NCVS Status and Progress Report
Volume 3/December 1992

The National Center for Voice and Speech is a consortium of institutions--The University of Iowa, The Denver Center for the Performing Arts, The University of Wisconsin-Madison and The University of Utah--whose investigators are dedicated to the rehabilitation, enhancement and protection of voice and speech.

Editorial and Distribution Information

Editor, Ingo Titze
Production Editor, Julie Ostrem
Technical Editor, Martin Milder
Editorial Assistant, Julie Lemke

Distribution of this report is not restricted.
However, production was limited to 1,000 copies.

Correspondence should be addressed as follows:
Editor, NCVS Status and Progress Report
The University of Iowa
330 Wendell Johnson Building
Iowa City, Iowa 52242
(319) 335-6600
FAX (319) 335-8851
e-mail titze@shc.uiowa.edu

Primary Sponsorship

The National Institute on Deafness and Other Communication Disorders,
Grant Number P60 DC00976

Other Sponsorship

The University of Iowa
Department of Speech Pathology and Audiology
Department of Otolaryngology - Head and Neck Surgery
Department of Preventive Medicine and Environmental Health

The Denver Center for the Performing Arts
Recording and Research Center
Department of Public Relations
Department of Public Affairs
Denver Center Media
Department of Development

The University of Wisconsin-Madison
Department of Communicative Disorders
Department of Surgery, Division of Otolaryngology
Waisman Center
Department of Electrical and Computer Engineering

The University of Utah
Department of Otolaryngology - Head and Neck Surgery

Memphis State University
Department of Audiology and Speech Pathology

The University of Illinois
Department of Speech and Hearing Science

NCVS Personnel

Administration

Central Office
Ingo Titze, Director
John Folkins, Deputy Director
Julie Ostrem, Program Assistant
Julie Lemke, Secretary

Area Coordinators
Research - Ingo Titze
Training - John Folkins
Continuing Education - Diane Bless
Dissemination - Ronald Scherer
Core Resources - Erich Luschei

Institutional Coordinators
The University of Iowa - Ingo Titze
The Denver Center for the Performing Arts - Ronald Scherer
University of Wisconsin-Madison - Diane Bless
University of Utah - Steven Gray

Investigators, Affiliates and Support Staff

Fariborz Alipour, Ph.D.
Barbara Bustillos
Patricia Benjamin, B.S.
Florence Blager, Ph.D.
Diane Bless, Ph.D.
James Brandenburg, M.D.
Larry Brown, D.M.A.
Kate Bunton, B.S.
Geron Coale, M.A.
Stefanie Countryman, M.A.
Heather Dove, M.A.
Kate Emerich, B.M.
Penny Farrell
John Folkins, Ph.D.
Charles Ford, M.D.
Steven Gray, M.D.
Judith Grayhack, Ph.D.
Chwen-Geng Guo, M.S.
Susan Hensley, M.A.
Marilyn A. Hetzel, Ph.D.
Margaret Hoehn, M.D.
Henry Hoffman, M.D.
Yoshiyuki Horii, Ph.D.
Joel Kahane, Ph.D.
Judith King, Ph.D.
David Kuehn, Ph.D.
Jon Lemke, Ph.D.
Julie Lemke
Erich Luschei, Ph.D.
Brian McCabe, M.D.
Deb Milbauer
Martin Milder, B.S.
Paul Milenkovic, Ph.D.
Kenneth Moll, Ph.D.
Hughlett Morris, Ph.D.
Jerald Moon, Ph.D.
John Nichols, B.A.
Lorraine Olson Ramig, Ph.D.
Julie Ostrem, B.S.
Mark Peters
Donald Robin, Ph.D.
Ronald Scherer, Ph.D.
Elaine Smith, Ph.D.
Marshall Smith, M.D.
Nancy Pearl Solomon, Ph.D.
Molly Strittmater, B.A.
Leatitia Thompson, Ph.D.
Ingo Titze, Ph.D.
Nancy Tye-Murray, Ph.D.
Vern Vail, B.S.
Katherine Verdolini-Marston, Ph.D.
William Winholtz, A.A.S.
Darrell Wong, Ph.D.
Raymond Wood, M.D.
Patricia Zebrowski, Ph.D.

Doctoral Students

Julie Barkmeier, M.S.
Todd Brennan, M.S.
David Druker, B.S.
Eileen Finnegan, M.A.
Mark Leddy, M.S.
Bruce Poburka, M.S.
Annie Ramos, M.S.
Eileen Savelkoul, M.S.
Alice Smith, M.A.
Ina Steinecke
Brad Story, B.S.
Kenneth Tom, M.A.

Postdoctoral Fellows

David Berry, Ph.D.
Young Min, M.D.

Visiting Scholars

Hanspeter Herzel, Ph.D., Germany
Meijin Nakayama, M.D., Japan
Tzu-Yu Hsiao, M.D., Taiwan

Advisory Board

Katherine Harris, Ph.D.
Minoru Hirano, M.D.
Clarence Sasaki, M.D.
Johan Sundberg, Ph.D.

Contents

Research papers submitted for peer review in archival journals

Forward

Greetings to all NCVS investigators and readers from other laboratories. This is the third progress report to our principal sponsor, the National Institute on Deafness and Other Communication Disorders. I am happy to say that the concept of a "center without walls" is working well. There is plenty of evidence of fruitful collaboration between the four sites. We just finished a consensus conference in Denver, the results of which have led to a publication and several new investigations. We appreciate the input from Madison, Utah, Denver and Iowa.

No updates on dissemination, training, and continuing education are included in this report. It came too quickly after the previous one. In the future, we will update these areas only once a year. On the other hand, research papers are increasing in such numbers that we may need to expand to quarterly issues. I hope that is true.

Ingo R. Titze, Director

December 1992

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 3
December 1992, 1-4

Morphologic Ultrastructure of Anchoring Fibers in Normal Vocal Fold Basement Membrane Zone

Steven D. Gray, M.D.
Department of Otolaryngology - Head and Neck Surgery, The University of Utah
Shirley S.N. Pignatari, M.D.
Department of Otolaryngology, Escola Paulista de Medicina, Sao Paulo, Brazil
Penny Harding
Department of Otolaryngology - Head and Neck Surgery, The University of Iowa

Abstract

Anchoring fibers are collagenous structures which help secure basal cells to the superficial layer of the lamina propria (SLLP). Transmission electron microscope (TEM) was used to study the morphology and organization of these anchoring fibers at the human vocal fold basement membrane zone. This study shows that anchoring fibers loop from the lamina densa through the lamina propria and reattach to the lamina densa. Collagen fibers from the lamina propria appear to pass through the loops created by the anchoring fibers. This relationship partially explains how the epidermis secures itself to the SLLP during vibration resulting from phonation. The population density of anchoring fibers and genetics is discussed.

Introduction

Anchoring fibers are collagenous structures composed mainly of type VII collagen (Sakai 1986) and are found in the basement membrane zone (BMZ). They have been described in oral mucosa (Susi 1967) and cervical epithelium (Laguens 1972), but the majority of the investigations have concerned the characterization of anchoring fibers in skin (Bentz 1983, Sakai 1986, Briggaman 1975, Tidman 1985). Katz (1984) described these fibers as originating in the lamina densa and looping to the dermis. They are not as large, as thick, and may not be as strong as the collagen fibers normally found in the lamina propria (Sakai 1986).

It is believed that anchoring fibers are one of the structures responsible for securing the lamina densa to the dermis (Tidman 1984, Tidman 1985). Several studies have demonstrated an

important role of these anchoring fibers in blistering diseases such as epidermolysis bullosa dystrophica recessive, which is characterized by an epidermal-dermal adherence failure (Briggaman 1975). A reduction or absence of the population density of anchoring fibers has been described in these diseases as the only consistent ultrastructural alteration found (Briggaman 1975, Tidman 1985).

To our knowledge no one has described anchoring fibers in the normal human vocal fold. Since the vocal fold has an epidermal-lamina propria interface, the anchoring fibers may have the function of attachment of the structures from the epidermis to the lamina propria.

The purpose of this study is to demonstrate the architecture of anchoring fibers, and morphologically characterize these structures in the normal human vocal fold basement membrane zone.

Materials and Methods

Patients

Three larynges were examined in this study.

Two larynges were autopsy tissue removed from 2 males, ages 60 and 51, who died from nonlaryngeal related diseases. The remaining one was removed surgically due to chronic aspiration (neurologic origin) from a 72 year old female patient. The fresh tissue was immediately immersed in glutaraldehyde solution, and the autopsy tissues were immersed within 10 hours postmortem.

Electron Microscopy (tissue processing)

The tissues were prefixed in 2.5% Glutaraldehyde/0.1 M Sodium Cacodylate buffer (pH 7.2) 1.5% Potassium Ferrocyanide. Acetone was used for dehydrating and finally, the tissues were embedded in spears resin. Ultrathin sections were cut from randomly selected blocks, stained with uranylacetate and lead citrate, and examined and photographed with Hitachi Electron Microscope (H-600).

Ultrastructural Findings

Electron microscopy (TEM) revealed the basement membrane zone in these normal human vocal cords to be similar to those described in our previous study in canine vocal cord (Gray 1988 1989). The deep layer of the epidermis is lined with basal cells laying on a basement membrane. In Figure 1 (see center-bound plate), TEM shows a normal epidermal-lamina propria junction of the human vocal fold. Notice the collagen fibers in the lamina propria. Beneath the basement membrane zone, the lamina propria is populated with collagen and elastic fibers. Some of these thicker collagen fibers are seen intertwining with these anchoring loops, forming a woven network of fibers (Figure 1).

The anchoring fibers can be seen looping from the lamina densa into the superficial layer of the lamina propria (SLLP) and reattaching to the lamina densa in Figure 2 (see center-bound plate).

Speculation exists regarding the mechanism of attachment between the lamina densa, anchoring fibers and the thicker collagen fibers of the SLLP. From these photographs, it appears that the SLLP collagen fibers actually loop through or wrap around the loop made by the anchoring fibers. This looping arrangement of the anchoring fibers likely provides increased structural integrity of this delicate tissue transition and interface known as the BMZ. It actually appears that two structures identified with TEM may pass through these loops: other anchoring fibers (characterized by

aperiodic banding) creating more loops, and collagen fibers (periodic banding) from the lamina propria. This creates a chain-like fence arrangement which allows the tissue to glide, bend, or compress while still maintaining structural integrity.

Discussion

The idea that anchoring fibers may be one of the principal agents responsible for securing the lamina densa to the dermis has been well demonstrated in skin BMZ, especially regarding blistering diseases (Briggaman 1975, Goldsmith 1983, Tidman 1984). However, the human vocal folds represent an epidermal-lamina propria interface, not an epidermal-dermal interface as is found in skin. This study demonstrates that the human vocal folds have a similar morphologic ultrastructural pattern as in the epidermal-dermal area.

Because the vocal cords need to withstand great vibratory and shearing stress, these anchoring-attaching structures may have an important role in the sequelae development of vocal fold nodules. Our preliminary study of nodules of human vocal cords (Gray 1989), has shown important disarrangement of the BMZ where the overall thickness appeared markedly increased. The anchoring fibers were oriented in all directions and not always linking the lamina densa to the lamina propria. In addition, some of them seemed partially formed and rudimentary, incapable of forming any secure attachment.

Tidman (1984) has related that the population density of anchoring fibers in the skin varies between body regions. They also counted the number of anchoring fibers an individual would have for a given area of the skin, and found that individual variation was considerable. We have attempted to do population density counts of anchoring fibers in the human vocal folds, by using morphometric techniques to count random areas along the BMZ. Unfortunately, we have not been able to get consistent results allowing the confidence needed to state a normal range of anchoring fiber population density. The inconsistent results are due to marked variation in anchoring fiber population density in between individuals and also location in the vocal fold. It appears that the number of anchoring fibers varies depending on the location in the fold, both from an anterior to posterior dimension as well as superior to inferior. Generally, the population density of anchoring fibers seems greatest in the mid-membranous fold area. This finding is consistent with Tidman (1984) showing that the population density of anchoring fibers is greatest in body regions where skin is subjected to shearing and stress motion (elbows, knees).

Summary

The population density of anchoring fibers per unit area is individually genetically determined; however, within each individual the population density varies depending on body site. (Briggaman 1975, Tidman 1985). Environmental influences have not been investigated. The effect of disease or trauma is not known. Further research in this area will help to understand the mechanisms by which pathologic reactions and sequelae originate from vocal cord injury. Are some individuals anatomically predisposed toward vocal injury since they have fewer attaching structures? How does the concept of genetic determination of anchoring fibers relate to other vocal fold structures and proteins? Are other vocal fold structures genetically determined and show similar individual variation? Are vocal abilities and talents in part genetically determined?

It is true that anchoring fibers represent a tiny piece of vocal fold anatomy and this study does not intend to exaggerate their significance. This study partially describes the anatomical

arrangement and attachment of the epidermis to the lamina propria. It was our hope that this report would prompt interest and discussion in individual variations of vocal fold architecture and composition which may have a genetic basis. The implications of that concept to those desiring a profession requiring vocal athletics is intriguing.

Acknowledgement

This work was supported by NIH grant K08 DC00036-05.

References

1.Sakai LY, Keene DR, Morris NP, Burgeson RE. Type VII collagen is a major structural component of anchoring fibrils. *L Cell Biol* 1986;103:1577-1586.

2.Susi FR, Belt WD, Kelly JW. Fine structure of fibrillar complexes associated with the basement membrane in human oral mucosa. *Brief Notes* 1967;686-690.

3.Laguens R. Subepithelial fibrils associated with the basement membrane of human cervical epithelium. *J Ultrastruct Res*, 1972;54:202-208.

4.Bentz H, Morris NP, Murray LW, Sakay LY, Hollister DW, Burgeson RE. Isolation and partial characterization of a new human collagen with an extended triple-helical structural domain. *Proc Natl Acad Sci USA* 1983;80:3168-3172.

5.Briggaman RA, Wheeler CE. Epidermolysis bullosa dystrophica recessive: A possible role of anchoring fibrils in the pathogenesis. *J Invest Dermatol* 1975;65:203-211.

6.Tidman MJ, Eady RAJ. Evaluation of anchoring fibrils and other components of the dermal-epidermal junction in dystrophic epidermolysis bullosa by a quantitative ultrastructural technique. *J Invest Dermatol* 1985;84:374-377.

7.Katz SI. The epidermal basement membrane zone structure, ontogeny and role in disease. *J Am Acad Dermatol* 1984;11:1025-1037.

8.Tidman MJ, Eady RAJ. Ultrastructural morphometry of normal human dermal-epidermal junction. The influence of age, sex, and body region on laminar and nonlaminar components. *J Inves Dermatol* 1984;83:448-453.

9.Gray S, Titze I. Histologic investigation of hyperphonated canine vocal cords. *Annals Otol Rhinol Otolaryngol* 1988;97:381-388.

10.Hirano M, Kurita S. Histological Structure of the Vocal Fold and Its Norman and Pathological Variations. In: Kirchner JA, ed. *Vocal Fold Histopathology: A Symposium.* College Hill Press, 1986:17-24.

11.Gray S, Titze I, Lusk RP. Electron microscopy of hyperphonated canine vocal cords. *J of Voice* 1987;1:109-115.

12.Gray S. Basement Membrane Zone Injury in Vocal Nodules. In: Gauffin J, Hammarberg B, eds. *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms.* Singular Publishing Group, Inc, 1989:21-27.

13.Biondi S, Biondi-Zappala M. Surface of laryngeal mucosa seen through the scanning electron microscope. *Folia Phoniatr (Basal)* 1974;26:241-248.

14.Tillman B, Pietzsch-Rohrschneider I, Huenges HL. The human vocal cord surface. *Cell Tissue Res* 1977;185:279-283.

15.Goldsmith LA, Briggaman RA. Monoclonal antibodies to anchoring fibrils for the diagnosis of epidermolysis bullosa. *J Invest Dermatol* 1983;81:464-466.

NCVS Status and Progress Report - 3
December 1992, 5-20

A Methodological Study of Hemilaryngeal Phonation

Jack Jiaqi Jiang, M.D., Ph.D.
Department of Otolaryngology/Head and Neck Surgery, Northwestern University
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

An excised hemilarynx setup was developed. The phonatory characteristics of nine excised canine larynges were examined. The left vocal fold of each larynx was then removed and substituted with a vertical plexiglass plate. The larynges were phonated again. Recordings were made of phonation threshold pressure, sound pressure level, average glottal flow, fundamental frequency, and amplitude of vocal fold vibration as observed with a video stroboscope. Measurements were made over a range of subglottal pressures. For the hemilarynx, simultaneous recordings of intraglottal pressure and vocal fold contact area were also made. It was found that amplitude and frequency of vocal fold vibration of the hemilarynx, as well as rates of change of amplitude and frequency as a function of subglottal pressure, were similar to those of the full larynx. Also similar were phonation threshold pressures and ranges of subglottal pressure over which the larynges phonated. The average airflow of the hemilarynx was about half that of the full larynx, and the sound pressure level was one fourth (about 6 dB less) in the hemilarynx, under similar conditions.

Introduction

Vertical hemilaryngectomy has been practiced widely for more than 40 years, ever since Alonso[1] developed the procedure to a practical level. Some patients are able to produce relatively normal phonation after one vocal fold has been removed. However, there are few kinematic or aerodynamic data available to assess the vibration of a single vocal fold against alternate structures. As is well known, a person with a single eye or a single ear can function relatively well, with some limitations. Although much research has been conducted on laryngeal phonation in general, little attention has been paid to single vocal fold phonation.

With recent developments in basic research and clinical measurements of vocal function, parameters such as vocal fold contact area and contact stress are of great interest because of their direct physiologic and clinical relevance[2]. The hemilarynx technique allows vocal fold contact area

profiles and contact stress to be measured directly during phonation. Experimentally, the procedure involves removing one vocal fold and replacing it with a glass or plexiglass prism containing either an electronic conductive surface or pressure transducers. However, before using the hemilarynx as a research tool to study phonation, the similarities and differences between the hemilarynx and full larynx need to be established. A better understanding of the physiology of single vocal fold phonation may then help clinicians improve surgical procedures, evaluate surgical results, and thereby achieve better phonation than is currently obtained. Additionally, from a research point of view, it is advantageous to observe the self-oscillating characteristics of a single fold because the second fold tends to obscure the lateral view and prevent observation of the vertical movements of the tissue.

The objective of this study was to develop a methodology for comparison of hemilaryngeal phonation with full larynx phonation. The hypothesis is advanced that, except for a 2:1 scale factor for airflow, the hemilarynx behaves similarly to the full larynx. Phonation threshold pressures, transglottal pressure and flow ranges, fundamental frequency, and amplitude of vibration are all hypothesized to be the same.

Background on Excised Full Larynx and Hemilarynx Experimentation

Research into laryngeal physiology with excised larynges has been systematically reviewed by Cooper[3,4]. According to this author, the first known experiments on voice production using an excised larynx preparation were conducted by Leonardo da Vinci (1452-1519). By introducing air into the lung and narrowing or widening the "fistula" at the exit of the windpipe, Leonardo obtained phonation. In 1741, over 200 years after Leonardo's work, the French physician and anatomist Antoine Ferrein (1693-1769) presented his studies on phonation of excised larynges[5,6]. Ferrein carried out extensive experimentation on excised human larynges, as well as those of dogs, bulls, pigs, and sheep. He used his own lungs as the air source. He found that vocal intensity was controlled by glottal width and air velocity. Fundamental frequency was found to be primarily controlled by vocal fold tension and possibly also by contraction of the vocal folds.

There were new studies on phonatory physiology during the 19th century, including judgments of phonatory intensity and frequency, by Liskovius[7,8] and Lehfeldt[9]. Most notably, however, Johannes Muller[10,11] observed the relationship between vocal fold tension, subglottal pressure, and fundamental frequency. The apparatus used by J. Muller was very similar to the modern apparatus of van den Berg and Tan[12], except that Muller used his own lung as the air source. Muller's apparatus simulated the contraction of all intrinsic laryngeal muscles isotonically or isometrically, except for the vocalis muscle.

In this century, van den Berg and colleagues[13,12,14] independently controlled airflow rate and configuration of the laryngeal cartilages in order to examine their effect on phonation, using excised human larynges. Van den Berg found that excised larynges could produce all the vocal registers. Longitudinal tension of the vocal folds was primarily responsible for determining vocal register and for controlling fundamental frequency within registers. Vertical phase differences in vibrations of the vocal fold were noted.

A similar experiment was also reported by Anthony[15]. In chest voice, fundamental frequency varied with subglottal pressure at a rate of 71.4 Hz/kPa, leaving vocal fold elongation constant. Excised larynx experiments were also reported by Hiroto[16]. Hiroto observed a wave-like motion of the vocal fold mucosa and proposed the mucovisco-elastic-aerodynamic theory of phonation.

Using an excised canine larynx, Baer[17] reported quantitative measurements of detailed mechanical vibration patterns. Small particles were attached to the vocal folds to serve as landmarks. The phonating excised larynges were observed stroboscopically to produce apparently stopped (or slow-motion) states, from both supraglottal and subglottal aspects. Measurements were made of the trajectories of the particles. Regulation of phonation was also observed. Baer found that points along the superior lip of the fold vibrate vertically while points in the lower margins move mostly horizontally.

More recently, Durham, Scherer, Druker and Titze[18] reported on a hemilarynx procedure. Sustained phonation was achieved in their experiment and the movement of the vocal fold was observed from both the top and the side. The study was methodolical only; therefore, no quantitative results were reported for sound quality, aerodynamics, and acoustics. Scherer, Druker and Titze[19] used an excised canine hemilarynx technique to measure vocal fold contact area and to compare it with the electroglottographic signal. In the Scherer et al. setup, the vocal fold was not self-oscillating, but driven by an electronic positioning arm[19].

Method

Larynges

Nine larynges were harvested 15 minutes post mortem from large (25-30 kg) mongrel dogs. Every larynx was used twice, once for full larynx vibration and then again for hemilarynx vibration. The larynxes came from experimental animals from several coronary research units at The University of Iowa. None of the animals had been intubated with endotracheal tubes. Most of them had been used for experimental procedures involving surgical impairment of coronary circulation for a duration of less than 3 hours. The coronary units were not always able to supply dependable age or sex information for the animals; therefore, larynges from either sex were used and age was not recorded. Sex does not appear to affect the range and quality of phonation for canines[18].

The length of the vocal folds in all the larynges was greater than 16 mm. Larger larynges were chosen for the following reasons: (1) A large larynx usually has a thicker vocal fold, which phonates better because of a lower phonation threshold pressure[19]. (2) It was easier to attached transducers and to make configurational adjustments on larger larynges.

Preparation of the hemilarynx

The following dissection procedure was used to avoid tissue damage. From a laryngofissure approach, the skin was cut and surrounding muscles were stretched. The trachea was identified from its easily recognizable ring structure, and dissected out. The trachea was cut 4-5 cm below the cricoid cartilage. The cutting end of the trachea was held by a 3-teeth traction forceps and retracted anteriorly and upward. The thyrohyoid and the sternothyroid muscle were cut. Another cut was made between the thyrohyoid membrane to disconnect the attached tissue, and the whole larynx was pulled out. Finally, the thyrohyoid ligaments (only 1-2 mm long in canines) and the rest of attached mucosa was cut so that the larynx was totally free from the body. The false folds were usually left attached to the free larynx until phonation, because they prevent the true vocal fold from drying out or being damaged during mounting of the excised larynx.

At least 4 cm of the trachea was left attached to the larynx for later ease of mounting. The larynx and the surrounding extrinsic muscles remained attached from the inferior edge of the hyoid bone to the cricoid cartilage.

After excision, the larynx was placed in a 0.67% saline solution[12] and stored in a refrigerator set at 40°F. The larynx was typically used in an experiment 12 to 36 hours after excision.

As described by Durham et al.[19], the remaining muscles were trimmed away from the thyroid cartilage and prepared for mounting. The cuneiform cartilages were removed. The epiglottis was removed at the level of the thyroid notch by cutting through the aryepiglottic folds. Large surgical scissors were used to cut horizontally around and through the thyroid cartilage about 5 mm above the level of the true vocal folds. The cut removes the false vocal folds and exposes the true vocal folds.

In preparing a hemilarynx immediately after a full larynx experiment, the left vocal fold, the left arytenoid cartilage, and the left half of the thyroid cartilage were removed (Figure 3; see center-bound plate) to facilitate mounting of the vertical plate. A piece of plexiglass, 9 mm thick, was carefully ground to a shape that exactly fit the left side of the subglottal space and the space vacated by the removed vocal fold, so that the remaining vocal fold achieved closure with the side of the vertical plate when the arytenoid cartilage was pushed (Figure 4 on plate). The gap between the cricoid ring and the curved side of the plate was sealed by tightly clamping a curved steel strip (1 mm thick) with a stainless steel screw driven through a hole on both the cricoid cartilage and the plexiglass plate. Occasionally, a clay-like gum was also used to prevent air from escaping.

The arytenoid cartilage was positioned for phonation with a 3-pronged device (Figure 4 on plate, middle-right of photograph). The 3-pronged device was attached to micrometers so that the vocal fold could be adducted by tightening the micrometer. A wood shim of a selected size was positioned between the arytenoid cartilage and the plexiglass plate (or between the two cartilages for full-larynx operation) to change the glottal configuration.

Another micrometer system was attached via a rod (upper-right of Figure 4 on plate) to the anterior tip of the thyroid lamina (by stitching). Elongation of the vocal fold was controlled by turning this micrometer. The length of the vocal fold was first set to the *in situ* length by the micrometer adjustment. This *in situ* length served as the reference length. Recorded micrometer values were used for subsequent positive or negative changes in vocal fold length.

To facilitate observation of vocal fold movement, a black 8-0 ophthalmic nylon suture with needle was used to stitch marks on the upper lip of the membranous portion of the remaining vocal fold (three tiny marks are barely visible on the vocal fold edge in Figure 4 on plate). The needle was pushed though the superficial layer of the vocal fold at a depth of about 0.5 mm, so that the stitch could be tightly attached to the vocal fold cover. This allowed visualization of a surface point without significant distortion and interference with vocal fold vibration.

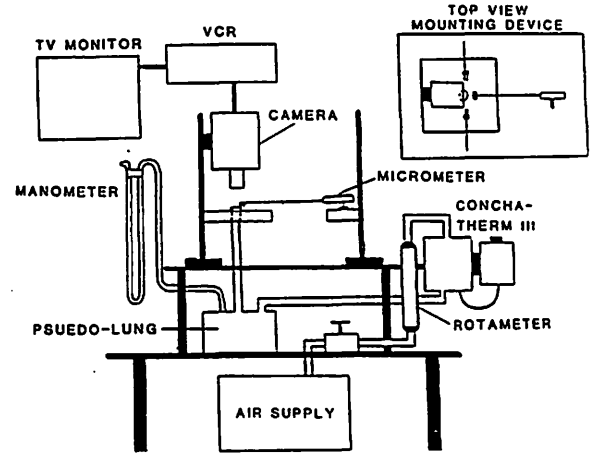


Figure 1. Illustration of the excised larynx apparatus (Durham et al. 1987, by permission).

The apparatus, larynx mounting, and control

The entire apparatus used for experimentation on a full larynx was described in detail by Baer[17] and Durham et al.[18]. Figure 1 illustrates the excised larynx apparatus. An Ingersoll-Rand (Type 30)

conventional air compressor was used to generate the airflow. Pressure was reduced to 15 psi using a Packer-Hannifin 49-21641 compressed air regulator. The amount of airflow used during the experiment was controlled by a valve. The flow was measured as it passed through a Gilmont rotameter-type flowmeter (J197). The input air was conditioned to 35-38°C and 95-100% relative humidity by two ConchaTherm III heater-humidifiers (Respiratory Care, Inc.) placed in series. It took about 15 minutes for the Conchatherms to heat and humidify the air to the above ranges. A Cole-Parmer Digi-Sense Thermocouple Thermometer (Model 8529-00) and a humidity probe (Check-It Electronics 424) were used to monitor the humidity 2 cm above the level of the vocal folds.

A plastic container, with a volume of approximately 5000 ml, was used as a damped acoustic resonator (a pseudo-lung) and a condensed water reservoir. This container was lined with soft urethane foam. It simulated the approximate open-end termination for acoustic waves propagating in a section of 3/4" copper plumbing pipe mounted vertically on this container. The larynx was connected with a section of trachea on the top of the pipe. The length of the pipe was 15 cm.

Subglottic pressure, vocal fold elongation, and prephonatory configuration were varied independently as control variables. Glottal flow, fundamental frequency, sound pressure level, and vibrational amplitude were recorded and analyzed for both full larynx and hemilarynx. The assumption of the functional similarity between the full larynx and the hemilarynx was tested with these variables.

Instrumentation, calibration, and data recording

An RCA color video camera (Model TC 5001) was used to record the glottal image from above. The image was viewed on a Panasonic 12-inch black and white monitor. The camera was mounted 15-20 cm from the larynx with the view angle normal to the glottis. The vocal folds were magnified 15 to 20 times when viewed on the monitor. In addition to the top view video recording, a second video camera (Sony DXC-102) with a 90 mm microlens was mounted at the left side of the larynx, perpendicular to the vertical plexiglass plate, so that the contact area image could be recorded and viewed on another Panasonic 12-inch black and white monitor.

A special video effect generator (Panasonic VY-922) was used to post both superior and lateral images on the same screen. The split image was refolded on a Panasonic PV1560 VHS VCR. On playback, the split image was viewed on a color Sony Trinitron 25-inch monitor. The edge-to-edge distortion of the TV screen measured less than 5%. The duration of each frame was about 33.3 ms. The speed of the video camera was 1/1000 second per frame. Before fixing the camera adjustments in each experiment, a piece of scale paper was placed on the vocal fold for magnification calibration.

A strobe light (Pioneer DS330-ST) was placed 10 cm from the glottis, at an angle of 45° from the vertical. The frequency of the strobe light could be adjusted manually to any value below 1000 Hz, or it could be triggered automatically by an outside electronic device. In order to obtain still video images, an electronic timing delay device was designed and built for triggering the strobe light. The trigger signal was a low-passed version (Wavetek-Rockland 432, 300Hz cutoff) of the acoustic signal. The stroboscope could thus be triggered at controlled adjustable timing delays. Stable images could be locked in at any phase of the phonation cycle. The flash firing, detected by a photo diode, was recorded simultaneously on the audio channel of the VCR so that corresponding phase information about the locked still video image could be obtained.

The acoustic signal was recorded using a Sennheiser MD441-U microphone, positioned 15 cm from the glottis. The axis of the microphone was at 45° from the long axis of the larynx in order to avoid airflow impinging directly onto the microphone. The signal picked up by the microphone was preamplified, then recorded by the left audio channel of the VCR. It was also monitored and analyzed on one channel of a DATA 6000 universal signal analyzer. Alternately, the signal was A/D converted and processed on a VAX 3200 computer.

Mean subglottal pressure in the pseudolung was measured with an open-ended water manometer (Dwyer No. 1211). The pressure was read out orally and recorded on the right audio channel of the VCR, which was being used as a chatter channel.

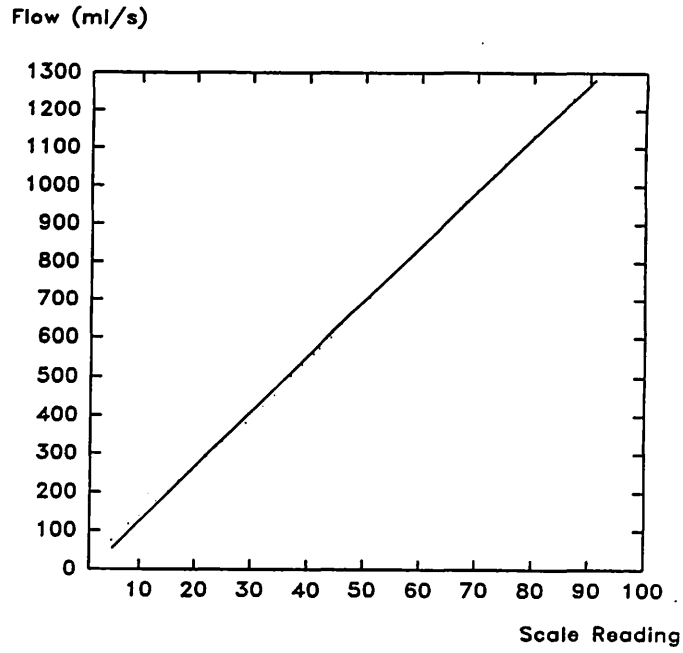


Figure 2. Calibration chart for Gilmont J197 flowmeter. The flowmeter is linear in the range of 30 ml/sec. to 1200 ml/s.

The microphone head of the sound level meter (Quest 215) was positioned 15 cm from the glottis and angled at 45°. The C scale was used, which filters out frequency components below 50 Hz. This low pass filtering decreases the artifact caused by low frequency air currents in the recording room. The sound pressure level was read out orally and recorded on the chatter channel of the VCR.

The average glottal flow, which was equal to the applied airflow from the compressor, was measured with a flowmeter (Gilmont J197). The measurement was made by taking the mid-ball reading from the flowmeter and looking up the flow on its calibration chart (Figure 2). The flowmeter was linear in the range from 30 ml/s to 1200 ml/s. The measurement read-outs were recorded on the chatter channel of the VCR.

Data Analysis

After laryngeal manipulations were completed, subglottal pressures, glottal flows, and the sound pressure levels were tabulated manually while listening to the chatter channel playback. The fundamental frequency of vocal fold vibration was determined from the acoustic signal with a Precision Data 6000 Universal Waveform Analyzer. Measurement of vocal fold movement was made by turning the jog dial of the VCR and reviewing the video tape frame-by-frame.

The amplitude of vocal fold vibration was measured on the video screen. As described, there was a phase difference between the top lip and the bottom lip. When the top lip and bottom lip were at the same distance from the midsagittal plane, the glottal area reached its maximum value. The *amplitude of vocal fold vibration* was defined as half of the maximum glottal width at this point. Although this is an underestimate of the maximum lateral tissue excursion, it was the most consistent measurement and served well for comparisons between the hemilarynx and the full larynx.

General Results

Gross performance of the hemilarynx

Self-oscillation of the vocal folds occurred when the subglottal pressure exceeded a threshold value. The typical phonation threshold pressure was 0.78 kPa, but this varied highly with fundamental frequency, as noted by Finkelhor et al.[20]. After a brief break in phonation, oscillation was sustained and the subglottal pressure remained at a level of 0.2-0.4 kPa higher than threshold pressure (for a constant flow source). Perceptually, the sound of the hemilarynx and the full larynx were similar, except that the full larynx was louder. With increases of subglottal pressure, average glottal flows, sound pressure levels, fundamental frequencies, and amplitudes of vocal fold vibration all increased. There usually was range of subglottal pressures above threshold that produced "good phonation". In this range, phonation had the characteristics of chest voice; that is, phonation was generally stable, loud, and rich in timbre. Phonation within this range was defined as *normal phonation.* As subglottal pressure continuously increased and exceeded the normal range, a second range was reached, typically between 2 to 3 kPa. In this range, vibration became very irregular and the sound became rough. The maximum subglottal pressure producing stable phonation was defined as *phonation instability pressure.* The tendency toward instability was observed for both excised hemilarynges and excised full larynges.

Vocal fold movement was observed in apparent slow motion under stroboscopic illumination. From the top view, vocal fold vibration of the hemilarynx was similar to that of the full larynx: the bottom lip of the vocal fold touched first, then the top lip, then both lips moved vertically upward, and finally the top lip separated gradually. When the top separated, the bottom lip was shadowed by the top lip and was invisible from the superior view. Before the top lip reached its maximum excursion, the bottom lip became visible and began to move toward the midline. Equivalently, a mucosal wave traveled from the bottom lip to the top lip.

From the side view in the hemilarynx, the bottom lip made contact with the vertical plane first, then the top lip made contact, then both lips moved up in the sagittal plane. The bottom lip then separated, followed by top lip separation.

The half glottal shape of the hemilarynx looked similar to that of the full larynx during vibration. This glottal shape was measured on the video monitor when lateral displacement reached its maximum value. Figure 3 shows differences in displacement between the hemilarynx and full larynx. The displacement function resembled a half sinusoid (coefficient of variation = 5.01%).

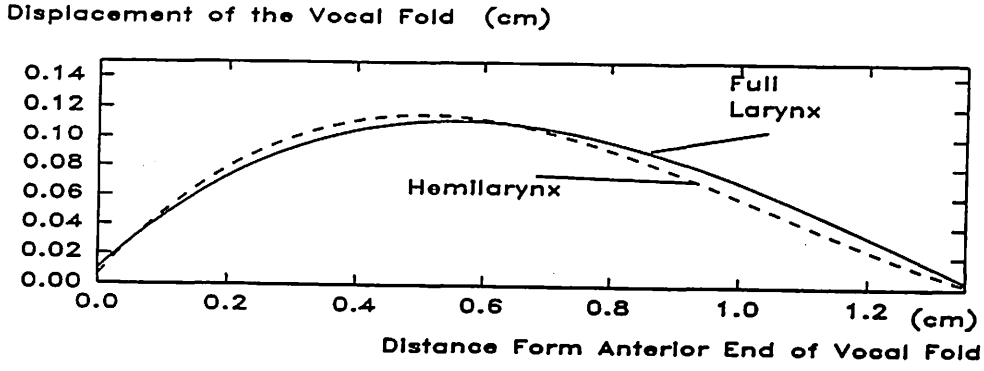


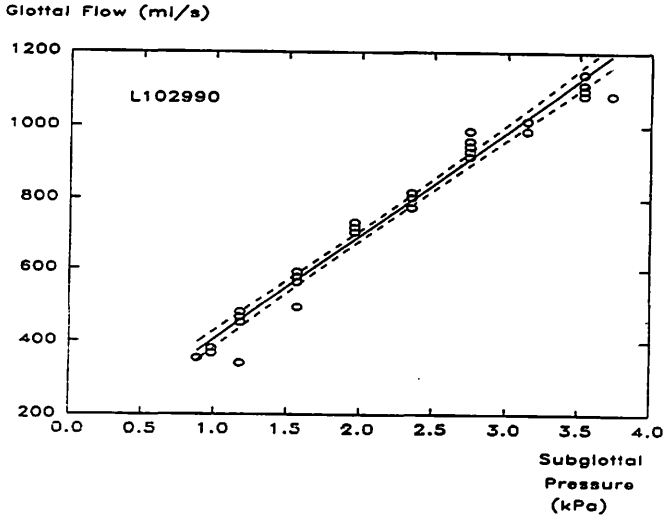
Figure 3. Comparison of vocal fold displacement between hemilarynx and full larynx as viewed from the top.

Repeatability of the measurements

Many factors are likely to affect measurements on excised larynges. These factors include reliability of the instrumentation, stability of the phonation setup, and stability of the tissue properties. A preliminary set of experiments were conducted to assess measurement repeatability prior to the full larynx-hemilarynx comparative measurements. Each measurement was repeated six times on randomly chosen larynges. The time interval between each measurement was about 10 minutes. Thus, the total duration for one repeatability assessment on a given larynx was about 60-70 minutes.

Figure 4 displays repeated glottal flow measurements at different subglottal pressures for one typical larynx. The 99% confidence range was also calculated and plotted. The repeatability error was less than 5%. The data fit a first-order regression line (r=0.98) in this range, but a higher order function is suggested by the deviations at the endpoints.

Figure 5 displays repeated sound pressure level (SPL) measurements at the same subglottal pressures, with a 99% confidence range. Measurement error was less than 3 dB.



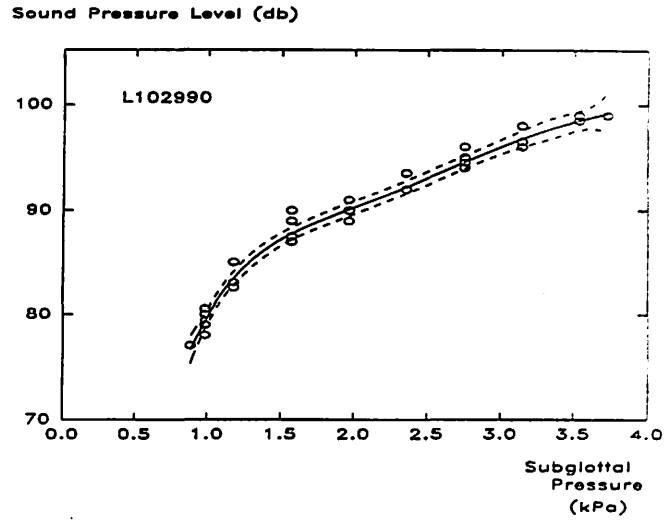
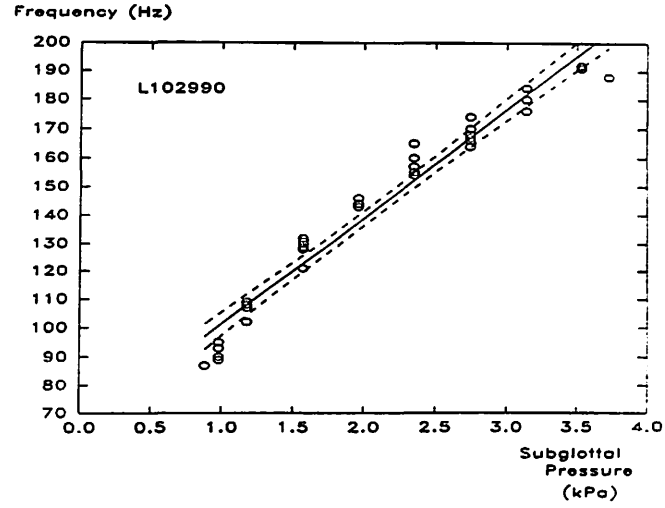


Figure 4 (left). Glottal flow vs. subglottal pressure with a 99% confidence range. Figure 5 (right). The six repeated data groups for sound pressure level vs. subglottal pressure with 99% confidence range.



