, under binary conditions it is impossible to tell if the contrasts are simply upper-lower rather than middle-other. Obviously, if the contrast is upper-lower, yet the relationships are interpreted as middle-other, conceptual harm will result. Accordingly, four studies will be utilized as the primary basis (but not the only basis) of the comments to follow: the two studies cited above, a follow-up on the Schoenhard/Hollien effort and the study reported by Vennard, Hirano and Ohala (1977).

First, taken as a whole, the available data (all sources) suggest that voice spectra (quality), fundamental frequency (pitch) and vocal intensity (loudness) all contribute to the concept/percept of vocal registers. In practice, however, most research involves a design where subjects phonate in adjacent registers with fundamental frequency and vocal intensity (to some extent anyway) held pretty much constant. Procedures of this type assume that register differences are products of different phonatory spectra. If this relationship holds, then so does the model that spectra/quality are the dominant features of registers and the fo and/or VI (vocal intensity) are subordinate. Yet, the available data argue that fo and spectra tend to be about equal in their contributions to a register concept/percept whereas vocal intensity makes a lesser contribution (about half of either of the other two). Thus, if our model holds, it would argue that perception of the middle register may not be as robust as are perceptions of the other two because the nature of the research carried out tends to obscure some of its more important features. In any case, this possible relationship must be kept in mind when otherwise relevant data are considered. Since there is little direct evidence available on the fundamental frequency-register link,[3] a spectral analysis of different register-related phonations is considered useful at this juncture.

One of the assumptions made by many vocal pedagogists is that the lower register exhibits many more harmonic partials (overtones) than does the upper register. Of course, in some instances this relationship results simply from the fact that one of the two tones is lower than the other. That is, if a measurement system computes partials to 6 kHz, an fo of 100 Hz will have 59 harmonic overtones whereas one with a fundamental of 300 Hz will exhibit only 19 even though the full frequency range of the two might be equally as extensive. However, what do the data show when fo is held constant across registers?

We have evaluated this relationship (Hollien and Keister, unpublished) and have found that upper register productions do not result in a total loss of the higher partials; rather there appears to be a drop in the energy level of the partials above 2 kHz for the upper register whereas this drop-off does not occur in the spectra of the lower register. Examination of Figure 23-2 will reveal this relationship. Here, it can be seen that the higher partials for the upper register exist but they are lower in energy level than are the upper partials for the lower register phonation (especially above P-12). However, these data do not agree very well with those reported by Large (1968) for the lower register. Accordingly, we followed up on this finding by replicating the spectral analysis procedures on the phonations of the eight female singers studied by Schoenhard and Hollien. The configurations seen in Figures 23-3 and 23-4 provide data on a two register (upper/lower) contrast by two different measurement techniques. Note first that the two sets of curves are roughly similar; that both tend to reflect the vowel being produced (the resonance regions reflect the lower formants). As with Figure 23-2, the major difference between the two registers occurs in the upper partials where the energy levels for those associated with the lower register are maintained whereas those for the upper register are not (especially above about HO-10). In any case, the spectral characteristics of the middle register now can be studied by contrasting them to the curves for the other two registers. However, from the data available, it appears that the individual middle register spectra do not follow either one or the other of the cited third type. Rather, some spectra appear to predict

---

Footnote:

[3] While evidence of an fo-register relationship tends to be indirect, it does exist (Cleveland, 1977; Hollien, 1974; Sundberg, 1979).

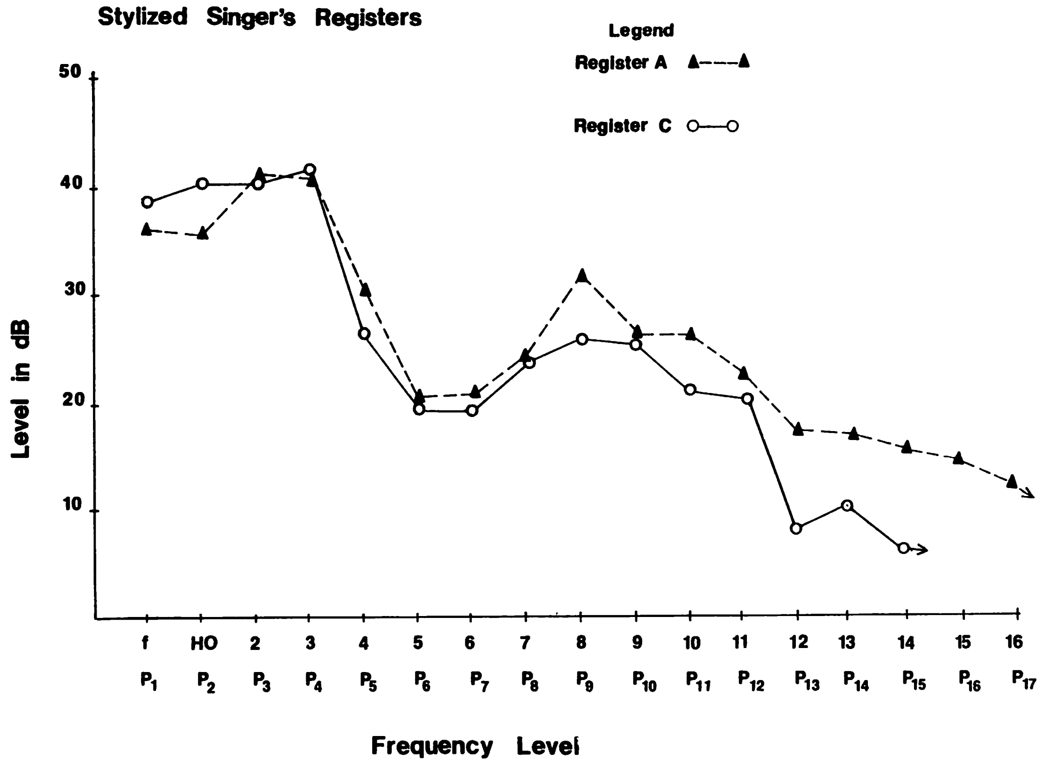


FIGURE 23-2 Spectral configurations of the upper (loft) and lower (modal) registers. Based on measurements made by Hollien and Keister for ten male and ten female singers.

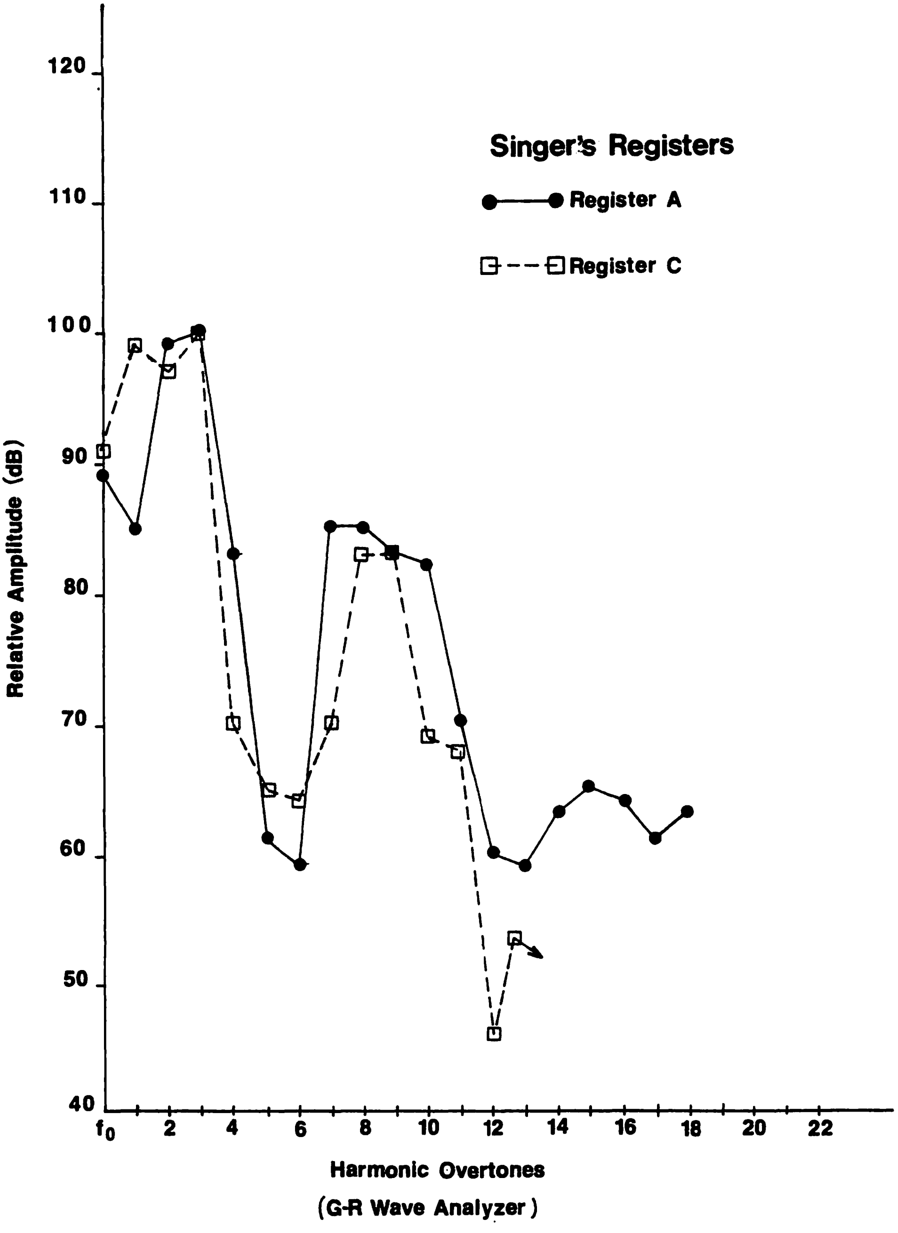


FIGURE 23-3 Spectral characteristics of two vocal registers produced by eight female singers. Measurements were made on a General Radio Wave Analyzer Model 1900-A.

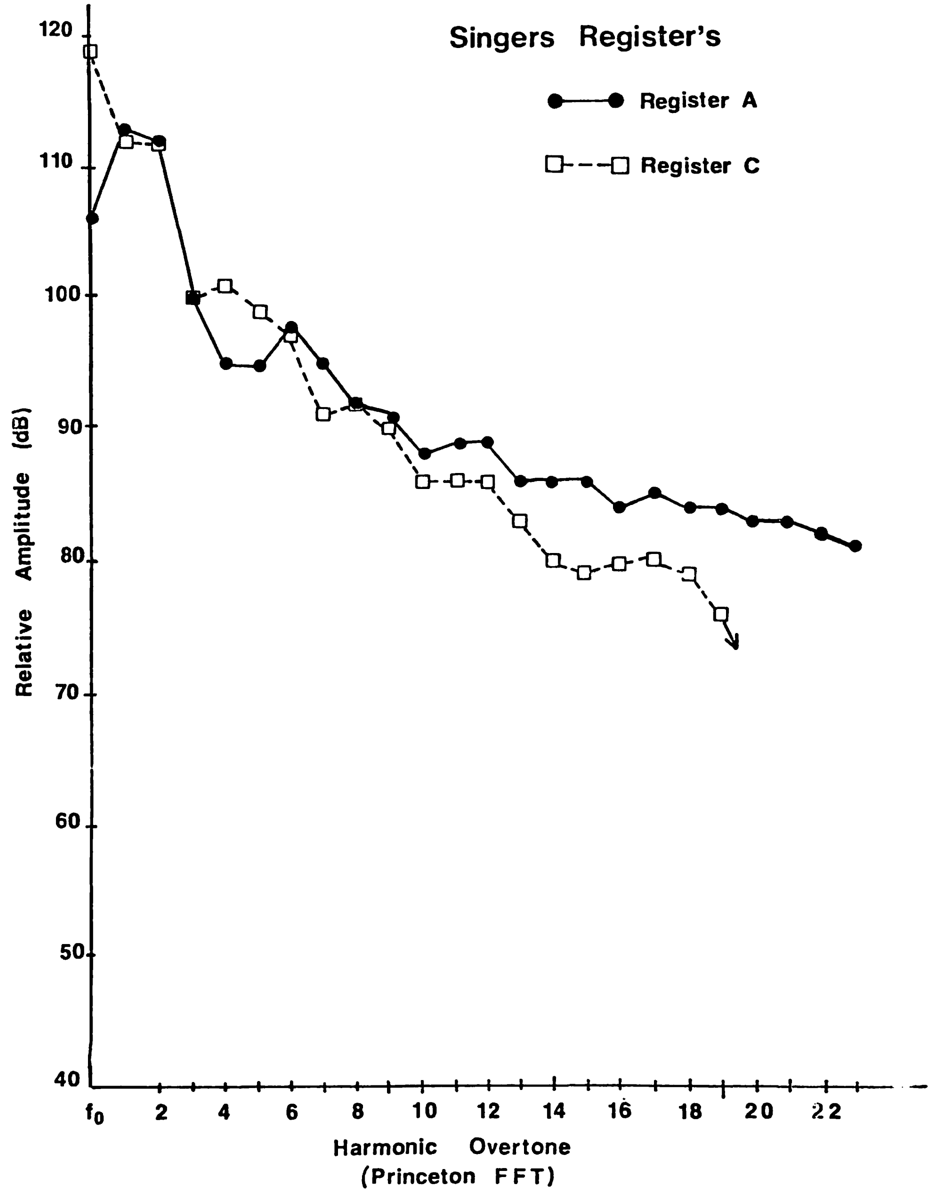


FIGURE 23-4

Spectral characteristics of two vocal registers produced by eight female singers. Measurements were made on a Princeton Model 4512 Fast Fourier Transform.

a lower-middle register relationship whereas others show similarity between the curves for the middle and upper registers. Yet a third type of curve can be noted — one that falls between those for the upper and lower registers. These relationships appear to be consistent with those reported by Large (1968) and Russo and Large (1978). In any event, they fail to suggest the presence of a voice register of laryngeal origin.

Vennard, *et al*, (1970) studied four singers who produced "scales, sustained tones and swell tones". They used some form of EMG (not well described at all) plus an aerodynamic procedure ("pneumotachograph") to evaluate these subjects and their singing. These authors suggest that aerodynamic factors were dominant for the upper register ("light" registration or falsetto) whereas the "myoelastic factors were paramount" for the lower register (heavy or chest registration). While these data appear to be somewhat overinterpreted, they do indeed provide some evidence of a laryngeal contrast between the upper and lower registers. Second, these authors correctly point out that no one (prior to them anyway) had been able to demonstrate the existence of a head (middle) register. They further contend that their data differentiate among the three registers as follows "Exertion of the cricothyroid was greatest for head, and about equal for chest and falsetto. Exertion of the vocalis was greatest for chest, less of head and least for falsetto. Air flow was greatest for falsetto, less for chest, and least for head." However, the authors also point out — and extensively so — that there were substantial differences among the singing/register/physiologic behaviors of their subjects. Given the differences and problems they outline, it is a little difficult to accept their generalizations. However, on the surface anyway, their discussions suggest the possibility that there is a middle register and that it is laryngeally linked. On the other hand, none of the 24 relevant studies reviewed by Hollien (1974) provided any evidence that a middle register exists — at least in the speaking voice. They did demonstrate that the physiologic correlates for the modal register showed systematic and orderly change, that most correlates were stable (but varied systematically) for vocal fry or pulse and that they were quite variable for loft (falsetto).

Other factors confound recognition and/or definition of a subtle operation such as middle register. Variations in larynx level (Shipp and Izdebski, 1975), spectral, vowel and/or vocal tract characteristics (Cleveland 1977; Sundberg, 1977), noise within phonation as a function of register (Emmanuel and Scarinzi, 1980) and, even variations in mouth opening (Sundberg, 1977). Indeed, the list is a long one as there are many other phonatory events or operations that could (and possibly do) differentially affect the perception or measurement of a middle register. It is a difficult entity to study; not nearly enough research has been carried out to provide a definitive answer to the questions: 1) does a middle register exist and 2) what is its nature?

CONCLUSIONS

In summary, it may be said:

1. There is substantial subjective opinion that a middle register exists. These opinions cannot be ignored.
2. There are some perceptual data that tend to support the existence of a middle register. However, these data are not robust.
3. There is very little acoustic and/or physiological evidence that would tend to support

the notion that a middle register exists. These data are the weakest of the several sets cited.

4. The nature of some of the projects/experiments designed to study the middle register could obscure its existence — so does the presence of a rather substantial number of variables that could act differentially upon it.

Accordingly, we are forced to conclude that the validity of a middle register cannot be established. At least, whatever the observed entity is, it cannot be shown to be a laryngeally linked voice register. Of course, it can be argued that the middle register does not exist but that the noted percept results from the efforts of a singer trying to produce the upper register at lower than reasonable frequencies (or the lower register at higher than reasonable frequencies) and, hence, the sung tone is so distorted that it appears to take on a separate identity. This possibility is an attractive one since it would account for all of the data reported in this review. Even the lack of consistency among the various findings is consistent with this postulate. That the "middle" register sometimes parallels the upper and other times the lower is a particularly compelling argument in this regard.

A fourth possibility also exists. It is that the entity of interest actually is a vocal tract based "register" and not laryngeally linked at all. This "dual register" concept is not a new one (Hollien, *et al*, 1976; Oncley, 1973) as it has been discussed previously. Again, the available data can be applied to this proposal and tend to authenticate it. Moreover, we have observed that certain premier singers, when singing scales, will hold quality for two or more octaves but can do so only by means of a marked and abrupt (every 3-5 tones) mechanical/physiological resetting of their vocal tract configuration. That is, they will produce an observable change in their oral-facial configuration every 3-5 tones — it is one that can be quite startling — yet the timbre of their sung tone does not vary as they change frequency. This observation, as well as most of the cited experimental data, are consistent with the concept that the middle register actually is a vocal tract related phenomenon, i.e., it results from a transfer function rather than a source function. Its variability and delicacy would argue especially against a laryngeal relationship.

In short, there are four positions that can be taken relative to the middle register: 1) that it does not exist, 2) that it is a source function (vocal) register, 3) that it is a transfer function (vocal tract) phenomenon or 4) that it actually is a distortion to the existing laryngeal registers. Unfortunately, the available data can be marshalled in support of any of these four positions. As such, it prevents this "in depth" review for being as definitive as we would wish.

REFERENCES

Appelman, R. (1953). "A study by means of planigraph, radiography and spectrograph of the physical changes which occur during the transition from the middle to the upper register in vocal tones," doctoral diss., Indiana Univ.

Appelman, R. (1967). *The Science of Vocal Pedagogy*. Bloomington: Indiana Univ. Press.

Appelman, R. (1981). "The NATS Bulletin Interviews D. Ralph Appelman," *The NATS Bulletin* 38, 5-7, 11-13.

Beard, C. (1980). "Recognition of chest, head and falsetto isoparametric tones," *The NATS Bulletin* 37, 8-14.

Bellussi, G. (1982). "Personal reflections on the question of vocal registers and proposals for the revision of terms," in *Report on Vocal Registers*, H. Hollien (Ed.), Collegium

Medicorum Theatri, New York.

Cleveland, T. (1977). "Acoustic properties of voice timbre types and their influence on voice classification," *J. Acoust. Soc. Am.* 61, 1622-1629.

Emanuel, F. and Scarinzi, A. (1980). "Vocal register effects on vowel spectral noise and roughness," *J. Comm. Dis.* 13, 121-131.

Fields, V.A. (1970). "Review of the literature on vocal registers," *The NATS Bulletin* 27, 37-39.

Garcia, M. (1840). *The Art of Singing*. Philadelphia: Oliver Ditson.

Garcia, M. (1840). *Ecole de Garcia: Traite Complet de l'Art de Chant*. Paris: Troupenas et Cie.

Hollien, H. (1974). "On vocal registers," *J. Phonetics* 2, 125-143.

Hollien, H. (Ed.) (1982). *Report on Vocal Registers*. Collegium Medicorum Theatri, New York.

Hollien, H. (Ed.) (1983). *Status Report on Voice Registers*. Committee on Vocal Registers, 9th Congress of Collegium Medicorum Theatri, Edinburgh, Scotland, U.K.

Hollien, H., Gould, W.J. and Johnson, B. (1976). "A two level concept of vocal registers," Proc. 16th Internatl. Cong. of Logopedics and Phoniatrics, Basel S. Karger, 188-194.

Hollien, H. and Michel, J.F. (1968). "Vocal fry as a phonational register," *J. Speech Hear. Res.* 11, 600-604.

Hollien, H., Moore, G.P., Wendahl, R.W. and Michel, J.F. (1966). "On the nature of vocal fry," *J. Speech Hear. Res.* 9, 245-247.

Johnson, B. (1982). "To have or have not — that is the question," in *Report on Vocal Registers*, H. Hollien (Ed.), Collegium Medicorum Theatri, New York.

Large, J. (1968). "An acoustical study of isoparametric tones in the female chest and middle registers in singing," *The NATS Bulletin* 25, 12-15.

Large, J. (1972). "Towards an integrated physiologic-acoustic theory of vocal registers," *The NATS Bulletin* 29, 18-25, 30-36.

Morner, M., Fransson, N. and Fant, G. (1964). "Voice register terminology and standard pitch," Speech Trans. Lab., Quart. Stat. Prog. Rep. 4, 17-23.

Nadoleczny, M. and Zimmerman, R. (1938). "Categories et registers de la voix," *Rev. Fr. Phoniat.* 23, 21-31.

Oncley, P. (1973). "Dual concept of singing registers," *Vocal Registers in Singing*. Large, J. (Ed.), The Hague: Mouton et Cie.

Preissler, W. (1939). "Stimmungange gattungen der menschlichen Stimme," *Arch. Sprach Stimmheilh* 3, 65-85.

Russo, V. and Large, J. (1978). "Psychoacoustic study of the Bel Canto model for register equalization," *J. Res. in Singing* 13, 1-25.

Ruth, W. (1963). "The registers of the singing voice," *The NATS Bulletin* 20, 2-5.

Schoenhard, C.A. and Hollien, H. (1982). "A perceptual study of registration in female singers," *The NATS Bulletin* 39, 22-28.

Shipp, T. and Izdebski, K. (1975). "Vocal frequency and laryngeal position by singers and nonsingers," *J. Acoust. Soc. Am.* 58, 1104-1106.

Sundberg, J. (1977). "Acoustics of the singing voice," *Scientific Am.* March, 82-91.

Sundberg, J. (1979). "Maximum speed of pitch change in singers and untrained subjects," *J. Phonetics* 7, 71-79.

Sundberg, J. (1982). "Registers," in *Report on Vocal Registers*, H. Hollien (Ed.), Collegium Medicorum Theatri, New York.

Titze, I. (1980). "Have we learned anything about registers?" Panel Discussion, Trans. 9th Symposium, Care of the Professional Voice, The Voice Foundation, New York, 130.

Vennard, W. (1967). *Singing, The Mechanism and the Technic.* New York: C. Fisher.

Vennard, W., Hirano, M. and Ohala, J. (1970). "Chest, head and falsetto," *The NATS Bulletin* 27, 30-36.

Wendler, J. and Seidner, W. (1982). "Considerations on vocal registers," in *Vox Humana, Papers in Speech Res.*, Univ. Jyväskylä 5, 154-164.

III

Biomechanics, Aerodynamics, and Acoustic Interactions of the Larynx

Section Editors

Paul Milenkovic
Ronald C. Scherer
Ingo R. Titze

24 ▪ MEASUREMENT OF VISCOELASTIC PROPERTIES IN LIVE TISSUE

Adrienne L. Perlman and Ingo R. Titze

INTRODUCTION

Having recognized the need for quantitative data describing the viscoelastic properties of vocal fold tissue, the first study specifically aimed at acquiring such information was done by Dr. Yuki Kakita and presented at the first vocal fold physiology conference in 1979, (Kakita, Hirano and Ohmaru, 1981). In that study, the investigators did isotonic measurements on isolated canine vocal folds ranging from 1 to 20 hours post mortem. This investigation provided basic information on the stiffness of the vocal fold layers, suggested a method for data analysis, and established a springboard for further study of the viscoelastic properties of vocal fold tissue.

STUDY I

In an effort to gain additional knowledge of the passive elastic properties of vocal fold tissue, Perlman, Titze and Cooper (1984), began investigation of isolated canine vocalis muscle. Isometric force elongation measurements were made for two tissue conditions. The first condition required maintaining the viability of the tissue by keeping it continuously submerged in an aerated electrolyte solution; the second condition used tissue in a post mortem state where the fluid balance was maintained by keeping it submerged in a .9% saline solution.

Tissue to be studied in the viable state was treated in the following manner: Each animal was anesthetized with 30 mg/kg sodium pentobarbital following which the larynx was excised and submerged in a Krebs-Ringer solution containing 115.50 mM NaCl, 4.64 mM KCL, 2.47 mM $CaCl_2$, 21.19 mM NaHCo3, 1.16 mM $MgCl_2 \cdot 6H_2O$, 1.17 mM NaH_2PO_4, and 13.26 mM glucose. This solution was continuously aerated with 95% O_2 and 5% CO_2 which yielded a pH of 7.4. The temperature was maintained at 37° ± 1°C. This solution provided an aerobic isosmotic environment in which the tissue remained viable for at least seven hours. The animal was sacrificed following excision of the larynx.

All dissection was performed in a heated bath with the larynx submerged in the aerated Krebs solution. Extrinsic muscles were removed from the thyroid cartilage and all structures above the level of the ventricle of Morgagni were cut away, thus exposing the true vocal folds (Figure 24-1).

In a manner similar to that used by Kakita, *et al* (1981), the thyroid cartilage was hemisected at the anterior notch and on dissection of the vocal fold, a wedge-shaped portion of arytenoid and thyroid cartilage was retained; this allowed for natural anterior and posterior boundary conditions on the isolated tissue preparation (Figure 24-2).

A 3-0 Tevdek polyester suture was then inserted through each cartilage wedge and secured. Experience has shown us that the commonly used braided silk suture has dynamically changing elastic properties that affect the validity of measurement. The polyester suture maintains more constant and predictable properties.

The viable tissue preparation was then mounted in an organ bath chamber and perfused with the aerated Krebs solution (Figure 24-3). Suture from one cartilage was fixed at one

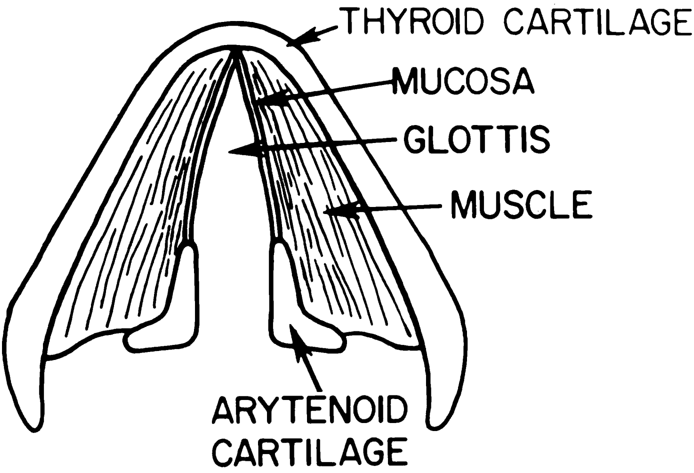


FIGURE 24-1
Schematic showing a transverse section of the larynx at the level of dissection.

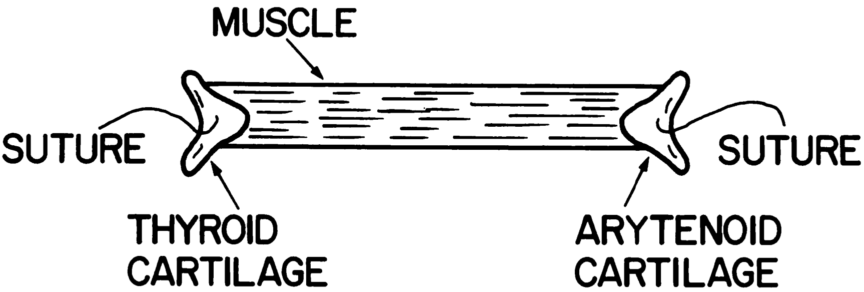


FIGURE 24-2
Schematic of a dissected specimen of vocal fold muscle showing the suture attachments through the cartilaginous boundaries.

end of the chamber and the other cartilage was connected to a Grass FT.03 force/displacement transducer fixed in a rack and pinon with a micrometer screw for length adjustment. The transducer signal was then amplified and displayed on a multichannel dynograph recorder.

In order to provide field stimulation to the preparation, a Grass S88 stimulator was connected to platinum ring electrodes in the chamber. Contractions resulting from stimulation were used to assess tissue viability.

The tissue samples studied in the non-viable condition were 6 hours and 19 hours post mortem. These were removed immediately upon demise of the animal. The protocol used with these preparations was identical to that used with the viable tissue except that saline solution replaced the electrolyte solution and there was no need to provide aeration.

The length of each preparation was adjusted in order to produce an initial tension of no more than 0.5 g. The samples were then allowed to equilibrate over 20 minutes before being subjected to isometric measurement.

Raw data were collected in the form of force as a function of time. Thus we observed the initial increase in tension followed by a gradual stress relaxation.

In order to normalize for differences in mass and cross sectional area, the responses were converted to stress-strain measures where:

$$\begin{aligned}\sigma &= \text{stress} = F/A \\ &= F\rho L_0(1+\varepsilon)/m\end{aligned}$$

This follows from a consideration of the total mass of the cylindrical tissue sample where,

$$\begin{aligned}m &= \rho AL \\ &= \rho AL_0(1 + \varepsilon)\end{aligned}$$

and ρ is the tissue density of 1.1 g/cm^3 (Greenspan and Edmands, (1971), A is the cross sectional area, L is any specified tissue length, L_0 is the resting length, and ε is the tissue strain ($\Delta L/L_0$).

Stress-strain curves for the viable and dead samples were obtained using two resting lengths. The first resting length was measured in situ and the second was measured 30 minutes following dissection of the tissue sample. Additionally, stress values were obtained immediately after stepwise elongation of the tissue, and after 20 minutes of stress relaxation. Results from that study showed that vocal fold elasticity depends on the level of strain, the elapsed time after elongation, the choice of resting length, and the condition of the tissue relative to it's viability.

Such results led us to the following conclusions:

1. As the elastic properties of viable and dead tissue differ significantly, developers of biomechanical models of vocal fold tissue vibration need to take this into consideration when using quantitative data reported in the literature.

2. The changes in stiffness resulting from stress relaxation requires rigid intrastudy consistency relative to the time at which stress values are obtained. We have found that 10% stress relaxation occurred within approximately 2.5 seconds following a manually controlled step function and continued for many minutes. Therefore, whichever time is chosen, the investigator needs to clarify that choice.

3. As stress is a function of the level of strain, the choice of an overall Young's modulus does not appear appropriate; rather, this value should appear as a function of strain as was presented by Kakita, *et al* (1981). These values were also influenced by the choice of resting length.

4. Choice of appropriate resting length was problematic. It was found that the viable tissue samples became 22½% to 41% shorter following dissection and dead tissue shortening ranged from 5% to 17%. The choice of the appropriate measurement for resting length appeared to be somewhat ad hoc, and in need of further investigation.

5. An additional concern was the great variability in stress values occurring at each level of strain. Although variability is characteristic of biological material, the investigators believed that with the creation of more refined experimental procedures, some of the causes of variability could be identified and possibly controlled.

Study II

A follow-up study looked at differences according to age and breed of the animal. In this investigation we used 7 samples from beagles each 10 months of age, 8 samples from mongrels assumed to be no more than ½ years old and 9 samples from old mongrels assumed to be older than 10 years of age.

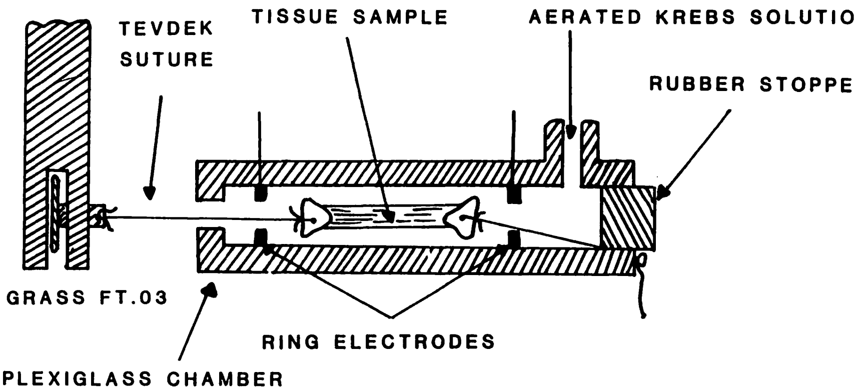


FIGURE 24-3
Schematic of organ bath chamber used in Study I.

The horizontal plexiglass chamber appeared to pose a problem in that the tissue was not pulled in a parallel fashion on elongation as is evidenced by the angle of the fixating suture. Therefore, the horizontal organ bath chamber was replaced by a water jacketed organ bath chamber that held the tissue vertically (Figure 24-4). A constant temperature was maintained in the jacket via an immersion circulator (Fisher Circulation, Model 73), and oxygenation was filtered through a sintered glass disc built into the base of the chamber.

Aware that the issue of resting length had not been resolved in the preceding investigation, we again chose to work with two different values. As it had previously been determined that the in situ resting length was less subject to variability than the dissected resting length, the in situ length was again selected as one of the conditions. This condition was defined as the distance from the point where the vocal fold attached to the thyroid cartilage, to the angle formed by the corniculate cartilage.

The second condition selected was termed the physiological resting length. This value was defined as that length whereupon supermaximum electrical stimulation of the in vitro preparation resulted in the maximum tension the muscle could develop. This length was measured in the water bath using a specially designed caliper. Due to the asymmetry of the samples, measurement was made on four sides and the average length computed. We found that the isolated muscle samples responded with the greatest tension to electrical stimulation at an average of 122 Hz.

Once the physiological resting length was established, the stress-strain curves were obtained from that setting. These curves were based on computation with the in situ resting lengths which were shorter than the physiological resting lengths. We chose to return to the in situ value for computation because this measurement was found to show less intersample variation than those values obtained in the organ bath chamber. In other words, the changes in length required by each sample in order to achieve the maximal tension on stimulation was more variable than the length measurements made in situ.

Figure 24-5 shows the mean stress-strain curves for each of the three groups of tissue samples. Tissue from the old mongrels was least stiff and that obtained from the young beagles was most stiff. When the curves were plotted after 5 minutes of stress relaxation the relationships appeared essentially the same. However, when the percent change in stress from elongation to 5 minutes post elongation was calculated, behavioral differences of the prepara-

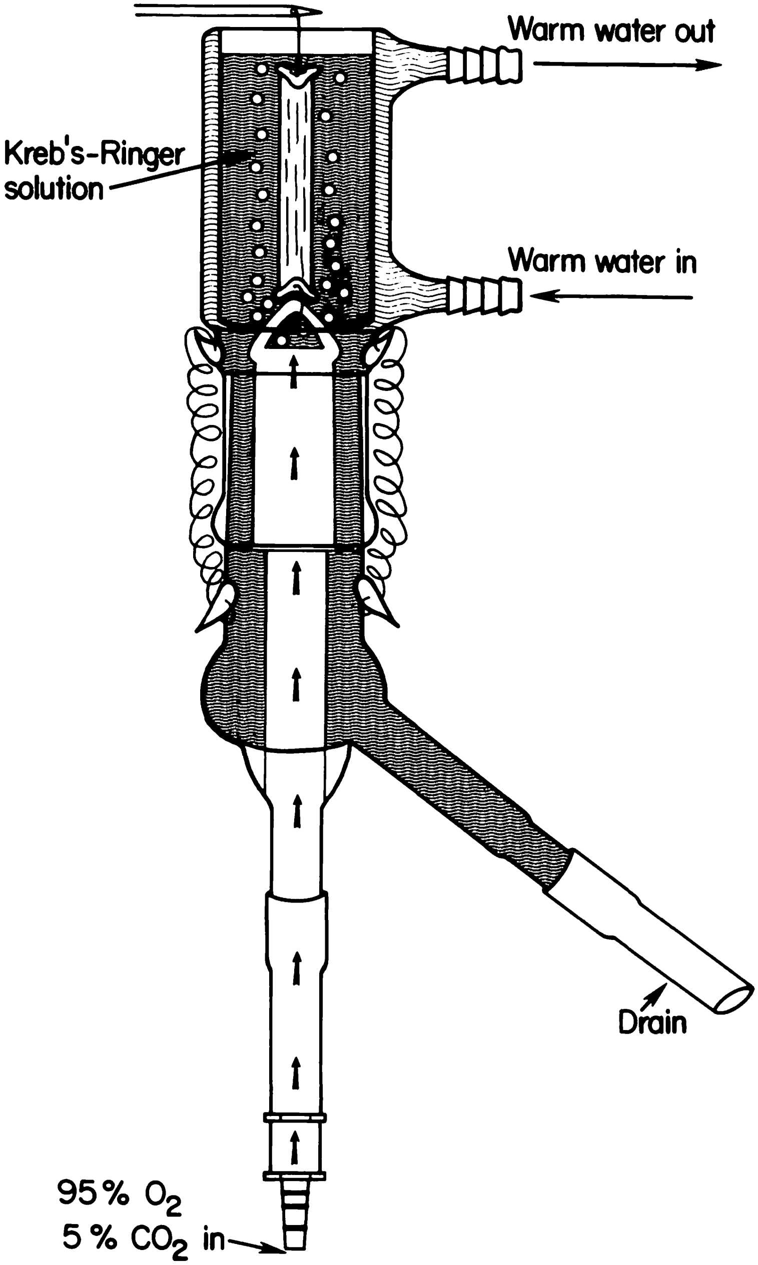


FIGURE 24-4
Schematic of water-jacketed organ bath chamber used in Study II.

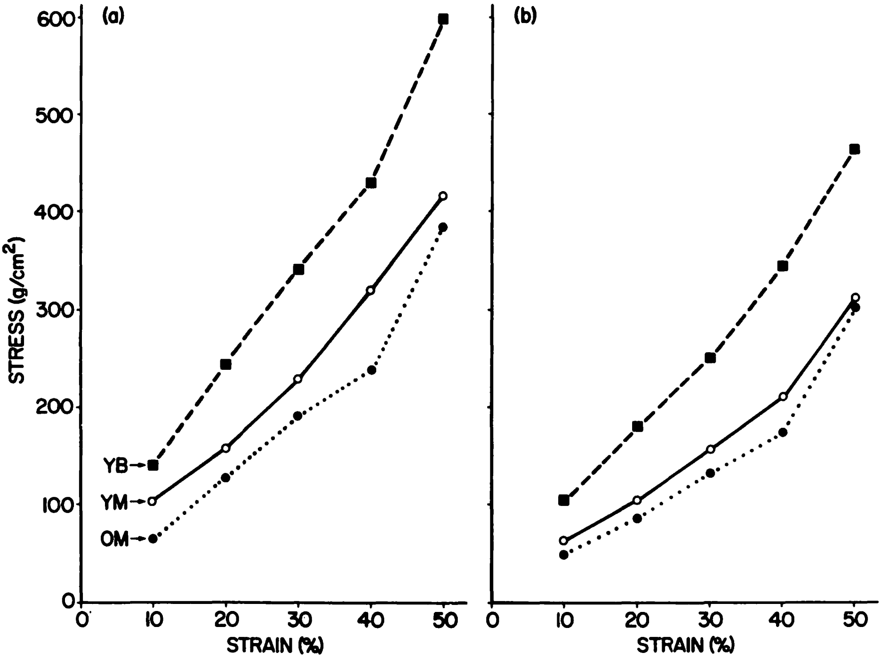


FIGURE 24-5
Stress-strain curves for vocalis muscle of young beagles (YB). Young mongrels (YM) and old mongrels (OM). (a) after a step elongation, (b) after 5 minutes of stress relaxation. Strain is computed on the basis of the in situ resting length.

tions were evidenced. The tissue from the old mongrels was more irregular in the amount of relaxation that occurred as a function of strain (Figure 24-6).

The three groups were compared relative to the variability at each level of strain (Figure 24-7). Using the coefficient of variation (V), the standard deviation of the stress value for each level of strain was divided by the mean stress for that level of strain. Again the old animals showed the most inconsistent behavior. The young mongrels were significantly less variable. Such differences as those observed in this investigation can be interpreted as supportive evidence of tissue alterations as a function of age, and may eventually prove valuable in the development of theoretical models of senescent voice production. Meanwhile, when using the canine model in acquisition of information on the viscoelasatic properties of vocal fold tissue, age and breed appear to be additional factors investigators must consider.

In an effort to define other parameters that can be controlled in the acquisition of such data, we are looking at the possible effects of sex, the possible release of residual neural transmitters following elongation, tissue geometry, and rise-time for the stepwise elongation. With an understanding of these parameters, standardization of a technique for measurement of the viscoelastic properties of vocal fold tissue will have been refined such that we are prepared to study the human model.

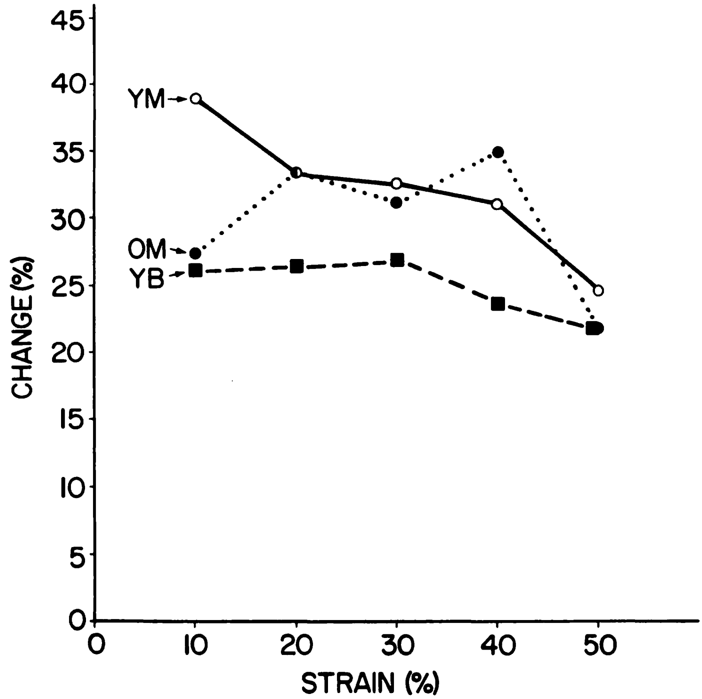


FIGURE 24-6
Percent change in stress from the step elongation to five minutes after the step elongation. Strain is computed on the in situ resting length.

REFERENCES

Greenspan, K. and Edmands, R.E. (1971). "The use of microelectode techniques in muscle," *Methods of Pharmacology* 1, A. Schwartz (Ed.), Appleton-Centry-Crofts, New York.

Kakita, Y., Hirano, M. and Ohmaru, K. (1981). "Physical properties of the vocal fold tissue: measurements on excised larynges," in *Vocal Fold Physiol.*, K. Stevens and M. Hirano (Eds.), Univ. of Tokyo Press.

Perlman, A.L., Titze, I.R. and Cooper, D.S. (1984). "Elasticity of canine vocal fold tissue," *J. Speech and Hear. Res.*, 27/2, 212-219.

DISCUSSION

B. WYKE: In view particularly of the changes (familiar to forensic pathologists) that take place in the protein structure and intercellular pH of muscle fibres after death, I doubt if studies on muscles made 6 to 19 hours after death (even if kept in saline solution - but non-oxygenated) are of much functional relevance. More relevant information should be obtained on the muscles removed under anaesthesia immediately into warm, oxygenated electrolyte solutions - but only if the muscles were continuously perfused with the same solution, since without this the viscoelastic measurements are being made on muscles with an empty

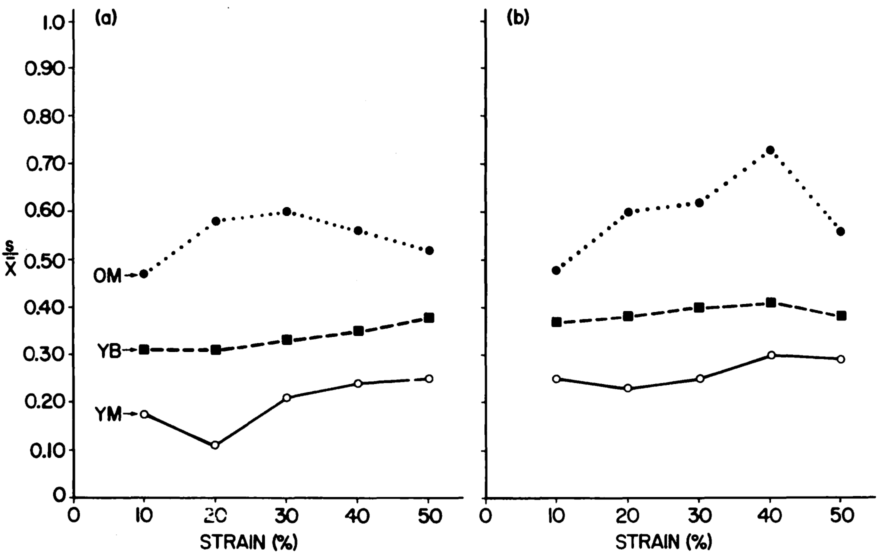


FIGURE 24-7

The coefficient of variation (V) where the standard deviation (s) of the stress for each level of strain is divided by the mean ($\bar{X}$) stress value for that strain. The smaller the V, the less relative sample variation. (a) after a step elongation, (b) after 5 minutes of stress relaxation. Strain is computed on the in situ resting length.

vasculature. Since the stiffness of muscles is influenced by their intravascular pressure and intramuscular blood volume, would not the significance of your studies be enhanced by continuous perfusion of the muscles from the time of their removal to the end of your observation? This could be achieved by microcannulation of one of the small arteries supplying the vocalis muscle at the time of surgical removal and perfusion from a perfusion pump set to operate at the dog's normal mean arterial pressure (about 80 mm Hg), which would require little modification of your existing apparatus.

A. PERLMAN: In regard to your first statement, I agree with you completely. In fact, that is precisely what our study was demonstrating. As so much of the data from previous laryngeal investigations have been obtained with dead specimens, we felt it was important to describe some of the differences in the stiffness characteristics of dead and viable tissue. The study by Perlman, Titze, and Cooper (J. Speech Hear. Res., 1984) goes into these differences in some detail.

Regarding your second statement, it would be ideal if we could study each body system in vivo, but that is obviously not possible. And so, one approach is to look at small parts of each system and gradually build upon them, in an effort to understand the whole. Therefore, we have applied a technique that is very standard in physiological and pharmacological research, and that allows us to look at the system in such a manner.

To perfuse through the circulatory system is a good suggestion. However, since the blood volume makes up approximatley 8% of the total fluid volume, it is doubtful the volume contributed by the capillary bed would change the stress-strain characteristics. Furthermore, the method by which we are preparing our sample does not lend itself to cannulation; there is no main arterial branch within the vocalis muscle, and the arterial supply to the vocal folds is external to the thyroid cartilage. As all but a very small wedge of cartilage and a strip of vocalis approximately 2 mm in diameter remain after dissection, preservation of the arterial system would be extraordinarily cumbersome, if not impossible. The method we use of restricting the diameter of the muscle does allow for adequate diffusion of the electrolyte solution. Even if it were possible to cannulate, the interruption in venous drainage would cause an abnormal drop in vascular resistance, rendering the effects of perfusion invalid; and maintaining a constant capillary volume would require precise and continuous measuring of the net movement of fluid out of the capillaries as well as measurement of fluid intake.

If one were to weigh the potential error resulting from a systematic volume loss or an error resulting from fluctuations in volume, I believe the first to be preferable, as volume corrections are fairly easy to make.

I. TITZE: I agree that the blood itself would not add any significant contribution to the elasticity, and that if one could measure the volumetric difference, then that correction would be a systematic correction that could be easily taken into account.

B. WYKE: Theoretically I would accept that, except that the studies I am familiar with comparing bloodless with perfused muscle in isolated chamber experiments, produced very significant differences in a number of viscoelastic properties, which is the reason I raised the matter.

D. ZEALEAR: I may have missed this. How did you measure stress and strain? Was it a step perturbation in length?

A. PERLMAN: We apply isometric steps in elongation, so the tissue is increased by one millimeter increments in length, and therefore we obtain force at one millimeter increments.

D. ZEALEAR: So you are looking at a steady state.

A. PERLMAN: That's right, isometric steady states.

25 ▪ ACTIVE LENGTH-TENSION CHARACTERISTICS OF NATURALLY ACTIVATED MUSCLE: MEASUREMENT AND IMPLICATIONS

Dom V. Finocchio and Erich S. Luschei

The relationship between the length of a skeletal muscle and the isometric tension that it can produce, usually called the length-tension (LT) cürve, has been known from the time of the earliest studies of muscle biomechanics. Generations of physiology students have attached the tendon of the gastrocnemius muscle of a frog to a rigidly-held force transducer and observed that the contractile force produced by tetanically stimulating the nerve to the muscle varies in a systematic way when the length of the muscle is changed. The force is near-maximal over a relatively wide range of central muscle lengths, but decreases rather steeply for very short or very long lengths. This LT curve has usually been explained by relating the structure of the muscle sarcomeres to the sliding-filament hypothesis of muscle contraction. In fact, the LT curve of single muscle fibers (Figure 25-1), measured with great precision and related to the directly-measured sarcomere length, has constituted an important piece of evidence in support of the sliding-filament hypothesis. While the LT curve has thus been quite important in helping to understand muscle contraction, comparatively little attention has been to the implication of the LT curve for understanding the control of limb movements. In general, there has been an assumption that normally-activated muscle would produce an LT curve like that seen with tetanically-stimulated muscle (see 35 impulses/sec curve of Figure 25-1). This curve has a broad peak for most muscle lengths in an intact limb. Thus, it is reasoned, isometric tension would only vary to a small degree for most functional positions, and would only become an important determinant of force at extreme positions.

The view that the LT curve of naturally-activated muscle has a broad, relatively flat central peak depends upon the assumption that the LT curves of sub-tetanically stimulated muscle are merely scaled versions of the curve seen with tetanic stimulation. An experiment by Rack and Westbury (1969) was the first to systematically analyze the LT curves of a muscle when it was stimulated at various rates. Low rates of whole-nerve stimulation would have produced a very unfused (rippled) force in which the large amount of internal movement in the muscle might have compromised the interpretation of the results. To avoid this problem, Rack and Westbury divided the nerve into 5 filaments and stimulated each of them individually in a sequential manner. A smooth (fused) contraction even at rates as low as 3 Hz was achieved in this manner. The results of their study of the LT curve of the cat soleus muscle is illustrated in Figure 25-1. Contrary to the assumption noted above, the LT curves are not scaled versions of one another. When stimulated at low rates, the LT curve is steep and relatively linear in the range of normal muscle lengths (see the curve for 5 impulses/sec in Figure 25-1). This result would suggest that the level of activation of a normally-innervated muscle would vary with limb position if the contracting muscle supported the same load at each position.

Limb mechanics also affect the degree of muscle activation that would be necessary to hold the same load at different limb positions. The mechanical advantage of the mandibular muscles does not change too much, however, for positions between occlusion (teeth together) and 18 mm of interincisal opening in the rhesus monkey. If the LT curve of monkey jaw muscles is relatively "flat" over this range, one would expect that holding a 600 gm load near occlu-

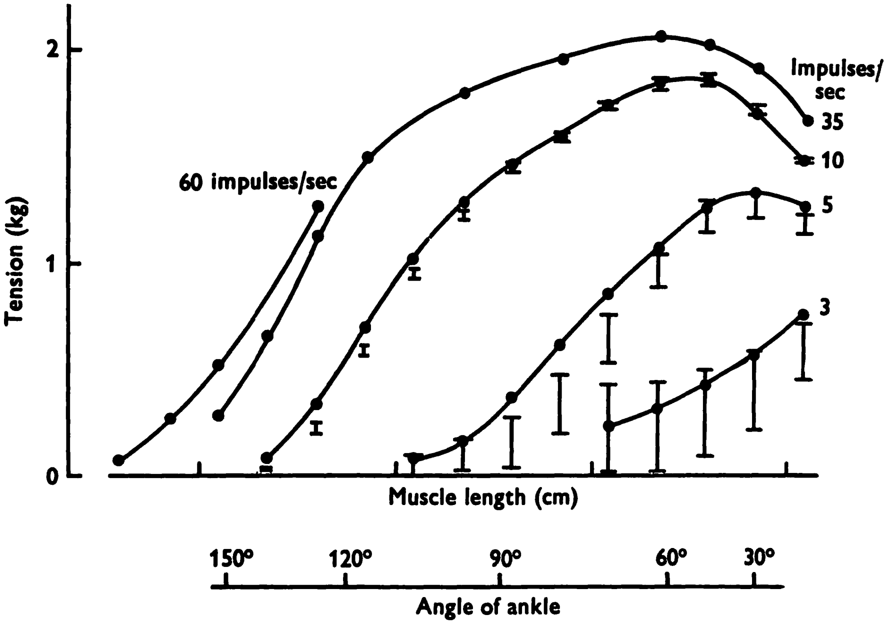


FIGURE 25-1
Length-tension curves of cat soleus muscle when stimulated at various rates, using distributed stimulation of the divided (five portions) motor nerve. Figure 8 from Rack and Westbury (1969).

sion would require about the same level of muscle activation as holding a 600 gm load with the teeth 18 mm apart. If one were to also take mechanics into account, assuming a flat LT curve would lead to the expectation that muscle activation would decrease for the position closer to occlusion because the jaw muscles have a better mechanical advantage near occlusion. The result shown in Figure 25-1 suggest the possibility, however, that the "natural" LT curve of jaw muscles might have a steep, positive, slope near occlusion. If this were the case, then one would expect to see the level of jaw muscle activation increase substantially as the same load is held closer and closer to occlusion.

We have had the opportunity recently to evaluate data that perhaps shed light on the LT curve of normally-activated jaw muscles. These observations are part of an experiment to determine whether or not jaw muscle spindle afferents are necessary for a monkey to control a specified jaw position against an electronically-controlled jaw-opening force. In order to answer this question, it has been necessary to carefully study the electromyographic activity (EMG) of jaw opening and closing muscles as a function of load and jaw position.

Three rhesus monkeys have been trained to control the position of their mandible by lifting the lever arm of a force-feedback servo system (Figure 25-2). The servo system exerted a specified, constant, opening force between 2 and 10 Newtons(N). The monkeys were required, in the part of the experiment to be presented here, to maintain the position of the lever arm such that a visual cursor, displayed on an oscilloscope in front of the monkey, remained within a stationary target zone for at least 2 seconds. After each successful attempt, the target zone was incremented up or down such that all jaw positions between 5 and 18 mm of interincisal distance were tested during a given loading condition. After this behavior

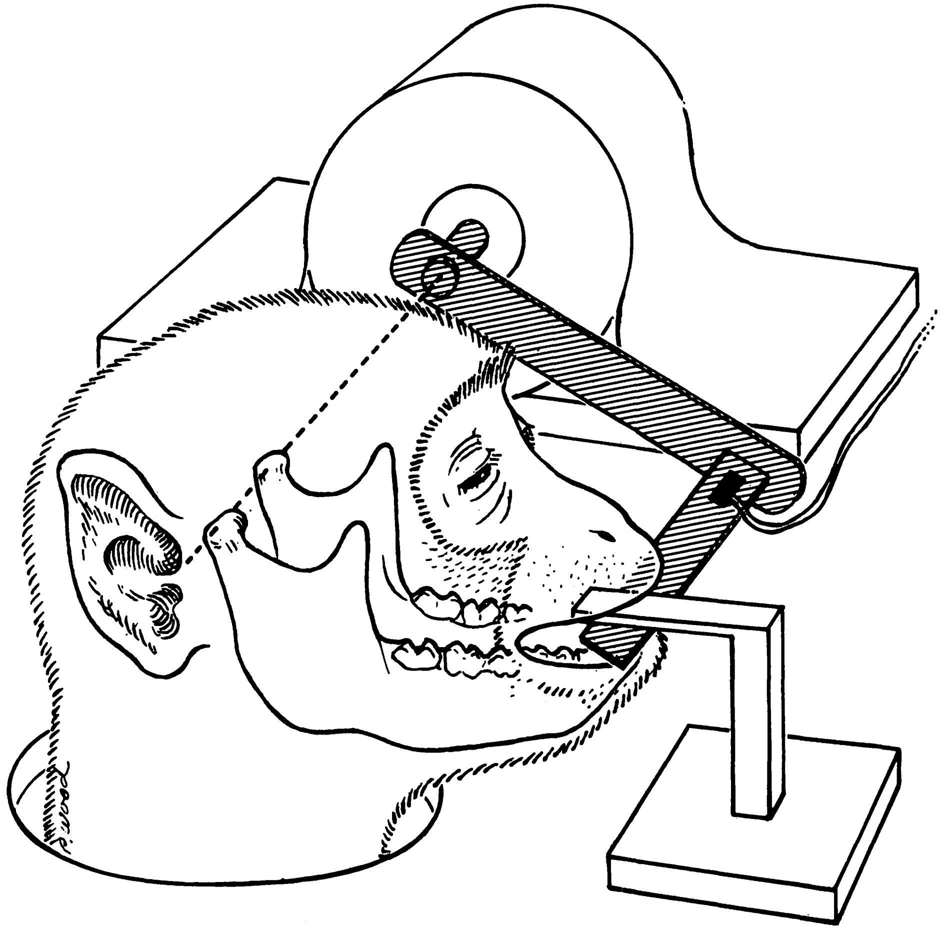


FIGURE 25-2

Simplified representation of the monkey working on the force-feedback servo system. As the axis of the motor was approximately co-axial with the monkey's temporomandibular joint, the lower beam followed the path of jaw opening. Plates around the upper and lower beams that forced the monkey to use its incisors are not shown. Strain gauges on the upper beam that sensed a significant downward force were used to enable reinforcement, i.e. the monkey could not fix the mandibular position and control the lower beam by moving its head.

was established, fine ductile wires were surgically implanted unilaterally, under halothane anesthesia, in the digastric, medial pterygoid, masseter and temporalis muscles. EMG activity was tape recorded from these muscles, as well as jaw position and loading force, while the animals performed the position tracking task under different loading conditions.

Records were analyzed by full-wave rectification and low- pass filtering of the EMG signals, and then digitizing the EMG, position, and force records on a laboratory computer. A "scrolling" program was used to select one second intervals at various jaw positions where the average position was steady (see Figure 25-3). The slope, y-intercept, and linear regres-

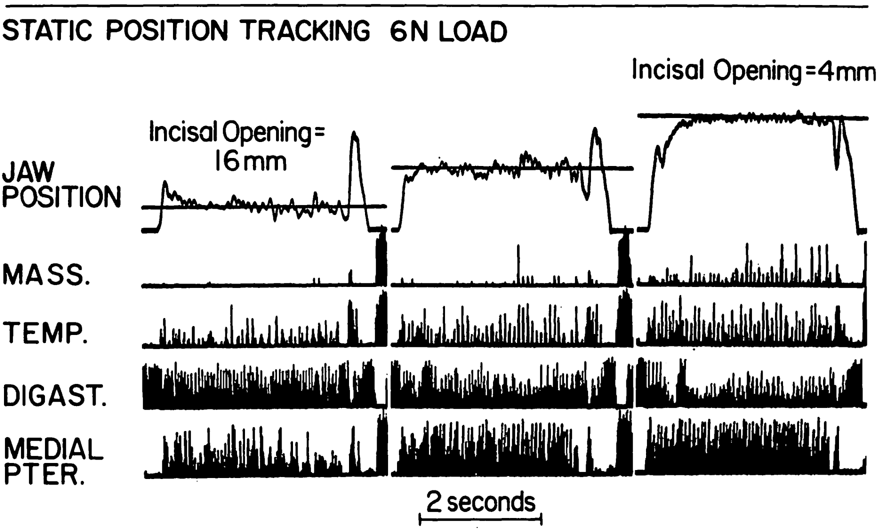


FIGURE 25-3
Repsresentative records of monkey 2 holding a 6N force at three different jaw openings. EMG records rectified but unfiltered. EMG amplifier gains were: masseter, 20K; temporalis, 20K; digastric, 5K; medial pterygoid, 10K. The horizontal line through the position trace represents the center of the target zone.

sion coefficient of the EMG amplitude versus interincisal distance was calculated for each jaw muscle. Scatterplots of these variables were also produced (Figure 25-4 and 25-5).

Figure 25-3 illustrates representative individual responses of monkey 2 holding a 6N load at three positions. When the jaws were open (leftmost response of Figure 25-3), the closing muscles (masseter, temporalis, and medial pterygoid) were less active than when the jaws were nearly closed. The activity of the anterior belly of the digastric muscle, an opening muscle, shows the opposite relationship to position, but the same relationship to muscle length. EMG amplitude of all the muscles at an intermediate position (middle response) appears to be intermediate to either more open or more closed positions. The bursts of muscle activity at the end of the responses were associated with the animal swallowing the squirt of apple juice used to reinforce the successful trial. These EMG responses are truncated because of saturation of the amplifiers. In order to record the muscle activity associated with these low-force responses, it was necessary to use amplifier gains of about 20,000. If we had used much lower gains, which would have been necessary in order to linearly record the EMG associated with chewing and swallowing, the low-level "positioning" muscle activity would hardly have been detectable. Another feature of the data that is illustrated in Figure 25-3 is that the jaw position is clearly controlled by a process of co-activating antagonistic muscle groups. This feature was always present in the records from all three monkeys, even when the load against closing was as high as 10N. During rapid phasic movements, however, as when chewing or swallowing, digastric activity was always out of phase with the activity of closing muscles.

The results of the quantitative analysis of the position-EMG relation for all of the major jaw muscles are shown in Figures 25-4 and 25-5 for monkeys 2 and 3, respectively. The results obtained with monkey 1 were similar, except that our EMG implants in digastric and medial pterygoid were not satisfactory for quantitative analysis. In general, the slope of the regres-

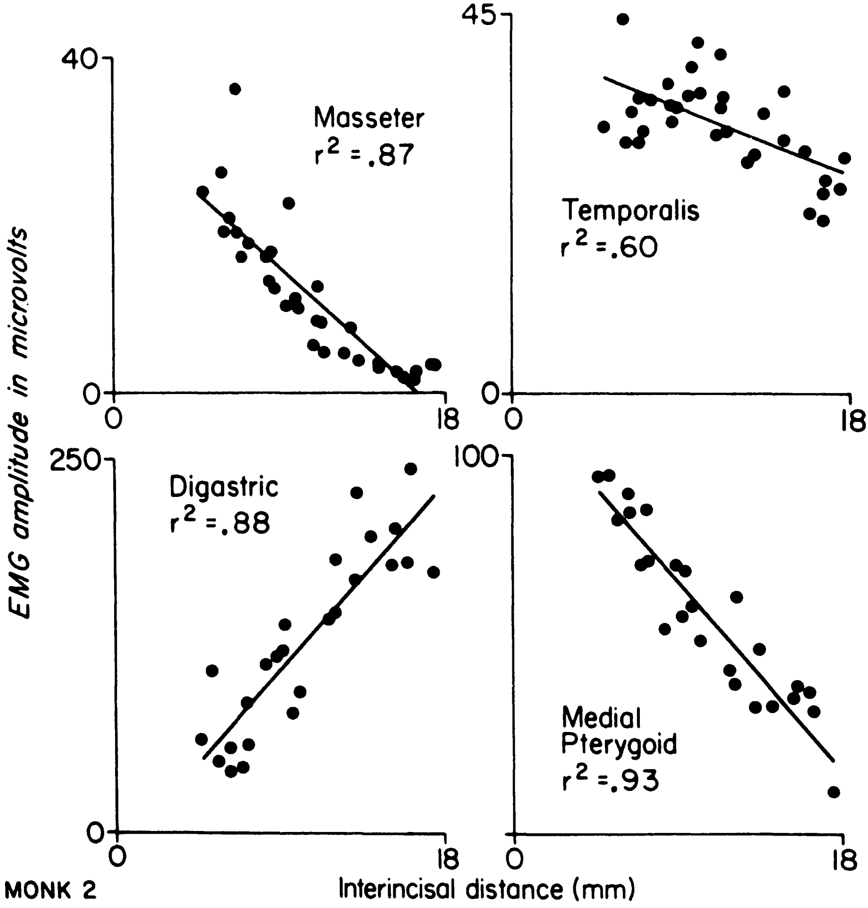


FIGURE 25-4

Scatterplots and linear regression lines for the EMG amplitude versus interincisal distance relationship for four jaw muscles of monkey 2. Load was 6N for these data.

sion lines of the EMG-position curves of the closing muscles increased (became steeper) in a monotonic manner with increasing loads. The changes were not subtle; typically the slopes with a 10N load were 2 to 3 times as steep as the slopes obtained with a 2N load.

These data would be consistent with the conclusion that jaw muscles operate on a steep LT curve when they are activated for maintaining jaw position against relatively small loads. This is not to say that this would be true at all levels of muscle activation. Chewing, for example, may very well work on a relatively "flat" LT curve near occlusion simply because the level of activation is closer to being "tetanic." The most general conclusion to be drawn from Figure 25-1 is that there is not just one LT curve; there are a family of them, each dictated by the level of activation.

Before the conclusion suggested above is accepted, it may be useful to consider alternative possibilities. One possibility is that the passive elastic support of the mandible is adequate to support a substantial fraction of the load when the jaws are open, so there is less

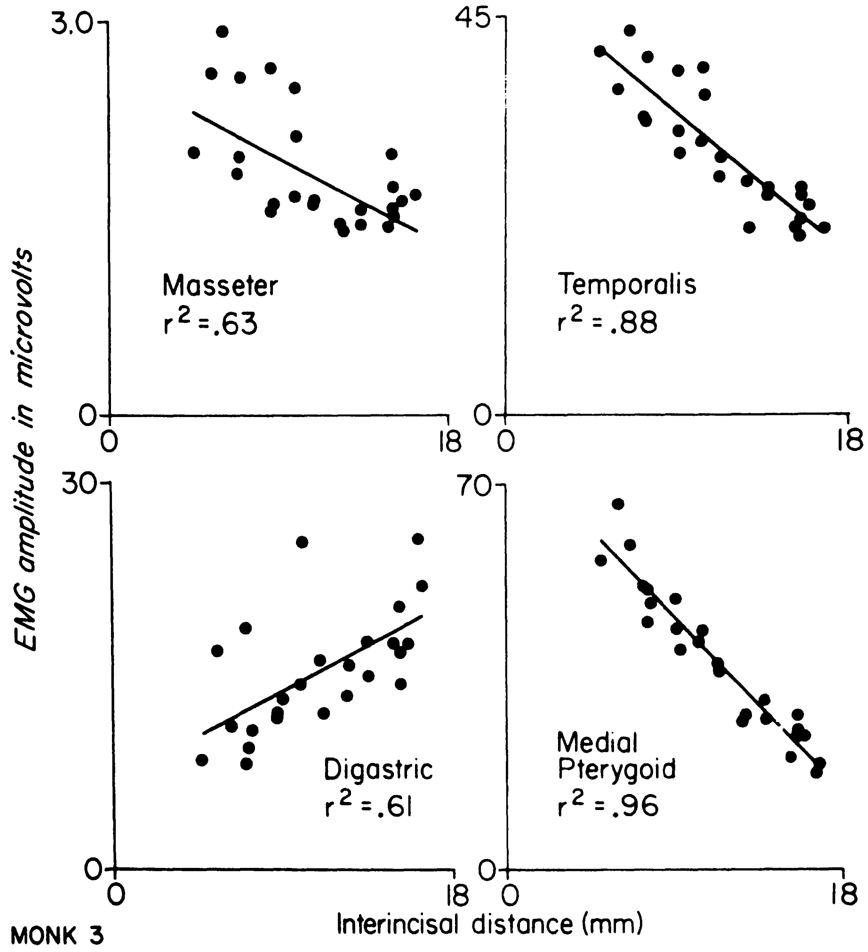


FIGURE 25-5
Same as Figure 25-4, except data is from monkey 3.

muscle activity at these positions. This is an unlikely hypothesis because when these monkeys were anesthetized deeply, until their muscles were totally relaxed, their jaw fell open of its own weight to an interincisal distance of 14 mm when their head and body were held in an upright posture. It only required an additional 20 gm hanging on the mandibular incisors to open the jaws to 18 mm. Another possibility is that the increase in the activity of jaw-closing muscles at more closed positions was to overcome an increased opening force produced by the digastric muscles. To make this possibility reasonable, it would be necessary to assume that digastric, too, was working on a steep, positive, LT curve. If the LT curve of digastric were "flat", then the fact that there was much less digastric activity when the jaw was closed than open would mean that the closing muscles were working against less rather than more opening force when the jaw was nearly closed. It is most parsimonious, however, to suppose that both opening and closing muscles work on a steep portion of their LT curve. Such an assumption is the basis for a current hypothesis concerning the manner in which animals could control the position of their limbs strictly from controlling the level of activity in muscle pairs (Polit and Bizzi, 1979; see Appendix below).

Muller and MacLeod (1982) recently presented an analysis of the biomechanics of the perioral muscles. As part of their experiment, they had subjects maintain the EMG level at a given value, and measured the isometric force developed by the contracting muscle when its length was varied. This relationship, denoted $\tilde{S}(x,e^*)$, is, in effect, a family of LT curves, each curve being related to a given level of muscle excitation. They found that such curves were steep and relatively linear over the normal operating range of the muscles. It is premature to suppose that all muscles usually operate over a steep portion of their LT curve, rather than across a broad peak, but this might be the case. At the very least, it would be prudent to regard muscle activity (EMG) as potentially a measure of the nervous system's attempt to control position of limbs or articulators.

ACKNOWLEDGEMENTS

We wish to thank Phyllis Wood and Pat Roberts for help with preparation of figures. This work was supported by National Institutes of Health Grants DE04884, RR00166, and RR00374.

APPENDIX

The model proposed by Polit and Bizzi may be understood in a qualitative manner by imagining that two muscles are having a "tug-of-war," and each has some constant, specified, level of activity. Suppose the point at which they are joined, which dicates the position being controlled, is far to one extreme. One muscle will be short and the other long. The long muscle generates more force than the short muscle, so the "tie point" moves. As the previously longer muscle becomes shorter it generates less force, and as the previously shorter muscle becomes longer, it generates more resistance to the movement. Eventually, the tie point reaches a position at which the force produced by one muscle just balances the force produced by the other muscle. This will be the "postural" position governed by these LT curves and these levels of muscle activity. This position will be intrinsically "regulated"; an attempt to displace the system from equilibrium will be met with increasing resistance as the displacement increases, and when the displacement force is removed, the system will return to its original equilibrium point. Whether this hypothesis is actually how posture is normally controlled has yet to be determined; it at least provides an explanation of how deafferented animals can control limb positions. One of the conditions that must be met for this hypothesis to be reasonable, however, is that muscles must work on relatively steep LT curves. The results of our experiment are consistent with that assumption.

REFERENCES

Gordon, A.M., Huxley, A.F. and Julian F.J. (1966). "The variation of isometric tension with sarcomere length in vertebrate muscle fibers," *J. Physiol.* (London) 184, 170-192.

Muller, E.M. and MacLeod, G. (1982). "Perioral biomechanics and its relation to labial motor control," paper, Acoust. Soc. Amer. meeting, Chicago, IL.

Polit, A. and Bizzi, E. (1979). "Characteristics of motor programs underlying arm movements in monkeys," *J. Neurophysiol.* 42, 183-194.

Rack, P.M.H. and Westbury, D.R. (1969). "The effects of length and stimulus rate on tension in the isometric cat soleus muscle," *J. Physiol.* (London) 204, 443-460.

DISCUSSION

J. FOLKINS: Have you made any attempts to calculate the changes in the force that should occur due to changes in the jaw position?
E. LUSCHEI: Yes I have. You can calculate the effective length-tension curve of these muscles from the data we have, simply by going back and finding the appropriate points on the appropriate graphs, and I've done that. It in fact shows that the length-tension property here is very steep. I would add that this conclusion was reached independently by Eric Muller and his colleagues in Wisconsin, looking at face muscles from a somewhat different view, and on a different muscle system. Their estimates of what are effectively the length-tension properties of muscles when they are functionally used to control the lip come out to be very similar to what our curves would be.
J. FOLKINS: Is your calculation based on pure mandibular rotation?
E. LUSCHEI: No, my calculation is based upon incisal distances. It could be corrected for that, but the results would not change significantly. These curves are very steep as you get close to occlusion.
J. FOLKINS: But you could measure a couple of points on the jaw and then you could calculate where the origins and insertions would change.
E. LUSCHEI: Yes, quantitatively you would need to do that. I have not done it.
O. FUJIMURA: This is a very interesting study, but I would like to make sure that there is no complication due to macroscopic structural problems. That is, if we look at the data for higher jaw position, we see that there is more force and more stability. That could possibly be explained if the monkey raises the tongue, and internally supports and stabilzes the position.
E. LUSCHEI: You suggest that perhaps the monkey is pushing on the roof of his mouth with his tongue, is that correct? We know that that is not the case because these monkeys continuously drink applesauce while they are doing the task with no change in the EMG while they hold the same position.
O. FUJIMURA: That is no guarantee, because you can leave a channel, particularly the midsagital channel, still supporting the tongue body against the palate on the side, which happens very typically for human articulation of front vowels.
E. LUSCHEI: That might occur at relatively high positions, but the highest position they are in is about 5 millimeters to occlusion. I have never actually measured that, but at the wider open positions, I don't think the animal could actually reach the roof of his mouth with his tongue, and I wouldn't see why he would do it.
O. FUJIMURA: No, I'm not suggesting that for low positions, but for low positions there might be an entirely different mechanism.
I. TITZE: Would you speculate about how the steepness of this curve would apply in the case of the vocal folds? Adrienne has pointed out that the typical rest length is considerably less than the physiologic length, and that the overall length change is perhaps on the order of 30% or so. Where would you imagine the operating point to be then in terms of what you have indicated here?
E. LUSCHEI: Reasoning from Rack and Westbury's result (1969, Fig. 1 of this presentation), one could speculate that the muscles that control the vocal fold would operate on a steep positive portion of the length-tension curve if, during phonation, the average rate of firing of motor units is low. One may note in Fig. 1 that the length-tension curve is steep

for stimulus rates of 3 and 5 per second even though the muscle is almost at its maximum physiological length (a length that the muscle cannot even achieve in situ). However, if the motor unit firing rates are very high during phonation, these muscles would be expected to be operating on a flatter portion of the curve, or even a portion with a negative slope. Thus, our guess has to be based on knowledge of the type of activation of these muscles. Further, we must keep in mind that we are assuming that laryngeal muscles are like slow limb muscle except for a scaling of the stimulus rates required for fusion.

I. TITZE: What would you suggest to be the best definition of rest length?

E. LUSCHEI: For the laryngeal muscles, I wouldn't care to offer a suggestion right now. For a limb muscle, it is probably the length utilized by that muscle for some large fraction of its ordinary use, 85% of its use, for example. That is, I think it is to be functionally defined rather than anatomically defined.

26 ▪ CONTACT PRESSURE BETWEEN THE VOCAL PROCESSES OF AN EXCISED BOVINE LARYNX

Ronald Scherer, Donald Cooper, Fariborz Alipour-Haghighi and Ingo Titze

INTRODUCTION

The range of vocal qualities from breathy to pressed may correspond to an increase in the adductory forces between the arytenoid cartilages. Furthermore, vocal pathologies which involve changes of the medial arytenoidal tissue may be associated with inappropriate levels or coordinations of adductory forces (von Leden and Moore, 1960). Also, the training of the professional voice may involve fine control of these forces. It is therefore of wide interest to examine the effects of changes in the levels of interarytenoid forces during phonation.

There appears to be little information on interarytenoid forces or pressures. Rethi (1897) placed a pressure measuring elastic instrument into the glottis of animals and stimulated the cricothyroid and posterior cricoarytenoid muscles. Highest values of closing forces were estimated to be approximately 1 kg. Kakeshita (1927) placed an inflatable rubber balloon within the glottis of canines and obtained pressures over 400 mm Hg during barking (also ref. Ranke and Lullies, 1953). Murakami and Kirchner (1971, 1972) placed a low frequency pressure catheter between the vocal processes of cats to measure pressure changes due to nerve stimulation. The catheter diameter was 4 mm, approximately half the length of the intercartilaginous glottis. The largest pressure measured was about 17 cm H_2O. The conditions were respiratory rather than phonatory. Other studies directly measuring interarytenoid forces have not yet been found.

Assuming that the basic motions of the vocal folds are sinusoidal, a relationship between the contact force and the frequency, effective mass, and vibrational amplitude can be obtained through the impulse-momentum law. The variation in contact force at the vocal processes should then be related to these vocal fold contact forces through the mechanical coupling of the vocal folds and the arytenoid cartilages.

Assume that the tissue motion is one-dimensional and sinusoidal, i.e.,

$$x = x_m \sin \omega t,$$

where x_m is the amplitude of motion, ω is the radian frequency, and t is time. The velocity of tissue motion is then

$$\dot{x} = \omega x_m \cos t,$$

and the magnitude of the velocity is

$$|\dot{x}| = \omega x_m$$

According to the impulse-momentum law,

$$F \Delta T = m \Delta v,$$

where F is the average force over the impulse interval Δt, m is the effective mass in motion, and Δv is the change in velocity of the mass.

Assuming the velocity goes to zero when the tissue reaches the midline position, the change in velocity (from maximum velocity just prior to impact) is

$$\Delta v = \omega x_m - 0,$$

so that the magnitude of the impulse force in the interval Δt is

$$|F| = m|\Delta v| / \Delta t = m\, \omega x_m / \Delta t$$

Therefore, contact force (and the measured contact pressure) appears to be proportional to the effective mass, vibrational amplitude, and frequency.

We suggest that if the vibrational amplitude of the vocal folds varies linearly with mean subglottal pressure (PS), then the peak-to-peak variations in vocal process contact pressure may also vary linearly with PS, as well as with frequency and effective mass. Our attempts to measure vibrational amplitudes directly have not been successful. In this report we would like to give some relevant contact pressure data.

Secondly it is reasonable to assume that a greater steady force of glottal adduction will result in a smaller effective value of pre-phonatory glottal width, and hence a greater glottal resistance to airflow. Shepherd (1974) measured flow resistance in an excised larynx at two levels of adductory force. The adductory force was created by pulling threads attached to lateral cricoarytenoid and thyroarytenoid muscle fibers superior to the region of the muscle processes. Shepherd's results indicated that flow resistance increased for the increase of adductory force at constant subglottal pressure values. We will hypothesize, therefore, that glottal resistance to airflow is related linearly to the mean DC contact pressure at the vocal processes.

EXPERIMENTAL METHODS

Tissue preparation

A bovine larynx was chosen for this pilot project because its size (approximately twice the size of a human larynx) was assumed sufficiently large to minimize abherent effects of the presence of the miniature pressure transducer within the cartilaginous glottis. The excised larynx was refrigerated for 2 days prior to experimentation. The tissue was cleaned of most extralaryngeal tissue leaving the PCA, CA, and CT muscles intact. The epiglottis was pulled superiorly and the quadrangular membrane and underlying tissues were stapled to the thyroid cartilage at the level of the superior portion of the arytenoid cartilages. This was done to prevent tissue collapse after removing the supralaryngeal tissue. The upper part of the thyroid cartilage and overlying tissues were removed about 2.8 cm above the vocal folds. An insulated laboratory clamp was placed posterior to the larynx so that the prongs could be adjusted to vary the medial-lateral position of the arytenoid cartilages (and thereby the adductory force at the vocal processes).

The data to report were obtained for runs on the second experimental day. The larynx was used for approximately 4 hours on the first day. The tissue was soaked in .9% saline solution overnight.

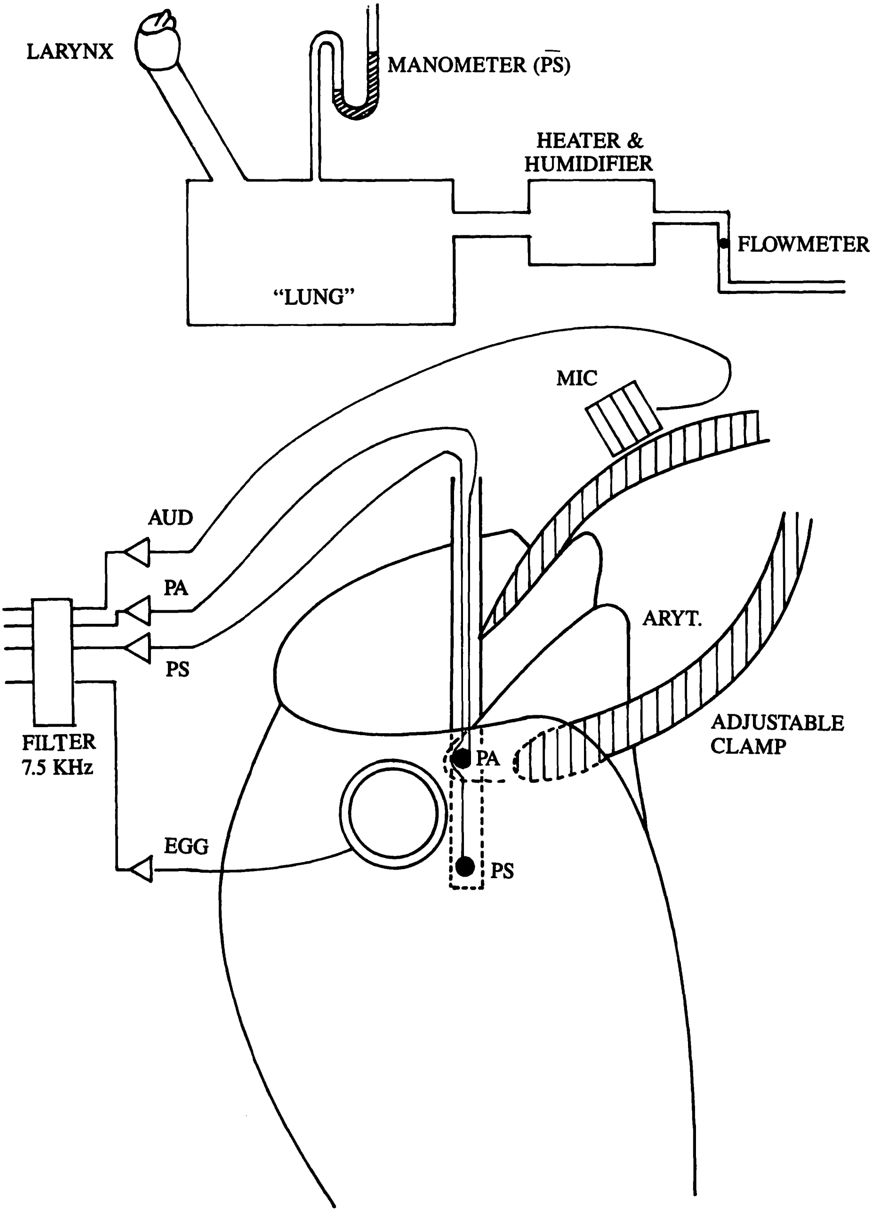


FIGURE 26-1
Schematic of the experimental arrangement

Tissue morphology

Although we have not performed extensive comparisons, the bovine larynx appears structurally different in detail from the human larynx. Not only does there appear to be greater thickness of mucosal tissue, but also there appears to be no ventricular fold (cf. Kurita *et al*, 1983). In addition, there is possibly a different orientation between vocalis and thyroarytenoid muscles. These differences may lead to different vibratory modes than expected for the human larynx. However, it may also be expected that the gross mechanical behavior of the bovine larynx is similar to that of the human and canine larynges which have been studied more extensively.

Subglottal system

Figure 26-1 illustrates the experimental arrangement. An 8 cm section of trachea was preserved and attached to the spout of a "pseudo-lung" container having a volume of approximately 4 liters. The first subglottal resonance of this system with larynx attached and vibrating was approximately 450 Hz. The cricoid cartilage was stabilized by placing nylon probes with projecting needles into the cricoid from 4 sides.

An air compressor system supplied constant air flow to a Gilmont rotameter used for measuring the volume flow. The air passed through an RCI Concha Therm III® servo controlled heater. This system is designed to create a 100% relative humidity condition through evaporation of water from watersoaked paper within the air channel. The air is heated by a heating column surrounding the air. The RCI system also offers a slow response temperature control of the air downstream at the site chosen for the location of a thermister probe. There was negligible resistance through the Concha Therm III® equipment.

Transducers and monitoring

As mentioned, mean air flow was measured by the rotameter. Mean subglottal pressure (PS) was measured by a water manometer. A tube for this purpose led to the pseudo-lung. A subglottal pressure correction was made to correct for the presence of a wire screen in the spout of the container. A Fourcin electroglottograph was attached to the external lamina of the thyroid. A Realistic Condenser Microphone (No. 33-1056A) was attached to the arytenoid clamp. During the experimental runs, the microphone was approximately 7.3 cm from the vocal folds.

Two Entran ultra-miniature pressure transducers (model EPIL-F080-2S) were used to measure the interarytenoid pressure and the subglottal pressure, respectively. The response of these silicon transducers is reportedly flat within ±5% between DC and 20K Hz. The linear range was up to ±140 cm H_2O.

Figure 25-2 shows the left side of the experimental larynx separated medially. The two Entran transducers were mounted on a metal shim and placed in the larynx. One transducer was placed on the vocal process and the other subglottally.

The pressure transducers were attached to the thin metal shim by means of double-sided gummed tape. The total thickness of each probe was approximately 0.1 cm. The A-P width of the shim at the vocal process was about .32 cm. The transducers were calibrated under constant air temperature conditions using static pressures. Recalibration over days and after repeated use was within 2%.

The tracheal temperature was measured using a fine-wire thermocouple inserted into the trachea through the tracheal wall and monitored with a Fluke 2240C Datalogger. The time varying signals (pressure between the vocal processes, subglottal pressure, electroglot-

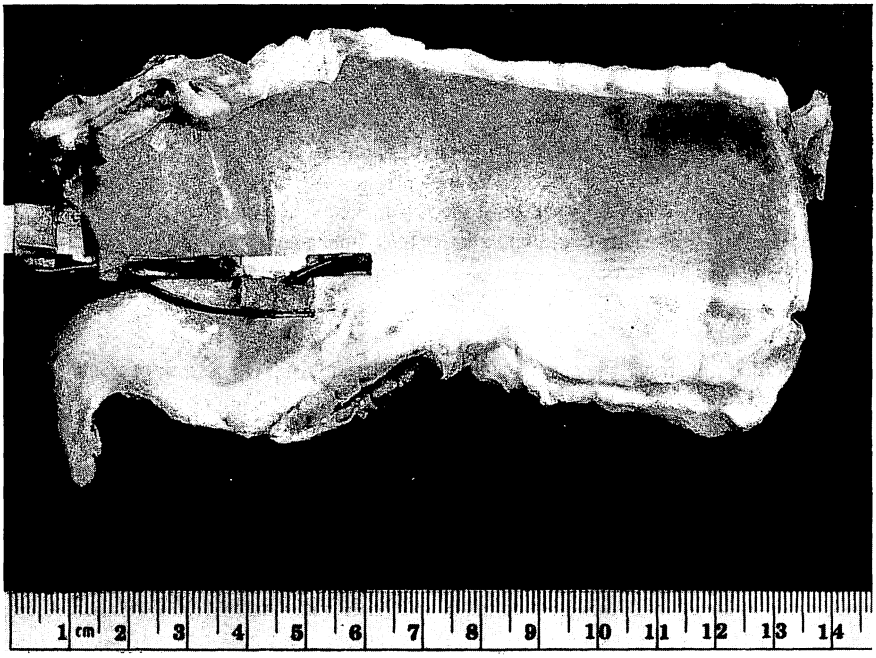


FIGURE 26-2
Left half of excised bovine larynx with pressure transducer placed on the vocal process and within the subglottal space.

tographic signal, and the audio signal) were monitored on a Data Precision model Data 6000 digital oscilloscope. The signals were amplified, low-pass filtered at 7.5 kHz, and digitized. A PDP 11/44 computer and a 12-bit LPA/11 converter were used. A sampling rate of 15625 samples per second per each of the 4 channels was used. Analysis of the digitized signal was performed using a VAX 11/780 computer and our own software display programs.

Measurement sequence

An experimental sequence generally was the following. The glottis was open with no phonation but with air flowing at a low rate. The small pressure transducers were positioned by eye at the vocal processes and electronically zeroed. The flow was then disconnected from the system so that the mean interarytenoid pressure at the vocal processes $(\overline{PA})$ could be set in the absence of vibration. The flow was then reattached. At the specified $\overline{PA}$ level a series of (2, 3, or 4) levels of flow and subglottal pressure were used. At each level the variables were measured or digitized. After the run for each specific $\overline{PA}$ level, the transducers were checked for zero shift.

LIMITATIONS OF THE STUDY

Phonatory conditions

The only external forces applied to the laryngeal tissues were the effective adductory forces produced by the arytenoid clamp. Phonation appeared to be in the "chest register"

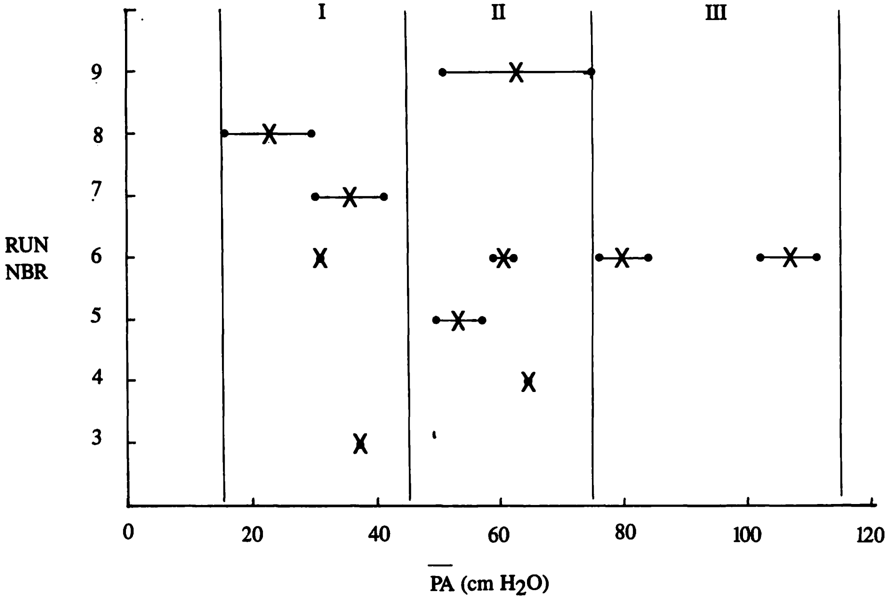


FIGURE 26-3
Levels of mean interarytenoid pressure ($\overline{\text{PA}}$) used in each experimental run. Three ranges of $\overline{\text{PA}}$ are designated, I: 15-45 cm H_20; II: 45-75 cm H_20; III: 75-115 cm H_20.

range because there was vertical phasing of the vocal fold tissue as well as a relatively long closed phase. There were no forces imitating active vocalis contraction, and no innervation of the larynx.

Pressure transducers

The Entran transducers initially were not sufficiently waterproofed. It was necessary to cover the transducer at the vocal processes with petroleum jelly to prevent water absorption. This coating did not appear to affect the vocal fold vibration characteristics.

The attempt was made to position the transducer near the anterior end of the vocal processes at a location previously marked with a pen. No method was used to ensure that the vocal process transducer was placed in the same position for each run. If the transducer were positioned too low, the subglottal pressure would be seen; if too high, a signal resembling the audio signal would appear.

The arytenoidal clamp was not attached to the arytenoids, but pressed against the lateral arytenoid surfaces. Any slipping of the clamp on the artenoids would have led to undesired changes in the interarytenoid pressure being measured.

The most serious limitation was the transient temperature sensitivity of the Entran transducers. Although the transducers were calibrated for static temperature effects, it was discovered that there was large sensitivity to abrupt temperature changes. Movement of the transducers from above the larynx to a position inside the larynx changed the ambient temperature from about 24°C to 34°C. A maximum estimate of the stabilization period was 3 minutes. A more typical stabilization period for the experiments would appear to be ap-

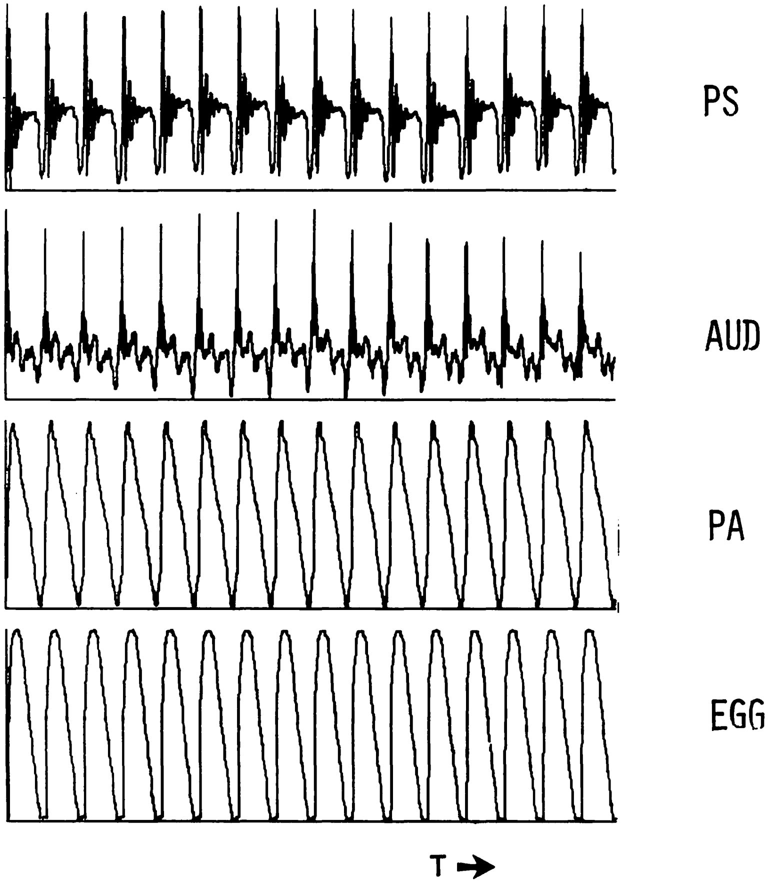


FIGURE 26-4
Example of scaled time-varying data display. PS, subglottal pressure; AUD, output acoustic signal; PA, interarytenoid pressure at the vocal process; EGG, electroglottograph signal. The case shown in the figure corresponds to $\overline{PS}$ = 11.7 cm H_2O, $\overline{PA}$ = 63.1 cm H_2O, and PA_{PKPK} = 24.5 cm H_2O.

proximately 1 minute. Accuracy of any specific value of interarytenoid pressure was estimated to be within ±6 cm H_2O (ref. Figure 26-3).

RESULTS AND DISCUSSION

Levels of interarytenoid pressure

Figure 26-3 shows the mean interarytenoid pressure levels $\overline{PA}$ used in the experiment according to the number of the experimental run. All runs except No. 6 were to have a specific PA level. The ranges shown indicate the change from the initial to final mean level for the given run. As stated before, this shift was probably caused by the transient temperature sen-

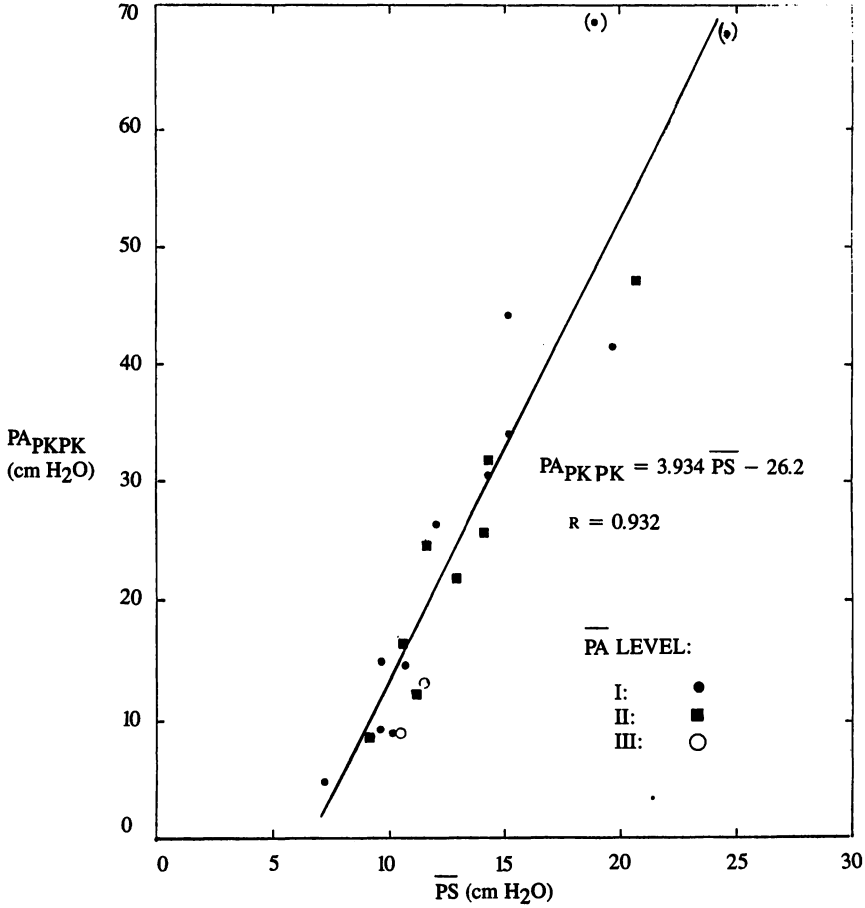


FIGURE 26-5
Relationship between mean subglottal pressure ($\overline{PS}$) and peak-to-peak variations of interarytenoid pressure (PA_{PKPK}).

sitivity of the Entran pressure transducer. The midpoints of the ranges were taken as the representative $\overline{PA}$ value.