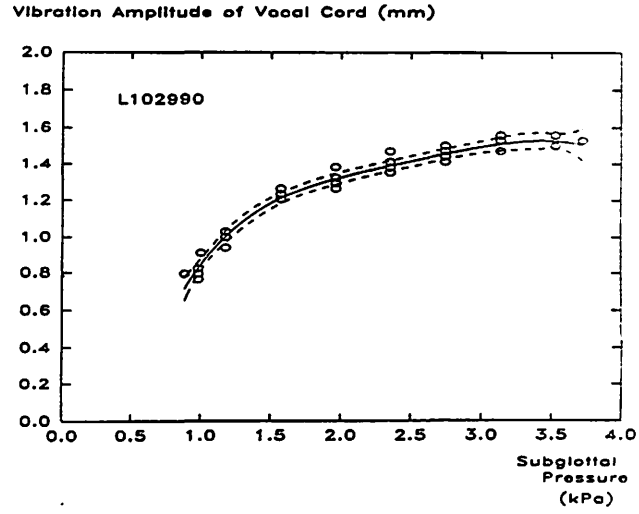


Figure 6 (left). The six repeated data groups for fundamental frequency of acoustic output vs. subglottal pressure with a 99% confidence range. Figure 7 (right). The four repeated data groups for the amplitude of the vocal fold vibration vs. subglottal pressure with a 99% confidence range.

Figure 6 displays repeated fundamental frequency measurements at the different subglottal pressures, with a 99% confidence range. This graph shows the measurement error was less than 10 Hz. The data can be approximated by a first-order linear model (r=0.98) in this range, but a second-order model could also be used if theoretical consideration would warrant it. The average slope is 40 Hz/kPa, which relates closely to the results of Baer[17] and Titze[21].

Figure 7 displays repeated measures of vocal fold vibrational amplitude at the different subglottal pressures, with a 99% confidence range. Only four measurements are plotted because the strobe light accidently did not trigger well in two larynges. The measurement error was less than 10%.

Although the within-larynx repeatability of the measurements was satisfactory, the variances of the measurements across larynges were large because there are great individual differences between larynges. Averages across larynges must be interpreted with caution, therefore.

Specific Comparisons Between Hemilarynges and Full Larynges

The purpose of this analysis was to compare parameters of the hemilarynx with those of the full larynx. Statistics were used to estimate measurement accuracy. The averaged value of each measurement, along with an error range, was calculated and plotted. Curve fitting was utilized to obtain an average trend, which represented the relationship between dependent and independent variables.

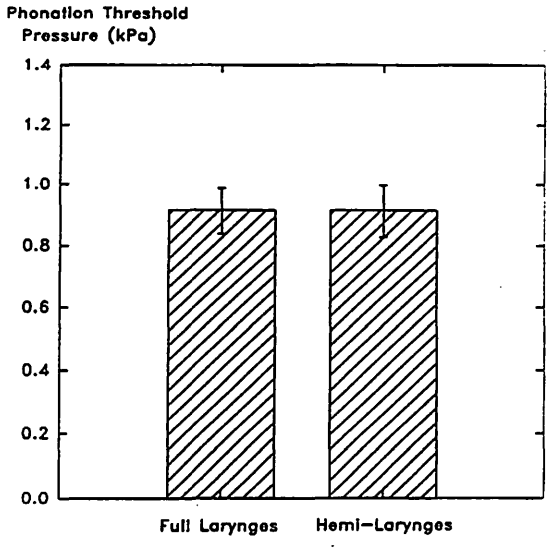
Phonation threshold pressure and range of subglottal pressure

Phonation threshold pressure was defined by Titze[19] as the minimum pressure required to initiate small amplitude vocal fold vibration. It depends on the glottal configuration (adduction, convergence angle of the glottis, vocal fold thickness) and biomechanical properties of vocal fold tissues (elasticity and viscosity). On purely theoretical grounds, Titze predicted phonation threshold pressures between 0.2-1.0 kPa.

In the current experiments, there was no detectable statistical difference between nine full larynges and their respective hemilarynx counterparts, both in terms of a mean value or a standard deviation (Figure 8). The range was similar to ranges found by Baer[17] and Finkelhor, Titze and Durham[20] for low elongation, but it was about 0.2-0.3 kPa greater than the threshold pressure reported by van den Berg[12] for an excised human larynx, and 0.3-0.5 kPa greater than Isshiki's[22] thresholds from a human subject. These results suggest that the human larynx (with living tissue) may have some advantage in ease of phonation over an excised canine larynx. Higher phonation threshold pressures in the excised larynx, which has no supraglottal vocal tract, may also suggest that the vocal tract plays a role in facilitating self-oscillation of the vocal folds.

It was found that a dehydrated vocal fold resulted in increased phonation threshold pressures[20,23]. A larynx with a damaged vocal fold also has higher phonation threshold pressures than a normal larynx [24]. In the same excised larynx study, there was the impression that increases in phonation threshold pressures affected the intensity range more than maximum intensity. This speculation may relate somehow to the strain/strangled voice, which is characterized by a hard-to-start phonation and a narrow intensity range, and especially by the loss of quiet phonation. The current experiment failed to show that under hemilaryngeal conditions, ease of the phonation is impaired.

Another important variable is phonation instability pressure. This is the subglottal pressure for which phonation becomes aperiodic at the high end of the intensity range. For the nine pairs of hemilarynges and full larynges, in which elongation and shim placement were controlled, the average phonation instability pressures in the hemilarynges were about 20% greater than those of the full larynges (Figure 9). Assuming that a systematic error did not play a role, it is possible that this difference in phonation instability pressure was due to differences between glottal flow velocity profiles. More studies are needed to investigate this difference in the future.



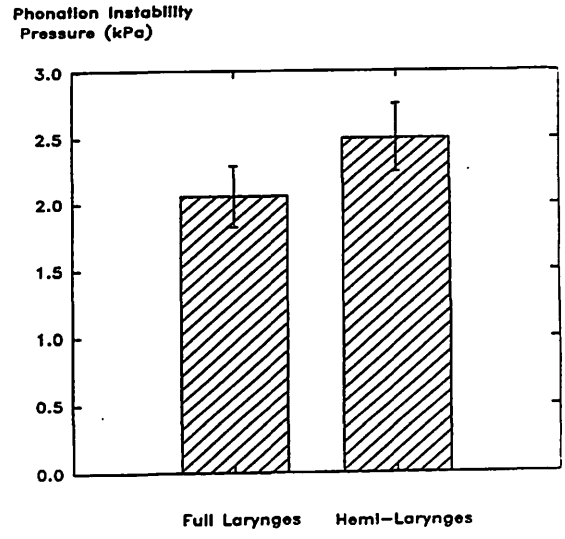


Figure 8 (left). Comparison of phonation threshold pressures between full larynges and their hemilarynx counterparts.
Figure 9 (right). Comparison of phonation instability pressure between full larynges and their hemilarynx counterparts.

Average glottal flow

Figure 10 shows average glottal flow versus subglottal pressure at constant elongation and prephonatory configuration. The microphone was positioned at 15 cm, 45° from the larynx. The overall average flow of the hemilarynx was about half that of the full larynx over the entire phonation range. The slope of the average glottal flow function was 0.3 l/kPa in the full larynx, about twice that of the hemilarynx. By way of comparison with earlier studies, this slope is the average of van den Berg's data (0.2-0.4 l/kPa). In the full larynx, average glottal flow had a positive linear relationship with subglottal pressure, and ranged from 0.2-1.2 l/s, while subglottal pressures ranged from 0.8-4.0 kPa.

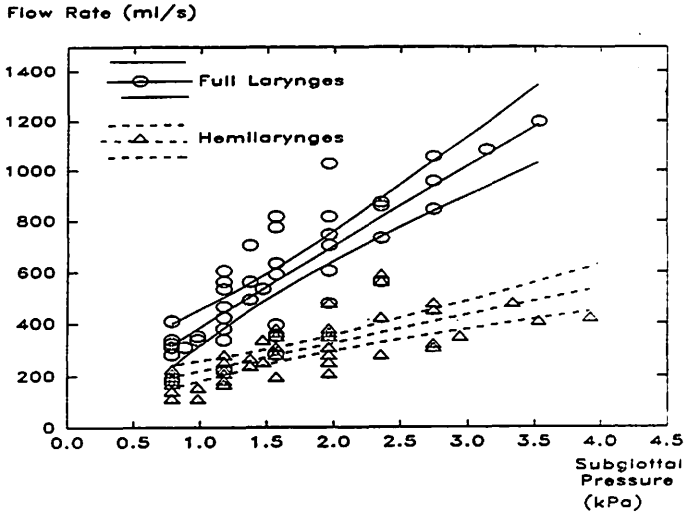
The typical range of flow in human subjects is 0.1-0.3 l/s[22]. In Baer's[17] experimental study on excised canine larynges, flow rate ranged from 0.1 to greater than 0.5 l/s. The present data displayed great concentration in the 0.2-0.5 l/s range. Although flow rates were higher than those typical for human phonation, they were similar to those of Bear and other reported data with excised larynges[12] and live canines[25].

Sound pressure output regulation of the hemilarynx

Figure 11 shows sound pressure level versus subglottal pressure at constant elongation and prephonatory configuration. The microphone was positioned at 15 cm, 45° from the larynx. The average sound pressure ranged from 80-96 dB in the full larynx and from 75-90 dB in the hemilarynx. The dynamic range of the sound level output was about 15 dB.

Assuming that glottal flow is nearly proportional to glottal area[26,27], and realizing that power is proportional to flow squared[28], the power of the acoustic output of the hemilarynx is predicted to be one-fourth that of the full larynx. In terms of sound pressure level, this corresponds to 6 dB, the difference observed.

According to informal clinical experience, the superior hemilaryngectomy patients seem to have phonation that is slightly softer than that of normal subjects, and perhaps a little more pressed. The present data may help explain this phenomenon. Since a hemilarynx produces 6 dB less power at the same subglottal pressure, patients may press harder to obtain comparable output. Alternately, if the same sound power is expected, the hemilarynx may need twice as much subglottal pressure to drive it, which may exceed the phonation instability pressure and cause unstable phonation.



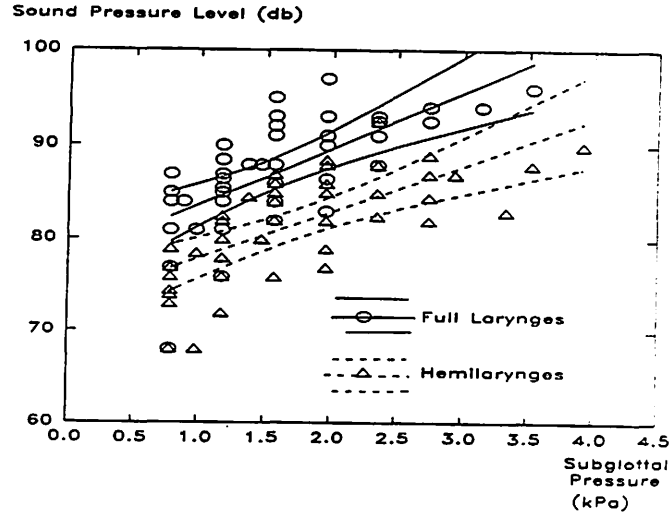


Figure 10 (left). Average glottal flow vs. subglottal pressure at the same elongation and prephonatory configuration for hemilarynges and full larynges (n=9). Figure 11(right). Sound pressure level vs. subglottal pressure for hemilarynges and full larynges at the same elongation and prephonatory configuration (n=9). The microphone was positioned 15 cm, 45° from the larynx.

Fundamental frequency

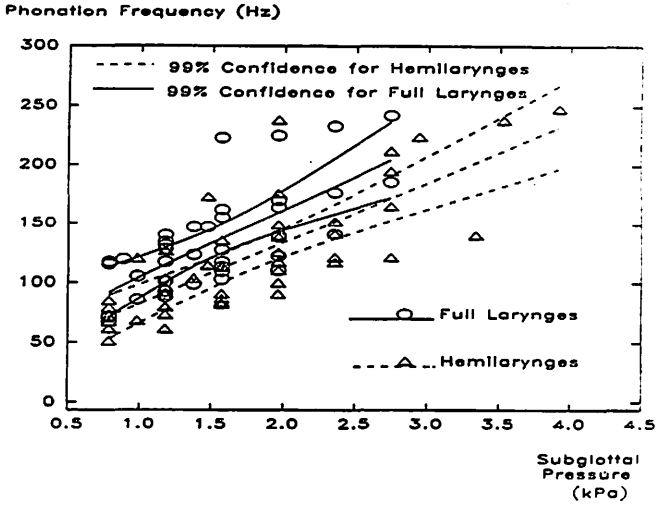
Figure 12 shows fundamental frequency versus subglottal pressure at constant elongation and prephonatory configuration. Averaged over the nine larynges, fundamental frequency had a positive linear relationship with subglottal pressure, ranging from 60 to 250 Hz, while subglottal pressure ranged from 0.7 to 3.7 kPa. The overall average frequency of the hemilarynx was slightly lower (about 25 Hz) than that of the full larynx over the phonation range. The slope of fundamental frequency increase over subglottal pressure was 50 Hz/kPa in the hemilarynx and 56 Hz/kPa in the full larynx. There was no statistical difference between them.

For the excised human larynx, the typical frequency range obtained by van den Berg and Tan[12] was 60-220 Hz in the subglottal pressure range of 0.5-2.9 kPa. In Baer's[17] data, subglottal pressure ranged from 0.5-1.7 kPa, with the greatest concentration in the 0.7-0.9 kPa range, whereas

F_0 ranged from 80-140 Hz, with the greatest concentration in the 80-110 Hz range. The frequency range that Titze[21] reported was 75-200 Hz. He measured and predicted the relation between subglottal pressure and fundamental frequency in phonation based on dynamic tension.

F_0-P_s slopes from the cited sources[13,17,21] all seem to cluster around 30-70 Hz/kPa. Our frequency data from both the hemilarynx and the full larynx were very similar, about 50 Hz/kPa. Titze predicted that the slope should decrease with vocal fold length, but this was not verified here.

Clinically, F_0 change does not seem to be a problem for partial vertical laryngectomy patients. Although there are no clinical data available on frequency regulation as a function of subglottal pressure for the hemilarynx, it would appear that no major abnormalities should occur, based on the present findings.



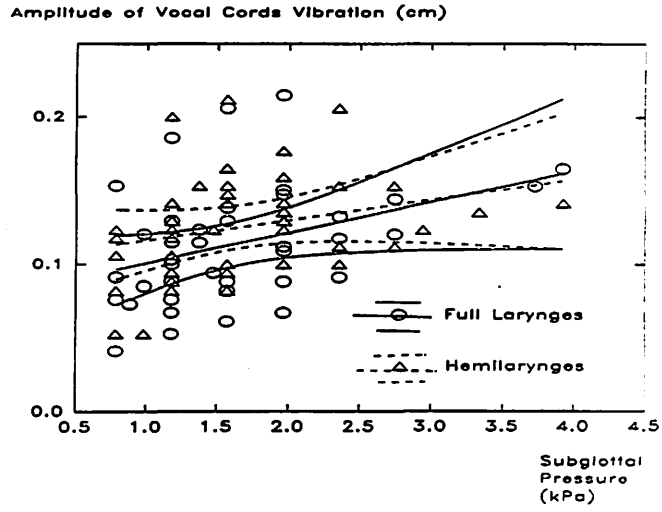


Figure 12 (left). Fundamental frequency vs. subglottal pressure for hemilarynges and full larynges at the same elongation and prephonatory configuration (n=9). Figure 13 (right). Vocal fold vibration amplitude vs. subglottal pressure for hemilarynges and full larynges at the same elongation and prephonatory configuration (n=9).

Vibrational amplitude

Figure 13 shows vibrational amplitude versus subglottal pressure at constant elongation and prephonatory configuration. Amplitude of vibration increased with subglottal pressure, and ranged from 0.05 cm to 0.2 cm in both the full larynx and the hemilarynx. The average amplitude of vocal fold vibration was about 0.01 mm (about 10%) greater in the hemilarynx than in the full larynx in the lower subglottal pressure range, and identical in the high subglottal pressure range. However, this difference was within the 99% confidence range, and was not statistically significant.

The effect of air viscosity on flow in the hemilarynx should be considered as one source of discrepancy. According to basic fluid dynamics, when a fluid moves over a solid surface, the relative velocity between the solid surface and the contacting layer of fluid is zero. The velocity at the surface of the vocal folds should be zero, therefore. According to velocity profiles along the glottis calculated by Alipour and Patel[29], the boundary layer of airflow is about 1/10 of the total glottal width. In the hemilarynx, because the glottis is the space between the vocal fold and a still vertical plate at the midline, the velocity profile has an additional boundary layer, and is not a mirror image of the profile of the full larynx. Even if the vocal fold were to vibrate at the same amplitude

and the glottal width were to be exactly half of that of the full larynx, the flows could not be identical in the half-glottis region. Due to nonlinear pressure-flow relations that depend on absolute values of glottal widths, glottal resistance would not scale in a simple 2:1 ratio. If pressures and flows are similar, therefore, some differences in amplitude of vibration can be expected.

In former experiments with excised canine larynges, a typical range of amplitude of vibration was reported to be 0.5-1.8 mm when elongation was zero[21]. This result is similar to the 0.5-2.0 mm range in Figure 13 The trend of the current data grossly fits the trend in Titze's data, as well as his mathematical predictions that amplitude varies with the square root of subglottal pressure[21]. The slope of the amplitude increase with subglottal pressure appeared to be slightly less for the hemilarynx than for the full larynx. This may again be caused by the nonlinear flow properties, but the exact differences are still topics which need more study.

Suggestions for Phonosurgery

A practical question is: What is the desirable procedure for reconstruction when one remaining vocal fold is normal? It is obvious that if the impaired vocal fold can be reconstructed such that normal geometric and biomechanical properties are regained, normal phonation will result. Another clear-cut situation is the one where one side of the vocal fold is unrepairable (such as in carcinoma of the larynx). In this situation, the current experiments suggest that it is still possible to produce near normal phonation if the single vocal fold can vibrate against a rigid vertical wall and the necessary adjustments on prephonatory settings of the remaining vocal fold can be achieved.

Situations between these extreme cases are more difficult to deal with. In previous studies[24,30], it was found that unilaterally scarred vocal folds, even though they still have a grossly normal shape, produce low intensity (on the average 12 dB less than normal) and unstable phonation with a high phonation threshold (about 1.2 kPa higher than normal). This might suggest that even if a vocal fold can be reconstructed grossly, but the biomechanical and geometrical characteristics of the fold cannot be restored, then the results of reconstruction might be worse than for a simple wall-like structure. According to computer simulations of vocal fold vibration[31] and clinical observations[32], asymmetries between the vocal folds can cause aperiodic voice. There are two choices for surgeons to make for treating a damaged vocal fold: (1) repair it to make it normal enough to cooperate with the opposite normal vocal fold and produce symmetric phonation, or (2) simply modify it as a wall-like structure which is stiff enough not to interfere with the vibration of the opposite vocal fold and sacrifice 6 dB intensity.

A fully controllable arytenoid cartilage is also important in reconstructive surgery. In this study, satisfactory phonation occurred only in certain combinations of arytenoid adduction, elongation, and tension of the top and bottom lip. A well reconstructed larynx should have three-dimensional movement control of the arytenoid cartilage, which is not only important for voice production, but also for respiration.

Conclusion

Based on the data of these experiments, the hemilarynx is similar to the full larynx in terms of phonation threshold pressure, phonation instability pressure, fundamental frequency, and vibrational amplitude. Differences were all less than 10%. Average airflow was scaled 2:1 and acoustic power was scaled 4:1, or 6 dB less in the hemilarynx than in the full larynx. These results suggest that an excised hemilarynx is a reasonable substitute for an excised full larynx for experimental purposes in which access to the medial plane of the vocal fold is needed.

Acknowledgement

This work was supported, in part, by a grant from the National Institute on Deafness and Other Communication Disorders, grant No. P60 DC000976-01.

References

1. Alonso, J.M. (1947). Conservative surgery of cancer of the larynx. Trans. Amer. Acad. Ophthal. Otolaryng., 51, 633.

2. Titze, I.R. (1984). Parameterization of the glottal area, glottal flow, and vocal fold contact area. J. Acoust. Soc. Am., 75, 570-580.

3. Cooper, D.S. (1986). Research in Laryngeal Physiology with Excised Larynges. In: Cummings C.W. (Ed.) Otolaryngology-Head and Neck Surgery, Vol. 3. St. Louis: C.V. Mosby, 1766-76.

4. Cooper D.S. (1989). Voice: A historical perspective. J. of Voice, 3, 187-203.

5. Ferrein, A. (1746). De la formation de la voix de l'homme. Suite des mémoires de mathématique et de physique tirés des registres de l'Academie royale des sciences de l'année MDC-CXLI. Amsterdam: Pierre Mortier, 545-79.

6. Ferrein, A. (1744). De la formation de la voix de l'homme. Histoire de l'Académie royale des sciences. Année MDCCXLI. Avec les Memoires de Mathématique et de Physique, pour la meme Année. Paris: I'Imprimerie Royale, 409-32.

7. Liskovius C. (1814). Dissertatio physiologica sistens theoriam vocis/Theorie der Stimme. Leipzig: Breitkoft und Haertel.

8. Liskovius, K. (1846). Physiologie der menschlichen Stimme. Leipzig: Verlag von Joh,. Amb. Barth.

9. Lehfeldt, C. (1835). Nonnulla de vocis formatione. Berlin: Typis Nietackianis.

10. Müller, J. (1837). Handbuch der Physiologie des Menschen, Vol. 2., Sect. I Von der Stimme und Sprache. Coblenz: Verlag von J. Holscher.

11. Müller, J. (1839). Über die Compensation der physischen Kräfte am menschlichen Stimmorgan. Berlin: Unger.

12. van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. J. Speech Hearing Res. 1, 227-244.

13. van den Berg, J., and Tan, T.S. (1959). Results of experiments with human larynges. Practa Oto-Rhino-Laryng., (Basel) 21, 425-450.

14. van den Berg, J. (1960). Vocal Ligaments Versus Registers, Current Problems in Phoniatrics and Logopedics, Vol. 1, pp. 19-34.

15. Anthony, J.K.F. (1968). Study of the Larynx II. Works in Progress No. 2, 77-82, Dept. of Phonetics and Linguistics, University of Edinburgh, Scotland.

16. Hiroto I. (1966). The mechanism of phonation - Its pathophysiological aspects. Oto-Rhino-Laryngol. Clinic, Kyoto, 39, 229-291. (in Japanese).

17. Baer, T. (1975). Investigation of Phonation Using Excised Larynges. Ph.D. dissertation, Massachusetts Institute of Technology. Boston, Massachusetts.

18. Durham, P.L., Scherer, R.C., Druker, D.G., and Titze, I.R. (1987). Development of Excised Larynx Procedures for Studying Mechanisms of Phonation. Technical Report VABL-1-1987.

19. Scherer R.C., Druker D.G., and Titze I.R. (1988). Electroglottography and Direct Measurement of Vocal Fold Contact Area. In O. Fujimura (Ed.), Vocal Fold Physiology: Voice Production, Mechanisms and Functions (pp. 279-292). New York: Raven Press.

19. Titze, I.R. (1988b). The physics of small-amplitude oscillation of the vocal folds. J. Acoust. Soc. Am. 83(4), 1536-1552.

20. Finkelhor, B.K., Titze, I.R., and Durham, P.L. (1988). The effect of viscosity changes in the vocal fold on the range of oscillation. J. of Voice, 1, 320-325.

21. Titze, I.R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. J. Acoust. Soc. Am., 85, 901-906.

22. Ishikki, N. (1964). Regulatory mechanism of voice intensity variation. J. Speech Hearing Res., 7, 17-29.

23. Verdolini-Marston, K., Titze, I., & Druker, D. (1990). Changes in phonation threshold pressure with induced conditions of hydration. J. of Voice, 42(2), 142-151.

24. Wexler, D., Jiang, J., Gray, S., and Titze, I. (1989). Phonosurgical studies: Fat-graft reconstruction of injured canine vocal folds. Ann. Otol. Rhinol. Laryngol., 98, 668-673.

25. Koyama, T., Harvey, J.E., and Ogura, J.H. (1971). Mechanism of voice production II. Regulation of pitch. Laryngoscope, 81, 47-65.

26. Flanagan, J.L. (1958). Some properties of the glottal sound source. J. Speech Hearing Res., 1, 99-116.

27. Flanagan, J.L. (1968). More-system interaction in the vocal tract. Annals N.Y. Acad. Sci., 155, 9-17.

28. Titze, I.R. (1988a). Regulation of vocal power and efficiency by subglottal pressure and glottal width. In O. Fujimura (Ed.), Vocal Fold Physiology: Voice production, Mechanisms, and Functions (pp. 227-238). New York: Raven Press.

29. Alipour, F. & Patel, V.C. (1991, April/May). A Two-dimensional Model of Laryngeal Flow. 121st. Acoustical Society of America Meeting.

30. Jiang, J., and Titze, I. (in review). Measurement of vocal fold intraglottal pressure and impact stress.

31. Ishizaka, K., and Isshiki, N. (1976). Computer simulation of pathological vocal cord vibration. J. Acoustic. Soc. Am., 60 1193-8.

32. Hanson, D.G., Gerratt, B., Kanin, R., and Berke, G. (1988). Glottographic measures of vocal fold vibrations: An examination of laryngeal paralysis. Laryngoscope, 98, 541-549.

NCVS Status and Progress Report - 3
December 1992, 21-37

Measurement of Vocal Fold Intraglottal Pressure and Impact Stress

Jack Jiaqi Jiang, M.D., Ph.D.
Department of Otolaryngology/Head and Neck Surgery, Northwestern University
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Intraglottal pressure was measured with a previously described hemilarynx procedure. Three phases were identified for intraglottal pressure: an impact phase, in which the vocal folds come into contact and produce a sharp pressure pulse; a pre-open phase, in which there is a progressive pressure build up due to increased exposure of the vocal fold surfaces to subglottal pressure; and an open phase, in which intraglottal pressure becomes aerodynamic and drops gradually from opening to closing. Impact pressure peaks were positively related to subglottal pressure, elongation, and adduction of the vocal folds. The midpoint of the membranous vocal fold received the maximum impact stress. The experimental results match well with predicted results based on analytical models and support a current theory of mechanical trauma leading to vocal nodules.

Introduction

Impact stress on vocal fold tissues is an important variable for understanding the mechanism of phonation and for determining the etiology of vocal nodules. The purpose of the present experiment was to measure impact stress in conjunction with aerodynamic pressure in the glottis of an excised hemilarynx[1]. *Impact stress* was defined as the normal stress on the contacting surfaces of the vocal folds. The stress component parallel to the surface (the shear stress) was ignored. Aerodynamic pressure was defined as the pressure on a vocal fold surface point when there is contact only with air. The following relations were investigated: (1) Impact stress and aerodynamic pressure versus subglottal pressure, (2) impact stress and aerodynamic pressure versus prephonatory configuration, (3) impact stress versus vocal fold elongation, and (4) spatial distribution of impact stress on the vocal fold.

To bring the measurement of impact stress into a historical perspective, a brief sketch of previous attempts to measure intraglottal pressure is presented.

Background on Intraglottal Pressure Measurement

Intraglottal pressure is the driving pressure for vocal fold vibration. Its observation in conjunction with vocal fold movement, contact area, and the acoustic output is essential to a full understanding of the mechanism of phonation.

Accompanying the development of a phonation theory, there have been a series of experiments measuring intraglottal pressure. During the 19th century, the study of laryngeal aerodynamics was approached primarily on the basis of physical models or experiments with excised larynges. According to Cooper[2], Harless[2] expressed regret that manometers did not have the high frequency response necessary to measure the dynamic variation of laryngeal air pressures during the glottal cycle; however, Harless hypothesized negative air pressures in the lower part of the glottis ("immediately below the fold") when the vocal folds were approaching each other at narrow glottal widths. In 1925, Tonndorf (cited in Cooper[2]) studied the Bernoulli principle with excised bovine larynges and found that, besides regions of strong positive pressure, there were also regions of strong negative pressure. According to Cooper, Weiss[4] published the first simultaneous cycle-to-cycle records of dynamic subglottal pressures. Weiss also recorded medial-lateral movement of the vocal folds by shining a narrow beam of light through the glottis, resulting in an image that permitted separate observation of the movement of right and left vocal folds.

Rethi (1897; cited in Scherer et al.[5]) placed a pressure-measuring elastic instrument into the glottis of animals and stimulated the cricothyroid and posterior cricoarytenoid muscles. The highest values of closing pressure were estimated to be approximately 75 kPa. Kakeshita (1927; cited in Scherer et al.[5]) placed an inflatable rubber balloon within the glottis of canines and obtained pressures greater than 53 kPa during barking. Murakami and Kirchner[6,7] used a low-frequency pressure catheter between the vocal folds of cats to measure pressure changes due to nerve stimulation. The largest pressure variation measured was about 1.68 kPa.

Scherer and Titze[8] placed a miniature transducer between the vocal processes of an excised bovine larynx and found that the maximum variation in contact pressure during phonation ranged from values less than the average subglottal pressure to values twice as large as the average subglottal pressure. In particular, the within-cycle variation of interarytenoid pressure was 1.5-11.3 kPa. Their results showed a positive relationship between peak-to-peak variations in interarytenoid pressure and mean subglottal pressure.

More recently, Reed et al.[9] obtained an intraglottal pressure waveform from a human volunteer. The waveform was similar to that reported by Scherer et al., namely, a pronounced initial peak and a long, rounded segment. Reed et al also found that there is a clear increase in the output of the intraglottal transducer with increases in the amplitude of the acoustic signal.

According to these studies, it would seem that current knowledge of intraglottal pressures would be advanced by a more systematic study that quantifies the relation between intraglottal pressure and phonatory control variables such as lung pressure, adduction, fundamental frequency, and mode of phonation.

Background on the Etiology of Vocal Nodules

Vocal nodules are the most common vocal disorders. They represent 2-3.9% of the entire ENT case load[10]. Since Terce first described the condition in 1868, discussion of etiology, histological nature, and therapy regarding vocal nodules has not ceased[11,12].

Chiari (1895; cited in Kambic[13]) and Epstein, Winston, Friedmann, & Ormerod[14] stressed the mechanical nature of the origin of nodules. These pathologic growths are now generally considered to be the consequence of mechanical trauma. A number of hypotheses have been proposed to explain the cause of vocal nodules and polyps[15]. However, no supporting data are strong enough to be used as evidence for the mechanical trauma hypothesis. The mechanical effects of vibration of vocal folds should be measured directly to clarify their influence on the tissue microstructure during phonation.

The following three hypotheses have been proposed regarding the types of mechanical influences that may be important.

(1) Mechanical pressure hypothesis: According to Arnold[16,17] nodules are the mechanical result of faulty or excessive vocal use. They may be likened to calluses on the hand or corns on the toes, the result of mechanical stress applied by tools or tight shoes. Sonnien et al.[15] hypothesized that microtrauma is more likely to be caused by pressing forces ('hammer-and anvil mechanism'). Vaughan[18] stated that vocal nodules develop from mechanical trauma caused by one vocal fold rubbing against the other. These authors all agree on the direct mechanical cause of hyperkinetic phonatory movements.

(2) Mechanical lifting hypothesis: Gery[19] stated that vocal nodules and polyps that occur at the junction of the anterior and middle thirds of the folds are due to mucosal distortion on separation, and not due to the trauma of collision. As the vocal folds abduct, a triangular wedge of mucous membrane forms. As the folds further abduct, this triangular wedge is the last point to separate and corresponds to the site of vocal nodules. As the process continues, the mucous membrane thickens, and eventually nodules occur.

(3) Accumulation hypothesis: By using a mathematical analysis and a physical model of vocal fold vibration, Jiang[20] found that the liquid on a vibrating band tends to accumulate toward the midpoint of the band. Clinically, vocal fold nodules also tend to occur at the mid-membranous fold. Combining the equation of fluid motion with the dynamics of string vibration, it was shown that fluid pressures build up towards the center of the vocal folds and that mechanical stresses may result from fluid accumulation.

Although all three hypotheses involve the claim that mechanical stress is the key to the etiology of vocal nodules, accurate measurement of such a stress has not been accomplished. The measurements presented here will help to set the stage for eventual formal testing of the above hypotheses.

Biomechanical Analysis and Prediction of Impact Stress

According to the body-cover theory[21,22], the vocal fold cover can be modeled as a vibrating ribbon. The mass M of a small element near the surface is:

$$M = \rho dx dy dz \quad , \tag{1}$$

where ρ is the tissue density and dx, dy, dz are spatial increments. Assuming the lateral displacement ξ of the tissue element to be sinusoidal in time during the open phase of the glottal cycle,

$$\xi = A \sin 2\pi f t \quad \left(0 < t < \frac{1}{2f}\right) \tag{2}$$

where A is the displacement amplitude and f is the frequency of vibration. The velocity of the element will be

$$v = 2\pi f A \cos 2\pi f t \ . \tag{3}$$

If we define the beginning of impact as the moment of peak velocity and the end of impact as the moment when the center of mass of the element is zero, then the change in velocity Δv during the impact is

$$\Delta v = 2\pi f A \ , \tag{4}$$

and, according to Newton's second law, the mean impact force is

$$F = (\rho dx dy dz)(2\pi f A)/\Delta t \ , \tag{5}$$

where Δt is the impact duration.

If the impact force is evenly distributed over the surface area *dydz*, the mean contact stress (pressure) on the element

$$P = (\rho dx)(2\pi f A)/\Delta t \ . \tag{6}$$

The thickness *dx* of vibrating tissue is assumed to be on the order of 0.1 cm[23] and typical amplitudes of vibration are also on the order of 0.1 cm. For a fundamental frequency of 200 Hz (in both the male and female range), a tissue density of 1.0 g/cm^3, and a Δt measured typically at 0.5 ms in this experiment, the predicted mean contact stress is on the order of 3.0 kPa.

If the anterior-posterior (*y*) variation of the ribbon-like displacement resembles a half-sinusoid[24,1]), then

$$P(y,t) = (2\pi\rho dx A f/\Delta t)\sin(\pi y/L) \tag{7}$$

Note that the predicted stress is propositional to vibrational amplitude and frequency. Note also that the predicted stress reaches its maximum at the midpoint of the membranous fold ($y=L/2$). At the *L/4* and *3L/4* points, the contact stress is

$$P_{1/4} = P_m \sin(\pi/4) \ , \tag{8}$$

or about 70.7% of the maximum contact stress P_m at the midpoint. We now describe some methods for testing the above relations experimentally.

Methods

Biological Tissue

Experimental data were obtained on five larynges. In addition, for normalized peak contact pressure measures, data from four larynges used previously[1] were included because these normalized data did not require a pressure and flow calibration. All larynges were harvested from animals euthanized for other experimental purposes. No animals were sacrificed for these experiments. There was no regard for sex or age of the animal.

Instrumentation, Calibration, and Data Recording

Details of the hemilarynx preparation procedure, the subglottal system, mounting apparatus, strobe light illumination, and audio and video recording systems are described elsewhere[1]. A brief synopsis is given here, however.

An apparatus was constructed with which vibration of an excised hemilarynx could be observed from the customary superior aspect (Figure 4, center-bound plate). The left half of the larynx was removed and replaced by a vertically-oriented plexiglass plate, held in place by two screws (top and bottom) as shown. Some additional mounting screws underneath the plate are seen shining through the plate. A three-pronged positioning device (right side in middle of photograph) was used to adjust the position of the arytenoid cartilage of the vocal fold. For adjustment of vocal fold length, a rod was connected via a 90° crossarm and suture to the anterior tip of the thyroid cartilage (top right of the photograph). The apparatus included instrumentation for recording and measuring the position of several observable landmarks (note three tiny black dots on vocal fold edge). Instrumentation for measurement of contact area profiles, intraglottal pressure, and contact stress is not shown for clarity.

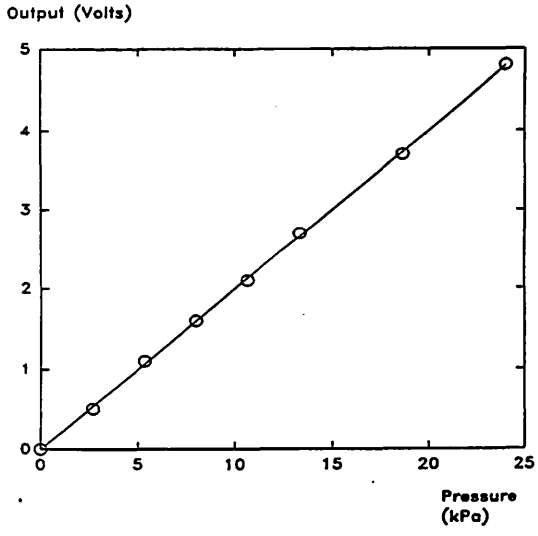
The vertically oriented plexiglass plate used in this experiment was modified from a plate used previously[1]. Three holes were drilled into the plate, in line with expected vocal fold contact locations (see center-bound plate, Figure 5, plate tilted to see holes from vocal fold side). The opening of the middle hole faced the midpoint of the membranous vocal fold. The opening of the anterior and posterior holes faced the one-fourth anterior and one-fourth posterior points on the membranous vocal fold. An Entran miniature pressure transducer (EPB-125-5) was mounted in one of these holes (the middle one in Figure 5), to measure the contact stress at any of the three locations. When one hole was in use, the other two holes were sealed with a modified flat-end drill bit (seen as black rods in the figure).

The stress-sensitive stainless steel diaphragm of the transducer and the flat end of a machine screw were flush with the surface of the vertical plate so that the vocal fold side of the vertical plate had a continuous flat surface. The effective stress-sensitive diaphragm surface was less than 7 mm^2, which is a fraction of a typical maximum vocal fold contact area (10 mm length x 3 mm depth = 30 mm^2). The surface area could therefore be approximated as a "point of contact" between the vocal fold and the diaphragm. During vibration, the rise time of contact was typically about 0.5 milliseconds, based on the measurements with a Data 6000 digital signal analyzer. This rise time is small in relation to an average 10 millisecond fundamental period of vibration. Thus, both temporal and spatial resolution of contact stress were obtained, although temporal resolution was much higher.

A sphygmomanometer (Nissei D-267038) was used to calibrate the Entran pressure transducer. The DC calibration result is shown in Figure 1 (below). The pressure signal was displayed and measured on the screen of a Data 6000 universal digital waveform analyzer. Since the frequency response of the combined transducer and amplifier (Entran IMV-15) was linear ($< \pm 0.5$ dB) from DC-40 kHz according to the technical specifications of the manufacturer, separate AC calibration was not considered necessary.

According to spectral analysis done on the Data 6000 Signal Analyzer, the actual frequency spectrum of the pressure waveform encountered in the experiment was 50 Hz.-10 kHz. This is not only within the frequency response of the device, but also of the Hi-Fi audio channels of a Panasonic 1960 VCR, which was used to record the signal. At least four audio channels were needed for recording the following signals: microphone, pressure waveform, strobe flash, and chatter signal. Since the VCR had only two audio channels, a 2x2 switch was used to select the appropriate channels. During the experiment, one switch combination was used to record a segment of phonation with two signals, then another combination was used to record a second segment with two other signals. Phase locking of all signals was obtained by combining first strobe flash with contact pressure, then contact pressure with microphone pressure. The phase error of two audio channels was less than 0.05 ms.

Because the pressure sensor was made of semiconductive material, there was a significant baseline drift of pressure over time. Such a drift would not have been recorded by the VCR. However, when the equipment was warmed up properly (> 30 min.) and the baseline was reset between trials (when the airflow stopped), the typical baseline shift over 5-10 seconds was less than 10% of the peak-to-peak signal. This duration allowed a segment of the pressure waveform to be recorded with a relatively dependable zero pressure baseline (Scherer 1990, personal communication). Moreover, the peak contact stress measurement was not affected by baseline shift because the important frequency components of the contact stress were within the audio range.



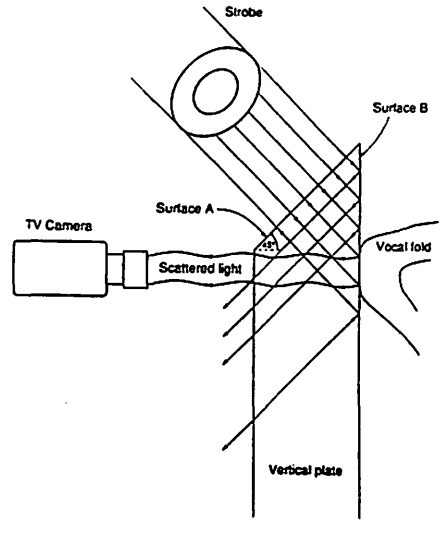


Figure 1 (left). Calibration for the Entran EPB-125-5 pressure transducer by use of a sphygmomanometer (Nissei D-267038). Figure 2 (right). Schematic showing the principle of illumination of the contact area. The illumination strobe light (Pioneer DS-303) is introduced at surface A, which is cut at 45° from surface B.