The $\overline{PA}$ values were grouped into 3 ranges as shown in Figure 26-3: I, 15-45 cm H_2O; II, 45-75 cm H_2O; III, 75-115 cm H_2O. The vocal processes were fully approximated for all runs.

Display of the time varying data

Figure 26-4 illustrates the display of the scaled time-varying data. The waveforms from top to bottom are the subglottal pressure (PS), the audio signal (AUD), the interarytenoid pressure (PA), and the electroglottographic signal (EGG). In this report we will not be concerned with the audio or EGG signals (ref. Scherer, Cooper, Alipour-Haghighi, and Titze, 1984). Our programs allowed us to analyze sections of figures such as these.

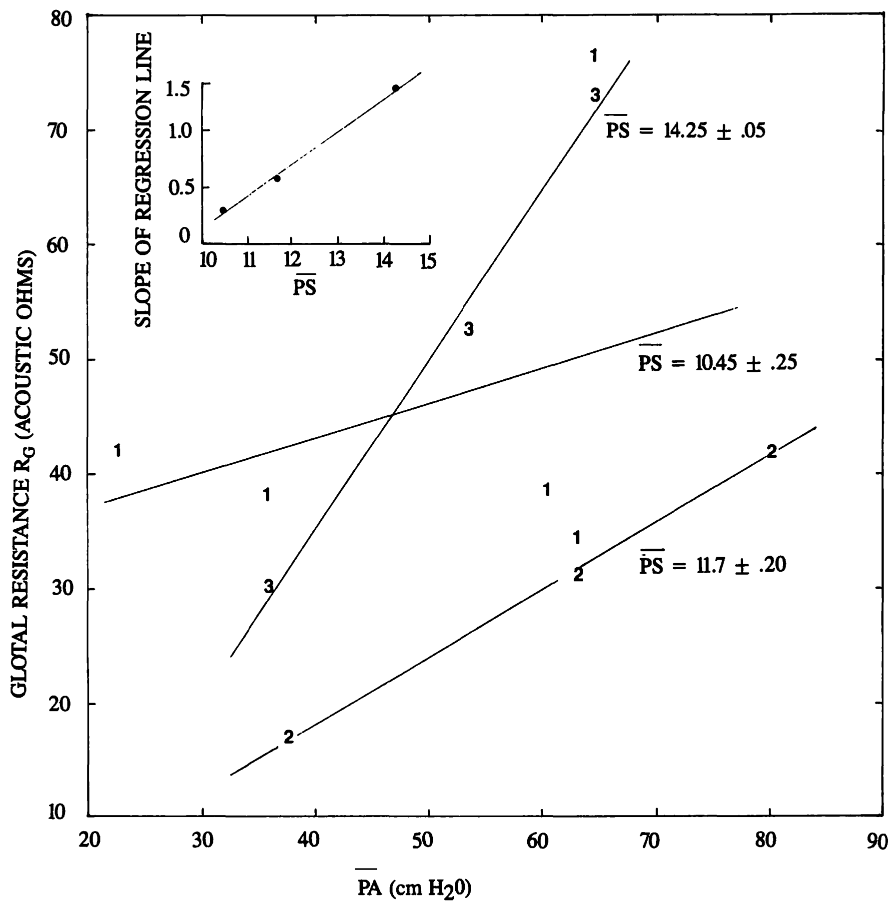


FIGURE 26-6
Relationship between mean interarytenoid pressure ($\overline{PA}$) and glottal flow resistance (R_g). R_g = $\overline{PS}/\overline{U}$, where $\overline{U}$ is mean flow.
27.

Figure 26-5 shows that peak-to-peak values of interarytenoid pressure (PA_{PKPK}) were strongly related to mean subglottal pressure ($\overline{PS}$). The linear fit appears to be strong in spite of the problematic reliability of run-to-run transducer replacement within the vocal process region. The figure suggests that there may be a threshold of $\overline{PS}$ below which the arytenoid cartilages are not forced to vibrate. It is noted that the values of PA_{PKPK} may be different from the human larynx values, although the range of $\overline{PS}$ values is reasonably similar.

The relationship shown in Figure 26-5 appears to be independent of mean interarytenoid pressure. According to the discussion in the Introductory section, the peak-to-peak variation in pressure at the vocal processes may be related to vibrational amplitude of the tissues, as well as to frequency and effective tissues mass. Subsequent studies should treat fundamental frequency as an independent variable and glottal tissue movement should be estimated. Furthermore, a more complete model of the arytenoid cartilage dynamics should be prepared.

Figure 26-6 shows the relationship between mean interarytenoid pressure (PA) and glottal resistance to airflow (R_g) measured in acoustic ohms [1 acoustic ohm = 1 (dyne/cm^2) /(cm^3/s) $\approx 10^{-3}$ cm H_2O/(cm^3/ s)]. R_g was measured as the ratio of the mean subglottal pressure to the mean airflow. In Figure 26-6, results are shown for 3 levels of subglottal pressure. Despite a small number of data points, the figure indicates that: (1) at a particular $\overline{PS}$ value, increasing interarytenoid pressure resulted in an increase in glottal flow resistance, and (2) for increasing values of subglottal pressure, the slope of resistance vs interarytenoid pressure increased. These results suggest that interarytenoid pressure may affect control over flow resistance, and this control may be more sensitive as subglottal pressure increases. Because of the paucity of data points, linear relationships as shown in Figure 26-6 may be inappropriate models, especially if mean interarytenoid pressures were allowed to go to zero. The outlander datum point for the smallest subglottal pressure case eludes explanation. A wider range of PA and $\overline{PS}$ values with more data points for each $\overline{PS}$ should be obtained in future work.

CONCLUSIONS

The pilot study reported here discusses relationships among interarytenoid pressure, subglottal pressure, and glottal resistance to air flow obtained within an excised bovine larynx. A linear relationship was found between the peak-to-peak variations in interarytenoid pressure (at the vocal processes) and mean subglottal pressure. This result may be relevant to the creation of vocal pathologies (Scherer *et al*, 1984). Future work should include kinematic measures of vocal fold motion, treat fundamental frequency as an independent variable, and test a suitable model of arytenoid cartilage dynamics.

Secondly, the simple notion of a linear relationship between mean interarytenoid pressure and airflow resistance at constant values of mean subglottal pressure appears to have some support from these data. The data are strictly preliminary, however. Furthermore, it is noted that in excised larynx studies of this type, there is no simulation of vocalis muscle activity, and no separation of the inherent effects from the interarytenoid and lateral crico-arytenoid muscles.

Thirdly, the Entran pressure transducers are useful in an experiment such as this, but difficulties such as undesirable temperature response and susceptibility to moisture must be overcome. Also, the artifactual effects of the presence of the transducers within the larynx must be studied further.

ACKNOWLEDGEMENT

This work was supported by the National Institutes of Health Grant NS-16320. We wish to thank John Barnes and James McGowan from the Respiratory Therapy Department, University of Iowa Hospitals and Clinics, for the use of their equipment, and John Dyson of the Medical Instrument Shop, Univeristy of Iowa Hospitals and Clinics, for construction of apparati.

REFERENCES

Entran Devices, Inc. 10 Washington Avenue, Fairfield, N.J., 07006, (201) 227-1002, Telex 130361.

Kakeshita, T. (1927). "Uber eine neue Methode zur Messung der beim Stimmbandverschluss wirkenden Krafte. I. Mitt," *Pflügers Arch. für die gesamte Physiologie* 215, 19.

Kurita, S., Nagata, K. and Hirano, M. (1983). "A comparative study of the layer structure of the vocal fold," in *Vocal Fold Physiology, Contempory Research and Clinical Issues*, D.M. Bless and J.H. Abbs (Eds.), College-Hill Press, San Diego, 3-21.

Murakami, Y. and Kirchner, J.A. (1971). "Reflex tensor mechanism of the larynx by external laryngeal muscles," *Ann. Otol.* 80, 46-64.

Murakami, Y. and Kirchner, J.A. (1972). "Mechanical and physiological properties of reflex laryngeal closure," *Ann. Otol.* 81, 59-71.

Ranke, O.F. and Lullies, H. (1953). *Gehor-Stimme-Sprache*. Springer-Verlag, Berlin.

Rethi, L. (1897). "Die Stimmbandspannung, experimentell gepruft," *Sitzgsber. Akad. Wiss. Wien, Math.-naturwiss.* Kl. III 106, 244.

Scherer, R.C., Cooper, D.S., Alipour-Haghighi, F. and Titze, I.R. (1984). "Vocal process contact pressures," in *Transcripts of the Twelfth Symposium: Care of the Professional Voice*. V. Lawrence (Ed.), The Voice Foundation, New York.

Shepherd, W. (1974). Comparison of Mechanical Performance of Normal and Simulated Non-normal Excised Human Larynges. Doctoral diss., The Univ. of Connecticut.

von Leden, H. and Moore, P. (1960). "Contact ulcer of the larynx: experimental observations," *Arch. Otolaryngol.* 72, 746-752.

DISCUSSION

M. HAST: Why did you use a bovine larynx in your study?

R. SCHERER: A bovine larynx was used for two primary reasons: availability and size. We wanted to maximize the difference between the size of the tissue and the size of the transducers. The transducers we used were the smallest I could find with sufficient characteristics to make the study feasible. We are now working more with dog larynges. Human larynges are difficult for us to obtain.

N. ISSHIKI: I'd like to make a comment from the clinical viewpoint. I think this kind of experiment is very important for understanding functional disorders. I sometimes perform operations requiring arytenoid adduction. For instance, a paralyzed vocal cord can be moved medially by surgically displacing the arytenoid cartilage. If we pull the arytenoid cartilage too much, the voice can become hoarse. As clinicians, we want to know what is the critical or optimal closure of the glottis without producing hoarseness. We have no information on that. In that sense, I would like to see you perform research on that aspect. What is the optimum pressure for producing voice? Three years ago I performed a similar experiment using an adult larynx, but without measuring contact pressure directly. We pulled the arytenoid cartilages using strings and weights. I am very anxious to know information about the contact pressure.

R. SCHERER: I think that it is not infeasible to attach a transducer such as the one we used to the end of a fiberscope in order to place the transducer between the arytenoids or even within the membranous portion of a human larynx. That is the direction in which we wish to go. The present study constitutes preliminary work in this area.

M. HIRANO: Were there any differences in acoustic output or vocal fold vibratory motion between the condition where there was no systematic relationship between the mean contact pressure and glottal resistance, and the conditions where the mean contact pressure was positively related to the glottal resistance?

R. SCHERER: We did not closely examine the vibratory motion of the vocal folds in this preliminary study. We were not sufficiently equipped to record stroboscopic views. It may be of interest, however, to note that the rise time (normalized by the cycle period) of the EGG signal during the closing portion of the glottis was not related to the mean vocal process contact pressure. This may imply that the contact pressures were so large that the study was relatively insensitive to certain motion aspects. Likewise, the acoustic spectra that we obtained were not consistently unique across the various conditions. We did not have the resources to measure overall acoustic intensity in this experiment.

C. LARSON: Did your clamp restrict movement of the arytenoid cartilages? To the extent that the arytenoids may participate in vibratory behavior, your clamp may have introduced an abnormal variable into your setup. Did you consider using a spring system to simulate muscle forces?

R. SCHERER: The use of an arytenoid cartilage clamp may effectively create the correct force vector for adduction, but, as you suggest, may also restrict the motion of the arytenoid cartilages. We initially secured sutures to muscle attachments in order to control adductory and vocal fold lengthening tensions. This was not completely satisfactory, perhaps because of the relatively stiff joints of the bovine larynx used. We then decided to use the clamp under the assumption that the basic relationships would be brought out by the study. For more subtle control of laryngeal configuration and greater insight into the coordination of muscle forces, we are planning to use the individual muscle simulation approach.

J. KAHANE: Have you thought about what effect the differences in bovine and human cricoarytenoid joint morphology and mechanics might have on your ability to extrapolate your data to human vocal function?

R. SCHERER: We have made the assumption that the differences between the bovine and human laryngeal morphology are not sufficiently great that the basic functional relationships would be significantly different. This assumption must be closely examined as further work is done, including a comparison of joint morphologies and mechanics.

We are inclined to think that there may be some distinct differences between the bovine and human larynx besides size. We typically see little ventricular fold tissue in the bovine larynx and a relatively large mucosal layer. Also we suspect that the orientation between the vocalis and thyroarytenoid muscles may be slightly different. On the other hand, we have some data to suggest that functionally the bovine larynx is similar to the human and dog excised larynges. For example, the relationship between fundamental frequency and subglottal pressure was found to be similar to Tom Baer's early data showing a 5 to 8 Hz/cm H_20 relationship, with asymptotic behavior at large subglottal pressure values. Also, the AC subglottal pressure variation was consistently about 35% of the mean subglottal pressure, a value that matches well the range of about 31% tp 37% found by Dr. Kitzing for human subjects.

27 ▪ MECHANICAL PROPERTIES OF THE HUMAN VOCAL FOLD — RESONANCE CHARACTERISTICS IN LIVING HUMANS AND IN EXCISED LARYNXES

Toshio Kaneko, Kensuke Komatsu, Haruhiko Suzuki, Toshiaki Kanesaka, T. Masuda, T. Numata and J. Naito

In the previous study, we have demonstrated a new method of measuring the resonance characteristics of the vocal fold in vivo and showed the resonance pattern during "quiet" respiration and in the phonation neutral gesture in normal subjects.

Before reporting the latest results, some figures demonstrated at Kurume, are partially presented here again.

APPARATUS AND METHOD

Figure 27-1 shows the equipment used. The apparatus are roughly divided into 2 systems, the ultrasonics and vibrator system. The equipment used for the ultrasonic system is almost the same as that used for ultrasonoglottography. However, the use of the equipment is quite different.

Figure 27-2 shows a block diagram of system for observing movement of the vocal fold in response to a vibrator system.

The echo from the marginal surface of the vocal fold was selected by an echo selector, was sent to a cathode rays oscilloscope and was swept by a slow sweep generator. As a recorder an electromagnetic oscilloscope was used. For observing the minute displacement of the vocal fold caused by the forced oscillation given to the framework of the larynx, we display the amplitude variation of echo signals as a function of frequency, that is variant M-mode display.

For the vibrator system a low frequency sinusoidal oscillator and mechanical shaker were utilized. The frequency of the oscillator was manually changed from 30 to 300 Hz.

RESONANCE PATTERN OF THE VOCAL FOLD IN NORMAL SUBJECT

Figure 27-3 shows resonance patterns of the vocal fold in two male adults during "quiet" respiration.

The differences of the vocal fold resonance among normal male adults, female adults and children were examined.

Figure 27-4 shows these results: in male adults, resonance frequencies of the vocal fold ranged from 91 to 145 Hz (average 128 Hz) and in female adults from 115 to 167 Hz (average 136 Hz). There seemed to be a trend indicating that higher resonance frequencies were observed in female adults. In children also, higher resonance frequencies were recorded.

During phonation at various pitches, on the contrary, the biomechanical factors of the vocal folds have to change depending on muscle control. Dynamic properties of the vocal folds during phonation under different conditions are important for analyzing the mechanism of phonation. For this purpose the resonance characteristics of the vocal folds in phonation neutral positions at several voice pitches were measured.

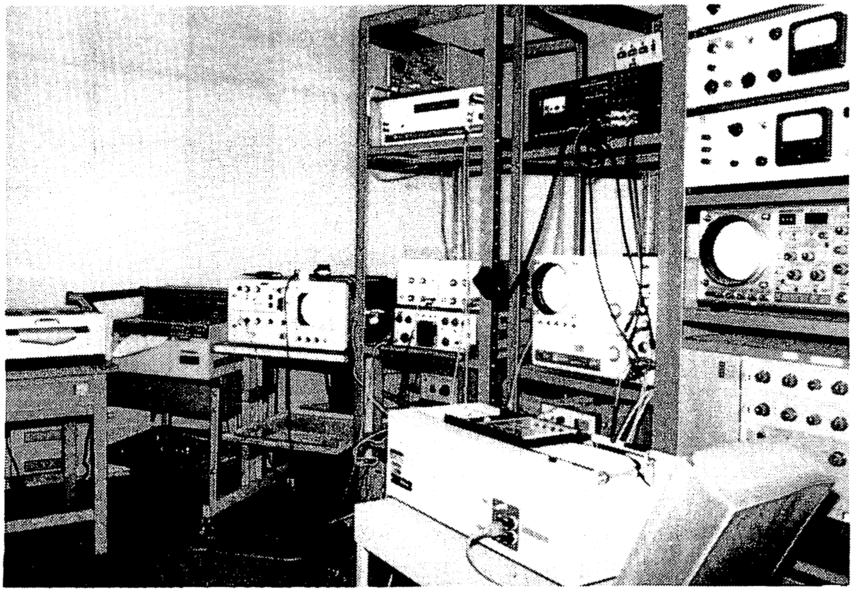


FIGURE 27-1
Apparatus.

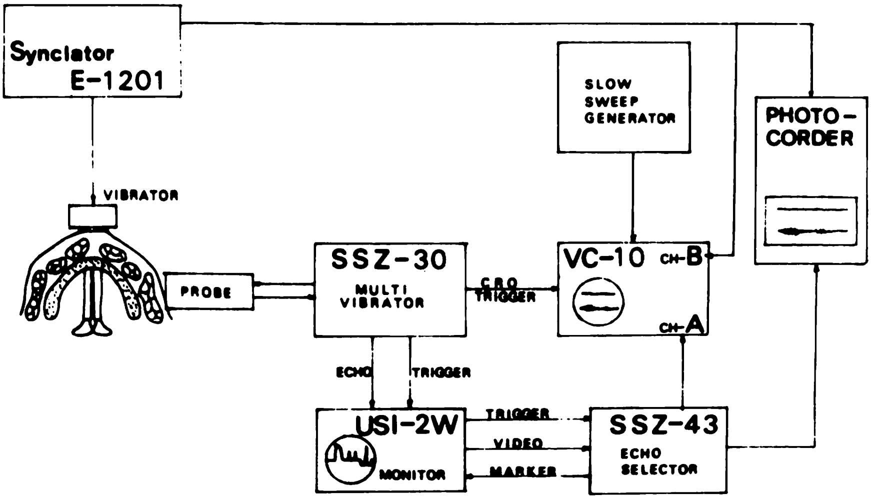


FIGURE 27-2
Block diagram.

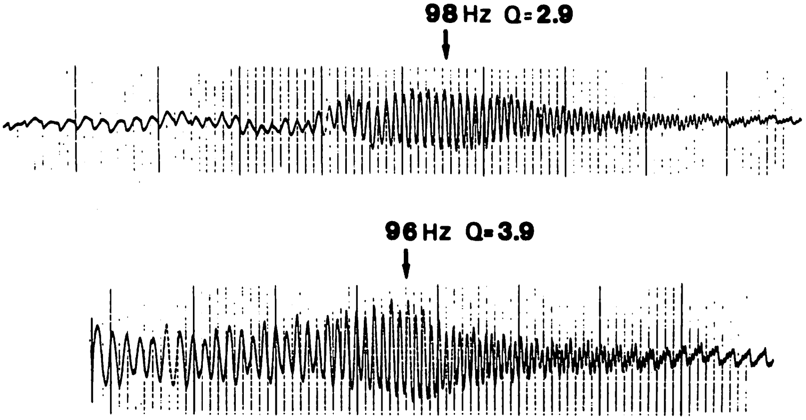


FIGURE 27-3
Two examples of resonance patterns for male vocal folds during quiet respiration.

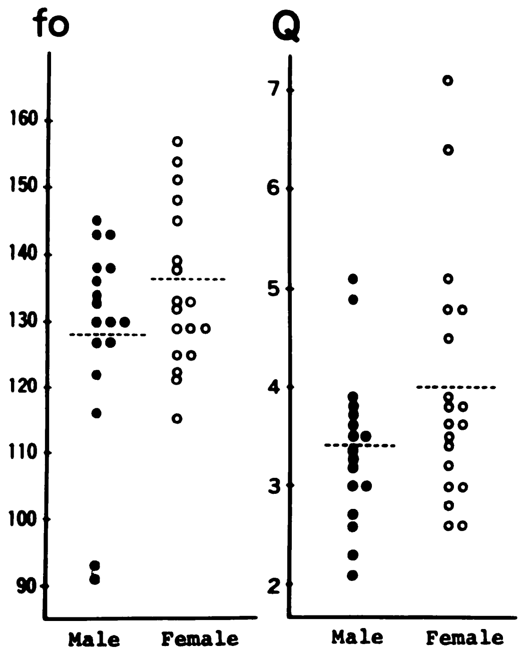


FIGURE 27-4
Resonance frequencies fo (Hz) and Q values during quiet respiration.

In Figure 27-5, the upper half shows a resonance pattern of a male vocal fold, just before phonation at a low pitch, in this case 100 Hz. A resonance frequency occurred at a point close to the phonation pitch, i.e. 100 Hz. The bottom half of the figure shows a resonance pattern of the vocal fold in the same subject, just before phonation at 150 Hz. Two distinct resonance frequencies can be observed in this case. The second resonance frequency occured at a point close to the phonation pitch, i.e., 150 Hz. Figure 27-6 shows resonance patterns of two female vocal folds. Two resonance points usually occured in female vocal folds. The lower resonance point was observed around 100 Hz, and the second one at over 200 Hz. The second resonance frequency is almost equal to the phonation pitch and the second resonance waveform is steeper than the first.

During "quiet" respiration the vocal fold is relatively relaxed. However, once the vocal fold is tensed, such as in the phonation neutral position, the mechanical properties are changed significantly, sex being a relevant factor.

RESONANCE PATTERN IN FRESH EXCISED LARYNXES

Similar examinations were performed in excised human larynxes with a cancer of the hypopharynx. The vocal folds of these specimens were intact.

The transducer was placed on the prelaryngeal muscles in order for ultrasonic beam to irradiate the margin of a vocal fold perpendicularly. The vibrator was pressed directly on the middle of the thyroid cartilage.

Effect of Cord Tension

Figure 27-7 shows the resonance pattern of the vocal fold in excised larynx without longitudinal tension (upper trace) and with tension of 160 gr (lower trace). A resonance frequency of the vocal fold in excised larynx without longitudinal tension occurred approximately at 100 Hz. When a longitudinal tension is applied, two resonance frequencies appear.

Figure 27-8 shows the relationship between the longitudinal tension and resonance frequencies. The stronger the tension, the higher the second resonance frequency was observed. In addition, the second resonance is sharper than the first. The lower resonance frequency, however, was almost constant although the longitudinal tension is applied.

Effect of Mass Factor

A small lead particle ranged from 10 to 65 mg was attached on the surface of a vocal fold, and resonance patterns of the vocal fold without/with a lead particle were analyzed in order to clarify the influence of mass factor.

Figure 27-9 shows the resonance patterns of the vocal fold without a lead particle (upper trace) and with a lead particle of 42 mg (lower trace). Lowering the resonance frequency was confirmed.

Figure 27-10 shows the relationship between loaded mass and resonance frequencies. Resonance frequencies without a lead particle and with lead particles of 30 mg, 60 mg occurred at near 100 Hz, 90 Hz and 80 Hz respectively. In addition, sharpness of the resonance has a tendency to decrease as the loaded mass increases.

From this data, the effective mass for human vocal folds can be estimated to be of the order of 0.127 gr. This value is incidentally the same order as deduced one in the earlier simulation by Dr. Ishizaka.

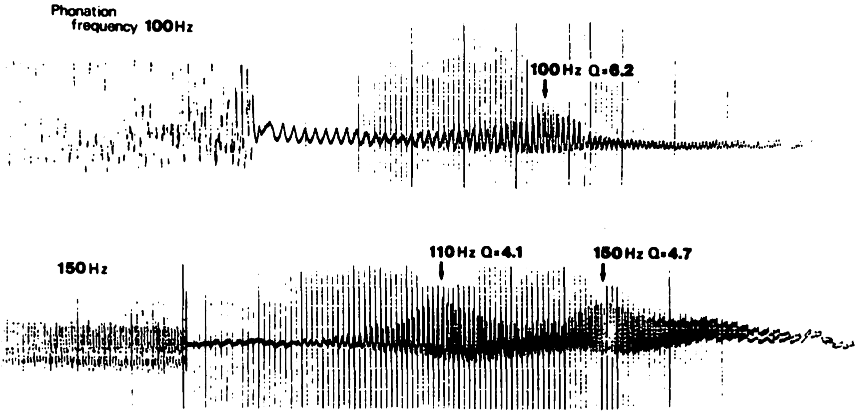


FIGURE 27-5
Examples of resonance patterns for male subjects with vocal fold in phonation neutral position.

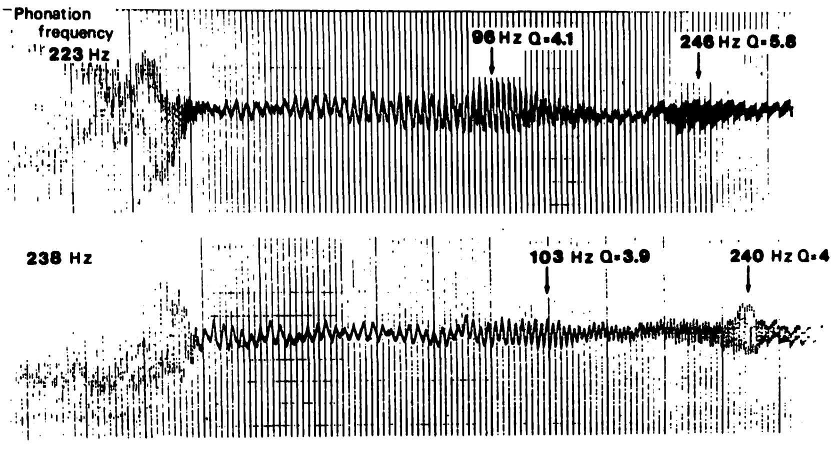


FIGURE 27-6
Examples of resonance patterns for female subjects with vocal fold in phonation neutral position.

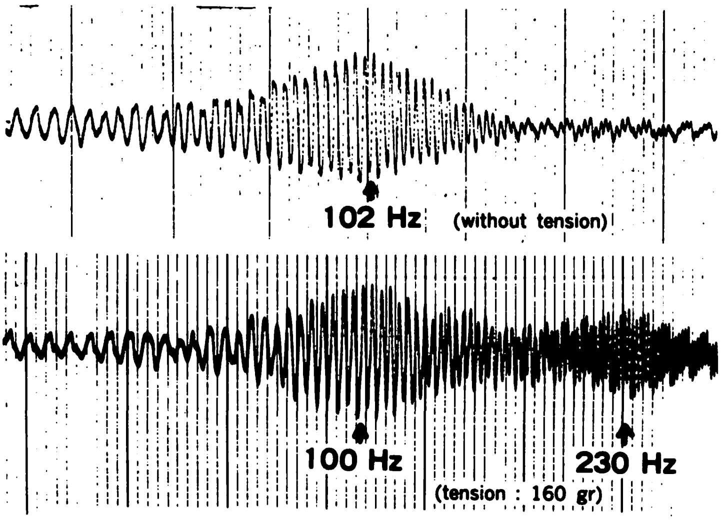


FIGURE 27-7
Resonance patterns without/with longitudinal tension.

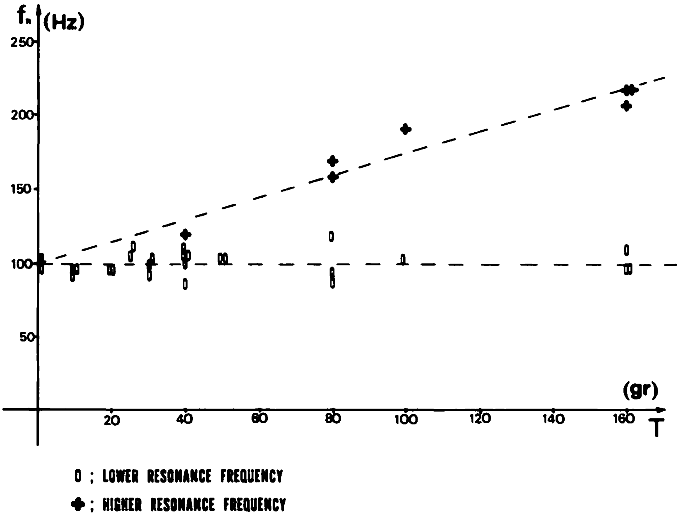


FIGURE 27-8
Relationship between longitudinal tension and resonance frequency.

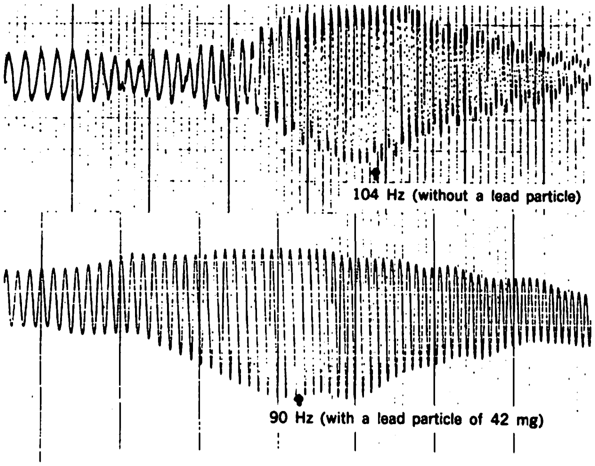


FIGURE 27-9
Resonance patterns of vocal fold in excised larynx without/with a lead particle.

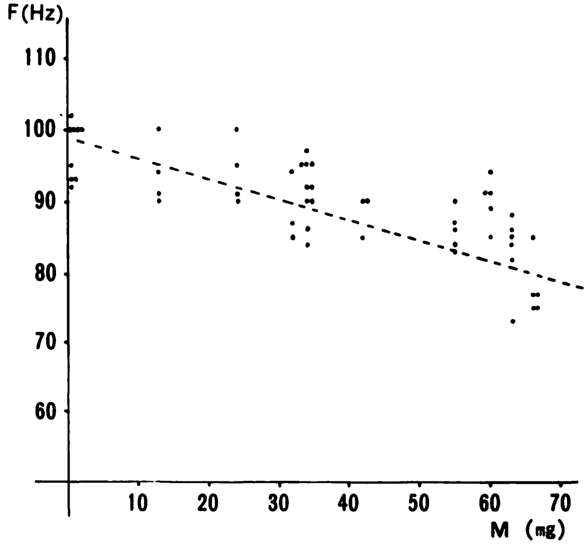


FIGURE 27-10
Relationship between loaded mass and resonance frequency.

RESONANCE PATTERNS OF THE VOCAL FOLDS WITH UNILATERAL VOCAL FOLD DISEASES DURING "QUIET" RESPIRATION

Resonance waveforms of both vocal folds with unilateral vocal fold polyp, recurrent laryngeal nerve paralysis and glottic cancer were measured during "quiet" respiration. These measurements make it possible to clarify a right and left difference of the mechanical properties of both vocal folds.

Figure 27-11 shows the results of resonance frequencies measured between normal vocal fold and with a laryngeal polyp in each patient. In the cases of polyp (10 males, 15 females), lowering resonance frequency of the vocal fold polyp in comparison with healthy side was detected. Average frequencies were f = 118.1 ± 12.0 Hz for the vocal fold with polyps and f = 131.8 ± 11.9 Hz for the normal side.

Figure 27-12 shows the relationship between the volume of polyp and the resonance frequencies. The larger the volume of polyp, the lower the resonance frequency is.

Figure 27-13 shows the results of resonance frequencies measured in normal side and vocal fold with a glottic cancer in each patient. In the majority case of glottic cancer (19 males only), rising resonance frequency of vocal fold with cancer in comparison with healthy side was detected. Average frequencies were f = 134.5 ± 16.9 Hz for the vocal fold with cancer and f = 118.5 ± 12.4 Hz for the other side.

Figure 27-14 shows the relationship between the mobility of vocal fold and the resonance frequencies. In glottic cancer, the mobility of vocal fold will be disturbed according to the infiltration of cancerous tissue into the deep layer. Average resonance frequency of glottic cancer with normal mobility, impaired mobility and fixation were 125.4 Hz, 131.5 Hz and 140.5 Hz respectively. The higher disturbance of the mobility, the higher resonance frequency is.

Figure 27-15 shows the results of resonance frequencies measured in normal side and vocal fold with recurrent laryngeal nerve paralysis in each patient. In the majority cases of recurrent laryngeal nerve paralysis (10 females, 4 males), lowering resonance frequency of the vocal fold with paralysis in comparison with the healthy side was detected. Average frequencies were f = 106.1 ± 20.0 Hz for the vocal fold with paralysis and f = 124.3 ± 20.7 Hz for the other.

Theoretically, the resonance frequency of the vocal fold is determined by its mass and stiffness. Lowering resonance frequency of the vocal fold with polyp is considered to be an effect of increased mass. Rising resonance frequency of the vocal fold with cancer, on the contrary, is due to increased stiffness. Falling resonance frequency of the vocal fold with recurrent laryngeal nerve paralysis can be attribute to decreased stiffness rather than atrophy, i.e., decreased mass.

MECHANICAL PROPERTIES OF THE VOCAL FOLD EXCITED BY A SINGLE RECTANGULAR PULSE IN PHONATION NEUTRAL GESTURE

Instead of above-mentioned method which provokes an oscillation of the vocal fold with continuous wave from 30 to 300 Hz, thyroid lamina was excited by a short rectangular impulse of 0.5 ms and an oscillation pattern of the free margin of the vocal fold was recorded.

The data so far obtained are only preliminary, and rigorous study of several conditions are necessary. The same results obtained in the phonation neutral gesture, however, will be presented here.

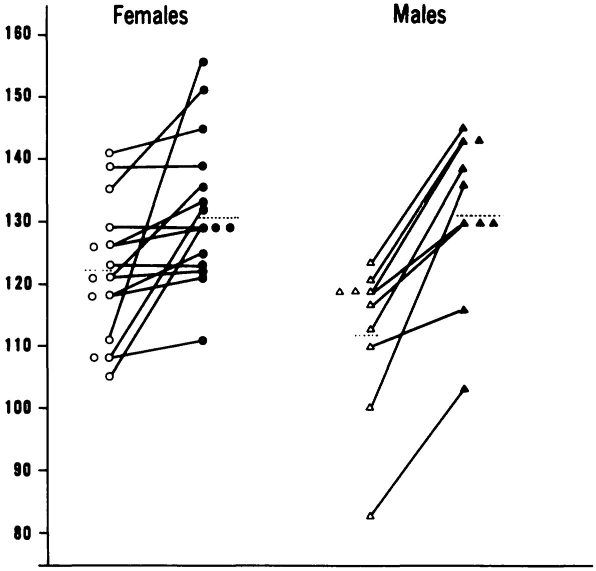


FIGURE 27-11
Resonance frequencies of normal vocal folds (● ▲) and vocal folds with polyp (O △).

		Average Resonance Frequencies of the Vocal Folds	
Volume of polyp	Number of the cases	with a polyp	normal
small	5	126.0 Hz	131.2 Hz
medium	14	118.1 Hz	132.1 Hz
large	4	116.3 Hz	137.5 Hz

FIGURE 27-12
Relationship between apparent volume of polyp and resonance frequency.

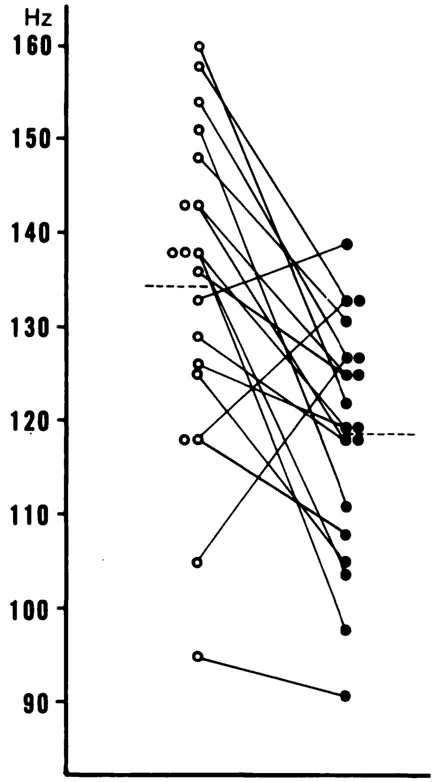


FIGURE 27-13
Resonance frequencies of normal vocal folds (●) and vocal folds with glottic cancer (○).

		Average Resonance Frequencies of the Vocal Folds	
Mobility	Number of the cases	with a cancer	normal
normal	11	132.2 Hz	119.1 Hz
impaired	4	134.8 Hz	115.5 Hz
fixed	4	140.8 Hz	120.0 Hz

FIGURE 27-14
Relationship between mobility of vocal fold and resonance frequency.

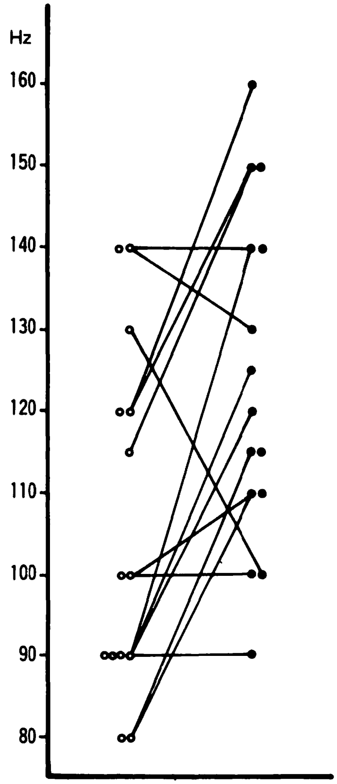


FIGURE 27-15
Resonance frequencies of normal vocal folds (●) and vocal folds with recurrent laryngeal nerve paralysis (O).

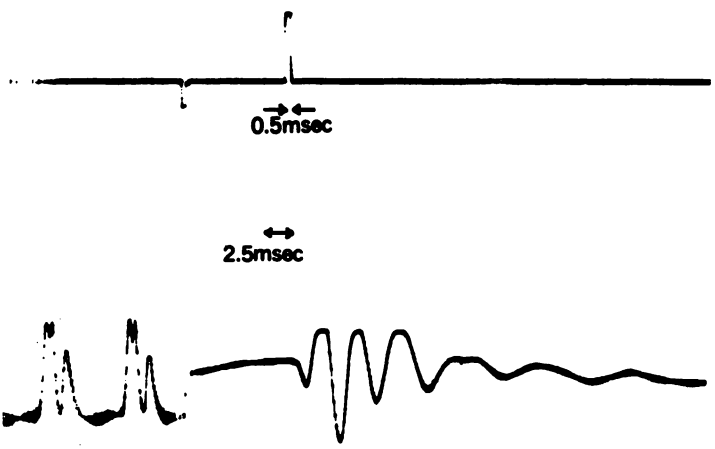


FIGURE 27-16
A single rectangular pulse stimulus of 0.5msec (upper trace) and an oscillation pattern of vocal fold (lower trace).

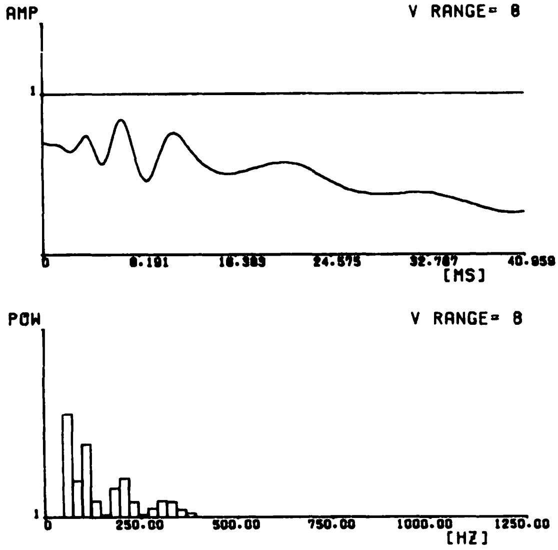


FIGURE 27-17
An oscillation pattern of vocal fold in phonation neutral position (upper trace) and a power spectrum analysed by FFT (lower trace).

Figure 27-16 shows a single rectangular pulse (the upper trace) and an oscillation pattern of the free margin of the vocal fold (the lower trace, right) and phonation pitch (the lower left). The portion of low frequency of this oscillation pattern varies according to phonation pitch, and the frequency of this part is almost equal to the phonation pitch.

Figure 27-17 shows an oscillation pattern of the vocal fold excited by a single pulse, just before phonation at 94 Hz (the upper trace) and a power spectrum analyzed by fast Fourier transform (the lower trace).

Although further work is needed, this method will be also available to evaluate the biomechanical properties of the vocal fold in living humans. The detail of these results will be presented in the near future.

REFERENCES

Kaneko, T., Komatsu, K. et al (1981). "Mechanical properties of the vocal fold - measurements in vivo," Stevens, K.N. and Hirano, M. (Eds.) *Vocal Fold Physiology.* Tokyo: Univ. of Tokyo Press, 365-376.

Komatsu, K., Kaneko, T. *et al* (1979). "Physical Properties of the Vocal Fold of Extirpated Human Larynx," The Japanese Society of Ultrasonics in Medicine (JSUM) Proceedings, 287-288.

Komatsu, K., Kaneko, T. *et al* (1982). "Physical Properties of the Vocal Fold of Extirpated Human Larynx," The Japanese Society of Ultrasonics in Medicine (JSUM) Proceedings, 263-264.

Ishizaka, K. and Flanagan, J.L. (1972)."Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* 51, 1233-1268.

DISCUSSION

S. HORIGUCHI: Is there any relationship between voice quality and the difference of resonance frequency and Q-value between one side and the other? And if there is some relationship, is your method useful for the determination or evaluation of voice quality?

T. KANEKO: Direct relationships between voice quality and resonance frequencies of the vocal fold cannot be determined by this experiment. The resonance characteristics were measured in phonation neutral gesture (without expiratory air flow). Further investigation is needed to clarify this problem.

S. HORIGUCHI: What do you think about the cause of diplophonic voice from your experiment?

T. KANEKO: I cannot give a specific answer to your question concerning the problem of diplophonic voice. However, the differences in resonances between the two vocal folds, as well as possible differences in resonance between the upper and lower lips of the vocal folds should be taken into consideration.

S. HORIGUCHI: Do you think your method is useful for detecting the degree of invasion of cancer?

T. KANEKO: Invasiveness of the cancerous tissue into the deep layer can be estimated by this method, especially in the case of the endophytic type of cancer.

I. TITZE: Since normal modes of a bounded viscoelastic system are to be expected, can you give us some idea about the standing wave mode patterns, particularly for the higher modes?

T. KANEKO: By our present method it is possible to measure the frequencies of the higher normal modes, but mode patterns themselves cannot be observed.

B. WYKE: Since the time scales were not indicated in slides 5 and 6, please state what were the time intervals between the onset of phonation and the elicitation of vocal fold resonance responses. In your paper you said only "just before phonation" - so how long is "just before", especially since your figures show the phonatory activity preceding the resonance responses. This is important in analyzing the processes of "prephonatory tuning".

T. KANEKO: The resonance waveform of the vocal fold was recorded in the phonation neutral gesture. That is, during sustained phonation of /i/, expiratory air flow was stopped with maintenance of the same laryngeal gesture, and the resonance patterns were obtained. The time interval of the elicitation of vocal fold resonance response and onset of phonation was less than 10 msec.

C. LARSON: That was a very nice study and it helped clear up one thing for me. I did something like this several years ago on myself. When I did it during vocalization, it changed the pitch of my voice. At the time, I wasn't thinking of it in terms of helping the resonance properties of the vocal folds. Instead, I was thinking that the vibration was exciting receptors, reflexively causing contraction of the muscles, and I think that is still a viable possibility. It is something that in this work you should be cognizant of, that is, that the vibration itself could excite receptors and cause muscle contraction which can affect your results.

T. KANEKO: It might be possible to detect the contraction of the muscle tissue, and determine whether the muscle tonus is different during phonation versus immediately after phonation.

O. FUJIMURA: Is the resonance being measured at the upper edge of the vocal folds or at lower levels? Do you have any optical observation of the vibration mode elicited?

T. KANEKO: In this experiment the resonance characteristics of the vocal fold in phonation neutral gesture were measured. It will be possible, however, to observe simultaneously the resonance characteristics of the upper edge and of the lower level of the vocal fold using two focused transducers with small diameter. Optical observation of the vibration mode excited by our method, however, has not yet been performed. Thank you very much for your comment.

28 ▪ GENERATION AND TRANSFER OF HEAT IN THE VOCAL FOLDS

Donald S. Cooper and Ingo R. Titze

When we turn to the older physiological literature relating to speech production, the importance of the development of modern rapidly-responding equipment is impressive. However, it is possible that the emphasis on the observation of rapidly-varying quantities leads us to overlook quantities which vary slowly and thus are considered less interesting from the dynamic point of view of modern studies of speech production. In studies of phonation on the basis of excised larynges, it has been found important to excite them with a stream of air which is heated to about body temperature, and humidified (van den Berg and Tan, 1959; van den Berg, 1960; Baer, 1975, 1981). Since phonation can be produced without the presence of other physiological systems which are crucial in the living organisms such as blood flow and innervation, it may be asked, what is the reason that these two quantities, heat and moisture, play such an important role in the successful production of phonation in the excised larynx, and presumably in the larynx in vivo? Is it possible that the main importance of heating the air blown through the excised larynx is simply to increase the amount of water which it takes up, and thereby prevent desiccation of the tissues?

There is certainly no reason to doubt the importance of tissue moisture to phonation. To begin with, water is the largest single component of the tissues in question, at least for muscle: from the data of Bowden and Scheuer (1960) on the wet and dry weights of intrinsic laryngeal muscles it can be calculated that vocalis muscle is composed of 79% water by weight, which agrees exactly with the usual finding for skeletal muscle in adults (Diem and Lentner, 1970). Leaving physiological functions aside, clearly the tissue desiccation which is a perennial hazard in work with excised larynges would lead to change of the volume and shape of the tissues. However, we can be more specific: Lecluse (1977) found that when vocalis tissue was allowed to dry in the air, it became stiffer (longitudinal elasticity). Baer (1975) has given in his dissertation many valuable details about the phonatory behavior of desiccated excised larynges.

However, there are also cogent reasons to expect that tissue temperature of the vocal folds plays a definite underlying role in the biomechanics of phonation. In the living organism, a number of quantities are temperature-dependent, such as the velocity of propagation of the muscle action potential (Stalberg, 1966), twitch contraction time (Buchthal and Schmalbruch, 1970), and twitch and tetanic tensions of muscle (Reichel, 1960), but these cannot be in question in the excised larynx. In most experiments on excised larynges, the body of the larynx itself has not been heated, but only the airstream, so that as we have found in our own studies of variation of the temperature in different parts of the vocal fold of excised larynges, the medial section of the fold is closer to the temperature of the airstream, while the lateral section of the fold is closer to the temperature of the ambient air. This suggests that we should turn to the structure of the medial section of the fold, as this has been studied by Hirano (1981). In it, a layer of epithelium covers the superficial layer of the lamina propria, composed mainly of collagen, which in turn overlies the intermediate layer, composed mainly of elastic fibers. In the deep layer of the lamina propria are found densely arranged collagenous fibers. More lateral of course, is the body of the muscle, but let us consider an important finding of Baer's dissertation (1975), the fact that he was able to remove the body of the muscle from excised canine larynges, leaving only the cover and transition layers, and still produce phonation. Thus the temperature effects in question may pertain most importantly to the mechanical characteristics of the tissue of the superficial part of the vocal fold, which has been shown to be the section most mobile during phonation (Baer 1975, 1981; Saito 1981). In fact, definite temperature dependence has been demonstrated in the mechanical properties of the tissues which compose this structure. Specifically, the elastic moduli of collagen and muscle decrease with temperature, while the elastic modulus of elastin increases with temperature. The effect is fairly modest for collagen, but the elastic modulus of muscle drops by about 40% when it is heated from room temperature to about 37°C, while the elastic modulus of elastin increases about 60% with the same temperature change (Apter, 1972); Figure 28-1). Apter's work was based on viable excised tissue, and the effects in dead tissue are somewhat different. However, it is highly probable that an important effect of the use of warmed air in experiments with excised larynges lies in the conduction of heat into the superficial layers of the vocal fold, the mechanical properties of which are thus thermally tuned, so to speak. It follows that in the living organism, the maintenance of the temperature stability of the tissues of the folds is closely related to their mechanical characteristics, passive as well as active, which in turn partly determine the nature of the resulting phonation. Consequently, it is worth while to ask what are the sources of heat in the tissues of the vocal fold in vivo, what are the internal heat transfer processes which determine the temperature gradient within the vocal fold, and what are the conditions of the inspiratory and expiratory air which affect the movement of heat into or out of the fold?

INTERNAL HEAT SOURCES IN THE VOCAL FOLDS

The larynx is remarkable for the systematic variation both of the internal sources of heat supplied to it and of the external environment, i.e. the respiratory airstream to which the larynx is exposed. The primary physiological sources of heat supplied to the larynx are the circulation of the blood and muscle activity. The blood vessels in the edge of the vocal fold are arranged with the vessels approaching it at the anterior or posterior end of the fold, and running parallel to the edge of the fold, with direct anastomosis between arterioles and venules. Vessels supplying the mucosa are distinct from those supplying the vocalis muscle (Hirano, 1977; Mihashi *et al*, 1981). In such a structure, the blood reaching the outer part of the fold is gradually cooled as heat passes through the arterial walls to the tissue, and blood returning from the periphery is partially rewarmed from the tissue, in the phenomenon of countercurrent circulation. Thus less heat is lost to the periphery than would otherwise be the case (Ruch and Patton, 1973).

In general, blood flow to skeletal muscles is observed to increase with their activity; however, blood flow to the vocal folds has been shown to decrease during vibration (Hiroto, 1969, 1971). Perhaps, as Hirano has suggested (1981), vocalis, like the tympanic membrane, is specially constructed to minimize the possibility of a resulting hypoxia. It is also possible that a decrease in bloodflow during phonation could at least partly reflect longitudinal tension from cricothyroid activity, since restriction of blood flow by externally applied tension on a muscle has been widely observed (Hudlicka, 1973).

A second source of heat in the vocal fold tissue is muscle metabolism. Resting heat production is small (about 2 mcal/ gm°min in frog sartorius muscle at 20°C, Dowben, 1980), but it increases during muscle activity. The time course of temperature changes is not simple or simply related to the electrical activity of the muscle (Abbott and Howarth, 1973; Carlson and Wilkie, 1974), but electromyographic measures may be taken to indicate the activities which lead to the metabolic generation of heat in the vocalis muscle. Cyclic variation of the

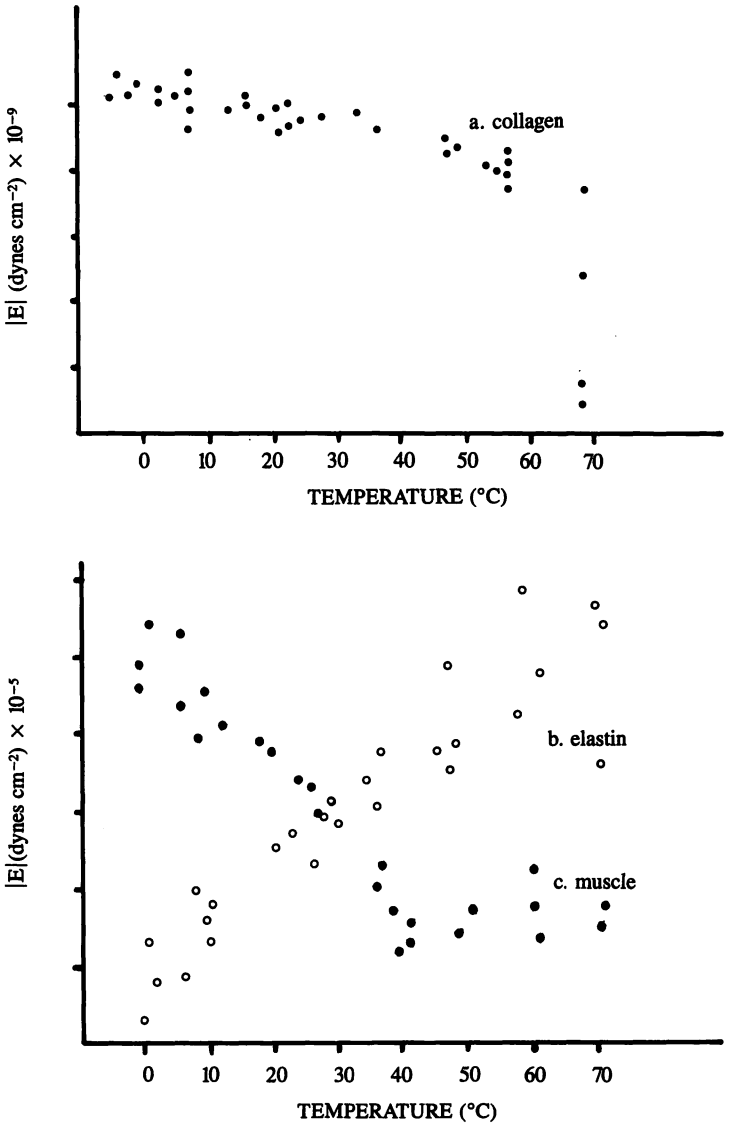


FIGURE 28-1

Effect of temperature on elastic modulus of collagen, elastin, and muscle. (From Apter, 1972, p.227. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.)

amplitude of EMG activity of vocalis occurs during respiration, its activity increasing at the onset of expiration and ending at or soon after the onset of inspiration (Wyke, 1974). The vocalis is usually active during phonation (Faaborg-Andersen 1965; Wustrow, 1970; Harris, 1981).

A third internal effect is thermoelastic heat, demonstrated by the change of temperature of a substance (organic or inorganic) dependent on its change of tension. For active frog and toad muscle, A.V. Hill found that when active muscle is stretched, it absorbs heat, and when it is released, it evolves heat. For resting frog muscle between resting length (l_0) and 1.25 l_0 it has been found that the muscle evolved heat when it was stretched, but cooled when it was released, while at greater lengths the effect was reversed (Hill, 1953; Woledge, 1961, 1963; Abbott and Howarth, 1973; Aidley, 1978). This phenomenon has not been investigated in any laryngeal tissue.

The possible cooling of the tissue clearly distinguishes thermoelastic heat from another effect proposed by Titze (1982), who has suggested that viscous loss in the vocal fold tissues may represent a significant amount of the available pulmonary power. If the energy dissipated is converted into heat, it may result in warming of the vocal fold. Titze has estimated that the average viscosity of vocal fold tissue is about 15-30 poise, lower in the superficial layers, and suggested that during a vertical vibration of one mm amplitude at 159 Hz, heating of the fold from viscous dissipation might amount to about .25°C (Cooper and Titze, to appear). Measurements by Cooper and Titze on vibrated excised vocal fold tissue appear to support this hypothesis, although the interaction of the effects of airstream temperature, tissue temperature, and ambient room temperature have so far made it difficult to measure this effect during phonation of excised larynges.

THE RESPIRATORY AIRSTREAM

We leave any systematic consideration of airstream thermal effects to specialists in fluid mechanics, and only venture to describe the systematic variation of the air temperatures to which the vocal folds are exposed as a matter of the stage in the respiratory cycle. Recently Cole (1982) has provided an illuminating picture of the modification of inspired air, on which the following remarks are based. The air to which the vocal folds are exposed is always at or near saturation, whether inspiratory or expiratory. However, its temperature varies appreciably. From room temperature, inspiratory air is warmed to about 33° by the time it reaches the larynx. Tracheal expiratory air appears to be near or slightly below body temperature. Because of the cooling by inspiratory air, the mucosa remain below body temperature, thus establishing a temperature gradient between the external air and the deep body temperature of the larynx and surrounding structures. If the vocal folds follow this general picture, then they are cooled by inspiratory air, but warmed by expiratory air and by the latent heat of evaporation from water which condenses in contact with the cooler mucosa.

HEAT TRANSFER IN THE VOCAL FOLD

In the above discussion we have summarized the sources of heat within the vocal folds, and in the respiratory airstream. Now it is time to consider the processes by which heat is distributed through the vocal folds, so that a temperature gradient is established between the lateral and medial sections of the vocal fold. In so doing we shall make use of a model constructed by Keller and Seiler (1971), which makes it possible to organize the elements we have described into a coherent framework. It is not claimed that this model is complete or definitive, but only that it is a useful point from which to start in the sense implied long ago by the French

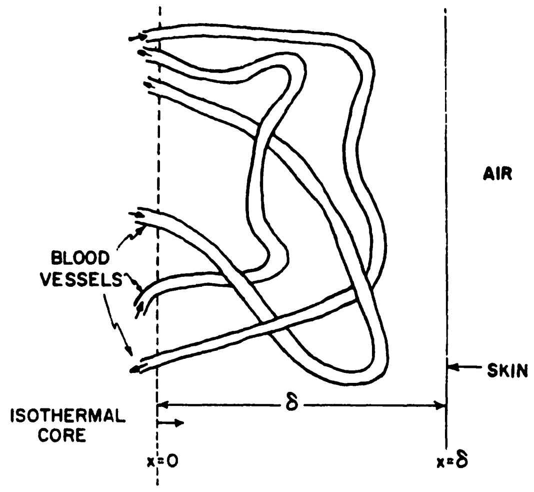


FIGURE 28-2
Schematic representation of vascularization and area of temperature variation in subcutaneous tissue region. (From Keller and Seiler, 1971. Reprinted by permission of the Amer. Physiol. Soc.)

phonetician Maurice Grammont, when he stated that "an incorrect theory at least leads to a correction, while the lack of a theory leads to nothing".

We shall describe this process in terms of a model of a small local region of the vocal fold, the thickness of which is small, and where the temperature gradient runs normal to the surface, so that the treatment can be one-dimensional. The model should combine the effects of heat conduction, convection, the generation of heat within the tissues by mechanical and metabolic processes, and countercurrent heat exchange. Such a region can be illustrated by Figure 28-2. The deeper region of the tissue will be referred to as the isothermal area, and the region where there exists a temperature gradient extending outwards to the mucosa will be designated the peripheral region, extending from $x = 0$ to $x = \delta$. In describing heat transfer it will be assumed that the prevailing conditions are defined not in terms of a dependence on time, but by the choice of the values of constants in the equations, so that the independent variable is distance, $\underline{x}$.

In Figure 28-2 we see an overall figure of the subcutaneous region, indicating the presence of vascularization and a temperature gradient between the isothermal core on the left and the respiratory air on the right. Designating the temperature of the core by T_c and the temperature of the mucosa by T_m, we can write the heat flux from the core to the mucosa as

$$H = \frac{k_{eff}}{\delta}(T_c - T_m)$$

with units of cal/(cm^2•s). k_{eff} is an effective thermal conductivity, with units of cal/(cm•s•°C), the specification of which is the objective of the analysis.

Now let us consider the heat transfer phenomena occurring within a unit of tissue. We are concerned with the steady state energy balance on a differential volume of tissue of unit cross-sectional area, located between the planes at x and $x + \delta x$, as illustrated in Figure 28-3. Three differential equations can be formulated, the first of which describes an energy balance on the tissue, the second of which describes an energy balance on the arteries, and the third of which describes an energy balance on the veins.

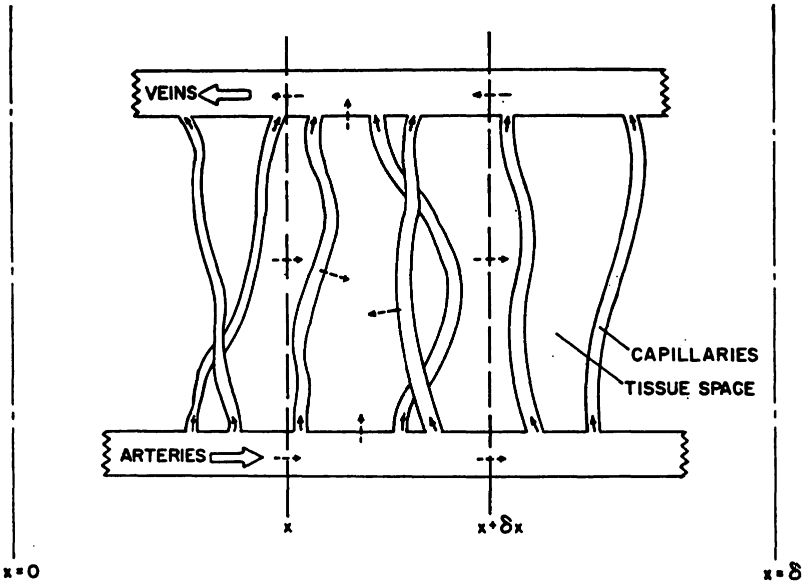


FIGURE 28-3
Element in subcutaneous tissue region. Direction of blood circulation is shown by solid lines. Paths of heat transfer included in analysis are shown by dashed arrows. (From Keller and Seiler, 1971. Reprinted by permission of the Amer. Physiol. Soc.)

These equations describe conductive heat flow across the x plane into the tissue volume and across the $x + \delta x$ plane out of the tissue volume. They also specify heat conduction from the arteries into the tissue space, and from the tissue into the veins. Heat transfer from the capillaries into the tissue space is described. Heat generation by tissue mechanical effects and tissue metabolism can also be included. The solution of these equations provides the basis for the calculation of k_{eff}.

It is not possible to go into the details of the formulation or solution of this set of thermal balances, but we will indicate some points which are significant for the process of heat transfer in the vocal folds. A first point is that the width of the peripheral region, described by the distance δ, is subject to variation. Its thickness depends mainly on the rate of blood perfusion of the tissue and tissue thermal conductivity, secondarily on the rates of heat production in the tissue and arterial precooling. The relative importance of the rate of heat production in the tissue could be increased by the mechanical sources of heat production inherent in an oscillating structure such as the vocal folds.

Since the objective of the analysis is the k_{eff}, it is significant that this has a limited range. Its minimum value is equal to the thermal conductivity of the tissue itself, while its maximum value is about four times the tissue thermal conductivity. This limits the effect of perfusion rate on peripheral heat flux, although the rate of perfusion can vary much more widely than the k_{eff}.

The specific value of k_{eff} is a function of three parameters which arise in the model. The specific form of this function depends on the assumptions made in the solution of the

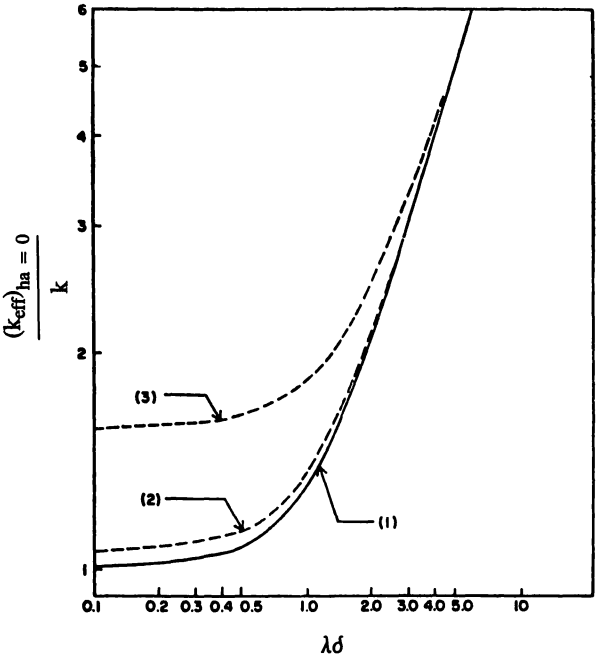


FIGURE 28-4

Nondimensional effective thermal conductivity as a function of the relative importance of capillary perfusion to thermal conductivity in heat transfer. Top line (3) shows the relationship in active muscle, the middle one (2) shows it in resting muscle, and the lowest one (1) shows it if metabolic heat is neglected. It is assumed as a limiting case that arterial-venous heat exchange (ha) is negligible. (From Keller and Seiler, 1971. Reprinted by permission of the Amer. Physiol. Soc.)

equations. The first parameter, $\lambda\delta$, is a ratio of capillary perfusion-induced heat transfer to conductive heat transfer. Thus its value would decrease during vibration, when Hiroto has observed that circulation to the vocal folds decreases (Mihashi *et al*, 1981). Consequently, during phonation, conductive heat transfer assumes greater importance in the vocal folds than during inspiration or expiration. It would also decrease under any other circumstances which specifically decreased the circulation to capillaries.

The second parameter, μ, is a ratio of the importance of arterial-venous countercurrent heat exchange to capillary perfusion in peripheral heat transfer. The last parameter, Φ, is a ratio of the heat generated in the peripheral region to the maximum quantity of heat which could be transferred by capillary perfusion. Clearly Φ would be smallest during inspiration, would increase as vocalis became active at the beginning of expiration, and would reach its highest value during actual phonation.

Consideration of a special case illustrates the sort of relationship which emerges from the analysis. For the case where there is no heat exchange between arterial and venous blood, Figure 28-4 shows the dependence of the dimensionless k_{eff}, divided by the thermal conductivity of the tissue, on the parameter $\lambda\delta$, the ratio of the importance of capillary perfusion to conductive heat transfer. The left-hand part of the curve corresponds to a condition of vasoconstriction, and the right hand part of the curve corresponds to a condition of vasodilatation. Clearly, under a condition of vasoconstriction, the difference between different levels of heat generation within the vocal folds would have a marked effect on the effective thermal coefficient. However, let us say that there occurred an increase in the ambient temperature of the folds. In this case there would occur a dilatation in the peripheral blood vessels, increasing the parameter $\lambda\delta$, and in this higher range of values the effect of the level of muscle activity on the effective thermal coefficient greatly decreases. This prediction of the model is substantiated by empirical findings in other body regions (Nielsen, 1969).

Some details of the above model are in need of modification. For instance, as the authors note, tissue thermal conductivity can be expected to vary as we move from one layer to another between the core and the periphery. The formal procedures for such analysis are available (e.g. Lih, 1975). Furthermore, the analysis highlights the importance of the studies of bloodflow in the vocal folds which have been carried out by Mihashi and his colleagues (1981), and adds a new significance to these studies. It is also clear that in order to implement such an analytic model, further studies must be carried out on such subjects as the capillary bloodflow, thermal conductivity of different tissue layers, and the temperature gradient of the vocal folds. However, we feel that it has been shown that even the present more limited model contains much that is illuminating in its systematization of the phenomena of conductive and convective heat transfer, countercurrent heat exchange, and the generation of heat in the tissues of the vocal folds.

How important can the effects of temperature on skeletal muscle function be? Clearly Apter's demonstration of the effects of temperature has considerable implications for work with excised tissue. If we extrapolate Apter's findings for viable excised tissue to the larynx in vivo, over the range of temperatures from 33°C to 37°C which is relevant in the vocal fold, the Young's modulus might vary about 20% over the range, and that of elastin about 12% (Apter, 1972). Similar effects can be expected for tissue viscosity, a property which is temperature-dependent (Ferry, 1980). Turning to active tissue properties, Stalberg (1966) found little effect of 4°C change in muscle temperature on the conduction velocity of the muscle action potential. However, Buchthal and Schmalbruch (1970) found that above 32°C there was a decrease in contraction time of 5% per degree centigrade for fast muscle, and of 4% per degree centigrade for slow muscle. Findings may be expected to vary for different muscles, and we cannot predict the effects to be found for laryngeal muscles. For this reason the question may merit further investigation.

In summary the tissues of the vocal fold vary in their mechanical properties as a function of temperature, and the different layers of the medial edge of the fold, which are crucial in phonation, are affected differently by changes in tissue temperature. Furthermore, the rate of generation of heat in the vocal folds on mechanical and metabolic bases varies systematically under different conditions of respiration and phonation. The transfer of heat from the tissues of the vocal fold to the periphery varies systematically with the above conditions, and with those of circulation. Although the temperature gradient of the vocal folds, and the phenomena of heat transfer in the peripheral region of the folds, are slowly-varying quantities, they constitute part of a substrate which contributes to the determination of the character of the more rapidly-varying quantities of phonatory waveform.

REFERENCES

Abbott, B.C. and Howarth, J.V. (1973). "Heat studies in excitable tissues," *Physiol. Rev.* 53/1, 120-158.

Aidley, D.J. (1978). *The Physiology of Excitable Cells*, 2nd Ed., Cambridge Univ. Press, Cambridge.

Apter, J.T. (1972). "Influence of composition on thermal properties of tissue," in *Biomechanics, Its Foundations and Objectives*, Y.C. Fung, N. Perrone, and M. Anliker (Eds.), Prentice Hall, Inc., Englewood Cliffs, NJ, 217-235.

Baer, T. (1975). "Investigation of phonation using excised larynges," doctoral diss., M.I.T.

Baer. T. (1981). "Observations of vocal fold vibrations: measurement of excised larynges," in *Vocal Fold Physiol.*, K.N. Stevens and M. Hirano (Eds.), Univ. of Tokyo Press, Tokyo, 119-133.

van den Berg, Jw. and Tan, T.S. (1959). "Results of experiments with human larynges," *Pract. Otol. Rhinol. Laryngol.* 21, 425-450.

van den Berg, Jw. (1960). "Vocal ligaments versus registers," *Curr. Prob. Phoniat. Logoped.* 1, 19-34.

Bowden, R.E.M. and Scheuer, J.L. (1960). "Weights of abductor and adductor muscles in the human larynx," *J. Laryngol.* 74, 971-980.

Buchthal, F. and Schmalbruch, H. (1970). "Contraction times and fibre types in intact human muscle," *Acta Physiol. Scand.* 79, 435-452.

Carlson, F.D. and Wilkie, D.R. (1974). *Muscle Physiology.* Prentice Hall, Englewood Cliffs, New Jersey.

Cole, P. (1982). "Modification of inspired air," in *The Nose: Upper Airway Physiology and the Atmospheric Environment.* D.F. Proctor and B. Andersen (Eds.), Elsevier Biomedical, Amsterdam, 351-375.

Cooper, D. S. and Titze, I.R. "Generation and dissipation of heat in vocal fold tissue," *J. Speach Hear. Res.* (in press.)

Diem, K. and Lentner (1970). *Scientific Tables.* Ciba-Geigy, Basel.

Dowben, R.M. (1980). "Contractility, with special reference to skeletal muscle," in *Medical Physiol.*, 14th Ed. Vol. I, V.B. Mountcastle and C.V. Mosby, (Eds.), St. Louis, MO, 82-119.

Faaborg-Andersen, K. (1965). *Electromyography of Laryngeal Muscles in Humans. Technics and Results.* Vol. 3 of Aktuelle Probleme der Phoniatrie und Logopadie, F. Trojan (Ed.), S. Karger, Basel.

Ferry, J.D. (1980). *Viscoelastic Properties of Polymers.* 3rd ed., John Wiley & Sons, Inc. New York, 43-44.

Harris, K.S. (1981). "Electromyography as a technique for laryngeal investigation," in *Proc. of the Conf. on the Assessment of Vocal Pathol.*, C.L. Ludlow and M.O. Hart (Eds.), ASHA Rep. 11, Rockville, MD.

Hill, A.V. (1953). "The 'instantaneous' elasticity of active muscle," *Proc. R. Soc.* B 141, 161-178.

Hirano, M. (1977). "Structure and vibratory behavior of the vocal folds," in *Dynamic Aspects of Speech Production*, M. Sawashima and F.S. Cooper (Eds.), Univ. of Tokyo Press, Tokyo, 13-30.

Hirano, M., Kurita, S. and Nakashima, T. (1981). "The structure of the vocal folds," in *Vocal Fold Physiol.*, K.N. Stevens and M. Hirano (Eds.), Univ. of Tokyo Press, Tokyo, 33-43.

Hiroto, I., Toyozumi, Y., Tomita, H., Miyagi, T., Kuroki, K., Koike, Y. and Matsushita, H. (1969). "An experimental study on the hemodynamics of the vocal fold during vibration," *Jap. J. Otol.* 79, 435-438, Cited after Mihashi *et al* (1981).

Hiroto, I. (1971). "The hemodynamics of the vocal fold and tympanic membrane during vibration," *Otologia* (Fukuoka) 17, 1-5. cited after Mihashi *et al* (1981).

Hudlicka, O. (1973). *Muscle Blood Flow.* Swets & Zeitlinger B.V., Amsterdam.

Keller, K.H. and Seiler, L., Jr. (1971). "An analysis of peripheral heat transfer in man," *J. of Appl. Physiol.* 30/5, 779-786.

LeCluse, F.I.E. (1977). *Electroglottografie*. Drukkerijelinkwijk B.V., Utrecht.

Lih, M.M.-S. (1975). *Transport Phenomena in Medicine and Biology.* Wiley-Interscience, New York.

Mihashi, S., Okada, M., Kurita, S., Nagata, K., Oda, M., Hirano, M. and Nakashima, T. (1981). "Vascular network of the vocal fold," in *Vocal Fold Physiol.*, K.N. Stevens and M. Hirano (Eds.), Univ. of Tokyo Press, Tokyo, 45-59.

Nielsen, B. (1969). "Thermoregulation in rest and exercise," *Acta Physiol. Scand. Suppl. 323.*

Reichel, H. (1960). Muskelphysiologie. Springer-Verlag, Berlin-Göttingen-Heidelberg.

Ruch, T.C. and Patton, H.D. (1973-74). *Physiology and Biophysics.* Vol. 2 and 3, W.B. Saunders, Co., Philadelphia.

Saito, S., Fukuda, H., Isogai, Y. and Ono, H. (1981). "X-ray stroboscopy," in *Vocal Fold Physiol.*, K.N. Stevens and M. Hirano (Eds.), Univ. of Tokyo Press, Tokyo, 95-106.

Stalberg, E. (1966). "Propagation velocity in human muscle fibers in situ," *Acta Physiol. Scand.* 70 Suppl. 287.

Titze, I.R. (1982). "Heat generation in the vocal folds and its possible effect on vocal endurance," in *10th Annual Symposium, Care of the Professional Voice*, The Voice Foundation, New York.

Woledge, R.C. (1961). "The thermoelastic effect of change of tension in active muscle," *J. Physiol.* (London) 155, 187-208.

Woledge, R.C. (1963). "Heat production and energy liberation in the early part of a muscular contraction," *J. Physiol.* (London) 166, 211-224.

Wustrow, F. (1970). "Elektromyographie im Bereich der Muskulatur des Kehlkopfes," in *Die Stimme und ihre Störungen, Vol. I of Handbuch der Stimm- und Sprachheilkunde*, by R. Luchsinger and G.E. Arnold, Springer-Verlag, Vienna.

Wyke, B.D. and Kirchner, J.A. (1976). "Neurology of the larynx," in *Scientific Foundations of Otolaryngology*, R. Hinchcliffe, and D. Harrison (Eds.) Heinemann, London, 546-574.

29 ▪ COLLAPSIBLE TUBE MODEL OF THE LARYNX

William A. Conrad

INTRODUCTION

The purpose of this paper is to show that an elementary one-mass model of the larynx can manifest a negative differential resistance. A supraglottal resistance is essential for the purpose. Introducing negative differential resistance leads to a statement of conditions for which the larynx will phonate. The condition, though well known in many branches of physics and engineering, is nonetheless unfamiliar to many people. I will therefore try to keep in view its definition in concrete terms while explaining its abstract consequences. Resistance is defined as pressure-drop divided by flow rate. Differential resistance is defined as change in pressure-drop divided by change in flow rate. When pressure-drop decreases as flow rate increases, the differential resistance is negative. Despite a superficial resemblance between the concepts of resistance and differential resistance, they are fundamentally different when the differential resistance is negative. A negative differential resistance can make a system unstable. The potential for instability is shown by steady flow conditions.

In this talk I shall use my experimental steady flow results on collapsible tubes as motivation for calculating similar relations for the larynx. Combining these calculations with data based on oscillations in collapsible tubes, I shall then make a plausible argument that phonation is a negative resistance oscillation. I will then make a conjecture about the false vocal folds and conclude with possible applications to clinical practice.

COMPARISON OF LARYNX, COLLAPSIBLE TUBE AND ONE-MASS MODEL

Figure 29-1 shows a cross-section of the larynx from Hollinshead (1982), a collapsible tube, and the fluid-mechanical one-mass model of Flanagan and Landgraf (1968). The feature they share is flow through a gap or slit whose cross-sectional area depends on the pressure in the gap. Because the mass of the wall plays no role for steady flow, it follows that all three structures are virtually identical. In a similar way, a dog larynx and a human larynx are nearly identical. In this type of system the fluid mechanics and wall mechanics are tightly coupled. Flow through the resistance downstream of the gap causes a back pressure which exerts a force on the wall that can displace it and thus change the area of the gap and its resistance to flow. For some conditions an increase in flow rate is associated with a decrease in pressure-drop across the gap.

STEADY FLOW IN COLLAPSIBLE TUBES

Let us now consider experiments with collapsible tubes. Figure 29-2 shows the apparatus used. A thin-walled rubber tube is mounted on rigid tubes surrounded by an air-tight box which can be pressurized. The external pressure P_e compresses the tube. Micrometer clamps upstream and downstream to the tube set boundary conditions and control flow. Flow is maintained by a constant level tank pressure P_T. The system is vented to the atmosphere.

There are obvious parallels to the larynx. The tank corresponds to the lungs, R_1 to subglottal resistance, R_2 to supraglottal resistance, and external pressure P_e roughly corresponds to muscle tension.

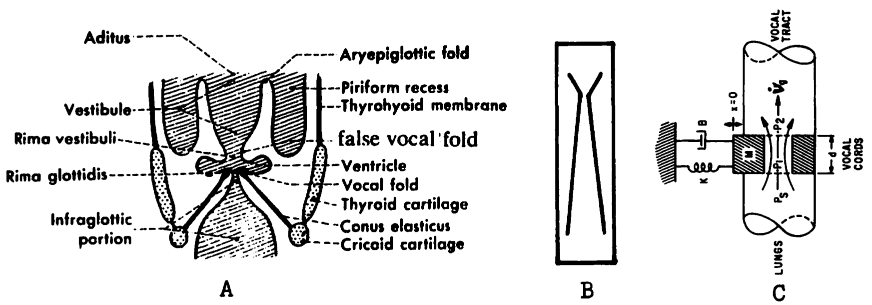


FIGURE 29-1

A) Frontal section through the larynx (Hollinshead, 1982). B) Collapsible tube with constant external pressure and constant downstream resistance. C) One mass model of Flanagan and Landgraf (1968). All these structures have flow through a gap whose area is pressure dependent.

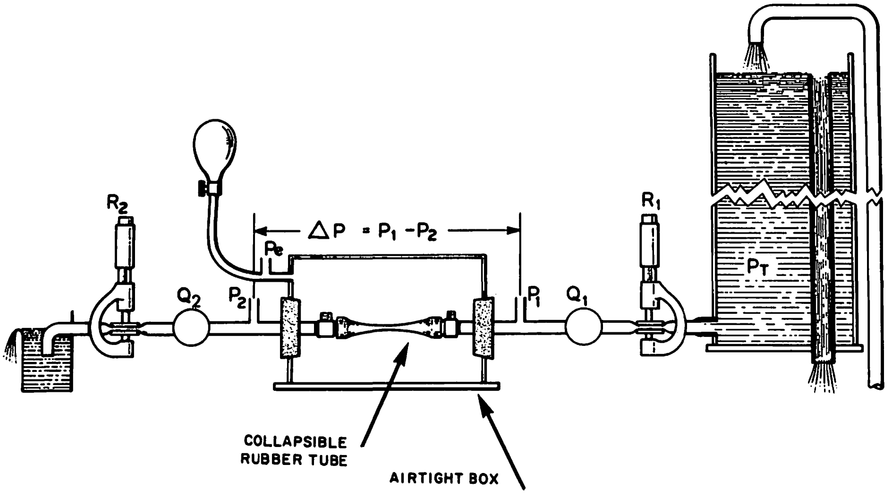


FIGURE 29-2

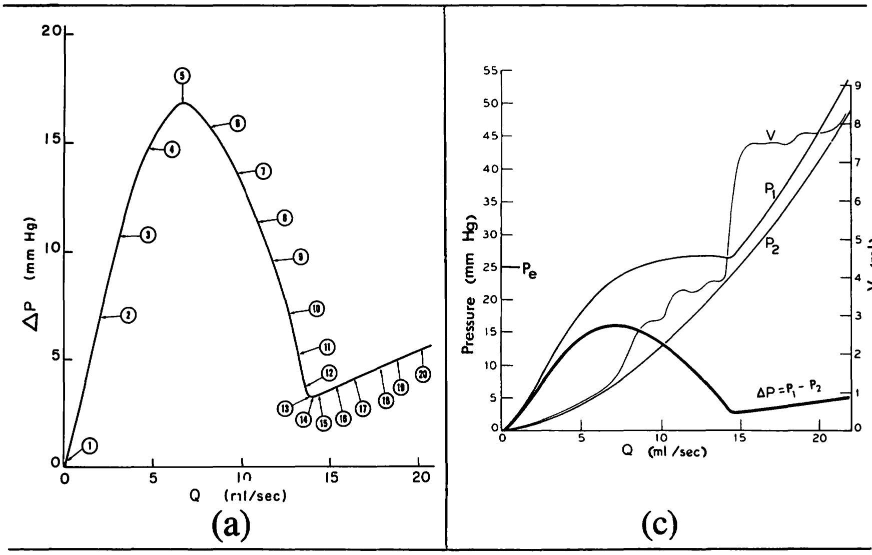
Model used to study the flow of water through a horizontal collapsible tube surrounded by a rigid air chamber. Open ports are pressure taps, circles are flow meters, constrictions are micrometer controlled (Conrad, 1969).

One mode of experimentation was to fix P_e and R_2 and then vary the tank pressure. The upstream pressure P_1 and downstream pressure P_2 (sub and supraglottal pressures) were not fixed but left free to vary. This is what acutally happens in the larynx in vivo. We measured the pressure-drop ΔP as function of flow rate Q.

Figure 29-3 shows pictures of the tube and the pressure-flow relations for steady flow. Note that there are three regions in the pressure-flow relations of the tube: open, partially collapsed, and completely collapsed, that is, collapsed over its entire length. In the open and collapsed regions pressure-drop is proportional to flow rate, that is, the tube is an ordinary resistor, with small resistance when open and a large resistance when collapsed. In the collapsed region flow is through small rigid side channels, as panel 00 shows. In the transition region, tube shape changes rapidly, and so also does tube resistance. Note that pressure-drop decreases as flow rate increases, that is, the differential resistance is negative.

RELATION BETWEEN COLLAPSIBLE TUBE AND LARYNX

Now let's move on to look at the relationship between collapsible tubes and the larynx. Suppose the tube were frozen in the shapes shown in Figure 29-3. The pressure-drop as a function of flow rate would be essentially like that shown in Figure 29-4. These curves represent the pressure-drop through a rigid orifice of decreasing cross-sectional area. The dashed curve denotes a collapsible tube. It is a trajectory through the curves for the frozen shapes. van den Berg (1957) studied flow through rigid models of the larynx which differed in cross-sectional area. The curves he obtained were of course similar to the solid curves in this figure. If he had used a flexible larynx and a supraglottal resistance I believe he would have found something like the dashed curve. This belief is based on the calculations in the next section. There I will derive the pressure-flow relations for the larynx using the one- mass model of Flanagan and Landgraf (1968) supplemented by a supraglottal resistance. Van den Berg's experimental results are derivable from the model.

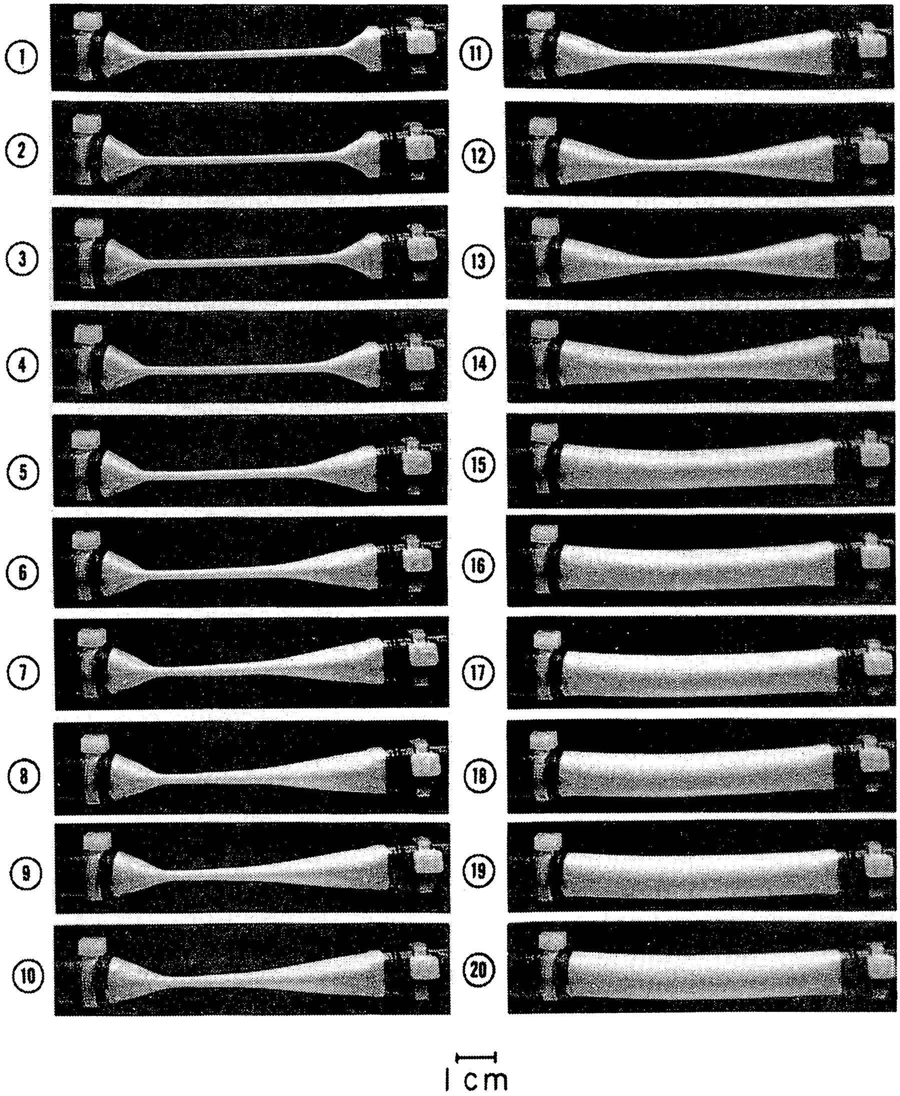


(a)

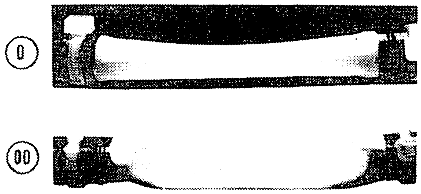
(c)

FIGURE 29-3

(a) Typical pressure drop versus flow or P-Q characteristic for a collapsible tube at constant external pressure and downstream constriction. (b) Side views of the collapsible tube at the indicated static operating points. The linear pressure-flow regions are associated with the collapsed or open tube, while in the nonlinear region, the tube is partially collapsed. (c) Upstream and downstream gauge pressures P_1 and P_2 and their difference ΔP as well as the volume V of the tube as functions of Q. P=25 mmHg, k = 0.08 mmHg/(ml/s)^2. (d) Zero flow conditions. Panel 0 is a side view for zero transmural pressure; panel 00 is a top view for P_{tr} = 30 mmHg; note patent side tunnels (Conrad, 1969).



(b)



(d)

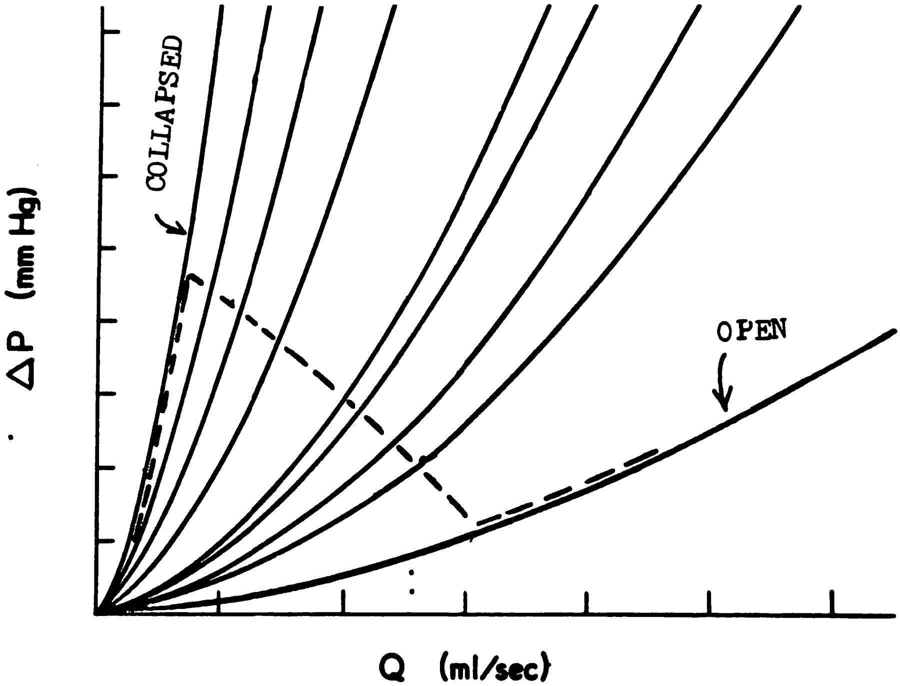


FIGURE 29-4
Solid lines represent *hypothetical* pressure-flow relations ΔP-Q curves, as if the shapes in Figure 29-3 were "frozen". Dashed line respresents ΔP-Q for a collapsible tube. It is a trajectory through the "frozen" shapes. The solid curves are similar to those obtained by van den Berg *et al* (1957) for a rigid model of the larynx.

DERIVATION OF PRESSURE-FLOW CURVES

Figure 29-5a shows a schematic diagram of the airway from the pulmonary pressure source to a supraglottal resistance at the entrance to the vocal tract. Figure 29-5b shows a steady flow resistive circuit representing the system, where U_g is the volume velocity, P_L is the driving pressure, and R_L is the subglottal resistance. The resistances for the larynx as shown are:

$$R_c = r_c \mid U_g \mid , \quad r_c = .68\rho A_g^{-2}$$

$$R_v = r_v, \quad r_v = 12\mu l_g^2 dA_g^{-3}$$

$$R_e = -r_e \mid U_g \mid , \quad r_e = \rho(1-A_g/A_v)/(A_g A_v)$$

$$R_f = r_f \mid U_g \mid , \quad r_f = .5\rho(.8A_f)^{-2}$$

I have used van den Berg's results for the kinetic flow-dependent resistance of the convergence to the glottis R_c and his relation for the viscous resistance R_v of the glottis itself. For the

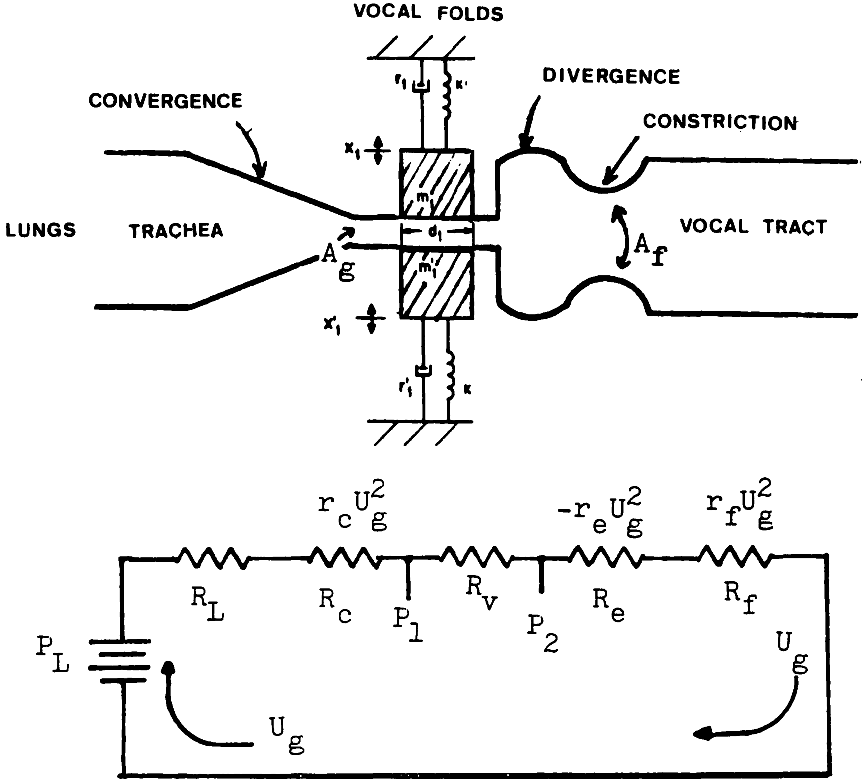


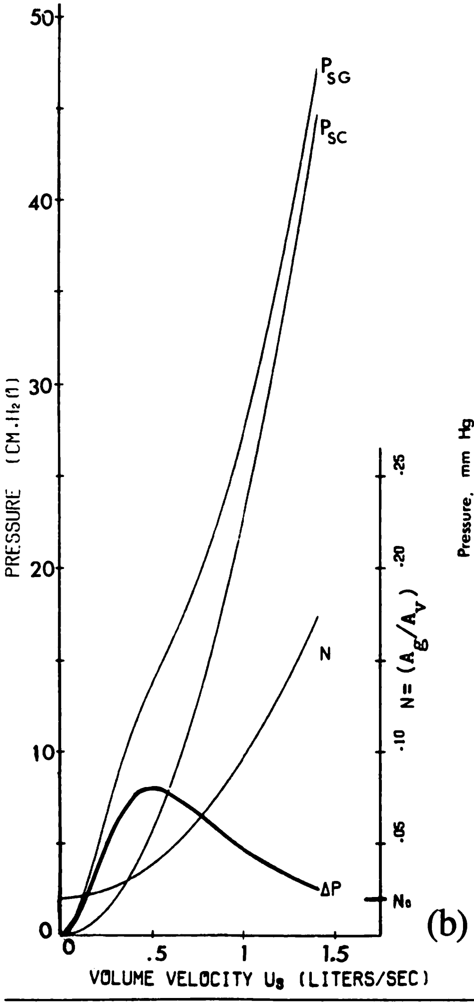
FIGURE 29-5
Schematic diagram of the larynx and equivalent resistive circuit for steady flow.

resistance of the divergence R_e, I used Ishizaka and Flanagan's (1972) value based on conservation of momentum. Note that the resistance R_e of the expansion is negative due to pressure recovery. It is not the same thing as a negative differential resistance. For the supraglottal resistance I used the same orifice formula as in Flanagan's (1958) paper on the glottis. In these formulas, A_g is the area of the glottis, A_v is the area of vocal tract, A_f is the area of the supraglottal constriction, l_g is the length of the glottal slit, d is the width of the folds, μ is the viscosity, and ρ is the density of the air.

Following Flanagan and Landgraf (1968), I assumed that the wall of the glottis was held in place by a linear spring of constant k. They also assumed that the force per unit area P_x acting on the wall was equalt to the mean glottal pressure (see footnote next page)

$$P_x = \tfrac{1}{2}(P_1+P_2) = (k/l_g d)x = Kx$$

where P_1 is the pressure at the inlet and P_2 that at the outlet of the glottis, x is the displacement of the wall, K is the normalized spring constant $k/l_g d$. It took the pressure of preparing this paper to make me realize that if the glottal area is assumed known, then all pressures and flows can be calculated. This is, of course, the method used by Flanagan and Landgraf



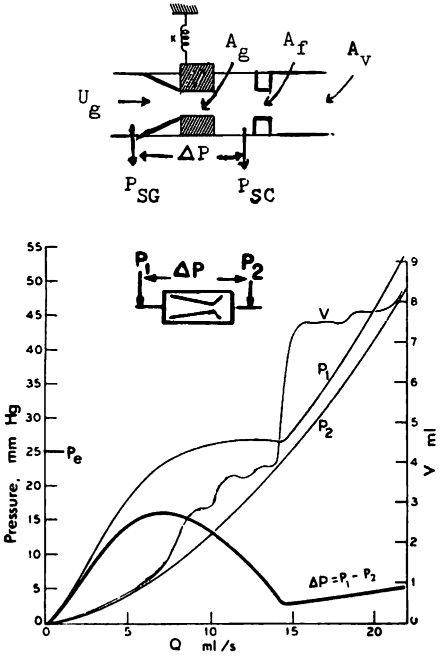


FIGURE 29-6B
Pressure-flow curves for the larynx plus a supraglottal resistance compared to similar curves for a collapsible tube. P_{SG} subglottal pressure, P_{SC} supraglottal pressure and $\Delta P = P_{SG} - P_{SC}$, $N = A_g/A_v$ or glottal area normalized to vocal tract area.

TYPICAL LARYNX PARAMETERS		
l_g	length	1.4 cm
d	depth	.3 cm
k	spring constant	8×10^4 g/sec^2
A_{go}	resting area	.08 cm^2
A_v	vocal tract area	4 cm^2
A_f	supraglottal constriction area	.2 cm^2

NORMALIZED PARAMETERS	
$N_o = A_{go}/A_v$	.02
A_f/A_v	.05
$K = k/l_g d$	2×10^5 g/(sec^{2}cm^2)

(a)

FIGURE 29-6A
Table of larynx parameters used to calculate pressure-flow curves.

and many others in making similar calculations. The volume velocity is easily calculated from the quadratic formula,

$$U_g = \frac{-\tfrac{1}{2}R_v \pm [(\tfrac{1}{2}R_v)^2 + 4(r_f - r_e)P_x]^{1/2}}{2(r_f - r_e)}$$

Note that for steady flow, the volume velocity must be zero when displacement x (and therefore P_x) are zero. The sign in front of the square root must be chosen as positive. The translaryngeal pressure ΔP is simply

$$\Delta P = (r_c - r_e)U_g^2 + R_v U_g$$

COMPUTATIONAL RESULTS

I have computed pressure-flow relations for the larynx assuming the values of the parameters shown in the table (Figure 29-6a). These are all typical values except for the area of the supraglottal orifice which is small. Figure 29-6b shows the results. Clearly visible is a region in which pressure-drop decreases as volume velocity increases; that is, the larynx has a negative differential resistance. Collapsible tube results are shown by way of comparison. The subglottal pressure P_{SG} and supraglottal pressure P_{SC} are also shown where P_{SC} is the pressure at the entrance to the supraglottal constriction. It is obvious that, qualitatively, the curves for the larynx and tube are similar even though they differ in details. As volume velocity increases, the pressure in the larynx and in the tube increases as shown by the inlet and outlet pressures P_{SG}, P_{SC}, P_1 and P_2. This increase in pressure is due to the increasing flow through the downstream or supraglottal resistance. The conduit opens with increasing pressure because it is flexible, not rigid. This can be seen from the volume V of the tube and cross-sectional area of the larynx represented by N, where N is the glottal area divided by the vocal tract area. Note that subglottal pressure P_{SG} increases more slowly than supraglottal pressure P_{SC}. This condition is essential in order to have a decreasing translaryngeal pressure ΔP. Tube and larynx are identical in this respect. Decreasing ΔP for increasing flow rate U_g - a negative differential resistance-depends on the flexibility of the walls and the presence of a downstream resistance.

The shape of the laryngeal pressure-flow curves depends on three factors: spring constant, resting opening and downstream orifice, all other things being equal. The wall force needed to achieve a given opening is a result of the pressure induced by the flow through the supraglottal resistance. Therefore, the greater the downstream resistance, that is, the

---

Footnote:

[1] Note added in proof. P_x can be easily derived in terms of the flow as follows:

$$P_1 = R_v U_g - r_e U_g^2 + r_f U_g^2$$

$$P_2 = -r_e U_g^2 + r_f U_g^2$$

$$P_x = \tfrac{1}{2}(P_1 + P_2)$$

$$\therefore P_x = \tfrac{1}{2}R_v U_g + (r_f - r_e)U_g^2$$

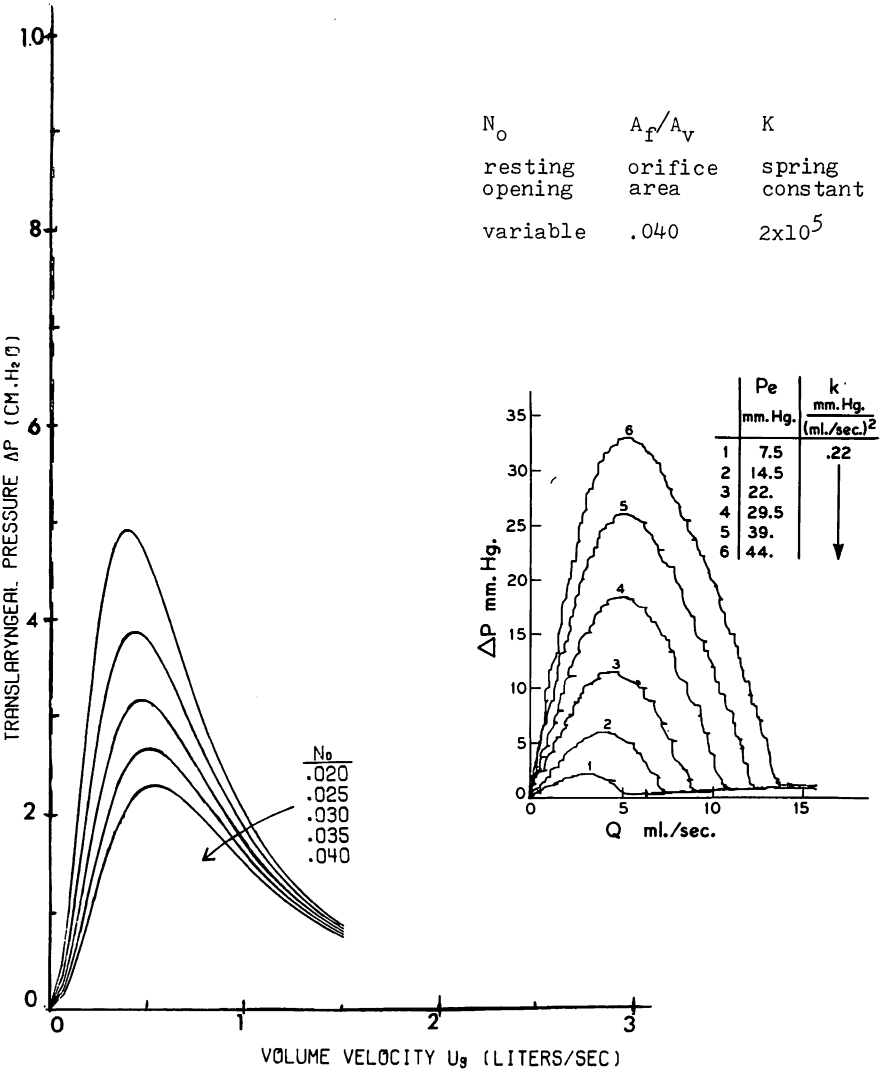


FIGURE 29-7

Family of translaryngeal pressure-volume velocity curves for variable resting opening, constant supraglottal orifice and constant spring constant.

smaller the area A_f of the orifice, the less the flow needed for a given pressure on the wall. This coupling of the wall to the flow will cause the pressure and volume velocity at the peak of the curve to change, for example, as spring constant, resting area and orifice area change.

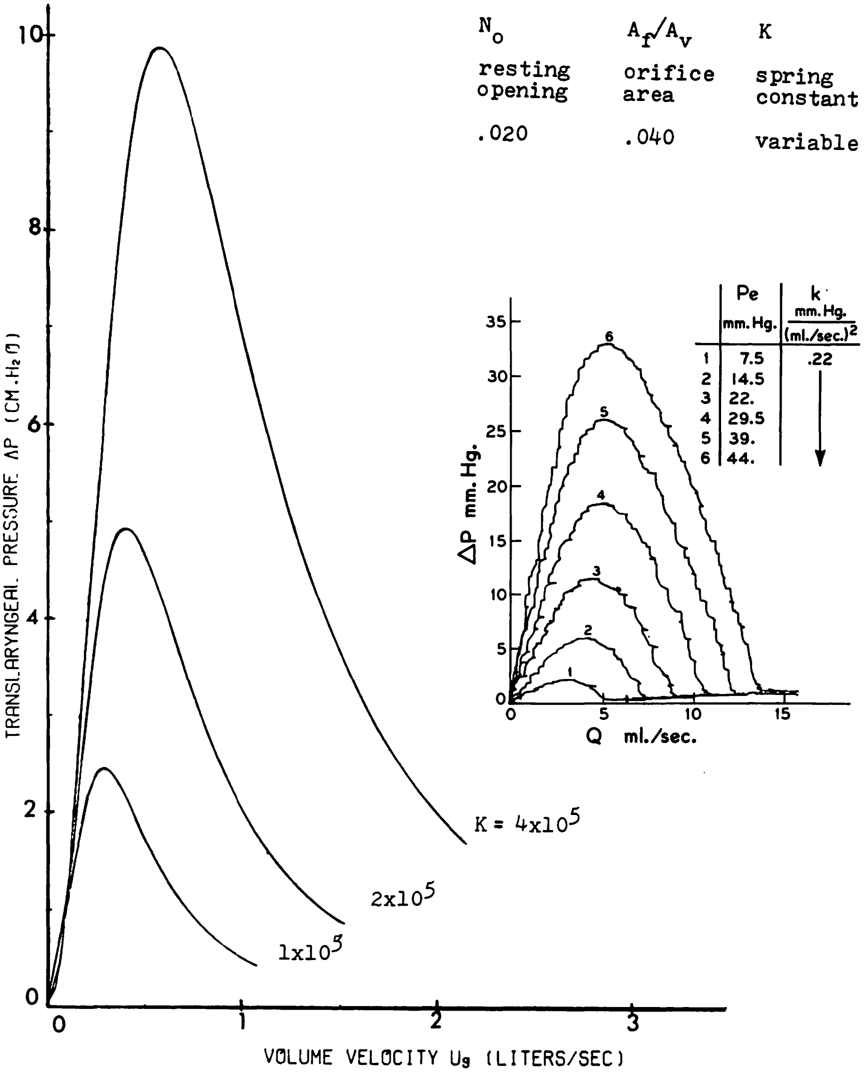
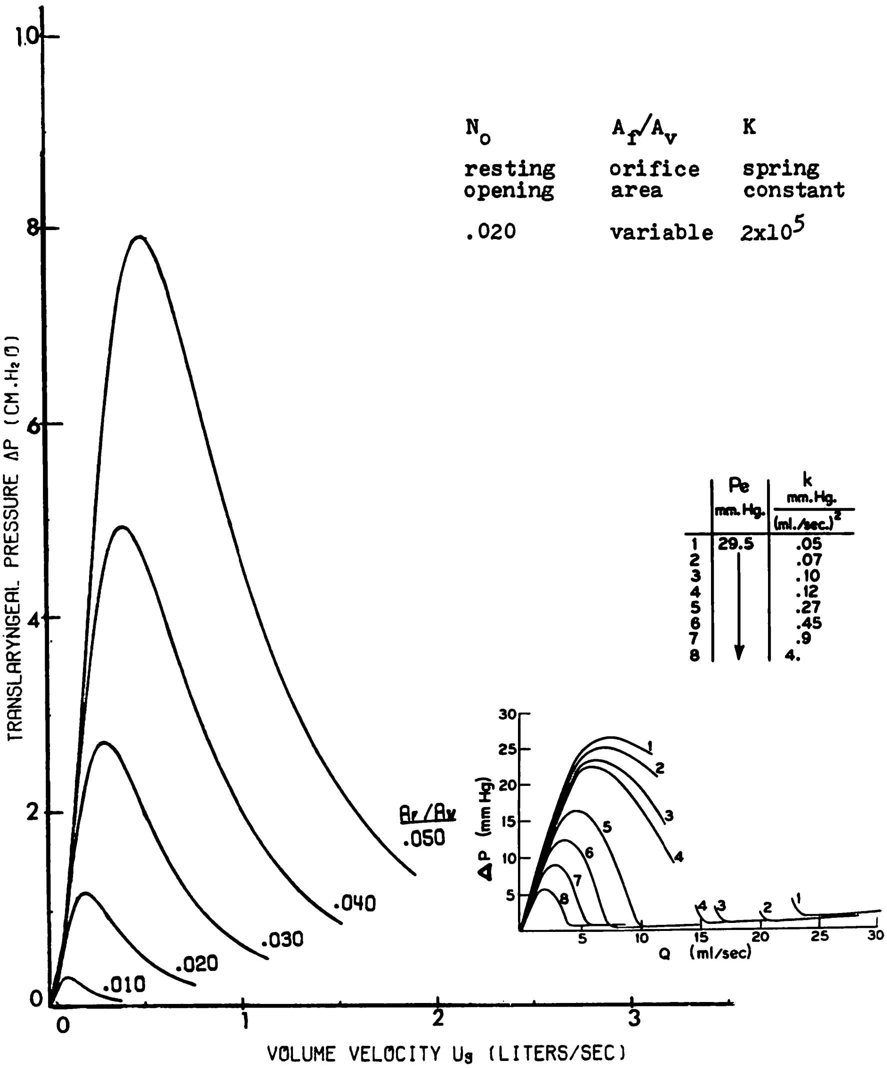


FIGURE 29-8
Family of translaryngeal pressure-volume velocity curves for constant resting opening, constant supraglottal orifice and variable spring constant.

Let me now show families of curves which illustrate the influence of these parameters.

Figure 29-7 shows a family of translaryngeal pressure versus volume velocity curves for variable resting opening, constant supraglottal orifice and constant spring constant. A fami-



N_0	A_f/A_v	K
resting opening	orifice area	spring constant
.020	variable	$2x10^5$

	Pe mm. Hg.	k $\frac{\text{mm. Hg.}}{(\text{ml./sec.})^2}$
1	29.5	.05
2	↓	.07
3		.10
4		.12
5		.27
6		.45
7		.9
8		4.

FIGURE 29-9

Family of translaryngeal pressure-volume velocity curves for constant resting opening, variable supraglottal orifice and constant spring constant.

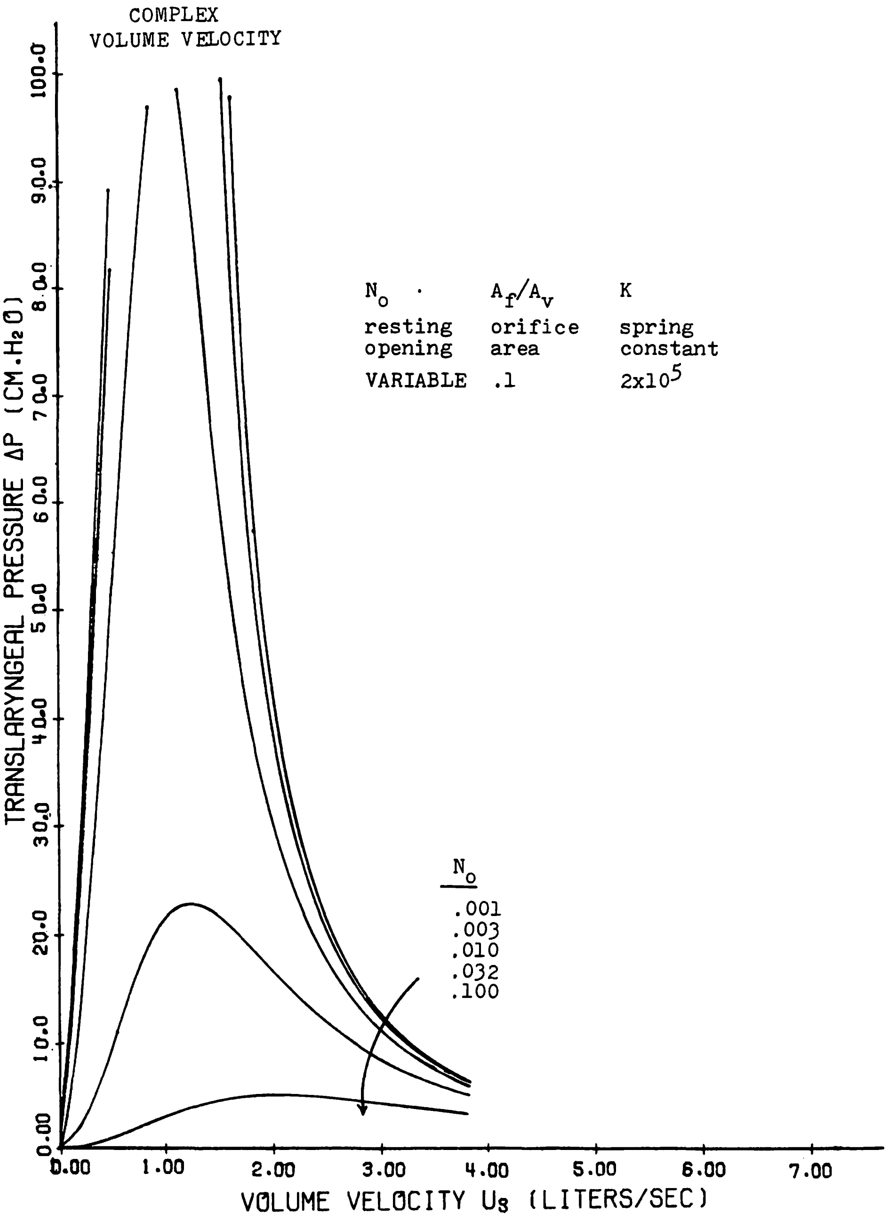


FIGURE 29-10
Pathological translaryngeal pressure-volume velocity curves for (small) variable resting openings, constant supraglottal orifice and constant spring constant. Volume velocity has complex values.

ly of collapsible tube curves with variable external pressure and constant downstream resistance is also shown. Clearly, both families rise and fall in qualitatively similar ways. The initial slope and peak pressure- drop increase with decreasing resting opening for the larynx and for increasing external pressure for the tube. However, the negative slope or negative differential resistance increases for the larynx but remains essentially constant for the tube.

Figure 29-8 shows a family of translaryngeal pressure versus volume velocity curves for constant resting opening, constant supraglottal orifice and variable spring constant. The same family of collapsible tube curves is shown as in the previous figure. Increasing the spring constant makes the system open less easily. The volume velocity and pressure to get comparable openings are larger. Peak pressure increases as does negative differential resistance.

Figure 29-9 shows a family of translaryngeal pressure versus volume velocity curves for constant resting opening, variable supraglottal orifice and constant spring constant. Collapsible tube curves for constant external pressure and variable downstream resistance are shown for comparison. As the area of the supraglottal orifice increases, that is, as its resistance decreases, the initial slope of the curves remains constant while the peak moves to higher volume velocities. This is similar to the collapsible tube result. Negative differential resistance for the larynx increases whereas that for the tube decreases.

Although the correspondence between the collapsible tube experiments and the calculations of the larynx model are encouraging, there are certain difficulties with these calculations. Figure 29-10 shows what happens for a supraglottal orifice of area one tenth the vocal tract area as the area of the resting opening of the glottis decreases to very small values. The peaks move to non-physiological pressures and volume velocities. Moreover, the volume velocity becomes complex. [Ishizaka (personal communication) suggests that complex flows be ignored because the assumed value of x does not apply.] Complex values are shown by breaks in the curves. The physical meaning of complex flows is not yet clear, but they are the result of assuming a supraglottal pressure recovery. If pressure recovery in the expansion region is actually less than expected and if the loss coefficient of the supraglottal orifice is changed, this will mend matters. These are questions for future study.

OSCILLATIONS

Figure 29-11 shows collapsible tube curves for constant P_e and variable downstream resistance. As downstream resistance decreases, indicated by decreasing k and curve number, the tube starts oscillating. Breaks in the curves mark the regions of oscillation. Note that the oscillation only occurs where the slope of the curve is negative. This point cannot be over- emphasized. It is the crucial focus of this presentation. Oscillations are always associated with a region of negative slope. The oscillation condition will be given in terms of this negative slope.

Figure 29-12 gives a detailed picture of the oscillations. Column (a) shows the oscillatory pressure-drop as a function of oscillatory flow rate at the downstream end in what is called a phase plane plot. Column (b) shows the build-up of flow rate in time and column (c) shows the oscillation on an expanded time scale. The phase plane plot shows that pressure and flow are out of phase as they must be in this type of oscillation. The circuit supporting the oscillation has both compliance and inertance. Since pressure and flow are out of phase, the tube cannot close in the same way that it opens. This is similar to the asymmetry in the opening and closing phases for the larynx as pointed out by Stevens (1977). Note also that as the operating point goes from peak to valley, the oscillation frequency decreases and their character changes from sinusoidal to nonsinusoidal. The decrease of frequency can be attributed to

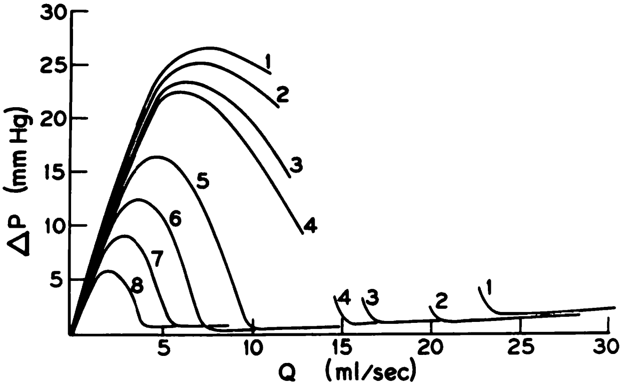


FIGURE 29-11

Family of ΔP-Q characteristics for constant external pressure and decreasing downstream resistance (decreasing k). Gaps in the curves represent regions of oscillation which increase as resistance decreases (Conrad, 1969).

changes in the nonlinear compliance of the tube. At the peak, the tube is collapsed and stiff, while at the valley it is open and soft. Titze (1976, 1980) suggested a similar reason for the change of laryngeal frequency with subglottal pressure.

CONDITIONS FOR OSCILLATION

Let us now examine the well-known conditions for oscillation for systems with this type of negative differential resistance. The condition applies both to the tube and the larynx. Figure 29-13 shows a piece-wise linear representation of the pressure-drop flow curves. Suppose we select a point on the curve and ask the question: what is the response to a positive perturbation in flow rate δq? At every operating point, that is, any point we may select, δq is associated with a pressure perturbation $\delta \Delta p$. $\delta \Delta p$ is positive in regions 1 and 3 but negative in region 2. Since differential resistance r is defined as the ratio of the pressure perturbation to the flow perturbation, then

$$r = \delta \Delta p / \delta q$$

is positive in regions 1 and 3 and negative in region 2. The differential resistance relates perturbations in flow rate to perturbations in pressure-drop. It is a so-called ac quantity which plays the role of resistance for perturbations. In this piece-wise linear representation, the differential resistance in regions 1 and 3 is the same as the ordinary resistance $\Delta P/Q$, but in region 2, the ordinary resistance decreases while differential resistance is negative.

When a system contains a negative differential resistance, self-excited oscillations are possible. The system must support a steady flow stably, that is, be stable for dc flow, and yet it must be unstable for nonsteady flow, that is, be ac unstable. The two requirements for oscillation are:

$|r| > R_{ext}$ ac unstable

$|r| < R_{dc}$ dc stable

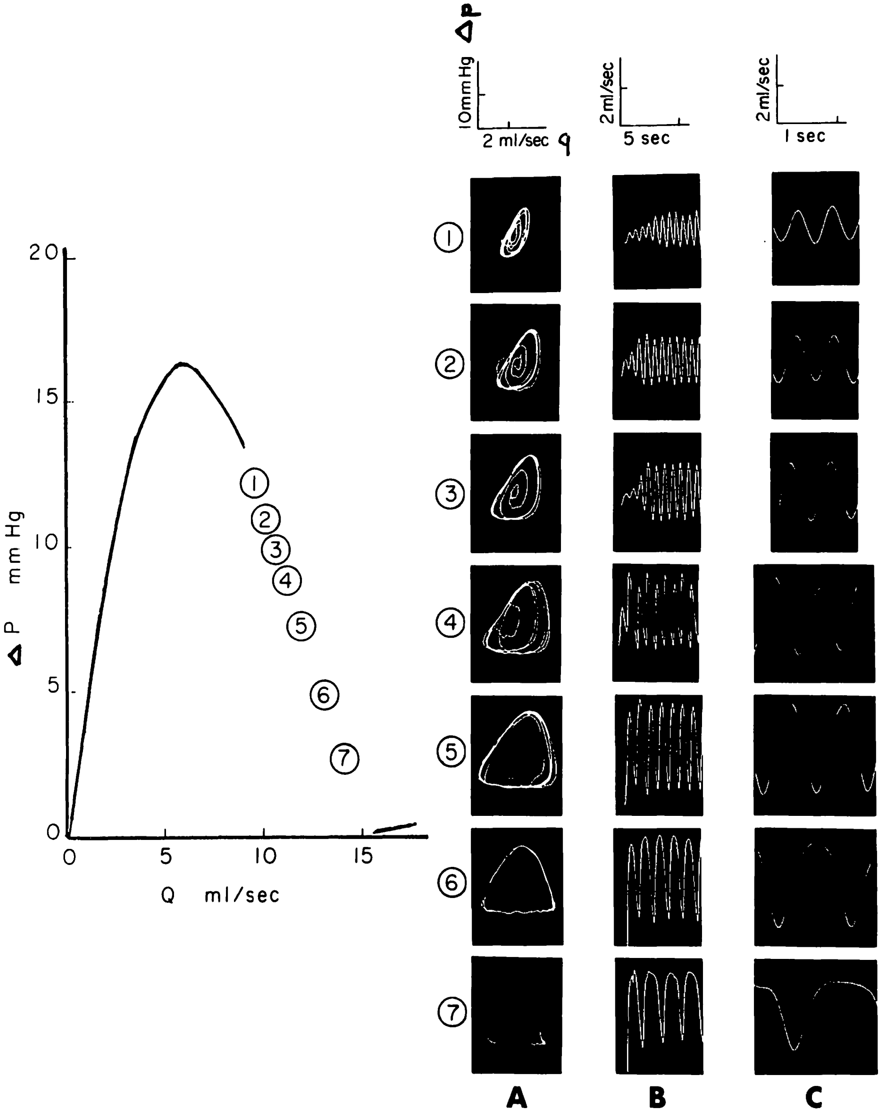


FIGURE 29-12

Relaxation oscillations in a collapsible tube as the static operating point moves through the region of negative slope. Column (a) shows the oscillatory pressure drop Δp across the tube plotted against the downstream oscillatory flow q in a phase plane diagram. Note the build-up of oscillation to a limit cycle. Column (b) shows the build-up in q as a function of time while column (c) shows a cycle of the oscillation on an expanded scale. As the operating point moves from peak to valley of the characteristic, the period increases, build-up is more rapid, and the oscillation is more non-sinusoidal (Conrad, 1969).

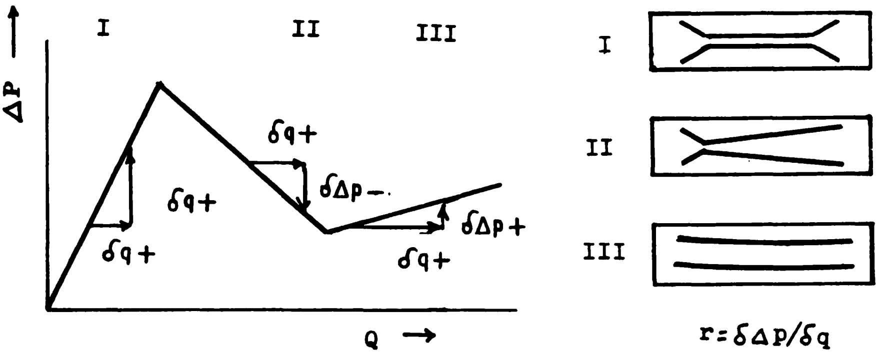


FIGURE 29-13
A positive flow rate perturbation $\delta\Delta p(-)$ in region II. The differential resistance $r=\delta\Delta p/\delta q$ is negative in region II.

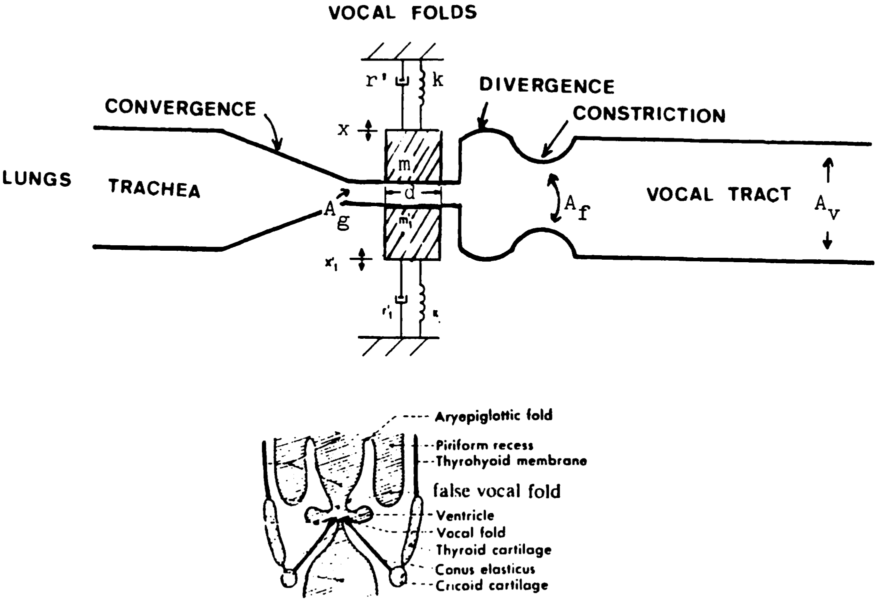


FIGURE 29-14
Schematic diagram of the larynx including a supraglottal resistance. The anatomical diagram shows that the false vocal folds could serve as the supraglottal resistance. (Hollinshead, 1982).

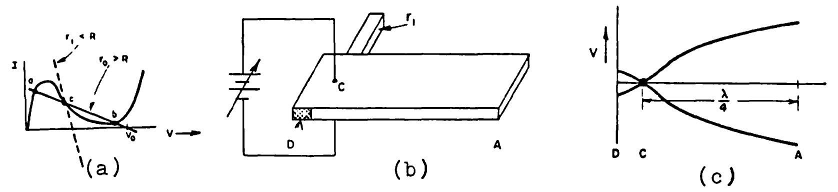


FIGURE 29-15
(a) Tunnel diode characteristic, (b) Stripline circuit showing tunnel diode at D and stabilizing resistor r_1, (c) standing wave pattern, open circuit at A. (Sommers, 1959).

that is, the absolute value of the negative differential resistance must be greater than the external resistance R_{ext}, yet less than dc resistance R_{dc}. R_{ext} is the actual resistance experienced by the perturbations in the sub and supraglottal circuits. R_{ext} is different from the steady flow resistance R_{dc} because of the compliance and inertance of the circuit. This oscillation condition applies to systems for which an increase in driving pressure opens the flow passage, so-called flow-controlled nonlinear resistances. The electrical analogue is a current-controlled nonlinear resistance. A general discussion of the oscillation condition is given by Reich *et al* (1966) in both analytical and graphical form.

APPLICATIONS OF LARYNX MODEL

Now that we have calculated pressure-flow curves for the larynx based on a widely accepted theory, let us consider some applications. Of course, these are purely speculative until the theory has been confirmed by experiment.

First, I would conjecture that the supraglottal resistance of the model is the false vocal folds. Figure 29-14 shows both a schematic and an anatomical diagram of the larynx. The ventricular or false folds are in the proper position and have the necessary compliance to serve as a supraglottal resistance. The combination is similar to a negative resistance tunnel diode oscillator with its stabilizing resistor shown in Figure 29-15. [Dr. Feder showed a film of the larynx of Rich Little imitating many speakers. It was obvious that supraglottal structures sometimes play a role in phonation. This was clear in the imitations of John F. Kennedy and Louie Armstrong and of Rich Little himself while speaking in his natural voice.]

Second, the calculations suggest that a larynx can be tested at surgery for potential phonation by steady flow means. One has merely to demonstrate that translaryngeal pressure decreases with volume velocity to assure that the larynx will phonate. Repeated testing, while constructing a prosthesis from the patient's own tissue, should be helpful. A steady flow test would also facilitate the development of a prosthetic larynx from artificial materials.

Third, pathological conditions of the larynx can be classified in terms of steady pressure-flow relations, though not for clinical practice. The classification could be useful when simulating pathological phonation on a computer.

Fourth, when teaching proper phonation, it may be helpful to focus attention on the supraglottal resistance after its site is identified.

The theory I have given for phonation is far from complete and in matters of this kind, practice is usually in advance of theory. Many practical men and women have brought the care of the voice to a high degree of excellence without this theory. Further development of this theory of phonation should parallel the familiar analyses of negative resistance oscillators with distributed circuits known in electrical engineering and acoustics. Why not try van der Pol's equation for the insight it can give?

ACKNOWLEDGMENT

My sincere thanks to my friend and colleague, Dr. David M. McQueen, who did the calculations and gave incisive editorial advice.

APPENDIX: GRAPHICAL STABILITY ANALYSIS

(See Reich *et al*, 1966; Chua, 1969; Cunningham, 1958)

The stability of a system containing a flow-controlled nonlinear resistance, or QNLR, can be discussed graphically in terms of load lines. The inset to Figure 29-A1 shows the circuit maintaining steady flow. We have

$$P_T = R_1Q + \Delta P(Q) + R_2Q$$

which can be rewritten as

$$P_T - (R_1 + R_2)Q = \Delta P$$

If we treat the left and the right hand sides of this equation as independent functions of Q, then the steady flow operating point is determined by the Q for which both functions are satisfied. The left hand side is called the steady flow or dc load line while the right hand side is the ΔP-Q curve or characteristic. As drawn, the operating point is stable, i.e. dc stable because the slope of the load line is more negative than that of the characteristic or $|r| < R_1 + R_2$. Nevertheless, the operating point can be unstable if R_1 and R_2 are effectively reduced by shunt compliances. The value of the differential resistance for the three regions of the piece-wise linear approximation for a collapsible tube is given in the box. For oscillations to occur, it is clear that the effective resistance R_{ext} must be less than R_2 since the negative differential resistance is at most R_2, i.e. we must have $R_2 > |r| > R_{ext}$. The compliance across R_1 and R_2 can convert the system from dc stable to ac unstable. A load line for the effective external resistance for nonsteady flow, the so-called dynamic load line, can be drawn. The figure shows a dynamic load line for an unstable system with $|r| > R_{ext}$.

REFERENCES

Attinger, E.O. and Anné, A. (1966). "Simulation of the cardiovascular system," *N.Y. Acad. Sci.* 128, 810-829.

Baer, T. (personal communication.)

van den Berg, Jw., Zantema, J.T. and Doornenbal, P., Jr. (1957). "On the air resistance and the Bernoulli effect of the human larynx," *J. Acoust. Soc. Am.* 29, 626-631.

Chua, L.O. (1969). *Introduction to Nonlinear Network Theory*. R.E. Krieger, Huntington, NY, Sect. 3-2-4.

Conrad, W.A. (1969). "Pressure-flow relationships in collapsible tubes," *IEEE Trans. Biomed. Eng.* BME 16, 284-295.}

Conrad, W.A. (1980). "A new model of the vocal cords based on a collapsible tube analogy," *Med. Res. Eng.* 13, 7-10.

Cunningham, W.J. (1958). *Introduction to Nonlinear Analysis.* McGraw-Hill, New York, Sects. 3.3, 3.7, 5.8, 6.5

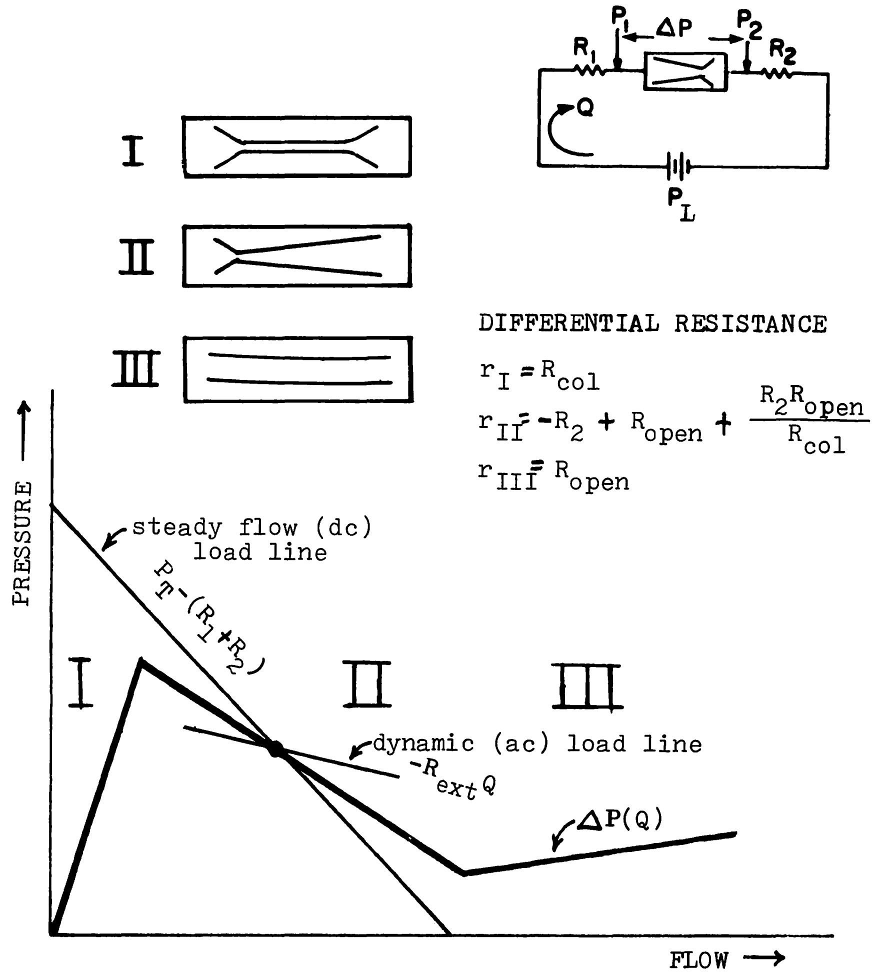


FIGURE 29-A1

Piece-wise linear ΔP-Q characteristic showing steady flow (dc) load line and dynamic (ac) load line. The system is dc stable, ac unstable because $|r_{II}| < R_1+R_2$ and $|r_{II}| > R_{ext}$ where R_{ext} is the effective ac resistance of the circuit external to the QNLR. r_I, r_{II}, and r_{III} are the differential resistances for a collapsible tube.

Flanagan, J.L. (1958). "Some properties of the glottal sound source," *J. Speech Hear. Res.* 1, 99-116.

Flanagan, J.L. and Landgraf, L. (1968). "Self-oscillating source for vocal tract synthesizers," *IEEE Trans. Audio and Electro-acoust.* AU-16, 57-64.

Hollinshead, W.H. (1982). *Anatomy for Surgeons: Vol. I, The Head and the Neck.* Harper and Row, New York, Chap. 8.

Ishizaka, K. (personal communication).

Ishizaka, K. and Flanagan, J.L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Sys. Tech. J.* 51, 1233-1268.

Ishizaka, K. and Flanagan, J.L. (1977). "Acoustic properties of longitudinal displacement in vocal-cord vibrations," *Bell Sys. Tech. J.* 56, 889-918.

Ramo, S. and Whinnery, J.R. (1953). *Fields and Waves in Modern Radio* Van Nostrand, Princeton, New Jersey, Sect. 11.10.

Reich, H.J., Skalnik, J.G. and Krauss, H.L. (1966). *Theory and Application of Active Devices.* Wiley, New York, Sect. 4.3

Rothenberg, M. (1981). "Acoustic interaction between the glottal source and the vocal tract," in K.N. Stevens and M. Hirano (Eds.) *Vocal Fold Physiol.* Toky: Univ. of Tokyo Press, 305-328.

Sommers, H.S., Jr. (1959). "Tunnel diodes as high-frequency devices," *Proc. IRE* 47, 1201-1206.

Stevens, K.N. (1981). "Vibration modes in relation to model parameters," in *Vocal Fold Physiol.*, K.N. Stevens and M. Hirano (Eds.), Univ. of Tokyo Press, Tokyo, 291-304.

Titze, I.R. (1976). "On the mechanics of vocal-fold vibration," *J. Acoust. Soc. Am.* 60, 1366-1380.

Titze, I.R. (1980). "Comments on the myoelastic-aerodynamic theory of phonation," *J. Speech Hear. Res.* 23, 495-510.

DISCUSSION

G. FANT: I am troubled by your assumption of a resistive supraglottal load. Standard theory provides an essentially inductive load. Does your model oscillate with an inductive load?

W. CONRAD: The answer is yes. I spoke of the resistive portion of the load alone for simplicity. The actual oscillation condition for a flow-controlled nonlinear resistance (QNLR) is (Conrad, 1980):

$$|r| > \mathrm{Re}Z, \qquad \mathrm{Im}Z = 0, \ \frac{\partial\, \mathrm{Im}Z}{\partial \omega} > 0,$$

where Z is the impedance of the circuit across the QNLR. Z has contributions from circuits exterior to the larynx and it should also include the transverse resonant circuit representing wall motion which I proposed in my 1980 paper on the larynx. The condition $\partial\, \mathrm{Im}Z / \partial \omega > 0$ ensures, by Foster's reactance theorem (Ramo and Whinnery, 1953), that the lumped equivalent circuit for Z can be represented by a series resonant circuit. The impedance Z should be dominated by the transverse wall impedance.

It is suggestive to look at the phase plane of collapsible tube oscillation, Figure 29-12, column (a). At operating point 7, the pressure change versus flow curve is a triangle. As the operating point moves up the column, the triangle becomes skewed to the right in a similar fashion to the flow-time curves calculated by Rothenberg (1981). As the operating point moves from the valley of the characteristic to the peak, the tube progressively collapses, see Figure 29-3 (b), and becomes more rigid. Since the inertance of the tube varies inversely as its area while compliance varies as the area, all other things being equal, inertance increases while compliance decreases. The skewing of the phase plane trajectory represents an effect of increasing inertance in the oscillations. This does not prove what happens as inertance external to the tube increases, but it is suggestive of the answer.

H. TEAGER: I'm fascinated by your work and think you have done a marvelous job. Am I correct in assuming that the kind of negative resistance oscillation you are describing here, the collapsible tube, is pretty much the same sort we used to make as kids when we blew up a balloon, held the top tight, and let it fly away?

W. CONRAD: Yes, that's exactly right.

H. TEAGER: In that case, there is no downstream resistance. It is forming a jet, and going off into the atmosphere. Have you thought about tying this to that?

W. CONRAD: I can't say anything about the jet phenomenon, because I haven't given it much thought, but I do feel that whenever you have a compliant structure, as we have in the balloon and the larynx and the lips and so on, you're going to get something like what I found, because there is always a little downstream resistance somewhere. And that gives enough back pressure to couple the wall to the flow, and then if things are right you will get the negative differential resistance.

O. FUJIMURA: In the early 60's, Svend Smith had a collapsible tube model of the vocal folds which he called a munyo. Isn't it that there is the required positive (load) resistance at the very edge of the narrowly constricted tube (which corresponds to the upper edges of the vocal folds)?

W. CONRAD: For a collapsible tube at least, the system becomes more and more unstable as you decrease the downstream resistance. I was extremely unhappy to discover that, in my calculations, the negative differential increased as the supraglottal resistance decreased. However, that decrease in the supraglottal resistance is due to the presence of the minus sign in front of the resistance of the expansion. So I think if we get rid of that or change that somehow, then the calculated negative differential resistance of the larynx will fall more in line with that of the collapsible tube. That ought to suggest that as you decrease the supraglottal resistance, the thing will oscillate, even in the limit when you have no downstream resistance. Note that Tom Baer (personal communication) pointed out at this meeting that the downstream resistance in my calculations can be considered as a fixed second mass of a two-mass model.

P. MILENKOVIC: The mass of the fluid in your test apparatus is large compared to the mass of the walls. Does this suggest that vocal cord mass is not a parameter in this system?

W. CONRAD: In the collapsible tube, the mass of the wall doesn't play any role. But in the larynx, it of course does play a role, and a very important role. I have proposed (Conrad, 1980) an equivalent circuit for the larynx in which the mass of the larynx determines the inertance of a transverse resonant circuit. The circuit, except for the QNLR, is similar to that for a thick-walled artery proposed by Attinger *et al* (1966). Ingo is going to show you something quite similar to this.

K. ISHIZAKA: Is the phase difference in the motion of the collapsible tube similar to that in vocal fold motion?

W. CONRAD: The answer is yes. A thin-walled collapsible tube has an asymmetry in the walls built into it and so the walls do move asymmetrically. As I have shown in unpublished experiments, a thick-walled rubber tube symmetrically clamped will behave like a rigid stenosis for steady flow, i.e., the pressure difference across the stenosis is proportional to the square of the volume flow, and there is no negative differential resistance. However, if some asymmetry in the clamping is introduced, a differential negative resistance is formed for steady flow and the tube will oscillate. The walls must move asymmetrically for an asymmetrically clamped tube.

30 ▪ MECHANISMS OF SUSTAINED OSCILLATION OF THE VOCAL FOLDS

Ingo R. Titze

ABSTRACT

Flow-coupled differential equations are derived for displacement of the vocal folds at the midpoint of the glottis. This allows comparison of the characteristics of oscillation with a simple mass-spring oscillator. Negative resistances and stiffnesses are derived for larger-amplitude vibration, but conditions of closure are not included. A force-displacement diagram highlights the energy exchange mechanism between the airstream and the vocal folds.

INTRODUCTION

The mechanism of self-oscillation of the vocal folds has been explained in terms of a negative aerodynamic coupling stiffness between the upper and lower portions of the vocal folds (Ishizaka and Matsudaira, 1972). More recently, Conrad (1980) has argued that the vocal folds and the laryngeal airway constitute a collapsible tube and that a negative dynamic resistance is produced by the yielding wall in the glottis when the supraglottal pressure is held constant. Common to both explanations is the fact that nonlinearities in the pressure-flow relationships can create nonlinearities in the effective stiffness, damping, and driving pressures. These nonlinearnities are responsible for instability in tissue displacement.

In order to conceptualize the self-oscillatory mechanism, Stevens (1977) and Titze (1980) have suggested that oscillation can occur whenever an asymmetry in the aerodynamic driving forces exists with respect to the opening and closing portions of the vocal folds. This asymmetry is usually facilitated by phase delay in movement between the upper and lower portions of the vocal folds, but it can also be enhanced by inertial acoustic loading of the vocal tract, indicating that vocal fold vibration is not independent of the vocal tract system.

Mathematically, the conditions for self-oscillation were established by Ishizaka and Matsudaira (1972) on the basis of the Routh-Hurwicz criterion for small amplitude displacement and by Conrad (1980) on the basis of zero net resistance in the flow circuit describing glottal aerodynamics. In this paper the two conditions will be shown to be part of a more general criterion for oscillation that involves nonlinear coupling between air and tissue movement and is not limited to small-amplitude displacement. Because the single-mass, single-spring oscillator is conceptually the easiest to understand, we will lump the effects of spatial variation of tissue movement (vertical phase differences within the glottis) into the framework of an equivalent one-mass system. Our main contribution will then be the derivation and discussion of the asymmetric driving pressure on the vocal folds at the center of the glottis, and the associated energy exchange between air and tissue.

LUMPED ELEMENT FLOW CIRCUIT

A lumped-element flow circuit that contains all the necessary conditions for oscillation is shown in Figure 30-1. The circuit is a low-frequency circuit in the sense that the compliance of the air is neglected. Thus, subglottal and supraglottal airways are represented by resistances R_1, R_2 and inertances I_1, I_2 respectively. A resistance R_c in the contraction (entry) region and a resistance R_E in the expansion (exit) region connect the upper and lower airways to

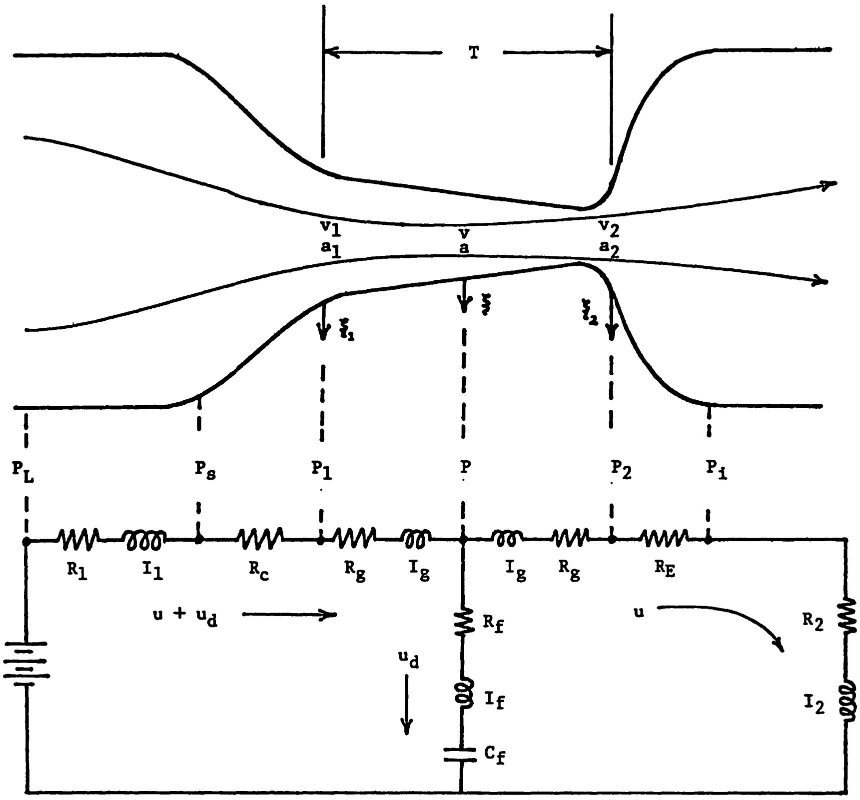


FIGURE 30-1
Lumped element flow circuit corresponding to glottal airway above.

the glottal region. The glottis is represented by a T-section that includes glottal air resistance and inertance R_g and I_g, respectively, and the yielding wall shunt elements R_f, I_f, and C_f. The latter represent the displacement flow resistance, inertance, and compliance of the vocal folds. It is important to note that the volume flow is not considered constant in the glottis. Above the midpoint of the glottis the flow is $u + u_d$, where u_d is the displacement flow, and below the midpoint the flow is the exit flow u.

Further definitions include a_1 and v_1 as glottal area and particle velocity at entry, a_2 and v_2 as glottal area and particle velocity at exit, and a and v as similar quantities at the midpoint. Since the length of the glottis is the thickness of the vocal folds, it is labeled T in Figure 30-1, and the length of the vocal folds (not visible in Figure 30-1) is designated L. This makes the medial surface area of the vocal folds LT, and the displacement flow

$$u_d = 2LT\dot{\xi} \qquad (30\text{-}1)$$

where $\dot{\xi}$ is the time-derivative of the lateral displacement of the vocal folds at the center of

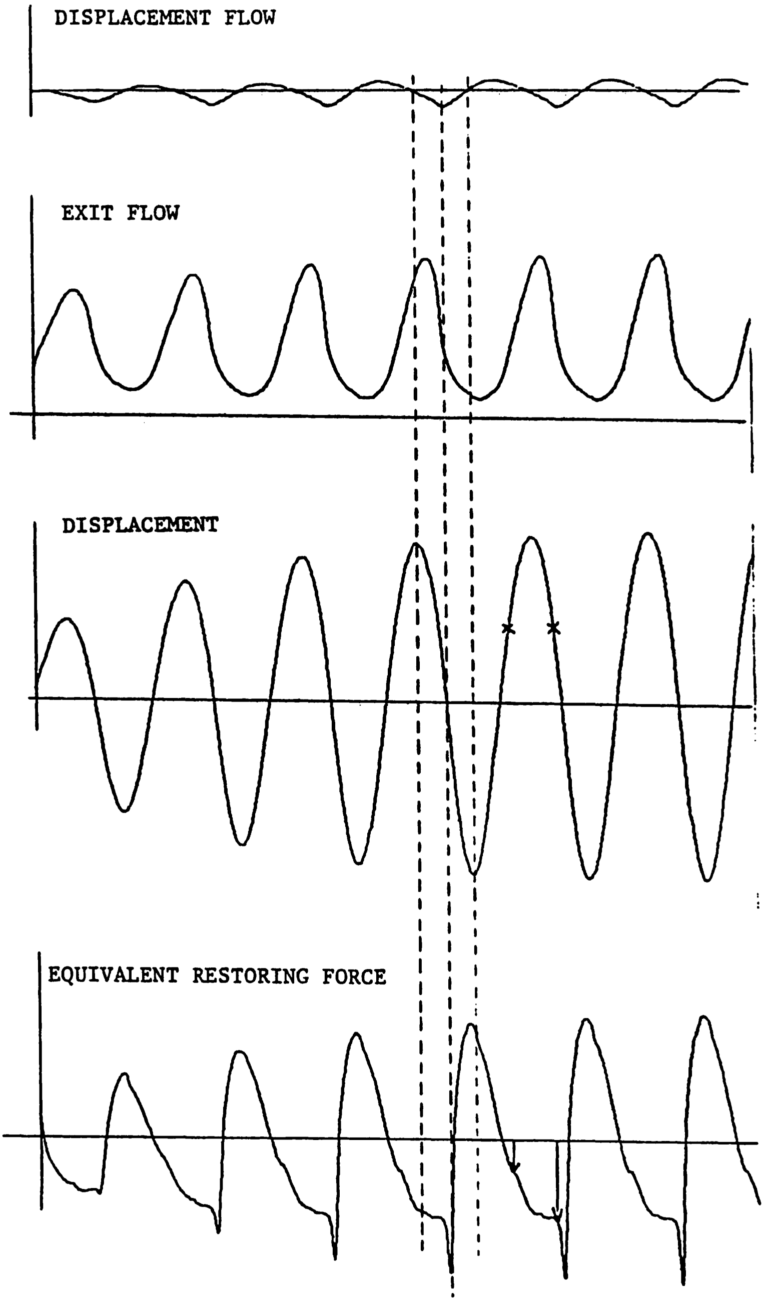


FIGURE 30-2

Simulation of displacement flow, glottal (exit) flow, midpoint displacement, and equivalent restoring force for a simple mass-spring oscillator. Total simulation time is 50 ms.

the glottis. Left-right symmetry has been assumed.

The values of the lumped flow resistances, inertances, and tissue compliance have been derived recently by the author (Titze, 1985) in terms of fundamental physical quantities, i.e., tissue and air density, tissue and air viscosity, imposed intrinsic muscle tensions and non-isotropic elastic constants. We will draw upon these results as necessary to elucidate the oscillation mechanisms, but a complete discussion is beyond the scope of this paper.

FLOW-COUPLED EQUATIONS OF MOTION

Two loop equations can be added to the differential equation (30-1) to describe the coupling between air flow and tissue movement. The upstream loop gives

$$\begin{aligned}&(R_1+R_c+R_g)u + (I_1+I_g)\dot{u} + (R_1+R_c+R_g+R_f)u_d \\ &+ (I_1+I_g+I_f)\dot{u}_d + (2LT/C_f)\xi = P_L\end{aligned} \qquad (30\text{-}2)$$

Here we have expressed the pressure across the tissue compliance as an integral of the displacement flow, which is easily evaluated from equation (30-1). The initial condition for ξ was set to zero.

The downstream loop equation is similarly written as

$$(R_2+R_E+R_g)u + (I_2+I_g)\dot{u} - R_f u_d - I_f\dot{u}_d - (2LT/C_f)\xi = 0 \qquad (30\text{-}3)$$

This equation can be solved for $\dot{u}_d$ and substituted into equation (30-2) to make the latter a function of u, U_d, and $\dot{u}$, but not a function $\dot{u}_d$. The three coupled differential equations can then be re-written as follows

$$\dot{\xi} = \frac{1}{2LT}u_d \qquad (30\text{-}4)$$

$$\dot{u} = [P_L - (R_1^*+rR_2^*)u - (R_1^*+R_f-rR_f)u_d - (1-r)\,(2LT/C_f)\xi] \,/\, (I_1^*+rI_2^*) \qquad (30\text{-}5)$$

$$\dot{u}_d = [R_2^*u + I_2^*\dot{u} - R_f u_d - (2LT/C_f)\xi] \,/\, I_f, \qquad (30\text{-}6)$$

where

$$R_1^* = R_1+R_c+R_g = \text{upstream resistance} \qquad (30\text{-}7)$$

$$R_2^* = R_2+R_E+R_g = \text{downstream resistance} \qquad (30\text{-}8)$$

$$I_1^* = I_1+I_g = \text{upstream inertance} \qquad (30\text{-}9)$$

$$I_2^* = I_2+I_g = \text{downstream inertance} \qquad (30\text{-}10)$$

$$r = (I_1+I_g+I_f) \,/\, I_f = \text{coupling ratio} \qquad (30\text{-}11)$$

Equations (30-4), (30-5), and (30-6) can easily be integrated by using a Runge-Kutta numerical technique. If initial values of u_d, ξ, and u are specified (usually all zero), the derivatives are evaluated and new values are obtained by integration for the next time step. It is important to compute the derivatives in the sequence given above, since $\dot{u}$ needs to be known before $\dot{u}_d$ can be evaluated. The flow resistances and inertances in equations (7-11) need not be constants, but can vary freely with u, u_d, or ξ. We will see shortly that they are in fact highly nonlinear functions of flow and displacement.

ASYMMETRIES IN TISSUE DISPLACEMENT

The non-uniform displacement of the tissue within the glottis (sometimes called vertical phasing) is primarily a result of mucosal surface waves propagating in the direction of the flow (the + z direction). Such travelling waves can be expressed in the general D'Alembert form

$$\xi(z,t) = f\,(t-z/c), \qquad (30\text{-}12)$$

where f can be any waveshape function and c is the surface wave velocity. It is clear that

$$\frac{\partial \xi}{\partial z} = -\frac{1}{c}\frac{\partial \xi}{\partial t} = -\dot{\xi}/c, \qquad (30\text{-}13)$$

regardless of the waveshape, but if we assume a linearly convergent or divergent glottal shape, the displacements at glottal entry and exit can be written as

$$\xi_1 = \xi - \frac{\partial \xi}{\partial z}\left(\frac{T}{2}\right) = \xi + \left(\frac{T}{2c}\right)\dot{\xi} = \xi + u_d/(4Lc), \qquad (30\text{-}14)$$

$$\xi_2 = \xi + \frac{\partial \xi}{\partial z}\left(\frac{T}{2}\right) = \xi + \left(\frac{T}{2c}\right)\dot{\xi} = \xi - u_d/(4Lc). \qquad (30\text{-}15)$$

Note the asymmetry in upstream and downstream displacement with respect to the midpoint displacement and velocity.

The nonlinear flow resistances R_c, R_g and R_E are complex functions of the Reynolds numbers at glottal entry and exit (Scherer, 1981: Gauffin, *et al* 1981). Typically, they can be expressed in terms of the kinetic pressure $1/2\varrho v^2$. Important quantities to consider, then, are the entry, midpoint, and exit particle velocities v_1, v, and v_2 together with the corresponding glottal areas a_1, a, and a_2 (Figure 30-1). If ξ_{01} and ξ_{02} are pre-phonatory displacements at glottal entry and exit, respectively, then the particle velocities can be written as

$$v_1 = \frac{u+u_d}{a_1} = \frac{u+u_d}{2L(\xi_{01}+\xi) + u_d/2c} \qquad (30\text{-}16)$$

$$v_2 = \frac{u}{a_2} = \frac{u}{2L(\xi_{02}+\xi) - u_d/2c}\,. \qquad (30\text{-}17)$$

These relationships demonstrate the important asymmetry that exists in the particle velocities (and hence the flow resistances) with respect to opening and closing of the glottis. During glottal opening u_d is positive. This decreases the upstream velocities and resistances while at the same time increasing the downstream velocities and resistances. Both of these changes have the effect of increasing the midpoint driving pressure P (Figure 30-1). During glottal closing, on the other hand, u_d is negative, creating the opposite effect of decreasing the midpoint driving pressure. This and other asymmetries in the driving pressure keep the vocal folds in oscillation, as we will demonstrate next.

ASYMMETRIES IN THE NET DRIVING PRESSURE

With the aid of Figure 30-1, the midpoint driving pressure can be written as

$$P = R_2^* u + I_2^* \dot{u} \tag{30-18}$$

$$= R_2^* a v + I_2^* \dot{a} v + I_2^* a \dot{v} \tag{30-19}$$

$$= 2L(R_2^* v + I_2^* \dot{v})\,(\xi_0 + \xi) + (I_2^* v/T) u_d \tag{30-20}$$

where the midpoint area $a = 2L(\xi_0 + \xi)$ has been used to relate the midpoint particle velocity v to the volume flow u, and the derivative of ξ has again been expressed in terms of the displacement flow.

Several asymmetries are immediately evident. Aside from the movement asymmetry that is reflected in R_2^* and I_2^* as suggested in the previous section, the term proportional to u_d in equation (30-20) will raise the driving pressure during opening and lower it during closing. Thus, even if no movement asymmetry exists (no vertical phasing), a mechanism for self-oscillation exists on the basis of downstream inertance. This can be either glottal inertance or vocal tract inertance, or both. The one-mass model of Flanagan and Landgraf (1968) produced oscillation on the basis of this inertance. As the vocal tract acoustic impedance becomes compliant, however, above the first formant frequency, the glottal inertance may not be sufficient to overcome the vocal tract compliance, in which case asymmetry in movement is essential to sustain oscillation.

NEGATIVE STIFFNESS AND RESISTANCE

By equating the pressure P from equation (30-20) to the total pressure across the shunt circuit in Figure 30-1 (as was previously done to obtain equation 30-3), it is possible to identify the effective flow-coupled tissue resistance and compliance as combined coefficients of u_d and ξ. This leads to

$$R_f^* = R_f - I_2^* v/T \tag{30-21}$$

$$2LT/C_f^* = 2LT/C_f - 2LR_2^* v - 2LI_2^* \dot{v}, \tag{30-22}$$

with a residual driving pressure of

$$P^* = 2L(R_2^* v + I_2^* \dot{v})\xi_o. \tag{30-23}$$

These relationships demonstrate that both negative resistance and negative stiffness terms exist. These terms are much more general, however, that the negative stiffness derived by Ishizaka and Matsudaira (1972) for small amplitude oscillation. In fact, we have shown (Titze, 1985) that their negative reacting stiffness is essentially the equivalent of the first pressure term in equation (30-23) for a specialized R_2^* that includes only the ideal Bernoulli pressure drop within the glottis.

RESULTS OF SIMULATION

The foregoing equations were programmed in FORTRAN on a VAX 11/780 computer.

A fourth-order Runge-Kutta intergration was used. Since we have not discussed conditions of closure of the glottis in the present theory, a sufficiently large pre-phonatory glottal width ($x_{01} = x_{02} = 0.05$ cm) was chosen to allow some leakage of flow during all parts of the cycle, as in breathy voice. Other parameters were $P_L = 8.0$ cm H_2O, $L = 1.4$ cm, $T = 0.3$ cm, $R_1 = R_2 = 1.0$ dyn-s/cm^5, $I_1 = .002$ dyn-s^2/ cm^5, $I_2 = .02$ dyn-s^2/cm^5, and glottal and vocal fold lumped elements as described in a more detailed treatment (Titze, 1985).

In order to relate the oscillation characteristics to those of an ideal mass-spring oscillator, the tissue restoring force and the viscous damping force of the vocal fold were added to the driving force PLT to obtain the total force F on the folds. According to Newton's second law, then,

$$M\ddot{\xi} = (2L^2T^2I_f)\dot{u}_d = F(\xi, \dot{\xi}, u,. . .); \tag{30-24}$$

where M is the equivalent mass of the folds. We will call F the equivalent restoring force because, in the simplest mass-spring mechanical oscillator, $F = -K\xi$, where K is the spring constant. This equivalent restoring force, together with the displacement ξ, the exit glottal flow u, and the displacement flow u_d, is plotted in Figure 30-2.

Several observations are important with regard to the relationships between these waveforms. First, note that displacement and displacement flow are 90° out of phase. This is expected, since the latter is proportional to the derivative of the former. Also note (dotted lines) that the peak of the exit flow lags behind the peak of the displacement. This has been attributed to air inertance and has been described in detail by Rothenberg (1981). Most importantly, however, the equivalent restoring force is seen to contain the proper asymmetries to keep the folds in sustained oscillation. For an ideal oscillator, the force would be sinusoidal and 180° out of phase with respect to the displacement $K\xi$). This general appearance is also observed, but nonlinearities are apparent. These help to overcome the viscous losses in the tissue. Note that, for an equivalent value of displacement during lateral and medial movement (marked crosses on Figure 30-2), the restoring force is greater for medial movement than for lateral movement (arrows in Figure 30-2). Although both forces are in opposition to the displacement in this case, the net force imparted in this half-cycle is in phase with the velocity, overcoming normal damping. There are cases, in fact, where the positive driving pressure overcomes the tissue restoring force during lateral movement, causing the net force to be in phase with velocity over most of the half cycle.

The pulse-like component at the lower extreme of the force in Figure 30-2 is a result of the nonlinear Bernoulli pressure. It occurs prior to the minimum displacement because the bottom of the folds reaches minimum displacement before the midpoint. The midpoint displacement is, of course, what is being plotted here.

Figure 30-3 shows a cyclic force-displacement diagram. The build-up and asymptotic conditions are clearly seen. This graph should be viewed with reference to a diagonal line that has a negative slope $K\xi$) in the case of an ideal oscillator. Again, the basic similarity is there, but the addition of two loops (a figure eight, roughly) is most revealing. Energy is supplied to the vocal folds by the airstream when the enclosed region is traversed clockwise (positive displacement), whereas energy is dissipated during negative displacement (counterclockwise traversing of the enclosed region). The magnitudes of the energies supplied and dissipated per cycle are, in fact, equal to the areas enclosed. During build-up of oscillation, the area on the right is greater than the area on the left, but the areas become exactly equal when oscillation is asymptotic.

The force-displacement diagram of Figure 30-3 can take on many different contours,

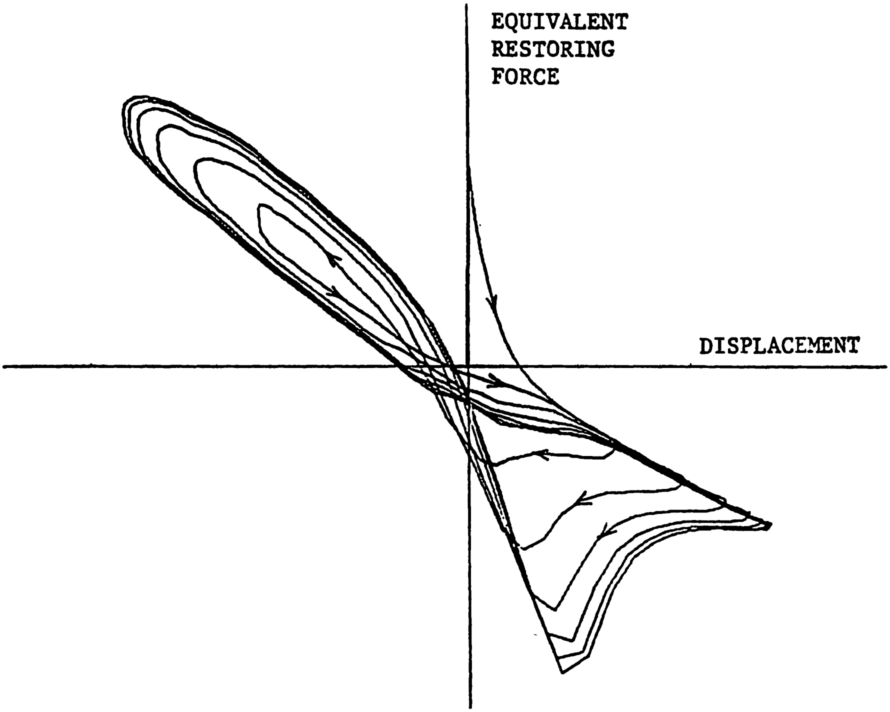


FIGURE 30-3
Force displacement diagram for the bottom two curves of Figure 30-2.

depending on the pre-phonatory shape of the glottis and any of the circuit elements in Figure 1. This is only a single case, and not necessarily a typical one, since vocal fold collision has not been addressed. Details for closure conditions are presently being investigated, however (Titze, 1985).

In summary, this investigation has born out previous hypotheses that (1) oscillation is dependent upon asymmetries in the net driving force, (2) flow-coupled nonlinearities exist in the form of negative resistances and negative stiffnesses, and (3) oscillation is dependent upon air inertances in the glottis and vocal tract as well as the asymmetry in movement often called vertical phasing.

ACKNOWLEDGEMENT

This work was supported by a grant from the National Institutes of Health, number NS 16320-03. The author also wishes to thank Dr. Fariborz Alipour-Haghighi for the details on the Runge-Kutta solution and other helpful suggestions.

REFERENCES

Conrad, W. (1980). "A new model of the vocal cords based on the collapsible tube analogy," *Med. Res. Eng.* 13/2, 7-10.

Flanagan, J. and Landgraf, L. (1968). "Self-oscillating source for vocal tract synthesizers," *IEEE Trans. Audio Electroacoust.* AU-16, 57-64.

Gauffin, J., Binh, N., Ananthapadmanabha, T. and Fant, G. (1981). "Glottal geometry and volume velocity waveform," *Vocal Fold Physiol. Conf.*, Madison, WI.

Ishizaka, K. and Matsudaira, M. (1972). *Fluid Mechanical Considerations of Vocal Cord Vibration.* SCRL Monography #8, Santa Barbara, CA.

Rothenberg, M. (1981). "Acoustic interaction between the glottal source and the vocal tract," *Vocal Fold Physiology*, K.N. Stevens and M. Hirano (Eds.) Univ. of Tokyo Press, 304-323.

Scherer, R. (1981). "Laryngeal fluid mechanics: steady flow considerations using static models," doctoral diss., Univ. of Iowa.

Stevens, K. (1979). "Physics of laryngeal behavior and larynx modes," *Phonetica* 34, 264-279.

Titze, I. (1980). "Comments on the myoelastic-aerodynamic theory of phonation," *J. Speech Hear. Res.* 23/3, 495-510.

Titze, I. (1985). "The physics of flow-induced oscillation of the vocal folds," in preparation.

DISCUSSION

G. FANT: Your model appears to be representative of breathy phonation as I have experienced from inverse filtering. The apparent feature is the location of the point of maximum slope change, and thus of excitation, which occurs downhill from the peak flow. The correspondence would be further illucidated if you differentiated your flow functions.

I. TITZE: That is a good point. We will attempt to do that in future reports in order to get a better comparison between your results and ours.

P. MILENKOVIC: You mentioned the mucosal wave on the vocal folds. Is this wave, or its effects, incorporated into one of the parameters of the lumped parameter model? Or is it dealt with apart from the model?

I. TITZE: It is incorporated in the lumped parameters via the displacement flow U_d. It turns out that the phase difference in movement between upper and lower portions of the vocal folds can be related to the displacement flow at the center, as detailed in the paper.

31 ▪ SOME OBSERVATIONS ON VOCAL TRACT OPERATION FROM A FLUID FLOW POINT OF VIEW

James F. Kaiser*

ABSTRACT

A number of ideas on the topic of the vocal tract, its operation, relevant models, and some important newly reported phenomena that yield a whole new way of viewing speech production in human beings are presented. These ideas center on considering the vocal tract operation from the air flow point of view. Separated flows in the forms of jets and vortices are argued to exist in the vocal tract during voicing. The interaction of these separated flows, i.e. the jet-cavity flow paradigm, is felt to be largely responsible for much of the fine structure of sound generation within the tract; this sound is in addition to the basic bulk glottal flow wave of pitch frequency. Experimental evidence of the existence of separated flow is advanced and a number of implications of this phenomena are discussed. Suggestions are made for further research experimentation. A comprehensive set of references to published related work is included.

INTRODUCTION

Two years ago at the 1981 Wisconsin Vocal Fold Physiology Conference H.M. Teager presented some very novel ideas and detailed measurements on the structure of air flow in the forward part of the vocal tract. These ideas were sufficiently different from those commonly assumed for the nature of air flow in the vocal tract that it was felt worthwhile to re-examine in some detail the more widely used vocal tract models and their underlying assumptions and to study the implications of these assumptions on sound generation. An end goal was to see if the new results could be easily incorporated into those existing models.

This paper presents a number of ideas on the topic of the vocal tract and some important newly reported phenomena that yield a whole new way of viewing speech production in human beings. It is hoped that by placing these ideas and the discussion together in a single paper, they will be examined carefully by those with considerably greater and more varied experience in the speech area than the writer.

To set the stage for the discussion of the new results we first examine the source-filter theory of speech production, its basic assumptions, and extensions to the more comprehensive vocal tract models. A number of observations and anomalies that are not simply or easily explainable with this theory are then listed and discussed. Some new results are described along with the experimental techniques involved. The anomalies are re-examined in light of the new results and a number of conjectures are advanced as to an additional mechanism for the production of sound in the vocal tract. Several possible experiments to test out some of the conjectures are also described.

These conjectures on additional mechanisms for the production of sound in the vocal tract are partially the result of my contact over the last three years with Dr. H.M. Teager of the Biomedical Engineering Section of the Boston University Medical Center, the researcher whose discoveries sparked a renewed interest in the operation of the vocal tract. Teager in his studies and extensive measurements on air flow in the vocal tract has found a number of previously unobserved phenomena inside the vocal tract during phonation which suggest that the currently accepted source-filter model of the vocal tract is inadequate as far as describing the details of what is really going on in the production of sound when we speak or sing. This knowledge of the presence of new phenomena has indicated to me the compelling need for a whole new look at the operation of the vocal tract to either prove or disprove the relevance and importance of these new mechanisms and conjectures. They are sufficiently different to cause considerable consternation to the experienced worker in the field. This paper then is a discussion of some of the ideas that have resulted from that contact and the thinking which followed.

MOTIVATION

In attending the most recent annual symposium on "Care of the Professional Voice" at the Juilliard School where scientists, MD's, speech therapists, surgeons, voice teachers, and professional singers and speakers interacted and exchanged ideas, it was sobering to observe the gulf that apparently still exists between the understanding of the scientists on one hand and the wealth of experience, pathologies, and practical knowhow of the therapists, teachers, speakers, and singers on the other. Many of the practitioners had difficulty relating to the scientific models which seemed to contribute only a modicum of understanding to their real needs as practising professionals. Questions such as 1) how does the musculature of the articulators and the surrounding regions effect changes in the perceived sound production, 2) what is the primary mechanism of sound generation and how does it appear in the model, 3) what physical and anatomical arrangements are characteristic or necessary for a good singer to possess, 4) what is a good list of do's and don'ts for a nose and throat surgeon to follow in order to minimize voice disruption in patients following surgery on the vocal tract, and 5) can different speech pathologies be determined or recognized from changes in speech quality, are but a small sampling of those I have heard raised and to which meaningful answers are sought. To these and other questions the scientific models often appeared to raise more questions than yielding understanding. Again the nagging question of why this should be so was ever present.

There are many experiences that each of us have had in listening to the nuances in speech and singing that beg for an explanation; immediate recognition of a speech change resulting from missing teeth or other dental work; speaking voices with a totally different character than those same voices in a singing mode; marked change in speech quality resulting from "dryness in the throat" or other pathologies, are but a few examples not satisfactorily explained by the source-filter paradigm. The question is why? What are the mechanisms involved?

The theory of hearing, a major area in itself, is intimately connected to the theory of speech production; one has only to note the strong parallelism in the respective models used for each to recognize this fact. Thus any new insight on the mechanism of speech production may also be important to hearing modeling work, and hence to speech recognition.

These and other factors have motivated, in one way or another, work in this area.

A SIMPLE OVERVIEW

At the risk of appearing overly pedantic, the following simple view of tract modeling is offered. Detailed modeling of sound production in the vocal tract forms the theoretical and analytic basis for almost all the speech processing systems in use and being designed. This work usually begins with measurement and analysis of the pressure wave which is received by a microphone placed some 3 - 15 cm outside the mouth. It is this pressure wave which is commonly referred to as sound or as the sound wave. It is the resulting microphone signal which is transmitted over the telephone network, or drives voice actuated equipment, or the pressure wave itself which propagates in the air to another human being as verbal communication. For additional information one can measure by various means (x-ray and optical) the motion of the more visible and active articulators (the tongue, lips, and jaw) and the fixed portion (the hard palate) to obtain a measure of the geometrical boundary of the tract at any instant of time. Good cross sectional shape information for the pharyngeal area is much more difficult to obtain. Coupled with high speed motion pictures of the vocal fold oscillation and the concomitant glottal time-varying area, one can postulate a fairly comprehensive model for the vocal tract. Using basic properties of the physics of fluids in essentially the steady state under small perturbations and restricting the analysis to one spatial dimension (namely transmission along the mid-sagittal line), one can write differential equations which describe the acoustic behavior in the absence of net air flow. This approach forms the starting point for almost all of the intense modeling work for the last forty years. Net air flow is taken into consideration only in the modeling of the vocal fold oscillation during phonation and in the modeling of the effects of the turbulence which occurs during frication.

The researchers who have done the detailed modeling have been primarily trained as electrical engineers who logically prefer to think of complex systems as networks with excitations and responses. The basic references on acoustics, virtually without exception, all use this analogous network or impedance approach to acoustic problems, it being the most versatile, essentially linear approach. The dictum of keeping the model as simple and as computable as possible dominates and is entirely consistent with the goal of development of models that can be implemented directly and also that give a measure of some understanding of the generation of the pressure wave we associate with speech. The early work of Wegel [14] and of Chiba and Kajiyama [1] is well known. This was followed by the notable work of Fant [4, 8, 10] and others on the source-filter theory of speech production. The sustained and cogent work of J.L. Flanagan [6, 7, 17, 25, 28, 31, 32, 34, 38, 39, 42, 43] and his colleagues at Bell Laboratories is the most complete example of vocal-cord/vocal-tract modeling incorporating a number of additional nonlinear fluid flow effects into the plane wave model that are missing in the more traditional source-filter model.

With this background let us first examine in some detail the traditional source-filter model of speech sound generation in the vocal tract.

THE SOURCE-FILTER MODEL

The source-filter theory of vocal tract operation has been around for nearly fifty years with most of the real development of the model occuring over the last twenty-five years [1-43]. The basic tenet of the theory is that the vocal tract can be modeled as a linear filter driven by a source (corresponding to the flow of air through the glottis) that is periodic for the voiced sounds and essentially random noise for the unvoiced sounds.

The basic assumptions underlying the source-filter model that permit the one-demensional wave equation representation are:

1. The fluid, air, is compressible obeying the gas law, PV^{γ} = constant.
2. Energy is propagated only by longitudinal motion of the fluid.
3. The fluid cannot support shear stresses nor propagate disturbances transverse to the main flow.
4. The fluid is assumed to be isotropic.

The equations relating flow and pressure as functions of time and distance along the tract from the glottis (the source) to the lips (the output) are set up by assuming adiabatic compression (the gas law), invoking Newton's law of motion in one dimension and conservation of mass (continuity). If the cross sections of the tract are assumed circular then these equations can be manipulated directly to yield the well-known Webster wave or Webster horn equations for pressure and volume velocity as functions of time and distance along the mid-line. Implicit in assumptions 2 and 3 above is that tract air flow during normal voicing (phonation) is laminar and has a velocity profile across any plane normal to the mid line that is uniform except very near the edges where it reduces rapidly to zero (the plane wave assumption). The pressure wave (sound) along the mid-line is assumed to be equal to the flow rate times the acoustic impedance at that point along the mid-line.

If the continuously varying circular cross sections are replaced by a finite sequence of constant diameter cylindrical sections then the continuous wave equation can be represented by a cascade of discrete equations closely resembling a ladder network in form. The elements (coefficients of the discrete equations) of this network are acoustic resistance, capacitance, and inductance whose values depend on the properties of the medium, air, and on the local cross sectional areas. Especially readable treatises on this model are Fant [8], Flanagan [25], and Rabiner and Schafer [44].

The model has been considerably extended through research on speech sounds to try to account for the observed output speech wave as measured by a microphone sensing changes in pressure in front of the mouth. Effects considered include compliance in the walls of the tract, interaction with the nasal cavity, radiation of sound from the walls and nose, frication induced noise in the vocal tract during unvoiced sounds resulting from turbulence generated in the flow in constricted areas, and Bernoulli effects (the nonlinear relationship between pressure and velocity along a streamline in steady, frictionless, incompressible flow) at the glottis.

A usual assumption in the source-filter model of the vocal tract is that the source and filter are essentially independent, i.e. there is negligible influence on the source flow wave by the "load" the vocal tract imposes. Flanagan [31] has studied this interaction and found it not always negligible; in fact the importance of this interaction was a primary motivating factor in his continuing extensive modeling work on the vocal fold oscillation mechanism and frication. There is currently a resurgence of interest in studying the various aspects of vocal tract loading and its effect on the vocal fold oscillation. Others, such as Stevens (discussion of [31], p. 16) feel vocal tract "loading" is only a second-order effect. Quasi-linearity also implies that during phonation any frequency component that appears in the output wave must have a corresponding component in the glottal wave; only the amplitude and the phase of the component can be changed by the vocal tract dynamics. (The exception to this is, of course, frication where the sound results from turbulence generated at the particular local constriction in the vocal tract; however, our discussion centers primarily on "voicing".) Slight nonlinearities in the tract can cause only a modest broadening of the spectral lines. As a result considerable effort has been expended to try to find the exact or true shape of the glottal flow wave [15, 17, 27-41] as this is the driving or input wave for the filter modeling the tract.

A number of excellent films [45-49] are available showing the graceful flowing vocal fold oscillations during phonation. Good two-dimensional information is available on glottal opening but not much is available in three dimensions where the third dimension is along the direction of flow. The assumption that is usually made here is that the vocal fold oscillation is of the relaxation type (derivable from Bernoulli's equation) and is driven by the lungs which, by means of the intercostal muscles, provide a source of constant pressure air in the subglottal region. Second, as far as the vocal tract goes, almost all the effort is spent on indirect ways of obtaining pseudo cross sectional areas as a function of distance along the tract supraglottally and in obtaining a good measure of the geometric shape of the vocal tract boundaries [18-23]. Third, the response of the vocal tract is measured solely by the pressure pickup at least 3 cm or more outside in front of the mouth. It is from these measurements that further refinements in the vocal tract model are prepared.

The best vocal tract models [43] with all their refinements are now so complicated that they require on the order of a hundred times real time to compute continuous output speech on a large modern general purpose computer. If special purpose array processors or vector machines are used then near real-time operation can be achieved. When mention is made of a possible inadequacy of the current model it is usually interpreted to mean that the best current model must be even further refined with an assumed corresponding increase in required computation. This is understandably met with resistance because the amount of computation-per-unit-time required by the current best model is already inordinately large making practical implementation not economically feasible at present. However this point should not dissuade one from seeking a clearer understanding of the precise mechanisms involved in the generation of the speech pressure wave.

The increasingly wide availability of computers and special purpose hardware devices, such as FFT boxes and array processors, have made the linear analysis of speech signals very easy to do. Unfortunately the source-filter model of the tract operation has also conditioned the kind of analysis that is performed on the speech wave; the methods currently in vogue include especially those that require the calculation of short time spectra. Emphasis is on spectral analysis and the concomitant averaging that is necessary to obtain a "meaningful spectrum". This has the effect of averaging out any very short term transients that occur within a single pitch period or differences between adjacent pitch periods. One should be very careful in using these tools to understand what kind of information is retained and what kind is eliminated. More on this point later.

To summarize briefly, the source-filter approach is essentially a linear system model with the primary burden of speech naturalness placed on the generation of a suitable glottal source wave and on coordinated control of the change in model parameters as articulation for connected speech proceeds. The synthetic speech generated by this model, while quite intelligible, is discernibly synthetic unless extreme care in transition adjustment is taken. The primary nonlinearities in the comprehensive model arise in trying to accommodate the nonlinear flow effects at the glottis (of the Bernoulli type) and at any other point in the tract where the cross section area is reduced (and the concomitant flow velocity increased) such that turbulence in the flow is generated; it is only at these points that the model assumes sound can be generated. What measurements there have been of flow in the tract have either been on rigid models of the glottis under steady state conditions or on the bulk volume flow out of the mouth by the indirect means of selectively measuring the pressure drop across a wire screen, the pneumotachograph approach [35], or on glottal flow inferred by graphical measurement of area from high-speed motion pictures and electroglottograph signals such as reported by Childers [49] and others. What seems not to have been done by the source-filter proponents in a thorough way is to actually measure the details of flow within the tract during voicing to substantiate or question the basic assumptions 2, 3, & 4.

Before some of the new results are described we choose first to try to develop a "mind set" by making a number of observations designed to stimulate both a broad and a detailed thinking about the sound production mechanism. For some of the observations the answers or explanations are seemingly straightforward; for others they are certainly not or are incomplete at best.

SOME OBSERVATIONS AND APPARENT ANOMALIES

For speech, the vocal-cord/vocal-tract model with all its improvements is felt by some to explain 90 to 95% of the observed phenomena; that is a rather high mark for any theory and model and is usually indicative that the central features of the physics are being taken into account. But what about the other 5 to 10% or more of the total effects? How do we explain them? Let us now look at some of these observations and so-called "anomalies".

1. In observing the time waveforms of continuous speech, why is the waveform not quasi-symmetric between plus and minus excursions? One can invariably tell when a speech wave is inverted. One also notes that as pitch is varied for voiced sounds, the general wave shape remains relatively fixed; only its terminating point is extended or contracted. How do we explain this in terms of the harmonic theory of periodically driven linear systems?
2. It is known that only a small fraction of the energy in the air flow through the glottis is converted into sound or acoustic energy during speech; see for example Bouhuys et al [50] and Titze [51]. Using the figures of Bouhuys the power passed through the glottis into the vocal tract is of the order of 0.1 to 0.5 watts during singing at moderate levels while the measured acoustic power out is only of the order of 0.1% to 1% of that amount; and singing tends to be more efficient than speaking in producing acoustic output. Where has the rest of the energy gone (the 99+%) and what is the mechanism of connversion into acoustic energy? How do the trained opera singers produce such intense and controlled sound, enough to fill the Met, with pairs of lungs not much different than our own? What is the mechanism of their more efficient production of sound? What is the explanation of "belting", i.e. the production of very high intensity vocal sounds and how is it accomplished without doing damage to the vocal folds?
3. In source-filter theory the LC product of the so-called acoustic inductance and capacitance is simply the reciprocal of the square of the velocity of sound in the medium. Therefore formant resonant frequencies in such a system should scale in proportion to the velocity of sound in the medium used. Why then when lighter gases such as helium are breathed do not the formant frequencies all move up in proportion to the sound velocity in the new medium? For pure helium the velocity of sound is approximately three times that of air yet the formant frequencies only move up less than a factor of two, and this in an irregular fashion, i.e. not all formants change by the same factor.
4. Other living creatures can produce sounds we clearly recognize as speech-like sounds [52-54]. The trained mynah bird for example can reproduce a number of speech phrases difficult to distinguish from those of its trainer. It does this with a breathing apparatus that is much smaller and quite different from that of the human and with an ear also very much different and much much smaller than our own by more than an order of magnitude! Also its brain is not much larger than a large pea! Klatt [52] has looked at parts of this problem in some detail but his explanations leave many unanswered questions. What is the mechanism of the mynah bird's sound production and why does it work so well with such a gross difference in physical size, arrangement, and control with respect to the human? What does this

example also say about our complex models of human hearing which require a well-developed cochlea and a considerable brain for speech recognition?

5. Human beings can make a wide variety of sounds with their articulatory apparatus such as pops, clicks, whistles, groans, numerous imitations of familiar musical instruments and animals as well as many others; for interesting and light reading see Newman's fascinating little paperback [55]. Many of these sounds are non-glottal in origin and hence do not fit in with the regular source-filter model. What is the mechanism of production of the audible acoustic wave and do not similar conditions also exist during speech production?
6. It is well known that speech can be passed through a nonlinear no-memory element which considerably distorts the speech wave but affects very little the intelligibility in the speech. See Licklider [56-58] on clipped speech and clipped differentiated speech. This nonlinear action changes considerably the balance in the speech spectra. Also voiced utterances may be considerably shortened by removing repetitive pitch periods without markedly affecting the intelligibility. This action removes the semblance of the steady state in the utterances; only the primarily transient behavior remains. Why is this so and how does one reconcile it with the source-filter model which is basically a steady-state model or with speech recognizers which depend on a minimal set of spectral templates?
7. Speech quality is affected markedly by "dry throats", nasal drip, missing teeth, excess saliva, etc. What is the mechanism of the effect? Why can we not speak or phonate well with dry cracker crumbs inside the mouth if the source of sound is at the glottis? Surface and space effects seem especially critical yet are apparently not easily accounted for in the model.
8. For sustained phonation of a fixed vowel how do we reconcile the supposed sharp corner in the glottal flow wave on closure that is "required" in order to generate the necessary high frequency energy to excite the higher formants, with the gracefully smooth, and visually observed, oscillation of the glottis? Is this more a property of the particular model than of reality? Nature seems to thoroughly abhor discontinuities.

One has only to read the discussion and conclusion sections of papers to realize that many serious and competent investigators, too, have realized there are many observable phenomena not completely or properly accounted for by the current vocal-tract paradigms. But where should one look for possible answers? What measurements are lacking?

It is the author's thesis that the first, and primary, place to look is inside the tract itself to see just what the fluid, air, is doing during speech production. It is in this medium that pressure waves and flow are created and transmitted to the outside world as sound. Virtually all previous work has concentrated on measuring the output sound, (the external pressure wave at the diaphragm of a microphone outside the mouth), the approximate shape of the tract boundaries, and, by secondary means, what the glottal source flow wave must have been to produce such an output sound with the known tract configuration. The primary burden has thus historically been placed at the determination of a suitable glottal source wave. Let us now look at some of what is known about flows in confined spaces.

SOME NEW RESULTS

While many investigators have looked in considerable detail at flow through the glottis, the means have been indirect. Either pressure measurements are made on fixed cross section models of the glottis and the flow inferred, such as [27], or area measurements are made

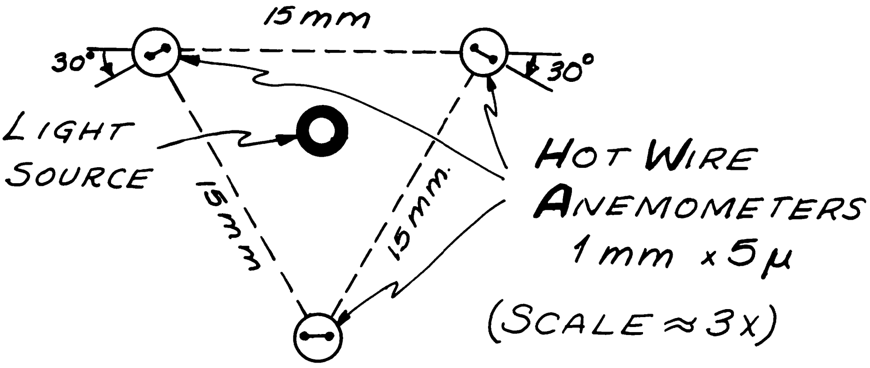


FIGURE 31-1
Hot Wire Sensor Array.

from video camera frames, or high speed movie frames, of the actual vocal fold oscillation and, with assumptions about transglottal pressure drop, a flow wave can be calculated [49]. Broad's 1979 tutorial [41] gives a very good summary of what is known about the vocal fold oscillation mechanism.

The first direct and detailed flow measurements of any consequence on the vocal tract and known to this author are those of Teager. He reported in a short corrrespondence item [59] on detailed hot-wire anemometer measurements made during the phonation of the vowel "I". His observations were rather startling, to say the least, and were initially met with an effort to disprove their veracity by faulting his experimental technique and claiming that if several "adjustments" were made in the data reduction to account for a postulated bias then it could be shown that the data really supported the source-filter model rather than being at odds with it. A visit to Teager's laboratory to observe his refined experimental technique and discuss further his findings convinced me that these results were real, reproducible, and only a small part of a much larger body of experimental data he had already obtained but not yet reported in the literature.

How were the measurements made and just what were these observations of his? The air flow measurements were made with an array of three hot-wire anemometers (5μ x 1 mm) arranged at the corners of an equilateral triangle 15 mm on centers in a vertical plane; the top two hot wires were inclined at ±30° from the horizontal and the bottom hot wire was horizontal. The entire array was positioned in a vertical plane about 10 mm in front of the cusp of the tongue and 10 mm below the hard palate. See Figure 31-1 for a sketch of the array. The ± 30° inclination of the two top hot wires was for the purpose of detecting differences in the right and left flows and helping resolve the differences due to any possible cross flow components. The hot-wire sensor bandwidth including the electronics was 0—5 kHz with calibration performed both before and after data taking (details are given in [76]). During the data taking while the vowel "I" (as in hit) was sustained, the entire array was slowly moved from left to right across the tract to detect if there were any major flow differences left to right. The published figures indicated that there were no appreciable left-right differences. Representative data from Teager [59], his Figure 1, is included here as Figure 31-2. The significant observations were, quoting from Teager [59], as follows:

1. None of the air flows or their first derivatives (acceleration components) are in

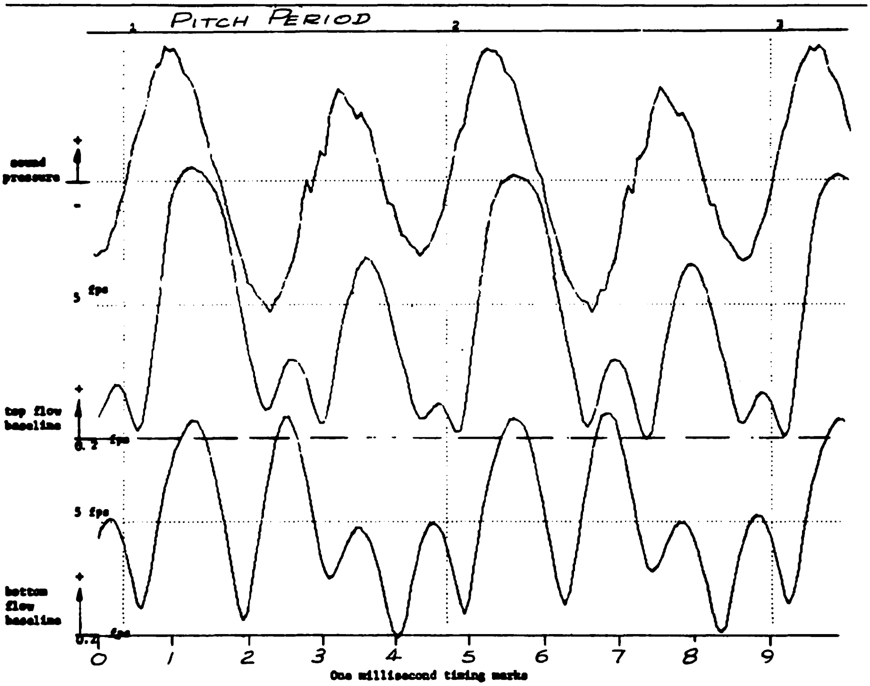


FIGURE 31-2
Representative simultaneous normalized sound and air flow for the vowel "I". Top trace: sound pressure. Middle trace: air flow velocity measured by anemometers at top rear of front oral cavity. Bottom trace: air flow velocity measured at bottom rear of front oral cavity. (From H.M. Teager, "Some observations on oral air flow during phonation," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-28, No. 5, 599-601, 1980.)

phase with the sound (dynamic pressure as measured with an electret microphone in the array); all three (sound, upper and lower flows) have formant structures which are visibly and measurably different.

2. Air flow differences between top and bottom sensors were manifestly too large to be consistent with either laminar flow or a passive acoustic filter.
3. The difference between these traces is consistent with a separated flow, specifically a sheet flow which alternates between palate and tongue at about the first formant rate of approximately 500 Hz.
4. Since, from basic physics, sound is produced from any rate of change of vector flow, it would appear that sound is (also) being actively generated within the front mouth cavity in this experiment.

These results for sustained voicing of the vowel "I" immediately indicate that flow within the vocal tract is anything but simple. The statement that none of the air flows or their first time derivatives are in phase with the sound (pressure wave) and that the air flows are *all* visibly and measurably different from the sound implies that the air flow is not behaving isotropically, i.e. uniformly across any cross section. If this is the case then the widely used concept of acoustic impedance, which depends heavily on the validity of the isotropic assumption, is open to question. With different flow rates present in different parts of the cross section which value is to be used in the calculation of acoustic impedance? This view is at first hard to accept in view of the so-called 90 + % efficacy of the source-filter theory and its modifications.

It might easily be argued that these are results that were obtained only in one region in the tract and hence are not representative of the flow as a whole; they could be just an aberration or second-order effect. It could also be suggested that these results were observed considerably far downstream from the glottis, the presumed sound source, and hence should have little effect on the actual sound.

But let us look more carefully and closely at what these results are indicating. First, the consistency of Teager's data over all of the 60 pitch periods shown is strongly supportive of his observations. If the effect were only an aberration then there should certainly be noticeable variation in sensor signal shape as the sensor array was moved about in the vertical plane normal to the mid-sagittal line, but this was not the case. Also the correlation of the first time derivative of the air flow with the sound should certainly have been much higher than was observed. The difference in the flow waves between top and bottom sensors, while periodic, is not consistent with the assumption of a single quasi-laminar flow region but strongly indicative of *separated flow* in the forward part of the mouth, as Teager suggests.

Separated flow implies that there are distinct regions of different flow behavior and that the flow is rotational, i.e. vortices are present [60-62]. Figure 31-3 is a sketch of air flow in the vocal fold region showing separated flow, the attachment region, and the concomitant vortices in the sinus Morgagni regions. Although with separated flow the flow velocity profile is strongly asymmetrical, as the sketch indicates, the pressure profile shows far less variation, there being little difference between the pressure in the sheet flow and the pressure in the adjacent quasi-static flow region.