For visual observation, the top of the vertical plate was ground to a prism in order to increase the contrast between contact and non-contact regions. The principle is illustrated in Figure 2. Stroboscopic illumination (Pioneer DS-303) was introduced at surface A, which was at 45° from surface B. If the medium on the right side of B is low-density air, total reflection occurs because plexiglass has a higher-density medium than air. However, if a wet vocal fold makes contact with surface B, the density of water is close to the density of the plexiglass and total reflection does not occur; therefore, light penetrates surface B and illuminate the contact region, which becomes bright and easily identifiable with scattered light received at 90° to the surface. The principle of total reflection versus partial reflection and partial transmission worked well for this experiment. To keep the vocal fold wet, saline was dropped onto the tissue during the experiment.

Measurement and Data Analysis

Pressure waveforms were displayed on the Data 6000 Signal Analyzer, either directly from the transducer-amplifier or from the audio channels of the VCR. They were also plotted on an HP 7475A Plotter. As stated above, any DC baseline of the pressure waveform could be measured only if records came directly from the transducer amplifier.

Peak stress was measured from the screen of the Data 6000. The stability of the peak stress was estimated on the basis of five repeated measurements in a single larynx. The measurement error was found to be less than 5%. The measurement of peak stress was dependent on the relative depth (flushness) of the sensor diaphragm with respect to the surface of the vertical plate, as shown in Figure 3. The results show that the error was less than 5% when the diaphragm of the sensor was between 0.0 and 0.5 mm above the surface of the plate. Recessed (negative) positions gave erroneous results. The test was repeated for a second larynx, as shown.

Vocal fold elongation was measured from the distance between the anterior and posterior stitch marks on the screen of the video monitor (these were the small dots shown in Figure 4 on plate). All the measurements were normalized with respect to the initial elongation between anterior and posterior stitch marks (before mounting, when the vocal fold was in the adducted position). The purpose of this normalization was to reduce the effect of individual larynx differences in vocal fold length.

Vocal fold adduction was determined by the distance between the arytenoid cartilages. In this experiment, the distance was controlled by placing various sizes of wooden shims between the arytenoid cartilages. The adductive force against the shim was maintained constant by a micrometer so that thicker shims caused less adduction. The thickness of the wood shim was used as a parameter to represent the degree of

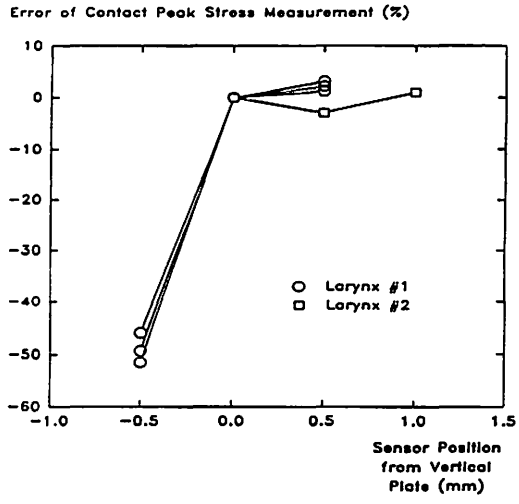
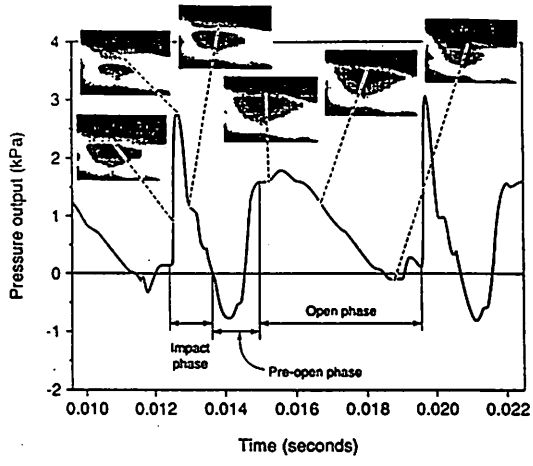


Figure 3. Measurement error versus sensor position from the vertical plate. The peak impact stresses were normalized to contact stress at the surface of the vertical plate (measurements were made on two larynges).

adduction. A "negative" shim size was defined as an overshooting of the micrometer after it passed the controlled standard force. This overshooting was measured in micrometer units. Typically, the best phonation was obtained at a shim size of 0.5 mm. This setting was called the *standard setting* and was used to normalize vocal fold adduction. There were three adduction conditions in this study: (1) the standard setting; (2) the standard setting plus 1 mm adduction (for which the shim size was -0.5mm); and (3) the standard setting minus 1 mm adduction (for which shim size was 1.5 mm).

The relation between peak stress and location of contact on the vocal fold was observed during optimal phonation, when the subglottal pressure was 1.2-1.6 kPa. Because of the large variation of this peak stress among trials and across larynges, each measured pressure was normalized to the peak stress at the midpoint of the vocal fold. Two normalized peak stresses, corresponding to anterior and posterior positions on vocal fold, were then studied by analysis of variance (ANOVA).



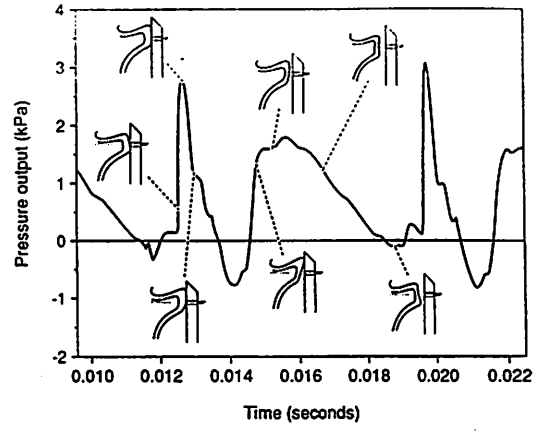


Figure 4 (left). Typical intraglottal pressure waveform with corresponding contact profiles in normal voice. Figure 5 (right). Typical intraglottal pressure waveform with sketches of corresponding vocal fold movement in coronal plane.

Results and Discussion

Intraglottal Pressure Waveforms and Vocal Fold Movement

Figure 4 shows a typical waveform with corresponding contact profiles and Figure 5 shows the same waveform with sketches of the vocal fold in coronal view. These two figures will be discussed side by side. Based on simultaneous observations of vocal fold contact conditions with a vertical plate and sensor, the intraglottal pressure waveform was classified as having three phases: *impact phase*, *pre-open phase*, and *open phase*.

The impact phase occurred when the vocal fold first made contact with the vertical plate and pressure sensor. Normal stress between the vocal fold and the vertical plate (or pressure sensor) rose quickly and assumed a peak value of approximately 3.0 kPa. This segment of the intraglottal pressure waveform was called *impact stress*. The impact phase finished with a quick stress relaxation.

The pre-open phase occurred when the bottom lip of the vocal fold initiated separation and the top lip was still in contact with the vertical plate. In this phase, a build-up of intraglottal pressure resulted from subglottal pressure being applied to the tissue, starting from the bottom and ending at the top lip of the vocal fold. The glottal shape was convergent in this phase. The pre-open phase ended with another large positive intraglottal pressure, in which the pressure sensor had no contact with the vocal fold and the pressure waveform represented subglottal pressure.

The opening phase occurred when vocal fold had no more contact with the vertical plate. This phase started with the opening of the top lip of the vocal fold. Pressure dropped over most of this phase. At the end of this phase, just before the peak stress of next cycle, the pressure was sightly negative. Much has been made in the literature about this negative (Bernoulli) pressure, but its importance is sometimes overstated.

Scherer and Titze[8] and Reed et al.[9] reported a similar pressure waveform from a full larynx, namely an initial peak and a long, round segment. However, the segment of the pre-open phase in their results was different from that shown in Figures 4 and 5. Both groups of investigators showed a slower pressure build-up in this phase. Since previous methodologies consisted of inserting a pressure probe into the glottis, the possibility for leakage flow around the probe existed. Incomplete closure at the top lip may have caused some difficulty in building up of intraglottal pressure, especially around the transducer.

A qualification needs to be made here about the combined video and audio data in Figure 4. Because a strobe technique was used, it was impossible to obtain sequential video frames during one pressure waveform cycle. In order to show the contact area better (without obstruction from the pressure transducer), images without the pressure transducer were obtained that had contact profiles similar to those recorded on the VCR. Thus, the video images in Figure 4 came from observations obtained over more than one cycle. They represent average results for stable phonation.

In falsetto mode, the impact stress and contact area were negligible. The pressure waveform represented only the aerodynamic pressure during phonation (Figure 6). This finding agrees with previous observations[25], in which the edges of the vocal folds did not meet at the midline. There is no major collision between the vocal folds in falsetto mode. One would reason, therefore, that falsetto users would have less incidence of nodules. This is supported by some clinical observations. Although there are no empirical data, a previous study (Wang 1985, personal communication) resulted in the conclusion that the Qing Yi player, one type of female Chinese opera character who uses only falsetto, has much less chance of developing a nodule than Hua Dan, another type of female Chinese opera character who uses chest voice. Further study on this topic needs to be done in the future.

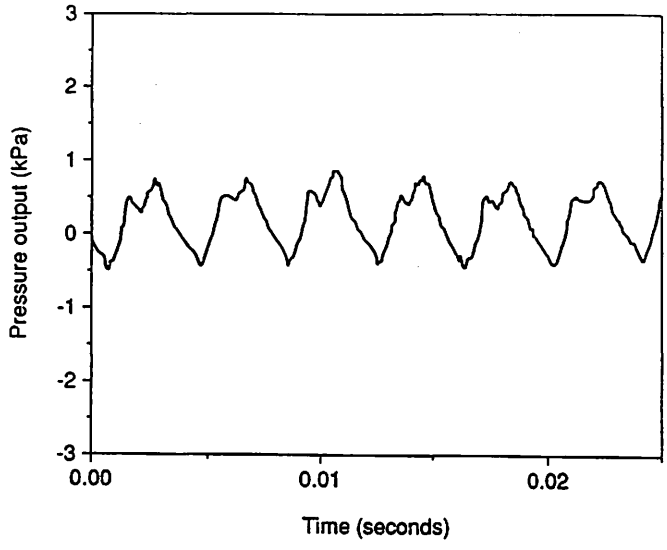
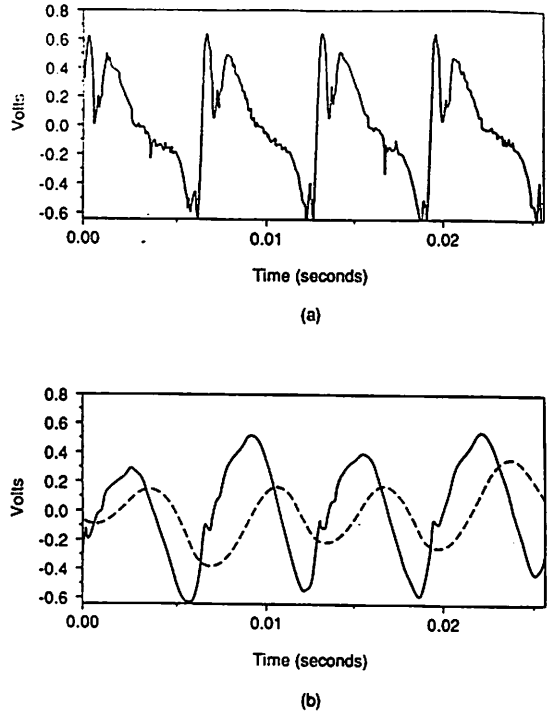


Figure 6. Intraglottal pressure waveform in falsetto voice. The impact stress peak no longer exists. The pressure waveform represents only the aerodynamic pressure during phonation.

Intraglottal Pressure Versus Acoustical Output

Figure 7 shows the intraglottal pressure waveform (part a) and the microphone waveform (part b) displayed simultaneously for typical phonation. Also shown in part (b) is an integration of the microphone signal (dashed line), which approximates the glottal flow. Because the microphone was only 10 cm from the glottis, the phase delay due to acoustic propagation is less than 0.3 millisecond or about 1/20 of a cycle. The acoustic pressure reaches its minimum value just before vocal fold impact and its maximum value during glottal opening. No vocal tract was attached, and hence a typical formant structure is not seen in the microphone signal. Some ripple is apparent, however, in the impact phase. It is possible that vertical tissue displacement causes a pressure fluctuation.



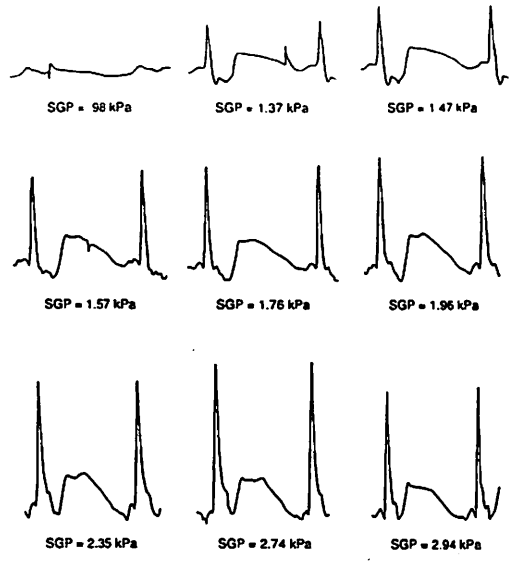


Figure 7 (left). Simultaneous display of (a) the intraglottal pressure waveform and (b) the acoustic microphone waveform (solid line) in normal phonation. The dashed line is the integral of the solid line, approximating the glottal flow. Figure 8 (right). Intraglottal pressure waveforms at the middle point of vocal cord versus subglottal pressure during normal phonation, with the prephonatory setting fixed.

Intraglottal Pressure Versus Subglottal Pressure

Figure 8 shows waveforms of the intraglottal pressure waveform at the midpoint of the vocal fold as a function of mean subglottal pressure during normal phonation, with all the prephonatory settings fixed. Peak impact stress increased as subglottal pressure increased. From these waveforms, Figure 9 was derived, which is a plot of peak impact stress versus mean subglottal pressure for five larynges. The slope, which is dimensionless (kPa/kPa), varied from 1.1 - 3.8, with an average value of 1.78. This is close to the 2.0 value reported by Scherer and Titze[8].

The question arises, should the relation between peak impact stress and mean subglottal pressure be linear? In previous work on excised larynges, the amplitude of vibration of the vocal fold was found to vary roughly as the square root of subglottal pressure at constant vocal fold length[26]. In this study, the relation between amplitude of vibration and peak impact stress was predicted to be linear (equation 6). The combined relation between mean subglottal pressure and peak impact stress should not be expected to be linear, therefore, even though a linear regression line was drawn. More work is needed to tease out the subtitles in variation of A, dx, *and* Δt in equation (6) with mean subglottal pressure. In particular, the impact interval Δt cannot be assumed to be a constant. Nevertheless, our preliminary results support the hypothesis that impact stress is positively related to subglottal pressure and vibrational amplitude.

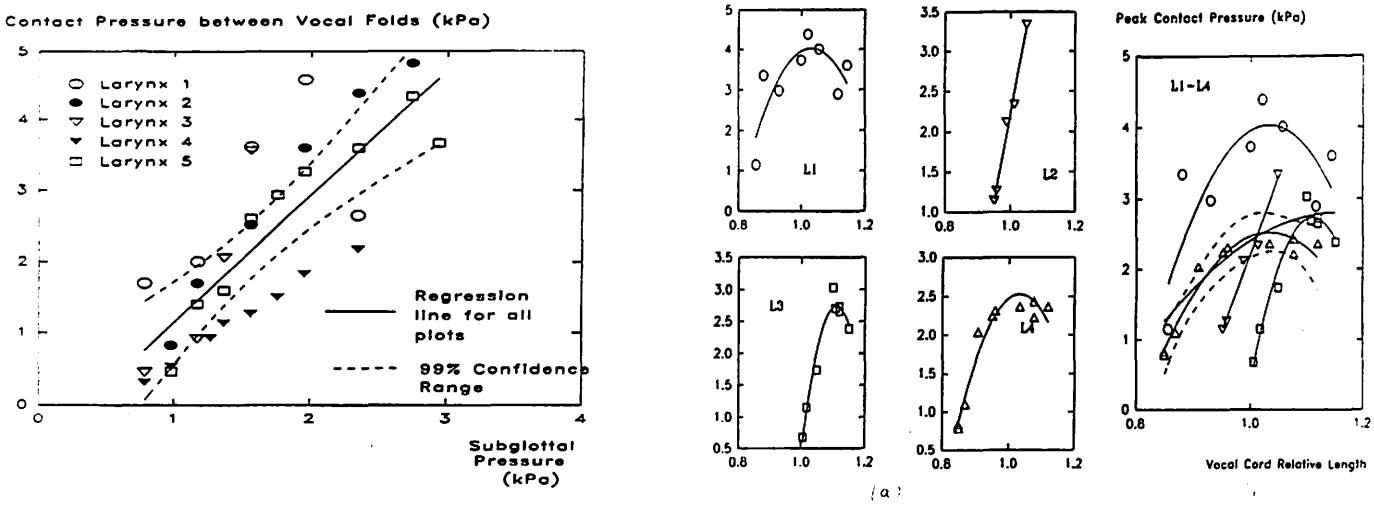


Figure 9 (left). Relation between peak contact stress and subglottal pressure for five hemilarynges. Figure 10 (right). Peak contact stress versus vocal fold elongation during normal phonation with prephonatory setting and subglottal pressure fixed. (a) individual larynges, and (b) composite figure (n=4).

Intraglottal Pressure Versus Elongation of Vocal Fold Elongation

Figure 10 shows peak impact stress for different vocal fold elongations during normal phonation, with adduction and subglottal pressure fixed. In general, vocal fold elongation resulted in an increase in peak stress and a decrease in duration of the peak stress. However, in general, when vocal fold elongation exceeded 5-8%, the peak stress began to decrease again. The effect of vocal fold elongation on the peak impact stress may be a combination of several factors. First, elongation increases the frequency of vibration, thereby increasing the impact force as quantified in equation (6). On the other hand, elongation also reduces the amplitude of vibration, thereby possibly decreasing the impact force. If at some length the amplitude reduction becomes more dramatic than the frequency increase, a turning point would be reached as shown.

Intraglottal Pressure Versus Glottal Width

Figure 11 shows the intraglottal pressure waveform at three different phonation modes: pressed, normal and breathy. These modes were simulated by different glottal widths (shim sizes),

under constant flow and fixed elongation. It appears that pressed voice has the greatest peak impact stress, followed by normal voice and breathy voice, respectively. In Figure 12, the peak impact stress is plotted as a function of shim size (glottal width). Clearly, the less the glottal width, the greater the peak impact stress. It was also observed that if glottal width is extended to a critical point (typically 1.0-1.5mm), the vibration mode often switched to falsetto, with a waveform as shown in Figure 6. Greater elongations usually caused an earlier switch. In contrast, there was no clear switching point between normal voice and pressed voice.

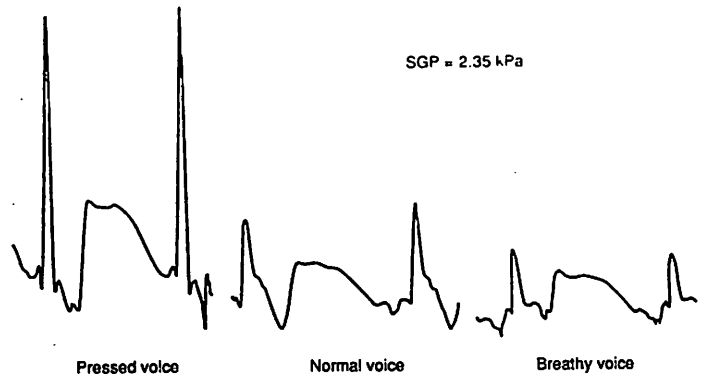
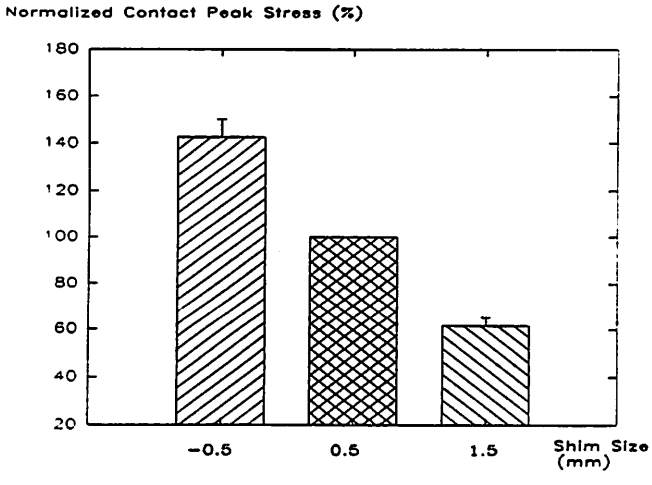


Figure 11. Intraglottal pressure waveform versus phonation type under constant flow and fixed elongation.

Intraglottal Pressure Versus Location on Fold

Peak stresses were measured at three different locations on the vocal fold. Figure 13 shows peak impact stress, normalized to the middle location, for five larynges. In normal phonation, with constant arytenoid approximation, constant vocal fold elongation and constant subglottal pressure, the peak stresses at the anterior and posterior quarter points of the membranous fold were about 62% of the stress at the midpoint. This corresponds relatively well to the 71% value predicted in equation (8), given possible errors in the precise location of vibrational endpoints of the vocal folds and possible deviations from a half-sinusoid model of the lowest string mode.



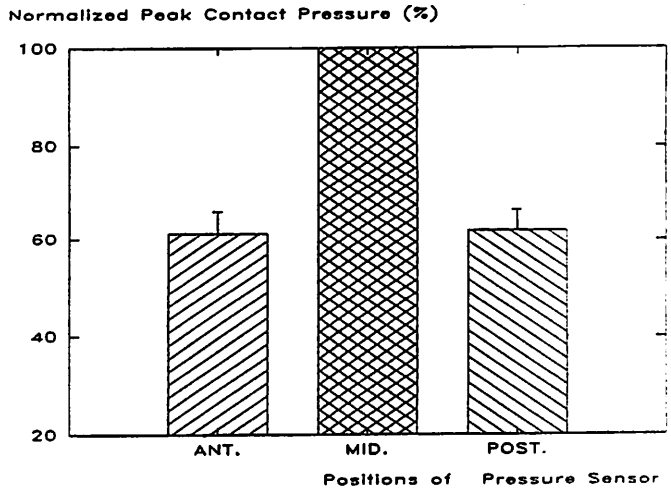


Figure 12 (left). Peak contact stress versus glottal width (shim size) during phonation under fixed glottal flow and elongation (n=9). Figure 13 (right). Normalized peak stress versus anterior-posterior location on vocal fold. The peak stress at the anterior and posterior 1/4 points of the vocal fold is about 62% of that at the midpoint (n=5).

Conclusions, Applications, and Limitations

The excised larynx offers a convenient method for investigation of impact stress in phonation because it allows for relatively easy manipulation of parameters. Based on experience gained with excised larynges in general and specific results of this study, it is appropriate to make some suggestions regarding clinical applications and future research.

Suggestions for Voice Abuse and Etiology of the Vocal Nodule

According to the present results, higher subglottal pressure, closer distance between arytenoid cartilages, and greater vocal fold elongation are independently and positively correlated with peak impact stress during phonation. It is also known that subglottal pressure is positively correlated with vocal intensity[27], elongation with pitch[28], and distance between arytenoid cartilages with "pressing" of the voice[29]. Most clinicians assert that loud voice, hyperadduction, and high pitch constitute potentially injurious abuse and misuse of the voice. Therefore, peak impact stress should correspond to vocal abuse and misuse. The present data support the mechanical trauma hypothesis of vocal abuse and the vocal nodule etiology.

It is important to determine a threshold value of impact stress that causes trauma. The measured peak stress was 2-4 kPa in normal phonation. Subglottal pressure, vocal fold elongation, and arytenoid approximation were all found to independently increase the impact stress. It is presumed (and clinically hypothesized) that the combination of a loud, strained, and relatively high pitched voice could generate excessive impact stresses during phonation. These high impact stresses acting on a delicate vocal fold mucosa, especially a low stress-resistant mucosa, or a predamaged mucosa, could possibly produce trauma and other tissue reactions, and finally vocal nodules.

The mechanical trauma hypothesis is also supported by some histological evidence. Kleinsasser[30] found that tissue regeneration is a major histological finding in vocal nodules. According to Gillman[31], regeneration means replacement of destructed tissues by connective tissue. Gillman divides the regeneration after an acute trauma into five chronological phases. Because the characteristics of all of the five regeneration phases can be found in the vocal nodule, Sonnien[15] concluded that traumatic influences extend over a long period of time. Because regeneration occurs after trauma, the implication is that the vocal nodule is caused by trauma. The observation of microstructural changes of an hyper-phonated vocal fold[32] also supports the observation that vocal fold trauma occurs after acute phonation.

The match between locations of nodules[33,10] with the location of the maximum impact force also supports the mechanical trauma hypothesis. Both occur at the middle of the membranous fold. Some nodules can be eliminated by speech therapy. An important aspect of such therapy is to direct patients to talk with less intensity, less adduction, and to use appropriate pitch. Presumably, the effect is to reduce the impact stress to the midportion of the vocal folds.

Limitation of Using the Canine Larynx

The purpose of the present experiments was to address questions related to human phonation. Canine larynges were used, however. Gross similarities between canine and human larynges allow the canine larynx to be used as a model for studying basic passive mechanisms of human phonation [34,35,36,21,37,38]. In addition to their similarity to human larynges, canine larynges are used because of their availability from other experimental protocols, requiring no additional sacrifice of vertebrate animals. However, there are some differences between the canine larynx and human larynx[36,38].

The most severe limitation is the absence of a vocal ligament in the canine. This difference alone might affect the generalizations made here. In further studies, human larynges might be considered.

Limitation of Using A Single Sensor

As predicted by Ishizaka & Matsudaira[39] Titze[40,26], intraglottal pressure varies spatially. During our experiments, stress was measured at three different positions by moving a single sensor. This manipulation could cause measurement error. A better solution would be to build an array of sensors to map out pressure distributions in the glottis. Construction details and cost were prohibitive for this experiment. In the future, however an array could be used to investigate strategies for adjustment of pitch and loudness that yield minimum stress patterns on the colliding tissue.

Acknowledgements

This work was supported, in part, by grant No. P60 DC00976 from the National Institutes of Health. Assistance in manuscript preparation and graphics was provided by Julie Lemke and Mark Peters, respectively.

References

1. Jiang, J., & Titze, I. (in review). A methodological study of hemilaryngeal phonation.

2. Cooper, D.S. (1986). Research in laryngeal physiology with excised larynges. In: Cummings, C.W. (Ed.). Otolaryngology-Head and Neck Surgery, Vol. 3. St. Louis: C.V. Mosby, 1766-1776.

3. Harless E. Die Stimme (1853). In: Wagner, R., Ed. Handworterbuch der Physiologie, Vol. 4. Braunschweig: F. Vieweg und Sohn.

4. Weiss, O. Uber die Entstehung der Vokale. II. (1914). Die Vorgange im ausgeschnittenen Kehlkopf. Arch Exp Klin Phon, 1:350-5.

5. Scherer R.C., Cooper D, Alipour F., & Titze I.R. (1985). Contact pressure between the vocal processes of an excised bovine larynx. In: I.R. Titze and R.C. Scherer (Eds.) Vocal Fold Physiology, The Denver Center for the Performing Arts, Denver, Co. pp. 292-303

6. Murakami, Y., & Kirchner, J.A. (1971). Reflex tensor mechanism of the larynx by external laryngeal muscles. Annals of Otology, Rhinology and Otolaryngology, 80, 46-64.

7. Murakami, Y., & Kirchner, J.A. (1972). Mechanical physiological properties of reflex laryngeal closure. Annals of Otology, Rhinology and Otolaryngology, 81, 59-71.

8. Scherer, R.C., & Titze, I.R. (1982, April). Vocal Fold Contact Stress During Phonation. Paper presented to the Acoustical Society of America, Chicago.

9. Reed, C., Doherty, T. & Shipp, T. (1990). Piezoelectric transducers for the direct measurement of medial forces in the glottis. Journal of the Acoustical Society of America, S1, 88, S151.

10. Nagata, K., Kurita, S., Yasumoto, S., Maeda, T., Kawasaki, H. & Hirano, M. (1983) Vocal Fold Polyps and Nodules, a 10 Years Review of 1156 Patients, Auris, Nasus, Larynx, Suppl. 10, S27-S35.

11. Brodnitz, F.S., & Froeschels, M.D. (1954). Treatment of nodules of vocal fold by chewing method. Archives of Otolaryngology, 191, 361-363.

12. Lancer, J.M., Syder, D., Jones, A.S., & Boutiller, A.L. (1988). Vocal fold nodules, A review. Clinical Otolaryngology, 13, 43-51.

13. Kambic, V., Radsel, Z., Zargi, M., & Acko, M. (1982). Vocal cord polyps: Incidence, histology and pathogenesis. Journal of Laryngology & Otology, 95, 609-618.

14. Epstein, S., Winston, P., Friedmann, I., & Ormerod, F.C. (1957). The vocal fold polyp. Journal of Laryngology & Otolaryngology, 71, 673-688.

15. Sonnien, A., Damste, P.H., Jol, J., & Fokkens, J. (1972). On vocal strain. Folia Phonatrica, 24, 321-336.

16. Arnold, G.E. (1962). Vocal nodules and polyps: Laryngeal tissue reaction to habitual hyperkinetic dysphonia. Journal of Speech and Hearing Disorders, 27, 205-217.

17. Arnold, G.E. (1963). Vocal nodules. In: Voice Problems and Laryngeal Pathology, by J. F. Daly (Moderator). New York State Journal of Medicine, 63, 3096-3110.

18. Vaughan, C.W. (1982). Current concepts in otolaryngology: Diagnosis and treatment of organic voice disorders. N. England Journal of Medicine, 307, 863-866.

19. Gery, P. (1973). Microlaryngostroboscopy and "singer's nodes". J. Otolaryngol. Sol. Aust. #: 525-527.

20. Jiang, J. (1985). Etiology of singer's nodules. A report of EENT Hospital, Shanghai Medical University, Shanghai, China.

21. Hirano, M. (1975). Phonosurgery: Basic and clinical investigations. Otologia (Fukuoka), 21, 239-240.

22. Titze, I.R. (1988). A framework of the study of vocal registers. Journal of Voice, 2, 183.

23. Saito, S., Fukuda, H., Kitahira, S., Isogai, Y., Tsuzuki, T., Muta, H., Takyama, E., Fujika, T., Kokawa, N., and Makino, K. (1985). Pellet tracking in the vocal fold while phonating experimental study using canine larynges with muscle activity. In: Vocal Fold Physiology, I.R. Titze and R.C. Scherer (Eds.), Denver Center for the Performing Arts, Denver, Co.

24. Titze, I.R. (1984). Parameterization of the glottal area, glottal flow, and vocal fold contact area. Journal of the Acoustical Society of America, 75, 570-580.

25. Rubin, H.J., & Hirt, C.C. (1960). The falsetto, A high-speed cinematographic study. Laryngoscope, 70, 1305-1324.

26. Titze, I.R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. Journal of the Acoustical Society of America, 85, 901-906.

27. Ishikki, N. (1964). Regulatory mechanism of voice intensity variation. J. Speech Hear. Res., 7, 17-29.

28. Hollien, H. (1962). The relationship of vocal fold length to vocal pitch for female subjects. The Twelfth International Speech and Voice Therapy Conference.

29. Scherer, R. and Vail, V. (1988). Measures of laryngeal adduction. J. Acoust. Soc. Amer. 84(S1), S581(a).

30. Kleinsasser, O. (1982). Pathogenesis of vocal fold polyps. Annals of Otology, Rhinology and Laryngology, 91, 378.

31. Gillman, T., (1968). Treatise on collagen, Vol. 2, In: GOULD Biology of Collagen, Academic Press, New York, 334.

32. Gray, S.D., Titze, I.R., & Lusk, R.P. (1987). Electron microscopy of hyperphonated canine vocal folds. Journal of Voice, 1(1), 109-115.

33. Rubin, H.G., & Lehrhoff, I. (1962). Pathogenesis and treatment of vocal nodules. Journal of Speech and Hearing Disorders, 27, 150-161.

34. Koyama, T., Harvey, J.E., & Ogura, J.H. (1971). Mechanism of voice production II. Regulation of pitch. Laryngoscope, 81, 47-65.

35. Koyama, T., Kawasaki, M., & Ogura, J.H. (1969). Mechanism of voice production I. Regulation of voice intensity. Laryngoscope, 79, 337-354.

36. Baer, T. (1975). Investigation of Phonation Using Excised Larynges. Ph.D. dissertation, Massachusetts Institute of Technology, Boston, Massachusetts.

37. Hirano, M. (1977). Structure and vibratory behavior of the vocal folds. In: M. Sawashima & F. Cooper (Eds.), Dynamic Aspects of Speech Production, (pp. 13-27). Tokyo: University of Tokyo Press.

38. Kirita, S., Nagata, K., & Hirano, M. (1981). A comparative study of the vocal folds. In: D.M. Bless & J.H. Abbs (Eds.) Vocal Fold Physiology, (pp. 3-21). San Diego, CA:College Hill Press.

39. Ishizaka, K., & Matsudaira, M. (1972). Fluid mechanical consideration of vocal fold vibration, Monograph 8, Speech Communication Research Laboratory, Santa Barbara, CA.

40. Titze, I.R. (1980). Comments on the myoelastic-aerodynamic theory of phonation. Journal of Speech and Hearing Research, 23(3), 495-510.

NCVS Status and Progress Report - 3
December 1992, 39-63

Evidence of Chaos in Vocal Fold Vibration

Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
R. J. Baken, Ph.D.
Department of Speech and Language Pathology and Audiology, Columbia University
Hanspeter Herzel, Ph.D.
Department of Physics, Humboldt University, Berlin, Germany

...And the earth was chaos and void.
Genesis I

The classical view of the natural world--which has strongly biased our perceptions of how things are and how they are supposed to behave--inclines us to categorize physical phenomena into two broad classes. Some processes are apparently regular, predictable, and (at least in some metaphorical sense) smooth. Others, however, are described as unstable, erratic, unpredictable, or random. Much of what we understand of nature concerns regular, "lawful" events, simply because they are amenable to relatively easy description and to classical modeling. Many important phenomena fall into the unstable/erratic class, however (fluid turbulence is a prime example). They have proven to be highly refractory, primarily due to the lack of an adequate conceptual base for structuring their investigation. Often they have been regarded simply as "noise". Thus we have had an overly simplistic dichotomy between predictability (determinism) and unpredictability (randomness).

Another view of the physical world has its origins in the seminal papers on topology by Poincaré (1881, 1882, 1885). The ability to simulate natural processes (with powerful digital computers) has resulted in the discovery that systems governed by simple mathematical laws can behave in a highly irregular (unpredictable) fashion. Although the response of such systems is *determined* entirely by internal properties and initial conditions, *predictability* over the long range is limited. The behavior can appear chaotic. Analysis of systems of this type has fallen under the

rubric of nonlinear dynamics. It uses geometric methods to move beyond simple time-course descriptions of a phenomenon, sketching a portrait of the system's overall dynamic properties. The portrait can in turn be dissected to reveal crucial bases of the system's behavior. We begin with some observations that relate to vocal fold vibration.

Chaos, Fractals and Vocal Fold Vibration

In vocal fold vibration, the displacement of isolated fleshpoints reveals a complex pattern of motion that varies from locus to locus (Baer, 1981a,b). Such motion can be simulated by computer (Titze and Alipour, in review). In Figure 4-1, trajectories of "fleshpoints" of a finite element simulation of vocal fold movement are shown (right side). There is an orderliness in the fleshpoint movements. After a start-up transient, the orbits are primarily elliptical. In Figure 4-2, however the trajectories are very disorderly, covering virtually every conceivable position. Not much has been changed in the model to go from one condition to the other. Only the Young's modulus of the cover has been reduced.

Tools of analysis of nonlinear dynamics may allow us to summarize many of the dynamic properties of the vocal folds in pictures that are comprehensible at a glance. Before proceeding with further analysis of vocal fold vibration, however, a simpler example is instructive.

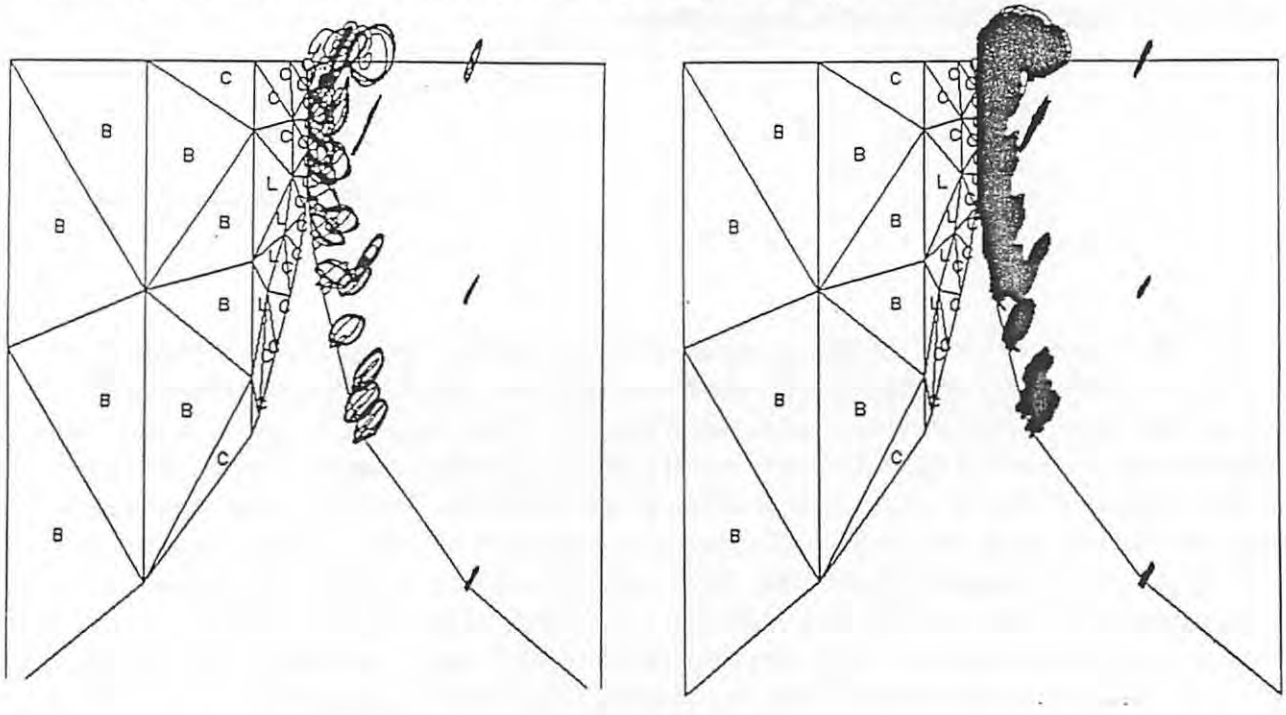


Figure 4-1 (left). Frontal section through vocal folds of a finite element computer simulation model. Left side shows tissue elements for body (B), ligament (L), and cover (C). Right side shows trajectories of nodal points during self-oscillation. Figure 4-2 (right). Same as Figure 4-1, but the Young's modulus of the cover was reduced from 1.5 KPa to 0.5 KPa.

A Simple Example of Chaos

Perhaps the best way of demonstrating what is at issue is to use an example that has become very familiar in the literature of nonlinear dynamics theory. Consider a recursive quadratic function

that generates sequences of numbers x_n by the relation

$$x_n = r(x_{n-1})(1-x_{n-1}), \tag{4-1}$$

This is known as the *logistic equation.*[1] Let the parameter r be set at 2.0. If we choose an arbitrary starting value for $x_{n-1} = 0.20$, we find that the following series of output values results: 0.320, 0.435, 0.492, 0.499, 0.500, 0.500... The recursive series converges relatively quickly on a final value. When the parameter r is set to 2.5, starting with $x_{n-1} = 0.20$ yields the convergent series 0.400, 0.600, 0.600... Such convergence suggests that the system modeled by the function could be described as "stable".

Consider, however, what happens when r is set at 3.4. The output series becomes 0.544, 0.843, 0.449...0.842, 0.452, 0.842, 0.452, 0.842, 0.452, 0.842... The output thus oscillates between two final values. If $r = 3.5$, the series converges to 0.382, 0.827, 0.501, 0.875, 0.382, 0.827, 0.501, 0.875..., an oscillation among 4 final values. Finally, when $r = 3.8$, the output is 0.789, 0.633, 0.883, 0.392, 0.906... without any suggestion of convergence. The very complex behavior of this simple recursive model is plotted as a function of the parameter r in Figure 4-3. Note that for part of the domain over which r varies, the function converges to a single value. As r increases, however, the output alternates between two stable values (for example, at r=3.25), or four (at r=3.5) or 8 or more.

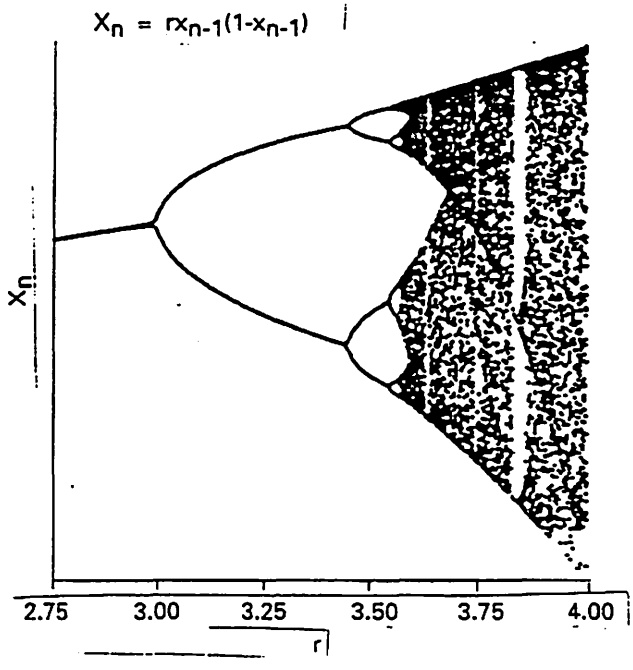


Figure 4-3. Bifurcation diagram for the logistic equation.