The source-filter theory on the other hand implicitly assumes irrotational flow in the tract, that is that vortices are not possible or present. The possible presence of separated flow and of vortices is not really a new idea. For example both van den Berg [27] and Broad [41] mention the possibility of vortices or eddies being present just above the glottis prior to the "attachment" of the glottal wave to the pharynx walls. They both assume a single glottal wave after attachment. Broad appears to assume the attachment takes place almost immediately above the glottis; he does not give any evidence to support this reattachment assumption. Teager's results tend to indicate the contrary obtains; reattachment is complex and varied, and separated flow continues through to the front part of the mouth where it is more easily observed by direct flow measurement techniques. Vennard [63] in his classic treatise on singing also notes that vortices are felt to exist in the tract; vortex is even an entry in his index!

Consulting the classic references on acoustics such as Morse [64] and Morse and Ingard [65] one can find almost no information on the role of flow in acoustics; total attention seems to be on the acoustic impedance concept with its implicit assumption of irrotational flow. Pierce [66] does have some comments and description on vorticity but in no way relates it to the human vocal tract. Yet when one consults the fluid dynamics literature, separated flow, jets, and vortices seem to abound in nearly every practical flow situation in one form or another. Nice laminar, well-behaved flows seem only to exist as textbook problems or as the result of very careful design and very steady flow conditions. One also notes there appear to be two schools of people in fluid dynamics, the experimentalists and the theoreticians; the experimentalists are the ones who carefully observe the dynamic behavior of flows in practical physical situations, conduct experiments to elucidate and clarify the observations, and who then design apparatus based on this broad experience and well established

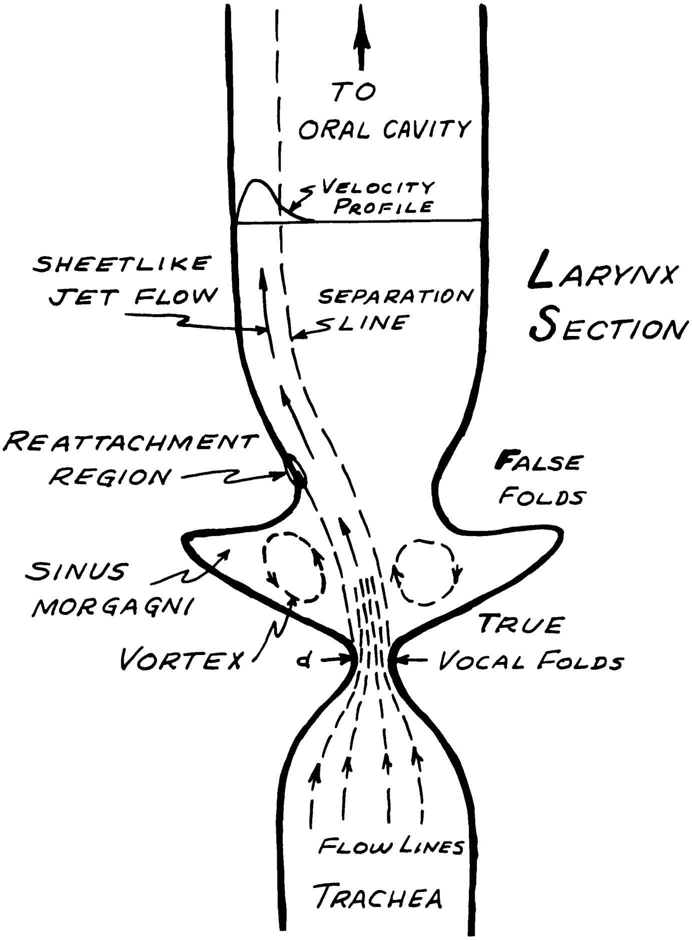


FIGURE 31-3.

principles; the theoreticians on the other hand are those who seem more interested only in solving the differential equations that purport to describe fluid flow in limited idealized situations; to verify their mathematically obtained results with suitably constructed experiments seems to be the exception rather than the rule.

In looking through the experimental fluid dynamics literature that can be related directly to the vocal tract problem, one finds a paucity of material. One name that does reappear several times is that of the Frenchman H.P.M Bouasse. He was an excellent experimental physicist who was both a superb teacher and a prolific writer. Bouasse authored more than fifty complete textbooks on physics and acoustics until his death in 1953. These works, in French, are extensive collections of experimental results and are fascinating to read. Unfortunately they are also rather difficult to obtain; the best holdings in the US are located at M.I.T. and the Library of Congress. His work on wind musical instruments [67] is especially interesting;

it points out the extreme importance of detailed and careful physical experimentation in the study of fluid-acoustic effects. The complexity of the interactions and the reluctance of the physical situations to yield to neat mathematical treatment is illustrated over and over again in his countless examples. His works on pipes and resonators [68], on jets and channels [69], on vortices [70], and on acoustics in general [71] merit careful reading. The complexities of flow separation and attachment are also found in the extensive literature on fluidics which employs the controlled jet as a means of power amplification [72]. Coanda, a Rumanian engineer who pioneered in aircraft and nozzle design from the 1930's through into the 70's, utilized the "wall attachment" effect, later named after him, in the design of many devices [73-75]. He has over 30 U.S. patents on many of these designs. Thus separated and attached flows are very real and are to be found in almost every physical situation involving moving fluids to some degree.

But what further experimental results have been reported that relate to these new flow ideas? Teager reported in more detail his experiments and suppositions at the Vocal Fold Physiology Conference at Madison, Wisconsin, June 1981 [76]. He further detailed his findings of separated flow, reattachment, and vortices in the vocal tract. He stated that he believes the basic mechanism is that of jet-cavity interaction and that "sound is generated throughout the vocal tract from the lobes of the lungs to the area beyond the lips, acting as a nonlinear coupled regenerative system". He again asked for experimental verification of refutation of these results by others. The very next paper at the conference by Gauffin et al [77] described in detail flow studies made on a fixed-opening model of the glottis; they were interested in the steady-state velocity profiles and pressure drops across the vocal folds as a function of glottal width and contour. Their model was scaled up by a factor of five (5) primarily to facilitate accurate pressure measurement along the fixed vocal fold contour. (This study and the associated technique were strongly reminiscent of the earlier work of van den Berg [27].) At the close of the paper Gauffin showed five 35 mm slides of the flow through their model under different conditions employing the Schlieren technique for flow visualization. All five of the slides showed very definite flow separation at the upper edge of the vocal folds, jet formation and subsequent attachment of the jet to only one wall of the tract (not simultaneously to both walls as Broad assumes), and the beginnings of vortex formation at the separation boundary! Figure 31-3, previously referred to, is also a reasonably accurate sketch of the general character of their obtained steady-state results. Even though there is some problem about the accuracy of simulation by a scaled-up model in such aerodynamic phenomena, these results were all in line with Teager's predictions and theory.

In an earlier talk by Proctor on control of breath [78], a film was shown of a long sequence of x-ray images of a side view of the head and neck of a person both speaking and singing. During the singing portion of the film one could see strong evidence of the presence of an intense flow along the roof of the mouth from the soft to the hard palate. The pulsatile flow was sufficiently intense to appear to cause the velum to close off the entrance to the nasal cavity and to direct the upward flow from the glottis through a 90 degree turn to flow along the soft and hard palates. The velum took the posture of an almost perfect quarter circle; one could virtually see it quiver in this position as the singing continued. Proctor did not draw attention to this observation.

In an ad-hoc session on flow in the vocal tract at that same conference E. Muller, of the University of Wisconsin, showed a number of slides of figures taken from flow studies in the respiration literature [79]. These included excerpts from Pedley, Sudlow and Schroter [80] which show separated flow and secondary flows are the rule rather than the exception in flow behavior in the lungs. See also [81]. Thus the air supply to the glottis is a complex flow with considerable secondary (normal to the primary flow) flow structure. The "necking down" of the flow subglottaly tends to smooth out (relaminarize) some of this complex flow structure but it is definitely there. In fact, were it not for the secondary flows the lungs would not be able to carry out there gas exchange function as efficiently as they do; secondary flows are critical to proper lung functioning. Following the showing of the slides considerable discussion ensued as to the presence of separated flow in the tract and its importance in speech production. For one reason or another many there still felt unconvinced of the role separated flow plays in sound generation; retreat was usually back to the more familiar source-filter paradigm.

The presence of attached sheet flows of considerable velocity in the mouth is further evidenced by the common, and somewhat embarrassing, experience of ejecting small droplets of saliva in the breath stream during speech generation. In this case the excess saliva has literally been torn from the surface by the local attached flow. It is also the interaction of the attached flow with any loose matter in the mouth that makes it very difficult to speak while eating crackers for example.

But the increasing amount of experimental evidence supporting the jet-cavity interaction paradigm and other related sources of flow noise (Ffowcs Williams [82]) is a fact that must be considered; it can no longer be neglected or passed off as "anecdotal". We now take the liberty of proposing the beginnings of a new approach to the theory of vocal tract operation based on the ideas of jet-cavity flow interaction and then reexamine some of the previous anomalies in light of this theory.

THE NEW THEORY

The basic thesis is that speech sounds are produced in the human by the complex interaction of sheetlike jet flows throughout the tract, that articulation therefore consists of conscious and learned control and direction of the glottal flow by the tract musculature, and that tract surface conditions do markedly affect the behavior of the attached flows. The zero formant, or pitch wave, for the voiced sounds is generated at the glottis but much of the fine structure of the higher formants are the result of the sheet-flow and cavity interactions. The mechanisms involved in sound production are necessarily robust in their form; there is usually no single, one and only way to produce a given sound. Nor is there only a single place where the sound can be generated in the tract (i.e. only at constrictions). The trained singer is well aware of this fact; the singer learns early where to "place" the sound generation for different effects.

Acoustic waves are generated during momentum exchange when flow waves intersect and thus interact. These same pressure waves are guided by the separation lines and are deflected by vortices in their path. In fact one method NASA [83] has used to detect the presence of an area of high vorticity in clear air is to bounce an acoustic wave off the area and measure the strength of the return. Fabrikant [84, 85] has found that the interaction of a sound wave with the synchronous particles of a rotating flow (cylindrical vortex) can induce sound *amplification*. Thus a knowledge of the flow, sparation lines, and the local vorticity, enables one to better locate the active region of sound generation and to estimate an effective radiating cross section.

That the mechanism is geometrically complex one has only to observe the wide variation in the posturing of one's articulation apparatus (the tongue, jaws, teeth, cheeks, and pharynx musculature) to generate sounds that are perceived as the same. People themselves differ very widely as to the shapes of the jaws, the hard palate, the teeth arrangement, the jaw alignment, the size and coverage of the lips, as well as the neural control of the many muscles that make up that part of the head; all have an effect on the quality of perceived speech as all can affect the nature of the air flow in one way or another. This is especially true of the musculature of the lower pharyngeal region. It is the professional speaker, the actor, the trained singer, the mimic, and the ventriloquist who are most aware of the broad range of flow control possibilities and the most adept at their control.

The medium transmitting the sound or pressure waves is the air in the tract and the consequent wave transmission follows directly along the lines of air flow and is guided to a considerable extent by the separation lines. Thus the nature of the flows must be known in some detail. Since the tract is geometrically rather complicated and can be changed rapidly in a nearly infinite variety of ways, it is necessary to develop a qualitative idea of the range of flow patterns that are generated when we phonate. Teager maintains [86] that each voiced phoneme has its own general characteristic flow pattern or type. Associated with each pattern is a time scale for the building up of the quasi-stable oscillation and a second time scale associated with the quenching or decay of the same. The flow pattern usually builds up in as little as 3-4 pitch periods with the detailed structure developing each pitch period. Some of the oscillations are of a regenerative type with additional energy being added or pumped in on each successive pitch period. The dynamics of these oscillations are also a function of the local geometry and its physical dimensions.

In [86] Teager makes some interesting observations when he gathers together the results from the analyses of the many different phonemes. He notes that at least four different situations obtain which relate a flow frequency measured well inside the tract (mouth) to a corresponding pressure (sound) frequency outside the mouth: 1) frequency f inside with no trace of f outside, i.e. an effective zero, 2) a frequency f outside but no corresponding f inside, 3) frequency f inside with frequency $2f$ outside, i.e. frequency doubling, and 4) frequency f inside and frequency $f/2$ outside, a frequency halving. He maintains that each of these phenomena can be explained in terms of specific jet-vortex interactions.

Teager also notes [86] that the pressure within the tract is very nearly uniform over any cross section normal to the predominant flow or mid-sagittal line. Hence he suggests that one must talk about energy in the flow rather than the pressure. Other than the ambient pressure and to a first order approximation a jet does not know what is going on in the space outside itself and vice-versa (e.g. see [60] pp. 171 ff); this also attests to the reasonableness of the approximate zero net pressure difference across any cross section. However the flow is far from being uniform across the cross section.

The pulsatile flow that develops at the vocal folds is of the general Bernoulli type with due account being given to the effects of separation on the exit flow and to the loading of the downstream portion of the tract. The extensive previous work on the vocal fold oscillation mechanism has taken the Bernoulli effect into account in the glottal opening but has not considered the effects of the supraglottal separated flow. The cavities just above the glottis and below the false folds on each side (the sinus and ventricle of Morgagni) tend to act as two chambers which can store flow energy during one part of the flow cycle and return that energy to the system in another part of the cycle; in one sense they each behave like a simple policeman's whistle without the little ball; very small changes in energy in the chambers can produce much larger changes in the main flow wave or waves in the form of a modulation. Teager has included an analysis of this effect for the simple whistle in an appendix in [76]. This effect is much more than a simple Helmholtz resonator; it is well known in fluidics and is the basic principle underlying the design of the fluid amplifiers [87-89].

Because the vocal tract is not ideally symmetric (between right and left sides), the attachment of the separated glottal jet to the pharynx wall may favor one side over the other or may alternate sides (giving a pitch-period waveform having an ABABAB form i.e. the odd-numbered pitch-period waveforms, A, or the even-numbered pitch-period waveforms, B, are more alike than adjacent pitch-period waveforms). Pathologies such as swellings or surface irritation in this area can thus have a marked audible effect on the sound by altering the asymmetries and thus changing the flow patterns.

It is interesting to draw a distinction between the two primary modes of using the vocal apparatus, *talking* and *singing*. *Talking* is characterized by a nearly free running pitch with the brain spending most all of its effort in simultaneously devising the word string and generating the requisite commands for the primary nerve signals to the articulatory musculature; a quasi-constant lung pressure, from intercostal muscle activity, yields a fairly constant pitch for the voiced sounds consistent with minor variations resulting from transient changes in the tract loading. These changes in loading are the result of changes in tract geometry and volume caused primarily by tongue and jaw movement. The pitch varies more widely when either more conscious attention is given to it or when lung pressures increases in an effort to speak words at a higher rate as when under tension. The division of time spent in talking among the three basic speech types, voiced, unvoiced, and silence, is crudely 60%, 25%, 15%. The voiced segments can be shortened considerably and still retain a very high level of intelligibility.

On the other hand *singing* is characterized by extended voicing. There is nothing more boring to the listener than a repeated time-invariant periodic wave as these sounds very rarely occur in nature. To hold the listener's interest the singer then must provide variation by exerting strong concentration on the voicing with careful attention to pitch, loudness, vibrato, and timbre. This is accomplished by proper accurate posturing and alignment of the trachea, larynx, and pharynx and also by the articulators, the tongue, jaws, and lips (as they define boundary shape, vocal tract volume, and the mouth opening shape). Proper control of the chest muscles to finely adjust the lung pressure differences and air flow rates is also essential. During performance the singer is usually voicing for 90-95% of the time. It is only during this other 5-10% of the time that the singer can perform the stops and plosives in order to make intelligible the sung words; this often requires over-accentuating the unvoiced sounds so that may be recognized. When viewed this way speech is a highly transient phenomena and singing is more or less a steadier state form of the voicing.

The source-filter paradigm gives only a broad outline of the general vocal tract shape that could yield a desired voiced sound, there being no unique single shape for each voiced sound. The same paradigm gives little hint as to what particular vocal tract shapes or arrangements are desirable from either a speaking or singing point of view especially in the pharyngeal region. On the other hand the jet-cavity interaction paradigm brings the 3-dimensional aspects of the tract anatomy into direct focus. The singer learns early through careful investigation and trial and error how to manipulate almost all of the musculature in the laryngeal and pharyngeal areas and what the effects are; in short the singer discovers how to hold the "nozzle" (the glottis) and how to position it for the desired effect. An interesting comment I have uniformly heard from a number of accomplished singers is that the better singers tend to have short necks rather than long necks; viewed from a flow point-of-view, concomitant with a short neck is a more favorable aspect size ratio to the pharynx region permitting more effective control of the glottal 'jet' and its tilting. It is very enlightening to reread the better books on singing technique and to analyze the description of the many details of the art of singing from the jet-cavity paradigm point-of-view. It gives a rather different and more understandable way to view the mechanics and physics of the situation. Sorting fact from fiction is still a difficult problem.

Cavagna and Margaria [90] noted that mean airflow through the glottis was relatively higher at low sound levels than it was at high sound levels. One partial explanation is that for the higher sound levels it could be the result of sizable vortex generation in the tract essentially building up to the point where the net tract flow is pinched down and thus reduced. It is a mechanism such as this that is felt to give opera singers their power and more efficient sound production; this improvement in efficiency is often by a factor of 10 or more. To facilitate a maximum coupling of the pressure wave to the outside the singer is often seen to distort the shape of his or her mouth accordingly.

There are cautions however as with any new theory. The Hamming-ism "Beware of finding what you're looking for" and the John Pierce quote "More things are known than are true" are especially appropriate. But even if we don't fully see and understand all the details of the fluid flow interactions at this time, we can observe some of their effects in the sound wave itself outside the tract.

The sound or pressure waves that are generated as the vortices and flows interact are inherently time synchronized with the bulk glottal flow wave which is the primary source of driving energy for the entire system. Thus apparent short bursts of higher frequency energy (the higher formants) can occur and usually do at about the same spot in each pitch period and they do depend also on the local physical geometry. Relative timing of these energy events then becomes critical and also a means for improving the discrimination between the different voiced phonemes. The power spectrum method is not too well suited to measuring this information because of its insensitivity to phase information; application results in effectively smearing out the detailed timing information and a loss of definition. What is needed is a better, more robust way to detect and track these rapidly appearing and disappearing frequency components, a way that is relatively insensitive to nonlinear-no-memory transformations of the amplitude scale.

A possible analysis technique that shows some promise of yielding the desired information is the application of the Wigner transformation. The set of three papers by Claasen and Mecklenbrauker [91-93] and the paper of Flandrin and Escudie [94] give the mathematical background for the technique along with many of its desirable properties. Chester [95] has applied the technique to the speech analysis problem with some measure of success. He points out that the Wigner distribution is more adept at analyzing transient type signals and displaying their properties in both the time and frequency domains than the more widely used Fourier transformation which is basically a steady state signal analysis tool. The Wigner distribution method also merits further attention to study the computational tradeoffs involved in its evaluation.

While the fluid dynamics literature is vast, most of it is concerned with the steady-state situation and with simple geometries to facilitate doing the mathematics. The increasingly wide availability of the less expensive minicomputers, special purpose computers, and VLSI circuits and an increased understanding of the relevant mathematics and effective computation algorithms is making possible the effective simulation of quite complex geometries. However, the material available on pulsatile flow and vortex interactions is rather small. The subject is beginning to receive increasing interest although the problems are very difficult mathematically [96-101]. The field is still very much one involving physical experimentation.

An area of physics that offers some means for explaining or at least analyzing the flow wave interactions with momentum transfer and conservation is that of "soliton waves". A number of good texts on the subject are just now becoming available [102-108] but this means of analysis is outside the scope of this paper.

OTHER OBSERVATIONS AND COMMENTS ON THE ANOMALIES

The non-symmetry between plus and minus values in the speech wave is a direct result of the non-zero net flow in the vocal tract; i.e. the glottal flow wave for expirational speech is essentially everywhere non-negative. This fact, in conjunction with the nonlinear pressure-flow relationships for separated flows, gives rise to the asymmetries observed.

The central difference between a speech wave produced by a loudspeaker and one produced by a human being is that in the latter case a net positive air flow through the tract is required as the source of energy for the flow-to-acoustic-wave conversion. That source-filter theory works as well as it does results from the fact that the tract may also be viewed as being excited by an ideal periodic flow wave that is planar and laminar and has an average velocity that is small compared to the velocity of sound; the theory of acoustic impedance can then be applied with many of its subtleties. This is the situation for example in the effective operation of the artificial larynx; the poorer quality in the sound results primarily from the difficulty in producing a good approximation to the glottal wave by this indirect means.

That the formant frequencies do not scale in direct proportion to the velocity of sound of the fluid present as the fluid is changed indicates that some of the assumptions underlying the standard acoustic impedance and the concomitant potential theory of flow are not being met. Teager's jet-cavity interaction paradigm [76] suggests that the formant frequencies should come closer to scaling as the square root of the sound velocity for the fluid present, i.e. for helium speech the formants could be expected to change by a factor of 1.7 rather than 3; he observed this in his experiments on the policeman's whistle when directly driven with helium. It is also interesting to reexamine Wathen-Dunn and Michaels' paper on the effects of gas density on speech production [109]; the data given in their Figure 5 fit the square root relationship much better than the linear relationship. Their detailed and careful experimentation covered a nearly 2-to-1 change in sound velocity. Thus the difficulties are traceable to the ignoring of the presence of separated flow; instead one needs to look at the dynamics of jets and their deflection and dispersion characteristics. The helium jet issuing forth from the glottis during voicing is much broader than the corresponding air jet and it also has a higher velocity. Also a larger fraction of the helium flow appears to go through the nasal cavity than for voicing with air. That the different formants move in an irregular way can be attributed to the dynamics of the different flow interactions they stimulate.

The mechanism of sound production utilized by the mynah bird is believed to consist primarily of a well controlled flow interaction in the region just above the syrinx. Perhaps this much simpler "vocal cavity" situation can form the basis for the beginnings of a meaningful mathematical model of flow interaction which could be simulated.

Much of the explanation of the myriad of non-speech-like sounds human beings can make with their articulatory apparatus can be traced to interesting flow interactions in various places within the vocal tract. Jets, vortices, and trubulence are felt to play the significant roles.

That speech can be passed through a no-memory nonlinear element with very little change in intelligibility can be viewed as additional evidence that it is not the detailed shape of the spectra of speech that is important but that it is the relative timing of the different flow interaction events and their pressure wave manifestations, the instantaneous formant frequencies, that are important for recognizing or distinguishing different phonemes. It is therefore necessary to track carefully within individual pitch periods the appearance and disappearance of energy as represented by the formants. This structure is re-established every pitch period for the voiced sounds. Slight perturbations in the position within the pitch period where the formant appears can cause the apparent appearance of a "zero" or "pole" in the spectra formed from an average over several pitch periods. This can help explain the so-called appearance and disappearance of the higher formants in the spectra with time although the formants are audible throughout. Teager [86] illustrates this effect with experimental data. The concept of "poles" and "zeros" can be viewed as more a property of the linear analysis tool, the DFT or FFT with its concomitant averaging, rather than the physical process itself.

Other evidence that tends to point in this same direction is the nature of the residuals in the LPC representation of the voiced segments [110]. While no definite pattern is seen to exist, the detailed residual wave shapes tend to suggest that there are multiple excitations occurring during each pitch period.

Some work has been done on jet-cavity interactions from a conservation of linear momentum point of view. The beginnings of a theory have been advanced by Elder [111-113] in his study of sound production in organ pipes. Here the pipe forms the primary cavity. This work goes beyond that of Helmholtz and is being continued by Elder who appears to be aware of some of the work of Bouasse.

The jet-cavity interaction theory of sound production in the vocal tract suggests that the articulators, especially the tongue, play a key role in both jet deflection and cavity resonance dynamics. Hence the interesting work by Nelson on minimal energy expenditure in the motor control processes of articulation [114] is a very meaningful and realistic approach to the actual situation. One needs now to see how the actual flow patterns are changed by the movements of the articulators, especially the tongue, during speaking and how the various cavity volumes are also changed by these same articulators. The ideas are complementary.

To facilitate further investigations into jet-cavity flow interactions it is absolutely necessary to be able to measure flow dynamically with some precision and flexibility. We now look briefly at one way this can be done.

FLOW MEASUREMENT TECHNIQUE

In order to obtain the wide bandwidth information on flow with a minimum disturbance to the flow the sensors must be very small and sensitive. The hot-wire anemometer fills these requirments very nicely. There is a large body of experimental technique that has been built up in both the aeronautical and the fluid mechanics areas [115-122]. With care, bandwidths up to 50 kHz or better can be obtained and errors below 5% readily achieved. The major problems come from the difficulties in positioning the probes inside the vocal tract, in protecting them from damage or destruction by accidentally touching the palate, tongue, or teeth, and in the prevention of a buildup of surface contamination on the hot-wires themselves. The sensors should therefore be protected by some form of shrouding. Since the hot-wire output is primarily proportional to the square root of the flow rate and since contamination of the hot-wire surface (as minute as it is) also affects its sensitivity, a comprehensive calibration procedure both before and after each test run is an essential part of the experimental procedure. This procedure should include calibration for both steady flows and for periodically, usually sinusoidally, varying flows.

Since the hot-wire anemometer has a cosine-squared sensitivity to flows coming in at an angle to the normal to the hot-wire, an array of two or three hot-wires must be used if direction information is also desired. The major manufacturers of precision hot-wire systems make both single and multiple wire probes in a variety of configurations.

SOME RECOMMENDATIONS

The jet-cavity interaction theory of sound production in the tract offers an interesting new approach to understanding the details of speech production. To develop further understanding of these new ideas much more experience is needed by the research community as a whole. Several activities which would be very meaningful at this point are:

1. Careful and detail flow measurements should be made to independently corroborate or disprove the published results of Teager [59, 76, 86] on a limited range of subjects, male and female, representative of the extremes of talker types.
2. To devise a mathematical way to represent the flow interaction and acoustic wave generation resulting from momentum interchange, simple single and double flow interactions should be studied with physical models. It may prove useful to augment this study by also using liquids where flow visualization is more easily accomplished. These studies should also include measurements on acoustic wave propagation both downstream and upstream in a developing flow as observations indicate that attenuation rate is different for the two directions.
3. A simple mechanical model of the mynah bird lung, syrinx, trachea, and mouth system would permit study of vowel generation with a view toward both mathematical algorithm development for computer speech synthesis and the construction of a simple speaking machine.
4. A critical reading and review of the comprehensive work of von Kempelen [123] on the construction of a speaking machine and the critical review of the treatises of Bouasse on acoustics, resonators, jets, and vortices should greatly increase our understanding of the critical flow issues.
5. Conducting a careful review of the speech disorder and dysfunction literature [124-128] and on the tract physiology itself [129-133] should help us see what new light the jet-cavity interaction theory sheds on the explanation of the pathologies, on their detection, and on their remediation.

These suggestions for further research possibilities are only a few of many that come to mind. Each reader, with his or her own particular range of experience and expertise in the speech area, should see additional possibilities for work that the new jet-cavity interaction idea suggests.

Other references which this author found useful included [134-148].

CONCLUSIONS

Separated flows in the forms of jets and vortices are argued to exist in the vocal tract during voicing. The interaction of these separated flows, i.e. the jet-cavity flow paradigm, is felt to be largely responsible for much of the fine structure of sound generation within the tract; this sound is in addition to the basic bulk glottal flow wave of pitch frequency. Experimental evidence of the existence of separated flow is advanced and a number of implications of this phenomena are discussed. The jet-cavity flow interaction paradigm removes an unnecessarily great burden from the glottal flow wave as having to provide all the higher frequency components found in each of the formants. Instead sound can be generated wherever flows interact with boundaries or other flows. The new approach also points out difficulties inherent in the use of the conventional spectral analysis methods, which are inherently linear tools, to analyze the output fine structure from essentially a nonlinear distributed system where modulation and demodulation effects abound.

The discussion of modeling was primarily qualitative. At this early point in the understanding of the new phenomena it is premature to offer any alternate computational models but one does now know a direction in which to look for improvement. The ideas of additional energy storage, modulation, and demodulation mechanisms inherent in separated flows with rotation are intriguing and offer fertile areas for additional investigation. But until and unless these flow interactions and their acoustic consequences can be adequately quantified, the highly developed models of Flanagan [43] should continue to be the best and most comprehensive available.

These ideas are only in their infancy but hopefully are reasonably coherent. They need careful examination by both the speech and the hearing communities for their relevance and utility. Although there are many unanswered questions remaining there is a definite new direction in which to look for possible explanations of long standing problems and so-called artifacts. Carefully devised and conducted experiments are absolutely necessary at this point.

There is so much that we do not know about the complex mechanism of flow interaction, acoustic wave generation, and pulsatile flow but it is hoped that the references offered will furnish a good starting point and help lead the interested reader to related fields where parts of the problem have been worked on. This should in turn lead to a richer understanding of the complexities of sound production in man and in the devising of algorithms to correctly identify those complexities.

ACKNOWLEDGEMENTS

I am deeply indebted to all those persons with whom I have discussed various parts of the general problem of this paper. The questions raised, the particular viewpoints expressed, the various observations made, the additional references provided, and the concerted thought conscientiously given by these people were greatly appreciated. The cogent criticisms of O. Fujimura, J.L. Flannagan and P.G. Simpkins on various aspects of this review were especially helpful. Perhaps most of all the thoroughness and depth of thought and seemingly endless patience given by Dr. H.M. Teager of the Boston University Medical Center to this major problem area and to my many questions merits special recognition and commendation. It is not often that one has the opportunity both to witness and to actively participate in the devising and formulating of a new approach which offers such an exciting new way to view a major outstanding problem in human communications.

*This work was done while the author was an employee of Bell Laboratories, Murray Hill, NJ 07974.

REFERENCES

1. T. Chiba, M. Kajiyama, *The Vowel, its Nature and Structure,* Kaiseikan Publ. Co., Tokyo, 1941.
2. H.K. Dunn, "The Calculation of Vowel Resonances, and an Electrical Vocal Tract", *J. Acoust. Soc. Amer.*, Vol. 22, No. 6, pp. 740-753, Nov. 1950.
3. G.E. Peterson, H.L. Barney, "Control Methods Used in a Study of the Vowels", *J. Acoust. Soc. Amer.*, Vol. 24, No. 2, pp. 175-184, Mar. 1952.
4. K.N. Stevens, S. Kasowski, C.G.M. Fant, "An Electrical Analog of the Vocal Tract", *J. Acoust. Soc. Amer.*, Vol. 25, No. 4, pp. 734-742, July 1953.
5. J.W. van den Berg, "Transmission of the Vocal Cavities", *J. Acoust. Soc. Amer.*, Vol. 27, No. 1, pp. 161-168, Jan. 1955.

6. J.L. Flanagan, "A speech analyzer for a formant-coding compression system", Sc.D. thesis, E.E. Dept., M.I.T., 1955. Appears also as Scientific Report No. 4, Acoustics Laboratory, M.I.T., May 1955.
7. J.L. Flanagan, "Automatic Extraction of Formant Frequencies from Continuous Speech", *J. Acoust. Soc. Amer.*, Vol. 28, No. 1, pp. 110-118, Jan. 1956.
8. C.G.M. Fant, *Acoustic Theory of Speech Production,* Mouton & Co., The Netherlands, 1960, reprinted 1970.
9. R.K. Potter, G.A. Kopp, H.G. Kopp, *Visible Speech,* Dover Pub., Inc., New York, 1966.
10. G. Fant, *Speech Sounds and Features,* MIT Press, Cambridge, MA, 1973.
11. J.M. Heinz, K.N. Stevens, "On the Properties of Voiceless Fricative Consonants", *J. Acoust. Soc. Amer.*, Vol. 33, No. 5, pp. 589-596, May 1961.
12. O. Fujimura, "Analysis of Nasal Consonants," *J. Acoust. Soc. Amer.*, Vol. 34, No. 12, pp. 1865-1875, Dec. 1962.
13. K.N. Stevens, "Airflow and Turbulence Noise for Fricative and Stop Consonants: Static Considerations", *J. Acoust. Soc. Amer.*, Vol. 50, No. 4, Part 2, pp. 1180-1192, 1971.
14. R.L. Wegel, "Theory of Vibration of the Larynx", *Bell Syst. Tech. J.*, Vol. 9, No. 1, pp. 207-227, Jan. 1930.
15. R.L. Miller, "Nature of the Vocal Cord Wave", *J. Acoust. Soc. Amer.,* Vol. 31, No. 6, pp. 667-677, June 1959.
16. M.V. Mathews, J.E. Miller, E.E. David, Jr., "Pitch Extraction Analysis of Voiced Sounds", *J. Acoust. Soc. Amer.*, Vol. 33, No. 2, pp. 179-186, Feb. 1961.
17. J.L. Flanagan, L. Cherry, "Excitation of Vocal Tract Synthesizers", *J. Acoust. Soc. Amer.*, Vol. 45, No. 3, pp. 764-769, Mar. 1969.
18. M.R. Schroeder, "Determination of the Geometry of the Human Vocal Tract", *J. Acoust. Soc. Amer.*, Vol. 41, Part 2, pp. 1002-1010, Apr. 1967.
19. P. Mermelstein, "Determination of the Vocal Tract-Shape from Measured Formant Frequencies", *J. Acoust. Soc. Amer.*, Vol. 41, pp. 1283-1294, May 1967.
20. B. Gopinath, M.M. Sondhi, "Determination of the Shape of the Human Vocal Tract from Acoustical Measurements", *Bell Syst. Tech. J.*, Vol. 49, pp. 1195-1214, July-Aug. 1970.
21. B.S. Atal, "Determination of the Vocal Tract Shape Directly from the Speech Wave", *J. Acoust. Soc. Amer.*, Vol. 47 (A), p. 64, Jan. 1970.
22. M.M. Sondhi, B. Gopinath, "Determination of Vocal-Tract Shape from Impulse Response at the Lip", *J. Acoust. Soc. Amer.*, Vol. 49, No. 6, Part 2, pp. 1847-1873, June 1971.
23. H. Wakita, "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Waveforms", *IEEE Trans. Audio Electroacoust.*, Vol. AU-21, No. 5, pp. 417-427, October 1973.
24. J.J. Fredberg, M.E.B. Wohl, G.M. Glass, H.L. Dorkin, "Airway area by acoustic reflections measured at the mouth", *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.*, Vol. 48, No. 5, pp. 749-758, 1980.
25. J.L. Flanagan, *Speech Analysis/Synthesis and Perception,*, Springer-Verlag, Berlin, 2nd edition, 1972.
26. M.M. Sondhi, "Model for Wave Propagation in a Lossy Vocal Tract", *J. Acoust. Soc. Amer.*, Vol. 55, No. 5, pp. 1070-1075, May 1974.
27. J.W. van den Berg, J.T. Zantema, P. Doornenbal, Jr., "On the Air Resistance and the Bernoulli Effect of the Human Larynx", *J. Acoust. Soc. Amer.*, Vol. 29, No. 5, pp.

626-631, May 1957.

28. J.L. Flanagan, "Some Properties of the Glottal Sound Source", *J. Speech Hearing Res.*, Vol. 1, pp. 99-116, 1958.
29. M.V. Mathews, J.E. Miller, E.E. David, "An Accurate Estimate of the Glottal Waveshape", *J. Acoust. Soc. Amer.*, Vol. 33, pp. 843, 1969, (talk).
30. J.E.Miller, M.V. Mathews, "Investigation of the Glottal Waveshape by Automatic Inverse Filtering", *J. Acoust. Soc. Amer.*, Vol. 35, pp. 1876, 1963, (talk).
31. J.L. Flanagan, "Source-system Interaction in the Vocal Tract", *Ann. New York Acad. Sci.*, Vol. 155, Art. 1, pp. 9-17, Nov. 20, 1968.
32. J.L. Flanagan, L.L. Landgraf, "Self-oscillating Source for Vocal- tract Synthesizers", *IEEE Trans. Audio Electroacoust. (Special Issue on Speech Communication and Processing - Part I)*, Vol. AU-16, No. 1, pp. 57-64, Mar. 1968.
33. K. Ishizaka, M. Matsudaira, "What Makes the Vocal Cords Vibrate", *Proc. 6th Intl. Congr. Acoust., Part II*, pp. 9-12, Aug. 1968.
34. K. Ishizaka, J.L. Flanagan, "Synthesis of Voiced Sounds from a Two- Mass Model of the Vocal Cords", *Bell Syst. Tech. J.*, Vol. 50, No. 7, pp. 1233-1268, July-August 1972.
35. M. Rothenberg, "A New Inverse-filtering Technique for Deriving the Glottal Air Flow Waveform During Voicing", *J. Acoust. Soc. Amer.*, Vol. 53, No. 6, pp. 1632-1645, June 1973.
36. V. Gupta, T.A. Wilson, G.S. Beavers, "A model for vocal cord excitation", *J. Acoust. Soc. Amer.*, Vol. 54, No. 6, pp. 1607-1617, Dec. 1973.
37. M.M. Sondhi, "Measurement of the Glottal Waveform", *J. Acoust. Soc. Amer.*, Vol. 57, No. 1, pp. 228-232, Jan. 1975.
38. J.L. Flanagan, K. Ishizaka, K.L. Shipley, "Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract", *Bell Syst. Tech. J.*, Vol. 54, No. 3, pp. 485-506, Mar. 1975.
39. J.L. Flanagan, K. Ishizaka, "Automatic Generation of Voiceless Excitation in a Vocal Cord - Vocal Tract Speech Synthesizer", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-24, No. 2, pp. 163-170, Apr. 1976.
40. M. Hirano, "Structure and Vibratory Behavior of the Vocal Folds", *U.S.-Japan Joint Seminar on Dynamic Aspects of Speech Production,* pp. 3-15, December 1976.
41. D.J. Broad, "The New Theories of Vocal Fold Vibration", *Speech and Language: Advances in Basic Research and Practice,* Vol. 2, Academic Press, Inc., pp. 203-256, 1979.
42. K. Ishizaka, J.C. French, J.L. Flanagan, "Direct Determination of Vocal Tract Wall Impedance", *IEEE Trans. Acoust., Speech, Signal Processing,* Vol. ASSP-23, No. 4, pp. 370-373, Aug. 1975.
43. J.L. Flanagan, K. Ishizaka, K.L. Shipley, "Signal Models for Low Bit Rate Coders", *J. Acoust. Soc. Amer.*, Vol. 68, No. 3, pp. 780-791, Sept. 1980.
44. L.R. Rabiner, R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1978.
45. D.W. Farnsworth, "High-speed Motion Pictures of the Human Vocal Cords", *Bell Labs. Record*, Vol. 18, No. 7, pp. 203-208, Mar, 1940; also 16 mm. film, same title, 22 min., silent, B&W, Bell Laboratories, 1939.
46. M. Hirano, "Official Report Phonosurgery - Basic and Clinical Investigations", *Otologia* (Fukuoka), Supplement No. 1, Vol. 21, pp. 239- 440, May 16, 1975; 76th Annual Conv. Oto-rhino-laryngological Soc. of Japan, Kurume Univ., Kurume, Japan, 1975.
47. O. Fujimura, "The Larynx in Speech Utterances", 16 mm film.
48. H. Fukuda, S. Saito, K. Kitahara, Y. Isogai, K. Makino, T. Tsuzuki, N. Kogawa, H.

Ono, "Vocal Fold Vibration in Excised Larynges with X-Ray Stroboscope and Ultra-High Speed Camera", Vocal Fold Physiology Conference, University of Wisconsin, Madison, WI, May 31 - June 4, 1981, College Hill Press, San Diego, CA 92120.

49. D.G. Childers, A.M. Smith, G.P. Moore, "Relationships Between Electroglottagraph, Speech, and Vocal Cord Contact", preprint, with 16 mm film of results, University of Florida, Department of Electrical Engineering, Gainesville, Florida, May 1982.
50. A. Bouhuys, J. Mead, D.F. Proctor, K.S. Stevens, "Pressure-Flow Events During Singing", *Annals of the New York Academy of Sciences*, Vol. 155, Art. 1, Special Issue - Sound Production in Man, pp. 165-176, Nov. 1968.
51. I.R. Titze, "Voice as a musical instrument", *J. Acoust. Soc. Amer.*, Vol. 65, Suppl. No. 1, p. S-93, Spring 1979.
52. D.H. Klatt, R.A. Stefanski, "How does a mynah bird imitate human speech?", *J. Acoust. Soc. Amer.*, Vol. 55, No. 4, pp. 822-832, April 1974.
53. C.H. Greenwalt, "How Birds Sing", *Scientific American,* Vol. 221, No. 5, pp. 126-139, Nov. 1969.
54. B. Webster, "When Songbirds Speak, Brain Should Get the Credit", *New York Times,* Section C, pp. C1 ff, May 18, 1982.
55. F.R. Newman, *Mouth Sounds,* Workman Publishing, New York, 1980.
56. J.C.R. Licklider, "Effects of Amplitude Distortion on the Intelligibility of Speech", *J. Acoust. Soc. Amer.*, Vol. 18, No. 2, pp. 429-434, Oct. 1946.
57. J.C.R. Licklider, I. Pollack, "Effects of Differentiation, Integration, and Infinite Peak Clipping Upon the Intelligibility of Speech", *J. Acoust. Soc. Amer.*, Vol. 20, No. 1, pp. 42-51, Jan. 1948.
58. J.C.R. Licklider, G.A. Miller, "The Perception of Speech", Chapter 26, pp. 1040-1074, in *Handbook of Experimental Psychology,* S.S. Stevens, John Wiley & Sons, Inc., New York, 1951.
59. H.M. Teager, "Some Observations on Oral Air Flow During Phonation", *IEEE Trans. Acoust., Speech, Signal Processing,* Vol. ASSP-28, No. 5, pp. 599-601, Oct. 1980.
60. H. Schlichting, *Boundary-Layer Theory,* McGraw-Hill Book Co., New York, 1968.
61. L. Prandtl, *Essentials of Fluid Dynamics,* Blackie & Son Ltd., London, 1952.
62. B.S. Massey, *Mechanics of Fluids*, Fourth Edition, Van Nostrand Reinhold Co., New York, 1979.
63. W. Vennard, *Singing: The Mechanism and the Technic,* 4th edition, Carl Fischer, Inc., New York, 1967.
64. P.M. Morse, *Vibration and Sound,* Second Edition, McGraw-Hill Book Co., New York, 1948.
65. P.M. Morse, U. Ingard, *Theoretical Acoustics,* McGraw-Hill Book Co., New York, 1968.
66. A.D. Pierce, *Acoustics / An Introduction to Its Physical Principles and Applications,* McGraw-Hill Book Co., New York, 1981.
67. H.P.M. Bouasse, *Instruments a Vent,* 2 volumes, Librairie Delagrave, Paris, France, 1929.
68. H.P.M. Bouasse, *Tuyaux et Resonateurs,* Librairie Delagrave, Paris, France, 1929.
69. H.P.M. Bouasse, *Jets, Tuyaux et Canaux,* Librairie Delagrave, Paris, France, 1923.
70. H.P.M. Bouasse, *Tourbillons, Forces Acoustique, Circulation Diverses,* 2 Volumes, Librairie Delagrave, Paris, France, 1931 & 1932.
71. H.P.M. Bouasse, *Acoustique Generale,* Librairie Delagrave, Paris, France, 1926.
72. J.M. Kirshner, Editor, *Fluid Amplifiers,* McGraw-Hill Book Co., New York, 1966.
73. I. Reba, "Applications of the Coanda Effect", *Scientific American,* Vol. 214, No. 6, pp.

84-92, June 1966.

74. A. Metral, "Sur un Phenomene de Deviation des Veines Fluides et ses Applications (Effet Coanda)", *Proc. Fifth Intl. Congress Appl. Mech.*, pp. 456-465, 1938.
75. B.G. Newman, "The Deflexion of Plane Jets by Adjacent Boundaries — Coanda Effect", *Boundary Layer & Flow Control,* G.V. Lachman, Editor, Vol. I, Pergamon Press, New York, pp. 232-264, 1961.
76. H.M. Teager, "The Effect of Separated Air Flow on Vocalization", Vocal Fold Physiology Conference, University of Wisconsin, Madison, WI, May 31 - June 4, 1981, College Hill Press, San Diego, CA 92120.
77. J. Gauffin, B. Nguyen, T.V. Ananthapadmanabha, G. Fant, "Glottal Geometry and Volume Velocity Waveform", Vocal Fold Physiology Conference, University of Wisconsin, Madison, WI, May 31 - June 4, 1981, College Hill Press, San Diego, CA 92120.
78. D.E. Proctor, "Control of the Breath", Vocal Fold Physiology Conference, University of Wisconsin, Madison, WI, May 31 - June 4, 1981.
79. E. Muller, "Particle Velocity Profiles and Separated Flows in the Vocal Tract", Vocal Fold Physiology Conference, University of Wisconsin, Madison, WI, May 31 - June 4, 1981, College Hill Press, San Diego, CA 92120.
80. T.J. Pedley, M.F. Sudlow, R.C. Schroter, "Gas Flow and Mixing in the Airways", Chapter 3, in *Bioengineering Aspects of the Lung,* edited by J. West, Marcel Dekker, Inc., New York, 1977.
81. J.B. West, P. Hugh-Jones, "Patterns of gas flow in the upper bronchial tree", *J. Appl. Physiol.*, Vol. 14, pp. 753-759, 1959.
82. J.E. Ffowcs Williams, "Aeroacoustics", *Ann. Rev. Fluid Mech.*, Vol. 9, #8110, pp. 447-468, 1977.
83. D.C. Burnham, "Review of Vortex Sensor Development Since 1970", *Proc. Aircraft Wake Vortices Conference,* March 15-17, 1977, Cambridge, MA, pp. 47-66; also available from NTIS as Report FAA-RD-77-68, J.N. Hallock, Editor, June 1977.
84. G.M. Golemshtok, A.L. Fabrikant, "Scattering and amplification of sound waves by a cylindrical vortex", *Sov. Phys. Acoust.*, Vol. 26, pp. 209 ff, 1980.
85. A.L. Fabrikant, "Sound scattering by a vortex", *Sov. Phys. Acoust.*, Vol. 28, No. 5, pp. 410-411, Sep. - Oct. 1982.
86. H.M. Teager, S.M. Teager, "A Phenomenological Model for Vowel Production in the Vocal Tract", *Recent Advànces in Speech Sciences,* R.G. Daniloff, Editor, College Hill Press, San Diego, CA, 1983.
87. ASME, *Fluid Jet Control Devices,* Proc. Symposium on Fluid Jet Control Devices, Amer. Soc. Mech. Engrs., 1962.
88. E.F. Humphrey, D.H. Tarumoto, *Fluidics,* Fluid Amplifier Associates, Boston, MA, 1965.
89. J.M. Kirshner, Editor, *Fluid Amplifiers,* McGraw-Hill Book Co., New York, 1966.
90. G.A. Cavagna, R. Margaria, "Airflow Rates and Efficiency Changes During Phonation", *Annals of the New York Academy of Sciences,* Vol. 155, Art 1, Special Issue - Sound Production in Man, pp. 152-164, Nov. 1968.
91. T.A.C.M. Claasen, W.F.G. Mecklenbrauker, "The Wigner Distribution — A Tool for Time-Frequency Signal Analysis, Part I: Continuous-Time Signals", Vol. 35, No. 3, pp. 217-250, 1980.
92. T.A.C.M. Claasen, W.F.G. Mecklenbrauker, "The Wigner Distribution — A Tool for Time-Frequency Signal Analysis, Part II: Discrete- Time Signals", *Philips J. Res.*, Vol. 35, Nos. 4/5, pp. 276-300, 1980.

93. T.A.C.M. Claasen, W.F.G. Mecklenbrauker, "The Wigner Distribution — A Tool for Time-Frequency Signal Analysis, Part III: Relations with Other Time-Frequency Signal Transformations", *Philips J. Res.,* Vol. 35, No. 6, pp. 372-389, 1980.
94. P. Flandrin, B. Escudie, "Time and frequency representation of finite energy signals: A physical property as a result of an Hilbertian condition", *Signal Processing,* Vol. 2, pp. 93-100, 1980.
95. D.B. Chester, "The Wigner Distribution and its Application to Speech Analysis and Recognition", Ph.D. Thesis, Univ. of Cincinnati, Department of Electrical and Computer Engineering, Cincinnati, OH, 1982.
96. E. Berger, R. Wille, "Periodic Flow Phenomena", *Ann. Rev. Fluid Mech.*, Vol. 4, #8027, pp. 313-340, 1972.
97. S.E. Widnall, "The Structure and Dynamics of Vortex Filaments", *Ann. Rev. Fluid Mech.*, Vol. 7, #8070, pp. 141-165, 1975.
98. D. Rockwell, E. Naudascher, "Self-sustained Oscillations of Impinging Free Shear Layers", *Ann. Rev. Fluid Mech.*, Vol. 11, #8135, pp. 67-94, 1979.
99. P.G. Saffman, G.R. Baker, "Vortex Interactions", *Ann. Rev. Fluid Mech.*, Vol. 11, #8136, pp. 95-122, 1979.
100. N.H. Fletcher, "Air Flow and Sound Generation in Musical Instruments", *Ann. Rev. Fluid Mech.*, Vol. 11, #8137, pp. 123-146, 1979.
101. G.B. Brown, "On Vortex Motion in Gaseous Jets and their Sensitivity to Sound", *Proc. Phys. Soc.*, Vol. 47, pp. 703-732, 1935.
102. R.M. Miura, Editor, *Backlund Transformation, the Inverse Scattering Method, Solitons, and Their Applications,* Springer-Verlag, New York, 1976.
103. A.R. Bishop, T. Schneider, Editors, "Solitons and Condensed Matter Physics", *Proceedings of the Symposium on Nonlinear (Soliton) Structure and Dynamics in Condensed Matter,* Oxford, England, June 1978, Springer-Verlag, Berlin, 341 pp., 39 papers, 1978.
104. K. Lonngreen, A. Scott, Editors, *Solitons in Action*, Academic Press, New York, 1978.
105. G.L. Lamb Jr., *Elements of Soliton Theory,* John Wiley & Sons, Inc., New York, 1980.
106. R.K. Bullough, P.J. Caudrey, Editors, *Solitons,* Springer-Verlag, Berlin, 1980.
107. M. Ablowitz, H. Segur, *Solitons and the Inverse Scattering Theorem,* SIAM, Philadelphia, PA, 1981.
108. G. Eilenberger, *Solitons, Mathematical Method for Physicists,* Springer-Verlag, Berlin, 1981.
109. W. Wathen-Dunn, S.B. Michaels, "Some Effects of Gas Density on Speech Production", *Annals of the New York Academy of Sciences,* Vol. 155, Art. 1, Special Issue - Sound Production in Man, pp. 368-378, Nov. 1968.
110. B.S. Atal, J.R. Remde, "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates", *Proc. 1982 ICASSP Conf.*, pp. 614-617, 1982.
111. S.A. Elder, "On the mechanism of sound production in organ pipes", *J. Acoust. Soc. Amer.*, Vol. 54, No. 6, pp. 1554-1564, Dec. 1973.
112. S.A. Elder, "Self-excited depth-mode resonance for a wall-mounted cavity in turbulent flow", *J. Acoust. Soc. Amer.*, Vol. 64, No. 3, pp. 877-890, 1978.
113. S.A. Elder, "Forced oscillations of a separated shear layer with application to cavity flow-tone effects", *J. Acoust. Soc. Amer.*, Vol. 67, No. 3, pp. 774-781, 1980.
114. W.L. Nelson, "Physical Principles for Economies of Skilled Movements", *Biological Cybernetics,* Vol. 46, pp. 135-147, 1983.
115. R.F. Blackwelder, "Hot-Wire and Hot-Film Anemometers", pp. 260-314, in *Fluid*

Dynamics, Methods of Experimental Physics Series, R.J. Emrich, Editor, Vol. 18, Part A, Academic Press, New York, 1981.

116. G. Comte-Bellot, "Hot-Wire Anemometry", *Ann. Rev. Fluid Mech.*, Vol. 8, #8088, pp. 209-231, 1976.
117. E. Ossofsky, "Constant Temperature Operation of the Hot-Wire Anemometer", *Review of Scientific Instruments,* Vol. 19, No. 12, pp. 881-889, Dec. 1948.
118. R.W.Ladenburg, "Hot-Wire Method", pp. 219-285, in *Physical Measurements in Gas Dynamics and Combustion,* Vol. IX, High Speed Aerodynamics and Jet Propulsion, Princeton University Press, Princeton, NJ, 1954.
119. J.M.L. Janssen, L. Ensing, J.B. Van Erp, "A Constant-Temperature-Operation Hot-Wire Anemometer", *Proc. IRE,* Vol. 47, No. 4, pp. 555-567, April 1959.
120. J.S. Lundsgaard, J. Gronlund, N. Einer-Jensen, "Evaluation of a constant-temperature hot-wire anemometer for respiratory-gas-flow measurements", *Med. & Biol. Eng. & Comput.*, Vol. 17, pp. 211-215, March 1979.
121. A.E. Perry, *Hot-Wire Anemometry,* Oxford University Press, Oxford, 184 pp., 1982.
122. H. Werle, "Hydrodynamic Flow Visualization", *Ann. Rev. Fluid Mech.*, Vol. 5, #8048, pp. 361-382, 1973.
123. W. von Kempelen, *Le Mechanisme de la Parole, suivi de la Description d'une Machine Parlante,* J.V. Degen, Vienna, 1791; (copies are held at United Engineering Center and the New York Public Library in French and German); an English translation is in press at the Venerable Press, 58 Edgemoor Road, Belmont MA 02178.
124. F.D. Minifie, T.J. Hixon, F. Williams, Editors, *Normal Aspects of Speech, Hearing, and Language,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1973.
125. E.D. Schubert, *Hearing: Its Function and Dysfunction,* Springer-Verlag, Vienna, 1980.
126. J.E. Bernthal, N.W. Bankson, *Articulation Disorders,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1981.
127. M. Hirano, *Clinical Examination of the Voice,* Springer-Verlag, New York, 1981.
128. C.L. Ludlow, M. O'Connell Hart, Editors, *Proc. of the Conf. on the Assessment of Vocal Pathology,* The American Speech-Language-Hearing Association, Rockville, MD, Dec. 1981; twelve paper on the mechanism and the assessment of phonatory function.
129. V.E. Negus, *The Comparative Anatomy and Physiology of the Larynx,* Grune & Stratton, Inc., New York, 1949.
130. B.R. Fink, *The Human Larynx / A Functional Study,* Raven Press, New York, 1975.
131. C. Meano, A. Khoury, *The Human Voice in Speech and Song,* C.C. Thomas Publ., Springfield, Ill., 1967.
132. G.J. Borden, K.S. Harris, *Speech Science Primer / Physiology, Acoustics, and Perception of Speech,* Williams & Wilkins, Baltimore, MD, 1980.
133. P. Ladefoged, *Three Areas of Experimental Phonetics,* Oxford University Press, London, 1967.
134. H.L.F.v. Helmholtz, *On the Sensations of Tone,* Dover Publications, Inc., New York, 1954.
135. A.H. Benade, *Horns, Strings & Harmony,* Anchor Books, Doubleday & Co., New York, 1960.
136. A.H. Benade, *Fundamentals of Musical Acoustics,* Oxford University Press, New York, 1976.
137. F. Winckel, *Music, Sound and Sensation; A Modern Exposition,* Dover, New York, 1967.
138. O. Fujimura, "Control of the Larynx in Speech", *Proc. Seminar on the Larynx and Language;Eighth Intl. Cong. of Phonetic Sciences,* Leeds, England, Aug. 1975.

139. J.R. Pierce, "Whither Speech Recognition?", *J. Acoust. Soc. Amer.*, Vol. 46, No. 4, Part 2, pp. 1049-1051, 1969.
140. J.V. Tobias, Editor, *Foundations of Modern Auditory Theory,* Volume II, Academic Press, New York, 1972.
141. F. Fallside, "Communication aids for the deaf", *Endeavour,* New Series, Vol. 3, No. 2, pp. 63-68, 1979.
142. J. Large, S. Iwata, "Aerodynamic Study of Vibrato and Voluntary 'Straight Tone' Pairs in Singing", *Folia Phoniat.*, Vol. 23, pp. 50-65, 1971.
143. R.F. Lyon, "A Computational Model of Filtering, Detection, and Compression in the Cochlea", *Proc. 1982 ICASSP Conf.*, pp. 1282-1285, 1982.
144. G.K. Batchelor, *An Introduction to Fluid Dynamics,* Cambridge University Press, Cambridge, 1967.
145. J. Lighthill, *Mathematical Biofluiddynamics,* SIAM, Philadelphia, 1975.
146. O. Fujimura, J. Lundqvist, "Sweep-Tone Measurements of Vocal-Tract Characteristics", *J. Acoust. Soc. Amer.*, Vol. 49, No. 2 (Part 2), pp. 541-558, Feb. 1971.
147. G.v. Bekesy, "Introduction", *Auditory System Anatomy Physiology (Ear),* Edited by W. D. Keideland, W.D. Neff, Springer-Verlag, Berlin, pp. 1-8, 1974.
148. G. Troup, "The Physics of the Singing Voice", *J. of Research in Singing,* Vol. VI, No. 1, pp. 1-26, Dec. 1982.

DISCUSSION

P. MILENKOVIC: No one is denying the existence of acoustic waves. The question is whether the acoustic source is distributed along the vocal tract or lumped at the glottis. No one is denying the existence of distributed sources either, as is the case with fricative sounds. The paper seems to argue that because the bulk flow contains so much energy, some of this energy has to leak out as a distributed source. My view is that since the flow velocity and rate of change of velocity are so much larger at the vocal folds, the bulk of source effects occur at the glottis in vowel sounds.

J. KAISER: There is no question that in voicing the primary source for the acoustic energy component at the pitch frequency is the flow wave produced at the vocal folds. What we suggest is that the higher frequency components of the voiced acoustic wave, i.e., the formant structure, can and do result from complex flow interactions involving momentum exchange which take place primarily downstream from the glottis. This removes an unnecessarily great burden from the glottal flow wave as having to provide all the high frequency components found in the formant structure. If higher frequency components are observed in the glottal pressure wave, this could be the result of a backward traveling acoustic wave from the place of flow interaction generating that component, i.e., an effect not a cause. These flow interactions are localized under steady conditions and have dynamics that are affected not only by the primary flow magnitude and direction but also by the local geometry. They are established or re-established each pitch period. The word "leak" is hardly appropriate for describing these interactions as Teager's results show that there is substantial flow energy complete with formant structure as far downstream as the mouth and that if sound were simply proportional to the time-rate-of-change of flow then sound intensities should be present which are more than several orders of magnitude greater than are observed; clearly much more is going on here than is being accounted for in the plane wave model.

G. FANT: A challenge of established views is always healthy, but I feel that our present fundamentals of acoustic theory of speech production are not shaken by your review. On the other hand, we would gain from more detailed insight into transverse wave propagation and high amplitude elaboration of the wave equation.

In my experience, mathematical reconstruction of voiced sounds from present interactive models provides a good fit to measured data. I do not agree that the points 1, 2, 3, and 8 in your written contribution present any anomalies.

J. KAISER: In discussing the operation of the vocal tract one must first make it clear what the objectives are motivating the work: most of the work to date has been directed toward the obtaining of workable mathematical models that can be used to compute an output time signal which corresponds closely to some characteristic (usually frequency) of the signal recorded at a pressure sensitive microphone outside the mouth. The best of these models are suitably complex and have enough adjustable parameters so that persons skilled in approximation techniques can reproduce almost any desired speech waveform by simply manipulating the parameter values. However that is not proof that these models are accurate representations of what is physically going on inside the tract.

On the other hand our investigations have been directed more at trying to understand the physical mechanisms underlying the generation of the complicated acoustic wave emanating from the mouth when we speak or sing: thus one must first understand the details of the air flow as that is the medium of transmission for any acoustic waves in the tract. Once this understanding is developed then, and only then, can one really begin to try to mathematically model what is actually going on. Thus when you speak of trying to obtain more detailed insight into transverse wave propagation, I would say let us look in more detail at the different flow patterns and mechanisms of energy storage and release that obtain inside the vocal tract; longitudinal and transverse may not be the best or simplest decomposition of the internal acoustic wave modes especially when vortices of different varieties are present. When you speak of high amplitude elaboration of the wave equation, I would suggest that we do not even understand sufficiently well the small amplitude case as the system is inherently nonlinear and similar phenomena occur at low amplitudes as well as at high amplitudes. As you know from the steady state glottal flow experiments of Gauffin *et al* performed in your laboratory, separated jet sheet flow attached to one wall of the tract model was the observed flow behavior over a range of velocities. Thus the flow regime is anything but linear suggesting that much more experimentation needs to be done.

A. BENADE: The Lighthill (convective) term of the second-order wave equation, which is proportional to $\sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} (\rho V_i V_j)$ can generate a large amount of signal at double frequency and simple $f_a \pm f_b$ regions of the spectrum.

We have seen enormous effects in the woodwinds of this type at comparable signal levels.

J. KAISER: I welcome your observation that a large amount of signal is generated in the woodwinds at frequencies not present in the driving flow. As to Lighthill's formulation, I do not understand it that well but believe it is concerned primarily with turbulent flow with very large velocity gradients, i.e., high Reynolds numbers. Since we are not postulating turbulent flow for voicing I am uncertain whether or not this convective term adequately describes the situation. Further modeling work certainly is merited.

A. BENADE: We have seen strictly nonturbulent second harmonic generation up to 40 percent of a sinusoidal excitation! With toneholds, etc. the effects can be enormous. (Inserted in proof).

O. FUJIMURA: There is no question that there are some deviations from the standard acoustic theory: the question is how much. For example, the formant frequencies may not be as predicted by Fant's theory because the acoustic parameters of the flowing medium varies depending on the speed of the flow, but we have to know (approximately) how much is the predicted deviation due to realistic flow speed variability across the cross section. I am all for experimenting to derive such quantitative estimates, even within the framework of the accepted theory.

J. KAISER: I strongly agree. Detailed and careful experimentation by other competent investigators is urgently needed at this point. The experiments should be both in vivo and in vitro and should consist of simultaneous direct macroscopic flow and pressure measurements. The investigators should continually be wary of "finding what they are looking for". In experimental fluid dynamics there are often many surprises awaiting the neophyte as well as the more experienced researcher.

32 ▪ ACTIVE FLUID DYNAMIC VOICE PRODUCTION MODELS, OR THERE IS A UNICORN IN THE GARDEN

Herbert M. Teager and Shushan M. Teager

Our objective this afternoon is to present a basis for a new model for speech production after presenting the evidence which renders current models untenable.

Two years ago in Wisconsin, and earlier in an IEEE paper, we presented some of our findings on separated flow. If one is to judge by citations or the replication of any of our hot wire experiments, positive results were almost nil. On the other hand, we have made some good friends and have been invited to present our latest findings to this distinguished body.

When one works in a field for some years and makes discoveries and observations that remain largely ignored, one has to either develop a philosophical attitude, or else give in to despair or acute paranoia. Seeing one's self as a penny-ante Galileo, unable to convince the papal authorities to go out and look through their own telescopes, while good for the ego, does not help get the message across, and perhaps leads to an overly pedantic and fatalistic attitude. The infrequent opportunities available to cite one's work in public turn into presentations, crammed with too many findings, that only make communication more difficult. Perhaps a more reasonable framework is to see the problem through the eyes and beliefs of the average workers in this field, who have been thoroughly conditioned by the standard textbook references which seemingly leave little room for any doubt whatsoever.

To better communicate with that listener, I have put myself in his or her position: namely the position I was in when I first began working with hearing and was mildly curious about possible non-linearities in speech. It has been suggested by friends that I sound like a revolutionary, and that I would be far better advised to keep my conclusions very mild and uncontroversial. I could not agree with them more. Being a scientific pariah is hardly a position a rational man would choose, and my original path was very conventional. After all, I started in the field as a respected electrical engineer with the standard faith in linear systems theory, with graduate and faculty credentials at the proper institutions to match. Many of the leaders in the field, were in fact, my graduate school class mates.

My initial findings and questions were modest ones, but I soon ran into a stone wall of what looked like doctrinaire opposition. Measured results that apparently disagreed with theory had to be either wrong, or of no significance. As a consequence I dug deeper both in my reading and experiments, to check other inalienable truths as well, which only radicalized me further.

I cannot go over the entire gamut of results, but I would like to review only a few of the salient discrepancies that convinced me of the work's validity. In doing so, I hope that you will listen with skepticism, but with an open mind, as you would to an explorer recounting his observations in a strange land. After all, when some nut tells you that what you thought was a rabbit in the garden is more like a unicorn, or a herd of elephants, you would be curious enough to look out the window, rather than refer to the literature for a dissertation on rabbits. This paper will thus be devoted to an overview of our past work and the presentation of a few new findings. For those who are desirous of more depth, we have available copies of our IEEE (1980) and Wisconsin (1981) papers, as well as our chapter in the forthcoming "Speech Science: Recent Advances", R.G. Daniloff, Editor (1985).

Our first enigma with speech production involved energy. The authorities seemed unanimous in asserting that losses in the vocal tract were small, and only a fraction of one percent of the expended mechanical lung input energy was involved in the rate of change of volume velocity at the glottis. Even that small amount of energy was further reduced by filtering in the vocal tract. No suggestions were ever evident as to where the missing 99.5% + went. Yet since it was 200 times greater than acoustic energy, it might be capable of producing quite large effects. Was there a loose cannon on the deck?

Our second set of problems came about when we began to work with speech signals both in the time and frequency domains. In any passive linear system such time and frequency observations should have been completely interchangeable, yet our observations indicated that they were not. The second formant, to cite but one example, could rapidly rise and fall one or more times within a single pitch period; yet the bandwidth, or Q, of the formant, as determined by a fourier analysis, would have mandated far longer rise and fall times.

Our third set of difficulties arose from an attempt to understand published work on diver's speech with high pressure Helium, where formants did not move up by the expected velocity of sound ratios. Being aware of the yielding wall hypothesis, Fant and Sonneson (1964), we ran an analogous experiment with pure Helium at atmospheric pressure in order to eliminate the effects of high pressure. We found that all formants shifted by a factor closer to 1.6 than the expected 2.9. Two significant points can be brought out from these results without any quibbling about loopholes in the theoretical derivations. These points are of value since responsible people in the field still seem to be under the impression that the yielding wall effect, postulated for high density vocal tract gases, predicts all the observed Helium speech effects. First: no one disputes that the effects of yielding walls would be significant only for formants below 200 Hz (Sondhi, 1974). Second: since a high density vocal tract gas produces a relative downwards formant shift, the reverse, or a relative upwards shift, should occur with low density atmospheric pressure Helium. The effects we measured, even for the lowest formant, were clearly in the wrong direction.

In the course of a recent extensive literature search to locate the source of the belief in yielding walls, we uncovered the papers of Morrow (1971) and Wathen-Dunn, *et al* (1968) which indicate the presence of a large number of other unsystematic formant shifts and losses with both high and atmospheric pressure gas mixtures of various sonic velocities and densities. Differing effects were noted by both authors, but were never explained or addressed again.

We began working with hot wire anemometry a decade ago with an unrecognized handicap that is shared with most workers in the field. We had virtually no physical feel for the pressures, flow rates and energies of speech, despite past work with signals from microphones. One of our biggest recent surprises could have been observed much sooner had it not taken us so long to get a feel for the magnitude and inter-relationships of dynes, pascals, rayals, ergs, and meters per second. An electrical engineer has a gut knowledge, or learns the hard way, of voltage, current and wattage magnitudes that are commensurate with different circuit elements. Whether it is because air flows are invisible, or because we have become spoiled with Db scales, FFT's, amplifiers, and microphones, such a feel is harder to come by for air measurements. While a physical feel for electricity might give one protection in differentiating between large and small effects, it was many years and lots of calibration procedures later before we developed a quantitative feel for vocal flows and pressures.

Before running experiments with advanced, crosswise moving trajectories of calibrated hot wire arrays, we had worked for years with single and multiple probes in various locations to learn the regions of interest for each vowel. Even those simple initial experiments yielded data which made it quite obvious that uniform, plane acoustical waves were incompatible with the observed separated flows. It was similarly clear that flows could possess rotations and did not necessarily repeat themselves from cycle to cycle. In our early papers, we stated that air flow was not only separated, but that the formant structure of interior flows and exterior sound were different. Further, the flow patterns, while consistent for any given vowel phoneme, differed radically across phonemes.

To elucidate what was going on within the mouth, we modified our instrumentation to include a calibrated array microphone and replaced one axial hot wire sensor by a radially sensitive one, and ran sets of mouth trajectories both along the assumed mean flow axis and crosswise in front of the teeth. The internal pressures we measured at normal speaking volumes were nowhere greater than 100 dynes/sq cm, yet the typical peak flow rate of 300 cm/sec (Teager, 1980) multiplied by the acoustic impedance should have yielded 12,000 dynes/sq cm. It took us several years to appreciate this major discrepancy.

From trajectories along the midline, we found that flow pulses were propagating at speeds comparable to, but less than the velocity of sound. The precision of our velocity measurements were too poor to resolve speeds within better than a factor of two, yet it also seemed that flow pulses, arising from the front and back of the mouth, moved at differing speeds and attenuations. Since the flows for "IH" and "EH" along the palate and tongue were manifestly different, we tried to understand the pressures that we assumed were driving them. Accordingly we set up two calibrated microphones in a differential arrangement to look for both the expected differential pressures between top and bottom, as well as along the top, mid, and bottom flow contours. The experiment produced a null result: pressures were uniform over cross sections within which separated flow occured. Further, differential pressures along the major flow axis were again 100 times smaller than predicted. After ruling out standing waves, we were left with a major dilema. What was driving the observed separated flow waves?

Acoustic impedance is not merely a nice number that relates pressures and flows in a sound wave. It also tacitly expresses the underlying assumed physical principle that a positive or negative going sound wave owes its motion to an equal interchange between compression and flow energies. Yet our measurements showed a flow wave that moved without a corresponding pressure wave. Since it clearly had kinetic energy to burn, and was losing part of that energy density downstream, it seemed as if we had discovered a new non-linear phenomenon which might be analogous to a soliton. We named this phenomenon a momentum wave.

There comes a time in any experimenter's life when it would be nice to ease up the pace of new and startling discoveries and find a usable theory to unify and guide his work. When basic physics seemingly fails, it is time to go back in history to locate the tacit assumption leading to the paradox. It took quite a while, but we finally found the oversight.

As most of you probably know, the Webster horn and other acoustic equations are based on the one dimensional wave equation:

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \qquad c^2 = \frac{1}{k\rho} \qquad (32\text{-}1)$$

which is easily derived from the equations of continuity and motion by differentiation:

$$k\frac{\partial p}{\partial t} = -\frac{\partial u}{\partial x} \qquad \frac{\partial p}{\partial x} = -\rho\frac{\partial u}{\partial t} \qquad (32\text{-}2)$$

While the classical references in acoustics such as Morse (1936) or Beranek (1954) allow that equations (32-1) and (32-2) can be slightly off because of second order changes in density accompanying pressure change, they are presented as applicable for all flow regimes.

With hindsight, the major pitfall can be seen in the second equation of (32-2) which is a restatement of the familiar F = mA. The applicable time derivative of velocity must be the total, not partial, derivatives with respect to time. A similar argument applies to the continuity equation. Ingard (1968) in his book with Morse, recognized this and corrected the second equations:

$$k\frac{\partial p}{\partial t} = -\frac{\partial u}{\partial x} - u\frac{\partial p}{\partial x} \qquad \frac{\partial p}{\partial x} = -\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right) \qquad (32\text{-}3)$$

Ingard then unwisely asserted that clearly, the previously missing term $u(\partial u/\partial x)$ is always negligible. Unfortunately, in separated flow it is the dominant term.

We have not pursued the solution of these oversimplified, one dimensional, equations except to note the following:

1. Acoustic impedance is inapplicable for predicting propagating flow-pressure relationships in moving air. It probably is still tenable for pure vibratory flows produced by loudspeakers and pressure propagation in semi-static air, but even these should be checked with hot wires.
2. Momentum wave velocities and attenuations will not be the same for the four cases of positive and negative going flow waves traveling up and down stream.
3. Wave velocity for each case is going to be a function of flow energy and surface geometry.
4. Including three dimensional and angular momentum effects will make the solution even more difficult. Unfortunately, the basic unsimplified Navier-Stokes equations are intrinsically unstable and unsolvable.
5. Past conclusions drawn from area function data are not necessarily incompatible with separated flow since we are dealing with wave motions whose speed is at least comparable to the classical velocity of sound.

Apropos of the invalidity of the classical wave equation for moving air, there is an old, unresolved controversy about long distance sound propagation. As some of you may know, one of my heros in the field of fluid-acoustic interactions is H. P. Bouasse, a Frenchman who died in 1953 after a prolific and unsung lifetime devoted to understanding the mechanisms of sound production. Most of his acoustic work was published in the thirties, but in a posthumously published volume (1962) he pondered about many aspects of sound propagation which still puzzled him.

During the war, for example, he could observe German bombing of a village in a valley from a vantage point on the surrounding hills a few miles distant. He wondered why it was that he could easily hear the explosions when the wind was still, or blowing toward him, but could not when the wind was blowing away. In a careful, systematic way, he computed the standard effects that might be responsible: doppler, refraction, scattering, etc., and concluded that under the circumstances they were all of second order, which left only the wind itself as the responsible agent.

We have become increasingly sheltered from the elements, and tend to think of a 10 knot wind as a solid mass moving at its mean speed with a velocity gradient over height. Of course this picture is not true, as any meteorologist working with wind anemometers will tell you. Winds tend to be separated flows, made up of threads and gusts which change quickly in direction, time and space, as any flier knows from his experience with cross wind landings in a a light plane. The explanation for Bouasse's dilema undoubtedly lies in asymmetry of wave propagation in separated flows.

Before going on to the model, let me share with you one other observation, obtained from the front of the teeth trajectories that casts doubt on the oscillating piston approach to mouth sound radiation.

Figure 32-1, is a front of the teeth, left to right trajectory for an "AH". There are many interesting features to note with respect to variability, long term modulations, and actual formant flows compared to the theoretically expected ones. The sidewise flow in the fourth trace clearly shows that the flow is incoherent, i.e., the formant flows are not in step across the mouth exit with respect to direction and amplitude. Further, from similar data taken along trajectories beyond the mouth, flow fluctuations remain quite large over several inches, even as the formant flow is dissipated by collisions and radiation. The old, common notion of having the wind blow away our words is thus probably quite literally true.

THE MODEL

In developing an alternative model for speech production, I shall here omit significant effects in the lungs, and only consider effects at the glottis and in the mouth. A common thread in our model is the variety of non-linear energy feedback and modulation effects that occur in a jet-cavity interaction. There are still major gaps in our understanding, primarily in the dynamics of rotating vortex flows.

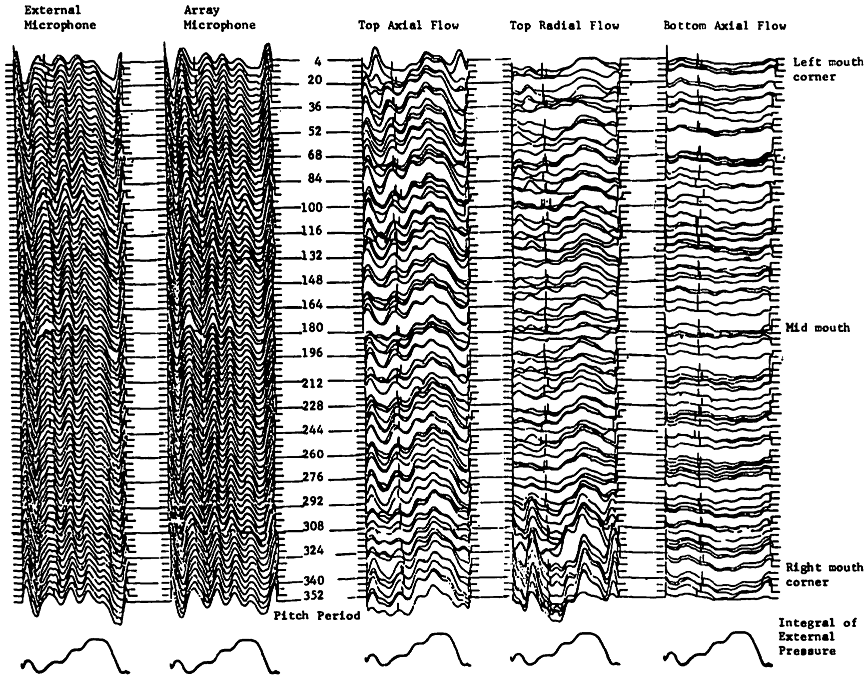


FIGURE 32-1
A front of the teeth, left to right trajectory for a sustained "AH".

First let us consider the salient physics of any unconfined static jet as it emerges from any nozzle, such as the lips of the glottis. For a more rigorous mathematical derivation, the reader should see Schlichting (1968). Let us assume that pressure upstream of the jet is a modest 1/100 of an atmosphere, i.e., 0.15 psi, 10 cm of water, 7.6 mm of mercury, or 10,000 dynes/sq cm, take your pick. The air leaving the nozzle in a direction dictated by lip geometry, will do so with a velocity of 4500 cm/sec, 150 ft/sec, or about 100 miles/hr. At those speeds, the jet will only spread and decelerate slowly as a function of gas composition and ambient geometry. The jet's spread, and thus its deceleration, is very much a function of induced, and other, surrounding air flow patterns. The essential point is that in spreading, the jet preserves total energy and momentum by increasing downstream volume velocity. One cubic centimeter of input 100 mph air can generate roughly 100 cc of 10 mph air downstream. In other words, more comes out than went in, because air can be drawn from the surroundings even though the free jet itself remains at ambient pressure. If the jet passes through a closed cavity, it will lower the cavity's pressure and exhaust it, i.e., the principle of an aspiration pump. Further, if the cavity is at all asymmetric, the jet will differentially exhaust one or the other side and will be thus "attracted" to the nearest wall, but the attachment may be unstable or oscillatory. One cannot have a separated flow without simultaneously setting up conditions for the formation of vortexes, which may be either axial, radial, or both.

For a real or artificial, unconfined glottis, the flow dynamics are obviously far more complicated than we have indicated. For the purpose of a jet cavity interaction, the glottis is providing a time-varying, directional source of separated air flow in the presence of vortexes modifying those flows with their own internal and external dynamics.

A jet-cavity interaction can result in an oscillator, a modulator, or a non-linear filter. It is important that we recognize the conditions necessary to produce each of these elements.

To make an oscillator out of a cavity there is basically one requirement. We merely have to regeneratively add energy when the energy is already high, or remove it over that part of the cycle where the energy is low. Energy can be stored in a cavity in basically two forms, kinetic or potential. Classical acoustics leads us to think of kinetic energy as volume velocity, and potential energy as pressure obtained from differential flow. When we include the jets and vortexes of separated flow, life becomes more complicated. A rotating flow can store kinetic energy, and then return it as pressure, with no differential flow. Since a vortex also carries angular momentum, it will speed up when compressed.

We are used to thinking about waves filling up space uniformly and passing through each other effortlessly. Flow waves however, do not easily pass through stream lines that intersect. Such an intersection acts as a non-linear yielding wall.

Making an oscillator is a question of using the geometry of walls and flows to couple energy into and out of a cavity at the proper time. One of the simplest of jet cavity oscillators is the whistle that we modelled and tested with air and Helium and discussed in the Wisconsin paper. In that whistle, a constant flow, tangential jet is deflected into and away from the cavity, resulting in a relaxation oscillator which alternately builds up and dumps cavity energy. That oscillator, because of a radial, rotating vortex within, also exhibited a strong amplitude modulation at the internal vortex precession rate.

A more common geometric arrangement within the vocal tract is that of a jet moving along a wall into a diverging and converging cavity. It is not difficult to arrange the input jet in such a way that the separated flow along the wall picks up a swirl, or angular momentum along the flow axis, just as rotating water in a sink collects angular momentum because the spinning fluid cannot escape out the central hole. Moreover, this axial vortex at the converging outlet can also act as a non-linear plug. When pressure in the chamber is high, outward flow is impeded by the compressed outlet vortex, but when pressure is low, the vortex expands and allows relatively more exit flow and an increase in vortex strength. Thus the system oscillates.

There are many other geometric arrangements such as the "bird whistle" (two washers in a cylinder) vortex whistles, or convergent coupled flow streams with a single exit such as aeolian whistles, bird syrinxes, etc., which are known to be oscillatory, but require further experimentation and analysis for a complete understanding. However, in each case, the basic energy storage and regenerative coupling elements are present.

It is in the area of energy modulators that separated flows have their greatest impact. Although it is reasonably clear that a separated flow crashing against a wall will instantly convert a large portion of its kinetic energy into pressure, separated flows are usually buffered from such catastrophes by secondary flows. On the other hand, if two non parallel flows are forced to collide, even inelastically, then the conservation of linear momentum will generate square law pressures. If these two flows are modulated by frequencies f_1 and f_2 the collisions can, and do, produce a gamut of non-linear frequencies: $2f_1$, $2f_2$, $f_1 \pm f_2$, etc. If the collisions only occur over part of a pitch period, even stranger formants will result. This can be a rationale for our observation that formant frequencies, not present in the flow, can be seen in output sound pressure.

The foregoing observations about oscillators can be applied to non-linear, time varying filters as well, since the distinction between the two is the degree of regenerative feedback. If the coupling is small, or only occurs over a part of the cycle, the oscillator will be degenerate ... a filter.

Finally, we do not dispute that small perturbations in flow and pressure may well be travelling forwards and backwards in the relatively still air portions of the tract in the classical manner, but the bulk of the energy is dancing to a different tune. We no longer need to look for a glottal flow wave for the generation and supression of harmonics. Geometry, in the form of wall shapes and separated flow effects, can serve quite nicely.

Dr. Kaiser has suggested the word paradigm for jet cavity interaction, and while I may not pronounce it properly, it certainly has a majesty to it which I would hate to have dispelled by my ignorance of what really happens under all conditions when a jet excites a cavity.

I have suggested that this business is a lonely one, and for many years I could only seek colleagues from the past, like Bouasse. In that connection I also owe a debt to Von Kempelen (1791) who has been patronized as an earlier worker with quaint ideas, now obviously obsolete. Not true. I am now building his speaking machine and carrying out many of his suggestions because I find myself in basic agreement with the observations of a pioneering experimentalist, who had a feel for the effects of air flow, even if he lacked modern instrumentation.

Our picture of vocal tract dynamics is still incomplete because in-vitro models for a variety of wall and flow geometries must still be built and tested in order to elucidate specific mechanisms, occuring in different areas of the throat and mouth, and how they interact when the systems are coupled. Such experimentation cannot yet be carried out in a controlled manner with humans, and until we have a far better feel for what is big and what is small; computer simulation of the necessarily simplified Navier-Stokes equations may do more harm than good.

REFERENCES

Beranek, L.L. (1954). *Acoustics*. McGraw-Hill, New York, 21.

Bouasse, H.P. (1962). "Complements de dynamiques des fluides et d'acoustique, (Oeuvres inedites)," Blanchard, Paris.

Fant, G. and Sonneson, B. (1964). "Speech at high ambient pressures," *STL-QPSR* 2, Royal Inst. of Tech. Stockholm.

Morrow, C.T. (1971). "Speech in deep submergence atomospheres," *J. Acoust. Soc. Am.* 50/3.

Morse, P.M. and Ingard, K.U. (1968). *Theoretical Acoustics.* McGraw-Hill, New York, 242.

Morse, P.M. (1936). "Vibrations and Sound," *Am. Inst. of Physics*, New York (1982 reissue), 222.

Schlichting, H. (1968). *Boundary Layer Theory*. McGraw-Hill, New York.

Sondhi, M.M. (1974). "Model for wave propagation in a lossy vocal tract," *J. Acoust. Soc. Am.* 55/5.

Teager, H.M. (1980). "Some observations on oral air flow during phonation," *IEEE Trans. on Acoust. Speech and Sig. Proc.* ASSP-28/5.

Teager, H.M. and Teager, S.M. (1981). "The effects of separated flow on vocalization," in *Vocal Fold Physiol. Conf.*, Madison, WI, College Hill Press, San Diego, CA.

Teager, H.M. and Teager, S.M. (1985). "A phenomenological model for vowel production in the vocal tract," in *Speech Science: Recent Advances*, R.G. Daniloff (Ed.), College Hill Press, San Diego, CA.

von Kempelen, H. (1791). "Le mechanisme de la parole, suivi de la description d'une machine parlante," J.V. Degen, Vienna. English translation in press. Belmont, MA, Venerable Press.

Wathen-Dunn, W. and Michaels, S.B. (1968). "Some effects of gas density on speech production," *Ann. of the N.Y. Acad. of Science, Speech Prod. in Man* 155, 368-378.

DISCUSSION

P. MILENKOVIC: The significant experimental finding that I see here is that the flows in the vocal tract are in the forward direction and that they have an AC component. Perhaps experimenters should measure these flows with a pneumotachograph type device with a wide enough frequency response to attempt to reproduce this finding by independent means.

H. TEAGER: One should be exceedingly cautious in using a pneumotachograph as a standard for air flow measurement. Even disregarding mask effects and frequency response limitations, it is not at all clear that such a pressure sensitive device actually measures pure flow. To my knowledge, calibration with large pulsatile flows have never been performed. (Editorial Comment: Readers interested in the calibration of a mask-type pneumotachograph with pulsatile flows are referred to Rothenberg (1973).)

T. BAER: There are differences of opinion about the validity of the hot wire flow meter measurement technique, especially in a low flow environment. As you know, the generally accepted theories of voice production predict that there should be reversing flow at the lips and in the mouth.

If there are, in fact, problems with nonlinearities or loss of frequency response around low flow velocities (as has been suggested and which appear to be indicated in a DISA catalog), it may be possible to reinterpret your data as showing reversing flow. If the data are reinterpreted in this way, some of the sharp corners that contribute to high calculated acoustic pressures outside the mouth would disappear. Detailed calibration data including the effects of the shroud, should be published. Also, have you taken into account a forward and a backward wave in making pressure calculations?

H. TEAGER: The questions raised regarding constant temperature hot wire anemometry

(CTHWA) are covered in the appendix, however, I should add that there is no such thing as a perfect measuring instrument available to anyone. Even optical instruments have more than twenty kinds of aberrations, yet that does not prevent people from building very precise optical gear. One checks the operation of equipment in regions where errors are suspected. When a negative flow is suspected, one tests for its existence by devising a probe geometry that is insensitive to negative flow and then looking for differences in outputs.

As to the questions regarding pressure calculations, classical acoustic flow waves are exactly matched by acoustic pressure through the proportionality of acoustic impedance in free space, thus a propagation flow and its associated pressure wave have the identical shape. It is the vector sum of forward and backward waves that yields a complex cavity impedance. Even in the case of a "standing wave" in a cavity, where forward and backward traveling waves are of equal magnitude, the peak pressure and peak flows located at nodes and antinodes respectively, have a constant phase relation, and are in the ratio of the free space acoustic impedance, in other words about 40 Rayals.

If measured flows and pressures are not related by acoustic impedance even in this extreme case, it follows that they are not related in the more common case of large forward waves and smaller backward waves.

G. FANT: The reason that you do not detect negative oscillations may be the consequence of rectification - they are absent from the signal, and this is just a limitation of the technique. Most of the items you list in your paper are qualitative in nature, and much of it is very interesting. I would like to encourage you to see how much you can improve existing theories in a quantitative manner. In my view, however, existing theories can account for essentials of speech signal shapes.

H. TEAGER: The acoustic theory cannot reconcile a measured flow velocity of 300 cm/sec with a measured pressure of 100 dynes per square centimeter - the pressure should be 12,000 dynes per square centimeter. A factor of 100 remains unexplained regardless of any postulated artifacts. From the outset, it has been the quantitative divergences of measurements from theory that has convinced us that effects such as flow separation and propagation of kinetic energy are too important to be neglected. It would be helpful if suggestions were made as to precisely what additional experiments and quantitative measurements would be deemed illuminating.

G. FANT: In inverse filtering measurements conducted with a pressure mask, we observe negative flows, and this negative flow comes up in the theory as well.

I have devoted a large amount of time to looking at your helium speech argument. Dr. Lindquist-Gauffin and I wrote a paper on speech in a high pressure helium atmosphere and obtained good experimental agreement with the acoustic theory.

H. TEAGER: With respect to speech with Helium and other gases, the "yielding wall" theory to which Dr. Fant refers was used to explain qualitative formant shifts for high density compressed gases in his and Lindquist-Gauffin's data.

However, in our own experiments with pure atmospheric Helium, as well as "Sealab" measurements with diving gas mixtures, formants above 500 Hz shifted upwards by 20 to 50% less than the expected velocity of sound ratios. Since the effects of a "yielding wall" in Dr. Sondhi's theoretical formulations come into play mainly for low frequency formants, high frequency effects of this magnitude are anomalous. The "yielding wall" hypothesis would be more secure if in-vitro experiments had been run with soft walled tubes, or in-vivo ones which changed the apparent throat thickness, such as phonating while submerged in a filled tub. Since our whistle experiments did produce the expected lower high frequency formant shifts purely as a consequence of separated flow effects, the issue should at least remain open.

M. ROTHENBERG: (Editorial Comment - The following question by Dr. Rothenberg has been abridged by him to preserve continuity of discussion. The issues raised are key issues and the entire question is given in the APPENDIX of this discussion.)

Many of the waveform characteristics that you allude to in your papers can be potentially ascribed to the waveform distortion that can occur when using a constant temperature hot-wire anemometer with small reversing flows. Could you give a more complete discription of the properties of your hot-wire anemometer system in the measurement ranges of interest and how they were obtained, including linearity, frequency response, effects of a shroud, and other distortion, and a technical description of the various components of your circuitry complete enough so that its performance can be verified independently?

H. TEAGER: (Editorial Comment - The following response has been abridged by Dr. Teager. The detailed response to Dr. Rothenberg's question [and to Dr. Baer's question on the validity of the use of a shroud to detect flow direction] is given in the APPENDIX to this discussion section.)

The absolute accuracy of our published flow measurements can be stated simply as follows: the flow vs. voltage calibration shown in our Wisconsin paper is repeatable, has no hysteresis, can be extrapolated down to zero flow, and is valid for all frequencies from 0-5 KHz within a conservative 5% bound on absolute error.

I. TITZE: I appreciate your courage to create this revolution in our field. You have a responsibility to develop a correspondence principle, similar to what one can rely on in relativistic or quantum mechanics, that harmonizes the new theory with the old. After all, much of what we observe has been adequately explained by existing theories. Can you develop a parameter, such as v/c in relativity, which brings the theories into correspondence?

H. TEAGER: Experimentalists have never been held "responsible" for the ultimate consequences of their findings, but are accountable, and have occasionally been praised, for the accuracy and elegance of their experiments. Conservative theoreticians have, on the other hand, generally avoided, or sought ad-hoc solutions to paradoxical findings. All this is very apparent in the development of Einstein's 1905 special theory, whose basis was the 1887 classical Michaelson-Morley null experiment to find the "Ether Drift" implied by Maxwell's equations.

While the essential feature of mass change, time dilation, and length contraction in "correspondence" form were formulated over the next few years by Fitzgerald, Lorentz and others, many classicists refused to believe that the "minor anomaly" was real, or could not be "explained". They never accepted Einstein's formulation based on the invariance of the velocity of light regardless of the relative speed of the observer and source. It was the young, emerging physicists however, who wrote the history. For every case in the development of science where a correspondence was found to bridge a gap, there were many others where, in hindsight, such a bridge could never exist, and a search would have been in vain.

If you are seeking a non-trivial correspondence, I do not know enough to formulate one. Trivially, as flow decreases separation effects disappear, but so does the voice. While years ago I still believed I was looking at a "special case", I have now come to believe that, based upon energy considerations, no correspondence is likely to be found.

The physics of separated flow precludes such an overlap, even if some limited measurements, such as Fourier transforms, are transparent to major distinctions. After all, a cavity resonator is not that far different from a cavity oscillator in the frequency domain and the geometrical parameters governing a resonator can, with little difficulty, be made to fit the frequency produced by a cavity oscillator. To fill in the theoretical gap, our observations must be considered seriously so that resources can be made available for duplication as well as for further work.

APPENDIX to Discussion Section for H. Teager's Paper

M. ROTHENBERG: I assume that you realize that there are a number of properties of the constant temperature hot wire anemometer that could interfere with its operation as a velocity sensing microphone for speech signals. The wire does not indicate when a flow reversal has occured: in electrical terms it provides at best a full wave rectified version of an oscillatory flow. Secondly, at forced convection velocities below 25 to 50 cm/sec, for a 5 micron tungsten wire, the cooling of the wire is increasingly dominated by free convection and by radiation. The linearization of the wire at these velocities may be influenced by air temperature and by other surfaces near the mouth, and it might change if the probe is moved about in the mouth. Also, as illustrated by Dr. Isshiki earlier in the conference, the frequency response of a constant temperature anemometer is degraded at lower flow rates. According to Rasmussen (1965), normal indirect methods for measuring frequency response can be overly optimistic at the flow rates to be expected with a flow microphone.

These problems can, at least in theory, be circumvented by either adding a known steady bias flow from a separate air supply, as Dr. Isshiki has done, or by using a movement of the wire of known velocity to add this bias, as I have attempted to do in experiments using a constant temperature hot-wire anemometer to measure air flow in the mouth. Both these approaches have been used before in other applications, though each has significant limitations when applied to measuring flow within the vocal tract.

In your papers you seem to be satisfied that you have circumvented all of these pitfalls without using a bias flow. In order to understand the significance of your recordings and to attempt to replicate them, I think it would be highly useful for you to supply the following information:

1. A reference to a discussion of the "shroud" method you employ to correct the anemometer output in the presence of reversing flows of magnitudes found in the vocal tract.
2. A description of the distortion, if any, that would remain in the waveform if the shroud method is used, and a physical description of the shrouds used for traces shown in your papers.
3. The lowest flow rate for which the output was linearized, and the method, if any, that was used to verify that the low-flow linearization did not change within the mouth.
4. The lowest flow rates measured in your experiments, the frequency response or response times at those low flow rates, and how they are obtained or estimated.
5. Any technical data not already mentioned in your papers that would be needed to replicate your tests, such as model numbers, type of plug-in units, overheat ratio used, feedback gain where this is adjustable, wire diameter and cable length.

H. TEAGER: The doubts that have been raised about the accuracy of CTHWA measurements must be exhaustively addressed in the light of their bearing on our conclusions. Our IEEE paper and the Wisconsin paper give additional details on all the relevant equipment and settings for duplication as well as static and dynamic calibration procedures.

Dr. Rothenberg's other questions, together with those from Dr. Baer on CTHWA, are answered below.

1. What does Constant Temperature Hot Wire Anemometry (CTHWA) measure?

A constant temperature hot wire probe system detects and measures resistance changes as a function of the convective heat loss due to relative motions of the probe and the surrounding air. Heat radiation and heat conduction are not factors in a properly designed system (Blackwelder, 1981). The hot wire's resistance is a linear function of its temperature which is maintained by a bridge circuit whose error voltage is amplified and fed back to drive the bridge. The system output is repeatable and monotonically 1:1 with the absolute magnitude of the flow velocity as long as wire temperature, wire resistance, wire thickness, ambient temperature and electronic system settings - such as gain- are held constant. If all these conditions are not met, known corrections can be applied. The system's output voltage difference is proportional to the square root of the velocity of flow. If desired, the system's output can be easily linearized, but the system should be calibrated. For static calibration, the flow velocity must be constant, quiet (in other words no "turbulent" flow fluctuations or ambient sounds), uniform over the entire wire area, of constant temperature and exactly known. These conditions are very hard to meet with an air supply at low velocity, and some misleading data have been published because of improper controls. To calibrate velocities less than 1 cm/sec, it is much better to move the wire in still air.

2. Does a significant loss of sensitivity occur at low flow rates? Are there pitfalls to the standard circumventions?

There is no significant loss of sensitivity in the physiological ranges of interest. Some of the expressed concern may stem from statements made in catalogs for simple, low-cost, constant current and hot film systems (DISA Catalog for Type 55D 80/81 Low Velocity Anemometer, p. 32) designed to read an average flow value. These systems are quite different from a CTHWA system and can have low flow sensitivity losses that need linearization. In its linearizer section DISA shows an unlabeled, hand drawn "J" shaped curve (DISA Catalog for Type 55D 80/81, p. 25, figure 25 curve) that clearly shows no sensitivity below 15 cm/sec. From the description of the experimental procedure (Almquist and Legath, 1965) with moving air, it was clear that the air flow was insufficiently controlled. A more controlled study, again made with moving air, shows a loss in sensitivity below 4 cm/sec (Blackwelder, 1959) for a 5 micron wire.

The theoretical reasons given in both papers (Almquist and Legath, 1965; Blackwelder, 1959) for the loss in sensitivity at low velocity is the postulated existence of a self-induced convection flow due to the buoyancy induced movement of the air heated by the hot wire. Both the theory and pertinent parameters are well covered by Blackwelder (1981) and by Compte-Bellot (1976). Basically, such added convection requires space, time, and still air to develop, as in the forced draft in a chimney flue. It is by no means certain that such induced flows can develop for very thin hot wires dissipating minimal heat energy in pulsatile flows.