For $r > 3.57$ stability seems to be lost. But the points in this region show a faint patterning, suggesting some underlying order. And an enlargement of part of this domain (Figure 4-4) shows occasional zones where stability is clearly present (as at $r \approx 3.84$). Furthermore, it can be shown that an increasing enlargement of any given region in the domain $3.57 < r < 4.0$ will show increasing numbers of such zones of stability. If the logistic equation were a model of vocal fold function, we might say that some adjustments produced stable phonation, some produce diplophonia, some "quadriphonia" or other stable but peculiar F_0 trains, and some adjustments produce aperiodicity.

[1] The logistic equation was originally proposed as a model of insect population changes, and has since been extensively analyzed by May (1974, 1976, 1980).

The disorganized regions are themselves quite interesting. It turns out that they are not nearly as unstructured as they seem. If, as shown in Figure 4-5, we plot just a few of the successive outputs of the recursion at a given r in the region of disorganization on the coordinates x_n, x_{n-1} (which define a "pseudo phase space") we find that they fall into a very discernible pattern. (Sequential points do not lie next to each other, however. They are scattered; only in the aggregate do they describe a geometric structure. In other words, it is not possible to connect successive points with lines and preserve the pattern shown.) The geometric shape that is formed is an *attractor*. It is, in some sense, a picture of the ordering principle that underlies the seemingly random sequence of outputs of the equation.

The logistic equation represents a purely deterministic system that has unpredictable outputs. But even the erratic outputs are clearly not random, as the attractor shows. The system is said (in the technical sense) to be *chaotic*. Any system is potentially chaotic if:

(a)it is deterministic and nonlinear;

(b)it shows sudden qualitative changes in its output, that is, it demonstrates *bifurcations* such as those shown in Figure 4-3 (the system may exhibit, at certain parameter values, nonperiodic behavior that appears as intermittant bursts of irregularity [transients], or the output may be exquisitely dependent on initial conditions and tiny parameter changes. For example, infinitesimal differences in r of the logistic equation may result in enormous differences in the regularity of the output);

(c)its representation in phase space shows fractal properties.

The essential properties of chaotic systems are discussed in a large number of texts that have appeared in the last few years. Among the better sources are Bergé, Pomeau & Vidal (1984), Moon (1987), Thompson and Stewart (1986) and Rasband (1990).

Criterion (c) above introduces a new concept --*fractal dimension*-- that has been shown to be of importance in understanding both structures and processes of the natural world. It will require a bit more explanation.

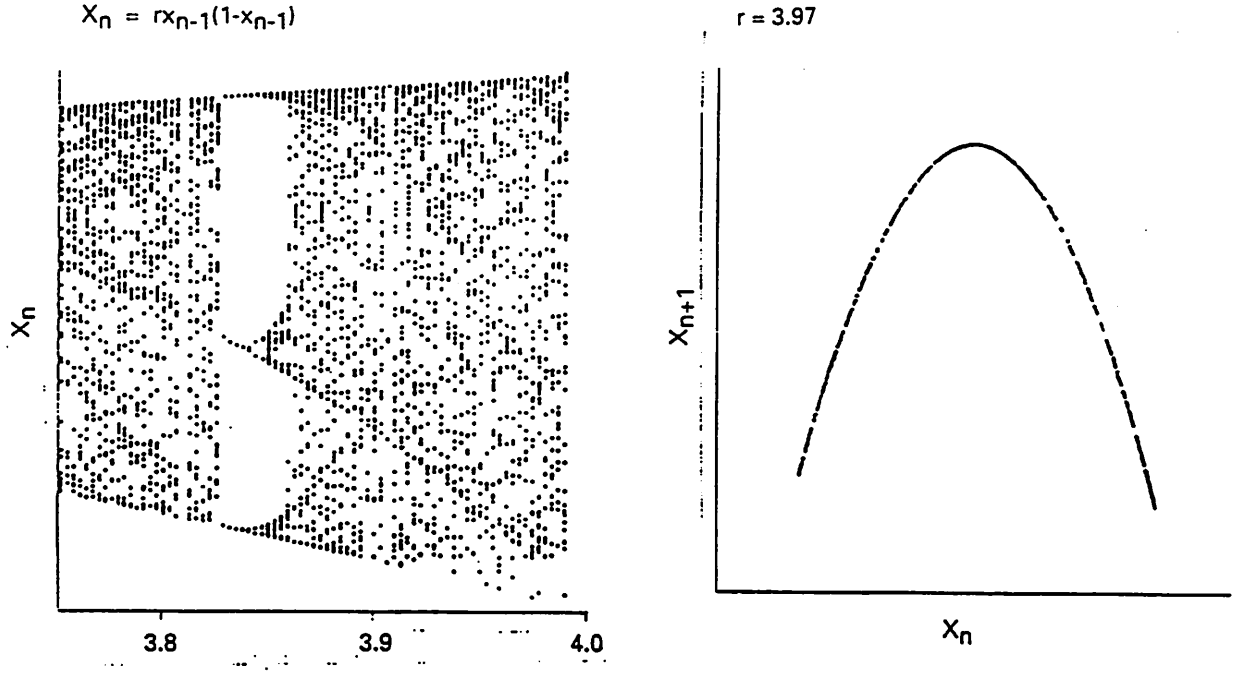


Figure 4-4 (left). Enlargement of a portion of the bifurcation diagram of Figure 4-3. Figure 4-5 (right). Attractor for the logistic equation.

Fractal Dimensions

Classical geometry limits dimensions to integer values. A point has dimension 0, a line 1, a plane 2, and so forth. The problem, however, has been that many geometric structures are not effectively described within those integer values. A classic example explored by Mandelbrot (1967) --who is largely responsible for the creation of the field of fractal geometry-- is the natural coastline of a country. Euclidean geometry holds that it is a highly irregular line (and hence its dimension = 1). In classical mathematics, this irregularity is made tractable by assuming that if any segment were sufficiently enlarged, straight line segments would appear. But this is clearly not the case for a coastline (or many other structures). An enlargement of a bay (an irregularity) shows smaller bays along its margins. These in turn have rocks and boulders that stick out to make the bays' edges erratic. The rocks themselves are jagged, and every jagged area is formed by the mineral grains of which the rock is made. The grains, in turn, are composed of aggregated crystals,. . . and so on. At every scale of magnification there is a similar degree of irregularity. A coastline is an example of a geometric structure that is "self similar at all scalings". If one tries to measure such a line, one finds that the length obtained depends on the size of the ruler used. Longer rulers follow fewer irregularities, and hence give a shorter length than do smaller rulers than can follow the line more closely. The result of this, as Mandelbrot demonstrates, is that the dimension of the coastline must be greater than 1. However, its dimension cannot be as much as 2, which is the dimension of a plane. Hence, the dimension of such a line must be *fractional*, and the line itself is said to be *fractal*. Any geometric structure (of any number of Euclidean dimensions) that shows such self-similarity is fractal. An entire geometry of fractality has been developed in the last several years (Barnsley, 1988).

There are a number of ways to estimate the fractal dimension of natural structures (Barnsley, 1988; Farmer, Ott, and Yorke, 1983; Froehling et al., 1981; Farmer, 1982; Grassberger, 1986). What that dimension represents is the degree of the structure's irregularity. Irregularity is different from variability. For example, the completely regular function $F(x) = \sin x$ has a very distinct variability, which could be expressed as the standard deviation of $F(x)$. But a plot of the points $F(x)$ is not at all self-similar and, on measurement, its dimension will be found to be 1, indicating that it really is a line.

The combination of nonlinear dynamics and fractal geometry has considerable potential for explaining many biological phenomena, a number of which are of particular importance in physiology or medicine. It has been found, for example, that strange attractors are associated with electroencephalographic activity and that they change, and their associated dimension is altered, in response to alterations of consciousness level, cognitive activity of the brain or the presentation of stimuli (Rapp et al., 1990; Samar and Rosenberg, 1990; Freeman, 1990). Chaotic behavior has also been observed in the ventilatory system (Mackey and Glass, 1977) and in the heart rhythm (Glass, Shrier, and Bélair, 1986; Goldberger, West, and Bhargava, 1986; Goldberger and Rigney, 1990).

Where there are fractal dimensions, there is the possibility of chaotic behavior. Glass and Mackey (1988), after a review of potentially or demonstrably chaotic functions in a number of abnormal physiological conditions, have proposed that disorders in which normal organization is disrupted and replaced by abnormal dynamics, such as Cheyne-Stokes breathing, certain cardiac arrhythmias, and epilepsy, may be most appropriately categorized as "dynamical diseases".

Recent Discoveries in Voice Source Characteristics

The vocal folds, together with glottal airflow, constitute a nonlinear oscillating system. This system is amenable to nonlinear dynamics analysis, which can provide an entirely different perspective on the underlying basis of phonatory function. This effort has just barely begun. Notable contributions are those of Pickover and Khorsani, (1986); Awrejcewicz, (1991); Wong, Ito, Cox and Titze, (1990); Mende, Herzel, and Wermke, (1990); and Herzel, Steinecke, Mende, and Wermke, (1991). The objective of these was to certify the applicability of nonlinear systems theory to vocal functioning and to derive initial suggestions about the dynamical nature of the system. Unfortunately, time series data (that is, data that come from a real physical system) are notoriously difficult to deal with. They inevitably include stochastic functions and cannot be made immune to the influences of contaminating variables. Thus, the work goes slowly and definitive answers even to the simple questions concerning applicability are not yet available. The findings already at hand, however, afford veiled glimpses of the value of a nonlinear dynamics approach to the voice. Of specific significance was the early modeling study by Isshiki and Ishizaka (1976), who showed that asymmetry in vocal fold vibration yielded subharmonic structure.

Herzel and his coworkers (Herzel et al. 1991; Mende, Herzel, and Wermke, 1990) have undertaken preliminary examination of the cries of newborn infants and have found numerous bifurcations in the sound pressure signal, including period doublings and sudden transitions to aperiodicity. Their analysis of these phenomena strongly suggest the presence of low-dimensional chaos, and they have established the consistency of these phenomena with two-mass models of the vocal folds. The more intensive modelling experience of Awrejcewicz (1990) has examined the trajectory of the vocal folds and supports the findings of Herzel et al. (1991). It would appear, then, that the vocal folds are, in principle, capable of chaotic behavior.

Baken (1990) has undertaken a preliminary examination of the fractal dimension (D_F) of normal vocal F_0 and amplitude. Using a box-counting algorithm for estimation of D_F, he has found that a data record of sequential period values had a D_F of 1.46, which was unaffected by speaker sex or by the mean F_0 of phonation. Records of sequential peak intensities of the same phonations had a significantly different mean D_F of 1.54, which also was unaffected by speaker sex or by mean vocal F_0. D_F was at best weakly correlated to measures of vocal variability such as relative average perturbation or shimmer. This work demonstrated that fractal geometry can be useful for measuring irregularity of vocal fold oscillation, independent of oscillatory variability (which, of course, can be the result of organized behavior).

Exploratory, and somewhat informal, examination of vocal F_0 and amplitude in cases of several different types of vocal disorder undertaken by Baken (in press) have demonstrated that the fractal dimension is often different from normal, but the differences do not seem related to traditional categories of laryngeal disorder. In fact, this result is not surprising: dynamical theory suggests that a different taxonomy of disorder may be necessary.

But is the vocal fold system potentially chaotic? In principle, it should be. Research now in progress should show whether it is in practice. There are, in the meantime, some suggestions that chaotic behavior does appear in certain circumstances. The baby-cry studies of Herzel and his coworkers certainly point in that direction, although it could be argued that the infant larynx and nervous system do not provide a suitable model for adult vocal production. Perhaps more telling is an investigation by Baken, Watson, and Dembowski, (in preparation) of F_0 variation in spasmodic dysphonic patients. It was designed as the proverbial "fishing expedition", based on the intuitive idea that interesting things might be happening to vocal F_0 in these patients. Recordings of several

male and female spasmodic dysphonic patients were drawn from the extensive collection that has been built up at the University of Texas at Dallas. Fundamental period values for all glottal cycles during sustained vowels were extracted using CSpeech, and a contour of these periods was prepared. Figure 4-6(a) is typical of one of the records that resulted.

What is interesting about this record is that it demonstrates period doubling bifurcations. Note that, at the start of the trial, successive periods are relatively stable, but after about 50 periods the contour oscillates between two values. This is seen in an expanded plot in Figure 4-6(b). Toward the end of the phonation another bifurcation occurs, and the period oscillates between 4 values (Figure 4-6c). These bifurcations are extremely abrupt, the transition from one state to another requiring only a few milliseconds. Another record (Figure 4-7) shows a more complex bifurcation, as the period undergoes a transient change from 2-cycle alternation to a more complicated pattern involving 8-10 periods (see expanded portion in part b).

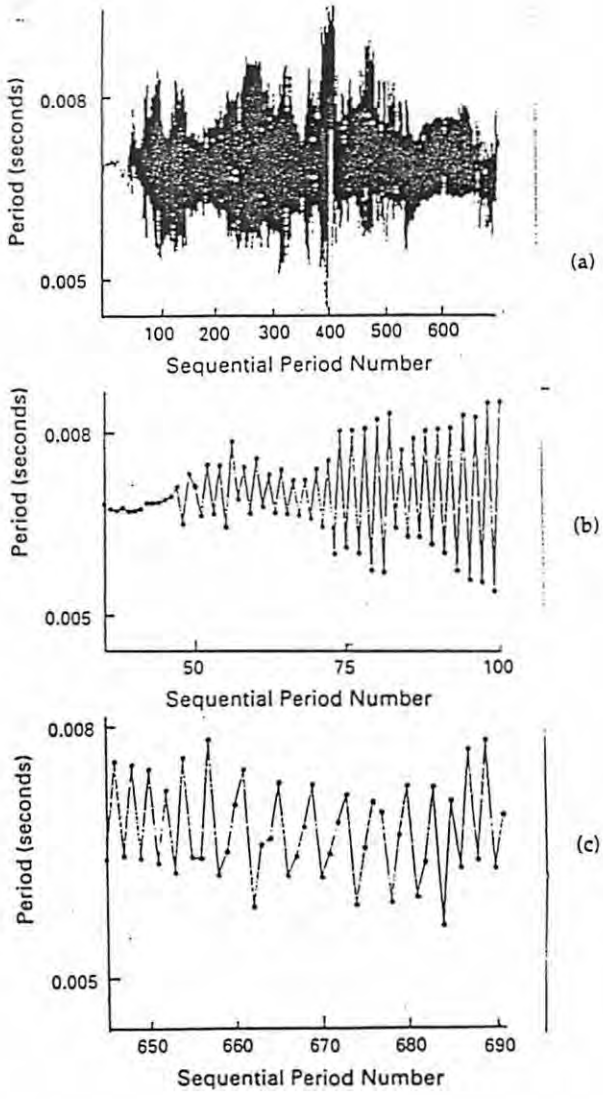


Figure 4-6. Fundamental period of vocal fold oscillation plotted as a function of period number. Spasmodic patient (a) full utterance, (b) expanded segment showing period doubling, (c) expanded segment showing period doubling changing to period quadrupling.

At this point it is appropriate to reflect on the possible contributors to irregularity in vocal fold vibration. Without attempting to sort out which contributors leads to which kind of pattern, we simply list a number of possible candidates.

1. Unsteadiness in muscle contractions in the laryngeal and respiratory system. In particular, the incomplete summation of muscle twitches in an attempt to form a "smooth tetanus" brings about a fundamental frequency jitter (Baer, 1981b; more recently, the process has been modeled by Titze, 1991).

2. Turbulence in the glottal airstream.

3. Vortex shedding and instability in the jet emerging from the glottis (this differs from the turbulence above). The jet may flip-flop from side to side, even if turbulence does not exist.

4. Asymmetry in the mechanical or geometrical properties of the two vocal folds. Usually, a dominant oscillation mode exists due to synchronization of two similar oscillators by the airflow, but excessive asymmetry may create desynchronization.

5. Nonlinearity in the mechanical properties of vocal fold tissues (the constitutive equation) and the pressure-flow relations. Nonlinearities complicate the mode structure of a vibrating system.

6. Coupling between the vocal folds and the vocal tract. Acoustic pressures in the subglottal and supraglottal region may play a part in driving the vocal folds. If these pressures change dynamically, oscillation may be perturbed.

7. Mucus riding on the surface of vocal fold tissue. The mucus could reorient itself from cycle to cycle, causing disturbances in the vibration pattern.

Several of these sources of irregularity can exist in combination with others. Some of them, like mucus and air turbulence, may result in high-dimensionality chaos; others, like left-right asymmetry, may lead to low-dimensionality chaos. It is important to study the sources one at a time to get a better understanding of their effect on vocal fold vibration. Some of this work can be done with physical models and excised larynges, where the effects of neural inputs, the vocal tract, or mucus can be selectively eliminated.

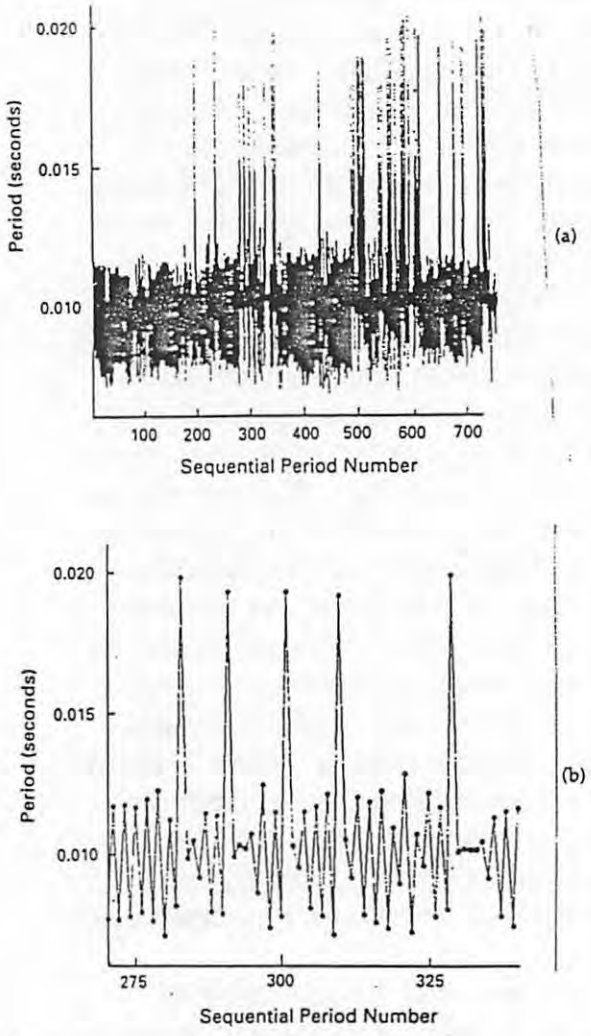


Figure 4-7. Fundamental period of vocal fold oscillation plotted as a function of period number. Spasmodic patient (a) full utterance, (b) expanded segment showing complex patterns involving multiple periods.

In the following section, some basic principles of nonlinear dynamics will be reviewed that will be helpful in interpretations of data obtained on humans, excised larynges, and computer simulation models.

Introduction to Nonlinear Dynamics

A *linear* dynamical system is described by equations of the form

$$\dot{x}_1 = f_1(x_1, x_2 \ldots x_n, \dot{x}_2, \dot{x}_3 \ldots \dot{x}_n, t) \tag{4-2}$$

$$\dot{x}_n = f_n(x_1, x_2 \ldots x_n, \dot{x}_1, \dot{x}_2 \ldots \dot{x}_{n-1}, t) \tag{4-3}$$

where $x_1, x_2 ... x_n$ are components of displacement or velocity and the dots indicate time differentiation. All of the displacements and all of the velocities are regarded as separate variables in time. Solution of these coupled equations is usually obtained by assuming complex exponential functions for x_i, finding the eigenvalues (which determine the characteristic frequencies of the system), and using the principle of superposition to express the total response as a sum of eigenresponses (transients) and driven responses (steady states).

For example, the dynamical equations for a second order mass-spring-damper system are written as

$$\dot{x}_1 = x_2 \tag{4-4}$$

$$\dot{x}_2 = [-b\dot{x}_1 - kx_1 + F(t)]/m \ , \tag{4-5}$$

where x_1 is displacement from equilibrium, x_2 is velocity, b is the damping coefficient, k is spring stiffness, m is mass, and $F(t)$ is the driving force. Solution of these equations is of the form

$$x_1 = Ae^{\gamma_1 t} + Be^{\gamma_2 t} + G(t) \ , \tag{4-6}$$

where A and B are constants (to be evaluated by initial conditions) and γ_1 and γ_2 are eigenvalues of the characteristic equation [found from 4-4 and 4-5 by setting F=0). The function $G(t)$ is the steady-state solution.

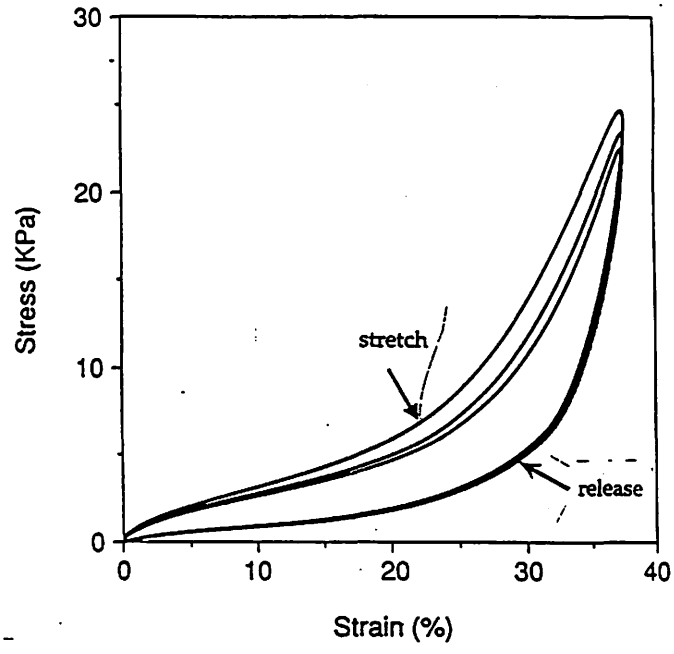
In *nonlinear* dynamical systems, the variables x_i and $\dot{x}_i$ in equations (4-2) and (4-3) form product terms (*e.g.*, $x_2\dot{x}_3$) or contain higher order terms (e.g., x_1^2). Solutions are much less straightforward and must usually be obtained numerically. More importantly, the nonlinear coupling between displacements and velocities makes the principle of superposition no longer applicable: *a large-signal response is not simply a scaled-up version of a small-signal response.* Different excitations may result in totally different time-courses, described by totally different mathematical functions. The system may be very sensitive to small changes in a parameter (such as a spring constant k or a damping coefficient b), or to an initial condition. Recall that this was demonstrated for the logistic equation (4-1) with the parameter r and with the starting value of x_{n-1}. [Equation 4-1 is a nonlinear differential equation written in discrete time steps x_n rather than in continuous form $x(t)$].

Nonlinearity in Vocal Fold Mechanics

Nonlinearity is exhibited in vocal fold mechanics in at least two ways. First, tissue deformation does not follow a simple Hooke's law for elastic springs. The restoring force $-kx_1$ in equation (4-5) is usually augmented by a quadratic or cubic function (involving x_1^2 or x_1^3 terms) in simple vocal fold models, or by an exponential term $e^{\alpha x}$ in more complex models. Second, the pressure-flow relation in the glottis in nonlinear. For steady flow and steady glottal configurations, the flow varies with the square root of the transglottal pressure. For oscillatory flow, the relation may become somewhat "linearized", but only over limited ranges (to be discussed later).

Consider first the mechanical properties of vocal fold tissues in more detail. A universal property of soft biological tissues is that the stress-strain curve is not only nonlinear, but also time-

dependent. There is no unique relation between mechanical load and tissue deformation unless a history of the deformation is included. Thus, if one step-elongates a piece of tissue to achieve a specific tension, the tension will not remain constant, but will relax with time. The combined nonlinear and time-dependent characteristics can be observed in a stretch-release experiment. Figure 4-8 shows results of such an experiment for the thyroarytenoid muscle. Three cycles of stretch and release, which occurred at a rate of 1 Hz, are shown. Note that the tissue exhibits less stress during release than during stretch at the same value of strain. This suggests that stress "leaks out" during the cycle. Not all the stored energy is given back. From a Carnot cycle point of view, the area inside the stretch-release loop represents the energy lost per cycle.



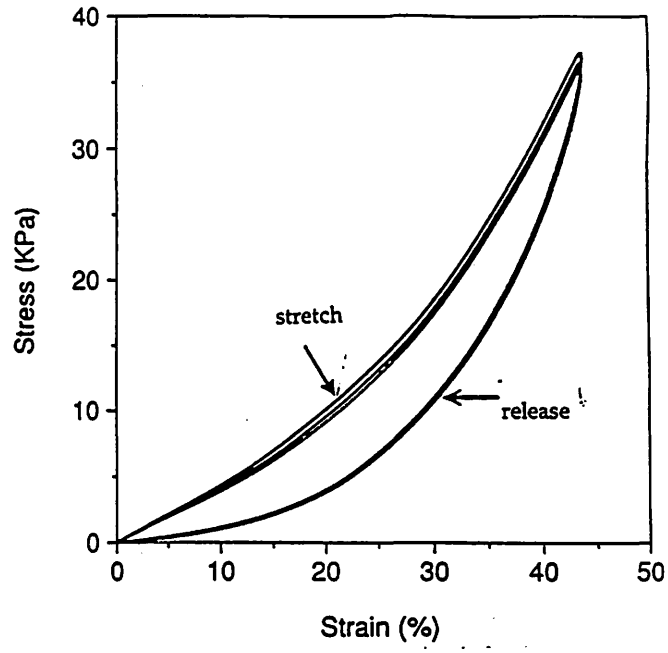


Figure 4-8 (left). Stress-strain curve for three cycles of stretch and release performed on the thyroarytenoid muscle.
Figure 4-9 (right). Stress-strain curve for three cycles of stretch and release performed on the vocal fold cover.

Figure 4-9 shows results of the same stretch-release experiment performed on vocal fold cover tissue. The main difference is the degree of nonlinearity. The "banana shape" has a gentler curve above the 30% strain level. But how does this nonlinearity affect phonation? The key parameter is the amplitude-to-length ratio (A/L) of the vocal folds. If the human larynx were to belong to a family of stringed instruments, its major peculiarity would be the shortness of the string in relation to its amplitude of vibration. For the dominant string mode, which is characterized by a half-wavelength between the anterior and posterior boundary attachments, the usual small-amplitude criterion is not met at all. A/L is between 0.1 and 1.0 (amplitudes in excess of 1 mm are observed for vocal fold lengths less than 1 cm). For a piano or violin string, on the other hand, the same ratio is on the order of 0.001. (A 1 mm amplitude of vibration for a 1 m piano string is not atypical). Many of the simplifying assumptions of elementary vibrating string theory (e.g. small angular deflections and constant tension) do not apply to the vocal folds. Tensions are dynamically varying, which produces mathematical nonlinearities in the effective restoring force (Titze, 1989).

Consider now the relation between transglottal pressure and glottal airflow. For a hard-walled constriction in a pipe,

$$P = \left(\frac{k}{2}\right)\rho U^2/A_g^2 , \tag{4-7}$$

where P is the pressure drop across the constriction (the glottis), ρ is the air density, U is the flow, A_g is the minimum cross-sectional area in the constriction, and k is a pressure coefficient. This nonlinear relation between transglottal pressure and glottal airflow, together with a tissue mode that allows for alternating convergent and divergent glottal shapes, provides the necessary instability for flow-induced oscillation (Ishizaka and Matsudaira, 1972; Titze, 1988).

For a yielding wall, the area A_g increases with pressure. Titze (1989) found that under dynamic conditions, the relation

$$A_g = CP^{1/2} \tag{4-8}$$

approximated the experimental results obtained in excised canine larynges, where C is a constant that varied with vocal fold length. This nonlinearity is related to the stress-strain nonlinearity of the tissue. The amplitude of vibration, and hence the glottal area, does not increase linearly with driving pressure because the tissue becomes stiffer under greater stretch. This may actually be an advantage from a control standpoint. A simple substitution of (4-8) into (4-7) reveals that

$$P = \frac{1}{C}\sqrt{\frac{k\rho}{2}}U , \tag{4-9}$$

suggesting a more linear relation between pressure and flow. Recent experiments by Sundberg, Titze and Scherer (1992) and Titze and Sundberg (1992) on singers have confirmed this relation, at least in some portions of the intensity and fundamental frequency ranges. Although this linearization may simplify the *control* process for phonation, it does not eliminate the nonlinearities. It simply "plays" one nonlinearity against the other. A sophisticated control system may look for regions of operation where complex relations between control variables can be simplified.

Attractors in Phase Space

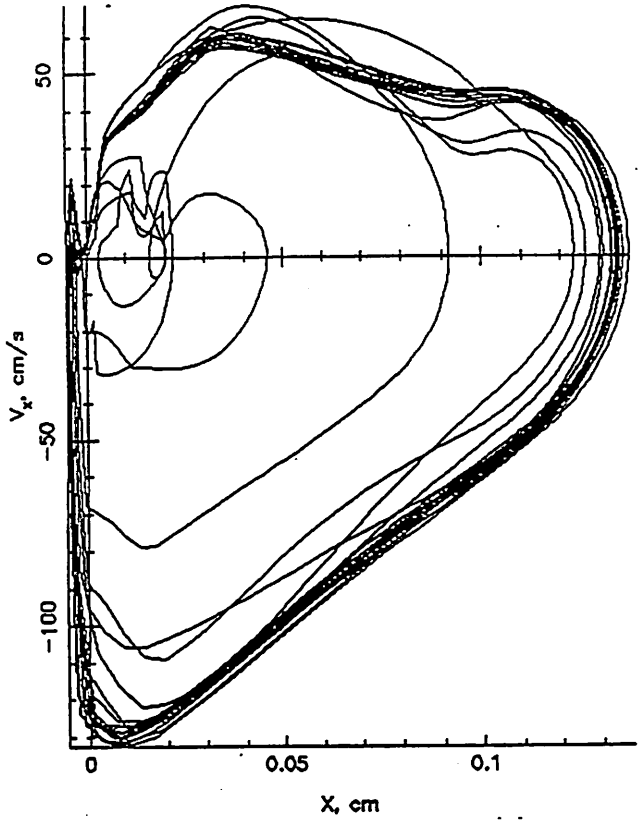
The crucial condition for the applicability of nonlinear dynamics is the dominance of a relatively low number of variables $x_i(t)$ ($i = 1, 2, ..., m$). Then, time-series can be projected into a *phase space*, which is spanned by these m coordinates. Under the assumption of fixed parameters (i.e. the external conditions are held constant) the m-dimensional vector $\underline{x}(t)$, termed *trajectory*, settles down on an *attractor* after some transient behavior (see Figure 4-10, which represents the lateral motion of one of the finite element nodal points of Figure 4-1. Lateral velocity is plotted as a function of lateral position). Note the transient behavior, indicated by orbits in the middle, followed by asymptotic behavior (nearly periodic) toward the outside. The outside orbits constitute the attractor. They determine the long-term behavior of the system.

A stylized attractor in phase space is shown enlarged in Figure 4-11. In addition to the overall phase portrait, *Poincaré maps* are often enlightening. Such maps are obtained if only the intersections of the orbits with a transversal section are studied. A variety of known attractor-types, together with their Poincaré maps, are illustrated in Figure 4-11:

i) stable stationary point (i.e. $x_i(t)$ are constant)
ii) stable limit cycle ($x_i(t)$ are periodic)
iii) stable n-torus ($x_i(t)$ is governed by n incommensurate frequencies)
iv) chaotic attractors ($x_i(t)$ nonperiodic)

A limit cycle in phase space corresponds to a fixed point of the Poincaré map, a 2-torus gives a closed curve and a low-dimensional chaotic attractor may lead to a pattern with a fractal structure.

Several attractors can coexist in phase space. Each attractor then has its own *basin of attraction*. The basin boundaries can be either smooth or fractal. In the latter case, *transient chaos* is possible, i.e. the trajectories change erratically over long periods of time and reach the attractor only after these extended periods. Some phonatory samples will be presented later that demonstrate the various dynamical regimes in phase space.



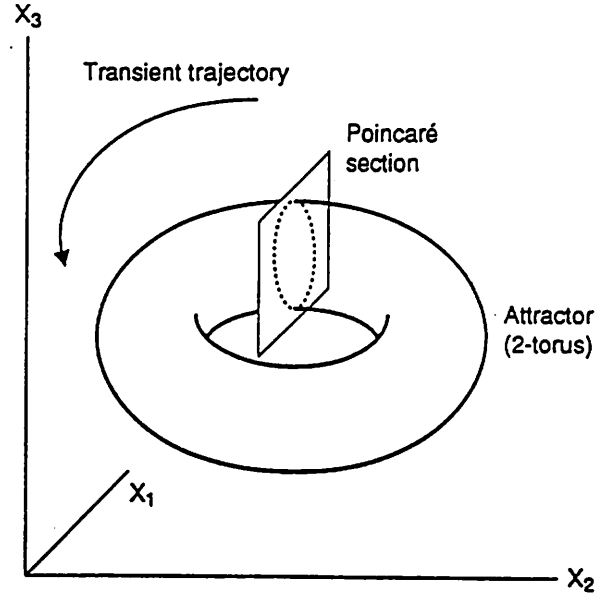


Figure 4-10 (left). Phase plane trajectory for one nodal point of Figure 4-1. Lateral velocity v_x is plotted against lateral displacement x. Figure 4-11 (above). Stylized visualization of an attractor in phase space, together with a Poincaré section.

Attractor Dimensions and Lyapunov Exponents

We now discuss the quantitative characterization of attractors. The selection of phase space variables x_i is somewhat arbitrary. However, certain quantities, such as attractor dimension,

Lyapunov exponents and the Kolmogorov entropy, are independent of the chosen coordinate system (Eckmann and Ruelle, 1985).

We start with some remarks on attractor dimensions. As already discussed earlier, the scaling behavior of various quantities can be exploited to define dimensions. The concept of *pointwise dimensions* turns out to be useful for the characterization of attractors in phase space: Let us assume that we have a large number of attractor points $\underline{x}(t_i) = \underline{x}_i$ (i=1,2,...,N) (each vector $\underline{x}_i$ corresponds to a point in the m-dimensional phase space). The number of neighbors of an attractor point $\underline{x}_i$ within a distance ε is termed local density $n_i(\varepsilon)$:

$$n_i(\varepsilon) = \sum_{j=1}^{N} \theta\left(\varepsilon - \| \underline{x}_i - \underline{x}_j \|\right), \tag{4-10}$$

where $\theta(s)$ is a unit step (Heaviside) function, defined with a dummy variable s as

$$\theta(s) = \begin{cases} 0 & \text{if } s \le 0 \\ 1 & s > 0 \end{cases} \tag{4-11}$$

If we find a power-law $n_i(\varepsilon) \sim \varepsilon^{D_p}$ over a certain ε-region, we can take D_p as an estimate of the pointwise dimension. For simple attractors this definition coincides with the familiar notion of a dimension (the "topological dimension"). For example a line gives $n_i(\varepsilon) \sim \varepsilon^1$ (i.e. $D_p = 1$) and a surface corresponds to $n_i(\varepsilon) \sim \varepsilon^2$. Chaotic attractors typically have noninteger dimensions greater than two.

In order to get reliable dimension estimates, the density $n_i(\varepsilon) \sim \varepsilon^{D_p}$ has to be averaged over many attractor points $\underline{x}_i$, leading to the correlation dimension D_2 (to be discussed later).

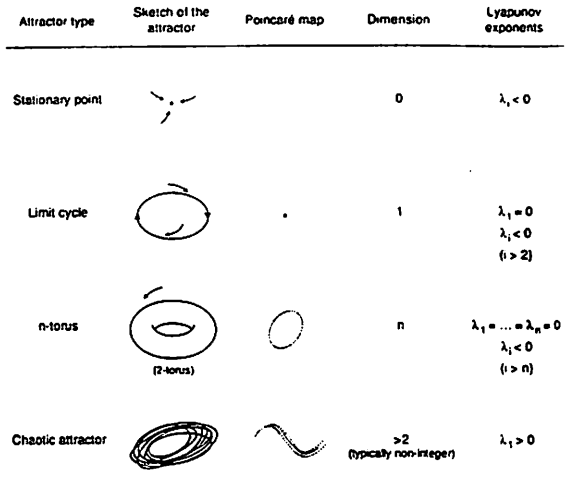


Figure 4-12. Various attractors and their characteristics.

Attractors can also be characterized by m Lyapunov exponents $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_m$. These quantities describe the stability properties of trajectories in different directions (see Eckmann and Ruelle, 1985 for details). A positive Lyapunov exponent γ_1 is of particular interest since its presence marks a system as chaotic. For $\gamma_1 > 0$ we find that sufficiently small deviations from a trajectory grow, in the mean, according to the function $\exp(\lambda_1 t)$, indicating a strong instability within the attractor. This inherent instability of chaotic systems implies limited predictability of the future if the initial state is known with only finite precision. Chaotic time-series have much in common with stochastic processes (decaying autocorrelations, limited predictability) but the detec-

tion of attractors and exponential instability (i.e. $\lambda_1 > 0$) allows chaos to be distinguished from noise.

Bifurcations

Up to now we have assumed fixed parameters, that is stationary external conditions. In these situations, the essential features of the dynamics are the attractors and their basins of attraction. If parameters of the dynamical system are varied, however, qualitative changes of attractors known as bifurcations are possible. For example, if one decreases the damping of an oscillatory system, one may switch from damped oscillation (fixed point attractor) to self-sustained oscillations (limit cycle). This transition is called a *Hopf bifurcation.* A limit cycle can bifurcate to a 2 torus, with two incommensurate frequencies, via a secondary Hopf bifurcation.

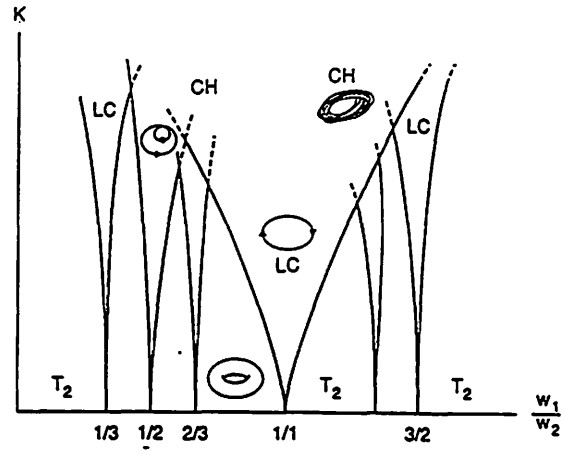


Figure 4-13. Schematic bifurcation diagram of coupled nonlinear oscillators and sketches of the corresponding attractors (LC is a limit cycle inside a resonance zone; T_2 is a 2-torus; CH is a chaotic attractor)

Another prominent kind of bifurcation is period-doubling (see e.g. Feigenbaum, 1983). In this case a limit cycle loses its stability and a periodic orbit of the double period is born. This bifurcation is accompanied by the appearance of subharmonics in the spectrum.

The chart of bifurcation lines in parameter space are termed *bifurcation diagrams.* As an important example we discuss the bifurcation diagram of coupled nonlinear oscillators (see Figure 4-13). The frequency ratio ω_1/ω_2 and the coupling strength K are taken as parameters. The qualitative behavior can be summarized as follows:

K = 0: For almost all ratios ω_1/ω_2 the frequencies are incommensurate and we have a 2-torus.

K small: Each rational value $\omega_1/\omega_2 = p/q$ (with p and q integers) leads to a resonance tongue Inside these resonance tongues the oscillators are *entrained,* i.e. they oscillate with a fixed phase shift. The attractor is therefore a limit cycle with long periods for large q. The width of these resonance zones shrinks drastically with increasing denomina tor q, that is, experimentally only the main resonances (1/1, 1/2, 2/3, ...) are visible.

K large: For increasing coupling strenth, the tongues overlap. This may lead to coexisting attractors and chaos. Within the resonances, chaos may occur via a cascade of pe riod-doubling bifurcations (Feigenbaum scenario; recall Figure 4-3, which is rotated 90° from Figure 4-13, with the parameter r begin similar to parameter k). Small changes of parameter values may result in drastic changes of the dynamic behavior.

Obviously, the bifurcation diagrams of nonlinear dynamical systems capable of producing deterministic chaos are extremely complicated. As a general result one can say that tori and period-doubling bifurcations are often precursors of chaotic dynamics. Particularly, coupled nonlinear oscillators exhibit complicated limit cycles (entrainment in resonances with $q \geq 2$), 2-tori and chaotic behavior.

From Observation to System Characterization

It was emphasized in the first section that nonperiodic measurements can be understood either as random processes or as manifestations of chaotic dynamics. In the case of random processes, the classical methods of time-series analysis, such as power spectra and autoregressive models, are appropriate. In the case of chaotic dynamics that can be modeled by relatively few variables, a qualitative theory of dynamical systems should be applied.

Phonatory Samples

Phenomena that are presumably manifestations of nonlinear dynamics (roughness, creaky voice, octave jumps, voice breaks, diplophonia) are frequently described in the literature (see e.g. Lieberman, 1963; Dolansky and Tjernlund, 1968; Koike, 1969; Monsen, 1979; Kelman, 1981; Imaizumi, 1986; Hirano, 1989). Particularly, newborn cries exhibit a lot of voice breaks and irregularities (Lind, 1965). In a recent paper (Mende et al., 1990) these observations are interpreted as bifurcations and low-dimensional chaos.

We now analyze some examples of vocalizations recorded at the university hospital (Charité) of the Humboldt-University (see Wendler et al., 1986 for details). The samples include normal as well as pathological voices. Each record of sustained vowels was classified perceptually according to its degrees of hoarseness (H), roughness (R) and breathiness (B). A four point grading was applied with 0 indicating normality, 1 slight, 2 moderate, and 3 extreme deviations.

The acoustic signals (sustained vowels at comfortable pitch) were digitized with a sampling rate of 20 kHz and 12 bit resolution. Figure 4-14 shows representative samples, labeled according to the diseases and the results of perceptual evaluation. The normal voice in Figure 4-14(a) is taken in the following as a standard of comparison. The pathological voices (parts b-d) exhibit large deviations from periodicity (jitter of more than 10% and shimmer from 15% to 50%). But it is evident that the deviations are by no means random. Therefore, the concept of nonlinear dynamics should be applied in order to detect the "hidden rules" that are related to attractors.

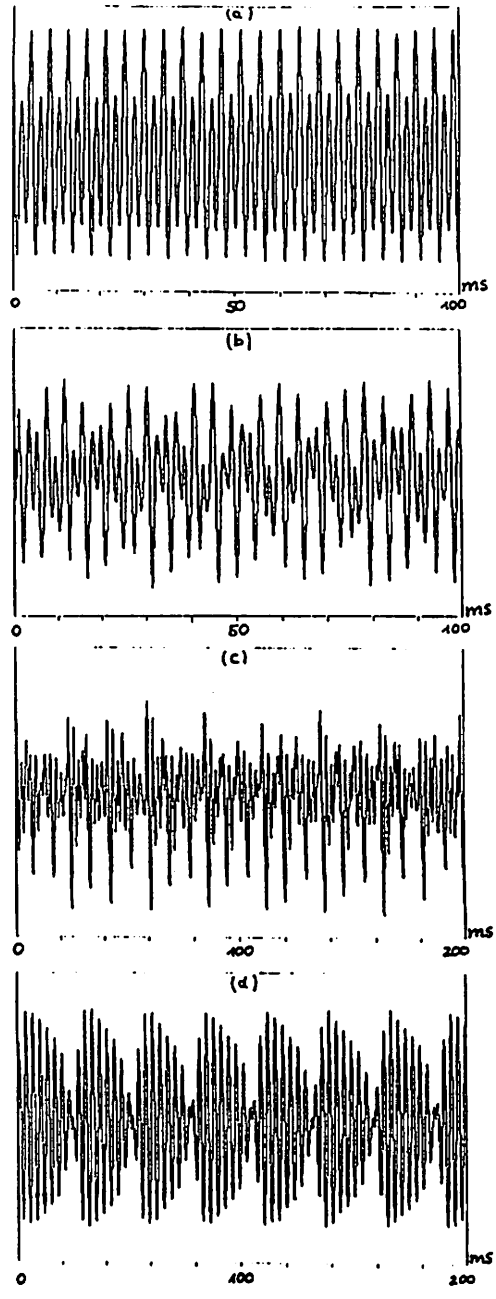


Figure 4-14. Acoustic waveforms from sustained vowels: (a) normal female voice; vowel "i"; (b) female voice, papillomas of the vocal folds; vowel "u"; (c) female voice, polyposis of the vocal folds; vowel "i"; (d) male voice, papillomas of the vocal folds; vowel "u".

Reconstruction of Attractors from Time-Series

In the previous section, the phase space spanned by the relevant dynamical variables x_i (i = 1,2,...m) was introduced. Ideally, these variables are positions and velocities of fleshpoints. In many laboratory situations, however, kinematic variables are not available. Furthermore, often only a single signal $x(t)$ is measured with sufficient accuracy (in our case the acoustic signal). Nevertheless, a pseudo phase space can be constructed easily in the following way (Froehling et al., 1981):

$$\underline{x}(t) = \{\, x(t),\ x(t+\tau),\ \ldots,\ x(t+(m-1)\tau)\,\} \tag{4-12}$$

A vector is reconstructed in "pseudo phase space" by m "delay-coordinates". The selection of a delay-time τ and an embedding dimension m is somewhat arbitrary, but attractor dimensions and Lyapunov exponents should not depend on τ and m, provided that m is sufficiently large (Eckmann and Ruelle, 1985). A variation of the embedding parameters can serve, therefore, as a consistency test of the algorithms discussed below.

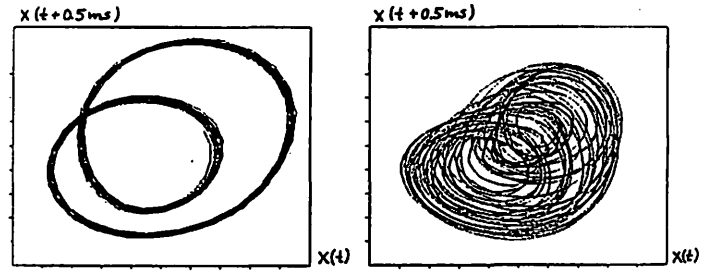


Figure 4-15. Projections of the samples in Figure 4-14(a) and (b) in pseudo phase space

In Figure 4-15 phase portraits in the pseudo phase space are shown for some of the signals. The periodic signal of the normal voice corresponds to a limit cycle with two loops, whereas the nonperiodic signal leads to a more complicated attractor.

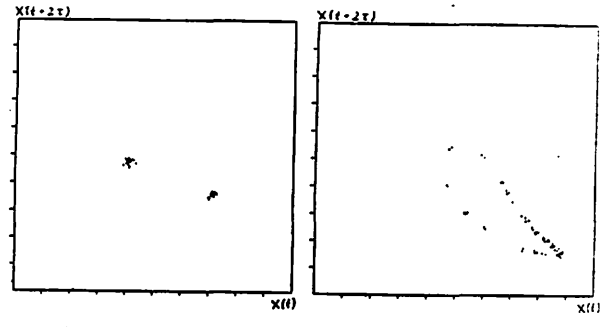


Figure 4-16. Poincaré maps of the samples in Figure 4-14(a) and (d) from intersections with the plane $x(t+\tau)=0$ for decreasing $x(t+\tau)$ (delay τ =1 ms and 0.9 ms).

The Poincaré maps in Figure 4-16 indicate a limit cycle with two intersections per period and a 2-torus. In the case of a randomly perturbed signal (e.g. a very breathy voice) an unstructured cloud of points would appear in the Poincaré map. Thus, the appearance of regular structures in phase portraits and Poincaré maps are first indications of low-dimensional dynamics.

Quantitative Characterizations of Attractors

Several techniques have been developed for the estimation of attractor dimensions (see e.g. Mayer-Kress, 1986). The most popular method is based on averaging the local densities $n_i(\varepsilon)$ introduced in the previous section:

$$\langle n_i(\varepsilon)\rangle \sim \sum_{i,j} \theta(\varepsilon - \|\underline{x}_i - \underline{x}_j\|) \sim C(\varepsilon) \sim \varepsilon^{D_2} \tag{4-13}$$

Here the brackets <...> denote averaging over many attractor points $\underline{x}_i$. The double sum is simply the number of distances between attractor points less than ε which is proportional to the so-called correlation integral $C(\varepsilon)$.

If $C(\varepsilon)$ scales as ε^{D_2}, then D_2 can be taken as an estimation of the correlation dimension (Grassberger and Procaccia, 1983). The estimate D_2 can be obtained from the slopes of log-log plots of the correlation integral $C(\varepsilon)$ versus the length scale ε. In the case of a low-dimensional attractor, one expects a scaling region with a constant slope D_2 (see Figure 4-17).

The dotted curves in Figure 4-17 refer to another method for the estimation of dimensions that is related to near-neighbor statistics (van der Water and Schram, 1988). If $r_k(\underline{x}_i)$ denotes the distance between an attractor point $\underline{x}_i$ to its k-th nearest neighbor, then the average distance depends on k as follows:

$$< r_k(\underline{x}_i) > \sim G(k,D)\, k^{1/D} \tag{4-14}$$

$$\text{with} \quad G(k,D) = \frac{\Gamma(k + 1/D)}{\Gamma(k)} \tag{4-15}$$

Again, the dimension D can be estimated from the slopes of the corresponding log-log plots. The prefactor G(k,D) is close to unity. Therefore, a first guess of D can be obtained by setting G(k,D) = 1. The resulting first estimation can be improved iteratively with the use of the Gamma function Γ.

The application of these techniques is demonstrated in Figure 4-17 for fixed delay-time τ and varying embedding dimensions *m* (other values of τ give comparable results). It can be seen that both methods give nearly identical slopes. As expected, the periodic signal of Figure 4-14(a) leads to a slope of about $D_2 = D = 1$, whereas the nonperiodic time-series of Figure 4-14(c) corresponds to the estimation $D_2 = D = 2.6$; a chaotic attractor is indicated.

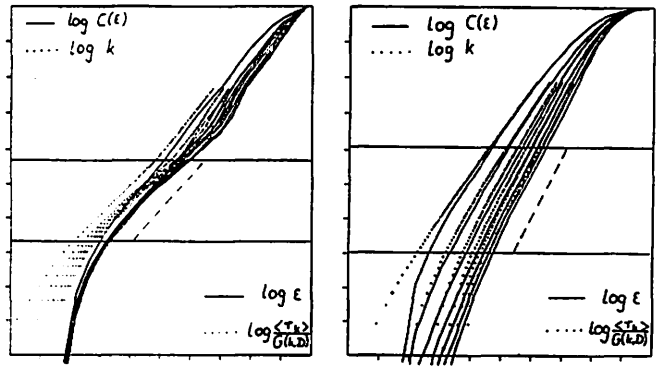


Figure 4-17. Log-log plots for the estimation of attractor dimensions derived from the records in Figure 4-14(a) and (b). Embedding dimensions: m=2,3,...8 (from above). full lines: correlations integral $C(\varepsilon)$ versus length scale ε. dotted lines: near-neighbor statistics (the horizontal lines delimit the approximate scaling region; the dashed lines indicate a slope of 1.0 and 2.6, respectively).

Now we turn to the estimation of the maximum Lyapunov exponent λ_1. Chaotic dynamics (i.e. $\lambda_1 > 0$) implies that nearby trajectories diverge (on the average) exponentially with a rate of λ_1. This property can be exploited to estimate the exponent λ_1. We use an algorithm by Wolf

et al. (1985) in a modified version (Herzel et al., 1991). The algorithm contains the following steps:

i) choose nearby points $\underline{x}_i = \underline{x}(t_i)$ and $\underline{x}(t_j)$ with an initial distance $d(0) = \| \underline{x}(t_i) - \underline{x}(t_j) \|$

ii) compute their distance $d(T) = \| \underline{x}(t_i+T) - \underline{x}(t_j+T) \|$ after an "evolution time" T

iii) find a new close neighbor of the reference point $\underline{x}(t_i+T)$ in such a way that the orientation of the separation vector is most nearly preserved (thus, a new initial distance d(0) results)

iv) average the local growth rate

$$ln \frac{d(T)}{d(0)}$$

over many separations. The "normalization" in the third step ensures that only small distances are probed for sufficiently small evolution times T and that the separation vector is oriented along the most unstable direction.

An indication of chaos is found if there is a linear growth in

$$\left\langle ln \frac{d(t)}{d(0)} \right\rangle$$

over a range of evolution times T. Moreover, one has to check that different embedding parameters τ and m lead to the same estimation of the exponent λ_1.

Figure 4-18 demonstrates robust estimation results. The analysis of the normal voice is consistent with the expected value $\lambda_1 = 0$ and the presumably chaotic time-series leads to $\lambda_1 \approx 0.1$ ms^{-1}.

Summarizing, we can say that the records of pathological voices in Figure 4-14 are, indeed, manifestations of low-dimensional attractors. Moreover, we have shown that time-series analysis inspired by nonlinear dynamics is applicable to phonatory samples.

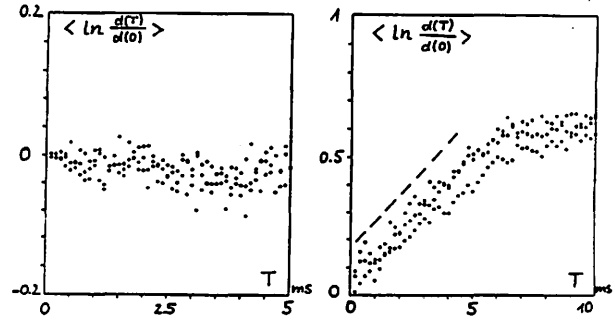


Figure 4-18. Average growth of the distance of initially nearby orbits with evolution time T. Three sets of embedding parameters are superimposed. (Figure a)Record from Figure 4-14(a) m=6 and τ =0.3 ms; m=6 and τ =0.5 ms; m=4 and τ =0.5 ms; (Figure b)Record from Figure 4-14(c) m=6 and τ =0.6 ms; m=6 and τ =1 ms; m=5 and τ =1 ms. The dashed line indicates a slope of γ_1 =0.1 ms^{-1}.

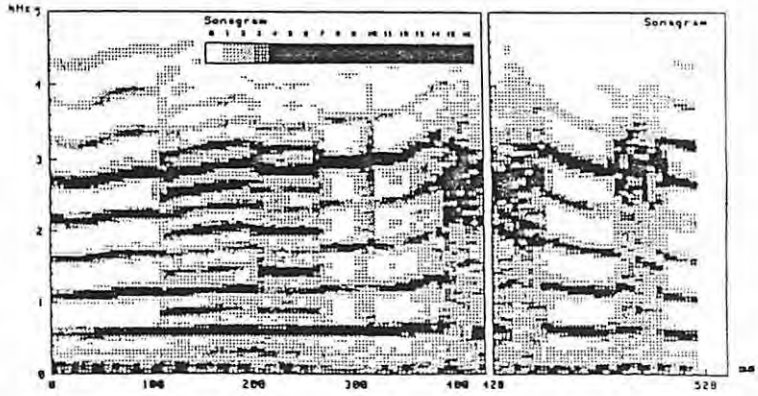
Bifurcations in Phonatory Signals

The characterization of attractors above was based on the assumption of stationarity. However, even during sustained vowels, changes of pitch and amplitudes are observed due to varying parameters such as muscle tension and

subglottal pressure. On one hand such instationarities complicate the attractor analysis. On the other hand, qualitative changes of the signals due to drifting parameters provide additional clues. These transitions can be interpreted as bifurcations of the underlying dynamical system. It was emphasized previously that bifurcation diagrams of potentially chaotic systems are extremely complicated. Consequently, we do not understand the entire collection of observed bifurcations in phonatory records. A few more examples will be given that demonstrate the rich variety of transitions.

Figure 4-19 gives a sense of the multitude of bifurcations possible in newborn cries. At the beginning of the spectrogram, only the fundamental frequency of about 500 Hz is present, together with its pronounced harmonics. Then, in connection with increasing energy and pitch, remarkably sharp transitions to more complicated states are visible. Particularly, the sudden appearances of subharmonics around 100 ms might be related to a jump from a 1:1 to a 1:2 resonance zone in the schematic bifurcation diagram of Figure 4-13. Such transitions to "subharmonic regimes" might be audible as octave jumps and are related to period-doubling bifurcations. Parametrically, the bifurcations may relate to dynamically changing tissue stress. In newborns, the amplitude to length ratio discussed earlier is extremely large, resulting in strong nonlinearities in the restoring forces.

It has been shown with the aid of Poincaré sections and dimension analysis that noisy segments of newborn cries can be interpreted as low-dimensional chaos (Mende et al., 1990). Although the spectrogram in Figure 4-19 is from a premature infant (28 weeks) with respiratory complications, we underscore that complex bifurcations and chaos were found in cries of healthy infants as well as in infants with complications (Lind, 1965; Mende et al., 1990).



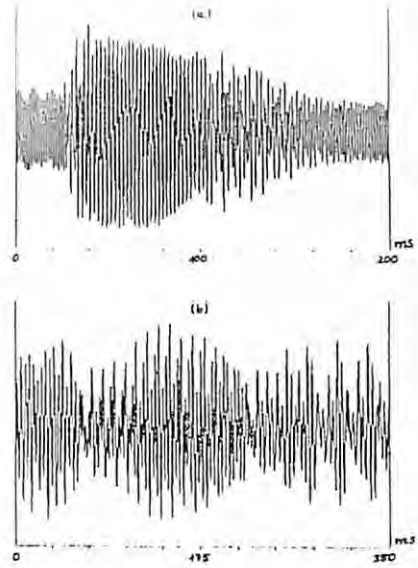


Figure 4-19 (above). Spectrogram from a newborn infant cry. Subharmonics and chaotic episodes are clearly visible. Figure 4-20 (right). Acoustic waveforms from pathological voices displaying bifurcations (see text): (Figure a)female voice, papillomas of the vocal folds (the same as in Figure 4-14(b); vowel "i"; (Figure b)male voice, with cancer of the vocal folds; vowel "e"

The record from a pathological voice in Figure 4-20(a) displays another interesting sequence of transitions. From a nearly periodic state, with a period of about 2.3 ms, a transition to a "low-frequency regime" with large amplitudes is observed. The reverse transition to a high-pitched

regime occurs via a segment of nonperiodic oscillation (from 180 to 280 ms), which could be interpreted as a "chaotic transient". The results are in some ways similar to those obtained on a spasmodic patient discussed earlier in Figures 4-6 and 4-7.

The voice of the cancer patient in Figure 4-20(b) is characterized by a strong instability, showing as multiple transitions during sustained vowels. The displayed signal starts with fast oscillations, then switches to a subharmonic regime, and it ends up in a slowly modulated shape.

Generally, many of the observed transitions remind us of the bifurcations of coupled oscillators. Thus, the relevant problem is to identify the relevant oscillating modes which are desynchronized due to diseases.

Concluding Remarks

The purpose of this chapter was to demonstrate the existance of low-dimensional chaotic dynamics in phonation. Attractors and bifurcations were identified using methods from nonlinear dynamics and simulaton models. Similar phenomena have been seen in earlier models of vocal fold vibrations (Ishizaka and Isshiki, 1976; Herzel et al., 1991).

Exploitation of these findings for clinical diagnosis is still a subject of future work. The estimation of attractor dimensions and Lyapunov exponents may be are too complicated for clinical applications, but the presence of low-dimensional dynamics can also be detected by much simpler quantities. In contrast to random perturbations, toroidal and chaotic behavior are characterized by strong correlations over several pitch periods. Thus, previously introduced methods for correlating pitch periods and amplitudes (Koike, 1969; Imaizumi, 1986) might be useful. Fundamental frequency or amplitude contours are intimately related to Poincaré sections introduced in this paper. Such plots are quite easily obtained from voice records and give hints about the underlying attractor. We point out, however, that the widely used perturbation measures jitter, shimmer or perturbation factor (Liebermann, 1963) are not appropriate for distinguishing chaos from random perturbations. They give only average deviations from a trend and are not sensitive to the salient patterns in the amplitude or F_o contour.

Summarizing, we emphasize that the understanding of pathological voices can benefit from the theory of nonlinear dynamics. Many observed phenomena can be understood as typical occurrences at the borderline of the 1:1 synchronization zone of oscillating modes. However, the precise identification of the essential modes and of the relevant nonlinearities is not yet clear. Specifically, a unification between analysis and synthesis must yet be achieved. A quantitative comparison of observed bifurcations and bifurcation diagrams from sophisticated models of vocal fold vibrations will, hopefully, give answers to the great variety of questions that still exist.

Acknowledgement

This work was supported, in part, by grant Nos. R01-DC00387-06 and P60-DC00976-01 from the National Institute on Deafness and Other Communication Disorders.

References

Alipour-Haghighi, F., and Titze, I. (1985). Viscoelastic modeling of canine vocalis muscle in relaxation. *Journal of the Acoustical Society of America, 78(6),* 1939-1943.

Alipour-Haghighi, F., Titze, I. (1990). Elastic models of vocal fold tissues. *Journal of the Acoustical Society of America,* (in review).

Awrejcewicz, J. (1990). Bifurcation portrait of the human vocal cord oscillations. *Journal of Sound and Vibration, 136,* 151-156.

Baer, T. (1981a). Observation of vocal fold vibration: Measurements of excised larynges. In Stevens, K. N. and Hirano, M. (eds.) *Vocal Fold Physiology* (pp 119-133). Tokyo: University of Tokyo Press.

Baer, T. (1981b). Investigation of the phonatory mechanism. *ASHA Reports, 11,* 38-46.

Baken, R.J. (1991). Géométrie fractale et évaluation de la voix: Application préliminaire á la dysphonie. *Bull. d'Audiophonogie. Ann. Sc. Univ. Franche-Comté 7(5-6),* 731-749.

Baken, R.J. (1990). Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice. *Journal of Voice, 4,* 185-197.

Barnsley, M. (1988). *Fractals everywhere.* New York: Academic.

Bergé, P., Pomeau, Y. and Vidal, C. (1984). *Order within chaos. Towards a deterministic approach to turbulence.* John Wiley and Sons, New York.

Dolansky, L., Tjernlund, P. (1968). On certain irregularities of voiced-speech waveforms, *IEEE Trans. AU-16,* 51-56.

Eckmann, J.-P., Ruelle, D. (1985). Ergodic theory of chaotic and strange attractors, *Rev. Mod. Phys. 57,* 617-656.

Farmer, J.D. (1982). Dimension, fractal measures, and chaotic dynamics. In Haken, H. (ed.), *Evolution of Order and Chaos* (228-246). New York: Springer-Verlag.

Farmer, J.D., Ott, E., and Yorke, J.A. (1983). The dimension of chaotic attractors. *Physica, 7D,* 153-180.

Feigenbaum, M. (1983). Universal behavior in nonlinear systems, *Physica 7D,* 16-39.

Freeman, W.J. (1990). Searching for signal and noise in the chaos of brain waves. In Krasner, S. (ed.), *The Ubiquity of Chaos* (pp 47-55). Washington, D.C.: American Association for the Advancement of Science.

Froehling, H., Crutchfield, J.P., Farmer, J.D., Packard, N.H., Shaw, S.H. (1981). On determining the dimension of chaotic flows, *Physica 3D,* 605-617.

Fujimura, O. (1968). An approximation to voice aperiodicity, *IEEE Trans. Audio Electroacoustics, AU-16,* 68-72.

Fujimura, O. (1990). Methods and goals of speech production research, *Language and Speech, 3(3),* 195-258.

Fujimura, O. (1992). Phonology and phonetics -- a syllable-based model of articulatory organization. *Journal of the Acoustical Society, Japan, (English) 13,* 39-48.

Glass, L. and Mackey, M.C. (1988). *From clocks to chaos.* Princeton, NJ: Princeton University.

Glass, L., Shrier, A. and Bélair, J. (1986). Chaotic cardiac rhythms. In Holden, A. V. *Chaos* (pp 327-356). Princeton, NJ: Princeton University.

Goldberger, A. and Rigney, D.R. (1990). Sudden death is not chaos. In Krasner, S. (ed.), *The Ubiquity of Chaos.* Washington, D.C.: American Association for the Advancement of Science, 23-34.

Goldberger, A., West, B.J., and Bhargava, V. (1986). Nonlinear mechanisms in physiology and pathophysiology: Towards a dynamical theory of health and disease. In Eisenfeld, J. and Witten, M. (eds.), *Modelling of biomedical systems* (pp 227-233). The Hague: Elsevier.

Grassberger, P. (1986). Estimating the fractal dimensions and entropies of strange attractors. In Holden, A. V. *Chaos* (pp 291-311). Princeton, NJ: Princeton University.

Grassberger, P. and Proccaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica 9D,* 189-208.

Herzel, H., Plath, P., Svensson, P. (1991). Experimental evidence of homoclinic chaos and type-II intermittency during the oxidation of methanol. *Physica, 48D,* 340-352.

Herzel, H., Steinecke, I., Mende, W., Wermke, K. (1991). Chaos and bifurcations during voiced speech. In E. Mosekilde, (ed.), *Complexity, Chaos and Biological Evolution.* , New York, NY: Plenum Press, 41-50.

Hirano, M. (1989). Objective evaluation of the human voice: Clinical Aspects, *Folia Phoniatr. 41,* 89-144.

Hirano, M., Kurita, S., and Nakashima, T. (1983). Growth, developing and aging of human vocal folds. In D. Bless and J. Abbs (eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues,* (pp 22-43). San Diego, CA: College Hill Press.

Imaizumi, S. (1986). Acoustic measures of roughness in pathological voice, *Journal of Phonetics 14,* 457-462.

Ishizaka, K., Isshiki, N. (1976). Computer simulation of pathological vocal-cord vibration, *Journal of the Acoustical Society of America, 60,* 1193-1198.

Ishizaka, K. and Matsudaira, M. (1972). Fluid mechanical considerations of vocal cord vibration. *SCRL Monograph 8,* Santa Barbara, CA, Speech Communication Research Laboratory.

Kahane, J.C. (1983). A survey of age-related changes in the connective tissues of the human adult larynx. In D. Bless and J. Abbs (eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues,* (pp 44-49). San Diego, CA: College Hill Press.

Kelman, A.W. (1981). Vibratory pattern of the vocal folds, *Folia Phoniatrica, 33,* 73-99.

Koike, Y. (1969). Vowel amplitude modulations in patients with laryngeal diseases, *Journal of the Acoustical Society of America, 45,* 839-844.

Liebermann, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathological larynges, *Journal of the Acoustical Society of America, 35,* 344-353.

Mackey, M.C. and Glass, L. (1977). Oscillation and chaos in physiological control systems. *Science, 197,* 287-289.

Mandelbrot, B. (1967). How long is the coast of Britain: Statistical self-similarity and fractional dimension. *Science, 156,* 636-638.

May, R.M. (1974). Biological populations with nonoverlapping generations: Stable points, stable cycles, and chaos. *Science, 186,* 645-647.

May, R.M. (1976). Simple mathematical models with very complicated dynamics. *Nature, 261,* 459-467.

May, R.M. (1980). Nonlinear phenomena in ecology and epidemiology. *Annals of the New York Academy of Sciences, 357,* 267-281.

Mayer-Kress, G. (ed.) (1986). Dimensions and entropies in chaotic systems, Springer-Verlag, Berlin.

Mende, W., Herzel, H., Wermke, K. (1990). Bifurcations and chaos in newborn cries. *Phys. Lett. 145 A,* 418-424.

Monsen, R.B. (1979). Acoustic qualities of phonation in young hearing-impaired children, *Journal of Speech and Hearing Research 22,* 270-288.

Moon, F.C. (1987). *Chaotic vibrations: An introduction for applied scientists and engineers.* New York: Wiley.

Pabon, J.P.H. and Plomp, R. (1972). Automatic phonetogram recording supplemented with acoustical voice quality parameters. *Journal of Speech and Hearing Research, 31,* 710-722.

Pickover, C.A. and Khorsani, A. (1986). Fractal characterizations of speech waveform graphs. *Computers and Graphics, 10,* 51-61.

Poincaré, H. (1881). Mémoire sur les courbes définies par une équation différentielle. *Journal de Mathématique, 3e sér., 7,* 375-422.

Poincaré, H. (1882). Mémoire sur les courbes définies part une équation différentielle. *Journal de Mathématique, 3e sér., 8,* 251-296.

Poincaré, H. (1885). Sur les courbes définies par les équations différentielles. *Journal de Mathématiques Pures et Appliquées, 4e sér. 1,* 167-244.

Rapp, P.E., Bashore, T.R., Zimmerman, I.D., Martinerie, J.M., Albano, A.M., and Mees, A.I. (1990). Dynamical characterization of brain electrical activity. In Krasner, S. (ed.), *The Ubiquity of Chaos* (pp 10-22). Washington, D.C.: American Association for the Advancement of Science.

Rasband, S.N. (1990). *Chaotic dynamics of nonlinear systems.* New York: Wiley.

Samar, V.J. and Rosenberg, S. (1990). A technique for the extraction of chaotic attractors from evoked response data: Toward the discovery of event related chaotic potentials. Presentation at the meeting of TENNET (Theoretical and Experimental Neuropsychology-Neuropsychologie Expérimentale et Théorique).

Sundberg, J., Scherer, R., and Titze, I. (in press). Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. *Journal of Voice.*

Thompson, J.M.T. and Stewart, H.B. (1986). *Nonlinear dynamics and chaos.* New York: Wiley.

Titze, I.R. (1988). The physics of small-amplitude oscillation of the vocal folds. *Journal of the Acoustical Society of America, 83(4),* 1536-1552.

Titze, I.R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America, 85(2),* 901-906.

Titze, I.R. (1991). A model for neurologic sources of aperiodicity in vocal fold vibration. *Journal of Speech and Hearing Research, 34,* 460-472.

Titze, I.R., and Sundberg, J. (1992). Vocal intensity in speakers and singers. *Journal of the Acoustical Society of America, 91(5),* 2936-2946.

Titze, I.R. and Alipour, F. (in review). Three source models for voice synthesis. *Journal of the Acoustical Society of America.*

van de Water, W., Schram, P. (1988). Generalized dimension from near-neighbor information. *Physical Review A, 37,* 3118-3125.

Wendler, J., Rauhut, A., Krüger, H. (1986). Classification of voice qualities. *Journal of Phonetics 14,* 485-488.

Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A. (1985). Determining Lyapunov exponents from a time-series. *Physica, 16D,* 285-317.

Wong, D., Ito, M.R., Cox, N.B., and Titze, I.R. (1991). Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases. *Journal of the Acoustical Society of America, 89,* 383-394.

NCVS Status and Progress Report - 3
December 1992, 65-79

Spectral Analyses of Activity of Laryngeal and Orofacial Muscles in Stutterers

Anne Smith, Ph.D.
Department of Audiology and Speech Sciences, Purdue University
Erich Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Margaret Denny, M.S.
Jennifer Wood, M.S.
Department of Audiology and Speech Sciences, Purdue University
Minoru Hirano, M.D.
Department of Otolaryngology, Kurume University Medical School
Steven Badylak, M.D., Ph.D.
Hillenbrand Biomedical Engineering Center, Purdue University

Abstract

Previous studies have reported that the disfluent speech of stutterers is often associated with tremor in orofacial muscle systems. In the present report, spectral analyses of the amplitude envelopes of laryngeal and orofacial EMGs revealed that tremorlike oscillations of EMG activity, similar to those observed in orofacial muscles, are also present in laryngeal muscles during stuttered speech. Furthermore, tremorlike oscillations in orofacial and laryngeal muscles appear to be entrained in some subjects. It is speculated that autonomic systems may provide a mechanism whereby oscillations in different muscle groups may become entrained.

Introduction

Stuttering is a disorder in which the motor outflow to the motoneuron pools recruited for speaking fails to generate normal patterns of speech movements. The precise etiology of this disorder is unknown, but recent multifactorial models implicate genetic, psychosocial, physiological, linguistic, and other factors as significant in the development of stuttering [1,2,3]. To understand the fundamental nature of this disorder and to develop successful therapeutic techniques, it would be important to specify the sources in the nervous system that are responsible for the failure of the motor command signals in stutterers' speech. Hypotheses about sources of inputs to motoneuron pools in stuttering must be based on precise descriptions of what actually is aberrant about muscle activity in stuttering.

A number of investigations from our laboratory and others have provided convincing evidence that speech breakdowns in stuttering (disfluencies) are often characterized by abnormal oscillations of EMG activity in muscles of the jaw, lips, and neck [4,5,6]. These oscillations of EMG typically occur in a frequency band of 5-15 Hz. To the extent that tremor has been studied in orofacial systems, it appears that this frequency band overlaps that of normal physiological tremor [7,8]. On the basis of such observations, it has been suggested that part of the disruptive drive to motoneuron pools in stuttering originates within the central and peripheral neural pathways that generate normal tremor [3,4,6].

To evaluate the relative importance of reflex, central, and biomechanical factors in the generation of tremor [9], an important experimental step is to determine if oscillations occur at common frequencies across effector systems. For example, Phillipbar et al. [10] found that tremor in eight Parkinsons patients did not show common spectral characteristics across limb, jaw, and vocal systems. This finding argues against a strong central component driving the oscillations in various effectors. Smith [6] found that spectra of amplitude envelopes of EMGs recorded from jaw, lip, and neck muscles could show common frequencies of dominant oscillation during intervals of stuttered speech. Denny and Smith [11] observed similar results in orofacial muscles during stuttered speech and reported that oscillations were highly correlated across muscles in some subjects. Denny and Smith also reported that oscillations in the 5-15 Hz band were not present in the fluent speech of these subjects. Thus, current evidence from recordings of orofacial muscles suggests that tremorlike oscillations grow large during disfluent intervals, but that abnormally high levels of oscillatory activity are not present during the fluent speech intervals of stutterers. Furthermore, in some stutterers, oscillations at a single frequency can be found in muscles of different articulators (e.g., jaw and lip), and the amplitude of these oscillations may be correlated across muscles over time.

The results reviewed above describe abnormal characteristics of activity of orofacial muscles during stuttering. An obvious question is whether other systems involved in speaking, laryngeal and respiratory systems, are disturbed in similar ways in stuttering. Thus, the focus of the present investigation is the analysis of activity of intrinsic laryngeal muscles during stuttering.

Despite the important role often ascribed to laryngeal function in stuttering, only two investigations have been reported in which activity of intrinsic laryngeal muscles was recorded during stuttering [12,13]. Neither study reported any qualitative or quantitative description of oscillations of laryngeal muscle activity associated with stuttering.

Tremor has often been studied by computing the spectrum of the amplitude envelope of the rectified, low-pass filtered EMG (e.g., the "demodulated EMG" of Elble and Randall [14]). Using this analysis technique, the present investigation was designed to determine (1) the normal spectral com-

position of the EMG envelope of laryngeal muscles during speech and other vocal tasks, (2) if stutterered speech is associated with abnormal dominant oscillations of laryngeal EMG, and (3) if oscillations of laryngeal and orofacial muscle activity are correlated.

Method

Subjects

Ten adult subjects were tested. Subjects N1-N3 were control subjects, two females aged 22 and 38, and one male aged 50. These three subjects exhibited normal speech with a mean speech rate of 147 words/min (s.d.=15).

S1-S7 were stutterers with a history of stuttering onset in childhood. Mean speech rate for the 7 stuttering subjects, including disfluencies, was 77 words/min (s.d.=25), and the mean number of disfluencies/100 words was 32 (s.d.= 11). On the basis of standardized measures[15], stuttering subjects were rated by a speech-language pathologist as severe (5 subjects) or moderate (2 subjects). In view of the invasive nature of the experimental procedures, mild stutterers were excluded from the subject population. Previous experience has shown that under similar experimental conditions, mild stutterers often do not have enough disfluencies on which to perform the data analysis. Subjects N1-N3 and S1-S4 were tested at the Purdue Speech Physiology Laboratory; data for S5-S7 were provided by Dr. Christy Ludlow of the NIH's Voice and Speech Laboratory.

Data Collection and Experimental Tasks (N1-N3 and S1-S4)

It was our goal to record the activity of two intrinsic laryngeal muscles, left thyroarytenoid (L-TA) and left cricothyroid (L-CT), and two orofacial muscles, left orbicularis oris inferior (L-OOI) and right levator labii superior (R-LLS). Analyses were planned for the two laryngeal recordings and for one channel of orofacial EMG.

For laryngeal recordings, electrodes were inserted into intrinsic laryngeal muscles with a 27g hypodermic needle by an otolaryngologist with extensive experience with laryngeal EMG. The bipolar hooked-wire electrodes consisted of two .002" diameter stainless steel wires with insulation removed from approximately 2mm at the ends of the hooks. Laryngeal insertions were made according to the procedures described in detail by Hirano [16]. These procedures included anesthetization of the skin at the insertion site with 1% lidocaine.

Successful placement of electrodes in TA or CT was assessed according to standard verification gestures for each muscle [16]. If electrode insertion did not result in successful placement according to the experimenters' on-line evaluation of patterns of EMG activation for the verification gestures, electrodes were removed. If the subject felt that he/she could easily tolerate another attempt, insertion was repeated on the same or the opposite side. The activity of orofacial muscles was recorded with surface electrodes taped to the skin overlying the muscle.

All EMG and other signals were recorded on FM tape (bandpass 0-1250 Hz) for off-line analyses. Other signals recorded simultaneously with the EMGs were the subject's voice and movement of the lower lip. The latter signal was transduced with a strain gauge cantilever attached to the vermilion border of the lower lip at midline [17]. The audio and movement signals were not analyzed, but served as aids to determine the time at which samples of EMG during fluent and disfluent speech and other experimental tasks should be extracted for analysis. In addition, a video recording of the subject (including face and torso) was obtained.

After verification of the EMG recordings, subjects were asked to perform pitch glides by singing from low pitch to high pitch and from high to low pitch. They were then asked to phonate the vowel "ee" (as in beet) as long as possible. Performance of each of these tasks was repeated several times. Subjects then engaged in conversational speech and read passages for approximately 30 minutes of continuous data collection. The conversations covered topics such as the subjects' history of stuttering, their employment, their hobbies, etc. Subjects also read several passages, including two passages from an anatomy text that included difficult vocabulary. At the end of the speech and reading tasks, subjects repeated the verification gestures so that the experimenters could determine if the laryngeal electrodes had been dislodged.

Data Collection and Experimental Tasks (S5-S7)

Data were collected for S5, S6, and S7 at the NIH Voice and Speech Section in an extensive protocol designed to provide baseline measures of laryngeal function prior to treatment by injection of botulinum toxin into the vocal folds. The protocol included a short speech sample that, for severe stutterers, contained enough disfluent speech for the present analysis. Because the goals of the NIH experiment were different, the data channels available were not identical to those recorded from the subjects at the Purdue facility, but the data included both orofacial and intrinsic laryngeal muscle recordings.

Methods employed at the NIH laboratory for laryngeal and orofacial EMG recording were essentially identical to those employed at the Purdue facility. Data were provided on FM tapes recorded with the same model recorder operated at the same speed as that used for the Purdue data collection.

Data Analysis

For each subject three channels of EMG and the audio signal were digitized on a laboratory computer. When possible, the EMGs selected were two laryngeal EMGs and an orofacial EMG showing consistent activity. The sampling rate was 2,048 samples/sec/channel. Prior to digitization the EMG signals were filtered, bandpass 50-1000 Hz for the laryngeal and 0-1000 Hz for the orofacial EMGs.

For each task performed by each subject, an attempt was made to digitize 20 intervals of one second duration. These intervals were taken for four conditions designated as:

FLU - the perceptually fluent speech of normal speakers and stutterers,
STUT - the perceptually stuttered speech of stutterers,
GLIDE - the pitch ascending and descending tasks,
"ee" - the sustained vowel.

Recorded muscles were not consistent across all subjects due to failure to successfully place electrodes in the target muscle or because some subjects were part of a different experiment (S5-S7). For all subjects, except S3 and S6, data were analyzed from two intrinsic laryngeal EMG channels and either orbicularis oris or levator labii. In the case of S3, two orofacial muscle channels (L-OOI and R-LLS) and one laryngeal channel (L-CT) were analyzed. For S6, the EMG of a lip muscle was not available, and the activity of medial pterygoid, a jaw-closing muscle, was analyzed along with that of TA and CT.

For normal speakers, the 20 one-sec FLU data files were taken randomly from the large samples of fluent speech recorded during the speech and reading tasks. The only criterion for selection of these intervals was one that was applied to all intervals selected for all tasks: that all EMG

channels appear to have some activity during the interval. The presence of EMG activity was verified by visual inspection of the records or in questionable cases, by a computer program that determined if the voltage was above an "activation threshold" for at least 25% of the record [6]. This criterion was applied because work from our laboratory has shown that spectral analysis of records with no EMG activity can produce spurious spectral peaks and cross-channel correlations due to the presence of noise that is common across channels.

Selection of the FLU data for the stuttering subjects was more problematic. Files were selected that (1) contained 1 sec of speech that two observers agreed contained no perceptible stuttering and (2) that were not immediately adjacent (within .5 sec) to a disfluent interval. Analyses of FLU speech were not completed for 4 of the stuttering subjects. These subjects had such severe stuttering that it was not possible to collect enough one-sec intervals for analysis within the above criteria. For each of the stuttering subjects whose fluent data were analyzed (S2, S3, S5), 20 FLU files could not be obtained, and the analysis was completed with fewer (at least 9) files.

For the STUT data, one-sec intervals that contained a disfluency were selected. These one-sec STUT intervals could be part of a longer disfluent event, or they could contain a disfluency with a duration < 1 sec. Because earlier work has shown that type of disfluency based on perceptual-linguistic categories, e.g., sound repetition, sound prolongation, interjection, is not related to the presence of oscillatory EMG activity, all types of disfluencies were included. Two judges listened to the audio for the selected intervals of perceptually FLU and STUT speech, and in the case of a disagreement, the interval was not included in the analysis. There was no consistent pattern of disfluency observed across the various speech conditions for the stuttering subjects. Some were more disfluent on the reading tasks, while others experienced more difficulty in conversational speech.

One-sec intervals for the pitch glide and sustained vowel tasks were selected randomly throughout the subjects' repeated performances of each task. Only the laryngeal EMGs were analyzed for these tasks, which often did not involve significant activation of orofacial muscles. Subjects S5-S7 did not perform these tasks.