To explore the low velocity sensitivity of a CTHWA, we set up an adjustable pendulum with a probe at its end, and monitored the output as a function of time. We could detect little deviation from the extrapolated static curve down to velocities of 0.4 cm/sec, at which point room and pendulum induced air currents plus ambient noises became sizeable factors. Obviously, since a CTHWA can be used as a sound detector, albeit at twice frequency, the sensitivity can never go to zero.

The commonly suggested methods to circumvent "sensitivity loss" such as oscillating probes and bias air flows, can lead to very large errors, data loss, and distortions.

An oscillating probe induces large secondary flows as it traverses an area in space. The flows induced in this manner may be comparable to existing separated flows, and may unpredictably interact with them. The area sampled cannot be expected to be isotropic, and in the presence of a pulsating outward flow, an oscillating system will produce misleading measurements.

A bias flow hides a far more insidious pitfall. Opposing air streams are not superimposable. The bias flow in which the sensor is embedded must be in the form of a jet which is itself separated, having a Gaussian spreading velocity distribution. An external flow cannot penetrate the separation easily if the flows are parallel. If the flows are perpendicular, a different characteristic, based on the "stiffness" of the jet, must be accounted for. Thus, because the sensor sees only the "bias flow", the output will be highly nonlinear in amplitude, frequency and relative flow direction. Unfortunately, and worst of all, in an effect known since Tyndall, where it was used as a "flame manometer" (Tyndall, 1867), a jet is highly sensitive to ambient pressures. It deflects, expands and contracts under pressure. Thus a "flow biased" CTHWA cannot cleanly measure flow, even a static one, without the intrusion of pressure effects.

3. Are there significant effects that degrade frequency responses?

As far as we have been able to determine using flow calibrated sirens (out to 4 K Hz) with flow rates in the physiological range of 10-300 cm/sec, frequency response is flat from 0 to 5 K Hz, and is independent of flow rate. There are no theoretical reasons, in a properly adjusted system, for any loss of frequency response below the upper pole of about 7 K Hz in still air. This pole increases with flow velocity, as shown in DISA's published frequency response data. These data are derived from a closed loop response test built into every commercial system by both TSI and DISA. The transient response test is the standard technique both for setting loop gains and ascertaining improper cable lengths and yields a conservative upper frequency.

Rasmussen (1965) asserts a loss of frequency response. In this experiment the author has placed his probe within a "bias flow" which was in fact a narrow, separated jet flow. For the reasons already covered, these measurements were far more indicative of the jet's response and sensitivity than of the hot wire system.

An in-vivo experiment can obviously be run over a range of voice intensities, and in the event of any strange flow-frequency response effects, repeat experiments at low and high voice intensities would show them if they were present.

4. How can reverse flow be distinguished? An elaboration on shrouding.

A CTHWA detects the absolute value of a flow, and it follows that if a flow reverses, a "full wave rectified" type output will be seen. Thus from the "AH" data shown in Figure 1, we concluded that the flow is always outwards due not only to the absence of twice frequency formant flows, but also from the known base line. A further test for the presence of a reverse flow is to note the instantaneous value of the flow measurement at the time of a potential reversal. If the flow is already large and positive, a reverse implies an improbable instantaneous jump to a large and negative flow. If doubts persist, a modified experiment with a shrouded probe can be run which will yield a very different result if reverse flow is present. Reverse

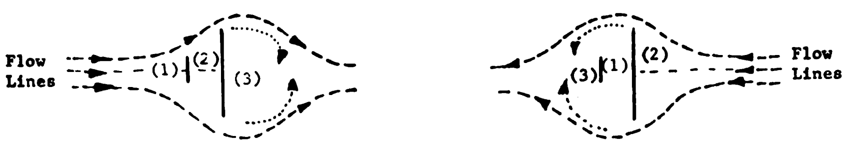


Figure 2 a
Forward Flow

Figure 2 b
Reverse Flow

(1) Hot wire (in plane)
(2) Barrier
(3) Region of vortex instability and minor noise generation from second order reverse flow

FIGURE 32-2.

flows were excluded because repeat modified experiments yielded no difference in output. Only the unshrouded, unlinearized, calibrated data was published. One can only have a reverse flow in the presence of a negative pressure gradient; mouth pressures during phonation are on the order of several mm of water above ambient pressure.

The idea of shrouding is obvious. A variation of our technique is described in the literature (Suhoke and Neuerburg, 1969). The efficacy of shrouding can be checked with pulsating flows from a calibration pump, a siren, or by simple hand motions of the sensor. We are quite aware that acoustic waves, within diffraction limitations, should go around barriers. Pulsatile flow waves however, do not turn corners easily, as illustrated in the simple diagrams in Figures 32-2a and 32-2b. The barrier shroud in 32-2a has little effect, while in 32-2b it acts as a noisy, lossy "diode" for reverse flows.

SUMMARY

The issue of possible artifacts has drawn attention away from measurements and conclusions that do not depend upon the exactitude of our experimental technique. Our earliest and most fundamental observation was that air flow in the mouth during phonation was unidirectional and separated. This is clearly shown in Figure 32-3 taken from a 1974 NIH proposal. In this figure the positive tongue and palate flows, associated with both high and low frequency formants, are clearly 180° out of phase. The baseline flow was 20% of peak, and had small subharmonic pitch period modulations. Our Wisconsin paper covers most of the details of how this data was taken with a single probe. If any interpretation, other than the one we made, is possible with this data, we, and anyone else who has studied it carefully, have been unable to suggest it. The more recent data shown in Figure 1 also indicate that large separated formant flows persist beyond the teeth.

We have affirmed that the seven dimensional (three linear velocity and three angular velocity, plus time) flow patterns are unique to each vowel, and have published four sets of manifestly different trajectory data to back up this postulate.

In this paper we have asserted that the measured formant flow would imply an absurdly high value of mouth pressure. The data on which this assertion is based have been lying dormant in the flow scales published in our 1980 IEEE paper.

Our final assertion, that flow and pressure formant frequencies are not one to one, is borne out by the Fourier analyses shown in our Wisconsin paper, and in the Daniloff chapter. Our work on fluid dynamic mechanisms and modelling has not been presented as fact, but rather as a consistent interpretation of the data shown. We still await suggestions for other, better models.

BIBLIOGRAPHY

Almquist, P. and Legath, E. (1965). "The hot wire anemometer at low air velocities," *DISA Information*, No. 2.

Blackwelder, R.F. (1981). "Hot wire and hot film anemometers," *Methods of Experimental Physics* 18, Fluid Mechanics, R.J. Emrich (Ed.) Academic Press, New York, 259-315, and "Low flow theory," 269-270.

Blackwelder, R.F. (1959). "Two dimensional convection from heated wires at low Reynolds numbers," *J. of Fluid Mech.* 6, 357-384.

Compte-Bellot, G. (1976). "Hot wire anemometry," *Am. Rev. of Fluid Mech.* 8, 209-231.

DISA Catalog (1982), Type 55D 80/81 Low Velocity Anemometer (p. 32), Figure 25 (p. 25).

Neuerburg, W. (1969). "Directional hot wire probe," *DISA information*, No. 7.

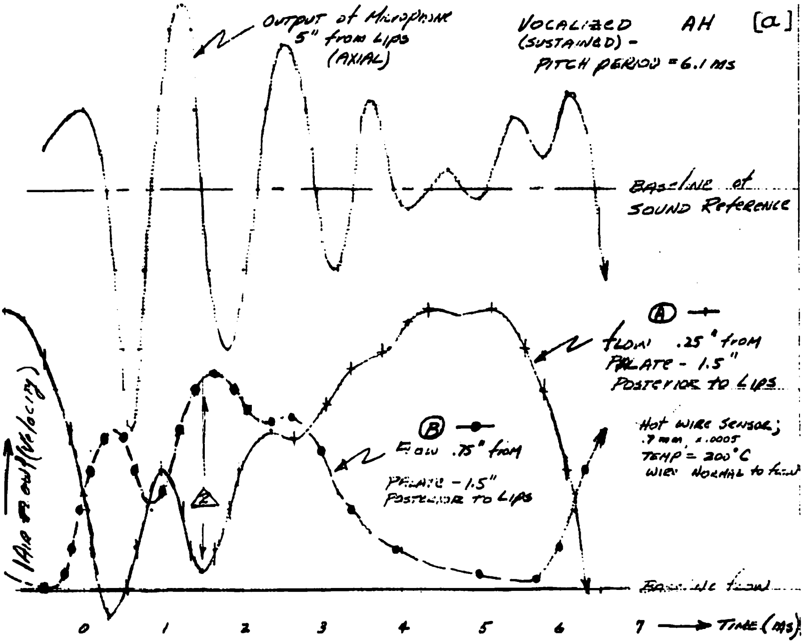


Fig. 3 - EXPERIMENTAL VERIFICATION OF SEPARATED FLOW.

1. SIMULTANEOUS FLOWS Ⓐ + Ⓑ, (SYNCHRONIZED WITH SOUND WAVE) ARE NON-SIMILAR.
2. FORMANT PERTURBATIONS OF Ⓐ AND Ⓑ ARE EXACTLY 180° OUT OF PHASE
3. FLOW AT EXIT OF NOSE AND MOUTH (NOT SHOWN) ARE SIMILAR TO, + IN PHASE WITH Ⓐ, BUT FORMANT PERTURBATIONS ARE FAR SMALLER (10% AS OPPOSED TO 50% OF PEAK FLOW DEVIATION).

INTERPRETATION:- THE DATA ARE CONSISTANT WITH A PULSED JET WHOSE AVERAGE FLOW AXIS IS CLOSE TO THE PALATE BUT WHOSE DIRECTION IS PERTURBED AT THE FORMANT FREQUENCY

10/74

FIGURE 32-3. Experimental verification of separated flow. From a 1974 NIH proposal.

Rasmussen, C.G. (1965). "The air jet hot-wire microphone: Application of hot-wire techniques to the measurement of steady and alternating partical velocity in air velocities below 2 m/sec," *DISA Information* No. 2, 5-13.

Rothenberg, M. (1973). "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," *J. Acoust. Soc. Am.* 53, 1632-1645.

Suhoke, R. and Neuerburg, W. (1969). "Directional hot wire probe," *DISA Information* No. 7, January.

Tyndall, J. (1867). *Sound.* New York.

33 ▪ SOME RESULTS ON THE ACOUSTIC AND AERODYNAMIC FACTORS IN PHONATION

Tirupattur Ananthapadmanabha and Jan Gauffin

ABSTRACT

Vocal fold vibrations are caused by the forces created by the flow of air through the glottis. In order to maintain such vibrations there has to be a transfer of energy from the flow to the vibrating folds. A necessary condition is then an asymmetry in the development of aerodynamic forces during the opening and closing phase of the vibratory cycle. There are mainly two mechanisms that can create such asymmetries. One is the effect of the acoustic loading of the glottis by the sub- and supraglottal cavities, and the other is the characteristic cross-sectional shape variation of the vocal folds during the glottal cycle. In this paper we present some results from calculations of the acoustic loading and discuss some results from composite models of the glottis.

INTRODUCTION

When air flows through the narrow glottis, the air accelerates exerting lateral forces (Bernoulli and convective forces) on the vocal folds. In response to these forces the vocal folds move and thus change the area and geometry which subsequently changes the forces, and so on. The pressure drop across the glottis will also vary during the vibratory cycle due to the effect of the acoustic loading by the sub- and supraglottal cavities. It is clear that such a complicated system is extremely difficult to describe in an accurate model. We have to make simplifications as are frequently made in works dealing with these problems. Examples of such simplifications are the use of one-dimensional flow theory in glottal aerodynamics, the two-mass model for vocal fold vibrations and the use of "equivalent area" (see Gauffin *et al*, 1981) for calculation of the relations between flow and area function.

The theoretical foundations for modeling the aerodynamics in the glottis are not well defined and we have to rely on empirical data. The classical experiments by van den Berg *et al* (1957) and the theory of phonation, van den Berg *et al* (1959), was a big step forward in the theory of phonation, but had important shortcomings. He used several simplifying assumptions: (a) the use of a rectangular glottis model, (b) the use of a single empirical formula for the one-dimensional flow, and for a wide range of pressure and flow values, (c) ignoring the effect of subglottal and vocal tract loads etc. Since his main interest was to capture first order effects, he considered the additional factors being of minor importance. He underestimated the acoustic loading effects of the trachea and the vocal cavity by comparing the dc resistance of the loads with the glottal resistance. Flanagan and Landgraf (1968) incorporated van den Berg's data in a one-mass model and found that the acoustic loading did not counteract the aerodynamic forces.

It is now well known that the vocal fold vibratory pattern can have a vertical phase difference in the movement of the upper and lower margins (Figure 33-1). In the one-mass model based on van den Berg's theory this is not represented. The mechanisms and functional role of this vibratory pattern are not yet fully understood. The question now arises if a model can be developed around van den Berg's basic concepts that can explain the characteristic

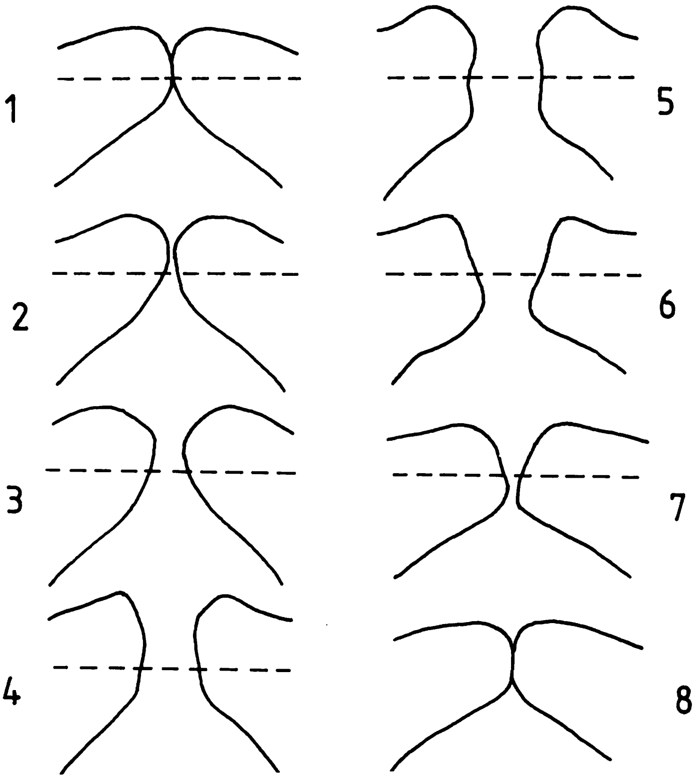


FIGURE 33-1
Schematic presentation of the glottal shapes in one vibratory cycle. Adapted from Hiroto (1980).

patterns of vocal fold vibrations and their significance. A promising approach in this direction is that of Ishizaka and Matsudaira (1972). Broad (1979), considered their theory as a conceptual breakthrough and provided a tutorial review on this subject.

Let us recall van den Berg's classical formula for the Bernoulli equation for the glottis:

$$\frac{k_1 \rho U_g^2}{2A_g^2} + \frac{12\mu dl}{A_g^3} U_g - \frac{k_2 \rho U_g^2}{2A_g^2} = P_{tg} \qquad (33\text{-}1)$$

where A_g is the glottal area, U_g is the volume velocity flow, P_{tg} the transglottal pressure, k_1 the entry coefficient, k_2 the exit recovery coefficient, ρ the air density, μ the viscosity, d the glottal depth, and l the glottal length.

According to the theory of Ishizaka and Matsudaira, when reduced to practical terms, the magnitude of the exit recovery coefficient k_2 is of the order of 0.1 and not 0.5, as given by van den Berg. Ishizaka and Matsudaira also consider a two-mass model. If the areas between the two masses are A_1 and A_2, and since the velocity particle of air is continuous across the sections, there has to be a pressure drop across the sections given by:

$$P_{12} = \frac{\rho U_g^2}{2} \left[\frac{1}{A_2^2} - \frac{1}{A_1^2} \right] \qquad (33\text{-}2)$$

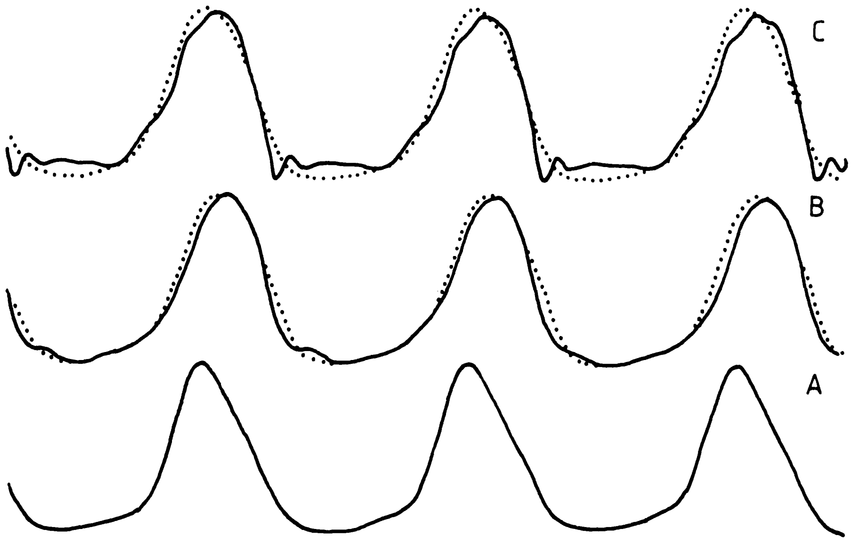


FIGURE 33-2
A: Glottal area from photoglottograph, B and C: Full lines are volume velocity from inverse filtering with flat spectrum criteria (B) and flat closed phase criteria (C). Dotted lines are calculated curves.

The entry coefficient k_1 is still assumed to be 1.37. The overall equation for the pressure drop across the glottis for static flow is given by:

$$\frac{0.37\rho U_g^2}{A_1^2} + \frac{\rho U_g^2}{2A_2^2} = P_{tg} = \frac{\rho U_g^2}{2}\left[\frac{0.37}{A_1^2} + \frac{1}{A_2^2}\right] \qquad (33\text{-}3)$$

where, for simplicity, we have omitted the exit recovery coefficient term and the viscous resistance term. In Table 33-I some examples are given of the effective kinetic resistance coefficient for converging and diverging glottis in a hypothetical two-mass model. As can be noticed this coefficient is very much dependent on the geometry of the glottis.

TABLE 33-1

Area ratio		Geometry	
	converging		diverging
1		1.37	
1.5	1.17		0.82
2.0	1.09		0.62

In the theory of Ishizaka and Matsudaira there is a step change in pressure within the glottis. If a continuous geometrical variation in the area is assumed, then Broad (1979) has shown that:

$$P(Y) = P_{sg}\left[1 - A^2(d)/A^2(y)\right] \tag{33-4}$$

where P(Y) is the pressure within the glottis at a section 'Y', A(Y) is the area of section 'Y', and A(d) is the area of the glottal outlet. Broad refers to the pressure immediately below the glottis as tracheal pressure denoted by P(T) and the pulmonic pressure is denoted by P_{sg} in his paper. With the section 'Y' chosen to coincide with the glottal entry, the subglottal pressure P_{sg} can be obtained. According to Broad, the subglottal pressure varies as the negative of the square of the glottal outlet area. He quotes Koike and Hirano's (1973) simultaneous measurements of glottal area and subglottal pressure as a validation of his derivation.

Ishizaka and Flanagan (1972) have incorporated the theory of Ishizaka and Matsudaira into a two-mass synthesis model with acoustic loading with which they have obtained many interesting results. In particular for our discussion, their simulation of subglottal pressure variations agrees well with the experimental recordings by Koike and Hirano (1973). The simulations of Ishizaka and Flanagan give results which show the combined effects of aerodynamic and acoustic factors. It is not possible to separate out the contribution from each of them.

Scherer (1981), Binh and Gauffin (1983) have made extensive measurements on models of the glottis with uniform and nonuniform geometries. They have obtained similar results. Their findings validate the formula for the exit recovery coefficient given by Ishizaka and Matsudaira. The intraglottal pressure drop profile based on a two-mass model is, however, not supported by the experimental data.

Titze (1980) is also critical of the two-mass model, especially regarding the aerodynamic forces maintaining the vocal fold vibrations. Some of the proposals of Titze, such as impulsive forces, may not be realistic. But, our interpretation of Titze is of a general philosophical concern. That is, any mechanical oscillator with low damping can sustain oscillations provided there is a net energy input to the system. Since the lung pressure is steady over a glottal cycle, there has to be an asymmetric driving force, or pressure distribution. This asymmetric force can be impulsive or distributed in time and may arise due to aerodynamic or acoustic factors. For a multi-element oscillatory system, one could start with any reasonably assumed hypothetical force changes and derive conditions for sustained oscillations (see Broad, 1979). Due to the lack of experimental support for the empirical formulae of Ishizaka and Matsudaira, and since the theory is applicable for small amplitude oscillations, Titze suggests looking into other than aerodynamic factors.

In the present paper we will first start out with the assumption that the glottal geometry is unimportant, mainly to provoke discussion, and then show the influence of acoustic factors. In the later section we will discuss the aerodynamic factors in connection with nonuniform glottal shapes and we will also discuss the relevance of the two-dimensional flow theory in connection with the different problems we want to solve.

ACOUSTIC FACTORS

In this section we will calculate the volume velocity flow through the glottis and the supraglottal and oral pressure variations starting out from the glottal area function. This approach is not new. Flanagan, for instance, used the measured glottal area function to numerically obtain the glottal flow (1968) and the intraglottal pressure variation (1959). Hiki *et al* (1970)

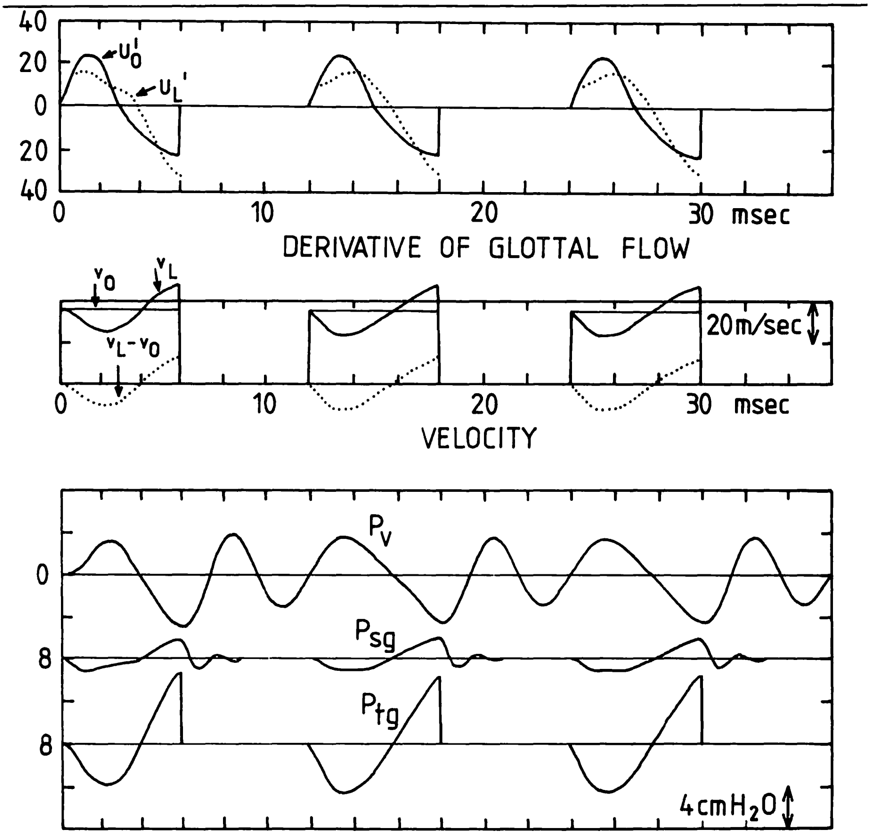


FIGURE 33-3
Model calculations of the short circuit and true glottal flow derivatives, particle velocity with and without load and the subglottal and oral pressure variations. Vowel /i/.

have used the ac impedance of the vocal tract and the subglottal system at the fundamental frequency.

Rothenberg (1980) proposed the use of glottal conductance and an inductive load. Fant (1979) gave an analytical approach based on assumed parametric glottal short circuit flow. Ananthapadmanabha and Fant (1982) gave both a numerical and analytical approach.

Ananthapadmanabha and Löfqvist (unpublished work) recorded simultaneously the glottal area (photoglottography) and the volume velocity through the lips by means of Rothenberg's mask. This glottal area function corresponds to a hypothetical one-mass model, the area of which is the projected area of the nonuniform glottis (minimum area). The following Bernoulli equation within the glottis is used:

$$\frac{1.1\rho U_g^2}{2A_g^2} = P_{tg} = P_L - P_T(t) - P_V(t) \qquad P_{sg}(t) = P_L(t) - P_T(t) \qquad (33\text{-}5)$$

where P_L is the lung pressure, P_T is the drop in trachea, and P_V is the drop due to the vocal tract load. The viscous resistance and glottal inertance terms have been omitted. The use

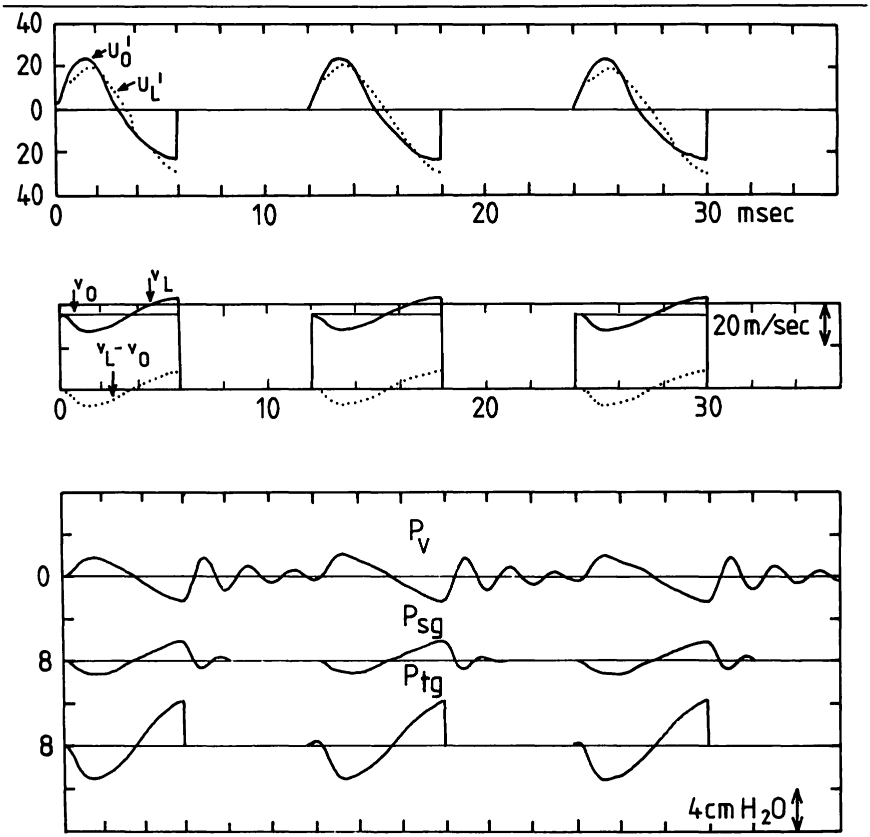


FIGURE 33-4
Model calculations of the short circuit and true glottal flow derivatives, particle velocity with and without load and the subglottal and oral pressure variations. Vowel /ε/.

of a fixed kinetic resistance coefficient implies that the effect of glottal geometry is neglected. A lung pressure of 10 cm H_2O and appropriate vocal tract and subglottal load elements are used. The glottal area is scaled so that the peak area corresponds to 20 mm^2. Based on the interactive model of Ananthapadmanabha and Fant (1982), the true glottal flow is calculated. The mask output is carefully inverse filtered using the INA interactive analysis program by J. Liljencrants at KTH. Two different criteria were used in the inverse filtering: 1. a flat closed phase criterion and 2. a smooth spectrum of the source. The calculated glottal flow is compared with the measured glottal flow in Figure 33-2. It can be seen that there is a good agreement between the two flow functions.

Based on Equation 33-5, the subglottal and oral pressure variations can be obtained as side products. These are shown for a parametric model of the glottal area function in Figures 33-3 and 33-4. The short circuit and true glottal flow derivatives, with load and no load particle velocity curves are also shown. For comparison, the measured pressure curves by Löfqvist and Kitzing are shown in Figure 33-5. The general agreement between the measured and calculated curves is good. For the subglottal pressure, an approximate relation is given by:

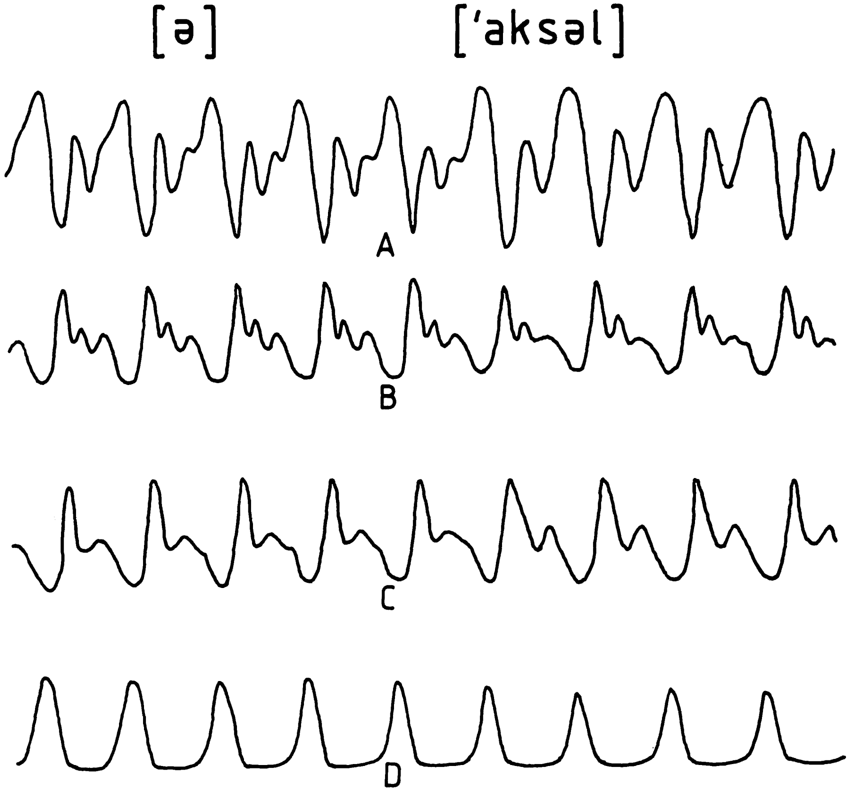


FIGURE 33-5
Simultaneous measurement of oral pressure, subglottal pressure, and glottal area. A. Oral pressure, B. Subglottal pressure, C. Transglottal pressure (B-A), D. Glottal area.

$$\begin{aligned} P_{sg}(t) &= P_L - L_{sg}\, dU_g(t)/dt \simeq P_L - L_{sg}\, dU_{sg}(t)/dt \\ &\simeq P_L - (\text{a constant})\, dAg(t)/dt \end{aligned} \tag{33-6}$$

Thus the subglottal pressure varies as the negative of the derivative of the glottal area function. This is in contradiction to earlier statements, e.g., Broad's citing of Koike and Hirano's measurements, where it is said that the subglottal pressure varies as the negative of the square of the area. Fant (1982) has derived an analytical solution to the combination of Equations (33-6) and (33-5).

Both the above calculations show that even when the aerodynamic effects of nonuniform glottal geometry are neglected, very good agreements between calculations and measurements can be obtained for true glottal flow and sub- and supraglottal pressure variations. However, it is wrong to draw the conclusion that glottal geometry factors have little effect. The glottal geometry determines the intraglottal pressure distribution and the mechanical stresses on the tissues.

AERODYNAMIC FACTORS

From the above discussion we can state the following:

A. The agreement between the experimental findings and the theory is reasonable for calculating the terminal variables, such as the glottal flow, subglottal and oral pressure measurements.

B. The intraglottal pressure drop profile cannot be predicted based on formulae derived using the one-dimensional flow assumption. This data can be obtained only experimentally, but, to derive theoretical formulae the two-dimensional flow theory has to be used.

C. Any theory on vocal fold vibration which requires the intraglottal pressure variations cannot be validated based on measurements of the terminal variables or synthesis results.

If the two-dimensional flow theory is used, the flow lines have to be known. There is no simple theory available for constructing these flow lines and for diverging sections of the glottis we have to rely on model experiments alone. This is especially true when separation of the flow occurs inside the glottis. We also have to piecewise approximate the cross-sectional shape with simple geometrical shapes. One such approximation may be using straight lines and circular segments. For converging sections, where the flow can be considered to be irrotational, we can then construct a first approximation of the flow lines using some simple rules, see Rouse (1946). These rules are that concave and converging boundaries (a in Figure 33-6) tend to create stagnation along the surface with higher velocity at the midline while converging areas with flat (b) or convex boundaries (c) tend to equalize the velocity distribution normal to the flow. Parallel, and in particular, diverging boundaries always create a rotational flow with concentration of the flow to the midline, and convective acceleration tends to separate the flow from the wall. An example is given in Figure 33-6B. In section (a) we should expect a slight increase in pressure along the boundary due to retardation of the flow and the convective acceleration is low. In section (b) the flow velocity is rapidly increasing and the velocity distribution normal to the flow is getting more and more uniform and at the exit of section (c), the flow profile will be approximately as indicated. Because of retardation of the flow along the boundary in section (d), the flow will not be irrotational and we cannot any more compare calculated pressures on different flow lines.

In the diagram at the bottom of Figure 33-6, the expected piezometric head along the center line and the boundary is indicated. The difference between these two pressure curves is, as explained above, caused by the velocity distribution and in section (c), where the velocity is high and the radius of the curvature is small, by convective acceleration. A decrease in the pressure along the center line is indicated in section (d). This is because we should expect a net acceleration in this section caused by the difference in velocity profile between the entry and the exit of that section.

At the abrupt exit of the glottis a laminar jet is created. The pressure on the surface of this jet is the same as immediately above the glottis. In Figure 33-6B, a diverging glottal exit is indicated. In this example the particle velocity is too high for the flow to follow the boundaries. Generally the flow resistance will then increase somewhat due to the eddies created at the exit. In glottal models this will be signaled by a higher noise level. If the expansion had been more gradual, the flow may have followed the boundary and some of the kinetic energy could have been gained back, which, in contrast, should result in a reduced total flow resistance. Even if the glottis is not divergent, separation can occur due to the convective acceleration if the entry section (c) has small enough radius.

To illustrate some of the above effects a series of experiments with a simple glottal model has been performed. In Figure 33-7, the differential pressure-volume velocity curves are shown

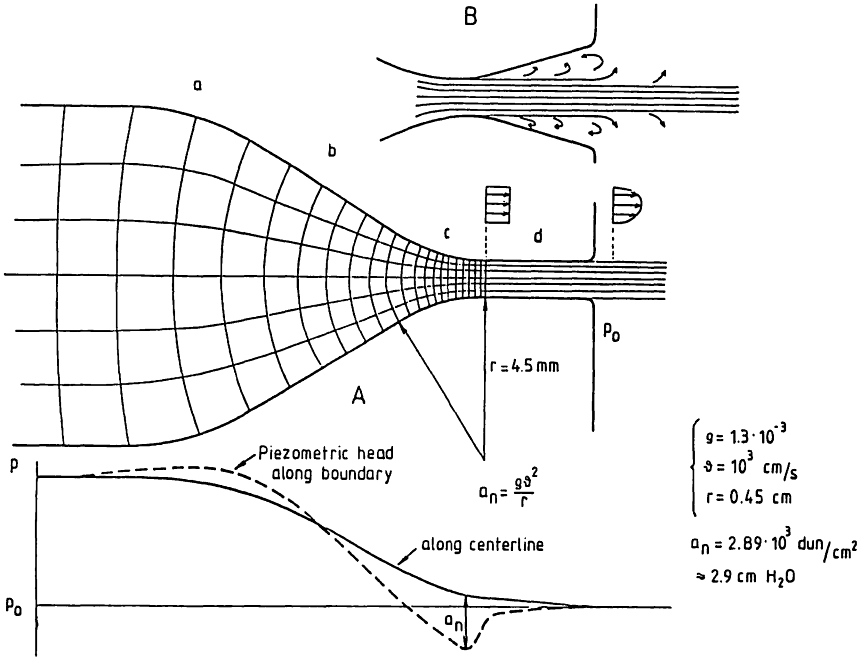


FIGURE 33-6
Flownet for a constriction similar to the glottis.

for five different configurations with the same minimum area. As can be seen (b) has a higher resistance than (a) and (c) because of the separation and instability of the flow already at the abrupt entry. By combining sections (a) and (b) as in (d), it is possible to make the glottal model more streamlined with less tendency for separation. The flow resistance then falls appreciably. A combination of (a) and (c) has a similar effect but only to a much smaller extent. Adding a third section at the inlet did not make any measurable difference. This indicates that the vocal fold shape at entry is only important up to a few times the minimum glottal area.

The recovery of energy downstream the glottis was very small in our models. When a vocal tract tube with a diameter of 2 cm was attached, a lowering of the flow resistance could only be found when the glottis was divergent, and it amounted to less than 4% for a glottal diameter of 0.4 mm. If a more accurate model of the vocal tract had been used with false vocal folds etc., probably more energy could have been recovered. This implies that a constant exit recovery coefficient cannot be used.

CONCLUSIONS

An urgent need is to set up critical standards to evaluate different models of the larynx. Various researchers with different formulations all claim to have obtained realistic results. First order effects can be achieved in most models. Also synthesis is not a good criterion for evaluation. Physiologically, we need to know more about the stresses on the vocal folds

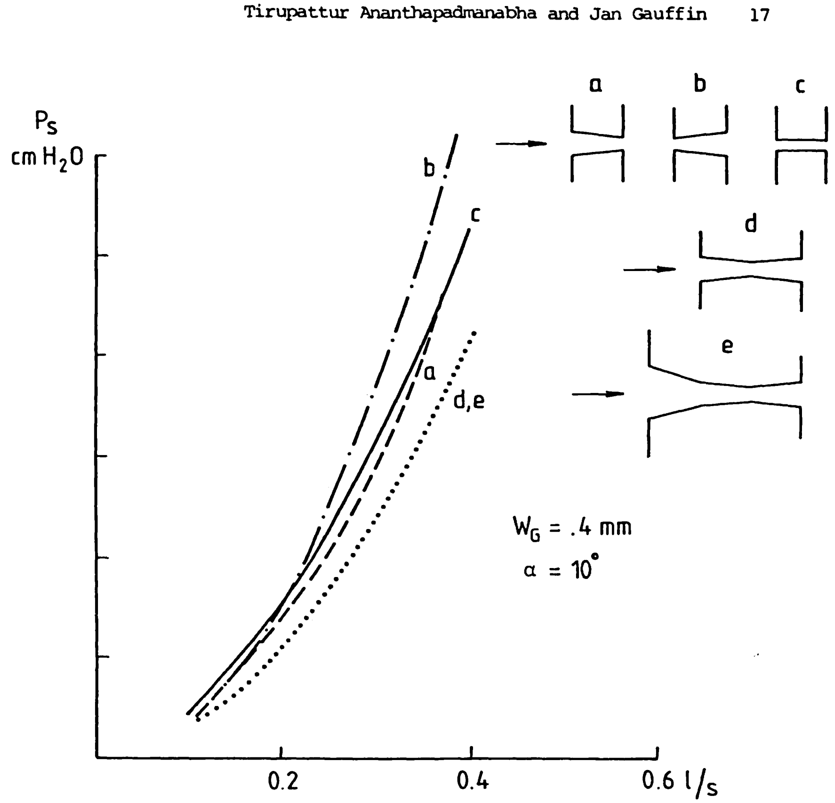


FIGURE 33-7
Measured pressure drop and volume flow in some glottal models with the same minimum area.

and the effect of the neuromuscular control. Computer simulations obscure the interrelations between aerodynamic, acoustic, muscular, and geometric factors, and tend to provide either geometrical patterns or synthetic speech. We need analytical modeling approaches, especially in the aerodynamic and acoustic stages, so that we gain a deeper insight into this fascinating mechanism, the larynx.

We feel, that for constructing better vocal fold models, we have to build our new theories around the flowline concepts sketched above.

REFERENCES

Ananthapadmanabha, T.V. and Fant, G. (1982). "Calculations of glottal flow and its components," *STL-QPSR* 1, 1-30.

van den Berg, Jw., Zantema, J.T. and Doornenbal, P. (1957). "On the air resistance and Bernoulli effect of the human larynx," *J. Acoust. Soc. Am.* 29, 626-631.

van den Berg, Jw., Zantema, J.T. and Doornenbal, P. (1959). "Myoelastic-aerodynamic theory of voice production," *J. Speech and Hear. Res.* 1, 227-243.

Binh, N. and Gauffin, J. (1983). "Aerodynamic measurements in an enlarged static laryngeal model," *STL-QPSR* 2-3, 36-60.

Broad, D.J. (1979). "The new theories of vocal fold vibration," in *Speech and Lang.: Advances in Basic Res. and Practice*, Vol. 2, Academic Press, 203-256.

Fant, G. (1979). "Glottal source and excitation analysis," *STL-QPSR* 1, 85-107.

Fant, G. (1982). "Preliminaries to analysis of the human voice source," *STL-QPSR* 4, 1-27.

Flanagan, J.L. (1959). "Estimates of intraglottal pressure during phonation," *J. Speech and Hear. Res.* 2, 168-172.

Flanagan, J.L. (1968). "Source-system interaction in the vocal tract," *Ann. of NY Acad. of Sciences* 155, 9-17.

Flanagan, J.L. and Landgraf, L.L. (1968). "Self-oscillating source for vocal tract synthesizers," *IEEE Trans. Audio and Electroacoust.* AU-16, 57-64.

Gauffin, J., Binh, N., Ananthapadmanabha, T.V. and Fant, G. (1981). "Glottal geometry and glottal flow," in *Proc. of the Vocal Fold Physiology Conference*, Madison, WI.

Hiki, S., Koike, Y. and Takahashi, H. (1970). "Simultaneous measurement of subglottal and supraglottal pressure variation," Paper, 79th meeting, ASA, DD4.

Ishizaka, K. and Flanagan, J.L. (1972). "Synthesis of voice sounds from the two mass model of the vocal cords," *Bell Syst. Tech. J.* 51, 1233-1267.

Ishizaka, K. and Matsudaira, M. (1972). "Theory of vocal cord vibrations," *Rep. Univ. Electro Commun.* 23, 107-136.

Koike, Y. and Hirano, M. (1973). "Glottal-area time function and subglottal-pressure variation," *J. Acoust. Soc. Am.* 54, 1618-1627.

Rothenberg, M. (1980). "Acoustic interaction between the glottal source and the vocal tract," *Proc. of Vocal Fold Physiology Conference*, Kurume, Japan.

Rouse, H. (1946). *Elementary Mechanics of Fluids*, Dover Publication, New York, 1978. Ch. II, VII.

Scherer, R.C. (1981). "Laryngeal fluid mechanics: steady flow considerations using static model," doctoral diss., Univ. of Iowa.

Titze, I.R. (1980). "Comments on the myoelastic-aerodynamic theory of phonation," *J. Speech and Hear. Res.* 23, 495-510.

DISCUSSION

R. SCHERER: The shape of the glottis, if there is a steady flow, will be important in determining the kind of flow regime within the glottis. Fluid mechanics literature indicates that for various geometries, either separation or attachment occurs, that is different stall regimes occur. It seems that these regimes take time to stabilize.

Are you, or perhaps anyone else in attendance, familiar with the kinds of time periods that would be necessary to establish the kinds of flows we are studying in the models? Or perhaps someone would want to comment on the validity of the models when we are not measuring the flows, only the pressures. All of this is a very crucial aspect in phonation and the resulting pressures acting on the surfaces.

J. GAUFFIN: I do not have an immediate answer to that question. I should expect the answer to have something to do with the particle velocities and the size of the model. If the velocity is very high and the model very small, one should expect flows to stabilize quickly.

A. BENADE: I don't have an answer, but perhaps this may help. The alto recorder works on the principle of an air jet that comes out over an abrupt break after which the air blows right across. If you round the corner to a couple of millimeters radius, you get a partial at-

tachment either flowing in or out. If you can start this configuration to oscillate, it will continue to oscillate once you get it going, but the flow tends to hang up on one side or the other.
J. GAUFFIN: We have to remember that the glottis is very small, and we have a very small radius of curvature in the flow as well as a high particle velocity. The flow separation should come much more easily in the glottis than in our scaled up models.

34 ▪ AIR RESISTANCE AND INTRAGLOTTAL PRESSURE IN A MODEL OF THE LARYNX

Kenzo Ishizaka

INTRODUCTION

Measurements of translaryngeal pressure drop and volume flow through models of the larynx having rectangular and diverging glottal ducts have recently been made (Scherer *et al*, 1981, 1983 and Gauffin *et al*, 1981). The measured data on the shale of a rectangular duct, (Scherer *et al*, 1983) are in good agreement with those calculated from the equations given by Ishizaka and Matsudaira (Ishizaka and Matsudaira, 1972).

The measured data on the diverging duct with a considerably large diverging angle of about 42°, however, show difference from those given by the equations of Ishizaka and Matsudaira, in which the diverging and converging ducts are approximated by the two parallel ducts (2-section approximation).

In this paper we will show that when the losses due to the area change in the diverging and converging ducts will be taken into account our equations can well match the measured data.

COMPARISONS BETWEEN MEASURED AND CALCULATED DATA

Scherer and Titze (1981) have shown several comparisons between their measured data

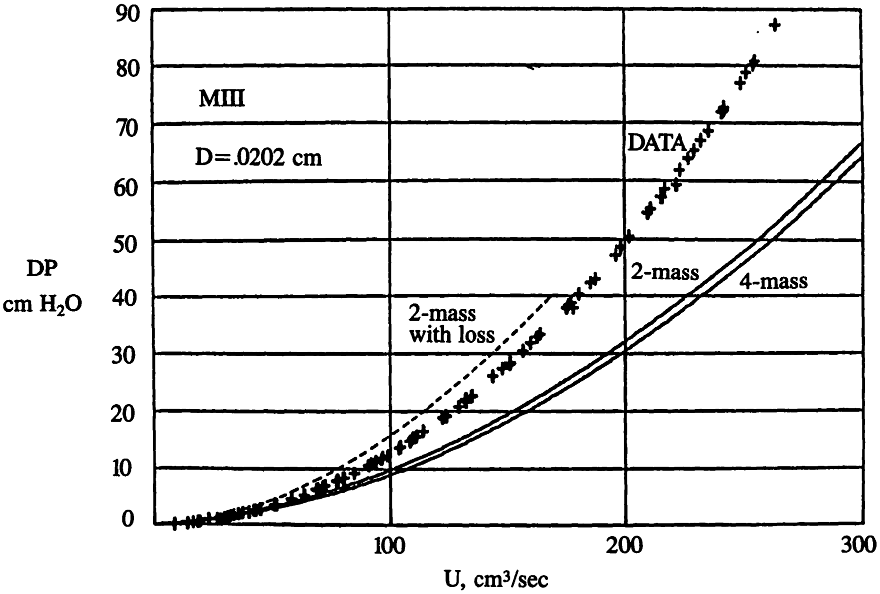


FIGURE 34-1
Translaryngeal pressure drop versus volume flow for a Model MI (after Scherer and Titze, 1981).

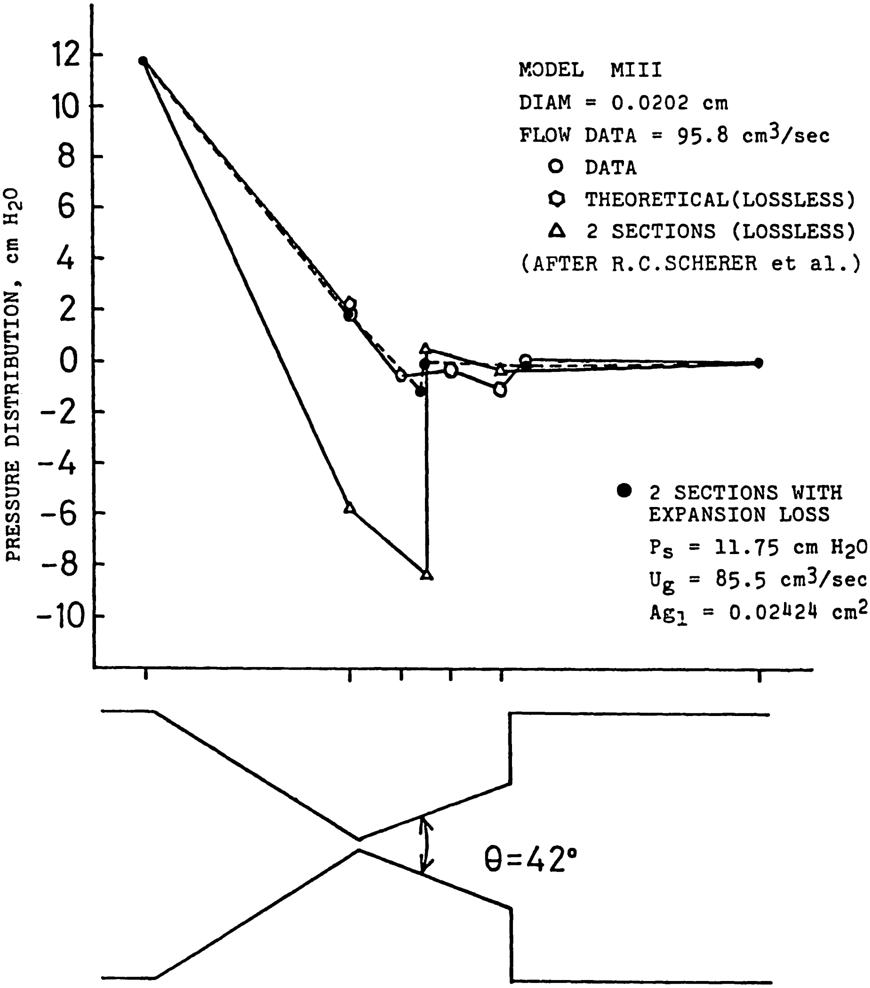


FIGURE 34-2

Pressure profiles for a Model MIII with glottal diameter d = 0.0202 cm (after Scherer and Titze, 1981).

on a MIII model having the diverging angle of 42° and calculated data from the equations of Ishizaka and Matsudaira, in which the losses due to the area change in the glottal duct are neglected (the equations for $\eta = 0$, where η is the loss factor).

Figures 34-1 through 34-4 are their figures in which the measured data are compared with the data calculated from the equations for $\eta = 0$. In the same figures also shown are the corresponding data calculated from the new equations involving the loss due to the area change in the diverging duct. In the present calculation, the loss factor η for the diverging duct is fixed at unity. (Our present purpose is not to fit our calculation exactly to their measured data but to show how the losses can be involved in our previous equations for $\eta = 0$ for better matching.)

Figure 34-1 shows the translaryngeal pressure drop vs volume flow. The calculated results

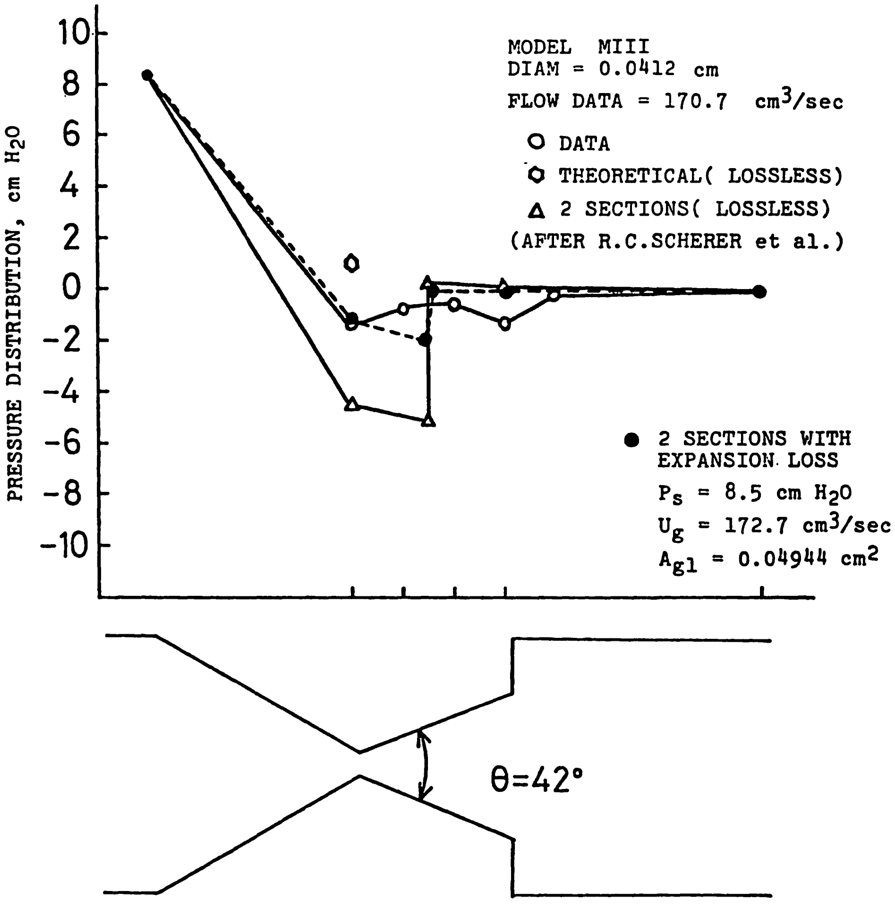


FIGURE 34-3
Same as Figure 34-2 with d = 0.0412 cm.

for $\eta = 1$ are shown by the dotted line. Figure 34-2, 34-3, and 34-4 show the pressure distribution in the model MIII with the glottal diameter 0.0202, 0.0412, and 0.0614 cm, respectively. The calculated results for $\eta = 1$ are shown by the closed circles in the figures. The calculated data for $\eta = 1$ are apparently closer to the measured data than those for $\eta = 0$, as seen in the figures.

LOSSES DUE TO THE AREA CHANGES IN THE GLOTTAL DUCTS

In the formulation of our previous equations, we have considered the losses due to the changes in cross-sectional areas as well as viscous losses. One of such losses is the loss caused by the sudden contraction of flow at the inlet of the glottis of small area from the trachea of wider area. Another is the sudden expansion loss in the upper larynx tube, into which the air flow suddenly emerges with high velocity from the glottis of a small area.

We, however, assumed that the difference in areas at the inlet and outlet of the glottal

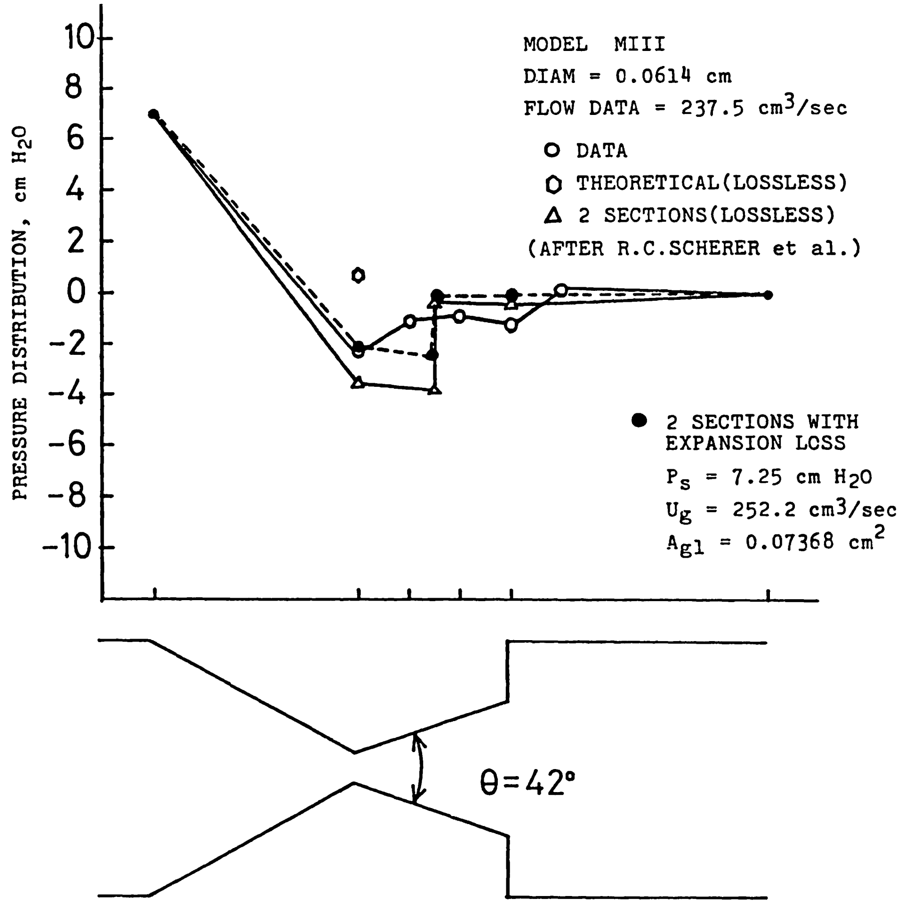


FIGURE 34-4
Same of Figure 34-2 with d = 0.0614 cm.

duct might relatively be small. The losses caused by the area changes in the glottal duct were then omitted for simplicity. It is apparent that if the diverging and converging angles are not small enough this assumption can no longer be valid.

The pressure losses in this case can simply be involved in the previous equations for $\eta = 0$ in the same manner as the losses caused by the sudden contraction at the inlet of the glottal duct and the sudden expansion of flow at the outlet of the glottis. The Bernoulli's equation for steady flow in the diverging and converging glottal duct becomes

$$p_1 + \rho v_1^2/2 = p_2 + \rho v_2^2/2 + h_k + h_v$$

where p and v are pressure and average particle velocity respectively, the subscripts 1 and 2 denote respectively the inlet and outlet of the glottal duct, h_k denotes the pressure drop due to the area change, and h_v the pressure drop due to viscosity between 1 and 2.

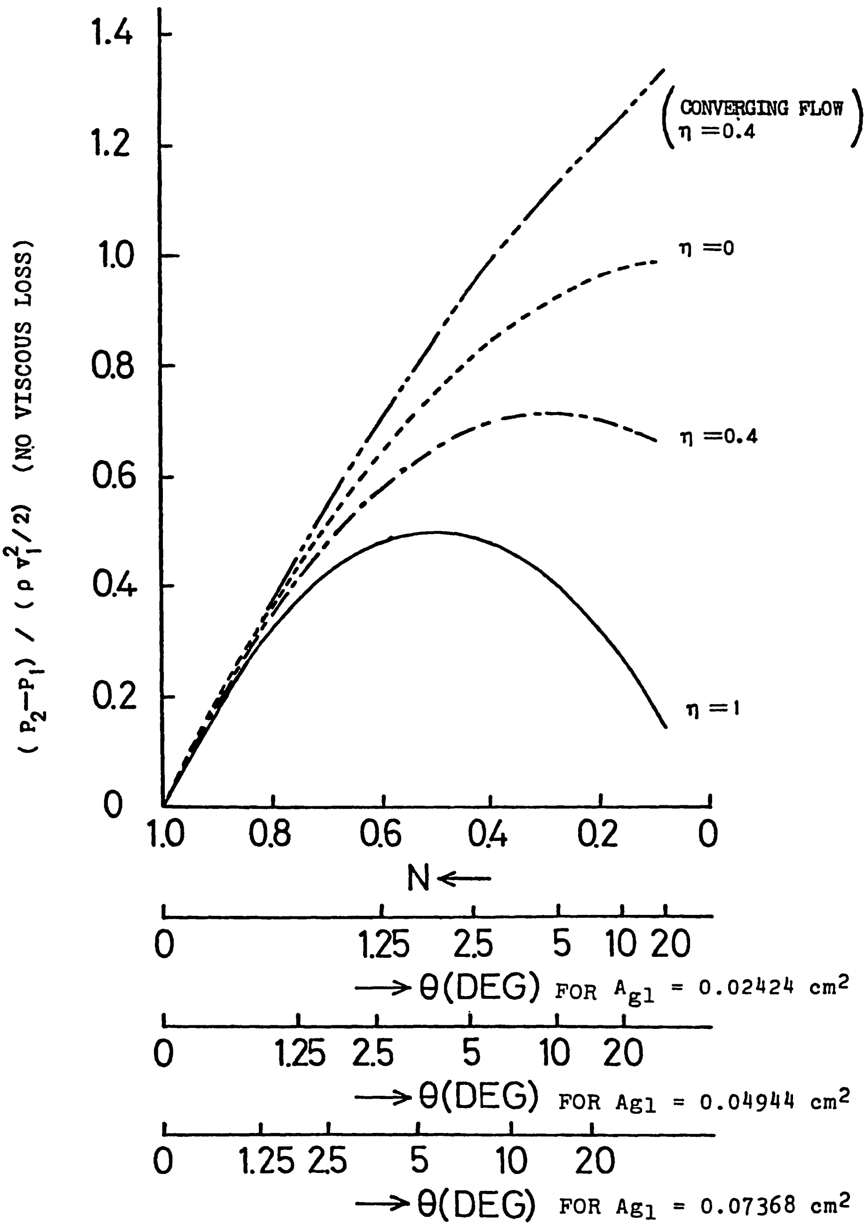


FIGURE 34-5

Normalized pressure recovery (and drop) as a function of diverging angle (and converging angle) of the glottal duct.

For diverging flow, $h_k = \zeta_d \cdot \rho v_1^2/2 = \eta_d(1 - A_1/A_2)^2 \cdot \rho v_1^2/2$,

For converging flow, $h_k = \zeta_c \cdot \rho v_2^2/2 = \eta_c(1 - A_2/A_1)^2 \cdot \rho v_2^2/2$.

where A_1 and A_2 are the cross-sectional areas at the inlet and outlet, respectively, and η_d and η_c are the loss factors for the diverging and converging ducts, respectively.

Applying the momentum law to the sudden expansion of flow, we have the loss factor $\eta_d = 1$. For sudden contraction of flow, the typical value for $\eta_c = 0.4$ to 0.5 according to Weisbach's experiments, (0.375 according to van den Berg's experiment). The loss factors η_c and η_d are not necessarily constant but can be approximated by fixed values for the limited range of diverging and converging angles of the glottal duct. It can roughly be estimated that η_d would take a value somewhat less than unity and η_c around 0.4.

PRESSURE RECOVERY IN DIVERGING DUCTS AND PRESSURE DROP IN CONVERGING DUCTS

Figure 34-5 shows pressure recovery (normalized to the kinematic energy of flow at the inlet) as a function of the diverging angle θ for $\eta_d = 0$, 0.4, and 1. The dotted line for $\eta = 0$ corresponds to that given by the previous equations for $\eta = 0$. For small values for $N = A_1/A_2$ or θ, there is, of course, no significant difference among the curves. One notices that for $\eta \neq 0$, the curves indicate the existence of the optimum diverging angle θ or N to attain the maximum pressure recovery. The basic principle is well known in experiments of pressure on diffusers in hydrodynamics. And in a conical diffuser, the most advantageous cone angle is about 8°, for which $\eta_d = 0.15$ to 0.2. The pressure drops (normalized to the kinematic energy of flow at the outlet) for the converging flow are also shown in Figure 34-5 for $\eta = 0$, and 0.4.

Figure 34-6 a and b show the pressure distributions in the larynx calculated for several diverging angles θ when $\eta = 1$ and 0, respectively. Again, one notices that for $\eta = 1$ the maximum negative pressure in the glottal duct is attained for $\theta = 5°$ to 10°. But for $\eta = 0$ (Figure 34-6b), the negative glottal pressure monotonically increases with increasing θ.

Using the equations with $\eta = 0$, and with $\eta_d = 1$ and $\eta_c = 0.4$, we have made computer simulation of voice production by means of the vocal cord/vocal tract model. The results of the computer simulation for the vowel /ɑ/ are shown in Figures 34-7 and 34-8. In order to assess the effects of the expansion and contraction losses in glottal ducts on the glottal volume flow U_g and output sound pressure P_{out}, the computer simulation both for $\eta = 0$, and for $\eta_d = 1$ and $\eta_c = 0.4$ was made for the same waveforms for A_{g1} and A_{g2}. Namely, the computer simulation for $\eta = 0$ was first made at the typical glottal condition and the data on A_{g1} and A_{g2} were stored. Then, for $\eta_d = 1$ and $\eta_c = 0.4$, the cord vibrations were not calculated but the stored data on A_{g1} and A_{g2} for $\eta = 0$ were read to compute U_g^* and P_{out}^*. A_g in Figure 34-7 is the minimum between A_{g1} and A_{g2}. The falling slope of U_g^* waveform is less steep than that of U_g, as Scherer and Titze (1981) suggested. As a result, the level of P_{out}^* is a little lower than that of P_{out}, because of the weak excitation of the vocal tract. This is also reflected in the amplitude spectra for the glottal volume flow and output sound pressure, as shown in Figure 34-8. Namely, U_g spectral amplitudes decay with about 13 db/oct at high frequencies, whereas U_g^* spectral amplitudes decay with about 16 db/oct at high frequencies. The peak levels of the 2nd and 3rd formants for P_{out}^* consequently are slightly lower by about 2 to 3 db than those of P_{out}. The difference in output sounds is also not significant perceptually.

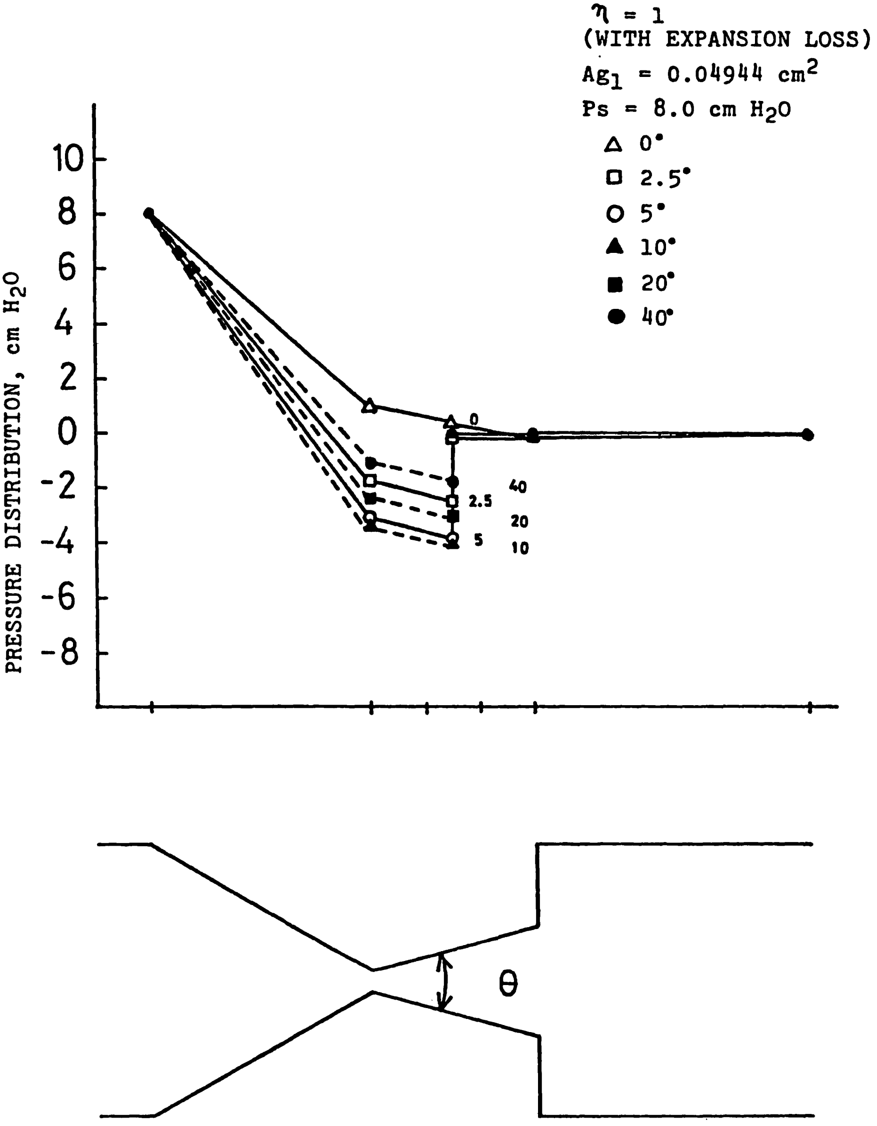


FIGURE 34-6a
Pressure profiles for several diverging angles of glottal ducts calculated for the loss factor η = 1 (a) and η = 0 (b).

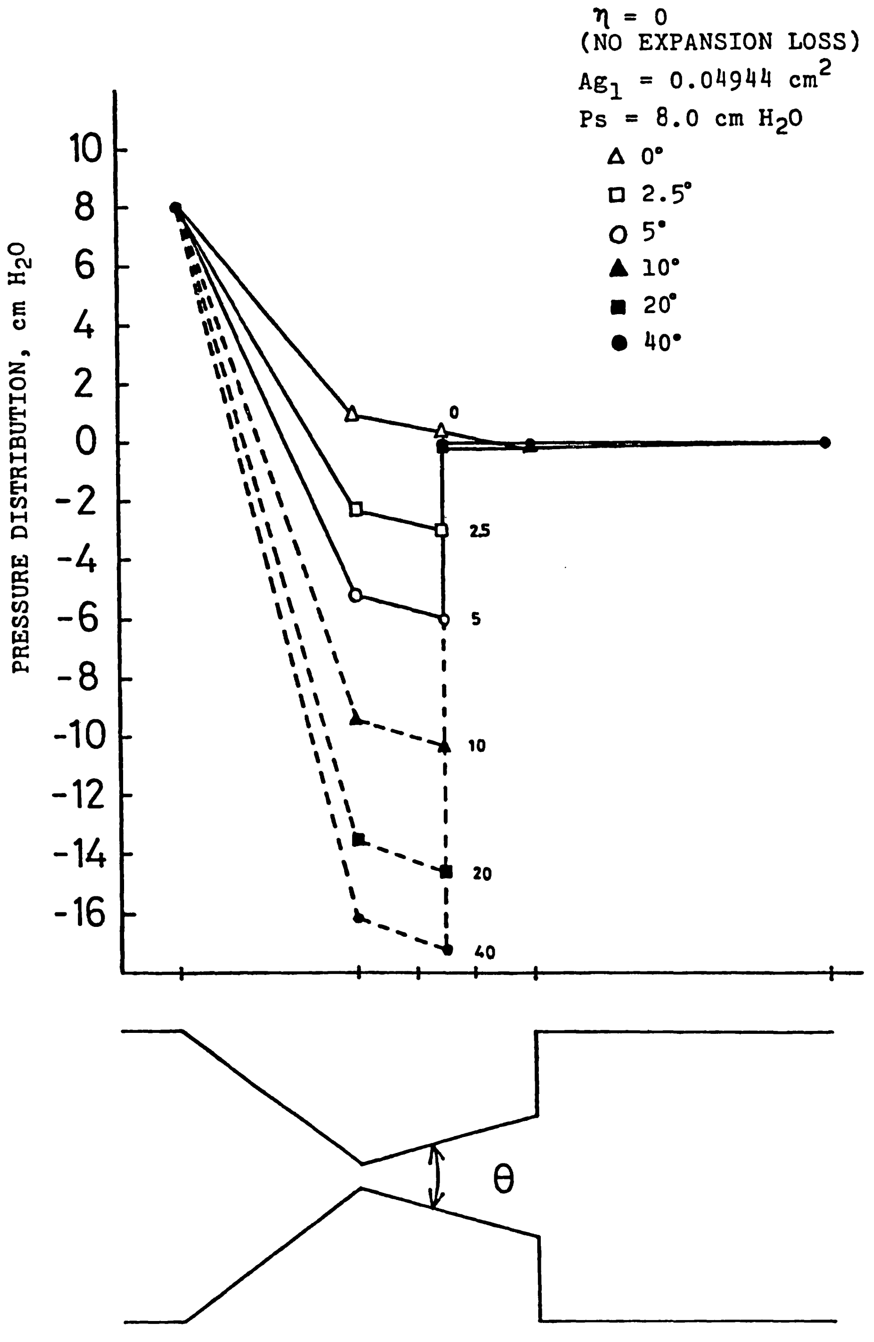

η = 0
(NO EXPANSION LOSS)
Ag1 = 0.04944 cm2
Ps = 8.0 cm H2O
△ 0°
□ 2.5°
○ 5°
▲ 10°
■ 20°
● 40°
PRESSURE DISTRIBUTION, cm H2O
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
0
2.5
5
10
20
40
θ

FIGURE 34-6b

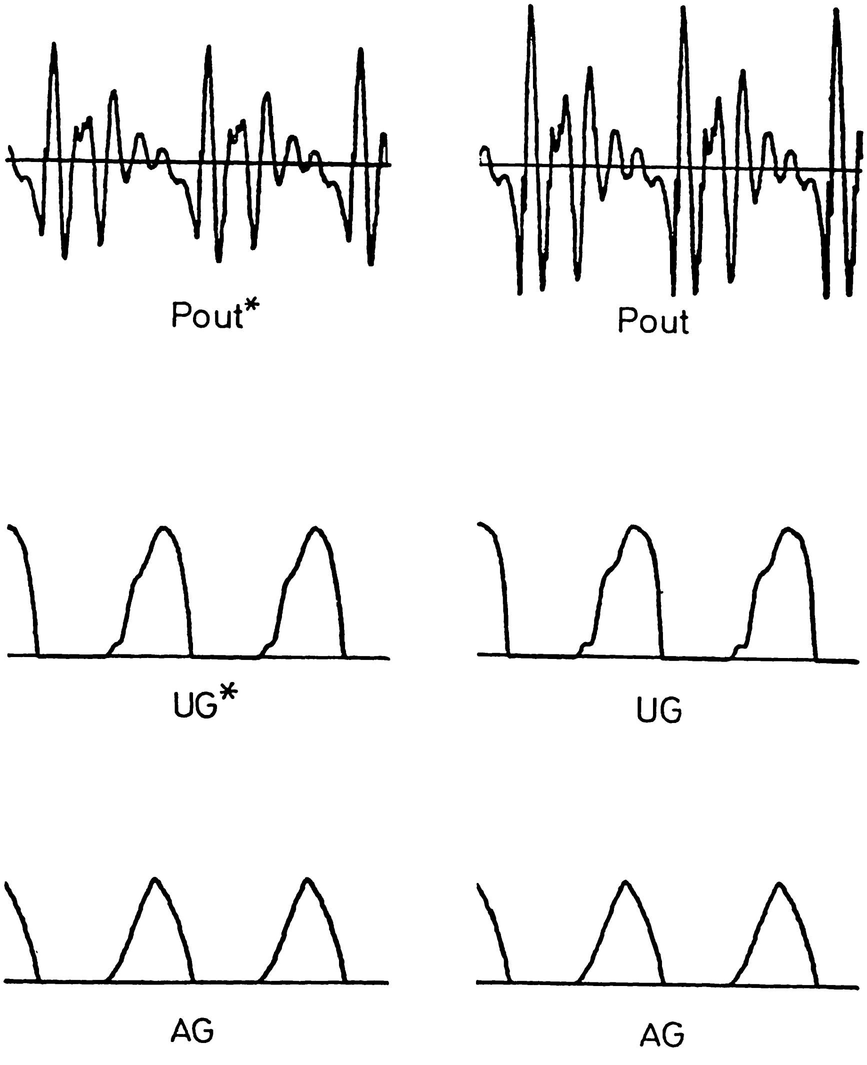


FIGURE 34-7
Results of the computer simulation for the vowel /a/ showing mouth-output sound, glottal flow and glottal area waves. *refers to $\eta_d = 1.0$ and $\eta_c = 0.4$, and no asterisk refers to $\eta = 0$.

CONCLUSION

We have shown that the losses due to the area change in the diverging and converging glottal duct could be involved in our previous equations $\eta = 0$ in the same manner as the losses due to the sudden contraction and sudden expansion of flow at the inlet and outlet of the glottis. The appropriate values for the loss factors η_c and η_d must be determined experimentally.

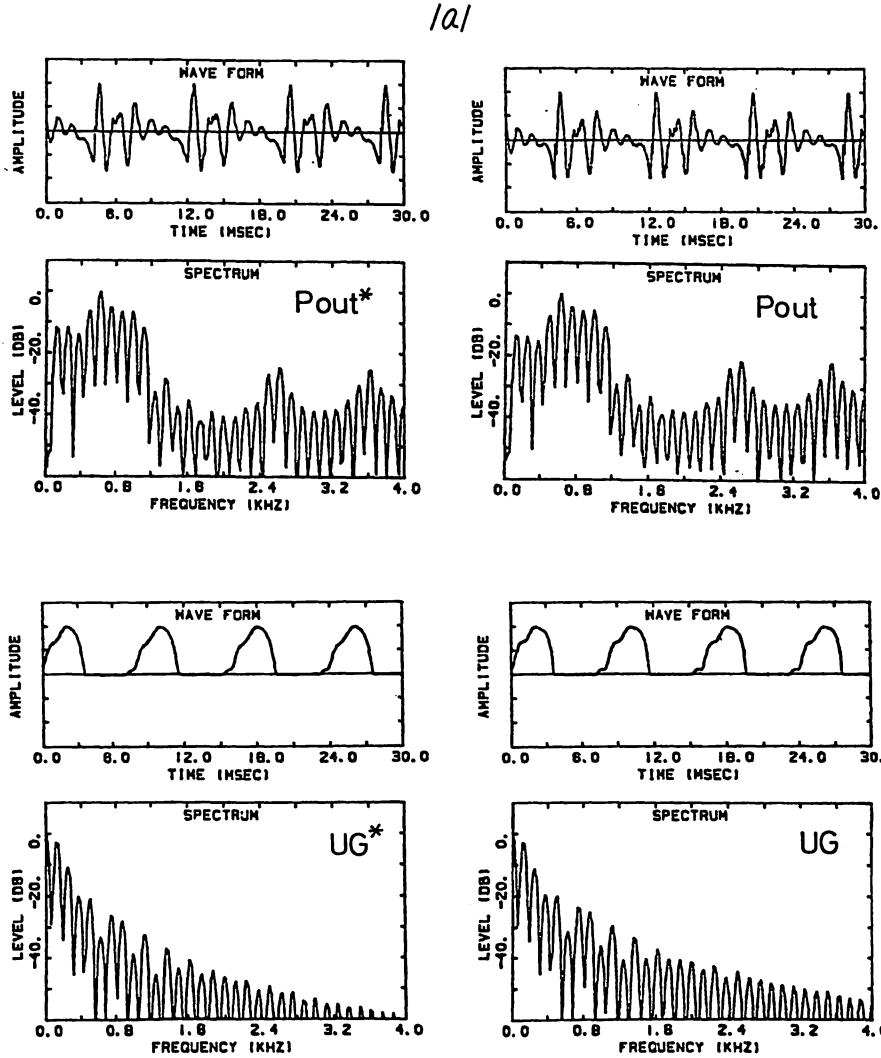


FIGURE 34-8
Spectra of mouth-output sound and glottal flow waves shown in Figure 34-7.

REFERENCES

Gauffin, J., *et al* (1981). "Glottal geometry and volume velocity waveform," *Conf. on Vocal Fold Physiol.*, Madison, WI.

Ishizaki, K. and Flanagan, J.L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* 51/6, 1233-1268.

Ishizaka, K. and Matsudaira, M. (1972). *Fluid Mechanical Considerations of Vocal Cord Vibration*. SCRL Monograph, #8, April.

Scherer, R.C. and Titze, I.R. (1981). "Pressure-flow relationships in a model in the laryngeal airway with a diverging glottis," *Conf. on Vocal Fold Physiol.*, Madison, WI.

Scherer, R.C., Titze, I.R. and Curtis, J.F. (1983). "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," *J. Acoust. Soc. Am.* 73/2, 668-676.

DISCUSSION

R. SCHERER: We are eager to test your new equations with the work we plan to do with the models that you saw in our lab. I have one minor comment, and it concerns the results with the latest model we are working with. We will report this data shortly.

This model, similar to one of ours from which you showed data, also had a 40 degree divergence angle. It was made more precisely. In that experiment, we constructed ten pressure holes along the vocal fold. Right past entry, the pressure went to the level that it attained all the way through the glottis and supraglottally. We did not see a pressure recovery within the glottis except for the largest diameters. When you get down to small diameters, it tends to level off.

Your paper is a nice piece of work that helps introduce more of the information on diffusers from mechanics.

35 ▪ AIR COLUMN, REED, AND PLAYER'S WINDWAY INTERACTION IN MUSICAL INSTRUMENTS

Arthur H. Benade

ABSTRACT

Musicians have always insisted on the importance of getting the proper shapes in a wind-player's air passages. For this reason, the apparent success of the current oscillation theory of reeds and musical air columns without inclusion of the player's windway effects became increasingly mysterious as the subject matured. Since this theory to date has been useful for guiding the construction of fine instruments, confidence in its techniques is sufficient to support a serious attack on the problem of extending it to include the player's windway. Major energy production occurs at frequencies where $A[(Z_u + Z_d)//Z_r] \gtrsim 1$. Here A and Z_r are the trans-conductance and impedance of the reed while Z_u and Z_d are the input impedances of the air columns looking upstream and downstream from the reed. Nonlinear effects couple these energy sources via heterodyne action, whether or not Z_u appears in the accounting. Meaningful extension of theory has been aided by the development of convenient pulse-echo/FFT-measurement techniques for the Z's of both the instrument air column and the player's windway. Most vowel (supraglottal) configurations give rise to Z_u peaks in the range of 450 to 1500 Hz that are able to play a significant role for instruments. The fact that these peaks do not coincide with speech formant frequencies has helped to confuse the situation, as has the fact that some players unconsciously exploit windway resonances, while many do not use them at all.

I. INTRODUCTION

This report is intended to provide an introductory account of how the player's own windway interacts with the reed and air column of a musical wind instrument. Our formal understanding of the reed/air-column interaction is extremely good today, to the extent that it is possible not only to describe the acoustical nature of the interaction but also to use it as an effective guide to the instrument maker in his labors to build a good instrument or to improve an already existing one. For this reason, our task is relatively simple: we need only to show how the additional complexities associated with the player's windway modify the mathematical physics of the simpler reed/air-column system and then examine the ways in which the modified system differs in its behavior from the one that has been well studied.

The reader's first reaction to the preceding paragraph may well be a remark such as, "Musicians for hundreds of years have insisted on the importance of the mouth and throat configuration of anyone who is serious about playing a wind instrument. How then can anyone claim to have understood a wind instrument to a useful extent <u>without</u> taking this fact into account? Furthermore, how can he then go on to announce the importance of the player's windway as a new discovery?" It is my hope that the answers to these important questions will themselves clarify the nature of just what it is that has become newly understood.

For many years I have stoutly told my musician friends (and myself, in my incarnation as a serious amateur player) that the role of the player's windway could only be clarified after the other, more easily visible contributors to the musical oscillation were properly elucidated.

As a matter of fact, today the question has inverted itself, taking the form: "How did such a largely influential part of the dynamical system remain incognito during the course of investigations covering many years in which changes of only two or three parts in a thousand of many other acoustical parameters could readily be associated with their dynamical and musical consequences?"

It will perhaps be useful to outline the preceding remarks in the following way before we look at the physics itself.

A. DOES THE LUNG-THROUGH-MOUTH AIRWAY SIGNIFICANTLY INFLUENCE THE PLAYING OF MUSICAL WIND INSTRUMENTS?
  1. MUSICIANS ARE UNANIMOUSLY OF THE OPINION THAT IT DOES.
  2. THE MUSICAL ACOUSTICIAN HAS TENDED TO IGNORE THE QUESTION, OR TO SET IT ASIDE AS A RELATIVELY SMALL INFLUENCE.

Item (2) above is a deliberate oversimplification. Measurements and speculations of an acoustical nature have been made over the span of many decades, but for various reasons no clear consensus has developed. The detailed recounting of this branch of history will not contribute appreciably to our present purpose, which is to give a compact description of what is known today, in a form that will (hopefully) be intelligible to a readership whose major

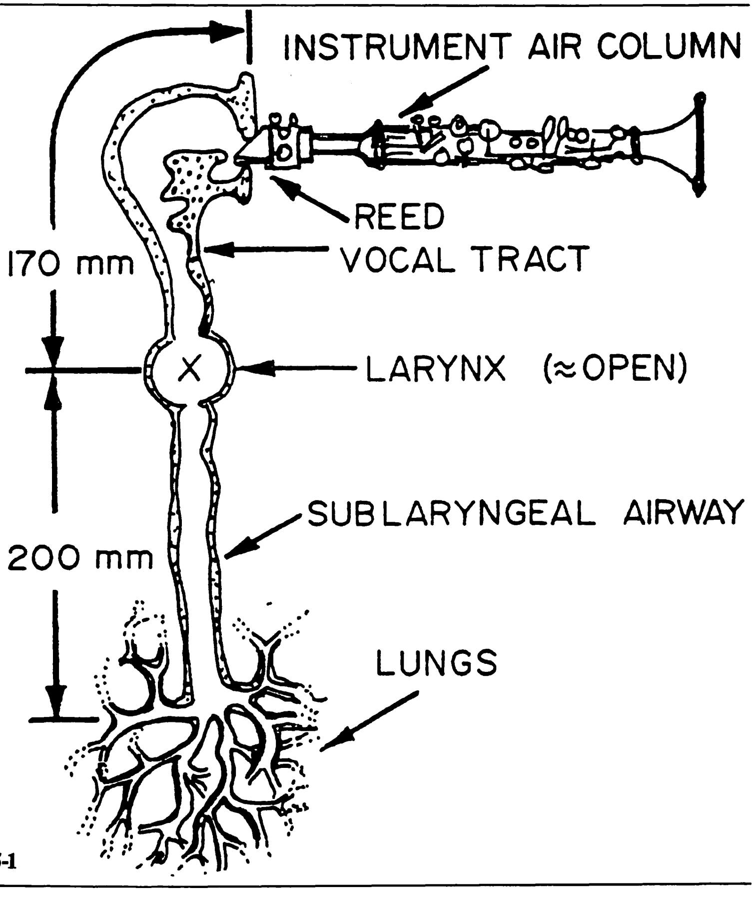


FIGURE 35-1

concerns are with the biophysics of the player himself rather than with the details of his interaction with a musical wind instrument.

I wish to make clear at this point in my introductory remarks that the present report is intended to be little more than an announcement of some of the recent results obtained in Cleveland. For the sake of brevity, I will therefore run the risk of frustrating my readers and annoying workers elsewhere whose results are not properly acknowledged. I shall, however, mention here by name those of my coworkers past and present who have made particularly large contributions (beyond the limits of any published work of theirs) to the insights reported here; these are Walter Worman, George Jameson, Stephen Thompson, and Peter Hoekje. The present report would not have become possible without their direct collaboration. This is true of George Jameson and Peter Hoekje in particular. Beyond this I will present only those bibliographical details that can directly aid the reader in his comprehension of the present discussion. A formal research report with proper acknowledgement and documentation is being prepared by Hoekje and myself for submission to the Journal of the Acoustical Society of America.

II. FORMULATION OF THE PROBLEM

Figure 35-1 shows the general nature of the dynamical system with which we are concerned. The system may be considered as being the concatenation of four major segments: the sublaryngeal airway (terminated its lower end by the player's lungs), the larynx (which in the present case is either wide open or partially closed in a manner that does not permit it to oscillate), the vocal tract (which is extensively adjustable via motions of the soft palate, tongue, jaws, etc), the reed of the musical instrument (whose operating point, damping, etc., are controlled by the player's lip position and pressure), and the musical air column (whose acoustical properties are controlled via the player's fingers on the various keys and/or tone holes).

In any musical wind instrument, whether woodwind, brass, voice (or even harmonica!), we find three interacting subsystems: an air passage from the wind supply (the player's windway or the organ pipe's foot and the wind chest below it), a flow-control device (the cane or lip reed of the orchestral wind instrument, the singer's larynx, the free reed of the harmonica, or the air reed of the flute player), and finally some sort of resonator and radiating system that ultimately couples with the room into which the sound is to be fed. Setting aside the flute family of instruments, the flow control device is a valve whose degree of closure is determined by the pressure difference across an operating surface.

Figure 35-2 shows two versions of the basic pressure-controlled system. One of the controlling pressures is maintained in part by the player's lungs and in part is produced by the acoustic disturbances taking place in the player's windway (PWW). The other pressure acting upon the reed-valve is found within the instrument's mouthpiece as a manifestation of the acoustical activity taking place in the instrument's air column (IAC). In Figure 35-2a the valve action is arranged so that an increase in the downstream pressure p_d leads to a greater flow. This arrangement is typical of all the orchestral reed woodwinds and of the organ reed pipes. Figure 35-2b shows, on the other hand, a system in which the valve action is reversed so that an increase of p_d decreases the flow u, as is typical in the orchestral brasses.

We will find it convenient to define directions in this essentially one- dimensional waveguide system with the help of the words "upstream" and "downstream," these being directly related to the direction of the normal flow of air from the player's lungs out into the room. Thus one of the flow- controlling pressures acts on the upstream side of the reed, while

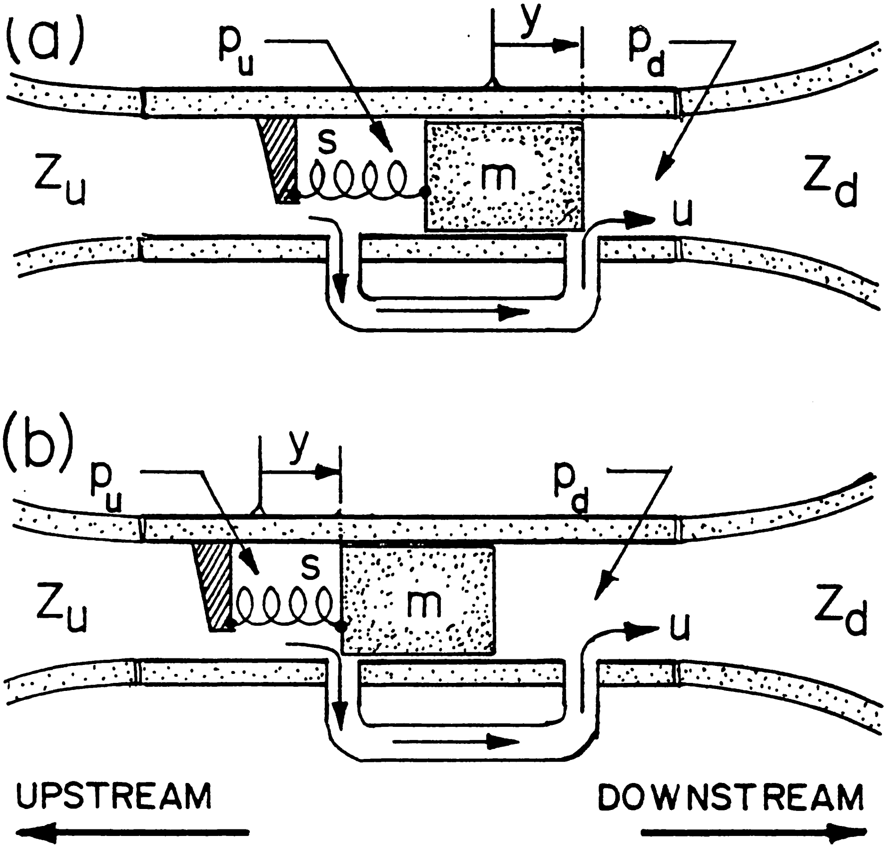


FIGURE 35-2

the other is exerted on its downstream surface. Terminology based on this convention prevents ambiguities of the sort that arise if one simply uses the words "up" and "down." For a clarinetist the airflow runs upward within the player, and then downward through his instrument. The problem of similar description of what goes on in a bassoon or tuba defies the imagination!.

We find it useful to characterize the PWW and the IAC via their impedances as seen by the flow controller. We will refer to the impedance looking upstream into the PWW as Z_u, while the impedance of the IAC will be denoted by Z_d. The reed itself requires two characterizations, since it plays two roles in the complete oscillating system. We define its acoustic impedance Z_r in terms of its displacement volume velocity when it moves in response to a driving pressure exerted on either one of its surfaces (see Figure 35-2); the other and perhaps more basic property of the reed is its flow-control characteristic, which is in general a nonlinear function. This flow-control characteristic is most conveniently specified by expressing the flow u as a Taylor expansion in the pressure difference p between the two sides of the reed, as given in Equation 35-1.

$$u(t) = u_o + A\ p(t) + B\ p^2(t) + C\ p^3(t) + + + + + \tag{35-1}$$

Because the reed assembly functions as a spring-mass-damper system, we recognize at once that Z_r shows a resonance property that makes it inversely proportional to the factor $D(\omega)$ that is set down as Equation 35-2.

$$D(\omega) = \left[\frac{\omega_r^2}{\omega_r{}^2 - \omega^2 + jg_r\omega}\right] \tag{35-2}$$

Here ω_r is the natural frequency of the reed and g_r is its half- power bandwidth. We notice, further, that since the actual flow rate of air through the reed depends on its position (and so only indirectly on the activating pressure), the flow-control coefficients are themselves resonant in their nature. That is, these coefficients may be written as the product of their low-frequency "steady-state" values (A_0, B_0, C_0, . . .) and the factor $D(\omega)$ defined above. This fact proves to be very important to our understanding of the way musical instruments are played. We can usefully remark here that A_0 is positive for the woodwind valve system of Figure 35-2a and negative for that belonging to the brasses, as in Figure 35-2b.

Let us now write down the pressure and flow relationships on the upstream and downstream sides of the reed, in terms of the impedances Z_u, Z_d, and Z_r. The positive direction of acoustic flow is defined to be the downstream direction of the DC flow from the player's lungs.

$$u = p_d/Z_d + (p_d - p_u)/Z_r \tag{35-3a}$$

$$-u = p_u/Z_u + (p_u - p_d)/Z_r \tag{35-3b}$$

The first term on the right side of each of these equations simply expresses the ordinary relation between pressure at the entryway of a waveguide and the flow that goes into that waveguide. The second term gives a measure of the flow that goes into the region vacated by the reed itself when it moves under the influence of the pressure difference that acts upon its two sides.

Equations 35-3a and 35-3b can be combined in an interesting and useful way: the flow u through the reed aperture can be expressed very simply in terms of the pressure difference p across the reed, as shown in Equation 35-4.

$$p = u(Z_u + Z_d)//Z_r \tag{35-4}$$

That is to say, the pressure difference across the reed is proportional to the sum of the upstream and downstream impedances, in parallel with the reed impedance (which tends to be very large compared with the other impedances, so that it has a secondary, though nontrivial, role in the oscillation process). We will use the unadorned symbol Z to represent this combined impedance.

As a preparation for the next step in the discussion, we should recapitulate the nature of the problem whose solution we are trying to outline. When a wind instrument is played, the upstream and downstream impedances (together with the reed's own impedance) are coupled via a flow-controlling valve to the player's lungs, which serve as the primary source of compressed air. The system is kept in oscillation by a feedback loop in which the net acoustical disturbance at the reed (i.e., the pressure difference across it) operates the flow controller, and the resulting flow serves as the excitory stimulus for the upstream and downstream waves.

Equation 35-1 provides us with a formal representation of the pressure-operated flow-control property u(p) of the reed, while Equation 35-4 provides in a very compact form the pressure response property of the entire airway system (PWW + IAC + REED) to a flow stimulus. We should notice that both of these equations relate the flow u, which is the same on both sides of the reed valve, to the pressure difference p across it. In other words, our analysis can be carried out in terms of p and u via the net Z and the "control polynomial" u(p), without our having to worry about the complications of the individual responses of our three subsystems to the flow which they jointly engender via a nonlinear coupling.

From the point of view of mathematical physics we have here an initial explanation of why effects produced by the PWW did not automatically destroy our ability to make meaningful calculations guided by, and checked against, experiments with reeds and various types of IAC—all that was necessary was that the PWW would not produce confusing and distracting effects. We were fortunate, indeed, over a period of many years that such was the case for long enough for us to get a firm grasp of the essential physics.

We turn now in the briefest way to a sketch of how the essential behavior of the system can be understood. Confining our attention for the moment to the case of strictly periodic oscillations in the system, we write the flow u(t) as a Fourier series:

$$u(t) = \sum u_n \cos(n\omega_0 t + \psi_n) \tag{35-5}$$

Here ω_0 represents the frequency of the tone being produced. Term by term this series represents the flow excitation spectrum being applied to the (PWW + IAC + REED) system. Given the (net) impedance $Z(\omega)$ of this system, we may write Z_n for its magnitude at the frequency $n\omega_0$ and ϕ_n for its phase. The pressure signal corresponding to u(t) can then be written down.

$$p(t) = \sum Z_n u_n \cos(n\omega_0 t + \psi_n + \phi_n) \tag{35-6}$$

As a matter of formal mathematics, Equations 35-1, 35-5, and 35-6 can be solved simultaneously to give the pressure spectrum across the reed for a given blowing pressure. While the detailed calculations are very tedious, it proves possible to extract a great deal of useful information about the system. This information, which can be readily checked against the behavior of real systems, depends much more on the overall mathematical structure of the problem than it does on the numerical values of the various parameters. That is, the salient features of the solution can be summarized very simply in a form that depends only on the systematic behavior of nonlinear trigonometric equations. Furthermore, when the complete story is in, we find (surprisingly enough) that the results show almost no sensitivity to the phases of the impedances, or of the reed resonance factor (Equation 35-2)! This is not to say that the phases are irrelevant or that they have random values—merely that the spectrum amplitudes are not sensitive to the phases of the Z_n's and the D_n's.