After digitization, the one-sec data files were full-wave rectified and smoothed with a 10.25 ms gaussian window. After this processing, the power spectrum of each one-sec record was computed (for details of the power spectrum computation see Smith and Denny [18]), and these spectra were averaged for each muscle, for each subject and task.

In addition, the coherence function was computed for within-subject pairs of EMG signals. Coherence ranges from 0-1.0 and is equivalent to the squared cross-correlation between power in two signals computed separately for each frequency [19]. If the power at a particular frequency covaries over the 20 one-sec samples of two EMG signals for the task, the coherence between the two signals will be a significant, nonzero value at that frequency.

Averaged power spectra and coherence functions were plotted for each subject and task. The locations of spectral maxima were determined, and frequency of occurrence of spectral maxima within three ranges, 0-4 Hz, 5-15 Hz, and >15 Hz, was tabulated. The occurrence of significant values of coherence within the 5-15 Hz band was also noted. Coherence functions calculated on less than 16 files were not considered to be meaningful, as limited sampling of noisy signals can lead to spuriously high coherence estimates [19]. For 20 files (20 sec of sampled data) with the present sampling parameters, coherence estimates > .08 are significantly different from 0 with $p < .05$; for 16 files this value is .10. For consistency, the more conservative limit for nonzero coherence, .10, was used to evaluate all coherence functions.

Results

Speech

Normal Speakers. Figure 1 shows the averaged spectra computed from the amplitude envelopes of the EMG for the three normal speaking subjects for all muscles. In all cases these spectra show very large peaks in the 0-4 Hz range, suggesting that during speech, the amplitude envelope of the EMG is consistently modulated in this frequency band. Summarized in Table 1 are the locations of the spectral maxima for the speech condition for the normal speakers. For 8 of 9 spectra, the maximum value falls between 2 and 4 Hz. No maxima were observed in the 5-15 Hz band; and only the CT spectrum of N2 shows a maximum above 15 Hz. The coherence functions computed for the 9 muscle pairs analyzed in the normal subjects' speech condition revealed no instances of significant, nonzero coherence in the 5-15 Hz band.

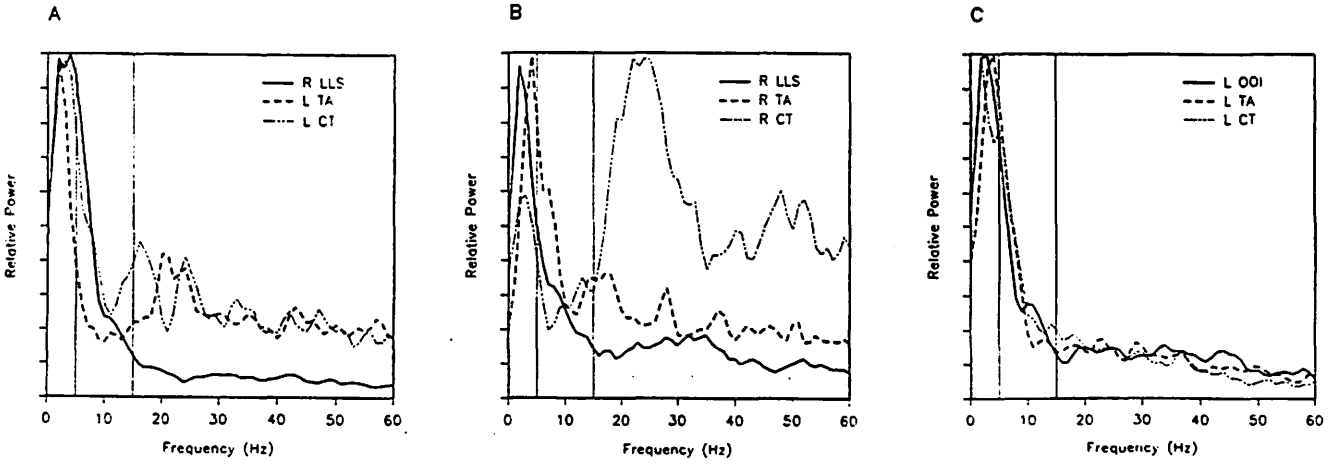


Figure 1. Averaged power spectra computed from the amplitude anvelopes of the EMG for the three normal speaking subjects N1 (panel A), N2 (panel B), and N3 (panel C). The 5-15 Hz band is marked by vertical lines in all spectra.

Table 1.
Location (frequency) of maxima in spectra of amplitude envelopes of EMGs during speech by normal speakers.

Subject	Muscles	Frequency Band (Hz) 0-4	5-15	>15
N1	R-LLS	4		
	L-TA	2		
	L-CT	3		
N2	R-LLS	2		
	R-TA	4		
	R-CT			22
N3	L-OOI	3		
	L-TA	4		
	L-CT	2		
% in each band		89%	0%	11%

Stutterers. Stuttering subjects showed a variety of patterns of laryngeal muscle activation during stuttering. Two examples are shown in Figures 2A and 3A, which contain EMG and audio records illustrating stuttered speech for S2 and S5. The records show 6 sec of data from a typical disfluency for each subject. Figures 2B and 3B contain the averaged power spectra for these two subjects computed from the STUT files for all muscles.

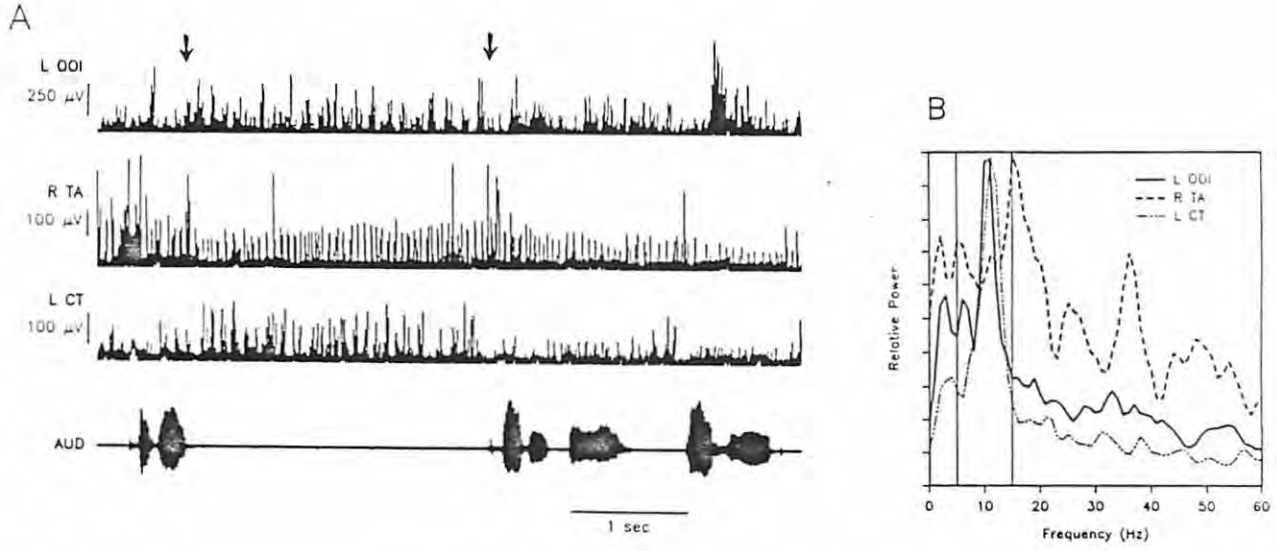


Figure 2. A: Typical pattern of muscle activity associated with disfluent speech of subject S2. The arrows indicate the interval in which stuttering occurred. B: Power spectra computed from amplitude envelopes of EMGs of S2 from the stuttered files

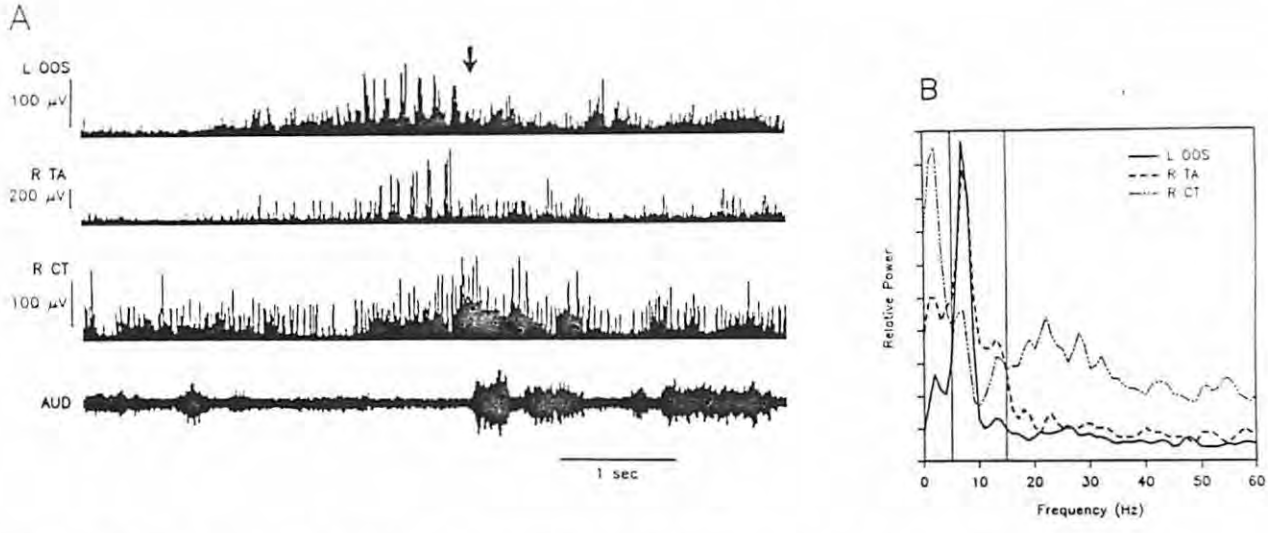


Figure 3. A: Typical pattern of muscle activity associated with disfluent speech of subject S5. The arrow indicates the end of a long interval in which stuttering occurred. The audio recording for this subject had a high level of background noise, but the onset of fluent speech (marked by the arrow) can still be detected visually. B: Power spectra computed from amplitude envelopes of EMGs of S5 from the stuttered files.

The time domain plots of data for Subject S2 (Figure 2A) suggest the presence of tremorlike oscillations for both OOI and CT during stuttering. The averaged power spectra for these muscles (Figure 2B) demonstrate that, during stuttering, both of these channels show a strong 11 Hz oscillation. Similarly, tremorlike oscillations are visually apparent in the OOS and TA channels for S5 (Figure 3A), and peaks at 8 Hz dominate the associated averaged power spectra shown in Figure 3B.

Table 2 indicates the distribution of the maxima of the spectra computed from stuttered speech intervals. Five of the 7 stuttering subjects had maxima in the 5-15 Hz band in at least one muscle. Forty-three percent of the 21 spectra had maxima in the 0-4 Hz range, while 48% had maxima in the 5-15 Hz band, and 9% had maxima above 15 Hz. Table 2 also contains the distribution of spectral maxima for the averaged power spectra computed from stutterers' fluent speech. Only one of these spectra (TA for S2) had a maximum in the 5-15 Hz band.

Table 2.
Location (frequency) of maxima in spectra of amplitude envelopes of stutterers' EMGs during stuttered speech (STUT) and fluent speech (FLU).

		STUT			FLU		
		Frequency band (Hz)			Frequency band (Hz)		
Subject	Muscle	0-4	5-15	>15	0-4	5-15	>15
S1	L-00I	1					
	L-TA			20			
	L-CT			18			
S2	L-00I		11		2		
	R-TA		15			15	
	L-CT		11		3		
S3	L-00I	2			2		
	R-LLS	3					
	L-CT	3			3		
S4	L-00I	2					
	L-LLS	4					
	L-TA	1					
S5	L-00S		7		2		
	R-TA		7		3		
	R-CT	2			2		
S6	R-MP		8				
	R-CT		5				
	R-TA		5				
S7	L-00S	2					
	R-TA	2					
	L-TA		10				
% in each band		43%	48%	9%	89%	11%	0%

Analysis of 3 muscle pairs for each of the 7 stuttering subjects yielded 21 coherence functions. The maximum value of the coherence function in the 5-15 Hz band for all muscle pairs for the STUT data was noted. Six of 7 stuttering subjects had significant coherence in the 5-15 Hz band in at least one muscle pair. Significant coherence occurred both within systems (e.g., TA/CT pairs) and across systems (e.g., OOI/TA). Figure 4 shows coherence functions computed for the OOI/CT pair of S2 and the OOS/TA pair of S5. It was noted above that these muscle pairs showed maxima at a common frequency in the power spectra (11 Hz for OOI and CT of S2, Figure 2B, and 8 Hz for OOS and TA of S5, Figure 3B). As Figure 4 indicates, the 11 Hz oscillation of S2 was not coherent, but the coherence function of S5 shows that the 8 Hz oscillations of OOS and TA were highly correlated.

Coherence analyses of stutterers' fluent speech were completed only for S3, who had enough files (19) for the coherence estimate. These functions revealed no significant coherence in the 5-15 Hz band for any muscle pair.

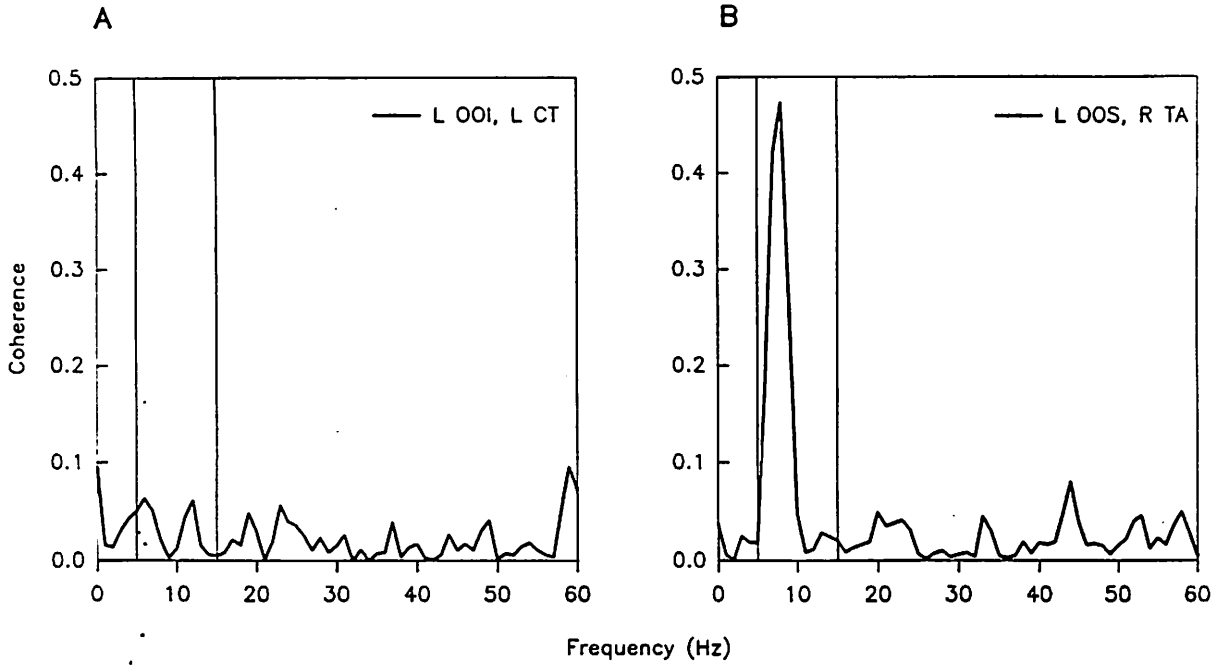


Figure 4. A: Coherence function computed for the L OOI and LCT pair of S2 on the basis the stuttered data. B: Coherence function computed for the L OOS and R TA pair of S5 on the basis of the stuttered data.

Pitch Glides and Sustained "ee" (N1-N3 and S1-S4). Table 3 indicates the location of the spectral maxima for the pitch glides and sustained "ee" performed by the normal speakers and stutterers S1-S4. For both groups the most common spectral pattern was one with a maximum above 15 Hz. This pattern is illustrated for the pitch glides of N1 and S1 in Figure 5, panels A and B. As shown in Table 3, maxima in the 5-15 Hz band did emerge for one muscle for three subjects (N2, N3, S2) in the sustained "ee" task. The CT and TA spectra for N2 for the sustained "ee" are shown in panel C of Figure 5.

Significant, nonzero coherence in the 5-15 Hz band was observed in two subjects in the pitch glide condition (N1, coherence = .11 at 11 Hz; S4, coherence = .21 at 6 Hz) and in two subjects for the sustained "ee" condition (N1, coherence = .12 at 15 Hz; S2, coherence = .20 at 6 Hz.

Table 3.
Location (frequency) of spectral maxima for power spectra computed for pitch glides and sustained "ee."

Subject	Muscle	Glide Frequency band (Hz) 0-4	5-15	>15	"ee" Frequency band (Hz) 0-4	5-15	>15
Normal Speakers							
N1	L-TA			39			24
	L-CT			28			23
N2	R-TA			18		14	
	R-CT			37			28
N3	L-TA	2					16
	L-CT	2			5		
% in each band		33%	0%	67%	0%	33%	67%
Stutterers							
S1	L-TA			36			27
	L-CT			32			24
S2	R-TA			20			20
	L-CT			19		14	
S3	L-CT			42			23
S4	L-TA			30	2		
	L-CT			23			18
% in each band		0%	0%	100%	14%	14%	72%

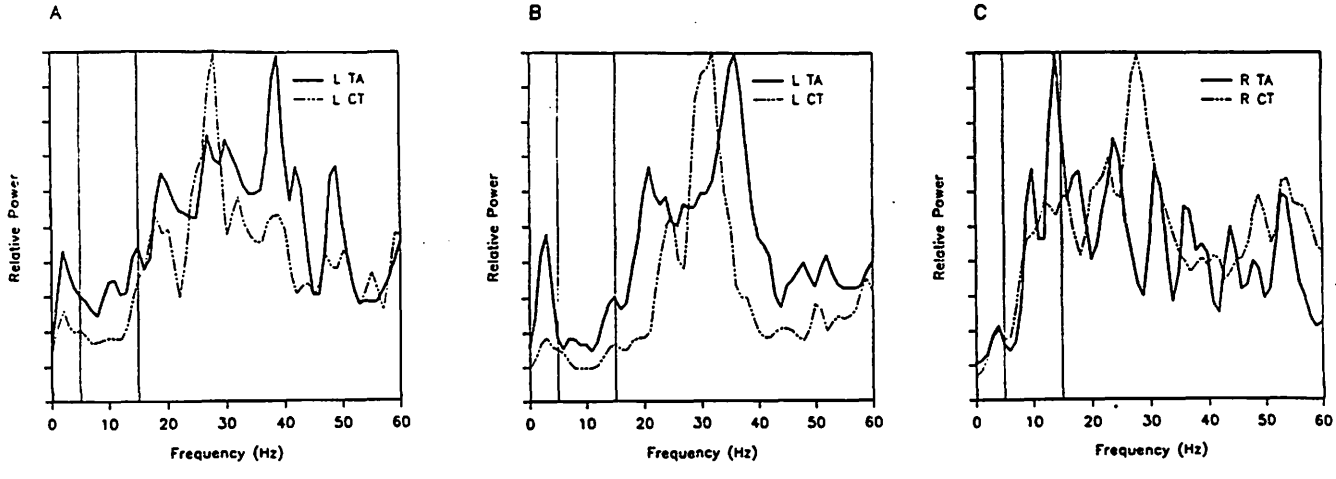


Figure 5. Power spectra computed from amplitude envelopes of laryngeal EMGs recorded during pitch glides of N1 (panel A) and S1 (panel B), and during sustained "ee" of N2 (panel C).

Discussion

Normal Spectral Patterns.

Because earlier investigations have not computed spectra of the EMG amplitude envelopes of laryngeal muscle activity, it was not known what spectral patterns would characterize the activity of TA and CT during the speech of normal subjects. During speech the position and stiffness of the vocal folds must be continuously adjusted to produce voiced and unvoiced sounds and to vary the pitch of the voice. TA plays a role in vocal fold adduction/abduction and in adjusting the internal stiffness of the vocal fold [20]. CT provides the major mechanism for adjusting the length of the vocal folds and thus is the muscle primarily responsible for changing the rate of vocal fold vibration [20].

Data recorded from the normal speakers in the present experiment clearly indicate that TA and CT activity is consistently modulated in relation to speech gestures. The dominant rate of modulation, 2-4 Hz, appears as the maximum value of the spectrum in most cases. Such spectral patterns are identical to those computed for the amplitude envelopes of orofacial muscle activity during normal speech [6]. The rate of modulation probably reflects the rate of production of primary speech gestures, and thus the rate of syllable production.

For subject N2, although there was a clear peak in the spectrum at 3 Hz (Figure 1B), the maximum of the spectrum occurred at 22 Hz. Spectra with maxima above 15 Hz have not been observed in the amplitude envelopes of orofacial muscles recorded during normal speech in the present or in a previous experiment [6]. It seems likely that these relatively high frequency maxima emerge in the laryngeal EMG envelopes when the recording is dominated by the firing of one or a few large motor unit spikes. In earlier studies in which the spectrum of the EMG amplitude envelope has been computed in speech [6 11], orofacial recordings have been obtained with broad-field surface or intramuscular electrodes (each wire inserted with separate needle). The resultant EMG was a gross interference pattern in which spikes of single units typically were not distinguishable. In contrast, the bipolar laryngeal EMG electrodes are inserted with a single needle and are likely to record from a much smaller population of motor units. Preliminary analyses of motor unit firing rates in the present data indicate that rates of 20-24 spikes/sec are typical.

Spectra with maxima > 15 Hz were the most common spectral pattern observed both for stutterers and normal speakers during performance of the singing and sustained vowel tasks. Thus, spectra for these tasks, unlike speech, are not normally dominated by a 0-4 Hz modulation of the EMG. For these tasks, spectra with maxima in the 5-15 Hz band were rare, with only three subjects showing such spectral patterns for one muscle in the sustained vowel condition.

Oscillations in the 5-15 Hz Band in the Stuttered and Fluent Speech of Stutterers.

The majority (5/7) of the stuttering subjects had spectral maxima in the 5-15 Hz band in at least one muscle during disfluent speech. Maxima in this frequency band were observed for both orofacial and laryngeal muscles. The results of the present investigation thus extend the findings of earlier studies of orofacial muscle sytems and lead to the conclusion that excessive oscillatory EMG activity can be present in laryngeal muscles during stuttering. It seems clear that patterns of muscle contraction necessary for normally coordinated speech movements would be disrupted by the synchronized firing of groups of motor units at rates of 5-15 Hz.

Also consistent with earlier investigations is the finding that not all stutterers show such oscillations [6 11]. Dominant oscillations in the 5-15 Hz band appear to be part of a neuromuscular pattern that characterizes stuttering, [6 11] but they are not a necessary condition for stuttering.

Disfluencies clearly can occur in the absence of excessive oscillations, thus other aberrant patterns of neuromuscular activity also must contribute to the breakdowns in speech production that characterize stuttering.

The fact that mild stutterers were excluded from the experiment might be interpreted to suggest that mild stutterers do not show tremorlike oscillations of EMG activity during their disfluent speech. This is not the case; earlier studies[6] have demonstrated tremorlike oscillations in EMGs of mild stutterers. Mild stutterers were excluded from the present experiment because they may exhibit little stuttering in a single, two-hour session. Because the central goal of the experiment was to analyze EMG during disfluent speech, we believed that stutterers should not be subjected to the present invasive methods unless it seemed highly likely that they would actually be disfluent during the recording session.

Given the severe nature of the stuttering exhibited by most of the subjects tested in the present experiment, it was difficult to obtain enough tokens of perceptually fluent speech for analysis. In the three stuttering subjects for whom power spectra were computed for fluent speech, the data suggest that oscillations in the 5-15 Hz band were reduced or absent during fluent speech. For example, S5 had spectral maxima at 7 Hz in both OOS and CT during stuttering, but in fluent speech, spectral maxima for these muscles were at 2 and 3 Hz respectively.

Oscillations in the 5-15 Hz band did emerge in the sustained vowel condition for two normal speakers and one stutterer. This may reflect the appearance of tremor during a task that required subjects to sustain phonation as long as they could. In any case, the fact that two normal speakers and one stutterer showed dominant oscillations in the 5-15 Hz band in the sustained vowel task would suggest that this was not an aberrant feature of stutterers' vocal motor behavior.

Correlations in Activity Across Muscles.

The present results demonstrate that oscillations in the 5-15 Hz band occur in laryngeal muscles as well as in orofacial muscle activity during stuttering. If a common source were driving these oscillations, the frequencies of oscillation should be the same, and the oscillations should covary in amplitude over time. As indicated in Table 2, in the three subjects who had oscillations in more than one muscle, spectral maxima occured at the same frequency for at least two muscles, both within system (S6, CT and TA at 5 Hz) and across systems (S2, OOI and CT at 11 Hz and S5, OOS and TA at 7 Hz). The existence of a common frequency of oscillation suggests a common driving source, but the coherence function provides a better test of a hypothetical common source. The coherence functions revealed that oscillations at a common frequency could be correlated or uncorrelated. The OOI and CT oscillations of S2 at 11 Hz showed zero coherence. In contrast, the 7 Hz, across-system oscillations of OOS and TA of S5 were highly correlated (coherence = .42), and the within system (TA and CT) oscillations of S6 showed the highest coherence value observed (.83). These mixed results are consistent with what has been observed across subjects with recordings of various orofacial muscles[6 11 21].

Potential Mechanisms Underlying Neuromuscular Oscillations in Stuttering.

Oscillations of muscle activity are common in both normal and pathological motor systems, and there are a large number of mechanisms that can produce neuromuscular oscillations [9]. At the present time, we have no data that would allow us to pinpoint with any degree of certainty the source of the neuromuscular oscillations that can characterize breakdowns in stutterers' speech. Certain aspects of the data, however, do provide a foundation for speculation about putative mechanisms.

Usually hypotheses about the source of neuromuscular oscillations attempt to distinguish the contribution of central and peripheral mechanisms. A strong central hypothesis to account for neuromuscular oscillations in stuttering would assert that a single source within the CNS generates the oscillations seen in various muscles within and across speech subsystems. The present data do not support a strong central hypothesis to account for oscillations across systems and subjects. In some cases, subjects had oscillations at different frequencies in different muscles. In other cases, oscillations could occur at a common frequency, but the oscillations were not correlated. A subject could, however, have oscillations in orofacial and laryngeal muscles that were highly correlated. Such results are consistent with the hypothesis that independent mechanisms drive oscillations in different muscles, but that these mechanisms can become entrained.

How does the oscillatory activity of a pool of motoneurons in the vagal motor nucleus become entrained with the oscillations of a pool of motoneurons in the facial nucleus? In addition to the neural systems that produce speech movements, the laryngeal and orofacial muscles are the targets of the output pathways of many other neural control systems. These include the systems involved in metabolic breathing, mastication, deglutition, and emotional expression. One possibility is that autonomic systems, for example, those involved in emotional expression, may provide a mechanism whereby oscillations can become linked across orofacial, laryngeal, and respiratory systems. With increased autonomic arousal, the neural centers involved in emotional expression may provide a positive feedback network to distributed motoneuron pools such that small, independent oscillations grow larger and ultimately become entrained.

Under this hypothesis, stuttering should be worse under conditions of increased autonomic arousal, and increased neuromuscular oscillations should be correlated with increases in autonomic arousal. On the basis of clinical evidence and subject reports, it is widely accepted that stuttering becomes worse under conditions of emotional stress [2]. In addition, in a study of 19 stuttering subjects, Weber and Smith [22] reported that the likelihood of disfluency and the severity of disfluency were positively correlated with the level of sympathetic arousal. Future work will be necessary to determine if the amplitude of neuromuscular oscillations in stuttering is correlated with levels of sympathetic arousal. Relevant to this argument is the demonstration of an increase in tremor in limbs of human subjects in response to increased levels of circulating adrenaline [23].

Conclusion

Stuttering is a complex movement disorder affecting speech motor systems. The present investigation extends earlier work by demonstrating that orofacial and laryngeal muscle systems can be disturbed in common ways in stuttering. As a final caveat, it is noted that the pattern of neuromuscular oscillations described in this experiment is only a part of the constellation of symptoms that constitute the disorder of stuttering. In no way is it implied that these oscillations are the cause of stuttering (see Smith [1] for a detailed discussion of this issue). Rather it is suggested that these oscillations are a common physiological correlate of stuttering, one that may ultimately provide a partial key to understanding the neural mechanisms that disrupt speech movement patterns in this disorder.

Acknowledgements

This work was supported by Grants DC00559 and DC00976 from the NIH National Institute on Deafness and Other Communication Disorders.

References

1. Smith A. Factors in the etiology of stuttering. ASHA Reports 1990;**18**:39-47.
2. Van Riper C. The nature of stuttering. Englewood Cliffs: Prentice-Hall, 1982.
3. Zimmermann G. Stuttering: A disorder of movement. Journal of Speech and Hearing Research 1980;**23**:122-36.
4. Fibiger S. Stuttering explained as a physiological tremor. Speech Transmission Laboratory-Quarterly Progress and Status Report 1971;**2-3**:1-24.
5. McClean M, Goldsmith H, Cerf A. Lower-lip EMG and displacement during bilabial disfluencies in adult stutterers. Journal of Speech and Hearing Research 1984;**27**:342-49.
6. Smith A. Neural drive to muscles in stuttering. Journal of Speech and Hearing Research 1989;**32**:252-64.
7. McFarland DH, Smith A, Moore CA, Weber, CM. Relationship between amplitude of tremor and reflex responses of the human jaw-closing system. Brain Research 1986;**366**:272-78.
8. Palla S, Ash MM Jr. Frequency analysis of human jaw tremor at rest. Archives of Oral Biology 1979;**24**:709-18.
9. Stein RB, Lee RG. Tremor and clonus. In:Brookhart JM, Mountcastle VB, eds. Handbook of physiology: The nervous system II. Bethesda:American Physiological Society, 1981:325-44.
10. Phillippbar SA, Robin DA, Luschei, ES. Limb, jaw, and vocal tremor in parkinson's patients. In:Yorkston KL, Buekelman DR, eds. Recent advances in clinical dysarthria. Boston:College-Hill,1989:166-97.
11. Denny M, Smith A. Gradations in a pattern of neuromuscular activity associated with stuttering. Journal of Speech and Hearing Research in press.
12. Freeman F, Ushijima T. Laryngeal muscle activity during stuttering. Journal of Speech and Hearing Research 1978;**21**:538-62.
13. Shapiro AI. An electromyographic analysis of the fluent and dysfluent utterances of several types of stutterers. Journal of Fluency Disorders 1980;**5**:203-32.
14. Elble RJ, Randall, JE. Motor-unit activity responsible for 8- to 12-Hz component of human physiological finger tremor. Journal of Neurophysiology 1976;**39**:370-83.
15. Riley G. Stuttering Severity Instrument for Children and Adults. Portland:C.C. Publications,1984.
16. Hirano M. Clinical examination of voice. In:Arnold GE, Winckel F, Wyke BD, eds. Disorders of Human Communication 5. New York:Springer-Verlag Wien, 1981.
17. Barlow SM, Cole KJ, Abbs JH. A new head-mounted lip-jaw movement transduction system for the study of motor speech disorders. Journal of Speech and Hearing Research 1983;**26**:283-88.
18. Smith A, Denny M. High frequency oscillations as indicators of neural control mechanisms in human respiration, mastication, and speech. Journal of Neurophysiology 1990;**63**:745-58.
19. Bendat JS, Piersol AG. Random data:Analysis and measurement procedures, 2nd ed. New York:Wiley,1986.
20. Hirano M, Kakita Y. Cover-body theory of vocal fold vibration. In: Daniloff RG, ed. Speech science: Recent advances San Diego:College Hill,1985:1-46.
21. Smith A, Denny M, Wood J. Instability in speech muscle systems in stuttering. In: Peters HFM, Hulstijn W, Starkweather W, eds. Speech motor control and stuttering, New York:Elsevier, 1991:231-42.

22. Weber CM, Smith A. Autonomic correlates of stuttering and speech assessed in a range of experimental tasks. Journal of Speech and Hearing Research 1990;**33**:690-706.
23. Marsden CD, Owen DA. The effect of adrenaline on physiological tremor in man. Journal of Physiology 1967;**188**:40-41.

NCVS Status and Progress Report - 3
December 1992, 81-95

Air Pressure Regulation During Speech Production

Jerald B. Moon, Ph.D.
John W. Folkins, Ph.D.
Alice E. Smith, M.A.
Erich S. Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Vocal tract pressures during speech tend to be maintained in the face of airway leaks that might be encountered by individuals with repaired palatal clefts. This study tested the hypothesis that such constant pressures can be explained as a consequence of constant pressure source characteristics of the respiratory system during speech production. This is in contrast to pressure regulation theories (e.g. Warren, 1986) which posit that active reflexive compensatory responses occur to regulate air pressure during speech. Four experiments were conducted to a) confirm the distinction between constant flow versus constant pressure sources using a plastic vocal tract model, b) compare air pressure and flow patterns associated with airway leaks in normal speakers to model data, c) study the effects of gas density in the vocal tract model on air pressure and air flow patterns predicted by a non-reflexive constant pressure source hypothesis, and d) compare air pressure and air flow patterns associated with airway leaks obtained from normal speakers using different gas densities to model data. Observed flow and pressure patterns fit a model of the speech production system as a constant pressure source. In addition, this explanation of the observed pressure and flow patterns does not require consideration of active compensatory respiratory, laryngeal, or upper airway system responses.

In 1986, Warren posited that pressure maintenance is a primary goal of the speech motor system. More specifically, Warren hypothesized that "speech conforms to patterns characteristic of

a regulating system" (p. 252) and that "compensatory behaviors in [individuals with a] cleft palate are manifestations of regulation and control strategies" (p. 252). That is, vocal tract air pressure is regulated during speech production to maintain a driving force above some minimally acceptable level. Regulation of speech pressure is controlled by active changes in respiratory effort or active alterations in resistance at one or more locations in the vocal tract (Warren et al., 1989a). Further, it has been suggested that speakers attempt to maintain stable air pressures even though auditory perceptual accuracy may be compromised as a result (Warren et al., 1989b).

The existence and utilization of an active pressure regulation system requires at least two events. First, there must be a pressure sensing system somewhere in the vocal tract. There is evidence of the existence of pressure sensors in the vocal tract (see Warren, 1986 for review). Second, online feedback of such pressure information must lead to activity modulating pressure in the desired amount and direction. Putnam et al. (1986) measured intraoral air pressure and air flow rate as normal talkers produced the syllable [p ʌ] across interlabial leak tubes of different cross-sectional areas. They observed that, while intraoral pressures dropped with increases in leak tube area, they did not fall below 4 cm H_2O. Increases in leak tube area were associated with increases in interlabial leak tube flow rate. Further, leak tube area had no effect on air flow rates observed during the vowel portion of the CV syllable. Putnam et al. concluded that these subjects increased respiratory effort in an attempt to regulate intraoral air pressure actively.

Warren et al. (1989a) compared air flow and air pressure patterns generated in the presence of velopharyngeal air leaks in a number of individuals to air pressures and air flows obtained using comparable velopharyngeal air leaks in a plastic vocal tract model. They argued that the model represented a passive system incapable of compensatory behavior. There was a 90% drop in model intraoral air pressure as the velopharyngeal opening was increased from 3 to 45 mm^2, while human subjects demonstrated only a 28% decrease. In order to obtain their model data, air flow rate was fixed at a level that produced typical intraoral air pressures with the 3 mm^2 velopharyngeal orifice plate in place. However, human air flow rates were observed to increase ten-fold over the range of velopharyngeal orifice areas measured. Warren et al. (1989a) argued that increased flows, at least at higher orifice areas, were due to an active respiratory response.

Warren et al. (1989b) investigated respiratory responses to velopharyngeal air leaks in normally speaking subjects. Speakers voluntarily lowered the soft palate during production of normally oralized speech sounds. A nonsignificant pressure drop from the oral to nasal condition was associated with a significant increase in expiratory volume. This was interpreted as an active respiratory response designed to maintain intraoral air pressure.

Additional support for active pressure regulation came from Warren et al.'s (1992) investigation of physiologic responses to an "imposed change in airway resistance" (p. 2947). Air pressure and air flow patterns were recorded from speakers with varying degrees of velopharyngeal dysfunction. In addition, air flows and pressures were recorded from normal subjects producing [pa] and [si] under varying interlabial bleed and bite block conditions. All three groups demonstrated relatively stable oral pressures as vocal tract leak increased in magnitude. Air flow rates increased with increases in vocal tract leak magnitude for the speakers with velopharyngeal dysfunction and with increases in interlabial leak tube size. Air flow rates did not increase significantly as a function of increases in bite block size. Warren et al. theorized that speakers in the bite block condition used a combination of respiratory and lingual compensations to maintain appropriate vocal tract resistance.

Another explanation for observed pressure stability is that pressure maintenance is secondary to auditory perceptual regulation by the system controlling speech production. That is, intraoral pressure might be modulated so that auditory perceptual accuracy is maintained. Auditory perceptual accuracy would then be regulated and intraoral air pressure would be controlled to reach the regulated goal. In the case of bilabial plosives in which inadequate intraoral air pressure would result in an perceptually inaccurate production, the only compensations available to the speaker would be altered respiratory drive, altered nasal or velopharyngeal valving, or alterations in lip closure (Netsell, 1990). Regardless of which strategy is chosen, the primary goal would be regulation of the percept.

Warren et al. (1984) used auditory masking and induced oral openings to study the compensatory strategies of normal individuals. While auditory perceptual accuracy was higher in the presence of auditory feedback, masking eventually resulted in the deterioration of auditory perceptual accuracy while peak intraoral air pressures stayed above 3 cm H_2O. The latter finding does not support auditory perceptual regulation as the primary goal.

Moon and Folkins (1991) tested perceptual regulation notion by altering auditory feedback on-line during plosive production. Perceptual regulation should have produced changes in intraoral air pressure. However, only large (30 dB) manipulations of the auditory signal were associated with significant changes in intraoral air pressure. Overall, the modulations of intraoral air pressure were not as large as Moon and Folkins expected, leading to the conclusion that auditory regulation alone does not explain pressure maintenance.

A third possible explanation for observed pressure stability does not require that pressure maintenance be a primary or secondary goal of the speech motor control system. The respiratory system can be considered as a musculo-elastic energy source that supplies a pressurized stream of air to the vocal tract. Resistance in the normal adult airway during respiration is less than 1.5 cm H_2O/LPS (Comroe, 1965). About one half of that is due to resistance in the nasal passages. In individuals with repaired palatal clefts, nasal resistance may be elevated to about 3-5 cm H_2O/LPS (Warren et al., 1969). Regardless, the output resistance of the normal lung-thorax unit is likely low compared to other resistances that may be encountered in the airway during speech production. On the basis of these aeromechanical properties, the human respiratory system may be characterized as a constant pressure source. That is, the constant pressure source has a low output resistance relative to any downstream resistances. Alterations in downstream resistance would be expected to have little effect on output pressure of the respiratory system while producing larger changes in air flow rate. A constant flow source, on the other hand, has a large output resistance relative to resistances likely to be encountered downstream. In a constant flow system, alterations in downstream resistance would be expected to have little effect on output volume air flow rate of the respiratory system while producing large changes in airway pressure.

The hypothesis tested in this investigation is that the human respiratory system acts as a constant pressure source during speech. In this case, pressure maintenance observed in association with airway leaks might be explained as a consequence of the inherent physiologic characteristics of the respiratory system and not as an active compensatory response. Four experiments were conducted. Experiment 1 was designed to confirm the distinction between constant flow versus constant pressure sources using a plastic vocal tract model. Experiment 2 compared air pressure and flow patterns associated with airway leaks in normal speakers to the model data from experiment 1.

In experiment 3 we studied the effects of gas density on model air pressure and air flow patterns predicted by the constant pressure source hypothesis. In experiment 4 we tested whether or not the gas density effects observed in experiment 3 would be observed in human subjects.

Experiment 1

In the first experiment, air pressure and flow patterns were recorded from a plastic model of the vocal tract using air delivered from either a constant pressure source or a constant flow source. Airway leaks were simulated using interchangeable plates.

Method

The plastic vocal tract model was designed after one originally developed Warren and DuBois (1964) and used in subsequent modelling studies by Warren and colleagues (eg. Warren et al., 1989b). The model dimensions were selected to simulate the volume and airway resistance of the normal adult vocal tract. The velopharyngeal orifice and oral port orifice of the model are adjustable using plates with holes of a known cross-sectional area. Since measures were to be compared to data collected from human subjects using interlabial leak tubes, the oral port opening of the model was manipulated in this experiment. Four orifice opening sizes were used. They were 3.14, 7.07, 12.56, and 19.63 mm^2. Two air sources were used. The constant flow source chosen was a regulated cylinder of compressed air. Regulated air cylinders have a high output impedance. The constant pressure source was a vacuum cleaner configured to use the blower output. The vacuum cleaner has a low output impedance. The fan speed of the vacuum cleaner was controlled using a variable AC transformer.

For each air source, the input flow rate to the model was first adjusted with the smallest (3.14 mm^2) plate in position to achieve an oral cavity pressure of 12 cm H_2O. The air flow rate was then left unaltered while 12 measures of oral cavity volume flow rate and intraoral air pressure were recorded for each of the four orifices.

Intraoral air pressure was sensed using a pressure transducer (Honeywell Microswitch PC01D) coupled to the model oral cavity. Air flow rates were sensed using a pneumotachograph (Hans Rudolf model 4719) and pressure transducer (Honeywell Microswitch PC01D36) coupled to the trachea of the model.

Pressure and flow transducers were calibrated using a U-tube water manometer (Fischer-Porter) and rotameter respectively. Pressure and flow signals were digitized on-line at 200 Hz and values were recorded directly from the computer screen.

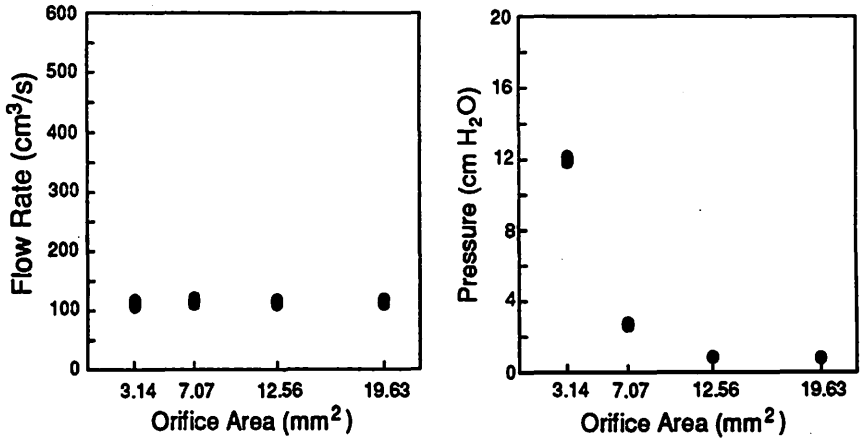


Figure 1. Airflow and pressure values recorded as a function of oral port orifice cross-sectional area using a plastic vocal tract model and compressed air (constant flow) cylinder.

Results

Individual data points obtained with the compressed air cylinder are shown in Figure 1. The tight clustering of the twelve data points for each orifice area reflects the consistency of the pressure and airflow relationship. It is clear that increases in orifice area are associated with a) relatively stable air flow rates, and b) decreasing intraoral pressure. These data reflect the air pressure and flow pattern expected of a constant flow source, and support the characterization of a compressed air cylinder as that type of air supply.

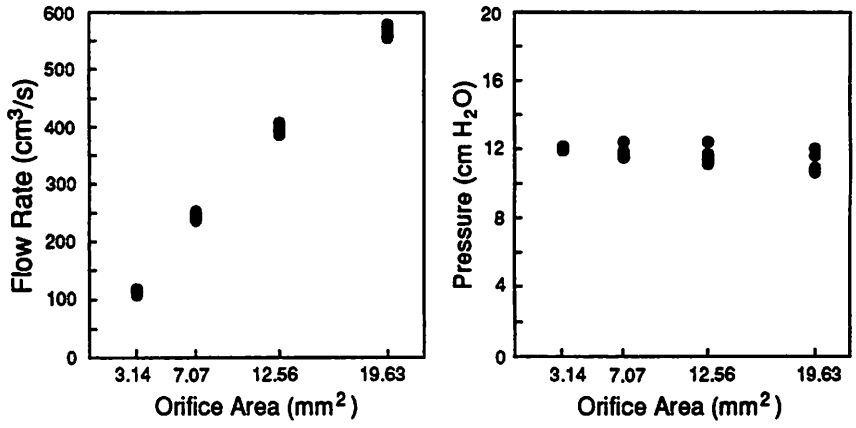


Figure 2. Airflow and pressure values recorded as a function of oral port orifice cross-sectional area using a plastic vocal tract model and vacuum cleaner (constant pressure) air supply.

Individual data points obtained with the vacuum cleaner are shown in Figure 2. Again, the twelve data points for each orifice plate are fairly tightly clustered. It is clear from Figure 2 that increases in orifice area are associated with a) relatively stable intraoral pressure, and b) increasing air flow rates. These data reflect the air pressure and air flow pattern expected of a constant pressure source, and support the characterization of the vacuum cleaner configuration as a constant pressure source.

Discussion

Figures 1 and 2 show that different air flow and pressure patterns can be achieved using the same vocal tract model and orifice configuration, but different air sources. The output resistance of a constant pressure source is much less than the downstream resistance. As a result, alterations in downstream resistance have little effect on airway pressure. However, they have a dramatic effect on airway flow rates, even though the input flow rate to the model is left unaltered. The output resistance of a constant flow source is higher than downstream resistance. As a result, alterations in downstream resistance have little effect on airway flow rates but a dramatic effect on airway pressure. Again, this occurs in the absence of any active manipulation of the air supply.

Experiment 2

In the second experiment, air pressure and flow patterns were recorded from human subjects as they produced speech through an interlabial leak tube similar to that used by Putnam et al. (1987). These measures were then compared to the model data collected in experiment 1.

Method

Five young adults (3 female and 2 male) with an average age of 24 years served as subjects. None reported any history of speech, language, or hearing problems. The speakers produced the syllable [p ʌ] under varying magnitudes of interlabial leak. Subjects produced the syllable with leaks created by a tube placed between the lips. Tubes with cross-sectional areas of 2.46, 19.63, and 50.27 mm^2 were employed. Thirty repetitions of the syllable were produced with each leak tube.

Order of application of the three tubes was counterbalanced. Air flow through the tube and peak intraoral air pressures were obtained from the middle twenty productions of each set.

Air flow was sensed by coupling the pneumotachograph and pressure transducer to the end of the leak tube. Intraoral air pressure was sensed by connecting a length of polyethylene tubing (1.5 mm i.d.) to the leak tube such that its inlet was positioned in the oral cavity. The tube was coupled to a pressure transducer. Air flow and pressure values were recorded on separate channels of a digital audio instrumentation recorder. Prior to measurement, air pressure and flow signals were low pass filtered (30 Hz) and digitized (500 Hz sample rate) using a laboratory computer. Peak airflow and peak pressure identified with [p] were measured.

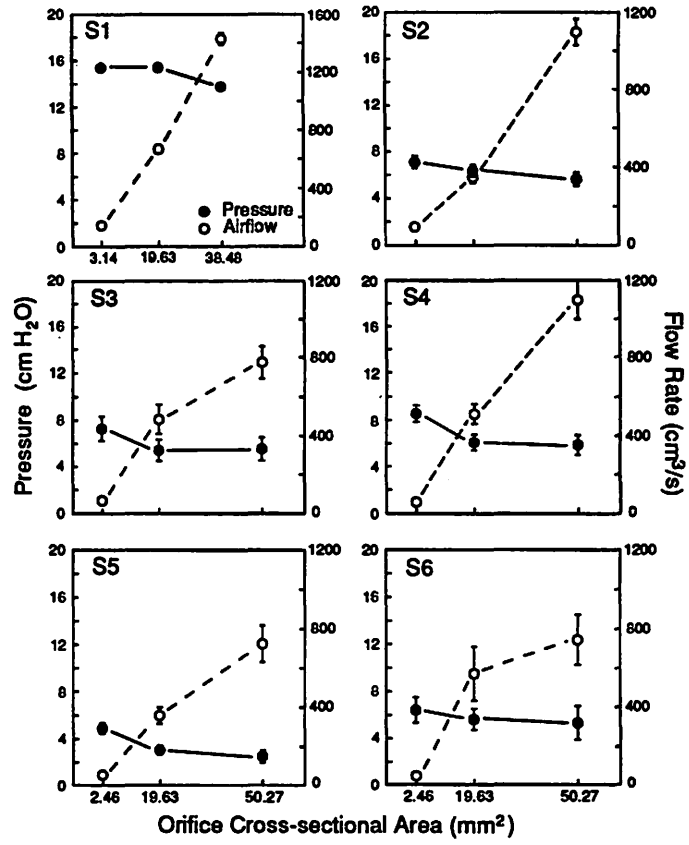


Figure 3. Mean and standard deviation values of intraoral air pressure and interlabial leak tube air flow rate recorded from human subjects.

Results

Means and standard deviations of intraoral air pressure and air flow through the leak tube are shown in figure 3. The five subjects recorded here are S2 through S6. A sixth subject (S1) will be presented below. For all subjects (S1 - S6), intraoral air pressure remained relatively stable as leak tube cross-sectional area increased. In addition, leak tube air flow increased dramatically as leak tube cross-sectional area increased. Finally, the standard deviations of the air pressure and air flow measures are quite low, demonstrating the consistency of the effect within each condition for each subject.

Discussion

Experiment 2 was conducted to compare human data to the data recorded from the plastic vocal tract model used in experiment 1. Comparison of the air flow and pressure patterns depicted in Figures 1, 2, and 3 reveals one obvious similarity: the air flow and pressure patterns depicted in Figure 3 closely resemble those obtained from the vocal tract model using a constant pressure air supply (Figure 2). This relationship is further substantiated by comparing S1 in Figure 3 to S2 through S6. While the pressures and flow rates reported for S1 are higher, the pattern of stable pressure and rising flow is the same. S1 is, in fact, the plastic vocal tract model from experiment 1 with an inflated balloon used as the air source. The inflated balloon is another example of a constant pressure source, and may be considered as a simulation of the elastic properties of the human lung-thorax unit. When the balloon is inflated and allowed to release its air through the vocal tract model, air flow and pressure patterns quite similar to those recorded from human subjects are ob-

served. It should be noted that the balloon pressures are higher than those recorded from S2 - S6. This merely reflects the degree of elasticity of the balloons, which generated higher pressures during their deflation. While absolute pressure and flow values recorded for S1 versus S2 - S6 differ in magnitude, pressure and air flow patterns do not.

Experiment 3

Pressures generated by air flowing through an orifice are related to gas density and orifice flow velocity by the equation

$$P \approx \rho V^2$$

In the case of our interlabial leak condition, P is intraoral air pressure (cm H_2O), ρ is gas density (grams/liter), and V is air velocity (cm/sec). The third experiment was designed to use gas density to manipulate intraoral air pressure and air flow patterns with the plastic model of the vocal tract. Specifically, in a constant pressure source system, variations in gas density should have little effect on airway pressure while airway flow rates should increase as leak orifice area increases. Further, based on the equation above, a reduction in gas density should result in a greater air flow rate for a given leak orifice area.

Method

The plastic vocal tract model utilized in experiment one was used in this experiment. Five orifice plates simulating an oral port were used with openings of 3.14, 7.07, 12.56, 19.63, and 38.48 mm^2. Two gases were employed: air with a gas density of 1.2 g/l and a mixture of Helium (80%) and Oxygen (20%) with a gas density of 0.42 g/l.

Balloons were used as the constant pressure source to facilitate the comparison of air to HeO_2. Preliminary studies showed that pressures generated by a given balloon remained stable across five inflations and deflations. Therefore, each balloon was sued five times and then discarded. Preliminary studies also showed that pressures generated by different balloons were also stable during their first five inflations and deflations. For each trial a balloon was filled with a constant volume of either air or HeO_2, coupled to the trachea of the vocal tract model, and allowed to deflate through the model. Orifice plates were attached to the model's oral port in counterbalanced order. Ten pressure and flow measures were made for each orifice plate under each gas density condition.

The same pressure and flow sensing instrumentation and data recording procedures used in experiment one were employed.

Results

Figure 4 shows means and standard deviations of intraoral air pressure and flow rate through the interlabial leak tube recorded in each of the gas density conditions. While air pressures were slightly higher for the HeO_2 mixture, the difference is less than 10%. Of greater interest is the observation that air pressures for both air and HeO_2 remain fairly stable as a function of increasing orifice cross-sectional area. Further, air flow rates of both gases increase as a function of increasing cross-sectional area of the leak tube.

While both air and HeO_2 flow rates increased, it is clear from Figure 4 that HeO_2 air flow rates increased at a greater rate than did the regular air flow rates. That is, a reduction in gas density

had the predicted effect on air flow rate through a given orifice. A more detailed comparison of the effects of gas density on the observed air versus HeO_2 flow rates was conducted using a modification of the $P \approx \rho V^2$ equation. We measured volume flow velocity (cm^3/sec) as opposed to flow velocity (cm/sec). Since flow velocity equals volume velocity/area, and since orifice area was held constant across gas densities (for each orifice plate), volume velocity may be substituted for velocity in the equation. The predicted ratio of HeO_2 flow to air flow for each orifice plate was compared to the measured ratio based on measured pressure, known gas density and measured volume air flow rate using the formula

$$\frac{U_2}{U_1} = \frac{P_2}{P_1}\left[\frac{\rho_1}{\rho_2}\right]^{1/2}$$

where U_1 and U_2 are volume flow rates (cm^3/sec) for air and HeO_2 respectively, P_1 and P_2 are measured air and HeO_2 intraoral pressures (cm H_2O), and ρ_1 and ρ_2 are air and HeO_2 gas densities (g/l). On average, the predicted U_2/U_1 ratios and measured U_2/U_1 ratios differ by less than 10%, which means that predicted gas density manipulations of air flow and air pressures generated in a constant pressure system were observed.

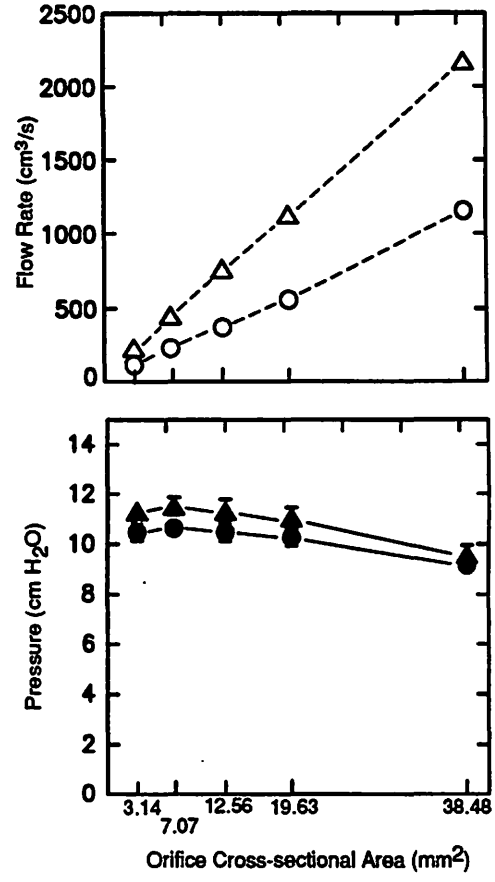


Figure 4. Mean and standard deviation values for airflow and pressure recorded as a function of oral port orifice cross-sectional area using a plastic vocal tract model and air versus HeO_2 delivered from balloons.

Discussion

The results of this experiment demonstrate that an elastic system such as a balloon acts as a constant pressure system and that pressure in a constant pressure system is not influenced by gas density. Gas density influences air flow rate in a predictable manner.

Experiment 4

The fourth experiment was conducted to determine whether the gas density effects observed in experiment 3 would be observed in intraoral pressure and interlabial leak tube flow rates in human subjects. Based on the results of experiment 3, it was predicted that intraoral air pressures generated by human subjects would remain relatively stable across leak tube cross-sectional areas for both air and HeO_2 conditions. It was also predicted that higher air flow rates would be generated in the HeO_2 condition, and that the magnitude of the flow rate increase could be explained by the relations between gas density, air pressure, and air flow rates existing in a passive flow system.

Method

Six young adult speakers (4 female, 2 male) who did not participate in experiment 2 participated in this experiment. Their average age was 28 years. None reported any history of speech,

language, or hearing problems.

As done in experiment 2, the speakers were asked to produce the syllable [p Λ] under varying conditions of interlabial orifice leak. Leaks were achieved by having subjects produce the syllable with tubes of 2.46, 19.63, and 50.27 mm^2 placed between the lips. Order of application was counterbalanced for the three tubes.

Two procedural alterations were made relative to experiment 2. First, 90 dB of white noise masking was used to negate any potential auditory feedback effects caused by the obvious Helium induced changes in acoustic spectra. The masking noise was delivered via insert earphones. Second, the two gas density conditions from experiment 3 were employed. Each gas was delivered to the subject using a face mask coupled to either the air or HeO_2 gas cylinder. Prior to a given trial, subjects were instructed to take five deep inhalations of the gas. This ensured saturation of the lungs with the gas being delivered. Subjects were not told which gas was being delivered to them from trial to trial. Twenty repetitions of [p Λ] were produced with each leak tube. Air flow rate through the tube and peak intraoral air pressures were obtained from the middle ten productions of each set. In some cases a trial had to be repeated in order to obtain a total of twenty repetitions. This occurred most often when subjects attempted to produce the CV syllable through the 50.27 mm^2 leak tube using HeO_2. In these instances, the middle five repetitions of each set were measured. Air pressures and flows were transduced, recorded, and analyzed in the same manner described in experiment 2.

Results

Means and standard deviations for intraoral air pressure and flow rate through the interlabial leak tubes are shown in Figure 5 for each gas density and leak tube condition. A number of observations may be made. First, a higher level of variability was observed here compared to the data recorded in experiment 2 (Figure 3). Second, intraoral air pressures tended to drop more as a function of increasing leak tube cross-sectional area compared to experiment 2. More importantly however, is the observation that similar air pressures were generally recorded for air versus HeO_2 while flow rates increased as cross-sectional area of the leak tube increased. Exceptions were observed for subjects 5 and 6 using the 19.63 mm^2 leak tube. It is also clear from Figure 5 that the HeO_2 flow rates were not systematically higher than the air flow rates, as observed in the model data (Figure 4).

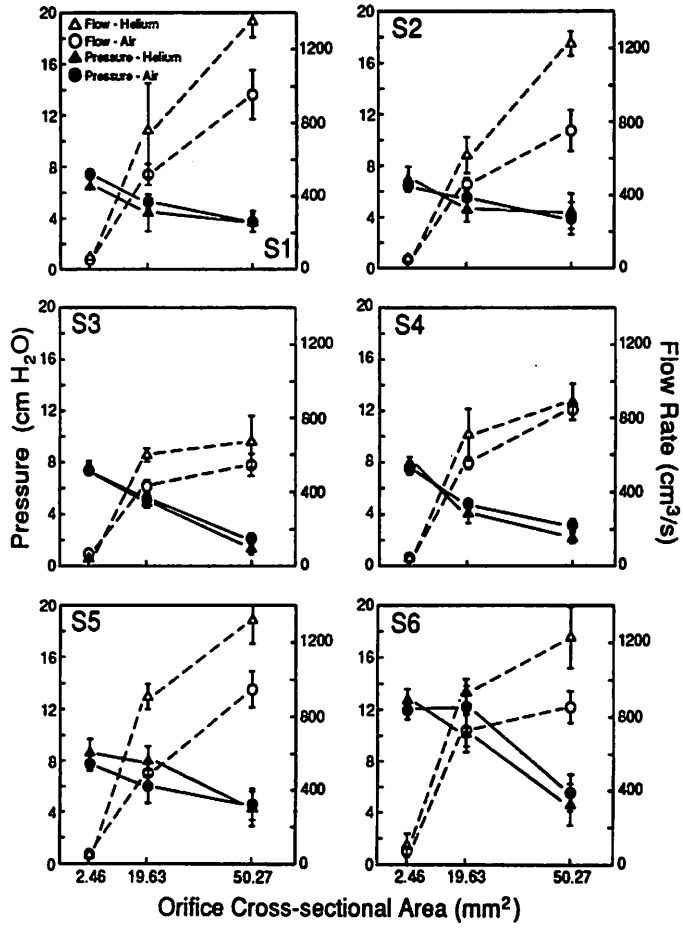


Figure 5. Mean and standard values of intraoral air pressure and interlabial leak tube air flow rate recorded from human subjects using air versus HeO_2.

Air versus HeO_2 air flow rates were compared using the formula employed in experiment 3. Exhaled gas densities measured for one of the investigators were 1.21 g/l for air and 0.46 g/l for HeO_2. When the flow ratio comparison described in experiment 3 was applied to the human data, average predicted and actual HeO_2 flow - normal air flow ratios differed by 69% for the 2.46 mm^2 tube, 9.6% for the 19.63 mm^2 tube, and 10% for the 50.27 mm^2 tube. The relationship observed for the small (2.46 mm^2) leak tube may be explained by the size of the tube itself. That is, the small cross-sectional area may have been flow-limiting, resulting in a low frequency response. As a consequence, the flow rates recorded for that tube may not have accurately reflected a difference between HeO_2 and air.

Discussion

Experiment 4 was conducted to determine whether human subjects would display the same pattern of constant pressure and rising flow as a function of gas mixture observed in the model data (experiment 3). In general, the data depicted in Figure 5 suggest that they do.

There are, however, some differences apparent in figure 5 that warrant discussion. First, a greater amount of variability in air pressure or airflow rate was observed at a given orifice size. However, this same phenomenon was observed in experiment 2 when comparing human to model data, and was not unexpected. that is, there are more inherent sources for variation in the human than in a plastic vocal tract model. More importantly, pressures (both air and HeO_2) were observed to drop, as a function of orifice area, to a greater extent in the human than in the model (experiment 3). In some cases, the drop was also greater than that observed in the humans subjects studied in experiment 2. There is, however, an important methodological difference between experiments 2 and 4. Experiment 4 was conducted using 90 Db of masking noise to prevent speakers from reacting to the obvious acoustic alterations induced by the HeO_2. No such masking was employed in experiment 2. If auditory feedback does play some role in the montoring of speech adequacy, it is feasible that the removal of auditory feedback could result in a reduction of average intraoral air pressure as orifice area increased.

With only two exceptions, air and HeO_2 produced essentially the same intraoral air pressures at a given orifice cross-sectional area. This, coupled with the observation of increasing air flow rate as orifice cross-sectional area increased, lend support to a constant pressure source view of the human respiratory system during speech. That is, an air pressure and air flow pattern similar to that observed in a passive vocal tract with a constant pressure source was observed in data collected from human subjects.

General Discussion

The integrity of the vocal tract airway may be compromised in a number of ways. In the case of repaired palatal clefts, velopharyngeal dysfunction and palatal fistula may result in unwanted leakage of air. Airway leaks may be induced in normal speakers by inserting leak tubes into the velopharyngeal aperture or between the lips. In each case, it has been observed that, during the production of bilabial plosive speech sounds, average intraoral air pressures tend to be maintained above a minimal level of about 3 to 4 cm H_2O. The physiologic basis for this phenomenon has been the target of this and numerous other investigations. Three possibilities were proposed by Warren et al. (1989b).

Warren and colleagues (Warren, 1986; Warren et al., 1989a, 1989b, 1992; Putnam et al., 1986) have postulated that air pressure regulation is a primary goal of the speech motor control

system. At least in the case of bilabial plosive production, pressure regulation is controlled by active compensatory manipulation of respiratory effort. Of the evidence presented in support of pressure regulation, the most notable is the observation that pressure drops associated with increasing the magnitude of an airway leak are much greater in a passive vocal tract model than in the human speaker. Further, increases in the size of an airway leak in the human speaker are also associated with increased air flow rate and air volume usage (Warren et al., 1989b;1992). However, the results of the present investigation do not provide support for Warren's hypothesis. Pressure drops recorded from human speakers in association with increasing the magnitude of airway leak did not differ from those recorded in a plastic vocal tract model when a constant pressure air source is used.

A second possibility is that pressure regulation is a secondary system goal. For example, Netsell (1990) and Moon and Folkins (1991) addressed the idea that pressure regulation was secondary to auditory perceptual regulation. Aerodynamic compensations would be induced by the requirement to produce perceptually accurate sound output. While Moon and Folkins (1991) demonstrated that intraoral pressure could be manipulated by altering auditory feedback, the results did not provide strong support for an auditory perceptual regulation hypothesis over a pressure regulation hypothesis. However, as discussed ealrier, some support for the role of auditory feedback may be found in the present investigation.

The results of this experiment suggest a third possibility. During speech, the human respiratory system may be modeled, at least to a first approximation as having a low output resistance relative to resistances likely to be encountered in the vocal tract during speech production (i.e. it would be a constant source of air pressure). Therefore, downstream resistances would by definition have little effect on respiratory system output pressure, but a more profound effect on airway flow rates.

The lung-thorax unit is an elastic system that, once displaced from rest, will generate subglottal pressures by recoiling. This recoil combined with respiratory muscle activity may allow the respiratory system to maintain pressure like the balloon used in this experiment. Specifically, observation of a typical pressure-volume curve depicting the recoil properties of the lung-thorax unit reveals that, in the lung volume region typically employed during speech production, recoil provides subglottal pressures above 5 cm H_2O (Comroe, 1965). Motor programs controlling the respiratory muscles during speech may supplement recoil forces to establish a steep pressure volume curve. This is in contrast to the notion suggested by Warren (1976) that changes in upper airway respiratory load require rapid respiratory adjustments, mediated by sensory information, to maintain airway pressures. While Warren does suggest that "simultaneous pre-programming for the respiratory muscles ... is also a possibility" (p. 133), he argues that sensory information is used in a servo mechanism to drive the respiratory system reflexively.

It is of interest to interpret some published data referred to earlier in conjunction with the pressure regulation hypothesis assuming a constant pressure air source. Based on the air pressure and air flow patterns observed in experiment one, it is apparent that earlier modelling data (e.g. Warren and Devereux, 1966; Warren, 1979) was obtained using a constant flow source. The potential effect that use of a constant pressure source might have had on the development of the orifice area estimation technique (Warren and DuBois, 1965) is of interest. Table 1 (see following page) depicts average estimated orifice area using the hydrokinetic equation calculated with constant flow and constant pressure source data from the present study. Similar orifice area estimations are obtained regardless of whether the constant flow or constant pressure source is used. All that differs

across the two air source types is the pattern of flow and pressure change associated with variations in orifice area size. As shown in the present investigation, this difference is significant however.

Active respiratory responses to increased airway leak were thought by Putnam et al. (1986) to result in maintained intraoral air pressures and increasing interlabial bleed flow rates. Stable air flow rates on the vowel of the CV syllables studied were thought to reflect "some accommodating adjustments in laryngeal airway resistance to normalize vowel flow in spite of the aeromechanical perturbations introduced during the preceding [p] and [s] segments" (p. 47). The relatively stable air pressure and increasing air flow rates observed by Putnam et al. (1986) during bilabial plosives can be explained by the constant pressure source hypothesis. Laryngeal vibration during the postconsonant vowel would be expected to produce a level of airway resistance that would be unaffected by interlabial bleed size since the oral port for the vowel would be open. The stable postconsonant vowel airflow rates observed in association with increases in interlabial bleed area would be expected given a constant pressure air supply and a consistent laryngeal airway resistance from production to production.

Table 1.

Orifice area values (mm^2) estimated using the hydrokinetic equation and air flow and pressure values from constant flow (air cylinder) versus constant pressure (vacuum blower) air sources.

Known Area	Cylinder Est Area	Cylinder % Error	Vacuum Est Area	Vacuum % Error	Cyl-Vac % Diff
3.14	3.18	1.53	3.20	2.16	0.82
7.07	6.87	-2.72	6.94	-1.79	1.10
12.56	12.21	-2.76	11.41	-9.10	-6.43
19.63	18.60	-4.86	17.29	-11.59	-6.96

Warren et al. (1989b) used respiratory induced plethysmography to demonstrate that lowering of the velum resulted in an increased expiratory volume during production of "papa". No differences in inspiratory speech volumes were observed. Further, no differences in expiratory duration were noted between normal and lowered velum conditions. These data may also be explained by considering the respiratory system to be a constant pressure source during speech. If vocal tract resistance decreases, the passive result would be an increase in expiratory volume. This is, in fact, what Warren et al. (1989b) observed.

Warren et al. (1992) used bite block and interlabial bleed tubes to demonstrate the role of both respiratory and lingual compensatory behaviors in maintaining intraoral air pressures during speech. While the relatively stable air pressures and increasing air flow rates observed during the interlabial bleed tube condition were attributed by Warren et al. to increased respiratory effort, they can also be explained by the constant pressure source hypothesis. Warren et al. (1992) attributed stable air flow rates observed in association with increases in bite block size to increased respiratory and lingual effort to prevent a drop in pressure. However, the same data can be explained by a combination of a constant pressure source and lingual activity designed to produce the appropriate

auditory perceptual output even though physiologic constraints may prevent the typical percept from being achieved.

As stated earlier, the most notable piece of evidence in support of active pressure regulation involves a comparison of human data to data collected from a plastic vocal tract model (Warren et al., 1989a). We have suggested that, based on air flow and air pressure patterns presented, a constant flow source was used to generate their model data. Warren et al. (1992) interpret their data to show that the human respiratory system is not a constant flow system. They argue, in fact, that it is a constant pressure system and that pressure constancy is mediated by reflexively modulated air flow adjustments. We would agree that intraoral air pressures are maintained by both normal speakers and those with velopharyngeal dysfunction in the face in varying magnitudes of vocal tract air leak. However, the arguments offered in support of active pressure regulation appear to be more consistent with a constant flow source characterization of the human respiratory system during speech. That is, intraoral air pressures would drop if air flow were not actively manipulated.

Summary and Conclusions

The results of the present investigation provide support for a different view of speech aerodynamics regulation. Our view of the respiratory system as a constant pressure source suggests that relative pressure stability observed in association with increasing the magnitude of an airway leak may be due to inherent physiologic properties of the human respiratory system.

Four modelling experiments were conducted in an attempt to gain a better understanding of the physiologic bases for pressure maintenance observed during speech production by individuals with velopharyngeal dysfunction. Pressure regulation and auditory perception hypotheses have been investigated in previous research. Data from the present investigation appear to provide support for a constant pressure source explanation of intraoral air pressure and air flow patterns observed in speakers with velopharyngeal dysfunction. More specifically, the observed pressure and flow patterns can be explained as a consequence of the constant pressure source characteristics of the respiratory system during speech. Our data may be taken as preliminary evidence of the ability to generate constant vocal tract pressure without the necessity for sensory information, including that which might elicit reflex responses of the respiratory system to alterations in vocal tract resistance.

The results of this experiment support a conceptualization of the human respiratory system as a balloon providing constant pressure regardless of changes in downstream resistance. In contrast, the active pressure regulation view of speech production is consistent with a conceptualization of the respiratory system as a compressed air cylinder. That is, a decrease in downstream resistance will result in a loss of pressure unless airflow rate is actively increased in response to sensory feedback. Resolution of these different explanations for pressure regulation will require systematic study of the short latency effects of vocal tract resistance perturbation on respiratory performance.

Acknowledgements

The authors would like to thank Professor David Cater, University of Iowa Department of Chemistry and Professor Ingo Titze, University of Iowa Department of Speech Pathology and Audiology, for their assistance. We would also like to thank Professor Don Warren, University of North Carolina Department of Dental Ecology for his comments on an earlier draft of this manuscript. This work was supported by NIH Grant DC00976. Portions of this manuscript were presented at the fall meeting of the Acoustical Society of America, San Diego, 1990, and at the annual meeting of the American Cleft Palate-Craniofacial Association, Portland, 1992.

References

Comroe, J. (1965). Physiology of Respiration (Year Book Medical Publ., New York).

Moon, J., and Folkins, J. (1991). "The effects of auditory feedback on the regulation of intraoral air pressure during speech," Jour. Acoust. Soc. Amer. 90, 2992-2999.

Netsell, R. (1990). Commentary on "Maintaining speech pressures in the presence of velopharyngeal impairment," Cleft Palate-Craniofacial Jour. 27, 58-60.

Putnam, A., Shelton, R., and Kastner, C. (1986). "Intraoral air pressure and oral air flow under different bleed and bite-block conditions," Jour. Speech Hear. Res. 29, 37-49.

Warren, D. (1976). Aerodynamics of speech production. In (N. Lass, Ed.) Contemporary Issues in Experimental Phonetics (Academic Press, New York), p. 105-137.

Warren, D. (1979). "Perci: A method for rating palatal efficiency," Cleft Palate Jour. 16, 279-285.

Warren, D., and Devereux, J. (1966). "An analog study of cleft palate speech," Cleft Palate Jour. 4, 38-46.

Warren, D., Duany, L., and Fischer, N. (1969). "Nasal pathway resistance in normal and cleft lip and palate subjects," Cleft Palate Jour. 6, 134-140.

Warren, D., and DuBois, A. (1964). "A pressure-flow technique for measuring velopharyngeal orifice area during continuous speech," Cleft Palate Jour. 1, 52-71.

Warren, D., Allen, G., and King, H. (1984). "Physiologic and perceptual effects of induced anterior open-bite," Folia Phoniatrica 36, 164-173.

Warren, D., Dalston, R., Morr, K., Hairfield, M., and Smith, L. (1989a). "The speech regulating system: temporal and aerodynamic responses to velopharyngeal inadequacy," Jour. Speech Hear. Res. 32, 566-575.

Warren, D., Morr, K., Rochet, A., and Dalston, R. (1989b). "Respiratory response to a decrease in velopharyngeal resistance," Jour. Acoust. Soc. Amer. 86, 917-924.

Warren, D. (1986). "Compensatory speech behaviors in individuals with cleft palate: a regulation/ control phenomenon?," Cleft Palate Jour. 23, 251-260.

Warren, D., Hinton, V., Pillsbury, H., and Hairfield, W. (1987). "Effects of size of the nasal airway on nasal airflow rate," Arch. Otolaryngol.-Head and Neck Surg. 113, 405-408.

Warren, D., Rochet, A., Dalston, R., and Mayo, R. (1992). "Controlling changes in vocal tract resistance," Jour. Acoust. Soc. Amer. 91, 2947-2953.

Williams, W., Brown, W., and Turner, G. (1987). "Intraoral air pressure discrimination by normal-speaking subjects," Folia Phoniatrica 39, 196-203.

NCVS Status and Progress Report - 3
December 1992, 97-115

Comparison of F_0 Extraction Methods for High Precision Voice Perturbation Measurements

Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
and
Recording and Research Center, The Denver Center for the Performing Arts
Haixiang Liang, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Voice perturbation measures, such as jitter and shimmer, depend on accurate extraction of fundamental frequency (F_0) and peak-to-peak amplitude of various waveform types. The extraction method directly affects the accuracy of the measures, particularly if several waveform types (with or without formant structure) are under consideration and if noise and modulation are present in the signal. For frequency perturbation, high precision is defined here as the ability to extract F_0 to ±0.01% under conditions of noise and modulation. Three F_0-extraction methods and their software implementations are discussed and compared. The methods are cycle-to-cycle waveform matching, zero-crossing and peak-picking. Interpolation between samples is added to make the extractions more accurate and reliable. The sensitivity of the methods to different parameters such as sampling frequency, mean F_0, signal-to-noise ratio, frequency modulation, and amplitude modulation, are explored.

Introduction

This paper is a continuation of several reports dealing with refinements of F_0 extraction for voice perturbation measures. Normative data on jitter (Horii, 1979) may need to be reevaluated in light of recent technological developments in recording and analysis. It will be shown that jitter in normal voices is less than half of what has traditionally been reported if more attention is paid to recording fidelity and extraction algorithms. This report is directed, therefore, toward high-preci-

sion F_0 extraction when perturbations are less than about 5%. It does not deal with sudden F_0 jumps (e.g., period doubling or intermittent phonation) found in some pathologic voices and in some types of vocal fry.

Various F_0 extraction methods have been summarized by Hess (1983). They can be classified into two major categories: event-detection methods and short-time average methods. Examples of event detectors are various forms of the peak-picking and zero-crossing algorithms. Examples of short-term average F_0 detectors are autocorrelation, minimal distance methods, ceptral analysis, and harmonic compression. For the latter class of methods, a window is applied to limit the signal over a few analysis cycles.

The objective of voice perturbation analysis is to determine the degree to which vocal fold vibration is aperiodic. The analysis can be applied to signals obtained from a microphone, an airflow mask placed over the mouth, an accelerometer attached to the neck, or an electroglottograph. Perturbation measures are defined as an average deviation from periodicity over several cycles.

Three factors need to be understood in voice perturbation measures obtained in the laboratory: (1) the physical perturbation created by vocal fold vibration; (2) the perturbation introduced by hardware due to additive noise and signal distortion; (3) the perturbation introduced by software (inexact extraction algorithms). The analysis objective is, of course, to find the perturbation due to the first factor, but unfortunately the additional undesired perturbations introduced by the second and third factors are often unavoidable, and one can only try to minimize their effects.

To examine hardware effects, Doherty and Shipp (1988) compared four different tape recording systems with direct analog to digital conversion (ADC). They found that digital recording systems do not significantly inflate normal vocal jitter and shimmer, but conventional analog recorders can. Work in progress on the effect of microphones on voice perturbation (Titze & Winholtz, in review) suggests that mouth-to-microphone distance can be critical for low perturbations.

With regard to the accuracy of extraction methods, Titze, Horii and Scherer (1987) determined that interpolation between samples is necessary unless high sampling rates (> 500 samples per cycle) are used. Milenkovic (1987) found that greater reliability and accuracy could be obtained by matching the entire waveshape across adjacent cycles rather than isolated events, like zero-crossings or peaks. In this waveform matching strategy, the F_0 difference between periods is decided by a least-squared difference between adjacent waveshapes. Deem, Manning, Knack and Matesich (1989) compared 12 different combinations of zero-crossing and peak-picking strategies (with and without interpolation and signal inversion), using direct ADC recording. They used sine waves and actual voice stimuli to compare the performances. For sine waves (perhaps as expected), zero-crossing with interpolation gave the lowest jitter values, an order of magnitude lower than peak picking without interpolation, which was the poorest. Peak picking for signals with broad peaks is notoriously poor because small amounts of additive noise can shift the location of the peaks. The present study builds on the work of Deem et al in the sense that it compares the waveform matching technique to the more traditional peak picking and zero crossing techniques.

It should also be pointed out, as Hillenbrand (1987) did with a formal investigation, that frequency perturbation, amplitude perturbation, and additive noise are seldom independent. Their origins cannot easily be determined in a voice signal because they involve source-filter interactions for which neither source nor filter is known precisely. In a synthetic signal, however, known amounts of modulation can be imposed at the source. The test is then to determine which algorithms can best recover the known frequency modulations.

Methods of F_0 Extraction

Since the subject of concern here is *high-precision* perturbation analysis, for which cycle-to-cycle variations are required, the short-term average methods are not applicable. An updated value of F_0 is needed for every cycle rather than over a window of several cycles. Therefore, the focus of discussion is on event detection methods. Three F_0 extraction methods are implemented and compared:

(1) WM: Waveform Matching over adjacent cycles, using a least squared error criterion
(2) PP: Peak-picking
(3) ZC: Zero-crossing after low-pass filtering

The WM method is strictly not a single-event method. Rather, it measures the combined change of many events from cycle to cycle. The PP and ZC methods do measure single events, however.

Although in principle there is no preference for extracting positive peaks over negative peaks, or positive-going zero crossings over negative-going ones, a choice is made here to adopt negative peaks and positive-going zero crossings. Negative peaks are usually closer to the major "excitation event" of the vocal tract, as are positive-going zero crossings when an electroglottograph (EGG) is used as the voice signal. A user can always invert the signals to test the procedures on opposite phases.

In all cases, interpolation between samples is used. It is known that interpolation will improve the accuracy of F_0 extraction when the sampling frequency is in the typical 10-20 kHz range (Titze, Horii & Scherer, 1987; Milenkovic, 1987).

Definition and Need for High Precision

To set up some guidelines toward a standard, we define high-precision F_0 extraction as the ability to resolve F_0 to 0.01% accuracy in a recorded speech signal. This means that a 1% perturbation should be resolvable to a 1% accuracy. It also means that a 0.1% perturbation should be resolvable to at least a 10% accuracy. This is not a strict requirement, because 0.1% F_0 perturbations are in the range of normal for sustained vowels.

To illustrate the point, Table 1 (following page) shows 120 frequencies extracted from 40 tokens of sustained [a] vowels (2s in duration) as uttered by 10 normal male and 10 normal female subjects. Each of the 20 subjects gave two tokens, one for a low pitch and one for a higher pitch. For each of the 40 tokens, the mean F_0 was extracted by the three methods: WM, PP and ZC. The algorithms used to extract the values will be described in detail later. By comparing the three adjacent columns in four quadrants of the Table (low F_0 males, high F_0 males, low F_0 females, and high F_0 females), it is seen that the methods agree in absolute F_0 to within 0.01%. Exceptions are marked with an asterisk (explained below). In most cases, there is a difference only in the fifth decimal place. Given this high level of agreement between WM, PP and ZC in terms of the mean F_0, one would expect a close agreement for the jitter values.

This is not the case, however. Table 2 shows the extracted jitter values for the identical tokens. Note that there are major discrepancies between the perturbations obtained with the three extraction methods. This suggests that different events jitter in different ways. Peaks jitter more than the zero crossings of the low-pass filtered waveform, which in turn jitter more than the least

Table 1.

Extracted F_o for identical tokens of a 2s [a] vowel, using waveform matching (WM), peak picking (PP) and zero crossing (ZC). Subjects were ten normal male speakers and ten normal female speakers.