Equations 35-7 and 35-8 will suffice here to indicate the nature of the playing pressure spectrum as measured across the reed. In particular, the fundamental component p_1, which is the pressure amplitude of the disturbance at the playing frequency, obeys an equation of the form

$$p_1 = \frac{Z_1[Bp_1p_2 + \tfrac{3}{4}Cp_1^3 + + +]}{1 - Z_1[A + 2Bp_o + + +]} \tag{35-7}$$

Similarly, the higher components have amplitudes that can all be written in the form

$$p_{n>1} = \frac{Z_n[(B \text{ etc})\ p_1^n + + +]}{1 - Z_n[A + 2Bp_o + + +]} \tag{35-8}$$

I want to point out that in these equations there is no explicit appearance of the phase shifts associated with the flow-control parameters or the impedances. Only the magnitudes are important when the oscillation is of periodic type.

We will postpone discussion of these results until we have sketched out a linear cousin to the analysis, in which we can see what happens to the n'th component of the pressure when looked at by itself, the inescapable nonlinear coupling between spectral components being represented by a flow source U_n that is "external" to the component in question.

III. A LINEAR COUSIN TO THE PROBLEM

Suppose that our system is running in a steady oscillation at the frequency ω_o, with a part u(t) of the flow being produced through the linear term Ap of the control polynomial, and part of it U(t) being externally imposed by an as-yet-unspecified source having the same periodicity. If we use the Fourier representation, the imposed flow may be written

$$U(t) = \sum U_n e^{jn\omega_o t} \tag{35-9}$$

and the pressure signal across the reed is

$$p(t) = \sum Z_n[u_n + U_n]e^{jn\omega_o t} \tag{35-10}$$

Equation 35-10 may be solved term by term for the flow component amplitudes in terms of the combined impedance Z_n and the corresponding transconductance A_n (evaluated at the frequencies ω_n of interest):

$$u_n = A_n p_n = Z_n A_n [u_n + U_n] \tag{35-11}$$

whence

$$u_n = U_n \left[(Z_n A_n)/(1 - Z_n A_n) \right] \tag{35-12}$$

Here and in the discussion that follows through Equation 35-13, the symbols A_n and Z_n have their ordinary complex representation; i.e., account is taken of both magnitude and phase.

Equation 35-12 has the familiar form that represents the current gain u_n/U_n of a feedback amplifier for which the open-loop gain is $Z_n A_n$. It is at once apparent, therefore, that each spectral component of the flow is self-sustaining as an independent oscillator if the real part of the open-loop gain is exactly unity. That is, an energy input provided by the external drive signal U_n is not needed to keep the oscillation going.

However, if the open-loop gain Z_nA_n is less than unity, the amplitude of the flow component u_n is proportional to U_n. Furthermore, the magnitude of u_n will die away exponentially in time if U_n is abruptly shut off, with a decay rate that is proportional to the discrepancy between unity and the magnitude of the real part of Z_nA_n.

If, on the other hand, the open-loop gain is greater than unity, an exponentially growing oscillation can take place with a growth rate that is once again proportional to the difference between unity and the magnitude of the real part of the open-loop gain. Under these conditions the feedback system is able (for the component in question) to generate more energy than it can dissipate, without need for an additional input via U_n.

So far as our present (oversimplified) model is concerned, we may summarize by saying that the oscillation of each spectral component is independent of all the others, and that it is inherently unstable. We are of course very much accustomed to this sort of instability, which is shared by all ordinary oscillators, and it is quite customary to recall the presence of some amplitude-dependent (nonlinear) additional damping which comes into play to stabilize the amplitude of a real oscillator.

In the multicomponent musical oscillator there are, to be sure, several amplitude-dependent sources of damping beyond that implied by Z_u, Z_d, and Z_r (turbulent damping, for example). There is, however, another way in which energy can be transferred in and out of each spectral component, a way that not only assures the stability of each component amplitude under much less stringent requirements on the open-loop gain, but also guarantees that the various amplitudes have a well-defined relationship to one another. This is of course an absolute requirement for a musical sound source whose tone color needs to be defined for each condition of playing chosen by its user. The fundamentally nonlinear nature of the control polynomial defined in Equation 35-1 shows (in simplest terms) that whatever pressure signal components p_n might be generated via the operations of the linear term in this polynomial, they will immediately breed contributions to the entire collection of flow components at all other harmonic frequencies according to the heterodyne (intermodulation) arithmetic that may be generalized for arbitrary exponents from the trigonometric relation

$$(M\cos P)\,(N\cos Q) = (MN/2)\left[\ \cos(P+Q) + \cos(P-Q)\ \right] \tag{35-13}$$

That is, the "externally imposed" flow components U_n that were introduced in Equation 35-9 may now be understood to represent in a very simple way (computationally useless but heuristically helpful) the transfer of energy from each modal oscillator to its brothers. It is no longer required that each component be precisely self-sustaining when looked at by itself; all that is required is that as a group the spectral components can jointly produce enough energy to supply their total energy expenditure to the outside world.

Our quasi-linear model provides us one more insight into the nature of the real-world nonlinear system: every spectral component is connected directly or indirectly to every other one, so that its phase is the resultant of many influences. The nature of the oscillation is such that there are <u>many</u> ways in which the actual phase of a given component can be reconciled with those of its conferers. Proper analysis shows that, as a result, the spectral amplitudes are determined almost exclusively by the magnitudes of the relevant Z and A, B, C parameters and not by their phase angles (Thompson 1978).

The discussion so far in this section has shown that energy production is favored at maxima of the $A(\omega)Z(\omega)$ product. In the woodwinds, A is very nearly A_0 over much of the spectral range because the reed's own natural frequency ω_r is relatively high (e.g., 2000-3000 Hz for a clarinet). This being so, energy production is favored at the impedance maxima of the PWW-IAC-REED system. This says (if for a moment we ignore Z_u and Z_r) that oscillation is favored at the normal-mode frequencies of the IAC taken with its reed end closed, as has been recognized for at least 200 years ("the clarinet plays as a stopped pipe").

Another implication of our discussion is that the overall energy production is largest if the impedance maxima are harmonically related to one another. This assures that each of the heterodyne frequency components generated from the harmonics of the played not finds itself matching one of the energy-producing impednace maxima and thus transferring energy to a productive place in the regenerative scheme. Let us put this in more obviously music-related words of the sort used prior to the explicit inclusion of PWW effects: a musical instrument whose impedance maxima (as modified by the parallel but large Z_r) are harmonically related is one that starts its tones well, produces a clear sound, provides controllable dynamics and stable pitches, and is otherwise a most attractive instrument in the hands of the player and in the ears of the listener. I have given a very extensive discussion of these matters in chapters 20 through 22 of my book (Benade 1976). Conscious recognition of the usefulness of accurate harmonic "alignments" of the air column resonances led (beginning around 1964) to a continuing evolution of laboratory and workshop techniques for the measurement and correction of the positions of the resonances belonging to essentially all the notes of an instrument's scale. The behavior of instruments adjusted by means of these techniques has been much admired by well-known musicians, and the techniques themselves are beginning to have a significant effect on the making of (at least artist-grade) instruments of all sorts today.

We had temporarily set aside the possibility that the ZA product could become large near the reed frequency ω_r, so that the harmonic for which $n\omega_0 \simeq \omega_r$ might contribute to the net energy production even though Z itself might not be large. While the book contains numerous qualitative remarks concerning the musical usefulness of this possibility in woodwinds, the detailed physics of it was not elucidated till later (Thompson 1979). For present purposes it will suffice to say that all really skilled woodwind players exploit the possibility of an extra energy source at ω_r by setting the reed frequency at some (any!) harmonic of the playing frequency in order to further stabilize and purify their sound production via the inclusion of an extra, accurately aligned participant in the "regime of oscillation." For brass instruments, the player must pay attention to ω_r, since the note he wishes to play is selected directly by arranging the lip-reed natural frequency to lie just below the fundamental of the desired tone. Further discussion of the curious dynamics of the brass instrument, with its reversed-sign value for the reed transconductance $A(\omega)$, would take us too far from the goals of this report. It will suffice for us to notice that the adjustability of the reed resonance frequency is a musically important resource for the woodwind player and an unavoidable necessity for the brass player. In both cases we find that a physiological adjustment is used as an adjunct to the mechanical controls provided by the player's hands on the keys, valves, and slides of his instrument.

We close this part of our thumbnail sketch of the (inherently nonlinear and therefore very stable) sound production mechanism of the orchestral wind instrument by pointing out once more that our understanding of it reached a highly developed state without any account being taken of the possibility that the player's windway could itself play a significant role. Our present analysis has shown that Z_u enters the dynamical equations in a manner that is entirely symmetrical with that of Z_d. To the scientist this means that he does not need to rework all his equations when he adds consideration of Z_u to his analysis of Z_d and Z_r: the symbol Z merely takes on a slightly different meaning. From the point of view of the musician it means that the player has one additional physiological adjustment-resource at his disposal (whose dynamical nature we now can see in a general way). For all of us, we have yet the question of how the dynamical effect of this resource could remain scientifically incognito for so long, a question to which a partial answer will be given below.

IV. SPECTRAL IMPLICATIONS

Now that we have sketched out the general nature of the nonlinear multicomponent regeneration process that functions in the orchestral wind instruments, we are in a position to examine the spectrum of the control pressure signal p(t), as given in Equations 35-7 and 35-8 above. Recall that in these equations we need only the magnitudes of the Z, A, B, C parameters! The first thing that we notice is that the denominators of these equations are almost exactly like the denominator of Equation 35-12, from which we learned of the crucial importance of the Z_nA_n product in controlling the amount of energy generation that can take place at the n'th harmonic. The only unfamiliar feature is the presence of other spectral components whose influence is added to the direct effect of the component in question. In Equations 35-7 and 35-8 these extra p_j's are the explicit representations (in an essentially exact formulation) of the "imposed flow" contributions that were introduced heuristically in Equation 35-9. Aside from this, the denominators have almost exactly the same meaning in the exact formulation that they did in our introductory version. We can see this explicitly in Equation 35-7, which gives information about the fundamental component of the spectrum. We begin by considering the form taken by this equation in the low-amplitude limit, where the quadratic and higher-order terms in the flow polynomial (Equation 35-1) have no role to play. Under these conditions, the fact that p_2 and other higher-order components are zero means that if there is to be any oscillation at all at the fundamental frequency, then $(1 - Z_1A)$ must vanish, exactly as we have come to expect.

We turn now to a consideration of the numerator of Equation 35-8. This shows a remarkably simple pair of overall relationships (that are well substantiated by experiment under suitable conditions), as we can see from the abridged version set down as Equation 35-14.

$$p_n = Z_n p_1^n \cdot (\text{other, slow-moving terms}) \tag{35-14}$$

The first of these relationships is that the general shape of the reed-drive pressure spectrum is well caricatured by the envelope of the controlling aggregate impedance, and the second is that the n'th pressure amplitude component is proportional to the n'th power of the fundamental component amplitude as this changes with the player's blowing pressure. In other words, as one plays a crescendo, keeping his embouchure and PWW constant, the oscillation "blossoms" from a nearly pure sinusoid into a waveform whose components grow progressively to the fully developed mezzoforte distribution implied by Equation 35-8. Playing louder yet causes the reed to close fully for a growing fraction of each cycle, giving rise to an entirely new type of spectral development that has its envelope determined by the duty-cycle of the puffs of air through the reed. Beyond this we need only to notice the exact parallelism of mathematical form in the denominators of Equations 35-7 and 35-8.

It is only a brief step now to a description of the two spectra (which can be measured) on each side of the reed: that is, the spectrum measured in the instrument's mouthpiece (as has been done for many years during the development of the basic theory outlined here) and the spectrum measured in the player's mouth. If we write $(p_n)_u$ and $(p_n)_d$ for these two pressure-spectrum components and recall that

$$Z_n = ((Z_u + Z_d)//Z_r)_n ,$$

then

$$(p_n)_u = u_n(Z_u)_n = p_n(Z_u/Z)_n \tag{35-15a}$$

$$(p_n)_d = u_d(Z_d)_n = p_n(Z_d/Z)_n \tag{35-15b}$$

If (as has been known for many years), Z_r is large enough to have only a small influence on the magnitude of Z, and if (as was presumed for almost as many years) Z_u is relatively small and featureless, equations like 35-7 and 35-8 appear to apply directly to the mouthpiece spectrum, calculated using Z_d obtained from measurements of the IAC. Experiments of this sort in fact have been done and have provided a significant fraction of the evidence that has to date supported our confidence in the theory as outlined. Notice once again our debt to the curious but fortunate accident that the influence of the PWW did not intrude upon our consciousness until we were ready to cope with it!

It has been a truism of the subject that changes in the mouthpiece pressure-spectrum amplitudes should directly reflect changes in the corresponding impedance peak heights, as is made explicit by the numerator of Equation 35-8 and the leading factor in Equation 35-14. It is but a short step from this for us to invoke the upstream/downstream symmetry of the system as justification for the idea that changes in Z_u produced by tongue and mouth movements by the player will produce exactly parallel changes in the pressure spectrum as measured in the player's mouth. However, it is not at once obvious what happens to the spectrum on one side of the reed as the result of changes in impedance on the other side.

Differentiation of the written-out form of Equation 35-15a with respect to Z_u, and of Equation 35-15b with respect to Z_d, gives us an explicit representation of these cross-influences. When this is done, a very surprising result is obtained:

TO FIRST APPROXIMATION, CHANGING Z ON ONE SIDE OF THE REED MAKES NO CHANGE IN THE SPECTRUM ON THE OTHER SIDE!

On closer examination we find that there are indeed small changes, especially if the perturbed spectral component is one of those for which the ZA product is nearly unity—if, in other words it is very nearly able to balance its own energy budget, and so support itself without feeding energy to, or absorbing it from, the other components.

We close this discussion of the overall theoretical formulation of the wind instrument regeneration process with a short summary of the major points, leaving the broader implications till after the presentation of some experimental data on the influence of the PWW on the playing regimes of real instruments. the first point which should be made is that the upstream and downstream impedances appear symmetrically in the theory. The second point is that everything about the oscillation is directly determined by aggregate Z as defined in Equation 35-4. The third point is that if the magnitude peaks of the aggregate Z function are harmonically related, the oscillation is stabilized, made clean and noise-free, and given a controllable nature that is favorable to good musical performance. The fourth point is that while changes in Z_u and Z_d alter the spectrum as observable on the same side as the changes are made, there is generally little or no change on the other side of the reed.

We may take item three above as giving an analytical indication of why a player might find it advantageous to manipulate his PWW. Similarly, item four can give us a hint as to why these effects were not immediately detectable in the course of ordinary research—measurements were only made on the downstream side of the reed!

V. IMPEDANCE MEASUREMENTS ON THE PLAYER'S WINDWAY

As has already been remarked, one of the reasons why many of us took it for granted that the PWW would have little effect on the basic regeneration processes of a musical wind instrument was the assumption that the multi- branched, softwalled air passages the player's lungs acted as an essentially reflectionless termination of the sub- and supraglottal airway. We were further encouraged in the belief that the upstream airway was unlikely to have an important role by the fact that the pipe foot and wind chest of a pipe organ have a relatively small physical (but not musically negligible!) influence on the sound and the stability of tone production. Twenty-five years ago this gave sufficient reason to move forward boldly, under the guidance of the writings of Henri Bouasse (Bouasse, 1929-30) and with the stimulation shortly afterwards of the accurate pioneering measurements of the clarinet reed's flow-control transconductance (A_0) carried out by John Backus (Backus 1963).

While precision measurements of IAC input impedances could be made from the earliest part of this active period (see the examples of measurement technique in Benade 1973), the necessarily slow frequency-sweep techniques then available could not be adapted to measurements on the highly variable PWW. The more recent arrival of convenient FFT procedures has led many of us to devise flow-impulse excitation methods, where the impedance is deduced from the Fourier transform of the pressure response signal. Members of my audience are far better acquainted with the history of this subject than I, so the present listing of references is only intended to indicate some of the earlier influences on my own thinking about this sort of procedure (Oliver 1964; Rosenberg and Gordon 1966; Fransson 1975; Dawson 1976; Krüger 1980). The remaining paragraphs of this section will be devoted first to an indication of the nature of the apparatus we have begun to use, then to the display of the PWW input impedance (Z_u) measured for various vocal tract configurations, and finally a description of some of the information that can be gained from them.

The impedance head used in our present experiments is of the sort shown in Figure 35-3 (Ibisi and Benade 1982). The primary sound source is a 27-mm diameter piezoelectric "beeper" disc bonded to the end of a short piece of 20-mm ID, 32-mm OD heavy-wall phenolic tubing by a bead of RTV rubber. The pressure signal is detected by an electret microphone whose 3-mm aperture looks into the tube only 12 mm from the face of the piezoelectric driver. If the piezoelectric transducer is considered to be a lossless single-mode harmonic oscillator, then a linearly rising ramp drive voltage will produce a single velocity pulse of the form

$$v(t) = V[1 - \cos(2\pi t/T)] \qquad (35\text{-}16)$$

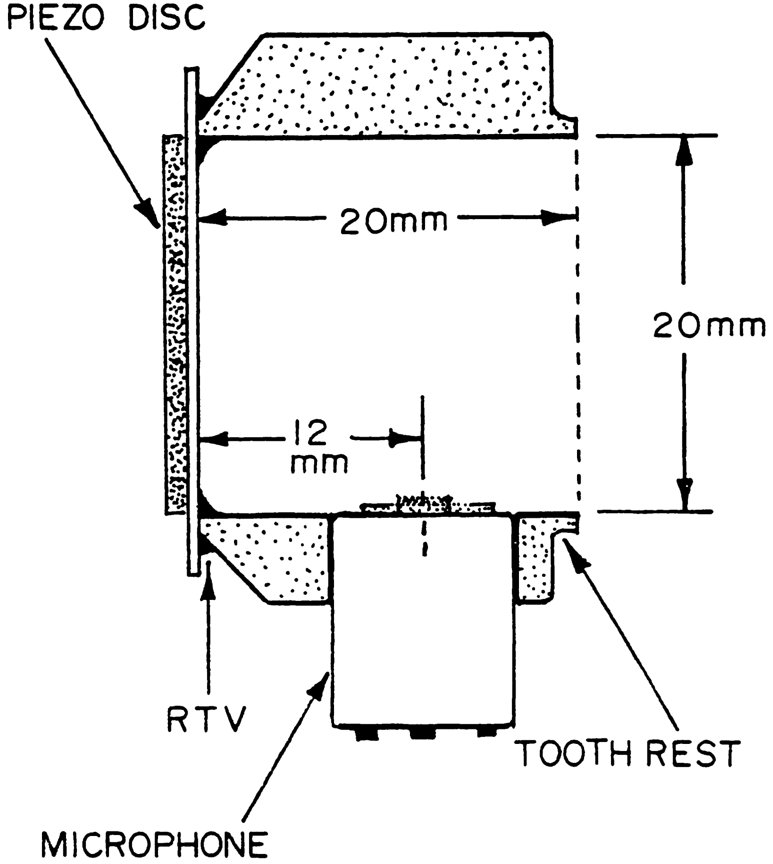


FIGURE 35-3

for 0 > t > T (zero otherwise), provided that the ramp duration T is exactly equal to the natural period of oscillation of the transducer. Only a slight modification of the drive voltage waveform is required to assure a very similar excitation velocity signal when account is taken of the fact that the transducer is a damped oscillator (a detailed report of these and other matters is in preparation for submission to JASA). Suffice it to say that our excitation pulse has a FWHM of about 0.083 milliseconds, so that FFT measurement of Z_u is possible without correction up to well beyond the 2500-Hz limit of our present major concern.

The upper part of Figure 35-4 shows the pulse-echo sequence observed when the driver is attached to a piece of 20-mm ID copper tubing open at the far end and 570 mm long. Notice that doubling the height of the initial pulse makes it a member of the alternating-polarity, exponentially decaying sequence of the later pulses, exactly as theory predicts. The lower part of the same figure shows the input impedance of this air column as calculated via FFT from the time waveform in the upper part of the figure. Both of these displays are plotted from data stored in a Hewlett-Packard 3582A real-time analyzer.

Figure 35-5 will orient us to the general magnitudes of the peak values of Z_u and Z_d

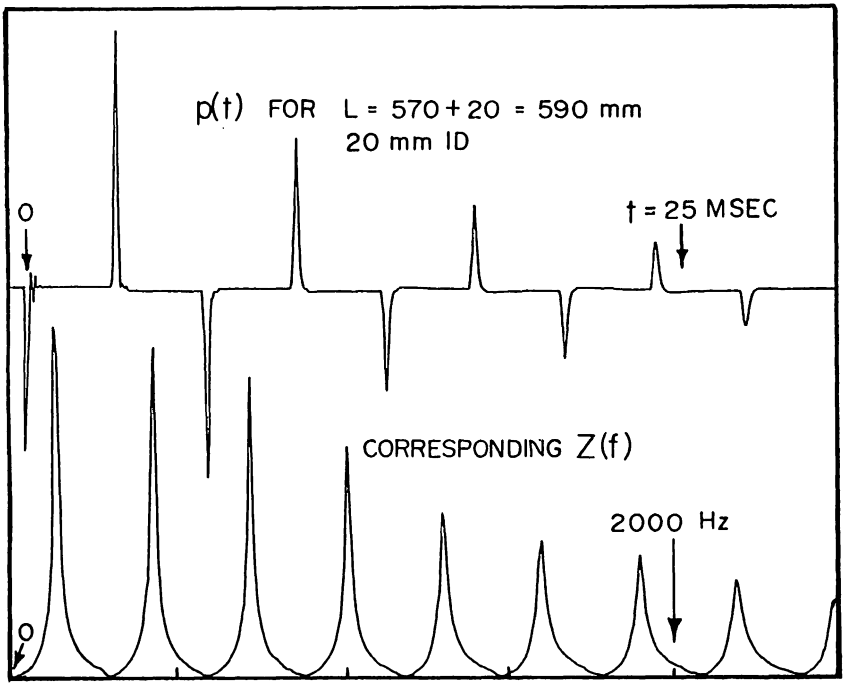


FIGURE 35-4

in a more-or-less musical context. The displays is linear in Z and in frequency. The tall peak visible in the neighborhood of 1300 Hz belongs to the measured Z_u under the following conditions: the subject has formed his vocal tract in the manner customary for articulating the vowel [ah] and the end of the impedance head has been inserted between his lips, with his teeth propped apart by the "tooth rest" on the impedance head to approximate the spacing they have on a clarinet. The less-tall sequence of resonance peaks that cross the entire figure reproduces the impedance curve for the 20-mm ID pipe shown in the preceding figure. The small circles placed above the peaks of this impedance curve show where they would be seen if the tube ID had been 15 mm rather than 20 mm. This 15-mm reference ID is chosen because it matches closely the size of the bore on a normal Bb clarinet (whose resonance peaks tend in fact to be less than about half as tall because of the additional damping associated with the complexities of the open and closed tone holes). It is clear that PWW impedance peaks can be very significantly taller than any of those that we might find in a real IAC. We will postpone any of the implications of this remark until Sec. VI, after we have looked at a little more data.

Figures 35-6, 35-7, and 35-8 show the measured Z_u curves (expressed logarithmically via the dB notation) as a linear function of frequency for the vowels [ah], [eh], and [ih]. In all cases the frequencies of the principal resonant peaks are marked, along with an indication of the wave impedance of a 20-mm ID tube, which we can use as a calibration value. It is not quite coincidental that in all three cases the low- and high-frequency limits of the measured curves match this reference value. We will return to a discussion of this phenomenon as soon as the figures have been described.

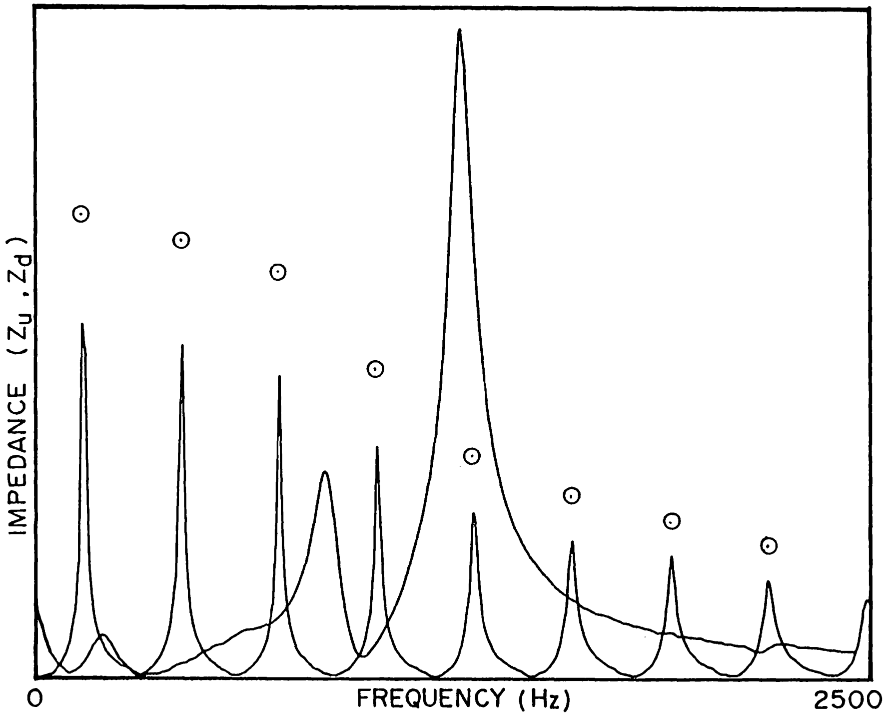


FIGURE 35-5

A dash-dot curved line will be seen in all three of these impedance diagrams. A separate curve of this type was originally drawn freehand on a copy of each resonance curve on the basis of criteria that will become clear very shortly. It was then verified that these three curves were all very similar. A single composite curve was then constructed by an informal averaging and smoothing procedure. It is this composite that is displayed.

In the theory of nonuniform horns it is convenient to define the wave impedance as seen at the input as the function $Z_0(\omega)$ that would be measured if the horn were to be given a reflectionless termination at its far end. This wave impedance can be real or imaginary, even if damping does not exist in the body of the horn. It is not difficult to show, then, that when the horn is given an arbitrary termination (real, imaginary, or complex), the measured input impedance will have maximum and minimum magnitudes that bound it in the following way:

$$(Z_{in})_{max} = [(1 + F)/(1 - F)]^{+1} \tag{35-17a}$$

$$(Z_{in})_{min} = [(1 + F)/(1 - F)]^{-1} \tag{35-17b}$$

Here the symbol F represents the fraction of the downward signal amplitude that returns to the input end after reflection from the termination at the other end. The attenuation considered here is due to all losses undergone in one round trip down and back in the waveguide plus the losses that take place for whatever reason at the termination itself. It is not possible to

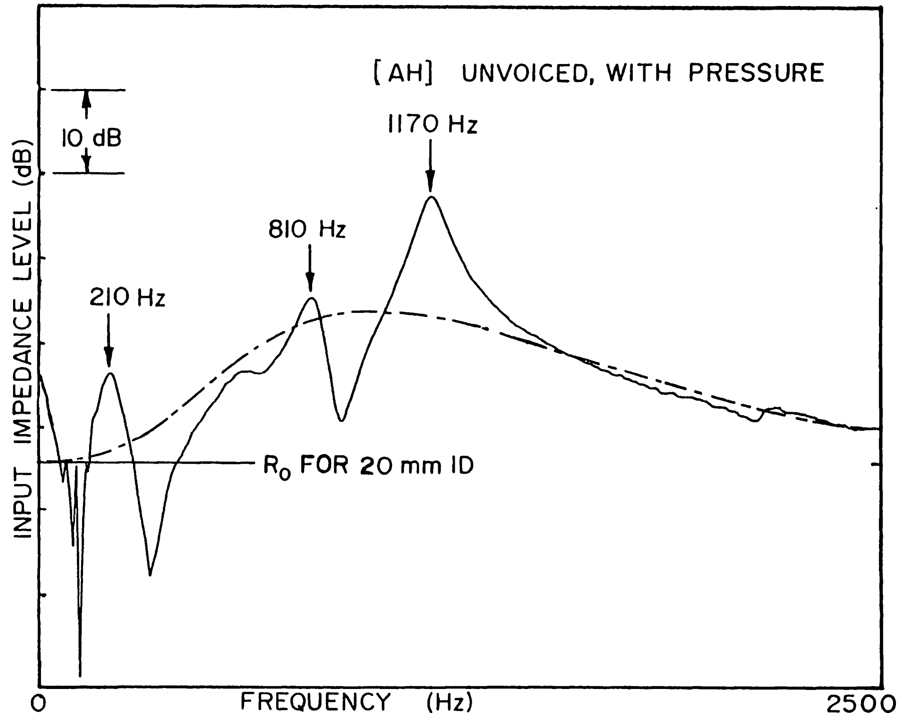


FIGURE 35-6

give a similarly straightforward account of the frequencies at which these extrema are found.

It is clear from Equations 35-17a and 35-17b that in many cases measurement of the peak and dip values of Z_{in} permits the determination of $Z_0(\omega)$-all one needs to do is to calculate the geometric means of adjacent pairs of Z_{max} and Z_{min}. The broken line in the figures is such a midline reconciled to give a curve for $Z_0(\omega)$ that is consistent with all the vowel configurations studied. The fact that various such configurations share a nearly common wave-impedance behavior is an interesting and surprising fact, concerning which we can extract some further information.

Horn theory tells us that in the limit of high frequencies, $Z_0(\omega)$ tends toward a value that is equal to the wave impedance R_0 of a uniform waveguide whose entering cross section matches that of the horn. Figures 35-6, 35-7, and 35-8 suggest on the basis of this property that the PWW has an entryway cross section that is close to that of a 20-mm ID pipe (our impedance head is short enough that its own cross section does not produce complications in any of the interpretations that we are making on the basis of general horn theory).

The fact that the peaks and dips in the measured Z_u's become less pronounced shows that the lungs (which serve as the termination of the PWW) are becoming less and less reflective at high frequencies. Thus, our original assumption that the PWW shows little or no resonant behavior is only justified in the limit of high frequencies.

We may usefully invoke horn theory to aid in the extraction of yet more information about the acoustical nature of the PWW. If the horn has a taper at its entryway (whether enlarging or contracting), then we find that the wave impedance rises proportionally to the frequency

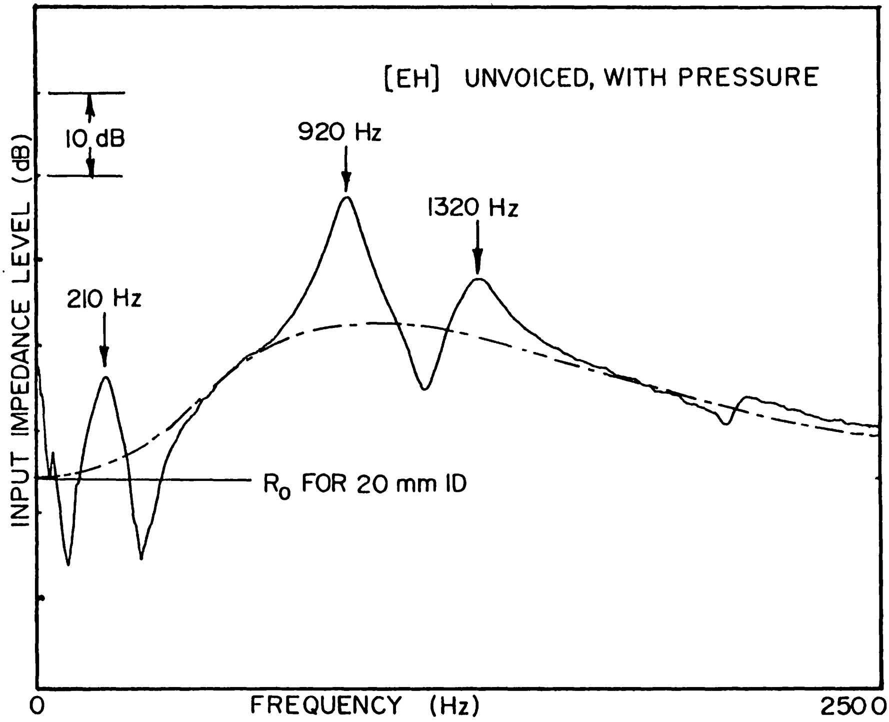


FIGURE 35-7

from a zero at zero frequency, and then that it ultimately levels out toward the value R_0 when $\omega x_0/c >> 1$. Here x_0 is the length of the apical cone implied by the horn's initial taper and its entryway cross section. The fact that our deduced curve for $Z_0(\omega)$ remains level at the value R_0 implies, then, that our PWW has very little taper in the region of the mouth (at least under the conditions of our measurements!).

Yet another piece of global information can be deduced from a study of our curves for the input impedance itself as measured for the PWW. If the PWW has an opening to the outer world anywhere along its length, then Z_{in} has a first-order zero at zero frequency. If, on the other hand, the PWW is air-tight except at its input end, the impedance has a simple pole at zero frequency. It is of course a truism that if the horn is given a nonreflecting termination, $Z_{in} = Z_0(\omega)$ at zero frequency (which limits to R_0). Inspection of the resonance curves of Figures 35-6, 35-7, and 35-8 therefore confirms that the player's lungs function as a closed and only somewhat reflecting termination for the PWW.

One final piece of information about the PWW can be gained from an examination of the pattern of its input impedance. It was, in fact, the feature that immediately called itself to my attention when the data were first in hand. All of the impedance curves showed a pattern of peaks and dips superposed on a broad hump centered in the neighborhood of 1000 Hz. Better put, there is a broad hump in the wave impedance itself, and this hump is fairly independent of the other details of PWW structure. This sort of overall pattern is reminiscent of the generic behavior of the input impedance for all valve settings of a brass wind instrument, where the explanation is well known (Benade 1976, sec. 20.4, p. 400). When ac-

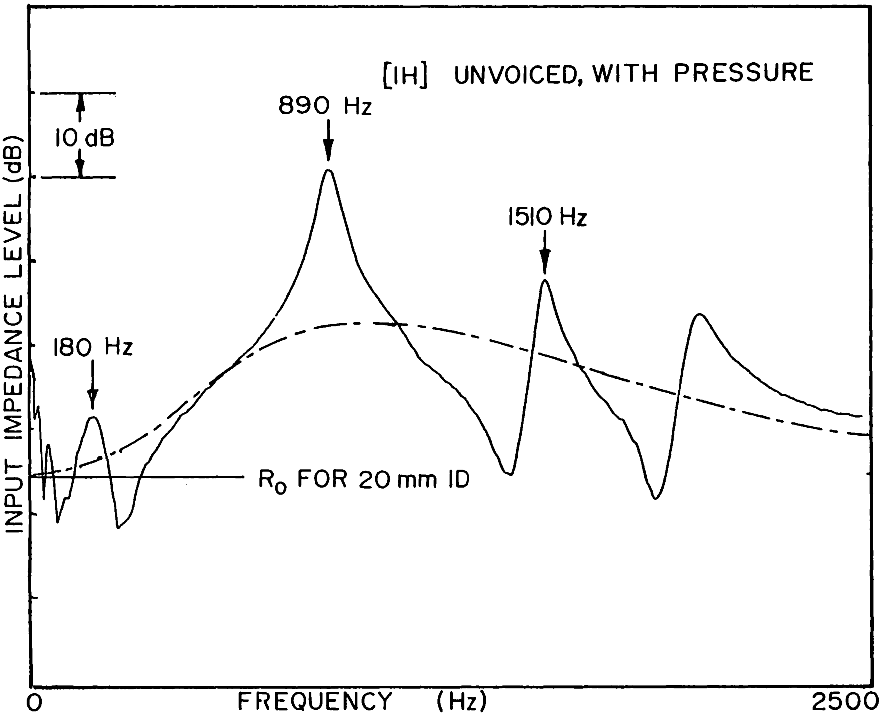


FIGURE 35-8

count is taken of the other systematic implications of the PWW shape as outlined above, one is led to suspect that it possesses a constriction a short way upstream, such that (speaking very informally), at the input end, a sort of Helmholtz resonator is constructed whose natural frequency matches that of the broad hump.

Figure 35-9 provides the first evidence of the plausibility of this sort of deduction. Here we see the input impedance of a very long piece of 20-mm ID tubing having an essentially nonreflecting termination. That is, we have a direct measurement of the wave impedance $Z_0(\omega)$. The horizontal line across the graph verifies that a uniform tube does in fact have a frequency-independent measured wave impedance. The upper curve shows the broad hump that was described in the preceding paragraph, as is produced by the introduction of a properly proportioned constriction at a suitable distance from the drive end. These proportions may be calculated from the position and the height of the hump, with no free parameters to take care of other features of the shape of the $Z_0(\omega)$ curve. There is a gratifying similarity between the shape of this curve and the (presumably analogous) freehand curve that was drawn on the earlier impedance curves.

Figure 35-10 shows the input impedance of an open-ended tube 1900 mm long into which the same constriction is installed. The $Z_0(\omega)$ curve from Figure 35-9 is superposed to show in summary form how precisely the general ideas of horn theory correspond with the experimental results. The high- and low-frequency limits of Z_{in} and the wave impedance are illustrated, along with the fact that on a dB plot, the wave impedance peaks and dips. The fact that in the present case the horn is open to the air is indicated by the low value of the

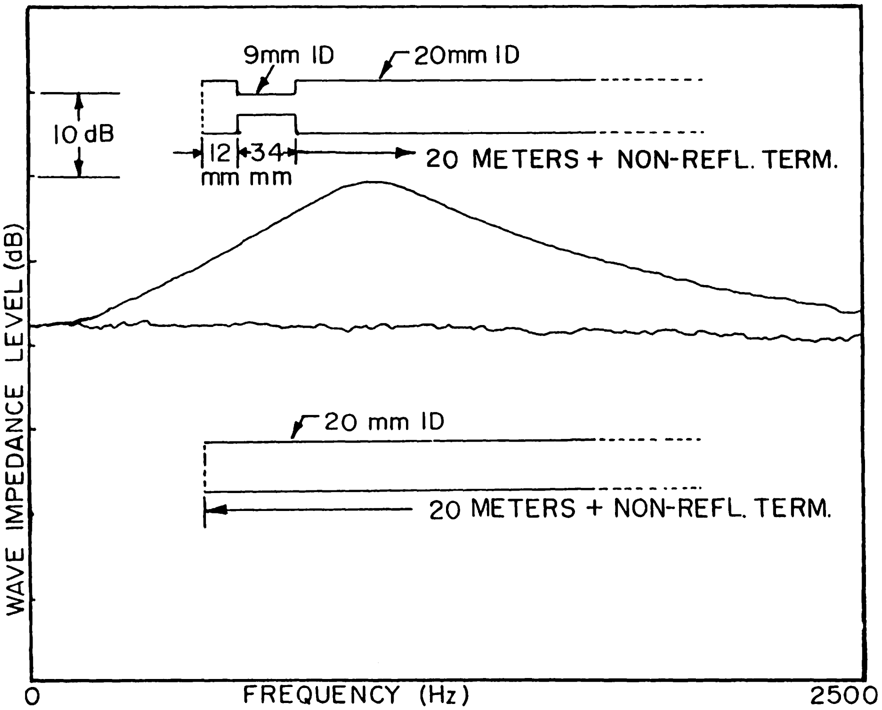


FIGURE 35-9

input impedance at low frequencies.

Figure 35-11 is a very similar illustration of the behavior of a nonuniform waveguide. The only essential feature in which this differs from its predecessor is the overall length of the tube, chosen this time to approximate the length of the PWW (as given in Figure 35-1). Despite the fact that the pipe is open at the far end, it is clear that the general pattern of peaks and dips is strongly reminiscent of those measured for various vowel configurations of the PWW. In particular, the mean spacing of the peaks is the same, as is expected for one-dimensional ducts of equal length.

VI. MEASURED SPECTRA

We will consider first the spectra measured on the two sides of the reed of a clarinet-like air column having three accurately aligned impedance maxima (located at odd multiples of 240 Hz). Above this (i.e., above a tone-hole-lattice cutoff frequency of about 1250 Hz), the impedance lies close to the value characteristic of an infinite line of 15-mm ID. This means that the IAC itself can generate energy strongly at the three odd harmonics of a 240-Hz playing frequency. Heterodyne transfer of energy between these harmonics constitutes the major source of energy in the entire oscillation. Absorption of energy at the impedance minima located at 480 and 960 Hz (even multiples of 240 Hz) and at all harmonic frequencies that lie above the tone-hole-lattic cutoff frequency assures amplitude stability.

On the upstream side of the reed we have the PWW, whose impedance function can be readily modified to give a wide variety of resonance peak distributions of the sort illustrated

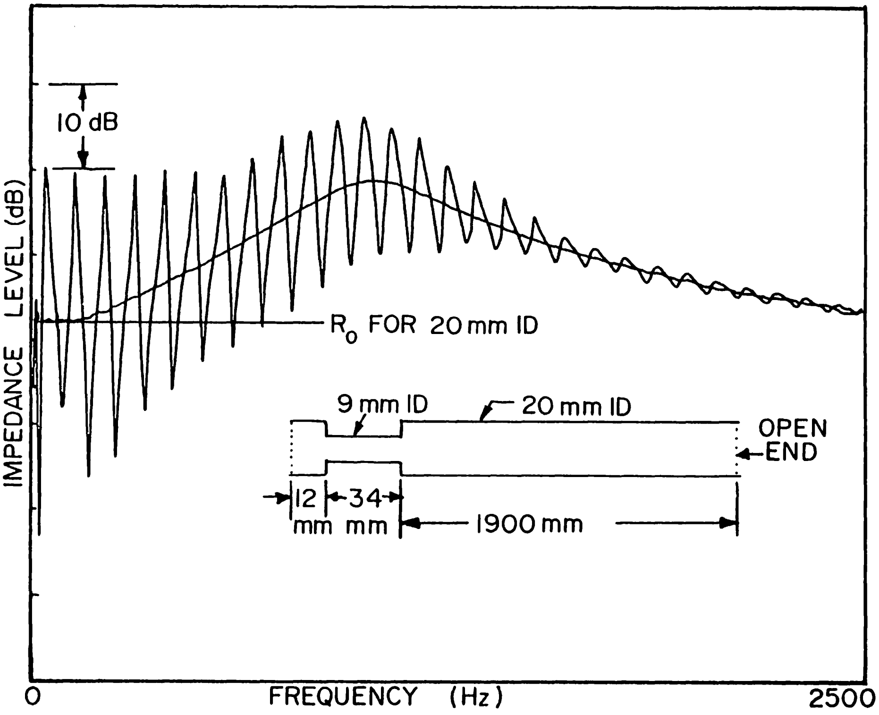


FIGURE 35-10

in Figures 35-6, 35-7, and 35-8 and discussed in Sec V. For present purposes it will suffice to recall that the tallest peak of $Z_u(\omega)$ tends to be two or three times as tall as the tallest peak of the IAC impedance curve.

The discussion in Sec. V showed that one obtains good, steady oscillation if the net impedance controlling the reed has a set of harmonically related peaks. We did not at that time face the question of what would happen to the system if (as in the present case) the net impedance has a set of harmonic peaks plus one or more inharmonic, "maverick" peaks as well. Our recent experiments show that (except under special circumstances to which we shall return shortly) a stable regime of oscillation normally sets itself up under the domination of the harmonic set of peaks. This "civilized" regime of oscillation involves energy production at a set of harmonic frequencies in the manner already outlined, WITH NO PRODUCTION AT THE FREQUENCY OF A MAVERICK PEAK! This seems a little surprising at first, but when we realize that if a sinusoid at the frequency of the maverick peak were to be produced, its nonlinearly generated harmonics, and the intermodulation product frequencies generated by it via nonlinear coupling to the harmonic components otherwise generated, would in general lie at points of the overall impedance curve where they would function as a heavy drain on the energy budget of the complete oscillation. It turns out, then, that such oscillations are not normally possible. As a result we find that the IAC dominates the sound production of our experimental system, as it does in the case of a real clarinet, with the PWW playing an auxiliary role in the manner which we are about to demonstrate.

The top half of Figure 35-12 shows the sound spectrum $(p_n)_d$ measured in the mouthpiece

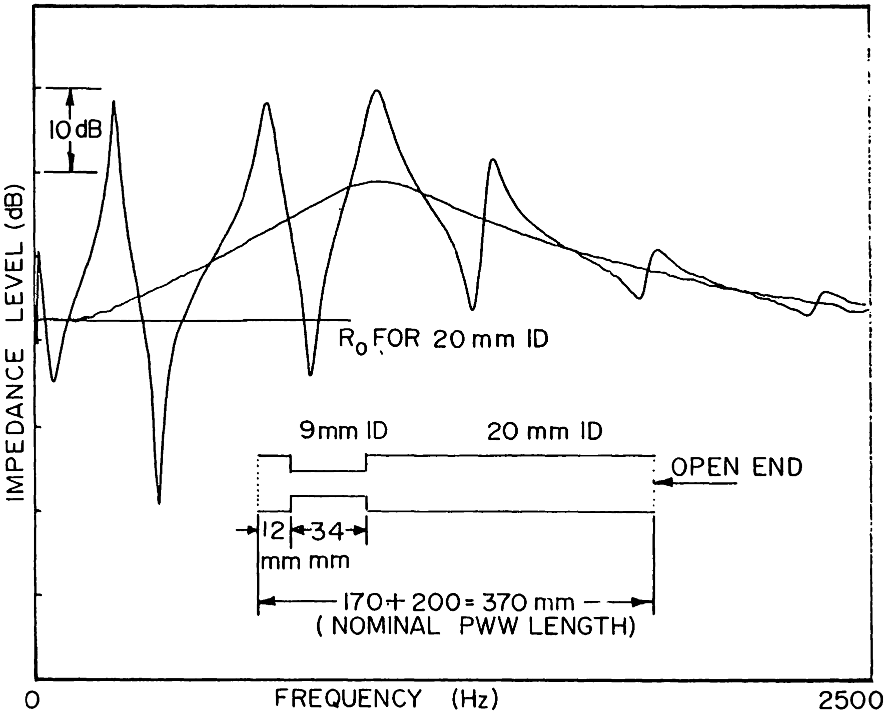


FIGURE 35-11

of our IAC under two experimental conditions. The lower half of the figure shows the corresponding two spectra $(p_n)_u$ measured in the player's mouth. The solid lines in both graphs shows the spectra measured under the "normal" condition that no resonances in the PWW are anywhere near the resonances of the IAC. The mouthpiece spectrum under these conditions is of exactly the form that has become familiar to us in our work with woodwinds over many years. Its shape is controlled by the impedance of the IAC, as we have become accustomed to take for granted (see Equation 35-8). We recognize, now, that the relative invisibility of the PWW generally comes from the fact that its resonances do not normally lie in places that give rise to the special circumstances referred to above, but not yet elucidated. The solid line in the lower part of the figure, which shows the nature of the corresponding spectrum in the player's mouth, does not have any particular features that we need to dwell upon at the moment.

If one listens to the signal picked up by a microphone within the player's mouth as he alters the configuration of his tongue, etc., it is at once apparent that one or another of the harmonics that belongs to the played note is boosted very considerably as a resonance of the PWW is brought into tune with it. The dashed lines in the two parts of Figure 35-12 show the nature of the spectral observed when the PWW has been configured so that the frequency of its major peak is brought into coincidence (at 960 Hz) with the fourth harmonic of the played note. We notice first of all that there is a 40-dB (100-fold) increase in the strength of this harmonic in the player's mouth, along with a considerable strengthening of all the neighboring harmonics as a result of intermodulation between this strong component and

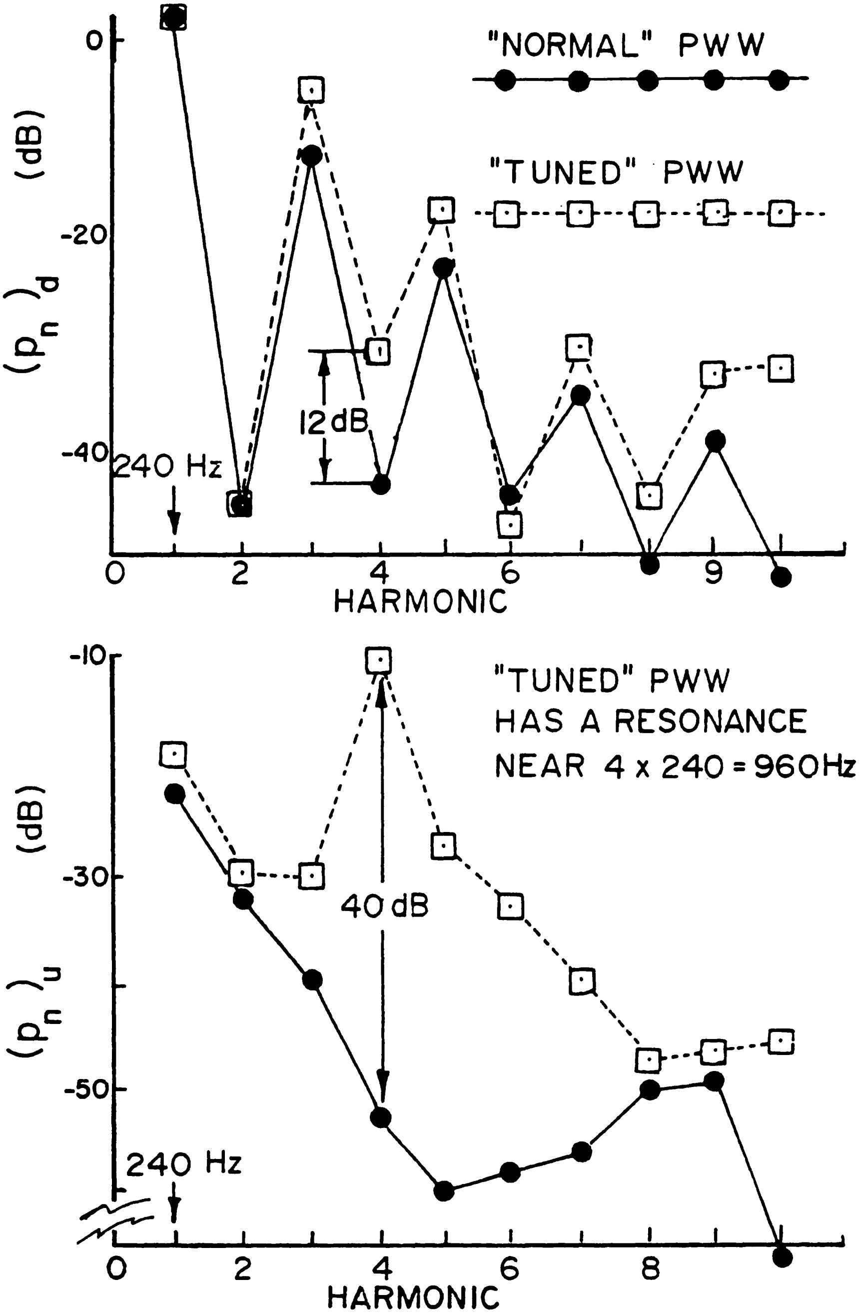

"NORMAL" PWW
"TUNED" PWW
$(p_n)_d$ (dB)
0
-20
-40
240 Hz
12 dB
0
2
4
6
9
HARMONIC
"TUNED" PWW
HAS A RESONANCE
NEAR 4 x 240 = 960Hz
$(p_n)_u$ (dB)
-10
-30
-50
40 dB
240 Hz
0
2
4
6
8
HARMONIC

FIGURE 35-12

its neighbors.

The spectrum measured within the mouthpiece of our "clarinet" under the special condition of harmonic tuning of the PWW (as given by the dotted lines in the upper diagram) shows no particularly noteworthy changes. The fundamental and second-harmonic amplitudes are essentially unchanged, while most of the other harmonics are strengthened by 5 or 6 dB (a doubling in amplitude). The fourth harmonic, which one might think should be greatly strengthened, has its level raised by only 12 dB (a four-fold increase in amplitude). These results illustrate and confirm the mathematical result described in Sec. IV to the effect that alterations in Z_u and in Z_d show their predominant effects on the same side of the reed as the changes are made, with only small changes being detectable on the other side.

One may confirm the perceptual smallness of the changes in the mouthpiece signal by listening to it while the PWW is being altered. One becomes aware only of a rather subtle change in the sound of the sort one is accustomed to hearing as a fine player makes an effort to achieve the best possible tone. In other words, the listener gets only a hint of the drastic changes that are taking place within the player's mouth. Our picture of what is going on as the clarinet reed collaborates with the IAC and PWW is at last becoming clear. Our understanding of the ways in which the PWW could be important in a general way without ever making its presence obvious is also beginning to develop. For example, proper alignment of a PWW resonance can improve a note by stabilizing the regime of oscillation and reducing the stray noise even though its influence on the externally perceived tone color may be quite small.

We now take our first step in describing the special circumstances under which the PWW can make its presence felt in an overt way. Consider a dynamical system in which the IAC is designed so as to give only a single strong resonance peak (an example of which is shown in the lower part of Figure 35-13, where the peak is at a frequency f_a = 340 Hz). If then the reed is instructed on its downstream side by such an air column and on its upstream side by some version of the PWW, there is no harmonic collection of resonance peaks that can cooperate in setting up a well-defined regime of oscillation and its corresponding harmonic spectrum. If the player will, however, explore the possible variations in $Z_u(\omega)$, it proves possible for him to find one or more cases where the tallest peak of the PWW resonance lies at such a frequency f_b that energy generation at this frequency and at f_a is sufficient to feed all of their intermodulation products, some of which may fall on top of one or more of the remaining (inharmonically positioned) resonance peaks of the PWW. The upper parts of Figure 35-13 show the enormously complicated mouthpiece and mouth spectra measured for such a special case, whose sound is recognizable by musicians to be of the complexly interwoven type that they are accustomed to calling a "multiphonic." Oscillatory energy is primarily produced near the frequency of the IAC impedance maximum and at one (or perhaps more) of the PWW resonance peaks. Because of the strong nonlinearity of the reed valve, these two or three primary spectral components have bred a whole host of intermodulation products. We have produced numerous versions of the IAC and PWW configuration described here, and are always able to produce multiphonics in exactly the same way. In all cases it has been possible to analyze the spectra in the manner already worked out for the analogous type of oscillation produced by an IAC whose resonances are inharmonic (Benade 1976, chap. 25).

There are two more special cases of IAC-PWW interaction that we need to consider. It proves readily possible for one to find configurations of the PWW in which the strongest of the resonance peaks is so placed that its coupling with the harmonically positioned resonances of the IAC can disrupt even a going regime of harmonic oscillation and replace

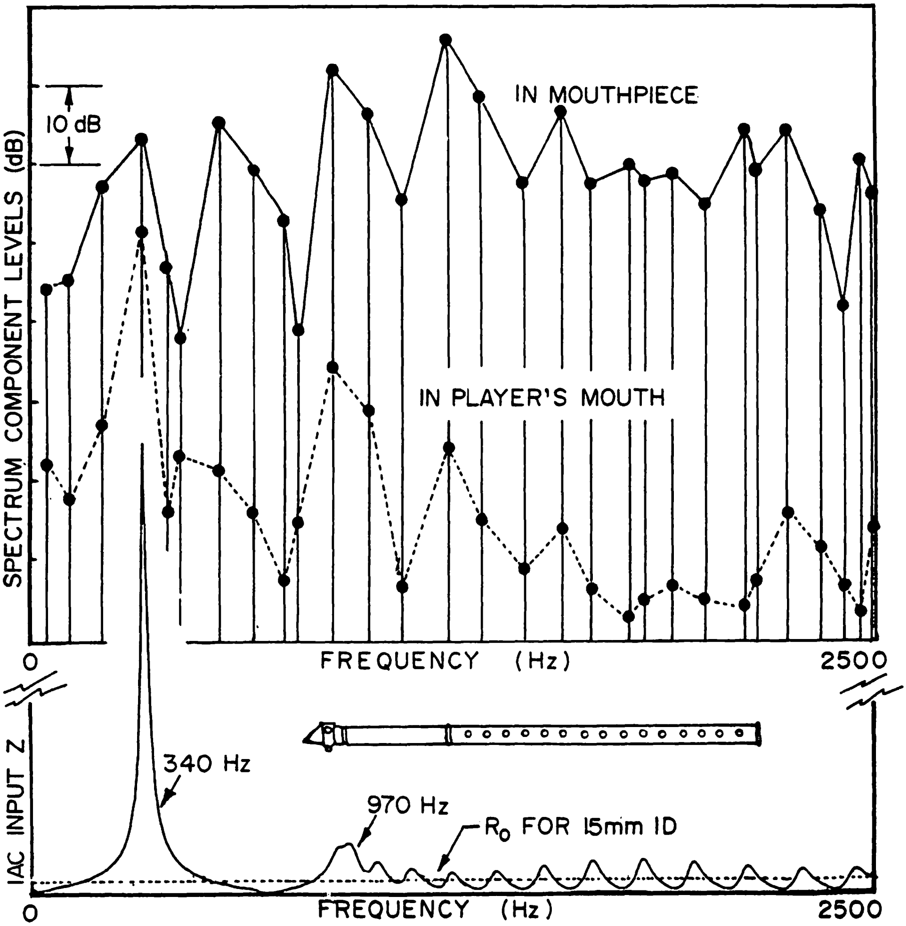


FIGURE 35-13

it with a multiphonic of rather considerable complexity. The precise dynamical requirements for this sort of transition have not been worked out so far, but it turns out that a player can, with only a little practice, learn to provoke such a multiphonic starting from almost any note of his normal scale. We will say no more about this class of phenomena (which needs considerably more study), turning instead to a dynamically much simpler, but musically more spectacular special case. A solution has been found to the rather challenging acoustical problem of designing an IAC (for use with a clarinet reed and mouthpiece facing) for which the input impedance Z_d seen by the reed is essentially resonance-free (basically real); the magnitude of this impedance is no more than about one tenth that of a normal 15-mm ID clarinet bore. Clearly, such an IAC can give no "instructions" to the downstream side of the reed, leaving it entirely under the influence of the PWW! Attempts to play on this air column are quickly rewarded with a variety of easily controllable tones whose pitches are determined directly either by the 450- to 1400-Hz range of easily available resonances of the PWW or by the 2000- to 3000-Hz range of reed resonance frequencies, where ZA can be large even when Z itself is no more than that provided at high frequencies by the PWW. In the first case, motions of the player's tongue, etc., vary the pitch, with very little change due to alterations of the embouchure tension that controls the resonance frequency of the reed itself. In the other case, the playing frequency depends on embouchure tension almost exclusively (via the reed frequency). It shows no influence from the (lower-frequency) resonances of the PWW that are controlled by the player's vocal tract configuration. The listener finds it very easy to recognize which one of the two ways of playing is in operation, and to recognize the transition of dominance from one to the other when the player manages to crowd one of them into the domain of easy playing of the other.

It is quickly apparent that a good reed-player has no trouble in learning to control the resonances of his PWW for quasimusical purposes. I possess a cassette tape on which microphone signals were recorded both from the upstream and the downstream sides of the reed while George Jameson performed (on demand, in quick succession, and with essentially no practice) a wide variety of familiar musical extracts, among which are themes from Haydn's Trumpet Concerto and Surprise Symphony, Mozart's Clarinet Concerto, the Sextette from Lucia di Lammermoor, and "Hearts and Flowers." In all cases the fluency of performance and accuracy of intonation was at least equal to that of any amateur whistler possessed of a good ear. It is essential to notice that I am NOT prepared to say that the ease with which this new mode of performance was learned gives evidence that the task is parallel to one that is familiar to a player in the normal course of his professional music-making! While it may turn out to be so after we have made a proper study of the question, we should remind ourselves at this point that the same player would probably have done equally well on a slide-whistle or a musical saw (which requires motor skills that are quite unfamiliar to him).

VII. CONCLUSION

For the present we will content ourselves with the knowledge that the reed has shown itself to be perfectly willing to take instruction from resonances on either its upstream or downstream sides. We have also learned that harmonically related resonances on the two sides of the reed can work together to give a steady and clean tone; that a maverick (inharmonic) resonance will not usually upset the steady operations of the reed controlled by a harmonically aligned set of resonances; and that, under special circumstances, it can disrupt the oscillation, producing instead a multiphonic of considerable complexity.

To the extent that our musical investigations have progressed, we have good reason to believe that most (though by no means all) present-day woodwind players use the resonances of their PWWs to "fill out," "clarify," or "fine- tune" the notes they play. Among the brass players the analogous behavior seems to be much more common. In any event, connections are beginning to be recognized between the age-old urgings of music teachers and the positions of various resonances in the PWW. We are also forced to be very much aware that the nature of these connections is not easily foreseen by the physicist. In particular, we must remember that it is utterly absurd to expect, a priori, that the positions of the vowel formant frequencies have any direct connection with the impedance maxima of the PWW as we have been discussing them. This is because the PWW resonances of interest to us are associated with the normal modes of an air column having a high impedance (i.e., a "closed" boundary at the player's mouth and a complicated boundary at the other end provided by the lungs, whereas the vowel formant frequencies belong with a (much shorter) air column that has a low-Z boundary at the mouth end and a high-Z (essentially closed) one at the larynx end!

In conclusion I will summarize the whole paper by outlining the reasons why all of these player's windway effects managed to stay out of scientific sight for so long. First of all, every player learns almost at once the easy task of avoiding "pathological" configurations of the PWW that give rise to those multiphonics that can take place despite the presence of aligned resonances of the IAC; these sounds can be recognized as the shrieks often produced by beginners.

Secondly, the audible effects of resonance alignment in the PWW are rather subtle and not readily recognized in the instrument's mouthpiece spectrum or in its concert-hall descendant. As we have seen, their existence was really only demonstrable once the overall picture had begun to be worked out. Many other subtleties of tone color and response had to be recognized and rationalized (if not quantified) before the PWW effects could be meaningfully studied.

A third reason, while quite personal to the present author, is quite typical of the cause of many inquiries in musical acoustics. At the beginning of my acoustical studies, I was a reasonably good player of many woodwinds, good enough to feel a little confidence in my judgements concerning tone, tuning, and response and their dynamical implications (subject always to revision by comments from professional players and by data acquired in the laboratory). This familiarity with instruments made it easy to devise check-experiments, and to notice phenomena that might or might not be ready for closer study. However, formal consideration of any physiological parameters had to be set aside in favor of the more accessible physics problem. In this spirit and under the stimulus of Bouasse's writings, we in Cleveland pursued the idea of aligning the IAC resonances and carried the physics far enough to guide my production of musically interesting clarinets, flutes, and other instruments (designed to suit my own taste, after which they generally won the approval of distinguished players). The fact was not recognized by me for several years that I was exploiting the reed resonance phenomenon via embouchure tension changes in the course of my playing and while I was making physical adjustments to the instruments. In retrospect, I realize how fortunate it was that my chief musical advisers concerning the scientifically most accessible clarinet family were all of the old school, who were accustomed to using these resonances in their own playing, so that my unconscious efforts were not contradicted or confused by my advisers' comments. (This resource now turns out to be exploited by only a certain fraction of the younger clarinet players, in part because today's commercial instruments are made in such a way that this kind of exploitation is difficult if not well-nigh impossible.) Exploitation of the reed resonance phenomenon is normal practice for the other woodwinds. In any event, once the cooperation phenomena were well in hand, it was possible to pin down the physics of the reed resonance effects (Thompson 1979) and to relate these to musical technique. Meanwhile, my musical friends kept insisting that embouchure tension was by no means the only means of physiological control available to the player and, from the start, I had my own feeling that they were right. As my own ability as a player grew, I became more and more aware of the possibilities and also increasingly confident (as other things fell into place) that the time was approaching that a worthwhile study could be made. The intellectual parallel of the PWW and the reed resonance studies was recognized and used as an encouragement. Before this, a certain amount of conscious brashness was required to simply leave something out of consideration until it forced its way into visibility!

The moral of the story is that progress in musical acoustics is (best? only?) made by a special kind of two-legged creature: one of its necessary legs is scientific, while the other is musical, and it is important that both of them be strong and in good working order.

Supported by a grant from the National Science Foundation.

REFERENCES

Backus, J. (1962). "Small vibration theory of the clarinet," *J. Acoust. Soc. Am.* 35, 301.

Benade, A.H. (1976). *Fundamentals of Musical Acoustics.* Oxford Univ. Press, New York.

Bouasse, H. (1920-30). *Instruments a Vent.* 2 Vols. Librarie Delagrave, Paris, I:115-116, 312-314, II:47.

Dawson, S.V. (1976). "Input impedance measurements of the respiratory system by impulse response," *Tech. Prog. Rep.* Harvard School of Public Health.

Fransson, F. (1965-75). A series of articles on the STL ionophone transducer published in the STL QPSR, Stockholm.

Fransson, F. and Jansson, E. (1975). "The STL ionophone transducer, properties and construction," *J. Acoust. Soc. Am.* 58, 910.

Ibisi, M.I. and Benade, A.H. (1982). "Impedance and impulse response measurements using low-cost components," *J. Acoust. Soc. Am.* 72, S63.

Krüger, W. (1979). "The impulse method of determining the characteristics of brass instrument response," *Proc. of the Musical Acoust. Conf.* (in German). Ceský Krumlov (Sept. 1979). Reprinted (in Czech) in Hudebni Nastroje 17, 52 (1980).

Oliver, B.M. (1964). "Time domain reflectometry," *Hewlett-Packard J.* 15/1. See also HP Application Note 75 "Selected articles on TDR applications.

Rosenberg, J. and Gordon, H. (1966). "The pulsed acoustic reflectometer," *Physics Lab. Proj. Rep.* Case Institute of Tech., Case Western Reserve Univ.

Thompson, S.C. (1978). "Reed resonance effects on woodwind nonlinear feedback oscillations," doctoral diss., Case Western Reserve Univ.

Thompson, S.C. (1979). "The effect of reed resonance on woodwind tone production," *J. Acoust. Soc. Am.* 66, 1299.

DISCUSSION

P. MILENKOVIC: Flute instructors also tell you to do things with your mouth. Do you have any ideas on how to attack this problem with the flutes?

A. BENADE: The effects of this kind are very, very small in flutes. I have said a few things about mouth interaction [A.H. Benade and J.W. French *J. Acoust. Soc. Am.* 37, 679, 1964], which are wrong. And John Coltman has said things [*J. Acoust. Soc. Am.* 54, 417, 1973] which I think now aren't quite right. They're not wrong. These effects are very small because it's longitudinal. They matter, of course. Now I can spot how a flute player has got his mouth by listening to it. The musicians we talked with differ about a very small effect compared with what I'm talking about here. [Added in proof: Walther Krüger of the DDR has recently shown large effects on a recorder.]

M. ROTHENBERG: I can't help asking this question because it's been on my mind for so long. Some years ago I did some experiments with a harmonica. Have you ever worked with that instrument?

A. BENADE: By chance a letter asking essentially this same question came from Roy Childs of the American Harmonica Society at just about the time all this was beginning to occupy my mind in a serious way. The harmonica reed is nominally of the "free" type with its own preferred frequency of oscillation. Working out the theory along lines very similar to those outlined in this paper shows (and experiment confirms) that oscillation is only possible when there is some inertance in the airway channel, and that the frequency of oscillation can be moved around by changing the viscous resistance (real) part of the airway impedance.

M. ROTHENBERG: I once heard somebody playing on the harmonica in a music store, making all this beautiful music. And I knew how to play "Oh Susannah!" you know. So I went up closer; I expected to see a gigantic machine with all sorts of buttons, but he had a little marine-band harmonica. So I took a course actually, with a friend of mine who plays harmonica, and I found out that you could do that reed bending. Furthermore, my experiments in blowing the harmonica with compressed air, using an adjustable acoustic loading behind the mouthpiece, also convinced me that much of the beautiful music that a good player can obtain from a small, single key harmonica is due to acoustical loading of the reed performed by appropriately shaping the lips, vocal tract and larynx. However, even though I had lived in the vocal tract, so to speak, for about 15 years, and had, I thought, a great deal of insight on what tract manipulations could cause significant changes in acoustic loading at the lips, after about 10 hours of expert instruction I could only manage to perform the very easiest examples of "reed bending" (which I would define as a pitch shift toward the opposing, supposedly inactive reed).

A. BENADE: Well, here's one thing that distracts people like us. The real part of Z down here talks to the imaginary part of the reed and vice versa in many of these situations. And so in pulling the frequency, that you can do if you put a handkerchief over the windway or something, you can move it a long way. We have backwards reflexes here. Thank you for raising the question because I had intended to give a little bow in the direction of Roy Childs. An Irish harmonica player also visited, and supplied a little of the stimulus.

36 ▪ THE VOICE SOURCE -THEORY AND ACOUSTIC MODELING

Gunnar Fant

INTRODUCTION

The acoustics of speech production is based on the concept of a source and a filter function - in a more general sense, a raw material and a sound-shaping process. In current models the source of voiced sounds is represented by a quasiperiodic succession of pulses of air emitted through the glottis as the vocal cords open and close and the filter function is assumed to be linear and time invariant. Speaker specific and contextual factors have not been given much attention. In spite of these shortcomings, speech synthesis has gained a fair quality but there remains much to improve in terms of voice quality and we still have difficulties in synthesis of female and children's voices.

In the last few years it has become apparent that in order to carry out a meaningful descriptive work we need a firmer theoretical basis of voice production including parameterization and data collection techniques, such as time- and frequency-domain inverse filtering and spectrum matching. The concept of source and filter will differ in a maximally true model of human voice production and in terminal analog synthesizers. Depending on the principle of synthesis, there exists a variety of combinations of source and filter functions that will provide one and the same, or approximately the same, output. Even in the "true" physiological and physically oriented model there exist alternative choices of source and filter function.

In general, source and filter have a mutual dependency or rather a common dependency on underlying phonatory and articulatory events mediated by mechanical and acoustical interaction. Thus, extreme articulatory narrowing disturbs the aerodynamic flow pattern and causes significant source changes. A vocal cord abduction changes not only the voice source but also the glottal impedance as part of the overall filter system.

SOURCE-FILTER DECOMPOSITION OF VOICED SOUNDS

The major theoretical complication in human voice production models is that in the glottal open state the sub- and supraglottal parts of the vocal tract are acoustically coupled through the time-variable and non-linear glottal impedance, whereas when glottis is closed the sub- and supraglottal systems execute approximately free and separate oscillations. Resonance frequency and especially band-widths may differ significantly in the two states, the major effects being seen as a "truncation" of formant time-domain envelopes in the open state, typical of F1 of maximally open back vowels (Fant, 1979, 1980, 1981, 1982; Ananthapadmanabha and Fant, 1982; Fant and Ananthapadmanabha, 1982).

The physically most complete speech production model that has been developed is that of Ishizaka and Flanagan (1972) and Flanagan *et al* (1975). With the two-mass model of the vocal cords incorporated in a distributed parameter system, their model does not have a specific source in the linear network sense. It is a self-oscillating and self-adjusting system, the main power deriving from the expiratory force as represented by the lung pressure.

In the work of Ananthapadmanabha and Fant (1982), the acoustical modeling of voice production starts by assuming a specific glottal area function $A_g(t)$ within a fundamental period and a specific lung pressure, P_l. The flow and pressure states in other parts of the system are then calculated by techniques similar to those of Ishizaka and Flanagan (1972)

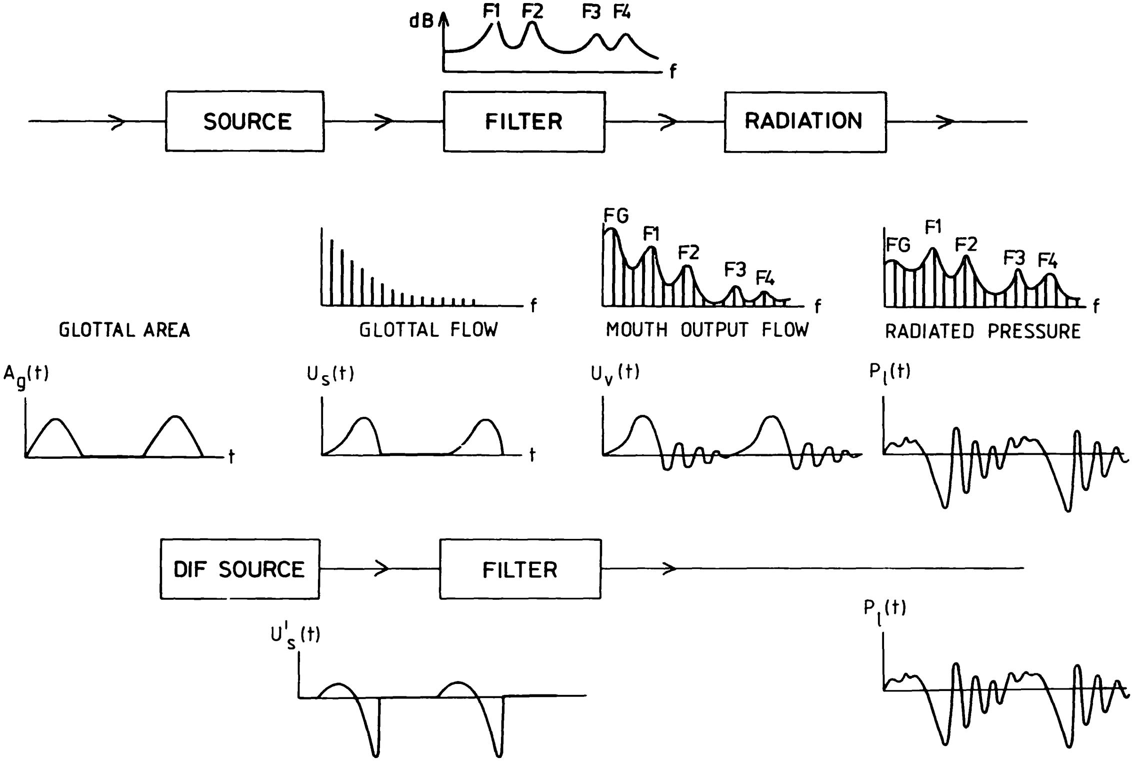


FIGURE 36-1
Source-filter representation of voiced sounds. The radiation transfer function is included as a differentiation of the source in the lower part.

leading to numerical determinations of the glottal volume velocity $U_g(t)$, the output volume velocity at the lips $U_0(t)$, and the sound pressure $P_a(t)$, at a distance of a centimeter from the speaker's mouth. By defining the filter function as the relation of $P_a(t)$ to $U_g(t)$ we define $U_g(t)$ as the source or we may go to the underlying glottal area function, $A_g(t)$ as conceptual reference which may attain the dimensionality of a flow source by being multiplied by a constant particle velocity $v_{go} = (2P_l/\rho k)^{1/2}$. This "shortcircuit" or "no-load" source

$$U_s(t) = v_{go}A_g(t) \tag{36-1}$$

differs from the true glottal source

$$U_g(t) = v_g(t)A_g(t) \tag{36-2}$$

in terms of the particle velocity function $v_g(t)/v_{go}$ which carries two main features. One is a pulse skewing effect (Rothenberg, 1980, 1981; Ananthapadmanabha and Fant, 1982; Fant, 1982) accounting for a latency of the peak of $U_g(t)$ compared to $A_g(t)$, a lower rise towards the peak, and a faster fall. The second feature is the existence of oscillatory ripple, usually of a frequency close to F1, superimposed on the glottal flow pulse.

We now have a choice of two different sources in the production process. With the more basic $A_g(t)$ proportional source we have a complicated non-linear time-variable filter function to cope with. Starting with the true glottal flow $U_g(t)$ as a source, the associated filter function is simply the supraglottal transfer function which is linear and only slowly time variable. On the other hand, the calculation of $U_g(t)$ is just as complex as calculating the output $P_a(t)$ and $U_g(t)$ contains properties of the entire sub- and supraglottal systems and is thus not invariant with the particular articulation.

The time- and frequency-domain, principal representation of voice production in Figure 36-1 incorporates these two alternative source functions and in addition the differentiated glottal flow $U_g(t)$ as a source which mathematically means that the differentiation inherent in the radiation transfer from lip flow $U_0(t)$ to sound pressure $P_a(t)$ is removed from the filter function and incorporated in the source function which enhances important properties of the source. The negative spike in $U'_g(t)$ representing the slope of $U_g(t)$ at closure is a proportionality parameter for all formant amplitudes (Fant, 1979; Gauffin and Sundberg, 1980). It may also be seen in the speech wave $P_a(t)$ at low F_O.

Another source characteristic apparent in the output at low F_O, high F_1, and low or medium voice intensity is a spectral maximum, F_g somewhere in the vicinity of one of the lowest harmonics, the relative amplitude of which stays rather invariant whilst F1 and higher formants gain in amplitude at higher voice effort (Fant, 1979, 1980). At a very low voice effort the source peak may dominate over F_1.

INTERACTION EFFECTS

The variability of the pulse skewing comparing different vowel articulations is illustrated by the model calculations of Figure 36-2 from Ananthapadmanabha and Fant (1982). Vowels produced with large vocal tract inductance, i.e., with a narrow but not extreme constriction, such as [a] with constricted pharynx and [i] with constricted mouth have a somewhat higher $U'_g(t)$ at closure than other vowels and thus, a greater intrinsic intensity. The variations are not large, of the order of a decibel.

Superposition of formant ripple from previous glottal periods may add to the complexity of the glottal flow waveform as illustrated by Figure 36-3. This effect may be prominent at

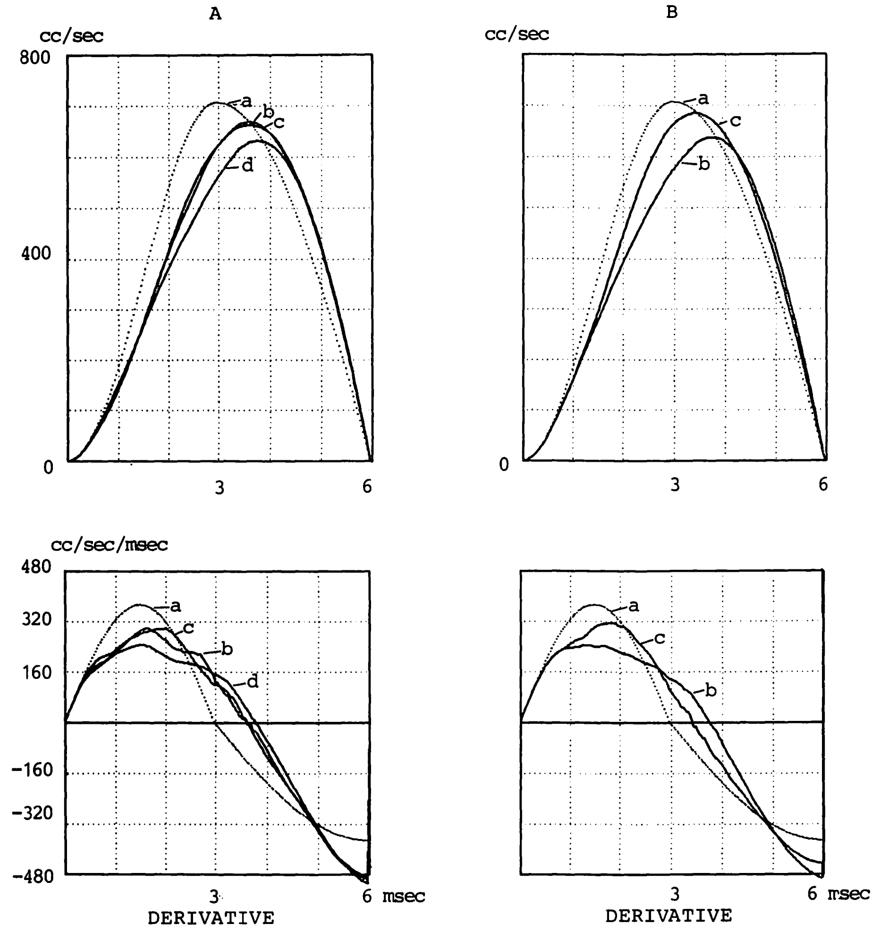


FIGURE 36-2

Glottal source pulse and its derivative for vowels. A) a) no load, b) vowel [a], c) vowel [o], d) vowel [u]. B. a) no load, b) vowel [i], c) vowel [ɛ] (from *STL-QPSR* 1/1982).

very low F_1 and relative high F_0 causing a seemingly random perturbation of glottal flow waveshape from one period to the next. The within a period induced ripple function already to be seen in Figure 36-2 is especially apparent in the model calculations of Figure 36-4 pertaining to a vowel [a] produced at $F_0 = 170$ Hz. A double peak is seen in the waveform of the differentiated glottal flow and the spectral correlates include a prominent antiresonance just above F_1. Such spectral irregularities vary with the duration of the glottal open period and thus with F_0, see Fant and Ananthapadmanabha (1982).

Even though the fine structure of the $U_g(t)$ source spectrum becomes complex, the time domain correspondence referring to the output sound may be simpler. Part of the glottal flow ripple is a mathematical prerequisite to a detuning and truncation of a formant oscillation when it reaches next vocal open period. One can visualize this as the addition of a compo-

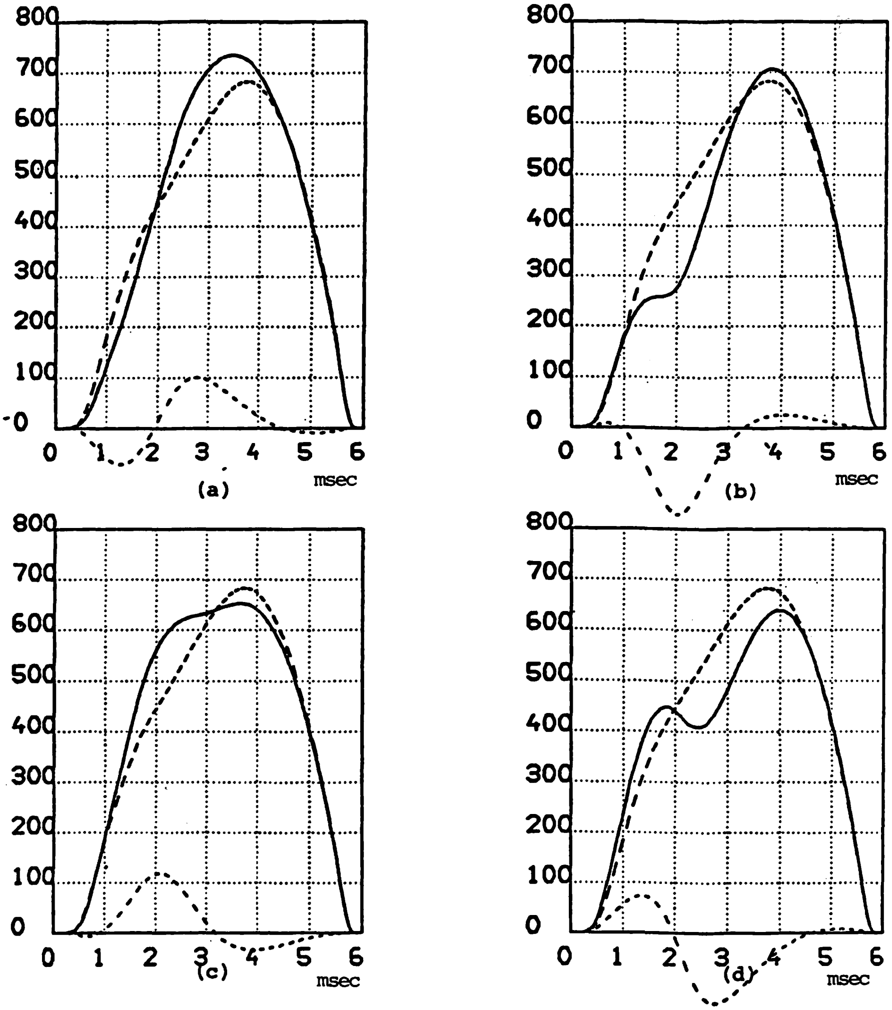


FIGURE 36-3
Effect of superposition on the glottal flow pulse. Dotted line, without superposition; solid line, with superposition. a) closed phase 6 ms, b) closed phase 7 ms, c) closed phase 4.75 ms, d) closed phase 3.625 ms. F1 VT load with F_1=220 Hz (from *STL-QPSR* 2-3/1982).

nent in the flow which will detune and partially cancel the formant oscillation which otherwise would have proceeded with a constant rate of damping.

This view also leads us to an approximate model where the source pulse is ripple free but properly skewed versus the glottal area function $A_g(t)$. Such a source supplied as a constant current source to the vocal tract supraglottal system in parallel with the time-varying glottal- plus subglottal impedance was adopted in the modeling of Fant (1979, 1981). During the glottal open phase the shunt impedance draws current which is the negative of the ripple component in the vocal tract input flow.

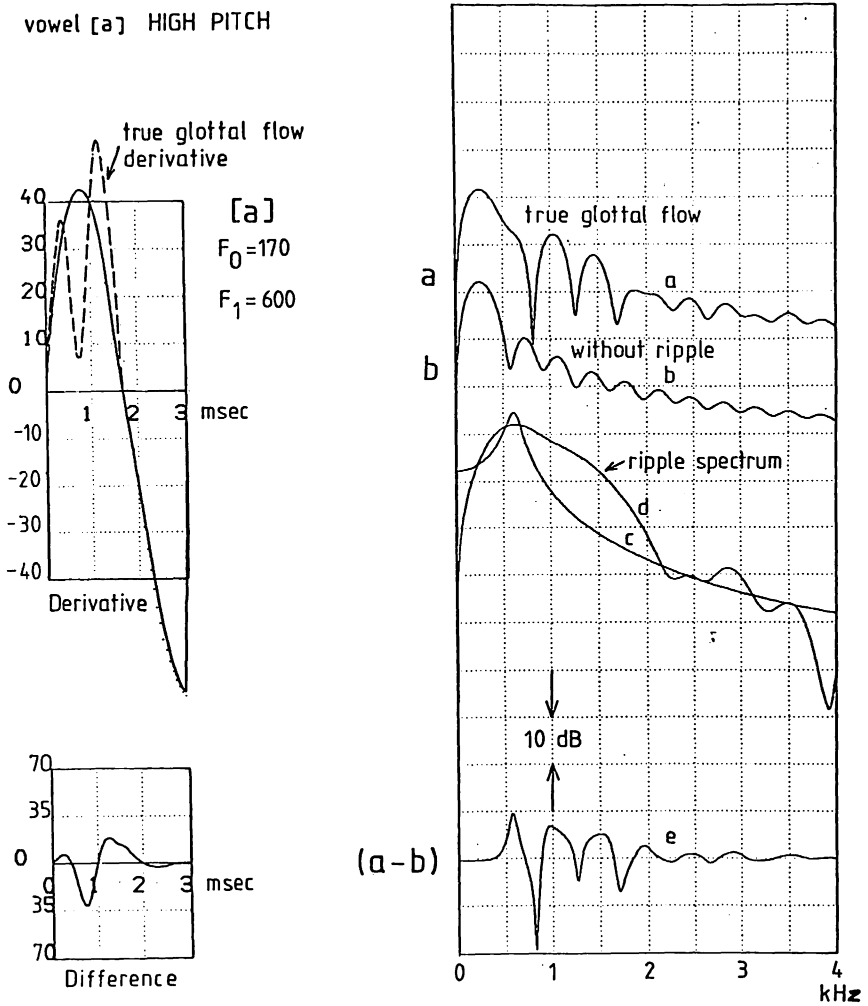


FIGURE 36-4

Example of differentiated glottal flow showing double peak. Vowel [a], F_1=600 Hz, F_0=170 Hz (from *STL-QPSR* 2-3/1982).

By removing the time variable shunt and omitting the pulse skewing effect, we end up with the noninteractive model of a terminal analog formant synthesizer excited by constant shape smooth pulses. In order to make this synthesis model more compatible with the true model, we could adjust the pulse skewing according to the particular vocal tract inductance and the particular glottal impedance. Furthermore, we need to introduce truncation and maybe also detuning by modulating, in the first hand F_1 and B_1 within the glottal open period. An alternative close to the "true" model is to start with an $A_g(t)$ source if known and to calculate the current into a Foster reactance network representation of the vocal tract load approximated by F1 and F2 parallel resonance circuits. This procedure first suggested by Mrayati *et al* (1976)

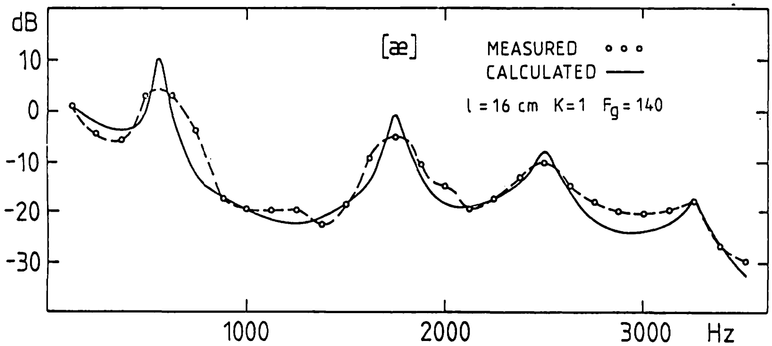


FIGURE 36-5
Spectrum matching with noninteractive model (from *STL-QPSR* 2-3/1982).

also takes into account the main pulse skewing. The load current is then transferred as the input to a conventional formant synthesizer.

We are now in a process of evaluating the perceptual importance of the interactive factors in production (Ananthapadmanabha *et al*, 1982). It appears that as long as relative formant amplitudes, including F_g are preserved, the perceptual differences are small. A preservation of the spectral fine structure between peaks or the corresponding time-domain properties of instantaneous formant frequency and rate of envelope decay appears to be less critical.

The difference between a measured harmonic spectrum of an [æ] vowel and a synthetic spectrum generated from a non-interactive model with a Fant (1979) source is illustrated in Figure 36-5. The bandwidth of formants was set to glottal closed average data of Fant (1972) which account for the overshoot of formant amplitude levels. The tendency of dispersion of spectral energy in the F1 range towards higher frequencies is typical of the detuning effect in the glottal open interval.

Pulse models. Breathy phonation

A parameterization of the glottal flow pulse or of the underlying glottal area function, see Figure 36-6, can be based on more or less established measures such as open quotient, speed quotient, and peak sharpness as defined by Lindqvist-Gauffin (1965). The measure "pulse asymmetry" which is almost the same as speed quotient is suggested as a contribution to the nomenclature.

Figure 36-6, drawn from the Fant (1979) model, does not include leakage. A more general model capable of representing non-abrupt return to the zero-line after the point of maximum discontinuity is included in Figure 36-7. The pulse is described in terms of parabolic consequtive parts which means that the flow derivative is a piecewise linear function and the second derivative a set of step functions and a spike at the point of maximum flow discontinuity. This model is outlined here for pedagogical purposes. Higher order power functions or connecting sinusoids would be more realistic (Ananthapadmanabha, 1984).

The need for the extension, Figure 36-7 of the Fant (1979) model, is apparent when studying inverse filtering functions from breathy phonations, such as the intervocalic voiced [h] in Figure 36-8. The differentiated glottal flow reveals the phase of continuing closing movement after the point of maximum discontinuity which perturbs the negative spike towards

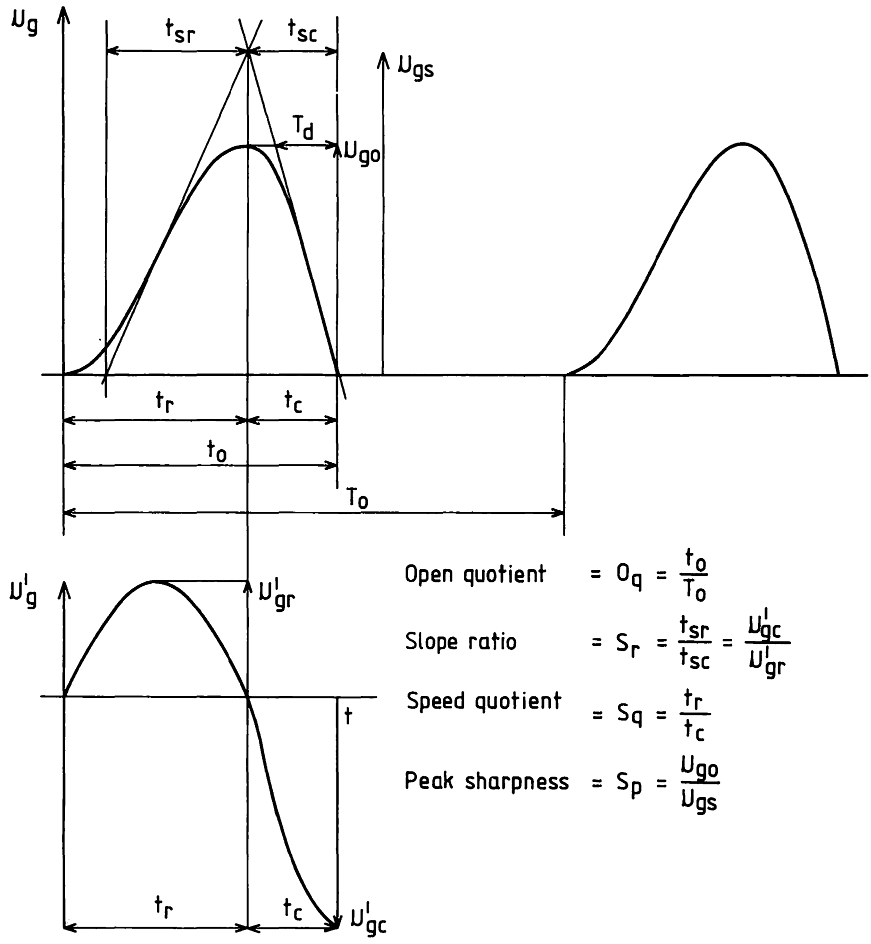


FIGURE 36-6
Pulse parameters for idealized case with complete closure for describing glottal flow or glottal area function.

a slightly smoothed triangular shape, thus providing an overall picture of a fullwave rectified pulse train. The reduced efficiency in formant excitation of this pulse shape combines with a substantial F1 bandwidth increase which typically reduces the F1 amplitude in breathy phonation. This is an example of covariation of source and filter functions.

Source spectrum shapes are not adequately described by a single slope value only. An important feature derivable from the particular combination of the F_g, FO, and K parameters (Fant, 1979) is the amplitude level at some higher frequency, e.g., at 1000 Hz. In connected speech, the F_g level stays rather invariant whilst formant levels in addition to F-pattern-induced variations tend to fall off as a result of a progressive abduction, e.g., at the termination of voicing. Similar effects occur as a reaction from supra-glottal narrowing when extended beyond that of close vowels. The amplitude level of F1 relative to that of F_g and also the absolute level of F_g are important correlates of breathy phonation (Bickley, 1982). Rules

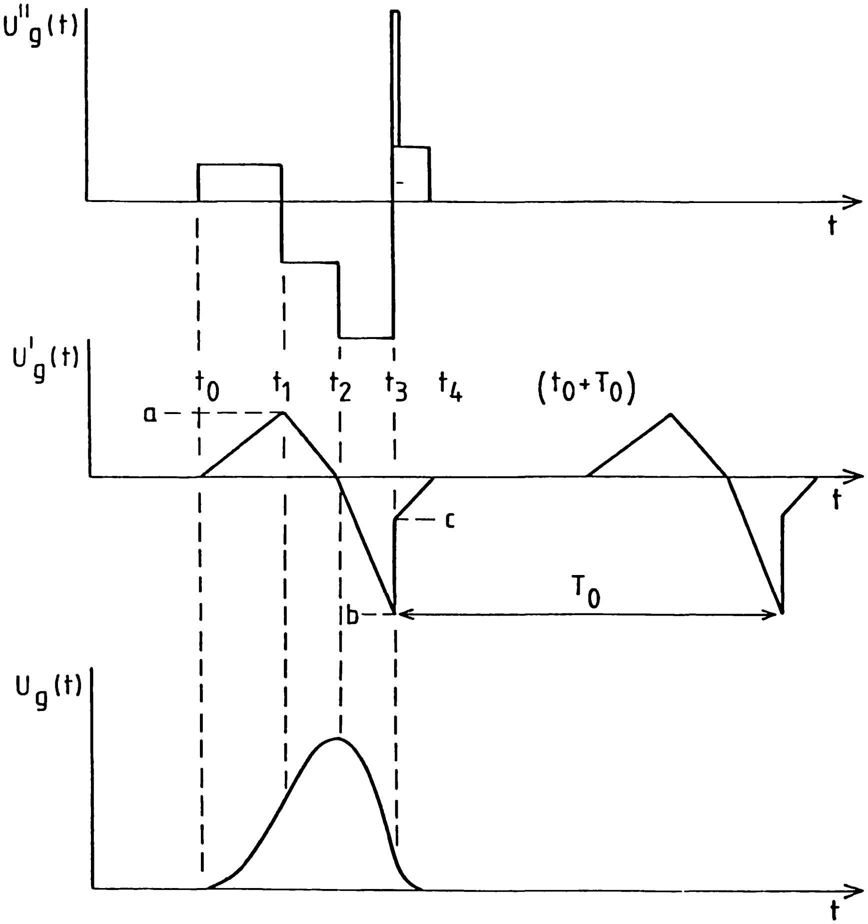


FIGURE 36-7
Stylized, more general model of glottal flow and its first and second derivative.

for predicting such relative amplitudes have to be based on the entire production process.

The relative role of lung pressure as a determinant of voice output intensity has been overestimated. It has been found experimentally that a doubling of subglottal pressure is associated with a raise of the overall sound pressure level by about 9 dB (Ladefoged and McKinney, 1963; Isshiki, 1964). However, if we assume that the glottal area-function and F_0 remain the same, the increase in particle velocity and thus in glottal flow from Equation (36-1) is 3 dB only. At increasing pressure the pulse skewing decreases and the net gain in flow derivative at closure would be slightly lower than 3 dB. What is then the origin of the remaining 6 dB? As shown by Fant (1982) it may be explained by several covarying factors. Increasing subglottal pressure increases the maximum deflection of the cords and we may expect a 3 dB increase of the overall scale factor of the glottal area function. Furthermore the pulse duration and especially the duration of the closing phase decrease and the fundamental frequency increases somewhat. There is also a small net in the pulse skewing related to

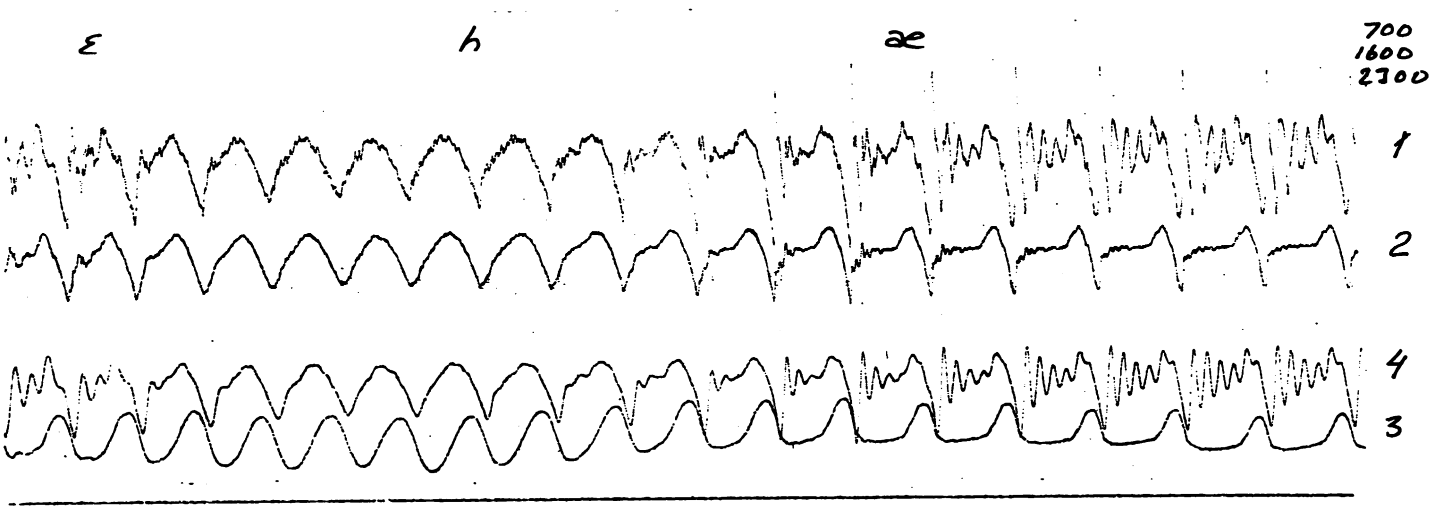


FIGURE 36-8
Intervcalic voiced [h], 1=oscillogram, 2=differentiated glottal flow from inverse filtering, 3=glottal flow, 4=same as 3 but without F1 cancellation (from *STL-QPRS* 2-3/1982).

the increase in rate of change of glottal area at closure. All these factors added together account for the missing 6 decibel. This example could illustrate a "flow" type of voice. If the cords are subjected to a medial compression counteracting the pressure increase the net gain in intensity would be related to a shortening of the glottal pulse rather than to an increase in maximum glottal area. In speech it is found that maximum glottal flow varies rather little with increasing voice intensity, Lindqvist-Gauffin (1965, 1970). Apparently there exists a large range of normal covariation of the physiological factors determining voice intensity.

These examples indicate the need of anchoring the rules of source dynamics on an insight in the underlying, more basic physiological parameters and production events. On the whole, it appears that abduction-adduction and other vocal cord activities carry the major part of the source dynamics whilst lung pressure variations are slower and carry less information.

A promising attempt of constructing a voice source with physiologically oriented parameters is that of Rothenberg *et al* (1975). We are now looking forward to widening our knowledge base to meet the demands of improved and better quantified production rules and synthesis strategies.

REFERENCES

Ananthapadmanabha, T.V. (1984). "Acoustic analysis of voice source dynamics," *STL-QPSR* 2-3/1983.

Ananthapadmanabha, T.V. and Fant, G. (1982). "Calculation of true glottal flow and its components," *STL-QPSR* 1/1982, 1-30.

Ananthapadmanabha, T.V., Nord, L. and Fant, G. (1982). "Perceptual discriminability of nonexponential/exponential damping of the first formant of vowel sounds," in *Proc. of the Representation of Speech in the Peripheral Auditory System*, R. Carlson & B. Granström (Eds.), Elsevier Biomedical Press, Amsterdam, 217-222.

Bickley, C. (1982). "Acoustic analysis and perception of breathy vowels," MIT, Speech Commun. Group Working Papers 1.

Fant, G. (1972). "Vocal tract wall effects, losses and resonance bandwidths," *STL-QPSR* 2-3/1972, 28-52.

Fant, G. (1979). "Glottal source and excitation analysis," *STL-QPSR* 1/1979, 85-107.

Fant, G. (1980). "Voice source dynamics," *STL-QPSR* 2-3/1980, 17-37.

Fant, G. (1981). "The source-filter concept in voice production," *STL-QPSR* 1/1981, 21-37.

Fant, G. (1982). "Preliminaries to analysis of the human voice source," *STL-QPSR* 4/1982, 1-27.

Fant, G. and Ananthapadmanabha, T.V. (1982). "Truncation and super-position," *STL-QPSR* 2-3/1982.

Flanagan, J.L., Ishizaka, K. and Shipley, K.L. (1975). "Synthesis of speech from a dynamic model of the vocal cords and vocal tract," *Bell Syst. Tech. J.* 54/3, 485-506.

Gauffin, J. and Sundberg, J. (1980). "Data on the glottal voice source behavior in vowel production," *STL-QPSR* 2-3/1980, 61-70.

Isshiki, N. (1964). "Regulatory mechanisms of voice intensity variation," *J. Speech Hear. Res.* 7, 12-29.

Ishizaka, K. and Flanagan, J.L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* 51, 1233-1268.

Lindqvist-Gauffin, J. (1965). "Studies of the voice source by means of inverse filtering," *STL-QPSR* 2/1965, 8-13.

Lindqvist-Gauffin, J. (1970). "The voice source studied by means of inverse filtering," *STL-QPSR* 1/1970, 3-9.

Ladefoged, P. and McKinney, N.P. (1963). "Loudness, sound pressure and sub-glottal pressure in speech," *J. Acoust. Soc. Am.* 35, 454-460.

Mrayati, M., Guerin, B. and Boe, L.J. (1976). "Etude de l'impedance d'entre du conduit vocal-couplage source - conduit vocal," *Acustica* 35, 330-340.

Rothenberg, M., Carlson, R., Granström, B. and Lindqvist-Gauffin, J. (1975). "A three-parameter voice source for speech synthesis," *in Speech Communication*, Vol. 2, G. Fant (Ed.), Almqvist & Wiksell, Stockholm, 235-243.

Rothenberg, M. (1981). "An interactive model for the voice source," *STL-QPSR* 4/1981, 1-17.

DISCUSSION

H. TEAGER: Can I presume from the mouth output flow shown in your figure that reversing (i.e. inward as well as outward) flows are a necessary feature of your vocal model?

G. FANT: Yes. This is so both in theory and practice, see e.g. my article in STL-QPSR 1/79. The simplest way to demonstrate the negative part of the mouth flow is to hook up a simple integrator to a condenser microphone placed close to the lips. In the time function as displayed by a scope, you see the F1 ripple starting with a negative component. It is also to be seen in the output of the Rothenberg mask. These are established facts. In a breathy voice, on the other hand, there is a superimposed airstream which generally is large enough that it exceeds the negative ripple amplitude.

P. KITZING: Can mathematical models of phonatory function explain the so called "dicrotic" wave shape in vocal fry?

G. FANT: Possibly. But I can only speculate that a closure, interrupted by a partial short opening after the lower parts of the cords have reached contact and before the upper parts close, might explain the phenomenon.

I. TITZE: Can you comment on how the closing slope (and hence the acoustic intensity) changes when Fo gets close to the first formant frequency? Is there a limit to the closing slope? In Rothenberg's model it can go to infinity, but I have heard you say that it cannot.

G. FANT: I do not have a simple answer. The flow will be highly influenced by F1, thus causing deviations from a straight line closure. The flow slope at closure cannot reach infinity unless the underlying glottal area function is a step function.

M. ROTHENBERG: The infinite terminal slope occurred in the solution of the nonlinear differential equation for the simplified source-tract system (a series configuration of a constant lung pressure P_L, and time-varying glottal resistance R_g, and a vocal tract inertance L_t) only if the glottal resistance is assumed not flow-dependent, and proportional to glottal area. When the glottal resistance is assumed to be inversely proportional to flow, as is believed to be the case for all except the smaller values of glottal area, Dr. Fant has shown that the terminal slope never becomes infinite, though it still does increase with L_t. If one adds to the flow-dependent glottal conductance a glottal resistance at small areas that is determined primarily by viscous losses, the terminal slope will increase considerably, since this conductance term is not flow-dependent. However, these more realistic models for the glottal source may still not yield an infinite terminal slope for finite L_t.

37 ▪ SOURCE-TRACT ACOUSTIC INTERACTION IN BREATHY VOICE

Martin Rothenberg

INTRODUCTION

It has been shown that nonlinear acoustic interaction between the glottal source and the low frequency inertance of the vocal tract (subglottal and supraglottal) during normal voicing can cause a skewing of the glottal flow pulse to the right, with a resulting increase in the high frequency acoustic energy generated at the instant of glottal closure and decrease in the high frequency energy generated by the opening of the glottis. For typical values of voice fundamental frequency and formant frequencies, these variations in high frequency energy help determine the energy at the second and higher formants, and therefore are an important determinant of voice quality. For open vowels, the degree of this type of source-tract acoustic interaction has been shown to depend on a dimensionless parameter which we have previously termed $\bar{L}_t$, the "normalized inertance", but for which we will hereafter use the symbol α, adapting a notation closer to that used by Fant (Fant, 1982), since the symbol $\bar{L}_t$ may imply that the parameter has the units of inertance (often symbolized L). The interaction parameter α is related to the total subglottal and supraglottal vocal tract inertance L_t as seen by the glottis by the relationship

$$\alpha = \frac{L_t \hat{G}_{gMAX}}{\tau_p/2} \qquad (37\text{-}1)$$

where τ_p is the duration of the glottal pulse, and $\hat{G}_{gMAX}$ is the maximum value of U_g/P_g that would be attained during the glottal cycle if there were no vocal tract loading, i.e. if L_t were zero, P_g being the transglottal pressure (constant and equal to the lung pressure P_L if $L_t = 0$) and U_g the glottal volume velocity (Rothenberg, 1981a; Fant, 1982).

In this paper we will show that during breathy voice produced by an abduction of the vocal folds, the vocal tract inertance has an opposite effect; it reduces the energy at the higher formants by smoothing the flow waveform. The point of excitation of the higher formants remains at the instant of glottal closure, but, as observed by Fant (1980) this instant is moved toward the middle of the falling segment of the flow waveform (instead of being at the end of the falling segment as in normal voice), as the fundamental frequency component of the flow waveform is delayed by the interaction of the vocal tract inertance and glottal conductance.

Moreover, it will be shown that the amount of the delay in the fundamental frequency component can be used to estimate α, the vocal tract interaction parameter during normal voice, and therefore can be used to test models for glottal source aerodynamics.

ANALYSIS DURING BREATHY VOICE

When the voice is made "breathy" by medial abduction of the vocal folds (and not by abduction of just the posterior interarytenoidal segment, as sometimes occurs (Rothenberg 1973)), the projected glottal area can be expected to have the form shown in the lower solid trace in Figure 37-1. During the "open" segments, labled 0, the folds are vibrating freely without any appreciable contact with each other. If there is a strong phase difference between the vibration patterns at the superior and inferior fold margins, the projected area (PGA)

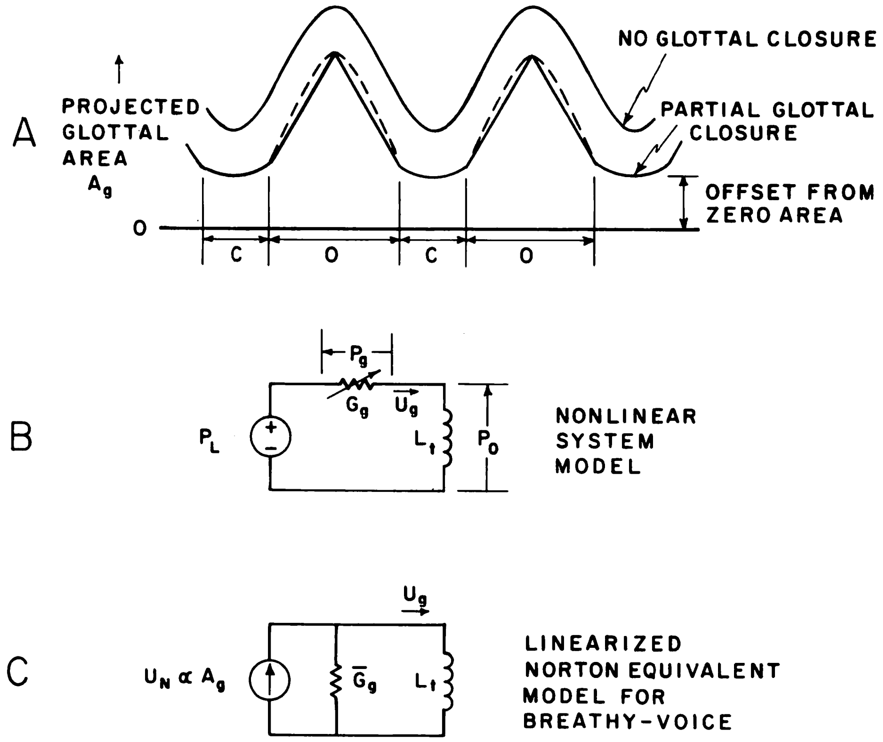


FIGURE 37-1

Development of a Norton-equivalent linear model for the glottal source during breathy voice. A. Idealized waveforms of projected glottal area for two degrees of vocal fold abduction. B. Simplified nonlinear interactive source-tract model. The incremental glottal conductance G_g is usually assumed to be proportional to A_g^2/U_g for large glottal areas (the flow-dependent case), and proportional to A_g^3 for vocal fold separations less than about 0.2 mm (the flow-independent case). C. A linearized small-signal or incremental model for the source-tract system during breathy voice.

pattern will tend to be more triangular, with the upper margins defining the PGA during the opening phase and the lower margins defining the PGA during the closing phase. At the apex, the upper and lower margins would be approximately equidistant. This phase difference is generally believed to be less with more tension in the vocal folds, and the PGA would then tend to be more smoothly varying, or sinusoidal, as in the dashed trace in Figure 37-1.

During the C or "closed" phase, some segments of the vocal folds have come into contact, and the variation of PGA is therefore more limited. However, the PGA waveform during the C periods will generally not be perceived as completely "flat" or constant in amplitude, since the segments of the vocal folds that are not in contact are still in motion. The existence of contact during this phase is easily verified by means of an electroglottograph. The fact that not all of the vocal folds are in contact can be verified from the offset from zero area of the measured PGA waveform, or the offset from zero flow of the glottal air flow waveform. All of these features are illustrated in the example presented in the following section.

In this paper we will be examining the relationship between the glottal air flow and glottal area functions, using the simple model shown in Figure 37-1, in which the subglottal and supraglottal systems are represented by only their low frequency inertance L_t. This condition is approximated when the lowest supraglottal and subglottal resonances are at least about three times the voice fundamental frequency. The glottal orifice is represented by its conductance $G_g = dU_g/dP_g$, and the lung pressure P_L is assumed to be constant.

Furthermore, it is assumed that the glottal conductance is determined by the PGA according to either the flow-dependent relationship $G_g \propto A_g^2/U_g$ derived from kinetic energy considerations and usually assumed to be valid at larger glottal areas, or the flow-independent relation $G_g \propto F(A_g)$ derived by considering the viscous losses that are assumed to predominate at small glottal areas (van den Berg, *et al*, 1957). According to van den Berg's model experiments, the function $F(A_g)$ is roughly proportional to A_g^3, with the viscous losses predominating at glottal areas less than about 10 or 15 % of the maximum area during the glottal vibratory cycle associated with normal, non-breathy voicing. When the folds are not parallel, i.e., when they are narrowly separated at some points along their length, and widely separated at other points, then both relationships may hold simultaneously, however this complication is usually ignored in a simple analysis.

In order to correctly predict the flow pattern that would result from a given area function in non-breathy voice (in which the glottal area and air flow waveforms have little or no offset from zero), it is necessary to consider the nonlinear relationship between area and flow represented by the differential equation for the model of Figure 37-1B. Due to the strong degree of nonlinear interaction that can occur at small glottal areas, no linearization of the system equations can be made without introducing a significant error. However, if in breathy voice the offset from zero area is large enough, we can advantageously use a linearized small signal analysis in which P_L and G_g are transformed into a Norton or flow-source equivalent network. The result is the model of Figure 37-1C, in which the fictitious flow function U_N appearing as a driving source represents the glottal flow U_g that would occur if $L_t=0$. The characterization of G_g usually assumed at larger glottal areas results in U_N being proportional to A_g. This is because when L_t is zero, the transglottal pressure P_g is equal to P_L and therefore constant. A constant P_g results in an essentially constant particle velocity v_g when the glottal area is large (Fant, 1960). If v_g is assumed constant over the glottal area, then U_g will be equal to v_gA_g, and therefore proportional to A_g. The bar over the dissipative element $\overline{G}_g$ indicates that it is an average value that is considered to be constant either at the value of dP_g/dU_g that occurs at the average glottal area during the vibratory cycle, or at the average value of dP_g/dU_g. These two values would converge as the offset from zero area became large compared to the variations in area over the glottal vibratory cycle.

In the form of Figure 37-1C, the system is easily identified as a first-order low pass system, with the waveform of A_g as input, and with U_g, the flow through L_t, as the output. The time constant of the system is given by the product $L_t\overline{G}_g$, and the -3 dB frequency is $1/(2\pi L_t \overline{G}_g)$.

To obtain order-of-magnitude estimates for the time constant and -3dB frequency of this source-tract low pass filter, we can estimate the average glottal conductance $\overline{G}_g$ as $\frac{1}{2}\overline{U}_g/\overline{P}_g$, where $\overline{U}_g$ and $\overline{P}_g$ are the average values during breathy voicing. The factor of ½ is introduced when the differential conductance is calculated under the flow-dependent conditions (Fant, 1960). Since $\overline{U}_g$ is roughly 1.0 liter/sec at normal voice levels, and $\underline{P}_g$ is closely equal to the average lung pressure, or about 6 cm H_2O during breathy voice, $\overline{G}_g$ would be about 1/12 liter/cm H_2O-sec. We have previously estimated (Rothenberg, 1981a) that for some vowels the supraglottal component of L_t can be as large as approximately .008 cm H_2O-sec^2/liter. The relative contribution of the subglottal system, which was neglected in that reference, can be inferred from the relative magnitudes of the excess pressures just below and above the glottis at the instant of vocal fold closure during normal voicing (Koike, 1981; Rothenberg, 1981b). Since the subglottal overpressure is about 50% of the supraglottal underpressure (very roughly), the estimate of L_t should be increased by a factor of 1.5, to .012. We may then estimate the source-tract time constant as:

$$\tau_{S\text{-}T} = \overline{G}_g L_t \simeq \frac{1}{12} \times .012 = .001 \text{ sec} \qquad (37\text{-}2)$$

The resulting -3dB frequency of the source-tract system is:

$$f_{-3dB,S\text{-}T} = \frac{1}{2\pi\, \tau_{S\text{-}T}} \simeq \frac{1}{2\pi\,(.001)} \simeq 150 \text{ Hz} \qquad (37\text{-}3)$$

For a first order low pass system, the asymptotic attenuation above the -3 db frequency is 20 dB per decade, or about 6 dB per octave. Thus we see that the low pass action of the source-tract filter can be expected to significantly attenuate all the harmonics of the fundamental frequency, and, in fact, result in some attenuation and phase shift of even the fundamental frequency component.

That the source-tract interaction during breathy voice is significant should not be surprising, since $\overline{G}_g$ is much larger than in normal voice, while L_t can be expected to change very little, and the interaction varies directly with the time constant $\overline{G}_g L_t$. However, we have just shown that the effect of this interaction on U_g during breathy voice is different than the effect it has on normal, non-breathy voice, in that it attenuates the higher harmonics for the breathy case, while it generally strengthens the higher harmonics in the non-breathy case; that is, the source-tract acoustic interaction tends to have an opposite effect on the high frequency spectrum in breathy voice, as compared to non-breathy voice. One consequence of this is that the interaction should tend to significantly increase the degree of spectral change at voiced-unvoiced transitions. We refer here to transitions in which the unvoiced speech segment is associated with a vocal fold abduction, and not with an adduction or glottal stop gesture.

Another effect to be expected during a voiced-unvoiced transition is a small decrease in the voice fundamental frequency (or increase in fundamental frequency during an unvoiced-voiced transition) as the phase shift at the fundamental frequency varies, since the frequency is the time derivative of the phase. The perturbation we refer to would be independent of, and would add to or subtract from, any changes in the timing of the glottal area oscillations. To obtain an estimate of the magnitude of the frequency perturbations, we can use the above rough estimate of 1 ms for the source-tract time constant during fairly breathy voicing, and note that this is the approximate delay at the fundamental frequency for fundamental frequencies not greatly removed from the source-tract -3dB frequency (150 Hz in the above calculation). During non-breathy voice, the phase shift between glottal area and glottal flow at the fundamental frequency is only about ½ msec for fundamental frequencies in this range (see, for example, the shift from the centerline of the peaks flows in the non-breathy simulation in Figure 37-6). If we assume that the transition from non-breathy to breathy voice occurs in about 30 ms, the average change in fundamental frequency during the transition would be (1-½)ms ÷ 30 ms, or about 1.5%, with a peak value of perhaps 2%.

Perturbations of the frequency of the fundamental component of this magnitude would be detectable instrumentally, though they may not be of great perceptual significance, coming as they do during a period of strong spectral change. In addition, pitch perception may often depend more on the periodicity of the glottal harmonics near the first and second formants, and therefore on the timing of the instants of glottal closure, than on the frequency of the fundamental component. However, the timing of the instants at which the vocal folds come in contact also tends to be perturbed by about one or two percent during a voiced-unvoiced transition, due to the fact that the contact occurs at a different point in the vocal fold cycle as the folds are abducted or adducted. As with the interactive effect, the perturbation would be a decrease in frequency during abduction and an increase during adduction.

THE SIMULTANEOUS RECORDING OF GLOTTAL FLOW, GLOTTAL AREA AND VOCAL FOLD CONTACT

To illustrate the different effect of source-tract interaction in non-breathy and breathy voice, we show in Figure 37-2 simultaneous recordings of inverse-filtered oral air flow, a photoglottograph signal and an electroglottograph signal. All three signals were recorded simultaneously on an F.M. tape recorder from a male adult subject pronouncing the Swedish sentence "Axel är här" ("Axel is here"), and written out on a Minograph ink jet chart recorder. The segments chosen for Figure 37-2 are the onset and offset of the /h/ in "här", including a few glottal cycles before and after the /h/. The horizontal arrow above the air flow trace indicates the approximate extent of the abductory movement for the /h/, as estimated from the three traces. The abductory movement encompasses about 18 glottal pulses.

The air flow was recorded from a mask system similar to one described previously (Rothenberg, 1977), with a response time of about 1/4 ms, and inverse-filtered using formant values that were an average of those for the vowel just before the /h/ in "här" and just after it. The orthographic "r" in "är" is not pronounced in this context, and both the preceding and succeeding vowels have a rather high F_1 (about 610 Hz and 710 Hz, respectively, for the sample recorded). Thus the estimate of glottal flow in the top trace of Figure 37-2 has a small remanent formant ripple during the closed-glottis periods of the nonbreathy voice associated with each vowel, and is probably set fairly accurately for the central segment of intervening /h/. The zero level for the flow trace was extrapolated from points of known zero flow before and after the sentence, and should be accurate within ± 2 mm, or 50 ml/s.

The delay introduced by the inverse filter reflects the sum of the glottis- to-mask transit delay and the delay due to any low pass filtering, and was measured to be 1.25 ms. The glottal air flow trace in the figure was shifted left by the equivalent of 1.25 ms to account for this delay, and should therefore be in correct time alignment with the photoglottograph and electroglottograph traces. The chart recorder response time was less than 1/4ms, and the time alignment of the three channels was adjusted to be better than 0.2 mm, which was equivalent to about 0.6 ms.

The photoglottogram in the middle trace was recorded by placing a strong light source on the neck near the cricoid cartilage and picking up the portion of the light visible through the glottis with a photodiode at the end of a catheter inserted through the nose and running down the posterior pharyngeal wall to a point about 1 or 2 cm above the glottis. The positions of the light source and sensor were adjusted until a good signal was obtained. Observation of the light coming through the glottis by means of a laryngeal mirror placed in the oropharynx showed that the glottis was not uniformly illuminated, and that a strong signal from the photodiode was consistently associated with a band of light across the glottis near its center. Thus a glottal "chink" between the arytenoid cartilages, for example, would not be detected in the photoglottogram. However, a uniform return of the air flow trace to near zero flow during normal voice for this speaker indicated that he had a fairly parallel action of the two vocal folds, and therefore that the medial separation measured by the photoglot-

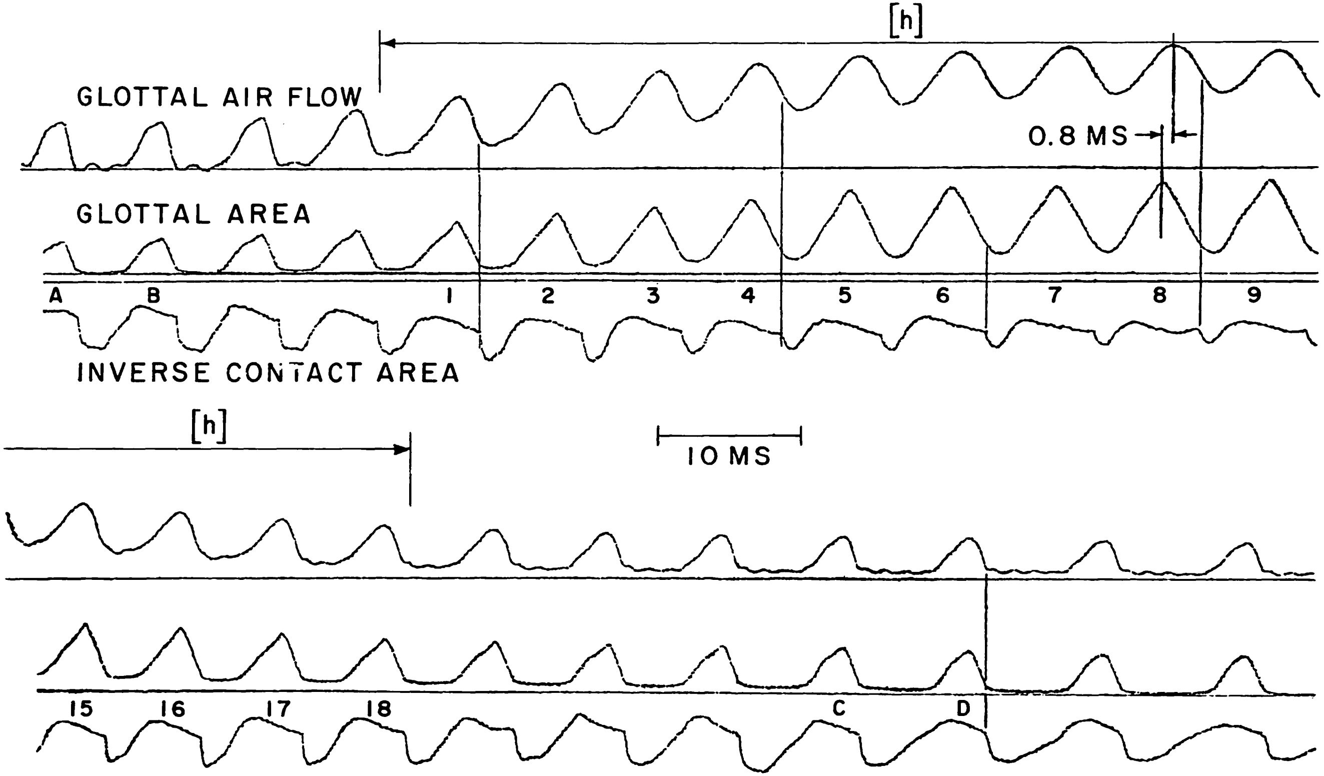


FIGURE 37-2
Glottal air flow (inverse filtered oral flow), projected glottal area (photoglottograph) and vocal fold contact area (electroglottograph) during the voiced to breathy-voiced transition at the word boundary between "är" and "här" of the Swedish sentence "Axel är här." ("Axel is here.").

togram would be closely proportional to the glottal area.

The response time of the photosensor was also checked at the light levels it was to be used at (since the response time of a photodiode varies with the light level) and found to be only one- or two-tenths of a msec. The step change in light required for the test was generated by a light-emitting-diode energized by a square wave generator. The upper zero line for the glottal area waveform was taken from the closed-glottis level of the adjacent vowels and is probably accurate. The lower zero line is taken from the initial glottal closure for the onset of the initial /a/ in "Axel", and probably has been affected more by drift than was the top zero line.

As one indication that the waveshapes of flow and area are mutually consistent, it can be noted that although the area trace tends to be triangular, there were some perturbations from a triangular shape near the apex in the non-breathy voice before the /h/. That these perturbations actually reflected changes in the area waveform, and were not just artifacts of the measurement system, is attested to by their presence in the inverse-filtered air flow trace, albeit in a smooth form due to the source-tract acoustic interaction.

The electroglottograph was a modified Laryngograph. The trace shown has the polarity of inverse vocal fold contact area, so that a positive deflection indicates a more open glottis. This polarity permits a more direct comparison with the area and flow waveforms. Though the electroglottograph trace appears to have a noise-generated cycle-to-cycle irregularity that is about 10 or 20 percent of the signal generated by vocal fold contact area variations, the timing of the rapid change in the waveform associated with the vocal fold closure is usually clear. As illustrated by the four long vertical lines, the instant of vocal fold closure comes closely just before the rapid drop in the IVFCA. This agrees with measurements from time-corrected inverse filtered air flow (Rothenberg 1981c) and, incidentally, verifies the accuracy of our method for correcting the timing of the air flow trace, which is the same as was used in that reference.

COMPUTATION OF α and L_t FROM MEASURED GLOTTAL AIR FLOW AND AREA

In this section we will compute the value of the interactive parameter α during non-breathy voice [1] from the glottal air flow and area waveforms in Figure 37-2, and show how the total vocal tract inertance L_t could be computed if P_L were also known, using various models for the glottal source. Since the time scales in the figure and the accuracy of the inverse filter settings were not optimum for this purpose, it is not expected that the results will exhibit the accuracy that can be attained in this type of computation. However, it is hoped that the resulting measurements will have some significance, and that the procedures will illustrate how such measurements can be used in the future to test models of glottal aerodynamics, and help evaluate the mechanisms behind voice quality differences.

Turning first to the non-breathy voice immediately before and after the /h/, we have labled the four non-breathy pulses closest to the abductory movement as A, B, C and D. It can be seen that the area waveform for these pulses is not symmetrical, even ignoring the perturbations at the apex. In each case, the rise time is about twice the duration of the fall time. A

Footnote:

[1] Since we will only compute α during non-breathy voice, for brevity we will often delete the modifier "for the non-breathy voice" when mentioning α, even though it varies from cycle-to-cycle with the degree of abduction.

"skew factor" σ characterizing this asymmetry in a manner convenient for our purposes is defined as:

$$\sigma = \frac{\text{rise time} - \text{fall time}}{\text{rise time} + \text{fall time}} \tag{37-4}$$

This factor can theoretically vary between +1 and -1, and is equal to about +1/3 for the non-breathy pulses marked A, B, C and D in Figure 37-2. (During pulses C and D, the glottal period has started to increase, but this has resulted only in a greater interpulse interval as compared to pulses A and B, with little change in the shape or amplitude of the pulses.)

It can be shown (Rothenberg, 1983) that if F_1 is much greater than F_0, as in our example, the interactive parameter α can be related to the waveshape of a non-breathy glottal air flow pulse by using the ratio of the magnitude of the maximum slope of the falling segment, signified S_F, to S_R, defined as the maximum slope of the increasing segment with the oscillations at the frequency of the first formant "averaged out". The initial half of the increasing segment tends to be of rather constant slope, except that the slope immediately following the onset tends to be reduced by factors such as the time-varying acoustic inertance of the glottis and the viscous losses in the glottis at small glottal areas. Therefore S_R could be measured at about 40% of peak flow amplitude, averaging over a short interval to eliminate the influence of oscillations at F_1 (assuming F_1 is much higher than F_0, as in this case). The slope of the falling segment is measured at its steepest point, which, from simple models for source-tract acoustic interaction during non-breathy voice, would be expected to occur at or just before the instant of glottal closure, and therefore at or just before the minimum level of flow attained during the glottal closing phase.

The relationships between the slope-ratio so defined and α are shown graphically in Figure 37-3. The solid line shows the relationship that would exist if the glottal conductance was proportional to A_g^2/U_g over its entire range (the flow-dependent case), and is derived from the solution to the resulting nonlinear differential equation (Fant, 1982). The dashed line is derived from a simple flow-independent assumption in which glottal conductance is assumed proportional to the area over its entire range.

Using the results of an analog simulation of a "combined" model (Rothenberg, 1983) in which the glottal conductance is flow-dependent at large areas and flow-independent at small areas, with $F_g \propto A_g$, it can be argued that (1) for a wide range of the interactive parameter α in the combined model, the increasing slope will be similar to that predicted by a flow-dependent model, and also similar to that predicted from a simple flow-dependent model if the interaction parameter in the flow-independent model is made equal to one-half α, (2) at small values of α the terminal slope of the falling flow segment is determined by the flow-dependent model, in which case the same factor of two for α is required to match the results to the slope predicted by the simple flow-independent assumption, and finally, (3) that at large values of α the final slope should be determined by the flow-independent model, with a crossover between the two cases apparently occurring at about $\alpha = 1$ for values of viscous (flow-independent) and kinetic (flow-dependent) loss similar to those measured by van den Berg and his associates (1957).

The heavy solid line in Figure 37-3 shows the result of these three conditions in the form of a plot of slope-ratio for the combined-source case as implemented in an analog simulation. Note that the slope-ratio for the combined source tends to break away from the plot for the strict-flow-dependent case when α is near unity. However, it should be emphasized that this plot is based on a number of coarse simplifications and assumptions that require experimental verification, especially the model used for the flow-independent glottal con-

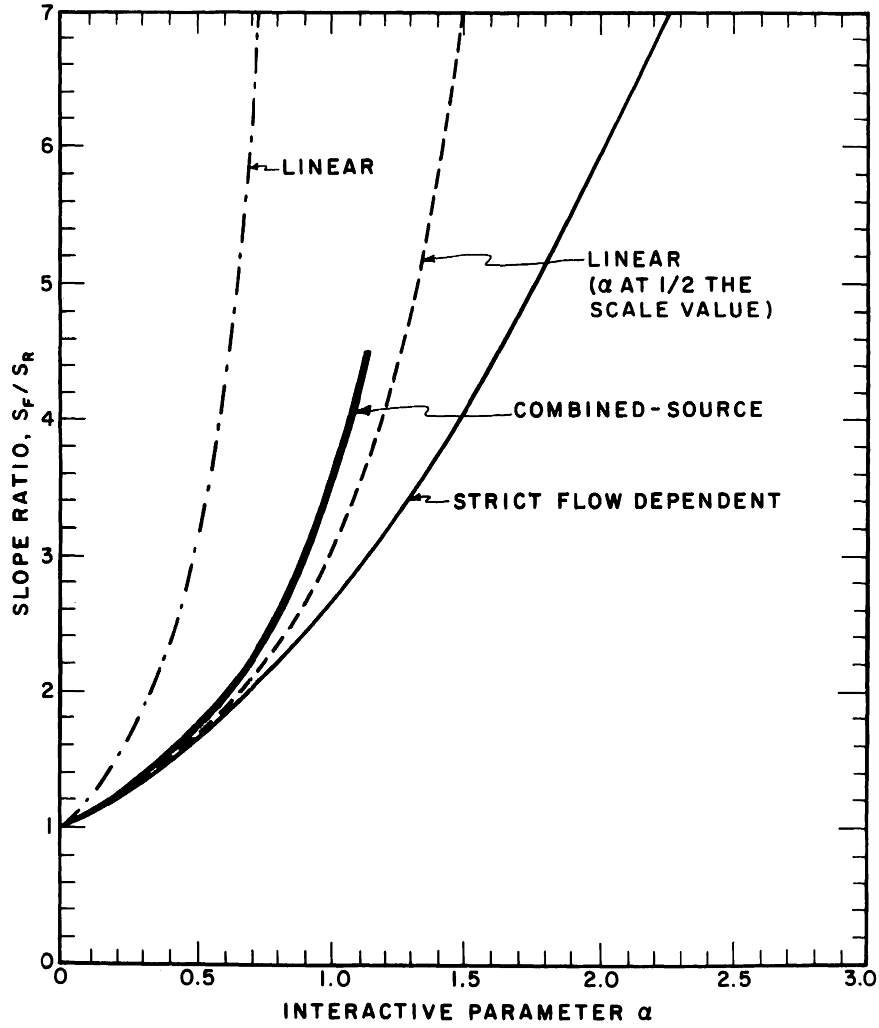


FIGURE 37-3
The slope-ratio S_F/S_R for a combined-source glottal model using estimates of viscous and kinetic glottal losses similar to those proposed by van den Berg, *et al* (1957). The slope ratios for a strict flow-dependent model and a linear flow-independent model are shown for comparison.

ductance at small areas, the precise nature of the glottal area function, and the glottal aerodynamics at small values of area during the closing phase. The plot is given here as a best available basis for predicting α from the slope-ratio.

The curves in Figure 37-3 are for a symmetrical triangular area function. (The combined-source plot was actually obtained using a half-wave rectified sinusoid for glottal area, but adjusted to approximate the results for a triangular function.) It can be shown (Rothenberg, 1981b and 1983) that for a non-symmetrical triangular area function in the simple flow-independent case, the slope-ratio can still be read from the curve in Figure 37-3 if the ordinate scale is interpreted as reading $\alpha + \sigma$ instead of α, where σ is the skew factor defined above. Thus for an area function with a value of σ of 1/3, S_F/S_R becomes infinite at $\alpha = 2/3$ instead of at unity, for the simple flow-independent case. Although no precise derivation can be presented at this time, it appears reasonable to assume that a skewing of the area waveform will have a roughly similar effect in the combined-source case, at least for small values of σ.[2]

With the above theoretical basis laid, we have only to measure the slope- ratio from Figure 37-2, use the heavy trace in Figure 37-3 to determine a value of α that would apply for a symmetrical pulse, and finally subtract $\sigma = 1/3$ to obtain a final estimate of α. Our graphical estimate of S_F/S_R is 3.5 for pulses A and B, 3.75 for pulse C, and 3.6 for pulse D. Taking into account (1) that the inverse filter was more accurately set for pulses C and D (there was less remanent F_1 oscillation during the closed-glottis period), and (2) that an inverse filter frequency setting that is too high (as in pulses A and B) will tend to lead to an underestimate[3] of the magnitude of the terminal slope S_F, we might accept an estimate for slope-ratio near the higher end of the range for these pulses, or about 3.7. This value of slope-ratio would be generated by a value of α of about 1.03 if the A_g pulse were symmetrical. Subtracting a factor of 0.33 to account for the tilt in A_g, we arrive at a final value for α of about 0.70.

This final value is entirely reasonable in terms of our previous estimates of a possible range for the interaction parameter. However, note that by having an asymmetric glottal area pulse that tilts to the right, the subject has, according to our model, increased his slope-ratio from about 2.15 (the "symmetric" ordinate for $\alpha = 0.70$) to a significantly higher value of 3.7.

We might also note here that to relate the interaction parameter α to the vocal tract inertance during the non-breathy and breathy voice, as we will do, the glottal pulses used should be produced with the same subglottal pressure, since α depends on unloaded glottal flow, and this varies with subglottal pressure. However, by using non-breathy pulses both before and after the /h/, and a linguistic context without a marked rise or fall in stress, we can perhaps assume that the "adjusted average" slope-ratio of 3.7 is generated by an average non-breathy subglottal pressure that is close to the value during the intervening breathy /h/.

Let us now consider the case of breathy-voice. In Figure 37-2 we have numbered as one-through-eight a sequence of glottal pulses of increasing breathiness. On an enlarged copy of Figure 37-2, we connected positive and negative peaks of the flow waveform, and the average value of flow at the center of each pulse. A slight adjustment of the negative peaks was made for the first few pulses, so as to obtain a value closer to the negative peak of the sinusoidal fundamental frequency component. After pulse three, the flow waveform itself is essentially sinusoidal.

The same procedure was followed for the area waveform, except that adjustment of the peak value to obtain the peak of the sinusoidal component was needed for all pulses. These

---

Footnotes:

[2] It is obvious that α cannot be adjusted by simply adding σ if the flow-dependent glottal source model prevails, since in this case the maximum falling slope must become infinite as σ approaches unity, but adding σ to α does not yield that result. The exact relation between σ, α and slope-ratio, and approximate relations for the combined-source glottal model, will be submitted for publication elsewhere (Rothenberg, 1983).

[3] Since the error component at F_1 has the phase of a cosine function when the inverse filter is set too high, it will obscure the final decay in air flow by making it appear delayed by 1/4 cycle at F_1.

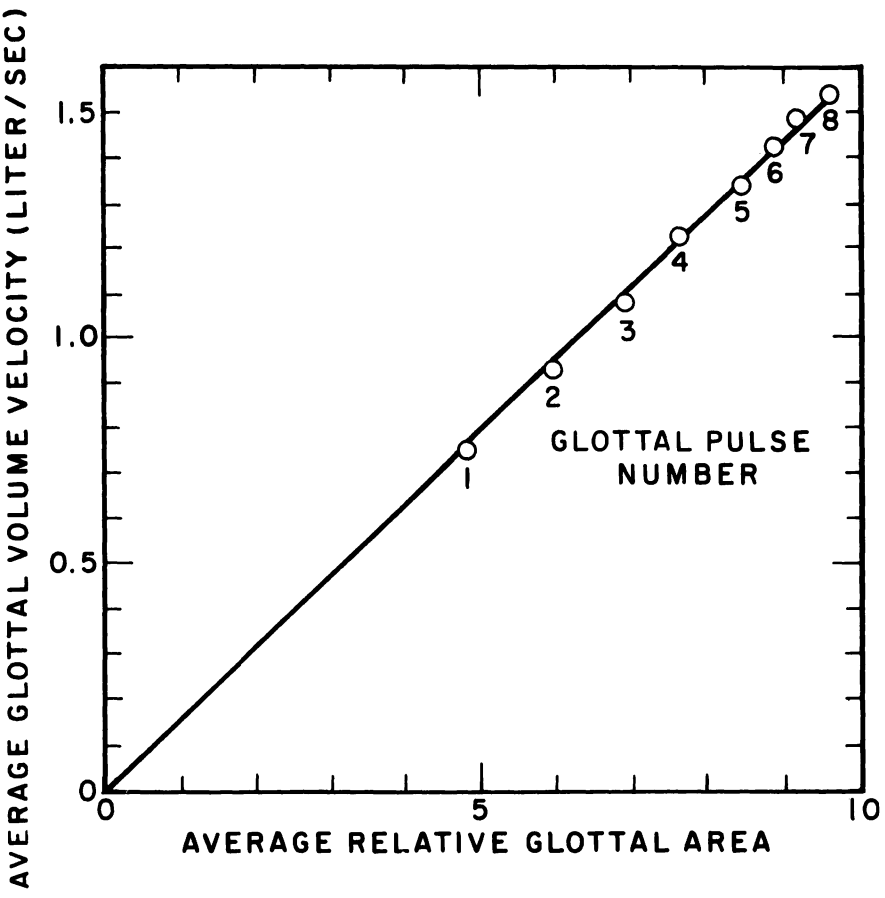


FIGURE 37-4
Average glottal air flow vs. average relative glottal area (in mm trace deflection) for eight consecutive glottal pulses in the transition from voiced to breathy-voiced shown in Figure 37-3.

judgements were made by the author by eye, on the basis of a long experience with Fourier analysis.

To test the consistency of the amplitude scales and zero settings, and to provide a calibration of the glottal area trace in terms of the resulting static glottal flow, the average area was plotted against average flow in Figure 37-4. The data fell close to a straight line going through the origin, indicating that volume velocity was closely proportional to the area as the area varied over a range of about 2:1. This would be the relation between flow and area predicted by the flow-dependent model for a constant average transglottal pressure, when the oscillations in flow and area around their mean values are small, and the time constant of the increase in average area is large compared to the time constant for source-tract interaction.

Checking to see if these conditions held, we find: (1) The 1/e time constant of the increase in average area, about 30 ms, is much larger than a typical source-tract time constant G_gL_t of about 1 or 2 ms. (See, for example, the computations in this paper for a typical value.) (2) From the linguistic context and the similarity of the glottal flow amplitudes before and after the /h/, we can assume that there was no great variation in subglottal pressure during this pulse sequence, though there may well have been a small variation of up to, say, five or ten percent that was not recorded in our procedure. (3) Since the variations in area and flow for the first three individual glottal pulses were certainly not "small" compared to the average values, we should expect some deviation from the relation $\overline{A}_g \propto \overline{U}_g$ for these pulses. For a more accurate estimate of the error caused by a non-negligible ratio of the "a-c to d-c" components, one would have to return to the underlying nonlinear differential equation.

Comparing the area and flow traces in Figure 37-2, we see that the flow appears to be a low-pass filtered version of the area, with the low-pass action increasing at larger glottal areas. This is precisely the relationship predicted above from the source-tract acoustic interaction. We will now proceed to show one way in which the underlying vocal tract inertance and the parameter α can be estimated from the phase lag or delay caused by L_t. This was done for the glottal pulse numbered 8.

In the small-signal Norton equivalent model for the glottal source shown in Figure 37-1C above, the "input" is a flow source representing the flow that would occur if the vocal tract impedance were zero. As generally assumed from the flow-dependent model, and verified for our example by the flow-area data of Figure 37-4, this source will have a waveshape similar to that of A_g for the larger values of area we are considering. The "output" is the glottal flow U_g, which is acted on by the low pass filter formed by $\overline{G}_g$ (the average small-signal glottal conductance G_g) and L_t. From linear system theory, we know that the time constant of this filter, $\overline{G}_g L_t$, can be computed from the phase shift θ at the voice fundamental frequency, F_0, from the relation

$$\tau_{S\text{-}T} = \overline{G}_g L_t = \frac{1}{2\pi F_0} \tan \Theta_{F_0} \qquad (37\text{-}5)$$

where θ_{F_0} is the phase difference between the F_0 sinusoidal components of U_g and A_g.

For pulse no. 8, the time lag $\Delta\tau$ between the F_0 components of U_g and A_g was measured to be 0.8 ms (after removing the 1.25 ms system delay in the air flow trace). This measurement was made graphically by bisecting a sinusoidal fit to each pulse (the short vertical lines in the figure). Because of the time scale in the figure, the accuracy of this measurement could only be considered to be about ±10%. The delay angle θ_{F_0} can then be computed as

$$\Theta_{F_0} = \Delta\tau \times 360\ F_0 = (0.8 \times 10^{-3})\ (360)\ (142) = 40.9° \qquad (37\text{-}6)$$

and from (5) above

$$\tau_{S\text{-}T} = \overline{G}_g L_t = \frac{1}{2\pi(142)} \text{Tan } 40.9° = 0.97 \times 10^{-3} \qquad (37\text{-}7)$$

The conductance $\overline{G}_g$ in this time constant is the small signal value, dU_g/dP_g, and for the flow-dependent case is approximately equal to ½ the quotient of $\overline{U}_g$ and P_L, with the factor of ½ being introduced in the process of differentiation (Fant, 1960). Reading $\overline{U}_g$ from Figure 37-4 as 1.53 liter/sec, we obtain in liter/sec, cm H_2O units:

$$L_t = \frac{\tau_{S\text{-}T}}{\overline{G}_g} = \frac{\tau_{S\text{-}T}}{\frac{1}{2}\overline{U}_g/P_L} = \frac{0.97 \times 10^{-3}}{\frac{1}{2}(1.53)/P_L} = 1.27 \times 10^{-3} \times P_L \qquad (37\text{-}8)$$

Using a typical value of P_L, say 7 cm H_2O, we see that L_t is approximately 9×10^{-3}. However, we will see that P_L is not needed in computing the interaction parameter α.

From the definition of α for non-breathy voicing (Equation 37-1),

$$\alpha = L_t \frac{\hat{G}_{gMAX}}{\tau_p/2} \tag{37-9}$$

where $\hat{G}_{gMAX}$ is the quotient U_{gMAX}/P_L that would occur if $L_t = 0$, at the peak area of a triangular glottal area pulse, and τ_p is the pulse duration, in this case about one-half of the glottal period, or 3.5 ms. U_{gMAX} was estimated to be 1.28 liter/sec by determining the peak of a non-symmetrical triangular pulse fit to a recorded non-breathy photoglottograph (relative area) pulse, and carrying over the peak relative area to flow, using Figure 37-4. The final computation yields

$$\alpha = L_t \frac{U_{gMAX}/P_L}{\tau_p/2} = 1.27\times10^{-3} P_L \frac{1.28/P_L}{1.75\times10^{-3}} = 0.93 \tag{37-10}$$

Since some of the above measurements could only be considered accurate to at best ±10% or ±15%, a result of $\alpha = 0.93$ can be considered a partial confirmation of the value obtained for non-breathy voice (0.70). However, more significantly, this example does illustrate that it is possible to obtain realistic estimates of L_t by measurements during breathy voice.

It could be noted here that this procedure could have been carried out from only the air flow trace and the vocal fold contact area, making a non-invasive measurement possible. The timing information required for the delay measurement could be obtained from the VFCA function, as long as some small period of vocal fold contact remained during the glottal cycle. When the period of vocal fold contact is only a small fraction of the glottal period, the most-closed peak of the VFCA waveform can be expected to align well with the minimum value of the PGA waveform. If P_L were available, then L_t could be computed fairly directly from the air flow and timing data. To obtain α, P_L is not required, however τ_p and a peak unloaded glottal flow U_{gMAX} must be estimated from the non-breathy voice segment for which α is to be relevant.

Finally, we note that L_t could also be obtained from the degree to which the ac-to-dc ratio in the flow trace is smaller than the ac-to-dc ratio in the area trace. However, since the dc or average flow is not determinable from the small signal model, we would have to go back to the nonlinear differential equation for the flow-dependent case to determine the correct relationship between these ratios and L_t.

This is illustrated in Figure 37-5, in which the input to an analog model for the glottal source and vocal tract, using the "combined-source" case of Figure 37-3, was given the form and ac-to-dc ratio (relative offset from zero) of A_g of pulse 8 in Figure 37-2, and L_t then adjusted to give the same ratio of ac-to-dc component for U_g as found in Figure 37-2, pulse 8. The value of L_t required to make this adjustment was similar to that computed above for Figure 37-2. The strict flow-dependent model gave a very similar result, since the glottal area never approaches the small values at which the combined-source and strict flow-dependent models differ most, however the value of L_t required was about 30% larger with strict flow dependence. The phase difference between the fundamental frequency components of A_g and U_g was also similar to that measured in Figure 37-2, for pulse 8, ~45° from the model vs. 40.9° from Figure 37-2. Though the difference could have been due to measurement errors, a difference of this magnitude, if substantiated in other measurements of the same nature, would indicate that there may be some significant inaccuracy in the models of laryngeal aerodynamics proposed so far. Alternatively, the discrepency could be due to an inaccuracy in the simple model we use for the vocal tract. For example, we have neglected the resistive and compliant elements of the vocal tract. Inclusion of the subglottal flow resistance in the

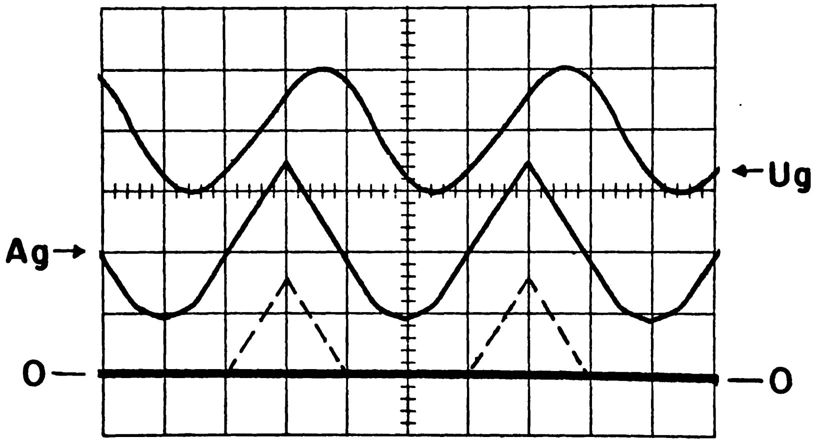


FIGURE 37-5
The solid lines are the results of an analog simulation of the flow and area traces in Figure 37-2 at the breathy glottal pulse number 8, using a combined-source model for the glottal conductance. The dashed A_g trace is the waveform used for the non-breathy simulation of Figure 37-6.

model might have reduced the phase shift caused by the vocal tract inertance; however, the proper representation of this resistance would require a consideration of the distribution of inertance and resistance in the trachea, bronchi and bronchioles, and an estimate of the degree to which the resistance is flow dependent at the flow levels involved.

The value of L_t required to give the relationship between area and flow shown in the simulated breathy-voice of Figure 37-5 can be determined by changing the simulated A_g signal to a symmetrical triangular pulse with no offset from zero flow, to simulate non-breathy voice, and measuring the slope-ratio of the resulting simulated glottal flow pulse. This has been done in Figure 37-6, for two glottal source models. In each case, the amplitude of the triangular A_g pulse was kept in roughly the same proportion to the breathy A_g as was the case for actual the non-breathy and breathy A_g waveforms in Figure 37-2. The resulting slope-ratios, as measured from the waveforms in Figure 37-6, are as follows.

Top trace: Strict flow-dependent source, S.R. = 2.4

Lower trace: Combined source, S.R. = 3.2

The increasing slope for the combined-source trace was measured using the protocol indicated in the definition of slope-ratio.

In Figure 37-3, a slope-ratio of 2.4 for the strict flow-dependent case and 3.2 for the combined-source case would both be generated by a normalized inertance of about 0.95. Since we have used a symmetrical A_g pulse of about the same duty cycle and relative amplitude as in the non-breathy voice of Figure 37-2, this simulation value of 0.95 should be comparable to the measured values of α, namely 0.93 from the breathy voice and 0.70 from the non-breathy voice.

Again we have obtained a result of the correct order-of-magnitude, but differing from the non-breathy voice measurement by a margin somewhat larger than that which can be safely attributable to measurement errors. If replicated in future similar measurements, this result would indicate some inaccuracy in the aerodynamic models used. For example, assuming for the non-breathy voice in Figure 37-2 a strict flow-dependent source, or a source with flow-

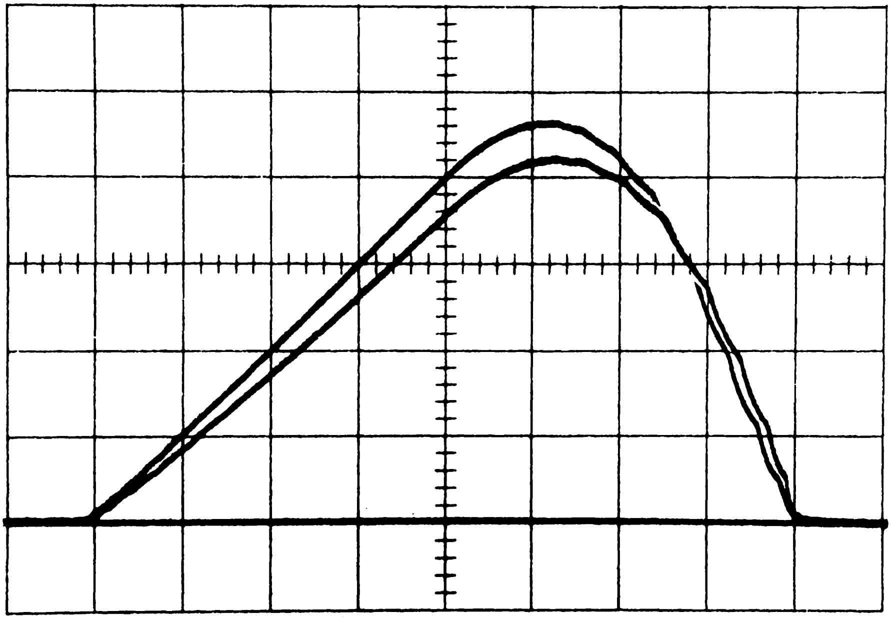


FIGURE 37-6
The smaller and more skewed trace is the simulated glottal air flow that results when using the same model parameters as in Figure 37-5, except that the average value of the simulated A_g waveform was reduced so as to form a symmetrical (skew factor=0) triangular pulse with a 50% duty cycle (the dashed trace in Figure 37-5). The relationship between the A_g pulses used for the glottal flow traces of Figures 37-5 and 37-6 was then approximately the same as between the breathy (pulse 8) and non-breathy pulses in Figure 37-2, except that the non-breathy pulses in Figure 37-2 had a skew factor of about 1/3. For the larger and more symmetrical trace, the conditions, including L_t, were the same, except that a strict flow-dependent source model was used. Scalloping of the decreasing flow was caused by the discrete implementation used for the time-varying glottal resistance.

dependence to smaller values of A_g than assumed in our combined-source model, would have yielded a measured interaction parameter greater than 0.70 and closer to the simulation value of 0.95. The adjustment made for an unsymmetrical area pulse could also be inaccurate.

The simulation results can be summarized by saying that they confirm the general validity of the models we have been using for glottal aerodynamics, and indicate that a more accurate simulation study of adjacent breathy and non-breathy voiced segments could be useful in deciding between the alternative aerodynamic models presently proposed. However, to be maximally useful, an analysis-by-synthesis study such as we have outlined above should be more accurately implemented, possibly using a digital formulation of the appropriate differential equations, with the actual A_g waveforms as "input" to the simulation, instead of the idealized shapes we have used here. The criteria we have used for goodness-of-fit, namely ac-to-dc ratio, phase shift at the fundamental frequency, and slope-ratio, are proposed here as convenient and meaningful measures, but by no means the only ones possible.

ACKNOWLEDGEMENT

The experiment from which Figure 37-2 was extracted was performed in cooperation with Drs. Gunnar Fant, Jan Gauffin, Peter Kitzing and Anders Löfqvist, while the author was a guest researcher at the Speech Transmission Laboratory of the Royal Institute of Technology in Stockholm, Sweden. The data from that experiment was largely responsible for directing my attention to many of the theoretical issues treated in this paper. During the work reported here, the author was supported by a research grant from the National Institutes of Health.

REFERENCES

van den Berg, Jw., Zantema, J.T. and Doornenbal, P. (1957). "On the air resistance and the Bernoulli effect of the human larynx," *J. Acoust. Soc. Am.* 29, 262-631.

Fant, G. (1960). *Acoustic Theory of Speech Production.* Mouton, S-Gravenhage.

Fant, G. (1980). "Voice source dynamics",*Quart. Prog. and Stat. Rep.*, 2-3/1980, Speech Trans. Lab., Royal Institute of Tech., Stockholm, 17-37.

Fant, G. (1983). "Preliminaries to analysis of the human voice source," *Speech Commun. Group Working Papers #3*, Res. Lab. of Electronics, MIT.

Koike, Y. (1981). "Sub- and supraglottal pressure variation during phonation," in Stevens, K. N. and Hirano, M. (Eds.) *Vocal Fold Physiology.* Tokyo: Univ. of Tokyo Press, 181-189.

Rothenberg, M. (1973). "A new inverse-filtering technique for deriving the glottal airflow waveform during voicing," *J. Acoust. Soc. Am.* 53/1, 1632-1654.

Rothenberg, M. (1981a). "Acoustic interaction between the glottal source and the vocal tract," in Stevens, K. N. and Hirano, M. (Eds.) *Vocal Fold Physiology.* Tokyo: Univ. of Tokyo Press, 305-328.

Rothenberg, M. (1981b). "The voice source in singing," in *Research Aspects of Singing*, Pub. #33 of the Royal Swedish Acad. of Music, Stockholm, 15-31.

Rothenberg, M. (1981c). "Some relations between glottal air flow and vocal fold contact area," Proceedings of the Conference on the Assessment of Vocal Pathology, Nat'l Institutes of Health, Bethesda, MD., *ASHA Reports* #11, 88-96.

Rothenberg, M. (1983). "The effect of flow-dependence on source-tract acoustic interaction," *J. Acoust. Soc. Am.* Suppl. 1, 73 (abstract). Complete manuscript to be submitted for publication.

DISCUSSION

J. BEAUCHAMP: I wasn't quite clear on what the controller on that resistor looked like.

M. ROTHENBERG: The analog electronic circuit used for simulating the combined-source case in Figures 37-3, 37-5 and 37-6 will be described in more detail elsewhere (Rothenberg, 1983). For all except the smallest areas and flows, a variable resistance in the electronic circuit made the small-signal differential conductance (the reciprocal of the resistance) proportional to the area squared divided by the volume flow. Since the conductance or resistance then depends on the volume flow rate, it is called "flow-dependent". This is the type of equation you get when the energy loss causing the resistance is dominated by kinetic energy losses. The flow-dependent losses are thought to dominate at the larger glottal areas and resulting higher volume flow rates. But at small areas and flows, the glottal conductance should, according to the theory resulting from van den Berg's early model experiments, be approximately proportional to A_g^3 and not depend on flow rate. However, to simplify the simulation I just made conductance proportional to A_g and not A_g^3 at small glottal areas. That was at areas below the equivalent of about 10 or 20% of the maximum glottal area in normal, non-breathy voicing. For the breathy voicing A_g^2/U_g dominated throughout most or all of the waveform, since the glottal areas and flow rates are greater than in normal voicing. Therefore including the flow-independent term at small areas only affected the non-breathy voicing. But if you took out the flow-independent resistance entirely and made the conductance proportional to A_g^2/U_g for all areas and flows, the non-breathy waveforms changed, and what may be significant in modifying van den Berg's theory, the waveforms changed in such a way that you obtained a little better fit to the measured values of α and L_t. But as I said in the paper, I don't know if that improvement was accidental or not.

38 ▪ EFFECTS OF THE SOURCE-TRACT INTERACTION USING VOCAL FOLD MODELS

Bernard Guérin

INTRODUCTION

We have studied two particular effects in source-tract coupling: the interaction on the fundamental frequency of the vocal cords oscillation and the influence on the flow waveform. These effects have been systematically taken into consideration in the two mass model of the vocal source and in the continuous two beam model of the vocal source model.

This kind of research is not new. In fact since Cuvier's (1907) observations, the acoustic influence of the vocal tract on the larynx operation phenomenon has been studied over and over again.

The work carried out by Weiss (1932), Kagen and Trendelenburg (1937) using additional pipes, show the interest of outlining the modifications of vocal cord vibrating mode (particularly F_O) when the supraglottal load varies. It seems that it was van den Berg (1955) who just summed up the problem in terms corresponding to the modern theory of speech production: a source-vocal tract interaction can take place if the vocal tract input impedance is not negligible compared to the glottal one. Following up on this work, Flanagan (1965) evaluates these two impedances and concludes: "...changes in the tract configuration generally do not greatly influence the operation of the vocal cords..." but he does not discard the possibility of such interaction.

The acoustic coupling phenomenon is not easy to isolate in natural speech and the use of additional pipes is limited to sustained vowel realizations only. Nevertheless, the effects observed when phonating in a gas environment other than air are a proof that seems hard to neglect. As a matter of fact, this shows that all the formant frequencies are shifted, as well as F_O. Beil (1962) carried out recordings of sustained vowels /i ε a y oe u o ɔ a e/ spoken by 5 speakers, first in air and then in a helium enriched gas environment and he noticed that F_O rises. The explanations that disregard acoustic coupling are not too convincing. Now, in this experiment, all observations would lead us to believe that the articulation configurations were the same, whether the subject was breathing air or not. In deep diving experiments using high pressurized gas mixtures have recorded increases in F_O as a function of time and depth that go up to 10%, (Picard, 1970). The data coming from voice measurements during singing, (Sundberg, 1979) also seem to corroborate this phenomenon. Finally, concerning the consequences of the coupling effects on F_O, it is well known that high vowels have a higher intrinsic pitch than low vowels. Generally, this phenomenon is explained by the mechanical interdependence between supraglottal configuration and larynx position. Larynx movements change the mechanical characteristics of the vocal cords and consequently the oscillation frequency.

But two other types of source-tract coupling can operate:

1. The vocal tract constriction introduces an air flow resistance which tends to raise the intra-oral pressure, thus reducing the pressure difference across the glottis. This effect, for high vowels, could account for a slight decrease in the fundamental voice frequency. But this result is contrary to the intrinsic pitch values.

2. The input impedance of the supraglottal cavities represents an acoustic load for the voice source.

Otherwise, two kinds of experimental results should be reported which could involve the importance of the acoustic load on the voice source frequency.

A first study on the characteristics of American nasalized vowels shows a gap between the fundamental frequencies of non nasalized and nasalized low vowels. Another study on French nasal vowels also shows a gap between the fundamental frequencies of nasal vowels and oral corresponding ones, (Carré, 1974).

In all cases, for oral vowels, nasalized vowels and French nasal vowels, the acoustic load on the vocal source changes with the vocal tract configuration and nasal coupling.

The studies which aim to know better the waveform of volume velocity are not recent: Lindquist (1965, 1970), Holmes (1962), Rothenberg (1973, 1977, 1981), Miller (1959), Mathews *et al* (1961), Miller *et al* (1963), Takazuki (1971), Colton *et al* (1976), Fischer *et al* (1975), Monsen *et al* (1977), Baer (1975), Farnsworth (1940), Sondhi (1975), Timcke (1958). Both the relative inacessibility as well as the extremely rapid and complex movements of the natural glottal wave make investigations of its shape very difficult. Therefore, a wide variety of quite different methods has been used in the past to investigate glottal vibration. To study sub- and supraglottal coupling, we propose to use two vocal fold models. Ten French vowels will be considered and the results will be compared with those obtained on natural speech by different methods mentioned above.

VOCAL TRACT-TWO MASS MODEL COUPLING EFFECTS

Model description

In order to study and incorporate aspects of coupling between the glottal source and the supraglottal cavities, a knowledge of the vocal tract driving point impedance is necessary. It is difficult to perform direct measurement of this impedance. Consequently, we have calculated it using a computer simulated lossy vocal tract model (Degryse, 1982). Special care was taken in introducing losses in this model; such as losses by heat transfer, by viscosity, as well as those due to wall vibrations and radiation impedance. It is relevant to note the obtained range of variation of the input impedance as shown in Table 38-1. In fact, the vocal tract input impedance is important for frequencies near the first formant, since its absolute magnitude is maximum at the formant frequency and can be equal to and even greater than the equivalent A.C. glottal impedance. Consequently, we only consider the first two formant characteristics for the model of the vocal tract input impedance.

The obtained results enable us to synthesize an equivalent circuit of the vocal tract driving point impedance (Mrayati *et al*, 1976). This circuit, where all the elements are controlled, is used as a load for a glottal source model. Figure 38-1 shows an oversimplified representation of the acoustic and mechanical parts of the model. We have also taken into account the subglottal cavity. A model of subglottal cavity is derived from Ishizaka's measurement (Al Ansari, 1981). Ishizaka's (1972) measurements of the input subglottal impedance of laryngectomized human subjects seemed to be an accurate one in comparison with van den Berg's measurements and Flanagan's estimations. These results show that this input acoustic impedance has three conspicuous resonance peaks. The resonance frequencies determined by Ishizaka's digital simulations, employing Weibel's anatomical model are adopted in the present investigation as a basis for our acoustical model of the subglottal impedance. Figure

38-2 shows the complete model including subglottal and supraglottal equivalent input impedances.

The vocal cord model is, in this section, the two-mass model (Ishizaka *et al*, 1972). Previous studies show that the model duplicates the principal features of human vocal cord behavior (Guérin, 1978; Guérin *et al*, 1980). The model may be controlled by two physiological parameters: the subglottal pressure Ps and the mass-tension factor Q (Guérin, 1978).

COUPLING EFFECT ON THE FUNDAMENTAL FREQUENCY

Theoretical investigation (Guérin *et al*, 1980)

Confining ourselves to the framework of a simplfied small signal two-mass model of the vocal source coupled with the vocal tract, we are going to study the influence of the input impedance of the vocal tract on the frequency of the vocal cord oscillation. To achieve this we have continued the work of Ishizaka and Matsudaira (1968). By considering the mechanical and aerodynamical system of equations we can study the variation of the oscillation

TABLE 38-1
Values of the elements of the input impedance equivalent circuit of the vocal tract.

vowel	Fl et F2 (Hz)	Rl et R2 (Ohm)	Ll et L2 (H)	Cl et C2 (F)
[u]	297 790	44.10^{5} 124.10^{5}	$6,43.10^{2}$ $9,17.10^{1}$	$4,465.10^{-10}$ $4,426.10^{-10}$
[o]	406 806	111.10^{5} 28.10^{5}	$5,36.10^{2}$ $2,57.10^{1}$	$2,868 .10^{-10}$ $1,516 .10^{-10}$
[c]	547 1019	221.10^{5} 43.10^{5}	$5,52.10^{2}$ $3,56.10^{1}$	$1,530 .10^{-10}$ $6,850 .10^{-10}$
[a]	686 1209	305.10^{5} 275.10^{5}	$4,64.10^{2}$ $1,65.10^{2}$	$1,156 .10^{-10}$ $1052 .10^{-10}$
[ɛ]	463 1905	88.10^{5} 97.10^{5}	$2,61.10^{2}$ $3,53.10^{1}$	$4,520 .10^{-10}$ $1,977 .10^{-10}$
[e]	423 2140	81.10^{5} 86.10^{5}	$3,17.10^{2}$ $3,08.10^{1}$	$4,465 .10^{-10}$ $1,796 .10^{-10}$
[i]	246 2285	34.10^{5} 317.10^{5}	$7,33.10^{2}$ $2,99.10^{1}$	$5,700 .10^{-10}$ $1,620 .10^{-10}$
[y]	282 1870	45.10^{5} 50.10^{5}	$6,84.10^{2}$ $1,64.10^{1}$	$4,654 .10^{-10}$ $4,420 .10^{-10}$
[ϕ]	388 1611	87.10^{5} 12.10^{5}	$5,06.10^{1}$ $7,28.10^{1}$	$3,320 .10^{-10}$ $1,340 .10^{-9}$
[œ]	541 1428	145.10^{5} 101.10^{5}	$3,35.10^{2}$ $5,05.10^{1}$	$2,580 .10^{-10}$ $2,460 .10^{-10}$

frequency with the nature of the input impedance of the vocal tract. Three cases are considered:

1. The load is of a capacitive type: it tends to increase the frequency of oscillation of the vocal source.
2. The load is of an inductive type: in this case, it tends to decrease the frequency of oscillation.
3. The load is resistive: as long as the resistance is below a certain value, F_0 decreases with the increase of the resistance value. Above this value, F_0 increases with the increase of the resistance.

These results are shown in Figure 39-3 where we consider the value of the first formant frequency in comparison with that of F_0. The above results were confirmed by an experiment carried out by Ishizaka *et al* (1972) in which they varied the length of an additional tube connected to the vocal tract and measured the corresponding value of F_0 (the first formant frequency is inversely proportional to the length of the tube).

When we consider the effects of the coupling with the subglottal cavities, the simulation show only some slight effects.

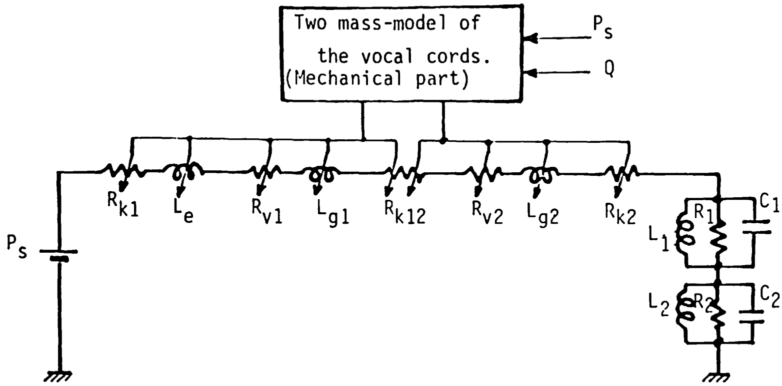


FIGURE 38-1
Two mass-model of the vocal cords loaded by an equivalent circuit of the vocal tract input impedance.

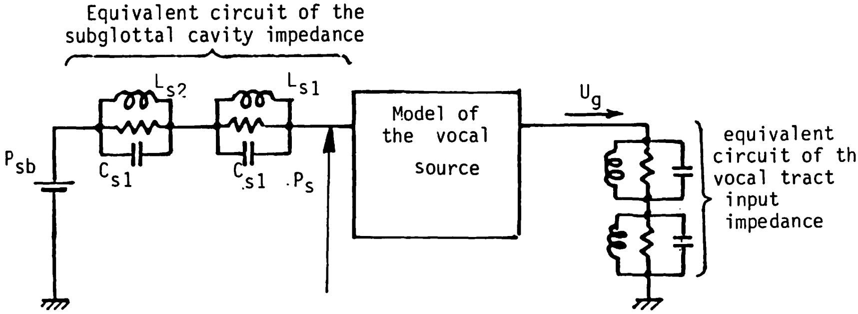


FIGURE 38-2
Complete model of the vocal source loaded by the subglottal and supraglottal cavities.

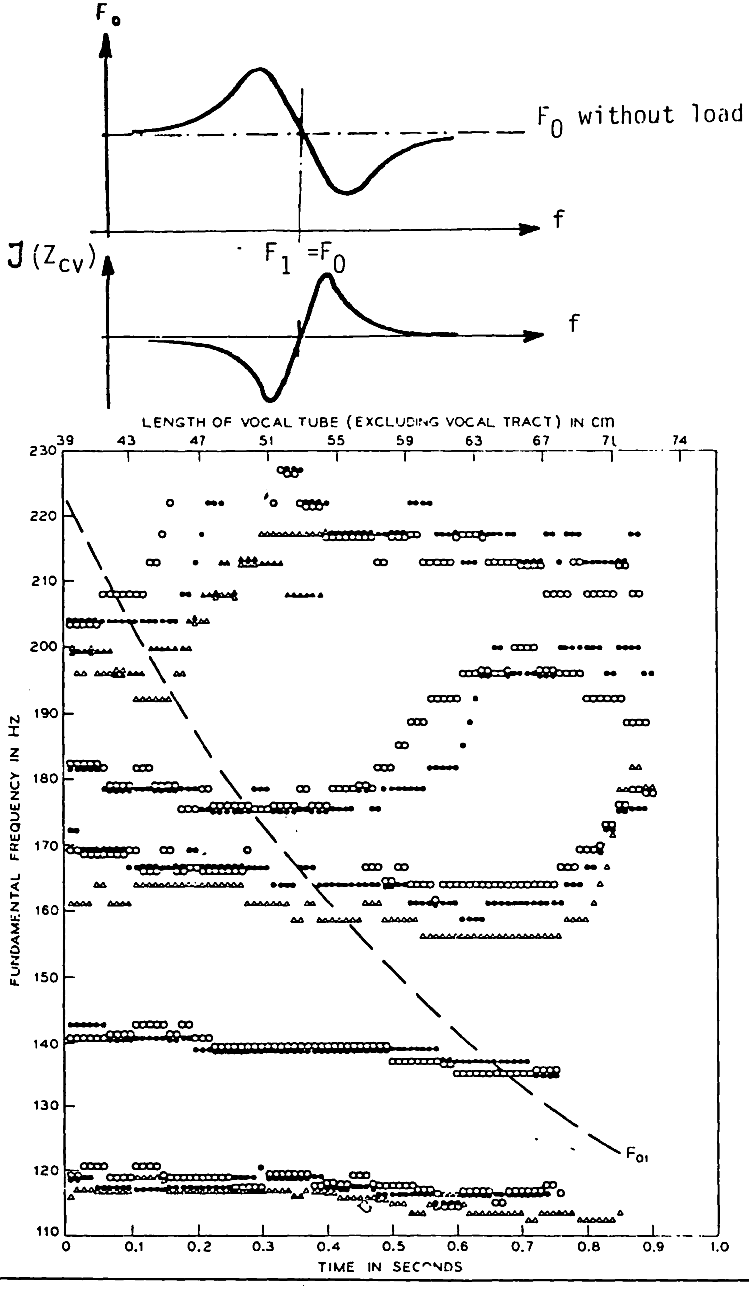


FIGURE 38-3

a)Theoretical fundamental frequency variations versus first formant frequency. b)Fundamental frequency measurements made for a human subject when the acoustic load on his vocal cords is varied. The broken line shows the first resonant frequency of the uniform tube. (From Ishizaka *et al* 1972).

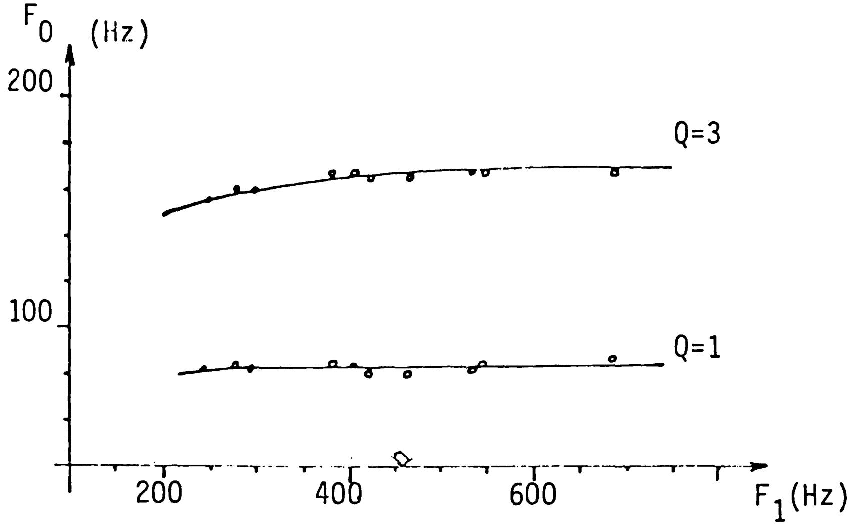


FIGURE 38-4
Frequency oscillation of the model of vocal source versus first formant frequency (simulation).

The simulation (Guérin *et al*, 1980; Al Ansari *et al*, 1981)

Concerning the ten French vowels, if we know the first formant frequency, we can trace the value of F_0 as a function of F_1. The results are shown in Figure 39-4. It is clear that F_0 increases slightly with F_1 which is contrary to the measurements of intrinsic fundamental frequency made on natural speech (but it agrees with the above theoretical investigation).

We therefore can suppose that the intrinsic value of F_0 is not determined by the acoustic coupling (see also Ewan, 1979) but by other more important effects which largely compensate for the effects of the acoustic coupling.

In other respects, the fundamental frequency F_0 of the vocal source output signal increases by a small amount when the subglottal impedance is implicated in the source circuit. This increase depends only on the lung pressure Psb. We expect that the slight difference is essentially due to the subglottal pressure variation Ps which reaches the maximum value at the glottis closing moment (Figure 38-5). It will be noted that the frequency of the subglottal pressure wave Ps is exactly the same as that of the larynx output wave, (Kioke and Hirano, 1973, 1981).

COUPLING EFFECTS ON THE GLOTTAL WAVEFORM

The shape of a typical glottal volume velocity is described in Figure 38-6. We define on the one hand, the open quotient O.Q. as the ratio of the open time and the wave period, and on the other hand the dissymmetry quotient as the ratio of the rise time over the fall time of the glottal wave.

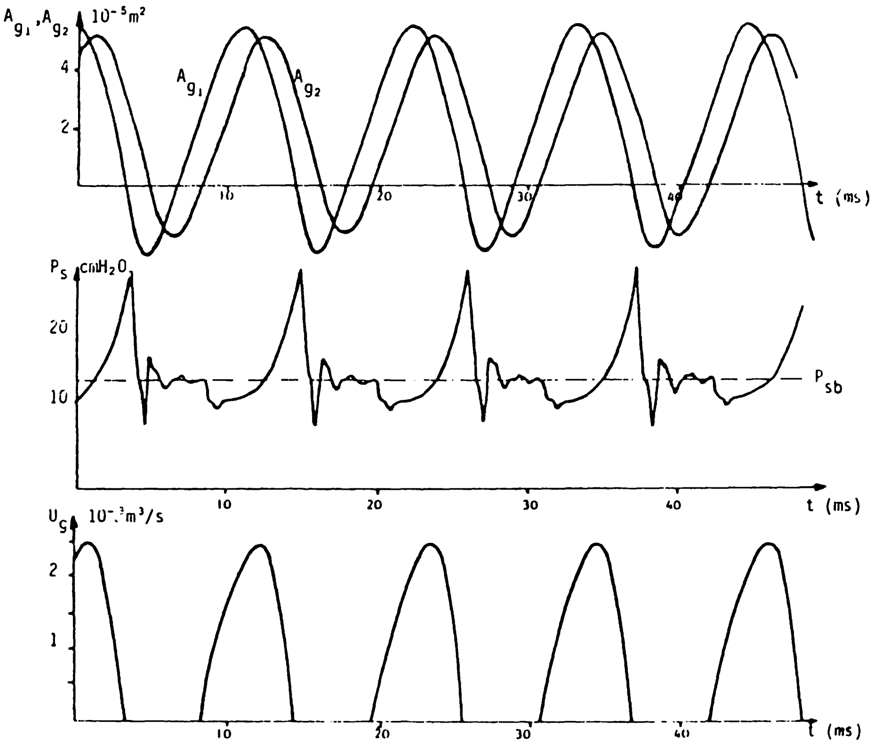


FIGURE 38-5
Variations of A_{g1}, A_{g2}, P_s, and U_g versus time.

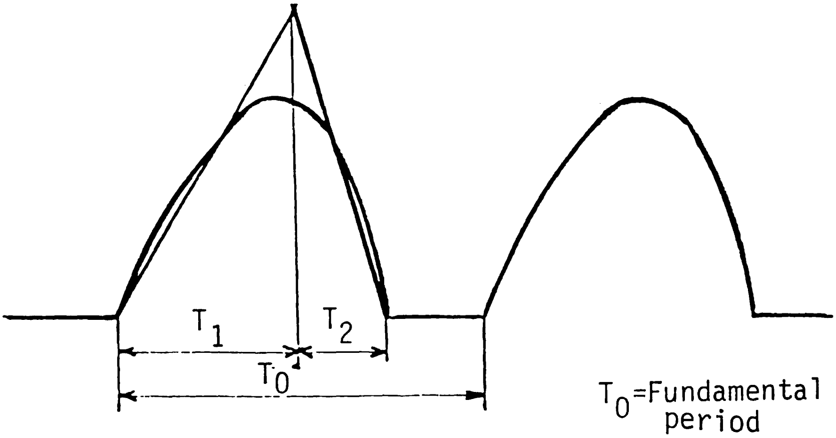


FIGURE 38-6
Triangular approximation of the glottal volume velocity. The open quotient is determined as O.Q.=$(T_1+T_2)/T_0$ and the dissymmetry quotient as D.Q. = T_2/T_1.

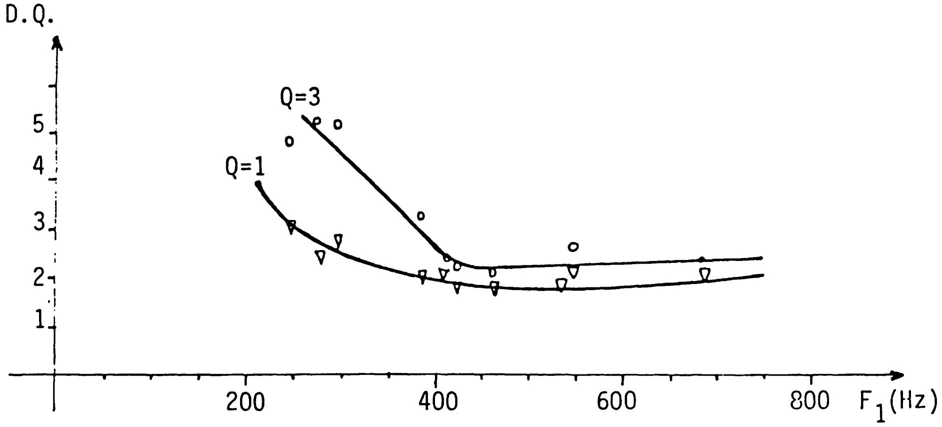


FIGURE 38-7
Dissymmetry quotient value versus first formant frequency of the vowels when the vocal source is loaded by the equivalent input impedance of the vocal tract.

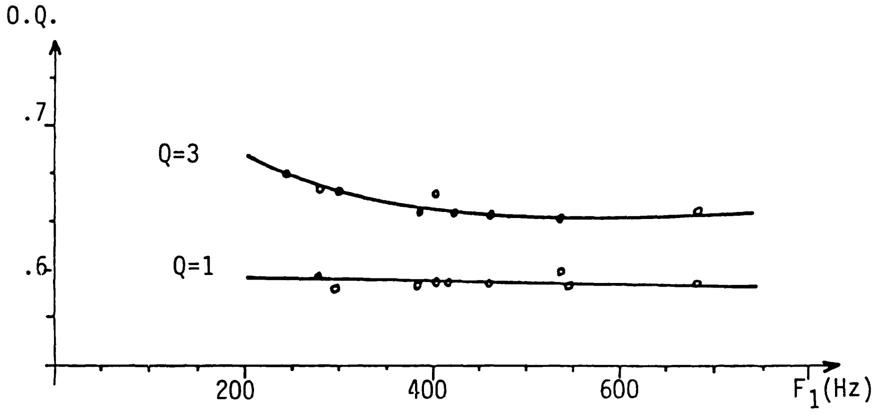


FIGURE 38-8
Open quotient versus first formant frequency.

Theoretical study. Rothenberg (1981)

In order to study the effects of the vocal tract input impedance coupled with the vocal source, we will simplify the problem by assuming that the glottal admittance Y_g is a pure conductance having a triangular evolution form. If the input impedance of the vocal tract is reactive, the results are shown in Figures 2 and 3 of Rothenberg (1981). In brief, we could say that the effect of the acoustic load is to increase the dissymetry coefficient D.Q.

Results

All the results obtained are shown in Figures 38-7 and 38-8. We could formulate the following remarks:

1. Vocal source-vocal tract coupling: The form of the flow wave could be strongly perturbed. We sometimes observe an additional overshooting at a frequency close to the first formant. This consequently increases the dissymmetry quotient D.Q., Figure 38-7 shows that the dissymmetry quotient increases when F_1 decreases. This is in perfect accordance with the theoretical results of the study carried out by Rothenberg (1980): the inductive part of input impedance increases when F_1 decreases. The opening quotient is less affected by this coupling (the incomplete glottal closure is not simulated). The mass-tension parameter Q must be equal to 3 in order to observe (Figure 38-8) an increase of O.Q. at low first formant frequencies. For all the vowels, we observe that the evolution of D.Q. with Ps and Q is observed to be of the same form that this obtained without the source coupling. An example is given in Figure 38-9.
2. Source-subglottal cavities coupling: The effects of this coupling are not visible except for the D.Q. which increases irrespective of the vowel studied. An example is given in Figure 38-10.

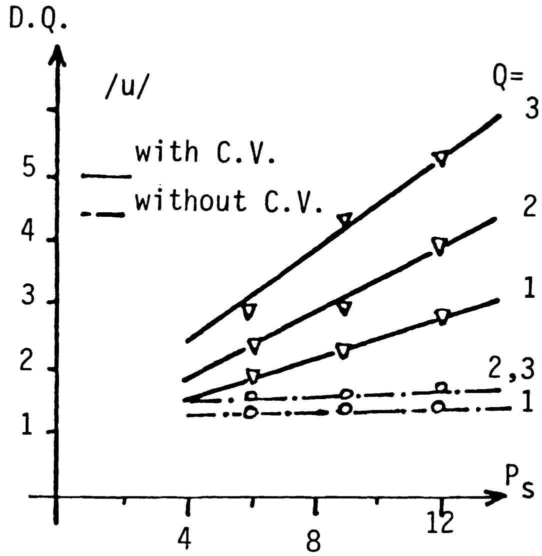


FIGURE 38-9
Variation of D.Q. versus P_s for the vowel /u/.

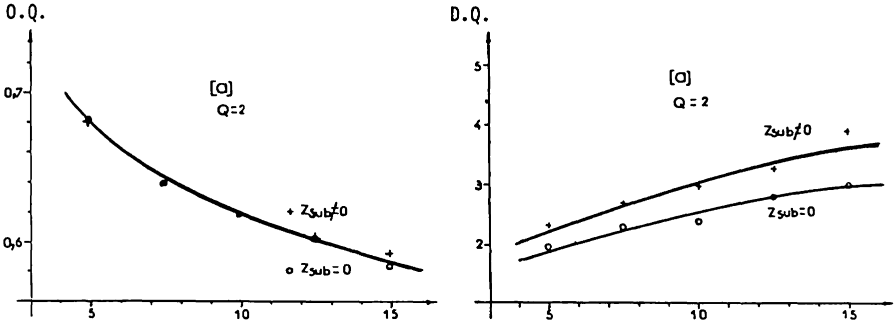


FIGURE 38-10
Effects of source-subglottal cavities coupling O.Q. and D.Q. for the vowel /a/.

Discussion

The above mentioned studies do not hold except for the temporal characteristics of the flow waveform. We could also in the same way consider the harmonic domain by calculating the average slope of its spectrum. Measurements have been taken for all the vowels. We could deduce some interesting results by comparing the value of α with the O.Q./D.Q. ratio as proposed by Takasugi (1977). Looking at Figure 38-11 we can verify that there is a correlation between α and O.Q./D.Q.: α increases when D.Q. decreases. This is explained by the following reasoning: the O.Q. is less sensitive to the acoustic coupling whereas the D.Q. is much more sensitive to it. In other words, when the D.Q. is large the spectrum of the flow wave is richer in high frequency harmonics than when D.Q. is small. Hence the D.Q. increase decreases the absolute value of α. The results are in accordance with the measurements of Takasugi carried out on natural speech.

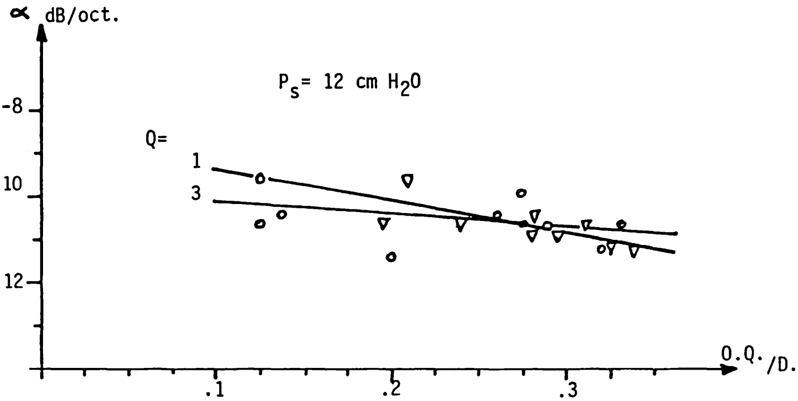


FIGURE 38-11
Correlation between the spectrum slope α and the ratio O.Q./D.Q.

Finally, when we take into account the two types of coupling (subglottal and supraglottal cavities) the effects could be summarized as follows:

1. The opening quotient O.Q. is not much modified when we consider the subglottal cavities and especially the vocal tract.
2. On the other hand, the dissymmetry quotient increases whenever we take into account a supplementary coupling.

CONTINUOUS MODEL OF THE VOCAL SOURCE

Description of the model (Descout *et al*, 1980; Perrier, 1982; Perrier *et al*, 1982)

The proposition of our model is the following: the two vocal cords are represented by a pair of beams coupled with distributed stiffness and dampers and submitted to the aerodynamic forces (Figure 38-12). These beams are biarticulated at the two extremities (like in the physiological domain at the thyroid and arytenoid). The model respects the continuity of the muscles. Vocal cords are assumed to be hinged beams. In a first step, we only consider the horizontal displacement. We apply the general equations and sign convention of the material resistance, the fundamental principle of the dynamic to a ds element leads to:

$$(EI)_i \bullet \partial^4 y/\partial x^4 + (N)_i \bullet \partial^2 y/\partial x^2 + (\rho S)_i \bullet \partial^2 y/\partial t^2 + (C)_i \bullet \partial y/\partial t = f(x,t) - g(x,t)$$

for $i = 1,2$ and when $f(x,t)$ and $g(x,t)$ are the force density (mechanical and aerodynamic forces).

We can obtain more detail in references: Descout *et al* (1980); Perrier *et al* (1982).

The numerical resolution of equations requires both temporal and spatial sampling. For the spatial derivative, we can use the center derivative definition and for the temporal derivative, we use the Newmark method.

First results

We have defined the variation ranges of the parameters as follows:

$$Ps : 4\text{-}16 \text{ cm } H_2O,\ EI : .25\ 10^{-6} - .165\ 10^{-5} \text{ U.S.I.},\ \rho s_1 : .45\ 10^{-2} - .16\ 10^{-1} \text{ kg/m}$$

Since the increase of the rigidity at the flexion corresponds to a vocal cords stretch, then it is accompanied by a decrease in the mass per unit length ρs_1. Taking into account the absence of a relation that correlates these two factors, we have arbitrarily supposed that a linear increase of EI1, corresponds to a linear increase of the mass per unit length. Figures 38-13, 38-14 and 38-15 show the influence of the subglottal pressure for different values of the vocal cord tension. A linear regression obtained from 105 simulations gives the following relation:

$$F_0 = 0.499\ 10^8\ EI_1 - 0.277\ 10^4 \rho s_1 + 0.207\ 10^{-1}\ Ps + Cte$$

The evolutions of the parameters characterising the glottal flow signal are the same type as these obtained by the two-mass model. Moreover, we studied the effects of the vocal tract input impedance on the frequency of oscillation of the model. Figure 38-16 shows the results

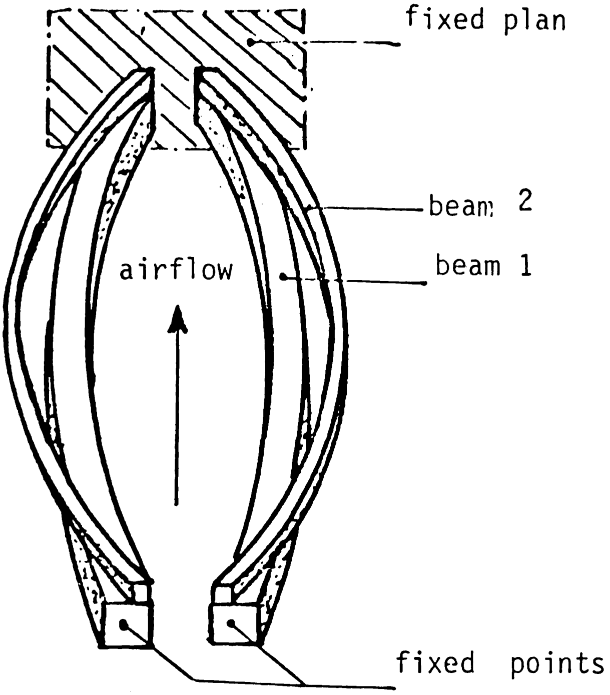


FIGURE 38-12
Continuous model of the vocal cords.

of simulating ten French vowels when we have taken account of non-linear coupling between both beams. We notice two kinds of phenomena: For values of Fl below 450 Hz we have a constant value of F_0 or a slight increase with the increase of Fl, but when Fl is above this value (450 Hz) F_0 always decreases with the increase of Fl. When no non-linear coupling is introduced, F_0 is quite constant. These last results are contrary to the theoretical results obtained from the two-mass model and also to the experiments made on natural speech. A set of simulations has been carried out with an artificial fixed second formant frequency, the results are slightly different: on the whole, we notice a very slight decrease in F_0 when Fl increases. The role of F2 does not seem to be negligible.

The fundamental theoretical study previously mentioned must be reviewed in order to take into account the first two formants. Moreover, we must keep in mind that the modeling of the aerodynamical effects is still rudimentary.

CONCLUSION

The results obtained by the aid of the two mass models of the vocal cords concerning coupling are in accordance with the measurements made on natural speech. The above discussion gives us the opportunity to suggest that the acoustic coupling cannot explain the differences in the intrinsic values of the fundamental frequency of different vowels.

Also, previous work (Guérin, 1978) shows that the coupling has no effect at all on the sound pressure level, and only the average damping differences at the time of propagation

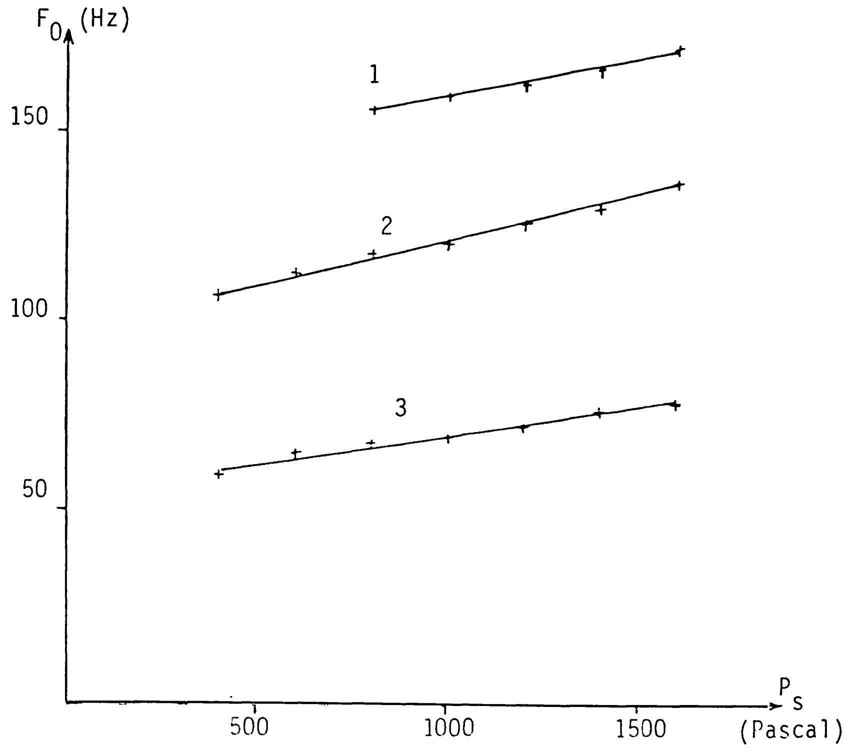


FIGURE 38-13
$F_O = f(P_s)$ for several values of EI and P_s. 1) $(EI)_1 = .165\ 10^{-5}$, $(P_s) = .43\ 10^{-2}$; 2)$(EI)_1 = .125\ 10^{-5}$, $(P_s) = .78\ 10^{-2}$; 3) $(EI)_1 = .45\ 10^{-6}$, (ps)$= .14\ 10^{-1}$.

in the vocal tract for different configurations are responsible for these intrinsic values.

The results obtained from the continuous model of the vocal cords lead us to different conclusions with regard to the intrinsic frequency of oscillation. A more profound and rigorous theoretical study would be necessary in order to form a set of representative results in accordance with the simulation results.

ACKNOWLEDGMENTS

We would like to thank Ingo Titze, W.A. Conrad and G. Fant for their valuable suggestions and helpful comments on an earlier version of this paper.

REFERENCES

Al Ansari, A., Guérin, B. and Degryse, D. (1981). "Subglottal impedance effects on the vocal source signal," 4th F.A.S.E. Symp., Venezia.

Al Ansari, A. (1981). "Etude du fontionnment et simulation en temps réel d'un modèle de la source vocale," doctoral diss., I.N.P. Grenoble.

Baer, T. (1975). "Investigation of phonation using excised larynxes," M.I.T. Thesis, Cambridge.

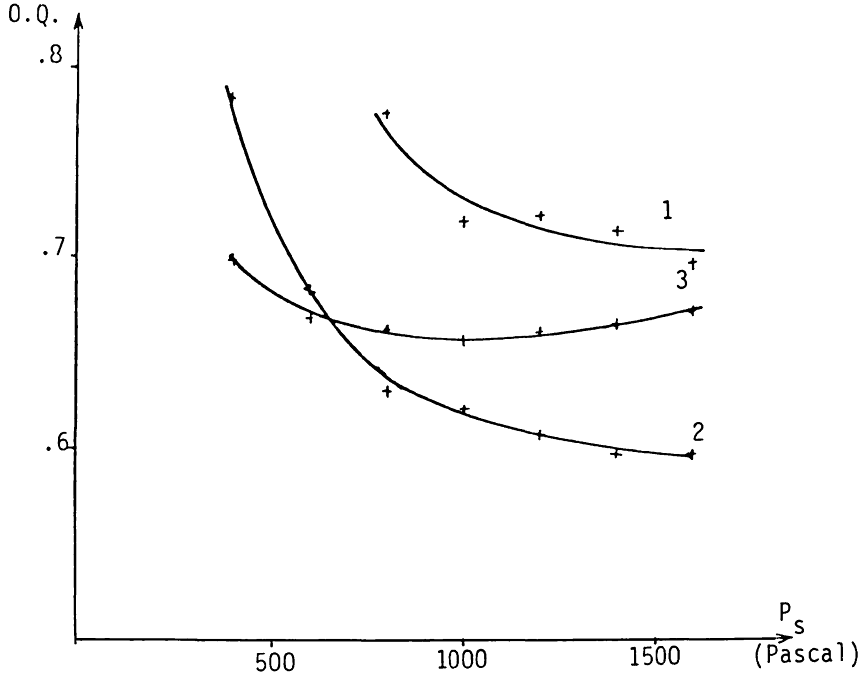


FIGURE 38-14
O.Q.=f(P_s) for the same conditions that are in Figure 38-13.

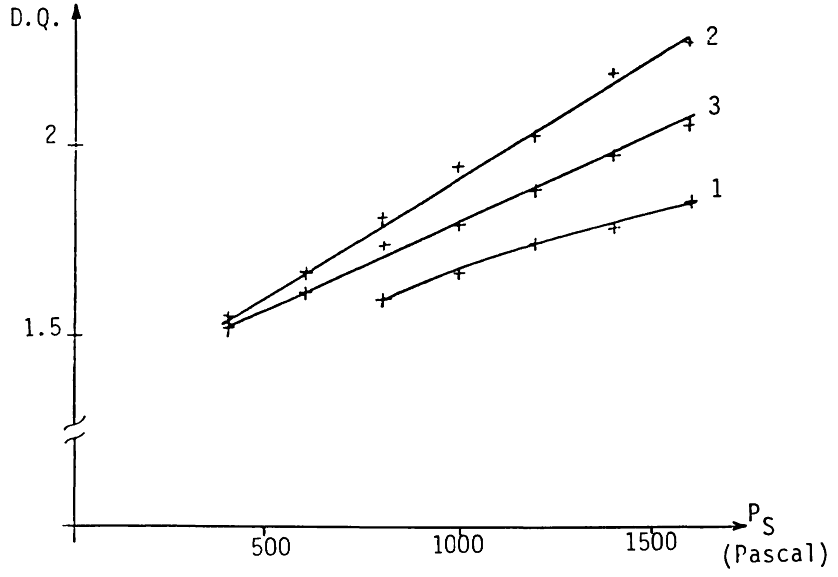


FIGURE 38-15
Effects of P_s on D.Q. for the same values of EI and P_s that are in Figure 38-13.

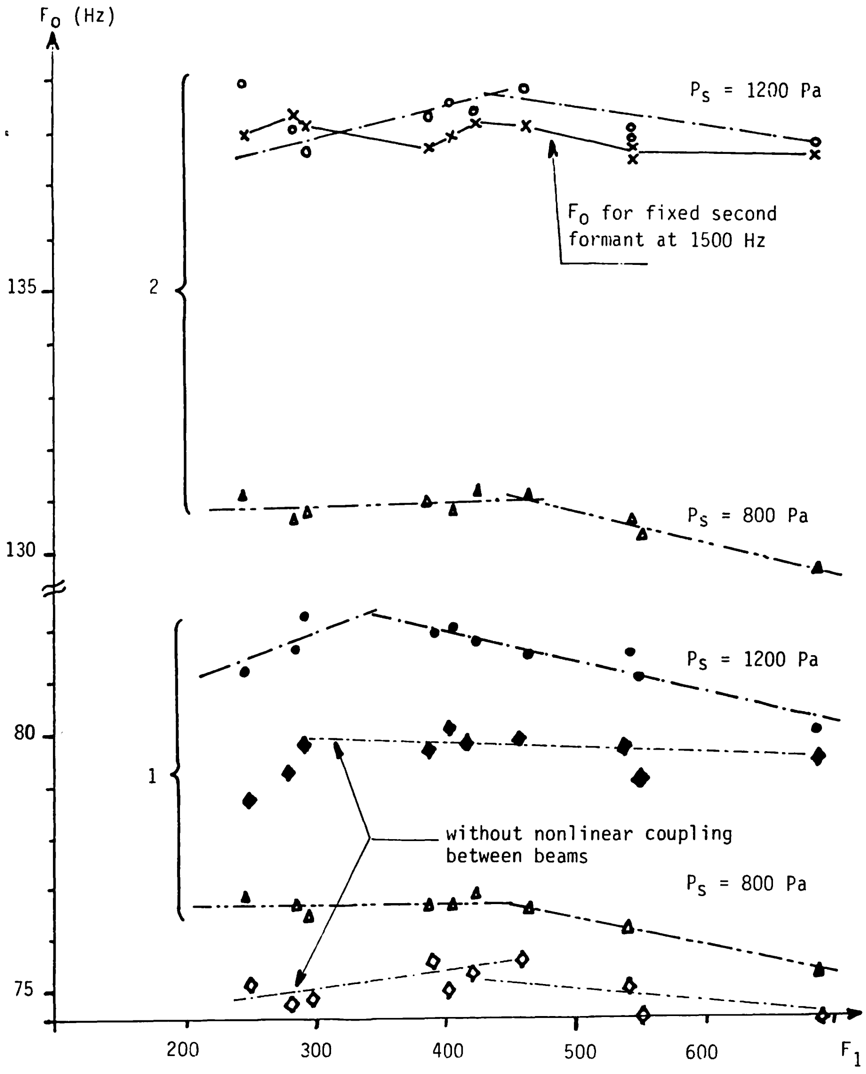


FIGURE 38-16
Fundamental frequency F_0 values as a function of the first formant frequency (simulated input impedance of the vocal tract with the first two formant). The vocal source parameters are: 1) $EI_1 = .65\ 10^{-6}$ and 2)$EI_1 = .15\ 10^{-5}$.

Beil, R.G. (1962). "Frequency analysis of vowels produced in a helium rich atmosphere," *J. Acoust. Soc. Am.* 34, 347-349.

van den Berg, Jw. (1955). "Calculation on a model of the vocal tract for vowel /i/ (meat) and on the larynx," *J. Acoust. Soc. Am.* 27, 332-338.

Carré, R. and Triboulet, J. (1974) "Acoustic characteristics of vowel nasalization," *J. Acoust. Soc. Am.* 55 S, 21.

Colton, R.H. and Estill, J. (1976). "Observation of glottal waveform in selected phonatory tasks," *J. Acoust. Soc. Am.* 59, S75(A).

Cuvier, G. (1907). "Leçon d'anatomie comparée." Paris.

Degryse, D. (1981). "Temporal simulation of wave propagation in the lossy vocal tract," 4th F.A.S.E. symp. Venezia, 193-196.

Descout, R., Auloge, J.Y. and Guérin, B. (1980). "Continuous model of the vocal source," *I.C.A.S.S.P.* 61-64.

Ewan, W.E. (1979). "Can intrinsic vowel F_0 be explained by source - tract coupling?" *J. Acoust. Soc. Am.* 66, 358-362.

Farnsworth, D.W. (1940). "High speed motion pictures of the human vocal cords," *Bell Lab. Record* 18, 203-208.

Flanagan, J.L. (1965). "Speech analysis, synthesis and perception," Springer Verlag, Berlin.

Flanangan, J.L., Ishizaka, K. and Shipley, K.L. (1975). "Synthesis of speech from a dynamic model of the vocal cords and vocal tract," *B.S.T.J.* 45, 485, 506.

Fischer, W.N., Monsen, R.B. and Engebretson, A.M. (1975). "Variations in the normal male glottal wave," *J. Acoust. Soc. Am.* 58, S41.

Guérin, B. (1978). "Contribution aux recherches sur la production de la parole-étude du fonctionnement de la source vocale-simulation d'un modele," thèse d'état, I.N.P. Grenoble.

Guérin, B., Boe, L.J. (1980). "Etude de l'influence du couplage acoustique source-conduit vocal sur Fo des voyelles orales," *Phonetica* 37, 169-192.

Holmes, N.H. (1962). "An investigation of the volume velocity waveform at the larynx during speech by means of an inverse filter," paper in Proc. of Speech Commun. Sem., B4, Stockholm.

Ishizaka, K. and Matsudaira, M. (1968). "What makes vocal cords vibrate?" paper in 6th Intern. Cong. Acoust. B13.

Ishizaka, K. and Matsudaira, M. (1972). "Fluid mechanical considerations of vocal cord vibrations," *Speech Com. Res. Lab.* Santa Barbara, CA. Monograph 8.

Ishizaka, K. and Flanagan, J.L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords." *B.S.T.J.* 51, 1233-1268.

Ishizaka, K., Matsudaira, M. and Kaneto, K. (1976). "Input acoustic-impedance measurement of the subglottal system," *J. Acoust. Soc. Am.* 60, 190-197.

Kagen, B. and Trendelenburg, W. (1937). "Zur Kenntnis der Wirkung von künstlichen Ansatzrohren auf die Stimmschwingungen," *Arch. Sprach-Stimmheilk* 1, 129.

Koike, V. and Hirano, M. (1973). "Glottal area time function and subglottal pressure variation," *J. Acoust. Soc. Am.* 54, 1618-1627.

Koike, V. (1981). "Sub- and supraglottal pressure variation during phonation," in *Vocal Fold Physiol.* K.N. Stevens and M. Hirano (Eds.) Univ. of Tokyo Press, 181-191.

Lindqvist, J. (1965). "Studies of the voice source by means of inverse filtering technique," 5th Intern'l Cong. Acoust. Liege, A35.

Lindqvist, J. (1970). "The voice source studied by means of inverse filtering," *Q.P.S.R. Speech Trans. Lab. Roy. Inst. Tech.* Stockholm, 3-9.

Mathews, M.V., Miller, J.E. and David, E.E. (1961). "An accurate estimate of the glottal waveshape," *J. Acoust. Soc. Am.* 33, 1976A.

Miller, R.L. (1959). "Nature of the vocal cord wave," *J. Acoust. Soc. Am.* 31, 667-677.

Miller, J.E. and Mathews, M.V. (1963). "Investigation of the glottal waveshape by automatic inverse filtering," *J. Acoust. Soc. Am.* 35, 1876A.

Monsen, R.B. and Engebretson, A.M. (1977). "Study of variations in the male and female glottal wave," *J. Acoust. Soc. Am.* 62, 981, 993.

Mrayati, M., Guérin, B. and Boe, L.J. (1976). "Etude de l'impédance d'entrée du conduit vocal. Couplage source conduit vocal," *Acustica* 35, 330-340.

Perrier, P. (1982). "Etude d'un modele continu des cordes vocales sous forme de deux poutres bi-articulées. Premières simulations," doctoral diss., I.N.P. Grenoble.

Perrier, P., Guérin, B. and Auloge, J.Y. (1982). "Simulation d'un modèle continu de la source vocal," *F.A.S.E./D.A.G.A.* 1059-1062.

Picard, S. (1970). "Amélioration de l'intelligibilite de la voix émise en atmosphere d'hélium." *Bull. Soc. Franc. Med. Subaquat. Hyperbare* 3, 5-8.

Rothenberg, M. (1973). "A new inverse filtering technique for deriving the glottal airflow waveform during voicing," *J. Acoust. Soc. Am.* 53/1, 1632-1645.

Rothenberg, M. and Zahorian, S. (1977). "Non linear inverse filtering technique for estimating the glottal area waveform," *J. Acoust. Soc. Am.* 61/4, 1063-1071.

Rothenberg, M. (1980). "Acoustic interaction between the glottal wave source and the vocal tract," in *Vocal Fold Physiol.*, K.N. Stevens and M. Hirano (Eds.) Univ. of Tokyo Press, 305-323.

Sondhi, M.M. (1975). "Measurement of the glottal waveform," *J. Acoust. Soc. Am.* 57, 166-178.

Sundberg, J. (1979). "Formants and fundamental frequency control in singing. An experimental study of coupling between vocal tract and voice source," *Q.P.S.R.* 1, 65-78.

Takasugi, T. (1971). "'Analysis by synthesis' method utilizing spectral features of voice source and measurement of glottal waveform parameters," *J. of the Radio Res. Lab.* 18, 97, 209-220.

Timke, R., von Leden, H. and Moore, P. (1958). "Laryngeal vibrations: measurements of the glottic wave-part I. The normal vibratory cycle," *A.M.A. Arch. Otol.* 69, 438-444.

Weibel, E.R. (1963). *Morphometry of the Human Lung*. Springer Verlag, Berlin.

Weiss, D. (1932). "Ein resonanzphanomen der singstimme," *Mschr. Ohrenheilk. Lar. - Rhinol.*, 66, 964-967.

DISCUSSION

G. FANT: Can you comment on fo change with increased vocal tract sub- or supraglottal load? I recall you have found a slight fo increase with such loading and an increased dissymmetry.

B. GUERIN: The dyssymmetry quotient is increased a little when the subglottal cavities are coupled. The variation of other characteristics like fo and O.Q. are not significant. The variation of fo with sub- and supraglottal coupling are in accordance with theoretical investigations. The inductive part of both cavities is the main cause of the increase of the dissymmetry quotient and of the modification of fo.

A. BENADE: How do you measure shifts in fundamental frequency when you change the formant frequency? Do you pull something away and watch the frequency change while someone sings?

B. GUERIN: I have not made experiments with natural speech for this paper. The results are based on simulation.

A. BENADE: I am suggesting that a simple experiment be conducted, to see what happens.

B. GUERIN: Yes, I have conducted an experiment where a small additional pipe is placed in front of the mouth. The subject speaks the vowel "ah". We rapidly spin the pipe, and we observe the same result.

A. BENADE: The spin would have to be very rapid.

B. GUERIN: Yes, it would.

A. BENADE: I am concerned that your model has no shunt impedance in the vocal fold region. While many people neglect this term, it plays an essential role in most sustained air flow systems.

B. GUERIN: It is true that there is no shunt impedance. With the glottal length being very short, we have neglected this impedance because many other aerodynamic aspects are approximate (the cross-section shape for example). In further simulation, the aerodynamic modelling will be reconsidered.

39 ▪ A SET-UP FOR TESTING THE VALIDITY OF THE TWO MASS MODEL OF THE VOCAL FOLDS

Bert Cranen and Louis Boves

INTRODUCTION

A better understanding of the behavior of the vocal folds is of importance to a number of disciplines, among which are medical diagnostics, communication engineering, and linguistics.

A direct approach to a study of the vocal folds is much impeded by the fact that their movements are difficult to observe directly. Moreover, the behavior of the vocal folds is very complex and difficult to understand since their motions are the result of a partly uncomprehended interplay of aerodynamic, viscoelastic, and acoustic effects.

The relative inaccessibility and the fact that measurements of one parameter of the voice production apparatus, while keeping all other parameters constant, are impossible in vivo, make research with the aid of a model attractive. A number of propositions with regard to modelling of the speech production apparatus have been made in the course of time. In these models each vocal fold was considered as one (Fant, 1960; Lieberman, 1967; Flanagan & Landgraf 1968), as two (Ishizaka & Flanagan, 1972), or as more than two masses coupled by means of springs and dampers (Titze, 1973, 1974), or as a viscoelastic continuum (Titze & Strong, 1975).

Modelling can be a powerful tool for isolating the most essential properties of a system. But, by modelling reality one also runs the risk of oversimplification: many observations may not be explained by too simple a model, or even worse, may be explained falsely. Thus, studying reality by means of a model has to be done with great caution. At all stages during simulation one has to be aware of the discrepancies which may exist between model behavior and reality. Only after having determined the conditions in which the model is a valid description of the system under consideration, outcomes of simulation experiments may be assumed to be reliable enough for them to be used in explaining related observations from reality.

To the best of our knowledge none of the vocal fold models mentioned has been subjected to explicit tests of their validity, at least not in another way than by attempts to establish whether the model can generate credibly looking flow pulses, or when loaded by a vocal tract, can produce vowel sounds of a normal quality. It is our intention to put the two-mass model to such a test, by means of a formal comparison of pressure and flow signals that can be measured at a number of points in the model with pressure and flow measurements carried out on normal subjects at corresponding points in the trachea and vocal tract. Specifically we intend to record pressure and flow just below and just above the glottis, together with the pressure (and perhaps the flow also) at the lips and additional information on the glottal area.

Our decision to test the fairly simple two-mass model instead of the continuum model (which undoubtedly has physiological validity over a much broader range) is based upon the fact that the number of degrees of freedom in the latter model most probably exceeds the number of criterion signals which we will be able to derive from our physiological recordings. Moreover, our measurements primarily bear on the acoustic/aerodynamic part of the vocal fold models, and it seems that both the continuum and the two-mass model share the same aerodynamic component.

The first part of this paper describes the set-up designed for making the physiological recordings. It contains a number of registrations made during a test-recording session during which an adult male subject with a normal voice produced a number of isolated vowels and a number of nonsense words of the form vCV and VCv, where v stands for an unstressed vowel, V for a stressed vowel, both taken from the set /a,e,i,o,u,y,ə/ and C for a consonant from the set /p,b; t,d; k,γ; f,v; s,z/.

Furthermore, a new calibration procedure for the pressure transducers is discussed. In the second part of the paper it is explained how we intend to extract volumeflow waveforms from differential pressure measurements. The first results of the proposed procedure are shown by means of some illustrations.

PHYSIOLOGICAL REGISTRATIONS

In order to obtain flow waveforms from pressure gradients below and above the folds. pressure recordings should be made at two different locations in the trachea and at two points in the vocal tract. For that purpose a catheter with four miniature pressure transducer elements has been used (see Figure 39-1). The catheter can be inserted through the nose between the vocal folds provided that the mucous membrane is anesthesized first. This procedure has a barely noticeable effect on phonation. The speech signal in front of the mouth is recorded by an ordinary condenser microphone. The information about glottal area variations and moments of opening and closing of the glottis may be retrieved from photoglottograph and electroglottograph signals. In our set-up both glottograms are made into AC coupled signals

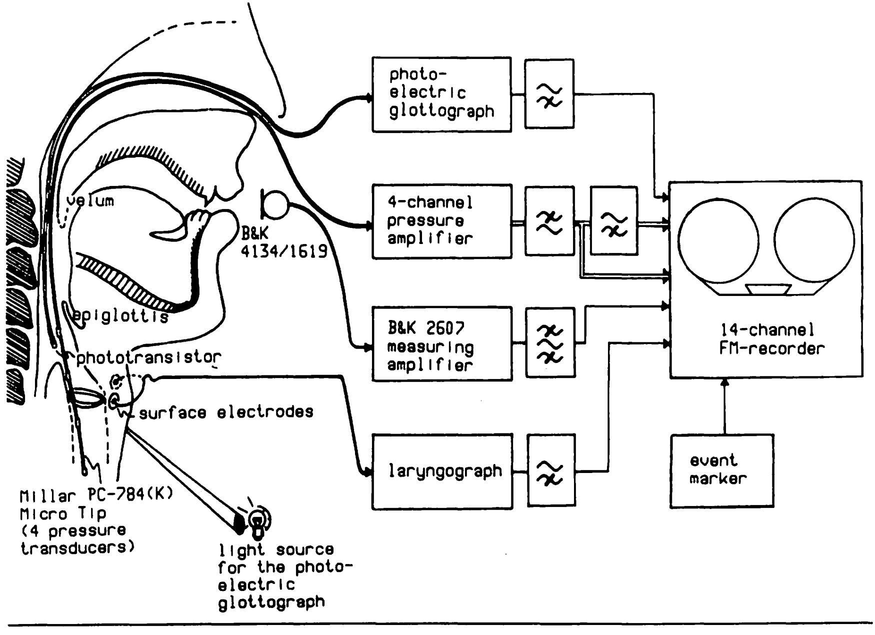


FIGURE 39-1
Schematic diagram of the experimental setup.

but the necessary high pass filtering is effected in such a way as to minimize phase distortion. Thus the glottograms give only relative information on vocal fold contact area and glottal area.

Previous experience with recording subglottal pressure showed that it is extremely difficult to find a satisfactory balance between the DC and AC components. DC variations can span a range of more than 10 cm water in normal phonation. Superimposed AC components typically amount to approximately 25% of the DC level. Therefore adjusting the gain to accommodate the DC part of the signals will reduce the AC components to only a small fraction of the dynamic range of the recording instruments. Emphasizing the AC components, however, calls for totally suppressing the DC information. As a solution to this dilemma we decided to record the four pressure signals using eight channels, four DC and four AC ones. This brings the total number of channels to record simultaneously to 11, to wit:

1-2) absolute DC pressure at two points in the trachea
3-4) AC pressure variations at two points in the trachea
5-6) absolute DC pressure at two points in the pharynx
7-8) AC pressure variations at two points in the pharynx
9) microphone signal just in front of the mouth
10) photoglottograph signal
11) electroglottograph signal

The signals are recorded on a 14 track instrumentation recorder. Afterwards they can be played back at a reduced speed allowing sampling and digitalisation by means of a general purpose mini-computer system.

The four pressure signals were obtained by means of a Millar catheter (type: PC-784[K]) with four miniature pressure transducers. The transducers are situated at equidistant intervals of 5 cm. The first transducer is located at the tip of the catheter (see Figure 39-2).

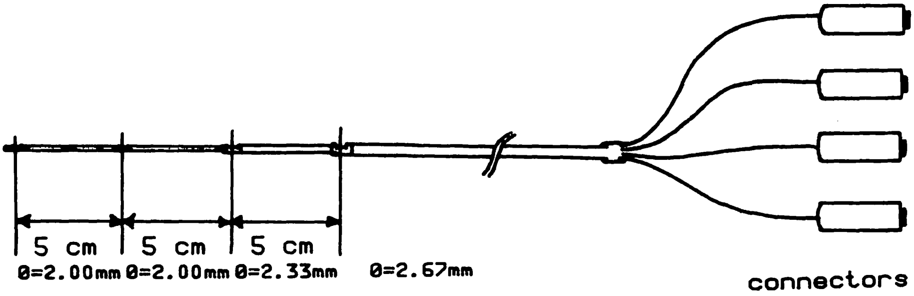


FIGURE 39-2
Dimensions of the Millar PC-784[K] catheter with its four pressure transducers.

In general, the outcomes of the measurements (Figure 39-3 through 39-6) tend to corroborate earlier observations. First of all, oscillatory behavior of the air masses around the glottis substantially increases the transglottal pressure at the moment of vocal fold closure. The customary formulation in the aerodynamic myoelastic theory of vocal fold vibration, which states that after closure the folds are pushed apart by the restored pressure, must clearly be amended in order to account for the fact that subglottal pressure is maximal at the moment at or just before closure. Explanations of tissue/air interactions solely on the basis of steady flow dynamics are not adequate.

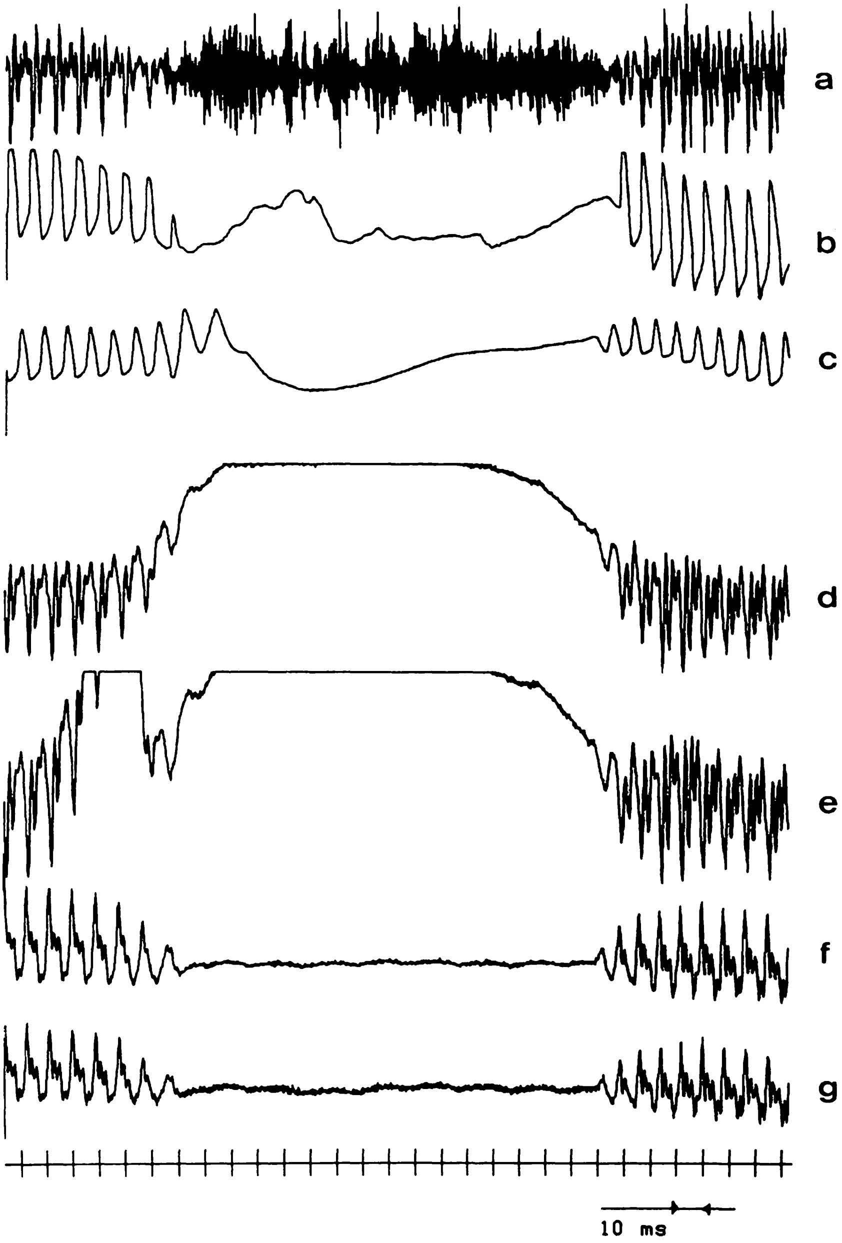


FIGURE 39-3
Seven simultaneous registrations during the production of /əsə/. a) speech signal just in front of the mouth. b) EGG. c) PGG. d) pressure signal in the upper part of the pharynx. e) pressure in the lower part of the pharynx. f) pressure just below the glottis. g) pressure in the lower part of the trachea.

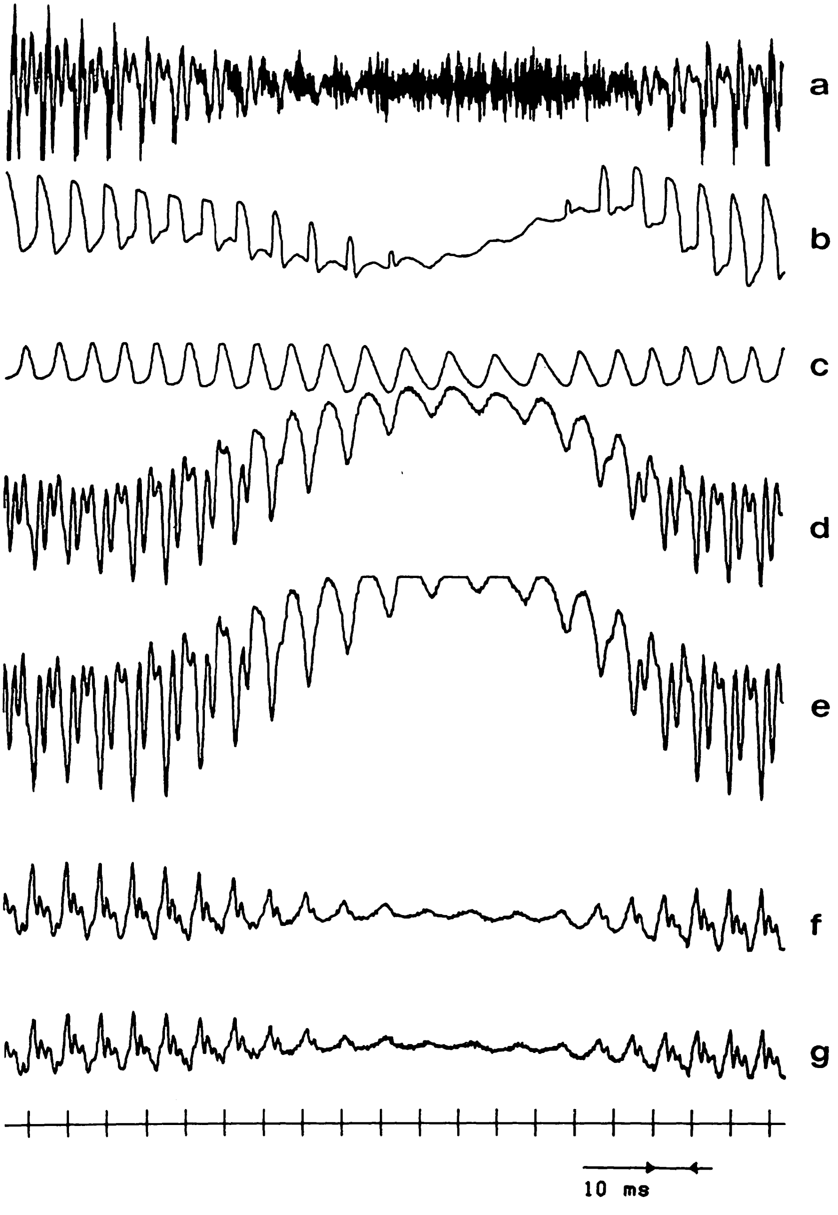


FIGURE 39-4
Seven simultaneous registrations during the production of /əzə/. a) speech signal just in front of the mouth. b) EGG. c) PGG. d) pressure signal in the upper part of the pharynx. e) pressure in the lower part of the pharynx. f) pressure just below the glottis. g) pressure in the lower part of the trachea.

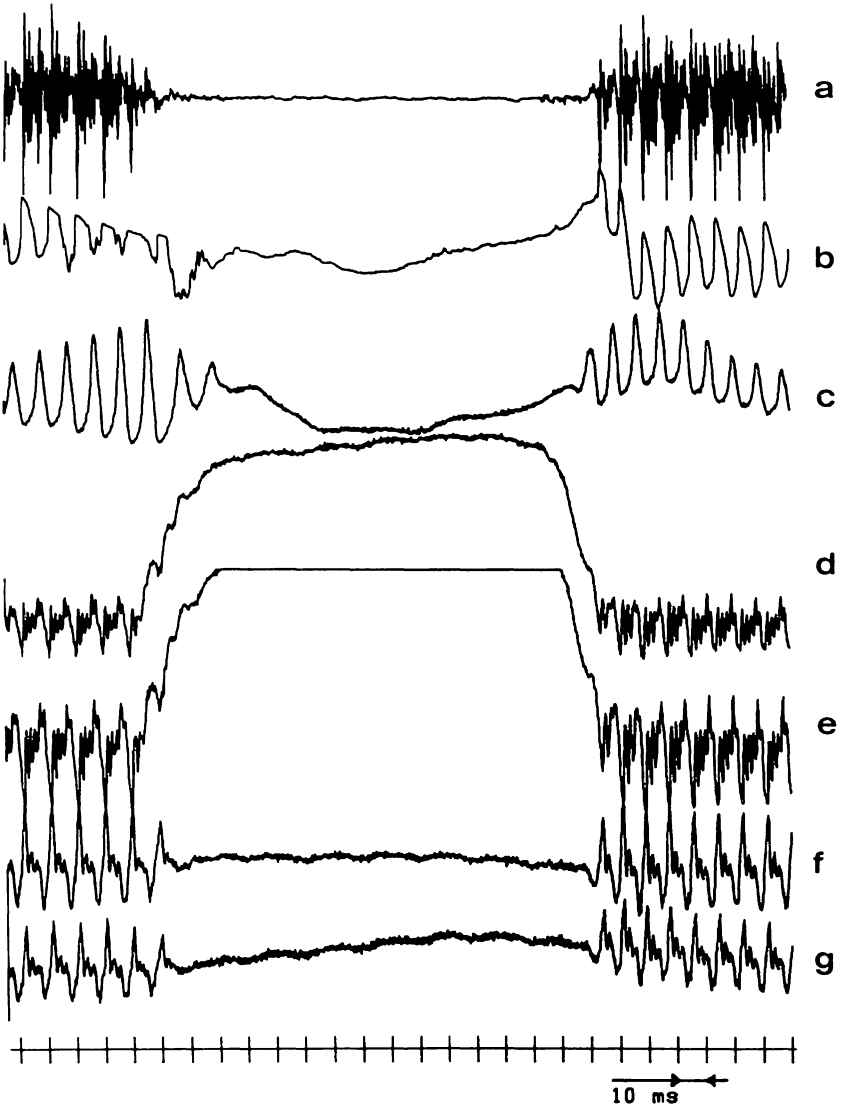


FIGURE 39-5
Seven simultaneous registrations during the production of /apa/. a) speech signal just in front of the mouth. b) EGG. c) PGG. d) pressure signal in the upper part of the pharynx. e) pressure in the lower part of the pharynx. f) pressure just below the glottis. g) pressure in the lower part of the trachea.

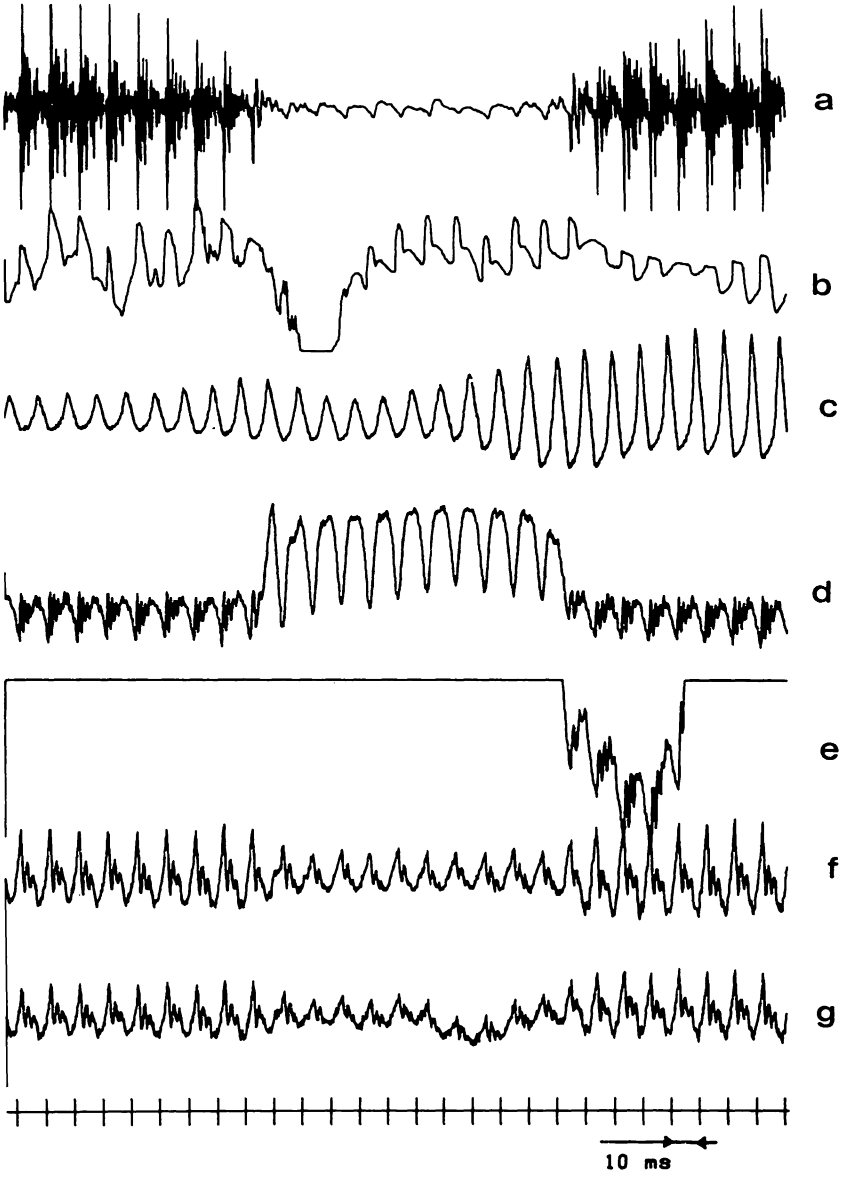


FIGURE 39-6
Seven simultaneous registrations during the production of /aba/. a) speech signal just in front of the mouth. b) EGG. c) PGG. d) pressure signal in the upper part of the pharynx. e) pressure in the lower part of the pharynx. f) pressure just below the glottis. g) pressure in the lower part of the trachea.

Furthermore, different measures of vocal fold motion may lead to different conclusions, especially with respect to voiced/unvoiced decisions. In the /əzə/ registration for instance (Figure 39-4) the acoustic speech signal looks hardly periodic, and the EGG also seems to indicate cessation of the regular fold motion. The PGG, however, keeps showing regular vibration throughout the consonant, be it that the form of the area pulses changes substantially. All pressure signals maintain periodic vibrations, though with markedly changed amplitude and frequency content. Similar findings were reported by Flanagan, Rabiner, Cristopher, Bock and Shipp (1976).

The photoglottogram of the /əsə/ utterance (Figure 39-3) suggests that the indisputable cessation of vocal fold vibration is caused by an abduction of the vocal folds. This confirms observations like those in Löfqvist and Yoshioka (1980). In order to enhance the usefulness of photoglottographic recordings it might be advantageous to combine AC and DC coupled signals too.

Finally it can be seen from the /aba/ registrations (Figure 39-6) that there may occur considerable larynx movements which give rise to problems in measuring the pressure: one of the transducers may get between the folds. The unsteady EGG was caused by electrodes which were fixed too loosely.

Up until now we have only considered timing relations and waveform features. An important question which still remains to be answered is: what absolute pressure measures can be obtained from the measurements?

Absolute in vitro calibrations of the pressure transducers have proven to be completely useless because of uncontrollable behavior of the transducers when exposed to a temperature shock. In order to avoid problems connected with the temperature sensitivity of the transducers we now propose an in vivo calibration procedure.

The measurements for the calibration can be split up into two parts. First, the subject opens the glottis without any appreciable breathing. In this situation all transducers will undisputably represent atmospheric pressure. The DC levels at the outputs of unbalanced amplifiers can then be recorded and later on be compensated for.

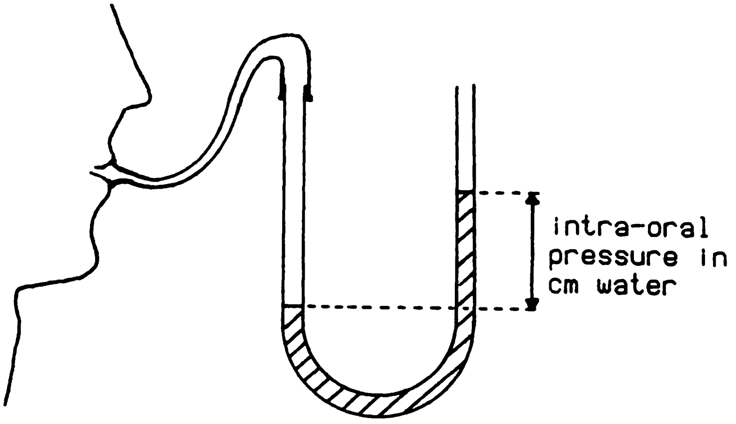


FIGURE 39-7
A water-filled manometer used for in vivo calibration of the pressure signals.

During the second part of the procedure the subject blows into one leg of an open manometer filled with water (see Figure 39-7). In doing so he tries to maintain a constant level of the water column during a period which is sufficiently long to allow vibrations of the watercolumn to damp out. A second person then reads the manometer, and his comment is recorded via the speech channel. This part of the calibration procedure enables the experimenter to compensate for possible sensitivity differences between the individual transducer/amplifier chains, and to construct an absolute pressure/output voltage graph for the DC coupled pressure channels.

This calibration procedure, which uses static calibration signals, is not suited for the AC coupled channels. We have good hope, however, that comparison of the energy in the AC part of the calibrated DC coupled recordings with the energy in the AC coupled versions may give sufficiently accurate estimates for differences in sensitivity for these channels too.

At this moment we did not yet use our measurements and try to calibrate our pressure recordings in an absolute sense. In the next part of this paper we will use an energy criterion instead when presenting the first results in our search for reliable flow waveforms.

PRESSURE GRADIENT IN RELATION TO VOLUME FLOW

For the moment we will direct our attention to acoustic phenomena, that is, we will not consider flow phenomena implying a net transport of mass. It is supposed that wave propagation can be described by linear differential equations and, as a consequence, that acoustic phenomena are superposed on and will not interact with steady flow.

With viscosity of a fluid neglected, the linearized equation of momentum becomes:

$$\rho_o \frac{\partial \vec{u}}{\partial t} = -\nabla p \tag{39-1}$$

and the equation of continuity:

$$\frac{\partial \rho}{\partial t} = -\rho_o \nabla \cdot \vec{u} \tag{39-2}$$

Here $\vec{u}$ denotes the particle velocity vector, p the pressure, ρ_0 the mass density of the fluid at rest, and ρ-ρ_0 the departure of the mass density from the state at rest. In case of plane wave propagation along the x-coordinate the unidirectional equations become:

$$\rho_o \frac{\partial \vec{u}}{\partial t} = -\frac{\partial p}{\partial x} \tag{39-3}$$

$$\frac{\partial \rho}{\partial t} = -\rho_o \frac{\partial \vec{u}}{\partial x} \tag{39-4}$$

From Equation (39-3) it can be seen that if the waveform of the spatial derivative of the pressure at location X_0 is known, the particle velocity waveform $u(X_0,t)$ can also be obtained, and hence, provided the crossectional area of the tube may be assumed constant, also the waveform of the volume velocity.

The spatial derivative of the pressure at location X_0 is defined as:

$$\left.\frac{\partial p}{\partial x}\right|_{X_0} = \lim_{\Delta x \to o} \frac{p(X_0 + \Delta x/2) - p(X_0 - \Delta x/2)}{\Delta x} \tag{39-5}$$

If Δx is taken much smaller than a quarter of the smallest wavelength we are interested in,

$\left.\frac{\partial p}{\partial x}\right|_{X_0}$ may be approximated by:

$$\left.\frac{\partial p}{\partial x}\right|_{X_0} = \frac{p(X_0 + \Delta x/2) - p(X_0 - \Delta x/2)}{\Delta x} \tag{39-6}$$

Thus, by integration of the differential pressure at location X_0 with respect to time, we obtain an estimate of the waveform of volumeflow at location X_0.

FLOW WAVEFORM ESTIMATION IN PRACTICE

Practically, the procedure of estimation of $\left.\frac{\partial p}{\partial x}\right| X_0$ appears to be not as straightforward as suggested in the former section. As a consequence of the fact that pressure signals recorded at closely spaced points in the same tube are nearly equal, and because the pressure signals have a rather low S/N-ratio to begin with (<40 dB), the S/N-ratio of the differential signal will be extremely low. Moreover, if one realizes that the transducers may have different sensitivities and/or different noise levels it will be clear that a severe estimation problem arises.

Obviously a small gain error in one pressure signal will give rise to an appreciable error in the differential signal estimate: the differential signal will be augmented with a waveform which is a constant factor times the pressure waveform. This reasoning clearly illustrates the need for an exact compensation of the individual sensitivity differences of the transducers. In order to solve this problem as well as possible for the time being without being forced to carry out extensive calibration procedures, we have adopted an energy criterion in the following way.

First the two pressure signals are band pass filtered by means of a digital linear phase filter with cut off frequencies at ca. 50 Hz and 1600 Hz. Next the signals are normalized with respect to their variance (total energy content). Note that this approach denies the existence of appreciable amplitude differences due to nodes and anti-nodes of standing waves. After normalization the signals are subtracted and integrated with respect to time. Results of this procedure are shown in Figures 39-8 and 39-9 for both under and above the glottis.

The integrated differential signals show clearly deterministic components. Whether the waveforms are good approximations of the real flow, however, still has to be established. It is likely that the energy criterion which was used to estimate the pressure gradient yields poor results, especially for the supraglottal waveforms.

Additional criteria will have to be formulated in order to decide whether our flow estimates are accurate. When physical flow calibration is impossible, the only way to accomplish this is to make cross checks with the predictions from models. Comparison of waveforms derived from the registrations and those produced by models of tubes terminated by proper termination impedances may provide us with the necessary data, at the risk of getting caught in circular reasoning.

CONCLUSION

In this paper we presented some preliminary results of an investigation into the relations between acoustic phenomena around the larynx and vocal fold motions. By means of a catheter with four pressure transducer elements the pressure at two locations in the trachea and at two locations above the glottis was measured. The speech signal just in front of the mouth,

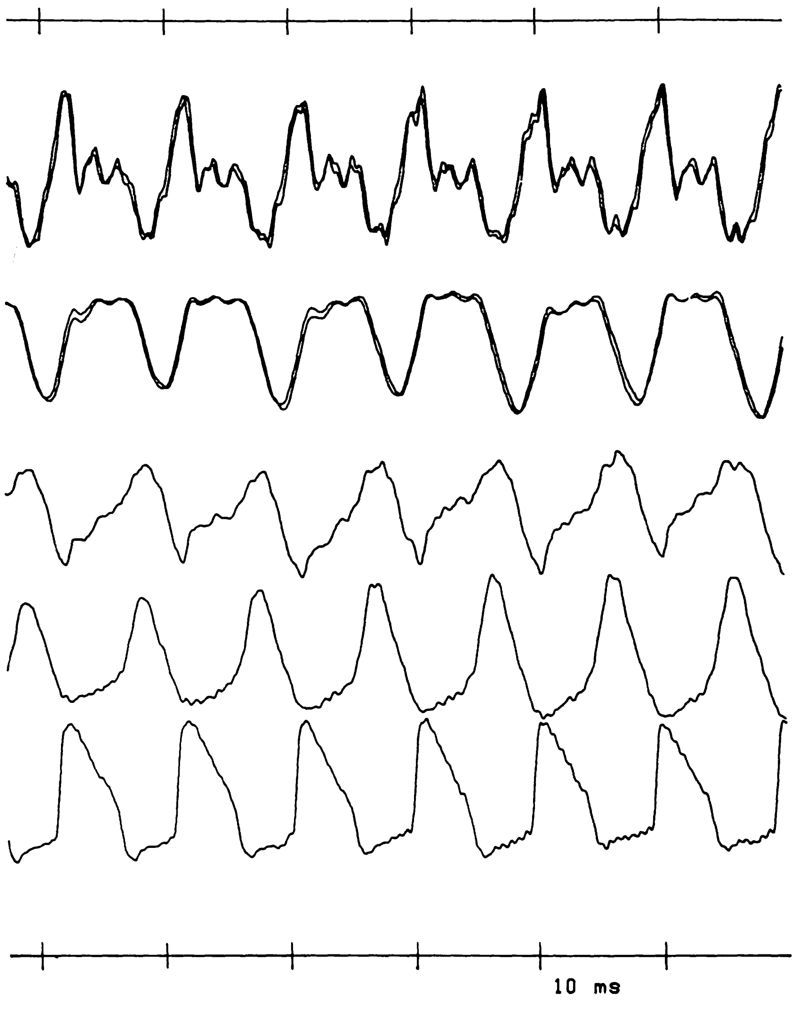


FIGURE 39-8
Simultaneous registrations during the production of /a/ and some derived signals. a) two subglottal pressure signals (normalized with respect to energy content). b) the integrated version of a). c) the difference signal of the signals in b). This corresponds to an estimate for the subglottal flow waveform. d). PGG. e). EGG.

the EGG, and the PGG were simultaneously recorded with the four pressure signals. It was shown that the procedure developed for recording these signals simultaneously is viable.

Preliminary attempts to process pressure difference signals in order to obtain estimates of the air flow in the trachea and in the pharynx have yielded promising results. A definitive proof of the contention that the time integral of differential pressure can be interpreted as flow has, however, still to be given. For giving a formal proof it will be necessary to get ac-

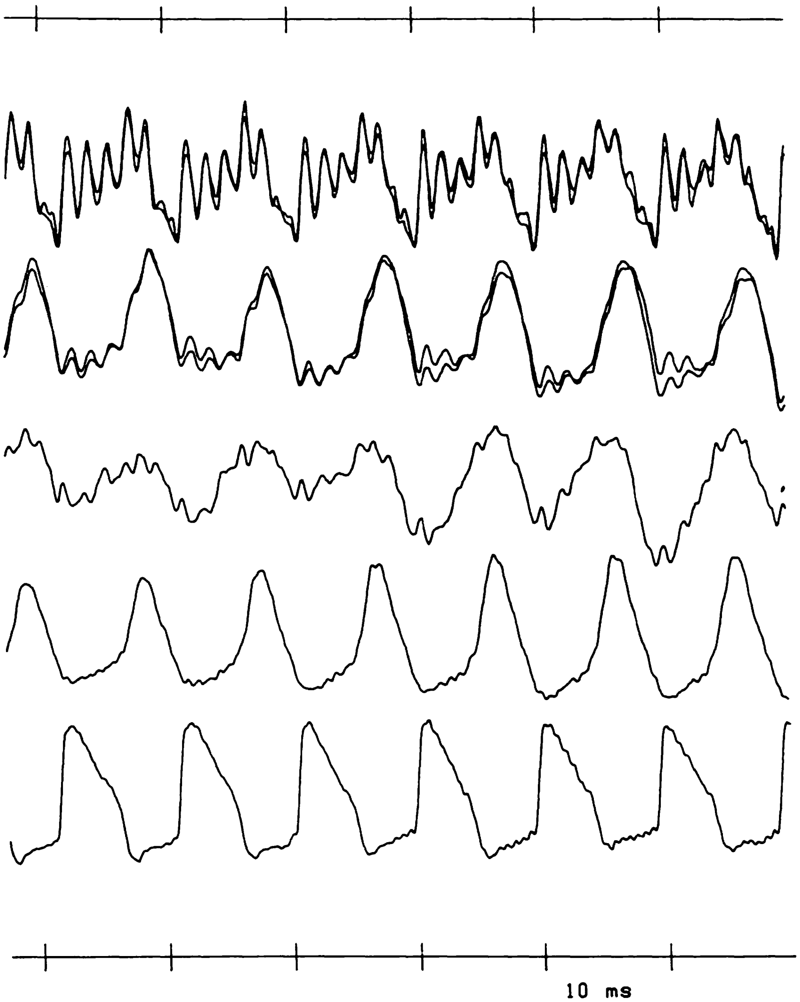


FIGURE 39-9

Simultaneous registrations during the production of /a/ and some derived signals. a) two supraglottal pressure signals (normalized with respect to energy content). b) the integrated versions of a). c) the difference signal of the signals in b). This corresponds to an estimate for the supraglottal flow waveform. d) PGG. e) EGG.

curate estimates of all sources of noise and error in the pressure measurement system. We expect to derive those estimates from careful in vivo calibrations of the pressure transducers which are planned for the near future. Calculated flow waveforms, together with measured pressure and glottal area waveforms, will be used in order to test the physiological reality of the two-mass model of the vocal folds.

ACKNOWLEDGEMENT

This research was supported by the Foundation for Linguistic Research, which is funded by the Netherlands Organization for the advancement of pure research, ZWO. Dr. C.I.E. Jansen of the University Hospital E.N.T. Clinic assisted in making the recordings.

REFERENCES

Cranen, B. and Boves, L. (1981). "The validity of a two mass model of the vocal folds in the field of linguistics - an experimental design," *Proceedings of the Institute of Phonetics Nijmegen* 5, 64-74.

Fant, G. (1960). *Acoustic Theory of Speech Production.* 's Gravenhage: Mouton & Co.

Flanagan, J.L. and Landgraf, L. (1968). "Self oscillating source for vocal tract synthesizers," *IEEE Trans. on Aud. and Electroacoust.* AU-16, 57-64.

Flanagan, J.L., Rabiner, L.R., Cristopher, D., Bock, D.E. and Shipp, T. (1976). "Digital analysis of laryngeal control in speech production," *J. Acoust. Soc. Am.* 60, 446-455.

Ishizaka, K. and Flanagan, J.L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* 51, 1233-1268.

Lieberman, P. (1967). *Intonation, Perception and Language.* MIT Press, Cambridge, Mass.

Löfqvist, A. and Yoshioka, H. (1980). "Laryngeal activity in Swedish obstruent clusters," *J. Acoust. Soc. Am. 68*, 792-801.

Titze, I.R. (1973). "The human vocal cords: a mathematical model (part 1)," *Phonetica* 28, 129-170.

Titze, I.R. (1974). "The human vocal cords: a mathematical model (part 2)," *Phonetica* 29, 1-21.

Titze, I.R. and Strong, W. (1975). "Normal modes in vocal cord tissues," *J. Acoust. Soc. Am.* 57, 736-744.

DISCUSSION

P. KITZING: First of all, I want to congratulate you on your fine measurements. I have had some experience with these types of measurements, and I can say with authority that it is not that easy to get this level of quality.

Perhaps you could use a flexible fiberoptics system: you could get rid of the rod and use a fiberoptic light fountain to put the light where you need it.

The light source is just one issue. Anybody who does these kinds of experiments should be advised to have a physician at hand. When I started such experiments, I was the physician, and I was the subject. I ended up with pneumonia from aspiration because of the local anesthesia you need. I would also suggest that you use an injection of atropine for this work. Without this drug there is a slight risk of spasm in the larynx.

L. BOVES: In response to your first remark, I would be very happy to have a flexible light guide. All of the flexible light guides available in our ENT clinic, however, are driven by AC power.

G. FANT: I feel that you could improve those flow functions by making use of inverse filtering. In theory, the signal you measure should have formant ripples superimposed, just as with the integrated speech wave itself. To perform this inverse filtering you would have to determine the appropriate transfer function that applies to the place where your signal is measured.

L. BOVES: I appreciate your suggestion. Determining the correct inverse filter, however, would not be very easy.

P. MILENKOVIC: Do mucosal secretions clog the microphones which you use, or is there some cleaning mechanism?

L. BOVES: Yours is a common question. We have never had a problem with this matter. The pressure in the trachea as well as through the glottal region is rather high, and a small amount of mucosal secretion should not interfere with the microphones too much.

R. SCHERER: This question concerns the severe problem of temperature effects with the silicon pressure transducers. If you have a transient temperature imposed on the transducer, it will react in a very sensitive manner - it takes some time for the transducer to stabilize. What if you took someone with a large capacity of air in the lungs who could expire the air relatively slowly? The subject would produce slowly moving air over a period of several seconds, not moving the air so fast as to deplete the lungs, so that the transducer would be in the right environment prior to phonation?

L. BOVES: That is not the problem. The problem is any time the catheter is brought from room temperature to say 37 degrees centigrade, it will react somewhat differently. As long as it reacts differently, you never can get an absolute calibration. You cannot predict how the sensors will behave the moment they are introduced to a subject.

R. SCHERER: My comment refers to the length of time it might take the transducer in an approximately 37 degree, 100% relative humidity environment, to stabilize sufficiently to be calibrated under those conditions. The idea is to reproduce that temperature and humidity either prior to inserting the catheter in your subject or while the catheter is in the subject.

M. ROTHENBERG: One may try to run the experiment in a sauna with high room temperature and relative humidity.

40 ▪ VOCAL TRACT AREA FUNCTION FROM TWO POINT ACOUSTIC MEASUREMENTS

Paul Milenkovic

ABSTRACT

A new method for estimating the vocal tract area function makes use of two acoustic measurements. One measurement is of far field acoustic pressure, the other of acoustic pressure measured intraorally. The calculation of the area function from two acoustic measurements is an extension of techniques for estimating the area function from far field acoustics alone. Use of intraoral pressure eliminates uncertainty in the area function estimate resulting from unknown glottal excitation and glottal loss. The measurement of intraoral pressure is made with a probe tube. Acoustic data is presented on a new probe tube configuration which eliminates undesired resonances.

INTRODUCTION

A variety of methods have been proposed to estimate the vocal tract area function from acoustic measurements (Atal, 1970, 1971; Sondhi, 1979, 1971; Wakita, 1973; Markel,*et al* 1976). The area function refers to the cross sectional area of the vocal tract as a function of position. These methods fall into two categories. In the first category, the vocal tract area function is estimated from an impulse response measurement taken at the lips (Sondhi,*et al* 1979, 1971). In the second category, the vocal tract area function is estimated from the acoustic speech waveform (Atal, 1970, 1971; Wakita, 1973; Markel,*et al* 1976).

The vocal tract impulse response is measured by applying an acoustic impulse to the lips with an impedance tube. To use the apparatus, one must articulate speech sounds while keeping one's lips pressed firmly against the impedance tube. A good seal must be maintained for the impulse response measurement to be accurate. One must also articulate speech without phonating to avoid creating acoustic interference.

The primary limitation of the impulse response measurement is in the conditions imposed by the impedance tube. One must articulate speech without phonating, and one's lips must be pressed against the impedance tube. Both of these conditions are a significant departure from conditions occurring in fluent speech.

The direct estimation of the vocal tract area function from the acoustic speech waveform avoids the fluency restrictions of the impedance tube. Even if the area function estimates are valid only for vowel sounds, the area function can be estimated for these sounds in fluent speech.

The area function is estimated from the speech waveform in two stages: the transfer function of the vocal tract is estimated from the speech waveform; the area function is in turn computed from the transfer function. Proper estimation of the transfer function requires measurement of both the input and the output of the vocal tract. The output of the vocal tract is given by the speech waveform. The input to the vocal tract is not measured—it is inferred by making an assumption about the spectrum of the glottal excitation driving the vocal tract.

The assumption about the glottal excitation spectrum is the primary difficulty with estimating the area function from speech. A common assumption is that the second derivative of the glottal airflow wave has a flat spectrum. This assumption has never been adequately tested because glottal airflow is difficult to measure directly. To the extent that the true glottal wave spectrum departs from this assumption, the estimate of vocal tract area function will be in error. This will in turn be reflected in errors in the vocal tract area function estimate (Milenkovic, 1981).

An alternative to making inferences about the vocal tract input is to measure it directly. This new approach involves measuring acoustic pressure both inside and outside the oral cavity. The vocal tract transfer function can be computed directly from the two measurements.

This paper presents a technique for computing the vocal tract area function from the pressure transfer function. This type of transfer function is the relation between the acoustic pressure wave measured in the far field and acoustic pressure measured just above the vocal folds. The importance of the pressure wave transfer function is that pressure is measurable with conventional microphones. Measurement of the intraoral acoustic pressure, however, will be a crucial part of the experimental procedure.

The glottal acoustic flow wave can also be computed from the two pressure measurements. The vocal tract admittance, seen looking out from the glottis towards the lips, can be computed from the transfer function. This admittance function is the transfer function of an inverse filter that recovers glottal flow from measured glottal pressure.

Acoustic pressure is measured in the region just above the vocal folds by way of a probe tube. A probe tube offers advantages of safety, reduced difficulty of placement, and reduced sensitivity to mucosal secretions over a miniature microphone. The disadvantage of a probe tube is that its frequency response has strong resonances in the frequency range of interest. This paper presents experimental verification of a probe tube design that avoids these resonances.

This paper is divided into three major parts. The first part is a review of the theory of how the vocal tract area function is estimated from the acoustic wave. The second part presents the technique for estimating the area function from the two acoustic measurements. The third part of the paper describes probe tube technology.

AREA FUNCTION FROM ACOUSTIC WAVE

Atal Formulation

The first technique for estimating the vocal tract area function from the acoustic wave is based on a formulation due to Atal (1970, 1971). The technique derives the vocal tract areas from the vocal tract transfer function, expressed as a frequency response. It is assumed that the vocal tract frequency response is represented by the speech spectrum or by some modification of the speech spectrum that corrects for the influence of the glottal wave.

The acoustic assumptions made in the Atal formulation are shown in Figure 40-1. The output of the vocal tract is taken to be the pressure wave occurring in the space outside the vocal tract. The input to the vocal tract is taken to be the flow wave occurring at the glottis. The vocal tract is assumed to be lossless and to support plane wave propagation.

The assumption of plane waves permits replacing the vocal system with a chain of two-port networks, also shown in Figure 40-1. The glottis is represented by a flow source. Both the vocal tract and the coupling between the lips and free space are represented by a two port network. The impedance of free space is represented by an acoustic resistance.

The Atal formulation expresses the driving point impulse response of the vocal tract, seen at the glottis, in terms of the vocal tract freqeuncy response. The driving point impulse

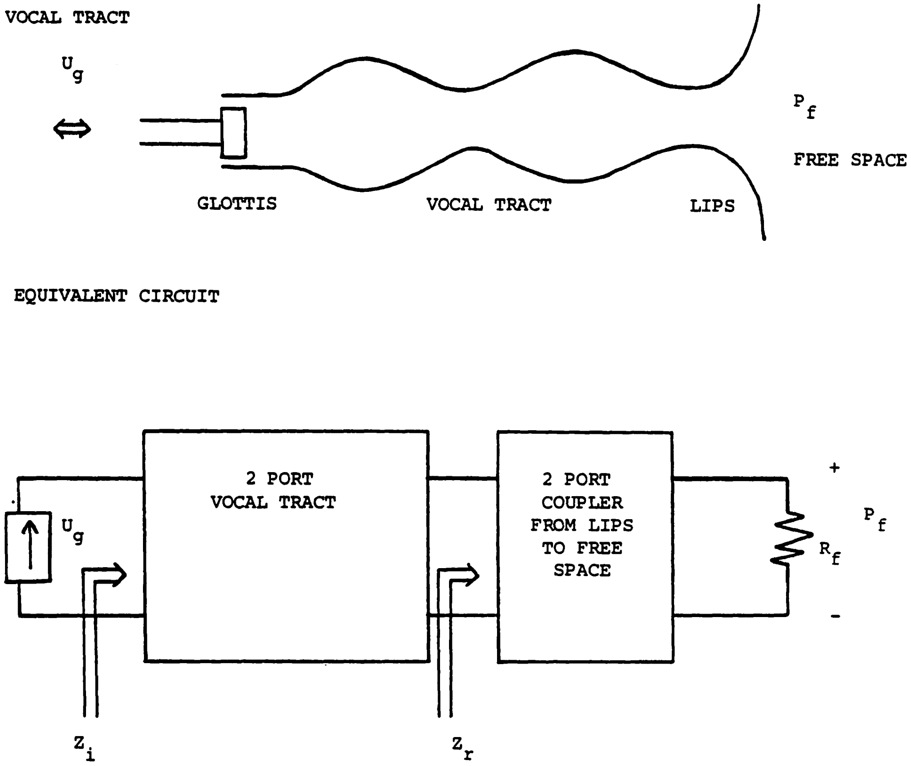


FIGURE 40-1
Atal formulation vocal tract model.

response is the pressure wave at the glottis that occurs when a glottal impulse is applied to the vocal tract. This impulse response uniquely determines the area function of a lossless vocal tract (Sondhi, 1971).

The impulse response comes from the frequency response by way of an energy balance argument. Assuming that the vocal tract is lossless, the power entering the vocal tract at the glottis equals the power being radiated into free space:

$$|U_g|^2 \operatorname{Re}(Z_i) = |P_f|^2 \frac{1}{R_f} \tag{40-1}$$

Rearranging terms gives:

$$\operatorname{Re}(Z_i) = \frac{1}{R_f} \frac{|P_f|^2}{|U_g|^2} \tag{40-2}$$

The quantity on the right hand side is simply the magnitude squared of frequency response relating glottal flow to far field pressure:

$$\mathrm{Re}(Z_i) = \frac{1}{R_f} S(f) \tag{40-3}$$

The lossless vocal tract has permitted us to express the real part of vocal tract impedance in terms of a proportionality with the vocal tract frequency response. The constant of proportionality is the impedance of free space. Note that the transfer function phase does not enter into the calculation.

The impulse response is obtained by taking the inverse Fourier transform of Equation (40-3):

$$h_e(t) = \frac{1}{R_f} R(t) \tag{40-4}$$

The left hand side is the even part of the impulse response. The right hand side is a constant times the autocorrelation function of the vocal tract transfer characteristic. This autocorrelation function is computed by correcting the autocorrelation of the speech wave for the effects of the glottal wave.

The even part of the impulse response h(t) is given by:

$$h_e(t) = \tfrac{1}{2}(h(t)+h(-t)) \tag{40-5}$$

The vocal tract is a causal system: $h(t) = 0$ for $t < 0$. This means that h(t) is recoverable from its even part. Vocal tract areas are uniquely determined by h(t) for a lossless vocal tract. The areas are determined sequentially, starting in the glottal region and working out towards the lips.

Discrete Section Version

There is a discrete section acoustic tube model of the vocal tract which lends itself to a discrete time formulation: this model is shown in Figure 40-2. This version of the model is useful for performing operations on a digital computer. The discrete time version of Equation (40-4) is:

$$h_e(nT) = \frac{1}{R_f} R(nT) \tag{40-6}$$

where T is the round trip flight time for a sound wave in one of the acoustic tube sections.

Consider the two discrete section models shown in Figure 40-2: Tube A and Tube B. Tube A is terminated in a radiation load, representing the acoustic coupling between the lips and free space. Tube B is terminated in a semi-infinite tube having a resistive characteristic impedance. Tubes A and B will have similar driving point impulse responses:

$$h_A(nT) = h_B(nT) \quad \text{FOR } 0 \leq n \leq N \tag{40-7}$$

where N is the number of discrete sections between the glottis and point C. The two tubes are indistinguishable by their impulse responses over this interval because sound waves that have been reflected back into Tube A by the radiation impedance have not made it back to the glottis in time NT.

The discrete case of the Atal formulation is often cast in terms of Tube B (Atal,*et al* 1971). An algorithm that computes areas from the impulse response, however, is equally valid for

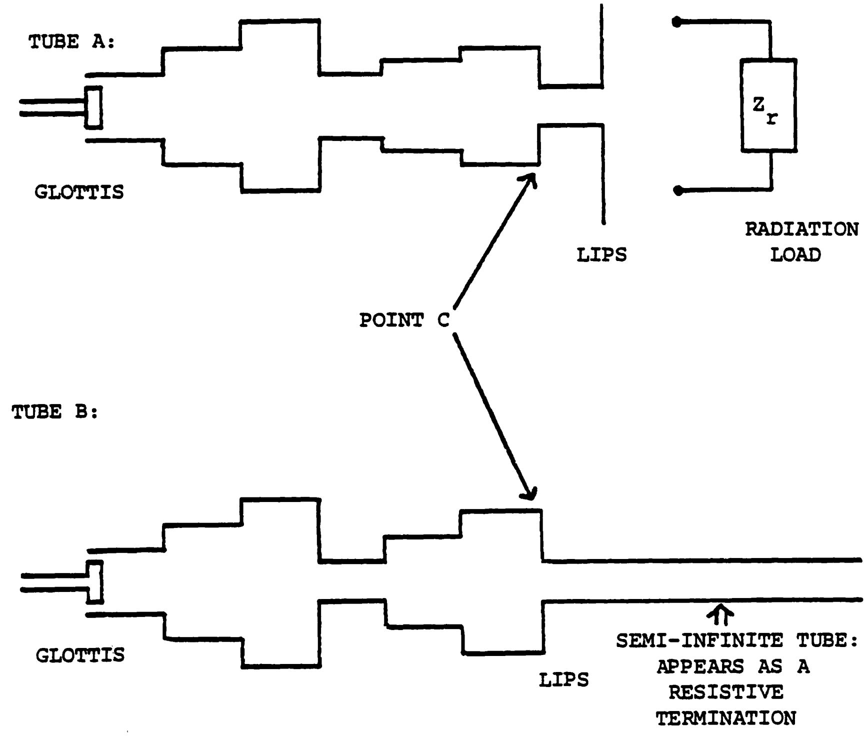


FIGURE 40-2
Discrete section vocal tract model.

Tube A for areas up to point C. This follows from the equality of the impulse responses for the two tubes in the interval [0,N]. It is possible to compute additional areas beyond point C in Figure 40-2. These areas will have a different physical meaning because they represent the radiation load.

The discrete case algorithm that computes areas from the impulse response is derived from the Durbin recursion (Makhoul, 1975). The recursion will generate reflection coefficients r along with intermediate quantities. The intermediate quantities are the successive orders of autoregressive coefficients a and the energy parameter e. The recursion is initialized by setting:

$$a_{00} = 1 \tag{40-8}$$

$$e_0 = h_e(0) \tag{40-9}$$

The Durbin recursion consists of evaluating the following four equations for successive values of i, starting with i = 1:

$$r_i = \frac{-\left(\sum_{j=0}^{i} h_e([i-j]T)a_{ij} \right)}{e_{i-1}} \tag{40-10}$$

$$e_i = e_{i-1}(1-r_i^2) \tag{40-11}$$

$$a_{io} = 1 \tag{40-12}$$

$$a_{ij} = a_{i-1\ j} + r_i a_{i-1\ j-1} \tag{40-13}$$

The areas are determined from the reflection coefficients r by:

$$S_{i+1} = S_i \frac{(1+r_i)}{(1-r_i)} \tag{40-14}$$

The areas are numbered starting with the glottis and working outward towards the lips. The reflection coefficients specify only area ratios. The absolute scale of the area function has to be determined by independent means. The absolute scale could be set by fixing the area of an acoustic tube section for which the area is known.

Wakita Formulation

There is another formulation for computing vocal tract areas from the acoustic speech wave. This formulation is due to Wakita (Wakita, 1973; Markel, *et al* 1976). A different set of assumptions is made as to the vocal tract glottal and lip terminations. These assumptions are outlined in Figure 40-3.

The glottal excitation is assumed to couple into the vocal tract as a forward going wave. All of the acoustic energy is assumed to leave the vocal tract as a backward going wave travelling down a semi-infinite tube. The characteristic impedance of this tube is the resistive impedance assumed by this model for the glottal termination.

The impedance at the lips is assumed to be an acoustic short. The model assumes that there is a coupling mechanism that produces an acoustic pressure in the far field in response to acoustic flow at the lips. Acoustic pressure at the lips is assumed to be zero.

Figure 40-3 gives a sequence of circuit models to relate the concepts in the Atal formulation to the Wakita formulation with its radically different boundary conditions. The Thevenin theorem is invoked to convert the glottal flow source-parallel impedance to a pressure source-series impedance. The reciprocity theorem is invoked to interchange the locations of the lip flow and the glottal pressure source. It is from the last circuit model that the energy balance argument can be invoked.

Power entering the vocal tract is equated to power leaving the vocal tract:

$$|U_g R_g|^2 \operatorname{Re} Y_i = |U_l|^2 R_g \tag{40-15}$$

$$\operatorname{Re} Y_i = \frac{1}{R_g} \frac{|U_l|^2}{|U_g|^2} \tag{40-16}$$

Taking the inverse Fourier transform gives:

$$h_e(t) = \frac{1}{R_g} R(t) \tag{40-17}$$

There are two differences with the Atal formulation. The impulse response h(t) is now the flow occurring at the lips that occurs in response to a lip pressure pulse applied to the

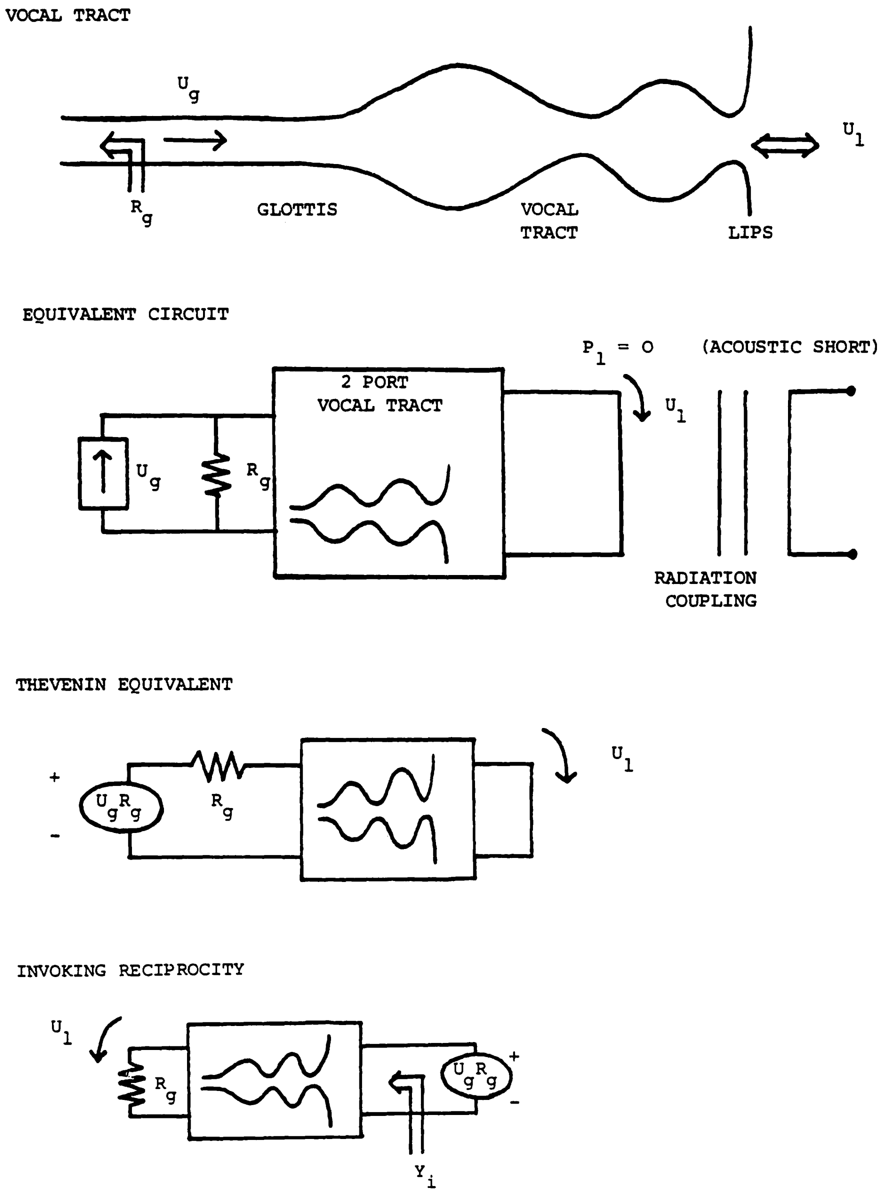


FIGURE 40-3
Wakita formulation vocal tract model.

vocal tract. The autocorrelation function R(t) is now the inverse transform of a glottal flow to lip flow transfer function. A different type of correction needs to be applied to the speech waveform autocorrelations to reflect the different vocal tract terminations in the Wakita model.

The Wakita formulation can also be expressed in discrete section form. The same reflection coefficients are computed using the Durbin recursion. The difference is that in the Wakita case the order of the reflection coefficients starts with the lips and works its way back toward the glottis.

In the Atal formulation, the reflection coefficients are determined in terms of the flow wave giving the formula:

$$r_i = \frac{S_{i+1} - S_i}{S_{i+1} + S_i} \tag{40-18}$$

while in the Wakita formulation, the reflection coefficients are determined in terms of the pressure wave giving the formula:

$$r_i = \frac{S_i - S_{i+1}}{S_{i+1} + S_i} \tag{40-19}$$

Starting from the same autocorrelation function R(t), the Wakita area function is obtained by taking the reciprocal of the Atal area function and then turning it end for end. As a result, the Atal and Wakita formulations give the same result when the log of area function is antisymmetric. For symmetric log area functions, the two formulations give opposite results.

The lack of agreement of the two area function formulations is one part of the problem. The underlying problem is that the proper correction to the speech autocorrelation function to obtain the autocorrelation function from which to compute areas is unknown. If correct yet differing corrections were applied to both the Atal and Wakita cases, both would give the correct area function—the Atal method computing areas from the glottis to the lips and the Wakita method computing the same areas the other way around.

Both area function formulations are incomplete because they require a transfer characteristic that is not measurable directly from the speech waveform. The next section describes a way of correcting this problem.

AREAS FROM TWO POINT ACOUSTICS

Figure 40-4 presents an acoustic model based on two point acoustic measurement. The basic philosophy is to take measurements that are physically practical and then to formulate an acoustic model based on them. Since pressure is convenient to measure, pressure measurements are used, one inside the vocal tract and one in the far field.

The glottis is represented by a flow source in parallel with an impedance which gives rise to an acoustic pressure at the input of the vocal tract. The particulars of glottal flow and glottal impedance do not enter the calculations because pressure is measured directly. Sound energy is conveyed losslessly by the vocal tract to the far field, where it is radiated.

Based on the lossless vocal tract assumption, the power entering the vocal tract is equated to the power leaving the vocal tract:

$$|P_g|^2 \operatorname{Re}Y_i = |P_f|^2 \frac{1}{R_f} \tag{40-20}$$

$$\operatorname{Re}Y_i = \frac{1}{R_f} \frac{|P_f|^2}{|P_g|^2} \tag{40-21}$$

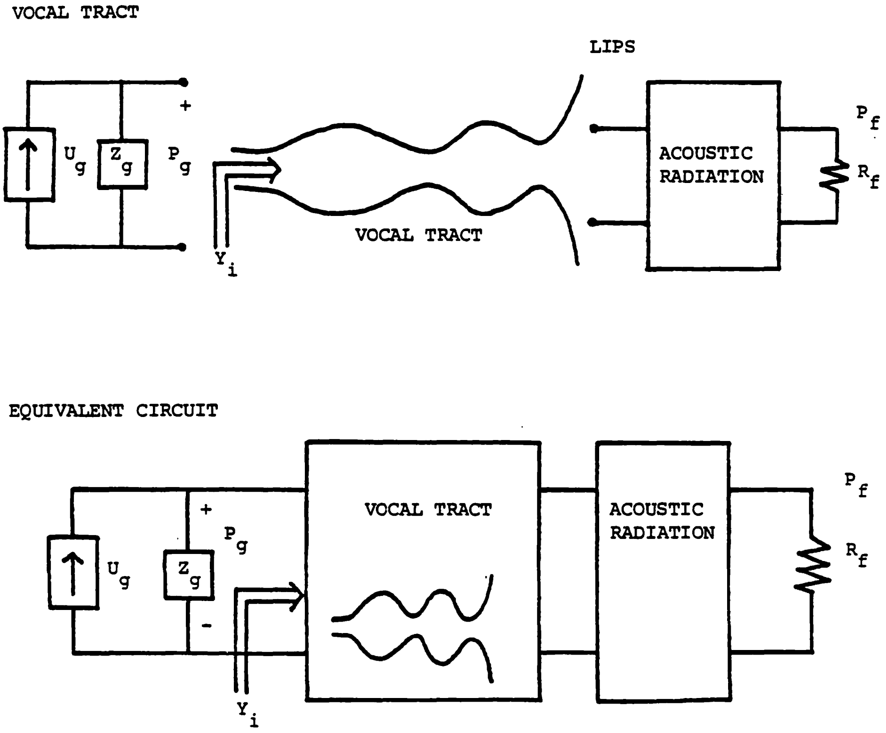


FIGURE 40-4
Vocal tract model for two point measurement.

Taking the inverse Fourier transform gives:

$$h_e(t) = \frac{1}{R_f} R(t) \tag{40-22}$$

The even part of h(t) is expressed as a proportionality with the autocorrelation function R(t). In this formulation, h(t) is a flow response to a pressure impulse as in the Wakita formulation.

The important consideration in the new formulation is that the required R(t) can be computed directly from the two pressure measurements. The Durbin recursion can be applied in the discrete case to compute reflection coefficients. Areas are computed starting with the glottis and working outward as in the Atal formulation. The areas are computed from the reflection coefficients, however, by Equation (40-19) of the Wakita formulation.

Once the energy balance arguments are understood for both the Atal and Wakita formulations, the new formulation is a natural extension. The new formulation, however, allows another item to be computed besides the area function. The transforms of glottal flow and glottal pressure and the vocal tract driving point admittance are related by:

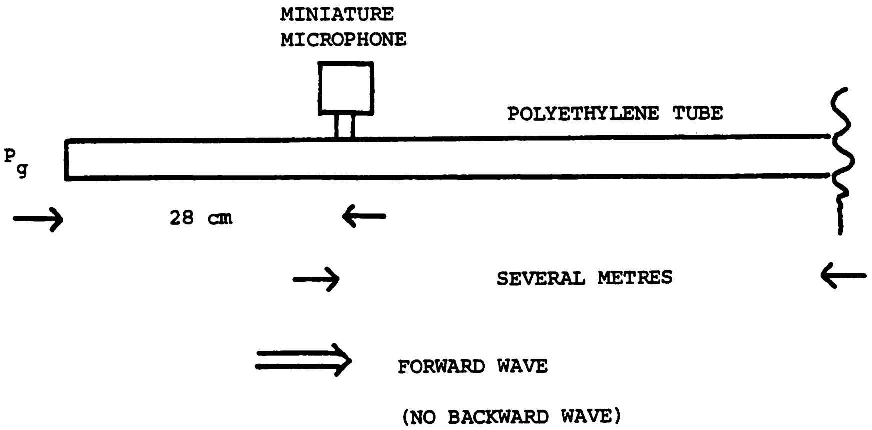


FIGURE 40-5
Reflectionless probe tube.

$$U_g = P_g Y_i \tag{40-23}$$

Taking the inverse Fourier transform gives:

$$u(t) = p(t) * h(t) \tag{40-24}$$

This formula indicates that by using the pressure measurement p(t) and the flow response h(t), one can compute the glottal flow. Pressure is measured by the intraoral microphone while the flow response is computed from the measured pressure transfer function by way of Equations (40-21) and (40-22).

The two point acoustic measurements offer a direct measure of vocal tract transfer function. It is possible to compute areas from this particular transfer function as an extension of the established techniques. It is also possible to compute glottal airflow using the vocal tract flow response that arises in these computations. The remaining question—how to measure intraoral acoustic pressure—is addressed in the next section:

INTRAORAL ACOUSTIC MEASUREMENT

The remaining issue is the measurement of acoustic pressure inside the oral cavity. There are two alternatives: placement of a miniaturized microphone at the point of measurement and remote placement of the microphone by way of a probe tube.

There are several factors which rule out placement of a microphone inside the oral cavity. There is a concern with safety. There is a risk, however remote, of aspirating this microphone, resulting in a serious medical emergency. There is the problem of keeping the microphone from being fouled by mucosal secretions. A risk of electric shock exists if microphone operating voltage appears on the microphone casing. Finally, the placement of the microphone may entail a high degree of discomfort.

A polyethylene probe tube has none of these problems. The tube is a single piece of high strength material, eliminating the aspiration hazard when properly employed by someone having clinical experience. There is no electric shock hazard as the tube is made of an elec-

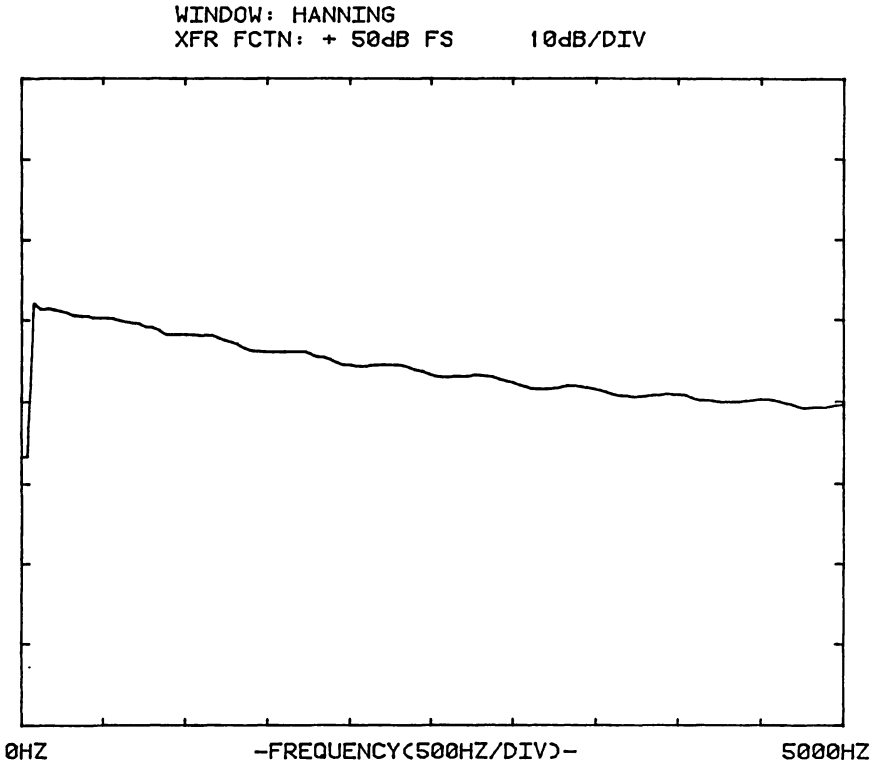


FIGURE 40-6
Measured transfer function of vented probe tube - microphone system.

trical insulator. Finally, the tube can be readily cleared of mucosal secretions by blowing air through it.

Most importantly, probe tubes are routinely placed nasopharyngeally in clinical work. A degree of discomfort is induced, but a patient is able to talk normally and fluently with a probe tube in place.

The usual clinical use of a probe tube is to measure pressure variation at low frequencies—to test the production of plosives. The area function estimation application requires pressure data in the acoustic frequency range. Most probe tube designs have several resonances in this frequency range. It would be possible to remove the effect of these resonances by properly filtering the recorded signal. This type of arrangement comes at the cost of signal-to-noise ratio.

A probe tube design which avoids resonances is shown in Figure 40-5. The basic principle is that a very long length of tubing—several meters—is used. Sound waves enter the probe tube and propagate in the forward direction. Because a narrow diameter tube is lossy, these waves dissipate before they reach the end. The probe tube appears to be semi-infinite because there are no waves travelling in the backward direction. With no backward waves there are no standing waves and no resonances.

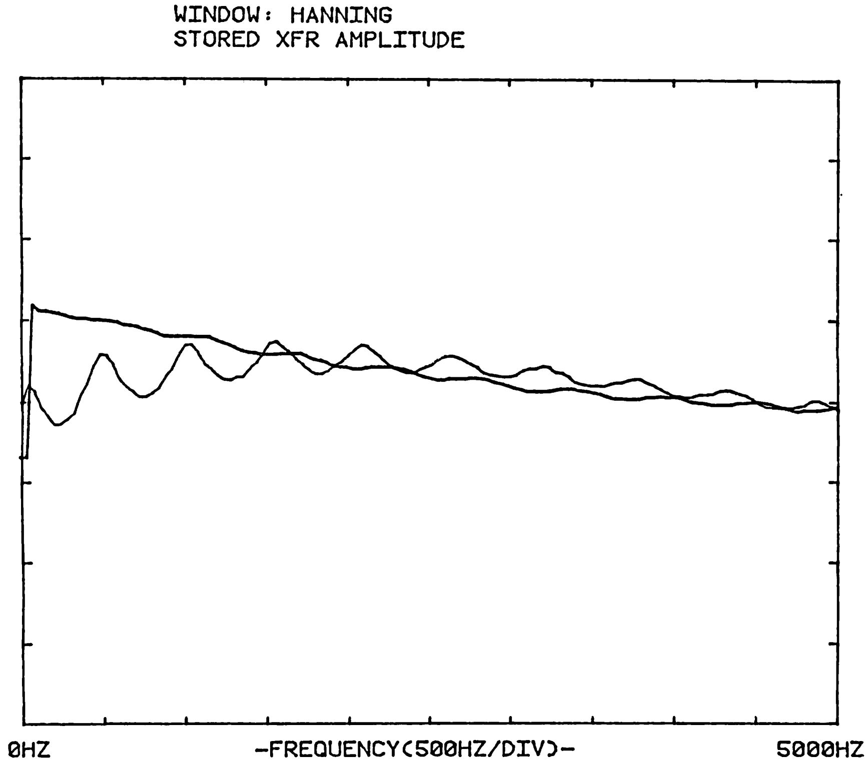


FIGURE 40-7
Measured transfer function of vented probe tube compared with unvented probe tube.

The pressure amplitude of the forward going waves is equal to the pressure at the entrance of the tube. This pressure is measured by tapping a miniature microphone into the side of the tube. A microphone should have a small diaphragm area to avoid perturbing the tube geometry.

Initial calibrations have been performed using tubing of 2.15 mm I.D. and by using a Knowles model BL-1785 ceramic microphone. The Knowles microphone was chosen for its unusual combination of wide bandwidth and small diaphragm area. The calibration technique is based on comparing the response of the probe tube system with a precision condenser microphone (Edmonds, *et al* 1971).

Figure 40-6 shows the frequency response curve of the probe tube system. Note the absence of resonances. There is roughly 10 dB attenuation at 5 KHz compared with the low frequencies. This attenuation is the result of the transmission line losses of the tube, losses that increase with frequency. The microphone was placed approximately 28 cm down the tube. Placing the microphone closer to the source or using a larger, less lossy tube would help.

Figure 40-7 shows a comparison of the "vented probe tube" with a probe tube where the microphone closes off one end. The same Knowles microphone was used in both cases. Note the introduction of resonances caused by the blind termination. These resonances would have been more severe if a microphone having a large dead volume were used, resulting in a Helmholtz resonator configuration.

SUMMARY

A technique of estimating the vocal tract area function from two acoustic measurements has been presented. The technique follows from considering how both the Atal and Wakita formulations for estimating areas can be derived from energy balance principles.

The philosophy behind the two point technique has been to find a transfer function that can be directly measured. The glottal pressure-far field pressure transfer function can be measured if intraoral pressure can be observed. A new method for estimating areas has been derived for the pressure transfer function case.

An added bonus of the two point measurement is that it is possible to extract glottal flow from the glottal pressure measurement. The inverse filter which does this is derived automatically in the course of estimating the areas.

The use of a probe tube is being advocated for obtaining intraoral acoustic pressure. The probe tube will result in some discomfort to subjects, but it is considered safe. The probe tube will permit a subject to speak normally and fluently, something not possible with impedance tube based area function estimation.

Data has been presented on a "vented probe tube" acoustic measurement apparatus. This data suggests that it is practical to measure intraoral air pressure in the acoustic frequency range.

REFERENCES

Atal, B.S. (1970). "Determination of the vocal tract shape directly from the speech waveform," *J. Acoust. Soc. Am.* 47, 65A.

Atal, B.S. and Hanauer, S.L. (1971). "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Am.* 50/2/2, 637-655.

Edmonds, T.D., Lilly, D.J. and Hardy, J.C. (1971). "Dynamic characteristics of air-pressure measuring systems used in speech research," *J. Acoust. Soc. Am.* 50/2/2, 1051-1057.

Makhoul, J. (1975). "Linear prediction: a tutorial review," *Proc. of the IEEE* 63, 561-580.

Markel, J.D. and Gray, A.H., Jr. (1976). *Linear Prediction of Speech*. Springer-Verlag, New York.

Milenkovic, P. (1981). "A systematic assessment of the accuracy of vocal tract area function estimates made from the speech waveform," unpublished doctoral diss., Caltech, Pasadena, CA.

Sondhi, M.M. (1979). "Estimation of vocal-tract areas: the need for acoustic measurements," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, ASSP-27/3, 268-273.

Sondhi, M.M. and Gopinath, B. (1971). "Determination of vocal tract shape from impulse response at the lips," *J. Acoust. Soc. Am.* 49, 1967-1983.

Wakita, H. (1973). "Direct estimation of the vocal-tract shape by inverse filtering of acoustic waveforms," *IEEE Trans. Audio and Electroacoustics*, AU-21, 417-427.

DISCUSSION

A. BENADE: These long probe tubes are something we have had fun with as well. You have to be careful because they are a little elliptical, so if there is any sound field along the length of them, they will absorb sound in a most unpleasant way from their entire length.

P. MILENKOVIC: Are you referring to the intraoral part of the tube or the part outside the vocal tract?
A. BENADE: Any part. These tubes receive everything. They have an enormous surface area.
P. MILENKOVIC: I see. You are saying that the fact that there is a long length of tube placed intraorally will result in microphonic pick- up of sound along the length of the vocal tract, even if the extraoral portion is acoustically insulated.
A. BENADE: The point is, whatever piece of the tube you consider, it has a huge area compared to the cross-section of the open end. It is possible to have an interference to signal ratio of over 10 dB if one is not careful.