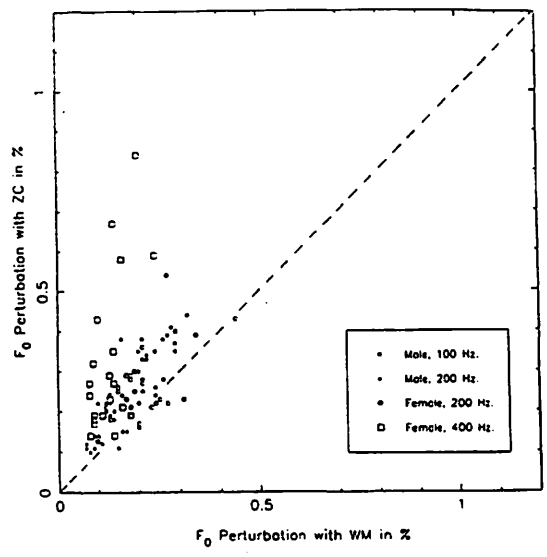
	Low F_o			High F_o		
	WM	PP	ZC	WM	PP	ZC
Male						
0	96.99	96.99	96.99	197.19	197.18	197.16*
1	98.99	98.98	98.98	184.69*	193.66	193.63
2	119.49	119.49	119.48	208.95	208.96	208.95
3	110.36	110.40*	110.36	165.04	165.04	165.04
4	112.69	112.69	112.68	220.37	220.47*	220.37
5	96.14	96.14	96.14	193.46	193.46	193.46
6	99.39	99.44*	99.40	196.46	196.47	196.47
7	101.24	101.23	101.23	197.89	197.89	197.89
8	107.41	107.34*	107.39	199.33	199.33	199.31
9	125.47	125.48	125.48	194.23	194.22	194.22
Females						
0	199.72	199.75	199.72	394.83*	394.90	394.89
1	195.80	195.80	195.79	393.03	393.04	393.04
2	197.86	197.87	197.87	396.51	396.50	396.51
3	195.53	195.53	195.54	391.84	391.86	391.86
4	199.70	199.72	199.72	394.85	394.86	394.87
5	207.01	207.02	207.02	420.13	420.17	420.17
6	198.87	198.88	198.88	401.08	401.14	401.14
7	233.39	233.40	233.39	387.50	387.51	387.51
8	194.80	194.80	194.80	395.16	395.17	395.18
9	209.99	210.01	209.98	491.72*	500.30	500.29

Table 2.

Jitter values for identical tokens of a 2s [a] vowel, using waveform matching (WM), peak picking (PP) and zero crossing (ZC). Subjects were ten normal male speakers and ten normal female speakers.

	Low F_o			High F_o		
	WM	PP	ZC	WM	PP	ZC
Male						
0	.20	.25	.30	.14	.28	.29*
1	.22	.31	.34	.56*	.25	.17
2	.29	.45	.37	.10	.19	.14
3	.72	1.84*	.78	.13	.21	.19
4	.24	.75	.26	.32	1.06*	.19
5	.21	.34	.28	.10	.20	.13
6	.56	.61*	.51	.07	.13	.12
7	.24	.38	.35	.20	.40	.16
8	.54	1.50*	.41	.17	.38	.15
9	.24	.50	.24	.12	.27	.22
Females						
0	.20	.34	.17	.20*	.70	.25
1	.20	.47	.30	.09	.23	.17
2	.18	.37	.28	.11	.32	.19
3	.17	.31	.29	.08	.34	.24
4	.34	.59	.39	.14	.34	.67
5	.18	.36	.19	.10	.46	.13
6	.15	.65	.26	.09	.62	.19
7	.25	.67	.23	.16	.44	.21
8	.13	.25	.18	.24	.39	.59
9	.23	.50	.21	.83*	1.19	.25

squared error between all the events (in waveform matching). This is shown graphically with two scatter plots, Figures 1 and 2. In Figure 1, jitter as extracted by the ZC method is plotted against jitter extracted by the WM method. Generally speaking, ZC values are greater than WM values, particularly for the high F_0 female tokens. In Figure 2, PP values are seen to be even larger than WM values.



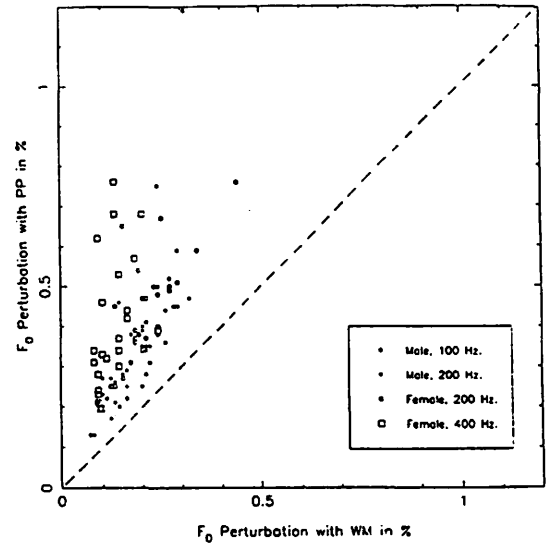


Figure 1. Fundamental frequency perturbation (jitter) as extracted by ZC (zero crossing) versus WM (waveform matching), performed on identical records of 20 normal subjects. The utterance was a 2s sustainted [a] vowel.
Figure 2.Fundamental frequency perturbation (jitter) as extracted by PP (peak picking) versus WM (waveform matching), performed on identical records of 20 normal subjects. The utterance was a 2s sustained [a] vowel.

The cases noted with an asterisk in Tables 1 and 2 were not included in the scatter plots. For these cases there was an extraction failure (to be discussed later). By showing only the "correct" cases, the relative differences between extraction methods can be assessed. For example, Figures 3, 4 and 5 show analysis details for one of the 40 tokens in Table 2, namely the high F_0 female token for which WM gave a value of 0.16, PP a value of 0.44, and ZC a value of 0.21.

Consider first the F_0 profile obtained by the PP method, shown in Figure 3 (next page). At the top is the F_0 contour for 775 periods. This contour contains both long-term and short-term fluctuations in F_0. Immediately below the contour is the second-order perturbation function (Pinto & Titze, 1990), which is basically the second derivative of the contour. It shows only the short-term fluctuations. In the bottom half of the figure, we see the autocorrelation function of the F_0 contour, a histogram of all F_0 values, and the power spectrum of the F_0 contour. All of these plots provide complementary information. They highlight different F_0 features. For example, the autocorrelation complements the second order perturbation function in that it isolates the long-term fluctuations instead of the short-term fluctuations. The histogram gives a quick view of the distribution of F_0 values and the power spectrum shows dominant modulation frequencies in the F_0 contour (e.g. 5 Hz).

Figures 4 and 5 show the identical segment analyzed with ZC and WM, respectively. The starting sample, ending sample, and number of cycles analyzed were exactly the same. Note the similarities in the long-term fluctuations (autocorrelation and lower region of the power spectrum) between all three figures. The long-term patterns are virtually indistinguishable from each other. However, the short-term fluctuations (the high frequency noise in the counter and the second-order perturbation) are remarkably different for the three methods. This is also seen in the fine detail of the histogram and the higher region of the power spectrum. We are lead to the conclusion that *the method of extraction is not simply a matter of precision, but a matter of fundamental choice about the microevents to be described within a cycle.* Peaks jitter more than zero crossings (on a low-pass filtered waveform), and the entire waveshape jitters less than either of the two.

Returning to the question of precision, however, it appears that differences of the type noted here cannot be resolved unless perturbations on the order of 0.1% can be extracted with at least a 10% accuracy. For the entire collection of 40 tokens, jitter with WM averaged at 0.175%, with ZC at 0.265%, and with PP at 0.388%. Standard deviations were 0.07%, 0.13%, and 0.17%, respectively. These values are all lower than the 0.5-0.7% range of average values reported by Horii (1979) for normal phonation. Although Horii used the PP approach, which gives the highest values, his earlier data were probably inflated by a factor of 2 due to insufficient sampling (without interpolation) and tape recording (Titze, Horii and Scherer, 1987). The normal jitter values reported here are more in line with those reported by Milenkovic (1987), who used the least squared waveform matching approach.

The foregoing preliminary analysis sets the stage for a thorough description of the methods. Following this description, a parametric study is conducted with synthetic stimuli.

Markers for Approximate Period Boundaries

One common procedure for all three methods is that rough period markers are first selected. This requires a preliminary pass through the data. The markers are placed on negative-going zero crossings (Figure 6, long vertical lines). If an EGG signal is available, this signal can be used for rough F_0 extraction because it has only two midlevel-crossings[1] per cycle. For microphone signals (if an EGG waveform is not available), the signal is first low-pass filtered (FIR filter, 200-300 points) to remove formant ripple. Typically, the cutoff frequency is selected between F_0 and 2 F_0. The periods between crossings are checked against a

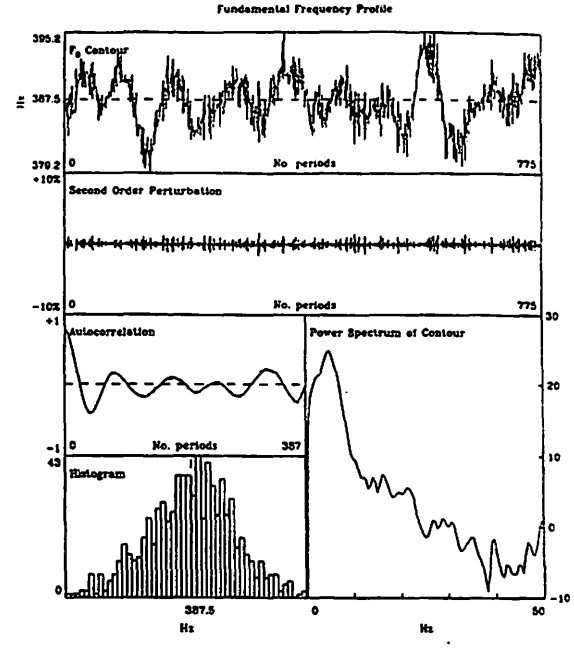
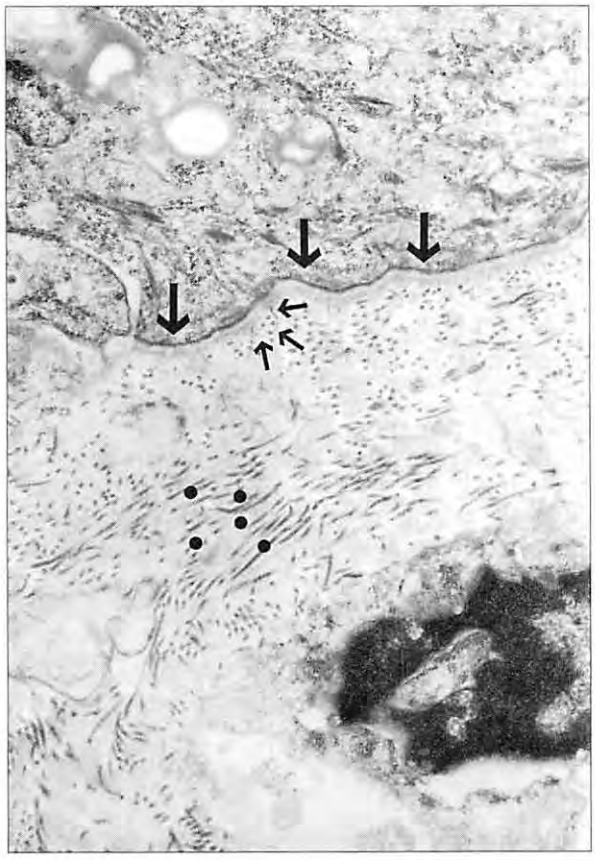


Figure 3.Fundamental frequency profile obtained with the PP method.

[1]Midlevel-crossing is used here instead of zero-crossing because some signals may have very good low-frequency response, exhibiting an average drift over many cycles. The midlevel value, defined as (absolute maximum + absolute minimum)/2 over the analysis segment, is then not necessarily zero. Alternately, the signal can be highpass filtered to remove the drift.



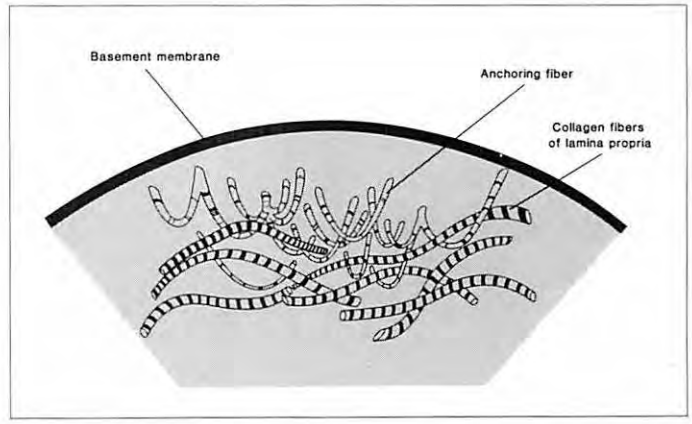
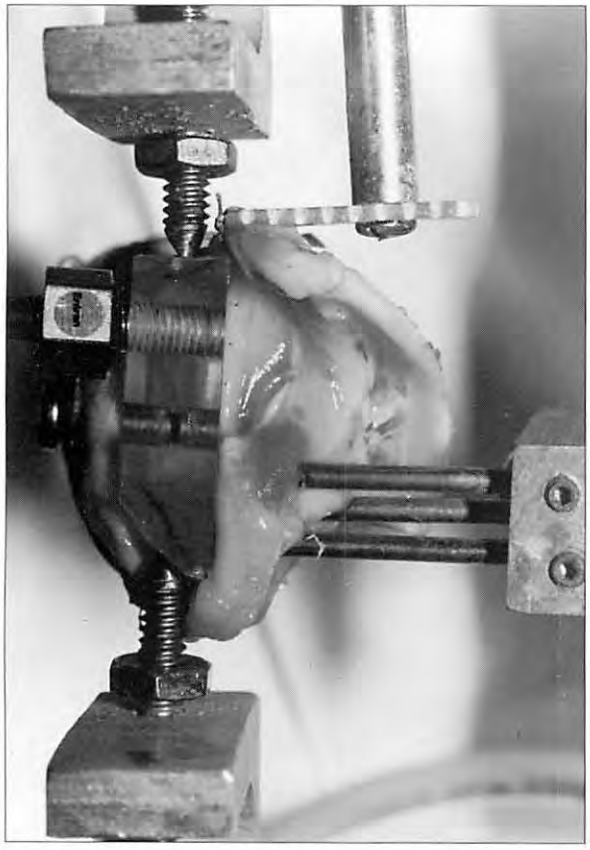
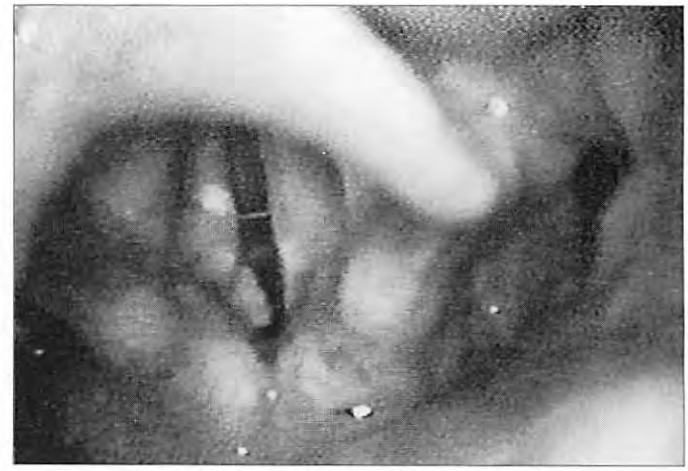
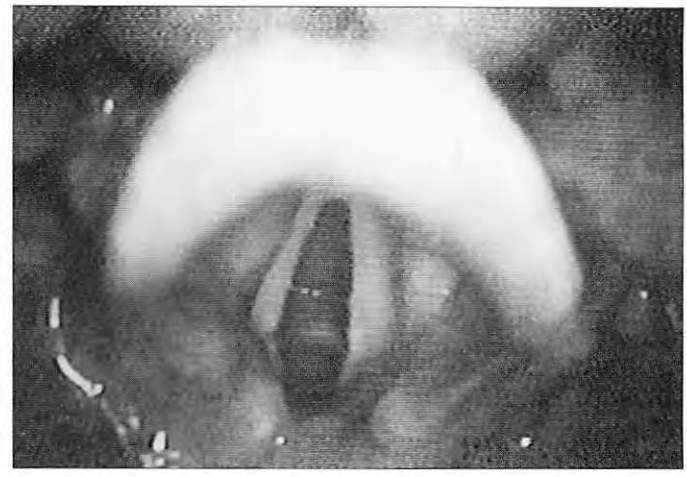


Fig. 1A (above) Epidermal-lamina propria junction of the normal human vocal fold. Note basal lamina (large arrows), collagen type III fibers of the superficial lamina propria (circles) and small anchoring fibers making small loops into the lamina propria (small arrows). (10,000x). Fig. 1B (top right) Schematic drawing of basement membrane/ basal lamina of cell, anchoring fibers which attach to the lamina densa of the basement membrane, and type III collagen fibers which pass through the loops of the anchoring fibers. Fig. 2 (center right) Electron micrograph of anchoring fibers in human vocal fold. Lamina densa (small circles), anchoring fibers (small arrows); type III collagen fiber (large arrows). Note fibers passing through loops. (50,000x). Fig. 3 (bottom right) Photograph showing lateral view of a hemilarynx (anterior to the left, posterior to the right). Half of the thyroid cartilage and the left vocal fold are removed to facilitate mounting the vertical plate.







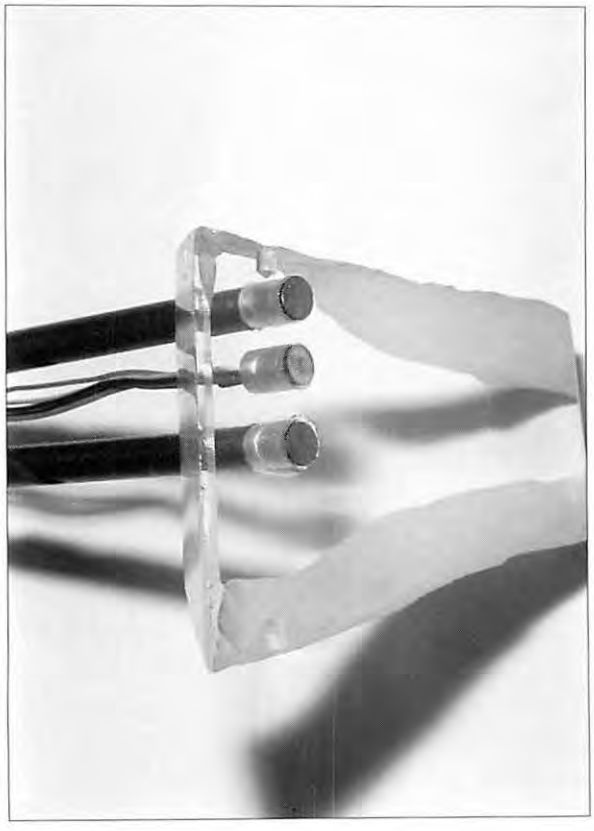
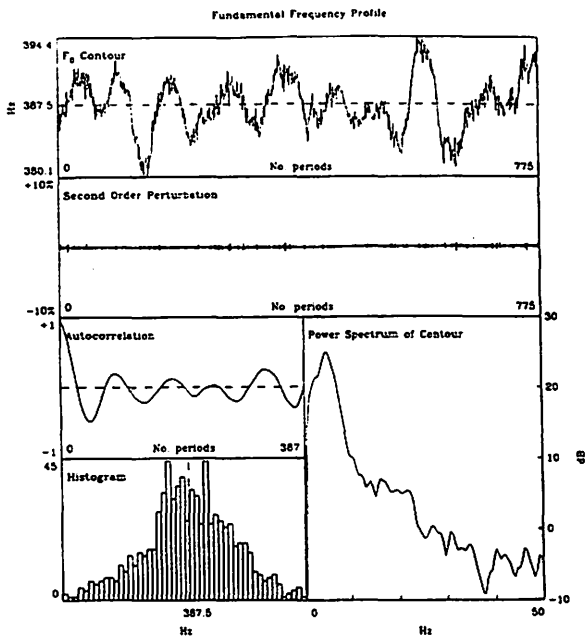


Fig. 4 (upper left) Photograph showing top view of the hemilarynx with plexiglass plate replacing the left vocal fold. The right fold is abutted against the plate, which is mounted by two screws (top and bottom). Additional mounting screws are seen shining through the plate from below. A three-pronged device (middle-right) is used to adjust the arytenoid cartilage, and a rod (upper-right) is used to adjust the vocal fold length. Fig. 5 (lower left) Photograph showing plexiglass plate from the vocal fold side. In the center hole is the stress-sensitive stainless steel diaphragm. Top and bottom holes are plugged with machine screws and rods inserted at the same depth making a continuous flat surface. Fig. 6 (upper right) Pre-surgical laryngeal appearance. Fig. 7 (lower right) Immediate post-surgical view.



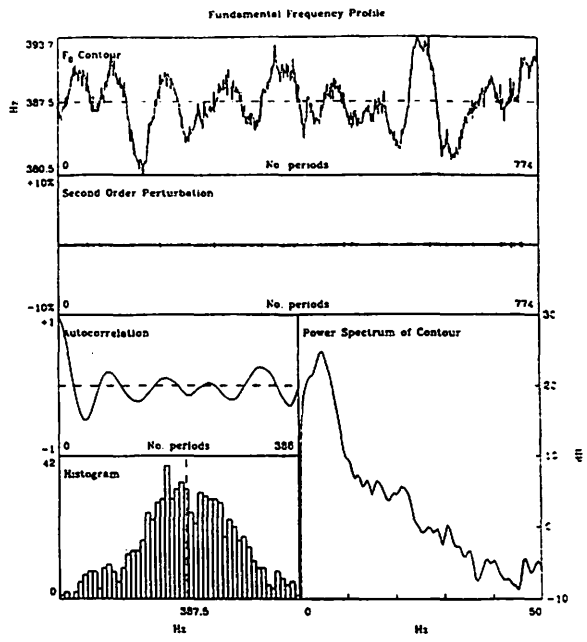


Figure 4 (left). Fundamental frequency profile obtained with the ZC method. Figure 5 (right). Fundamental frequency profile obtained with the WM method.

user-specified frequency range (e.g. 50 Hz to 400 Hz). In the event that a period falls outside of this range, an error flag is set and several consecutive periods can be averaged over a given number of cycles (e.g. 10-20). The crossings are then re-marked according to the mean for the segment. Any period falling out of the specified frequency range is not counted towards the mean. This eliminates any undesired bias toward highly irregular periods. The feature was implemented to study pathologic voices, where a few "rough" cycles are often seen among a string of well-behaved ones. It has less relevance to the present study in which perturbations are kept small. No subharmonics or sudden discontinuities in voicing are included here.

After the rough cycle division markers have been selected, one of the three methods is used to find the periods more accurately, usually *on the non-filtered waveform*. The markers extend across both filtered and unfiltered waveforms simultaneously. The phase relationship between the markers and specific events in the unfiltered waveform is totally arbitrary. It does not need to be known. The markers serve only as a "one cycle window" wherein the search for events is conducted.

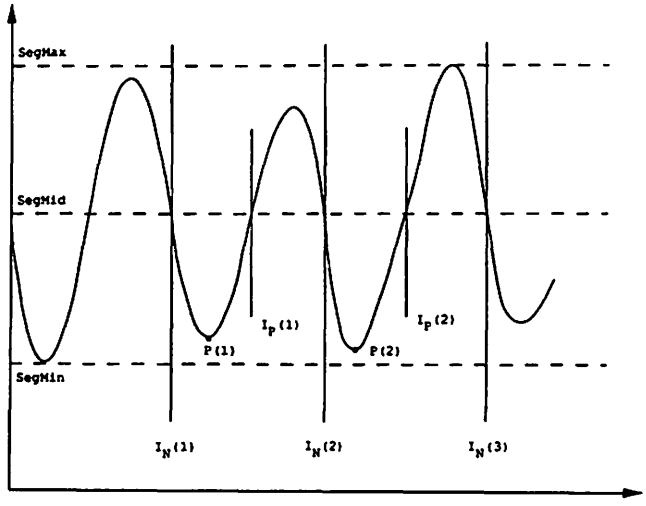


Figure 6. A signal showing how the cycle markers are placed.

The WM method

This method is a special case of the method described by Milenkovic (1987). Milenkovic's amplitude scaling factor is not used for computation of the least squared difference between adjacent waveforms. The procedure starts by searching for the location of the absolute minimum between the first two rough markers [Figure 6, point P(1)]. A point P(2) between the 2nd and 3rd markers is then found such that the mean squared error between the two adjacent waveforms is minimal. This procedure is repeated until all cycles are processed.

Mathematically, the method can be explained as follows. Assume buff(k) represents the data buffer, N_c+1 the number of rough markers (where N_c is the number of cycles), $I_N(i)$ the array of integer rough markers (negative zero-crossing), P(i) the array of integer negative peak locations, and $F_o(i)$ the fundamental frequency array. Then the procedure is:

(1) Find the negative peak location P(1) between $I_N(1)$ and $I_N(2)$.

(2) For i = 2 to N_c+1

(a) Initial guess: $P(i) = I_N(i) + (P(i-1)-I_N(i-1))$

(b) Set lower and upper looking limits:
$J_1 = P(i) - PERC * (P(i) - P(i-1))$,
$J_2 = P(i) + PERC * (P(i) - P(i-1))$, where PERC is a given parameter, e.g. 0.05-0.15 (5-15% change).

(c) Find a point J_m between J_1 and J_2 so that $ERR(J_m)$ is minimal, where

$$ERR(j) = \frac{1}{j-P(i-1)} \times \sum_{k=P(i-1)}^{j-1} (buff(k+[j-P(i-1)]) - buff(k))^2 \qquad (1)$$

(d) If $J_m = J_1$ or J_2, expand the search beyond J_1 or J_2 until a minimal ERR(j) is found.

(e) Let $P(i) = J_m$. Note that P(i) is an integer; therefore, P(i)- P(i-1) gives an estimate of the pitch period with accuracy limited by the sampling frequency.

(f) Interpolation: To find a better estimate, fit a second order polynomial to $ERR(J_m-1)$, $ERR(J_m)$ & $ERR(J_m+1)$ to find the minimal location. The distance between the minimal location and J_m is

$$\delta = -0.5 \times \frac{ERR(J_m+1) - ERR(J_m-1)}{ERR(J_m+1) - 2 \times ERR(J_m) + ERR(J_m-1)} \qquad (2)$$

(g) The (i-1)th cycle frequency is then

$$F(i-1) = \frac{sampling\ frequency}{P(i) - P(i-1) + \delta} \qquad (3)$$

(3) Adjust P(i) to the closest negative peak location. This is for the purpose of overlay ing the current cycle with the next one in a consistent way, starting at an integer sample point.

It is seen that this method of detecting the "true period" depends on an average agreement of many events that take place within the cycle. It will be seen later that the method is relatively insensitive to additive noise, but does fail for large fluctuations in F_o.

A variation of the WM method is to begin the waveform matching at zero crossings instead of peaks. The program looks for a positive zero-crossing location between the first two rough markers ($I_p(1)$ in Figure 6), and then tries to find a point between the 2nd and 3rd markers so that the mean squared error between the two cycles is minimal. This procedure is repeated until all cycles are processed. Mathematically, it is similar to the foregoing procedure, except:

(1) In step 1, find the positive zero-crossing location instead of the negative peak loca tion.

(2) In step 3, adjust to the closest positive zero-crossing location instead of the minimal peak location.

The PP Method

This method applies peak picking to the signal. A cycle-to-cycle allowable percent change is also implemented to reduce extraction failures.

Assume buff(k) to be the signal array, N_c+1 the number of rough markers, $I_N(i)$ the integer array of rough marker locations (negative zero-crossings), P_i the current integer negative peak location, $P_r(i)$ the array of real (decimal) negative peak locations, and $F_o(i)$ the fundamental frequency array. Then mathematically, PP can be explained as follows:

(1) For i=1 to N_c+1

(a) Let $J_1=I_N(i)$, $J_2=I_N(i+1)$ as illustrated in Figure 6.

(b) Find the location P_i of the negative peak between J_1 and J_2. Also find the location of the positive peak between these markers.

(c) If the locations are at J_1 or J_2, or P_i is too close to $P_r(i-1)$, shift the searching segment $[J_1,J_2]$ backward and forward alternatively, and go back to step (b).

(d) Interpolation: fit a second order polynomial to buff(P_i-1), buff(P_i) and buff(P_i+1) to find the real-valued minimal location $P_r(i)$:

$$P_r(i) = P_1 + \frac{-0.5 \times (buff(P_1+1) - buff(P_1-1))}{buff(P_1+1) - 2 \times buff(P_1) + buff(P_1-1)} \quad (4)$$

(e) For i>1, the (i-1)*th* cycle frequency is

$$F(i-1) = \frac{sampling\ frequency}{P_r(i) - P_r(i-1)} \quad (5)$$

(2) Check the array F(i) against a given allowable cycle-to-cycle percent change PERC. Assume ERR(i) is a logical array to indicate if an incorrect pitch frequency was found in (2).

(a) Find the smallest J_m between 1 and N_c-1 so that the absolute value of $F(J_m+1)$-$F(J_m)$ is less than or equal to $F(J_m)$ multiplied by PERC.

(b) If J_m is not found, all cycle frequencies are incorrect, or PERC is too small. Quit here or select a bigger PERC & go back to step (a).

(c) For i = 1 to J_m-1, ERR(i) = .true.
(d) ERR(J_m) = .false.
(e) For i = J_m+1 to N_c, ERR(i) = abs(F(i)-F(i-1)) - LE (PERC*F(i-1))
(f) If ERR(i) is true for a cycle, that cycle will not be used to calculate perturba tion parameters.

The ZC Method

The ZC method is used only on signals that have two zero crossings per cycle. This can be a low-pass filtered version of the speech signal or an EGG signal. The locations of positive zero-crossings are found between consecutive markers.

Assume again that buff(k) is the signal.

(1) For i = 1 to N_c+1

(a) Let J_1=I_N(i), J_2=I_N(i+1)

(b) Find the location P_i of a positive-going zero-crossing between J_1 and J_2. Actually, P_i is found as the midpoint crossing because the segment may have an offset. The condition:

(buff(P_i) .le. SegMid) .and.
(buff(P_i+1) .gt. SegMid)

is used, where SegMid is the midlevel of the segment (Figure 3).

(c) If P_i or the locations of extrema are at J_1 or J_2, or P_i is too close to P_r(i-1), shift the searching limits [J_1,J_2] forward and backward alternatively, and go back to step (b).

(d) Interpolation: fit a straight line to buff(P_i) and buff(P_i+1) to find the real-valued crossing location P_r(i):

$$P_r(i) = P_1 + \frac{SegMid - buff(P_1)}{buff(P_1+1) - buff(P_1)} \tag{6}$$

(e) For i>1, the (i-1)*th* cycle frequency is

$$F(i-1) = \frac{sampling\ frequency}{P_r(i) - P_r(i-1)} \tag{7}$$

(2) Check the array F(i) against a given allowable cycle-to-cycle percent change PERC. Same as step (2) of the PP method.

This concludes description of the three pitch extraction methods. The test signals will now be discussed.

Test Signals

Two types of test signals were used to check the performance of the F_o extraction methods: sine waves and mouth pressure analogs. Amplitude modulation, frequency modulation, and additive white noise were imposed on these signals.

Sine waves

Sine waves were a special case of generic waveshapes that included skewing and truncation (to approximate glottal area and vocal fold contact area for other applications). The basic waveform as a function of phase for each period is:

$$S_b(\alpha) = \begin{cases} 0.5\,[\,1-\cos(\frac{1+r}{2}\alpha)\,]\;, & 0 \le \alpha < \frac{2\pi}{1+r} \\ 0.5\,[\,1-\cos(\frac{1+r}{2r}\alpha + \frac{r-1}{r}\pi)\,], & \frac{2\pi}{1+r} \le \alpha \le 2\pi \end{cases} \tag{8}$$

where r is a parameter to control the amount of skewing to the left ($r \geq 1$). When $r = 1$, $S_b(\alpha)$ is a pure sine wave with a baseline offset.

PO analogs

The basic glottal flow waveform as a function of phase α is chosen to be

$$S_b(\alpha) = a \cdot \alpha^b \cdot \exp(-c \cdot \alpha) \quad 0 \le \alpha \le 2\pi \;, \tag{9}$$

where the parameters a, b, and c are defined as

$$b = \log_e(5000) \,/\, [\,1 - \log_e(1+r)\,] \tag{10}$$

$$c = b \cdot (1+r) \,/\, (2\pi) \tag{11}$$

$$a = (c \cdot \exp(1) \,/\, b)^b \tag{12}$$

The parameter r is again used to control the amount of waveform skewing ($r \geq 1$), which was set to 3 for these PO analogs. The phase α is replaced by 2π-$\alpha(t)$ so that the skewing is to the right.

With these parameters, the following extreme values are realized:

$$\alpha_{max} = \frac{b}{c} = \frac{2\pi}{1+r} \qquad \textit{phase at max flow} \tag{13}$$

$$S_b(\alpha_{max}) = 1 \qquad \textit{maximum flow} \tag{14}$$

$$S_b(2\pi) = \frac{1}{5000} \qquad \textit{fully decayed flow} \tag{15}$$

An instantaneous amplitude A(t) and phase $\alpha(t)$ are determined by the modulations (see below). With these modulations, the glottal flow signal is

$$U_g(t) = A(t) \cdot s_b(\alpha)\,|_{\alpha=2\pi-\alpha(t)} \tag{16}$$

Following the model by Rabiner and Schafer (1978), the mouth pressure PO signal can be obtained by combining the glottal flow with the vocal tract filter V(z) and the radiation filter R(z).

The vocal tract filter V(z) is modeled by

$$V(z) = \prod_{k=1}^{M} V_k(z) = \prod_{k=1}^{M} \frac{1 - 2r_k\cos(\theta_k) + r_k^2}{1 - 2r_k\cos(\theta_k)\,z^{-1} + r_k^2 z^{-2}} \tag{17}$$

where each pair of r_k and θ_k determine a formant structure with given frequency and bandwidth. For this study, we used five formants. Their frequencies and bandwidths are given in Table 3.

Table 3.
Formant frequencies and bandwidths for synthetic signals.

Formant	Frequency	Bandwidth
1st	650.3	94.1
2nd	1075.7	91.4
3rd	2463.1	107.4
4th	3558.3	198.7
5th	4631.3	89.8

The radiation filter R(z) is modeled by

$$R(z) = R_o(1 - z^{-1}) \tag{18}$$

where R_o is a constant. The PO analog signal is then defined as

$$S_o(t) = U_g(t) \otimes h(t) \tag{19}$$

where h(t) is the impulse response of the filter (H(z) = V(z) R(z), and $\otimes$ indicates convolution). Figure 7(b) shows the PO analog signals for the nominal parameters chosen in this study (no modulations).

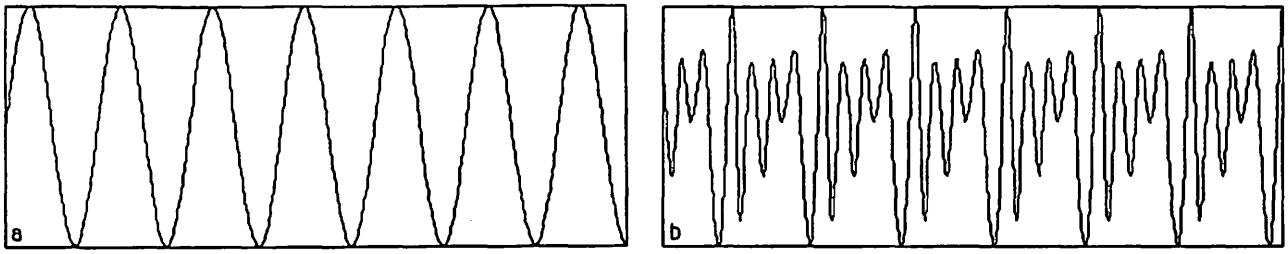


Figure 7. The test signals used. (a) Sine wave (b) PO (oral pressure) analog wave.

Modulations

Four types of modulations are identified: (a) none; (b) random; (c) sinusoidal; (d) triangular. The modulation function is defined as:

$$modfn(type, f_m, t) = \begin{Bmatrix} 0 & , type=none \\ n_1(t) & , type=random \\ \cos(2\pi f_m t) & , type=sinusoidal \\ 1-4\,|\mathrm{mod}(f_m t, 1)-0.5|, & type=triangular \end{Bmatrix} \quad (20)$$

where f_m is the modulating frequency and $n_1(t)$ is noise from a random number generator.

For frequency modulation, the instantaneous frequency and phase are

$$f(t) = f_0\,[1 + k_f \cdot modfn(FMType, f_{fm}, t)\,] \quad (21)$$

$$\alpha(t) = 2\pi \cdot \mathrm{mod}\left[\int_0^t f(\tau)\,d\tau\,,\ 1.0\right]\ , \quad (22)$$

where f_0 is the center frequency, f_{fm} is the modulating frequency for FM, and k_f is a modulation index. For AM, the instantaneous amplitude is

$$A(t) = A_0\,[1 + k_a \cdot modfn(AMType, f_{am}, t)\,]\ , \quad (23)$$

where A_o is the nominal amplitude, F_{am} is the AM frequency, and k_a is the AM index.

When noise is added, the signal assumes the form

$$S(t) = S_0(t) + k_n \cdot n(t)\ , \quad (24)$$

where

$$S_0(t) = A(t) \cdot S_b(\alpha)\,|_{\alpha=\alpha(t)} \tag{25}$$

is the signal without additive noise, n(t) is the random noise term, and k_n is a parameter to control the amount of noise. The signal to noise ratio is

$$SNR = 10\log_{10}\frac{Average\ Signal\ Power}{Average\ Noise\ Power} \tag{26}$$

where the powers are determined by root mean squared (RMS) calculations on the signals.

Analysis Results

The WM, PP and ZC algorithms described earlier were programmed in FORTRAN and installed in a software package called GLIMPES (Glottal Imaging by Processing External Signals). To compare performance between the extraction methods, the mean rectified cycle-to-cycle difference in F_0 was calculated as a jitter measure over a window of N cycles. This measure is often used as a definition of the mean first order perturbation (Pinto & Titze, 1990) and is defined here as

$$PF1 = \frac{1}{N-1}\sum_{i=1}^{N-1}\frac{|f(i+1) - f(i)|}{0.5 * [f(i+1) + f(i)]} \times 100\ , \tag{27}$$

where f(i) is the fundamental frequency extracted for i*th* cycle. Note the PF1 is normalized to a short-term mean F_0 to express PF1 in percent. The short-term (2 cycle) average is used in anticipation of applications where F_0 has unidirectional trends.

Since all the test signals were simulated by computer, the theoretical cycle-to-cycle F_0 values and the theoretical PF1 values were known. Specifically, the theoretical values were found as follows:

(1) When there is no frequency modulation, the theoretical F_0 stays constant for all cycles at the specified value.

(2) When frequency modulation is present, a new cycle is started every time the instanta neous phase changes by 2π. To match the extracted frequencies to the theoretical ones, GLIMPES outputs the starting time in its extraction, i.e. the starting sample for the first extracted cycle. This starting time is then refined through interpolation on the theoretical instantaneous phase curve until the first few extracted pitch frequen cies match well with their corresponding theoretical values. Once the starting time is synchronized, all other theoretical frequencies are identified.

The first order perturbation factor PF1 will now be reported for several parameters and imposed perturbations. Six comparisons will be made for each set of parameters:

(1) WF method applied to PO analog (symbol ◇)
(2) WF method applied to sine wave (symbol ■)
(3) PP method applied to PO analog (symbol ▲)
(4) ZC method applied to low-pass filtered PO analog (symbol △)
(5) ZC method applied to sine wave (symbol ◆)
(6) ZC method applied to low-pass filtered sine wave (symbol □)

Fundamental Frequency

The first test was to find out how the mean value of the F_o affects PF1. To do this, we used the two test signals with no frequency modulation, no amplitude modulation, high signal-to-noise ratio SNR (80 dB), and fixed sampling frequency (20 kHz). An FIR lowpass filter of order 201 was used for cycle marking as described earlier. F_o was varied in 50 Hz steps from 50 Hz to 500 Hz. The results are shown in Figure 8 and can be summarized as follows:

(1) Changing the fundamental frequency does not have a profound overall effect on the performances of the four methods. However, great fluctuations exist in the curves because integer relations between sampling frequency and F_o give much lower val ues. When sampling frequency is a multiple of F_o and there are no modulations, the waveforms are repeated exactly from cycle to cycle in even numbers of samples, so the extracted F_o stays the same. In other words, no interpolation between samples is necessary. For 20 kHz sampling, the non-integer frequencies are 150 Hz, 300 Hz, 350 Hz, and 450 Hz. We regard these as the better test cases.
(2) The jitter values for any WM test case were in the 0.0001% to 0.001% range (the theoretical value is zero, of course).
(3) The PP and ZC methods gave poorer results, with most of the values in the 0.01% to 0.001% range, and some above 0.01%.

The above results apply only to cases where no modulations are imposed. Further generalization awaits data to be presented later.

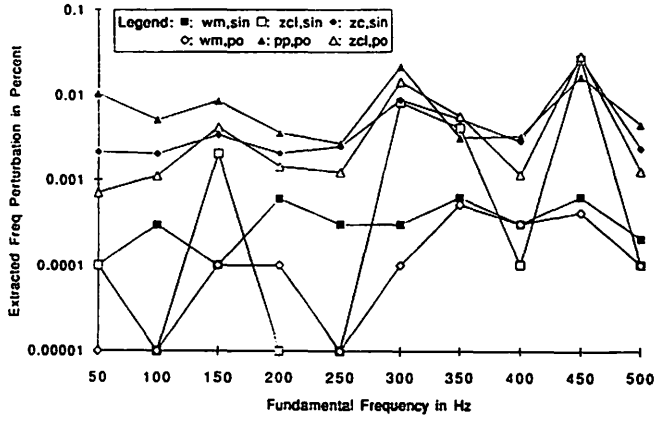


Figure 8. Extracted F_o perturbation in percent vs. fundamental frequency F_o in Hz, with fs=20 kHz, SNR=80 dB, and no modulation.

Sampling Frequency

A test was used to check the effect of sampling frequency f_s on PF1. Results are shown in Figure 9. Without modulations and with F_o=150 Hz, SNR = 80 dB, and f_s changing from 20 kHz to 50 kHz, it was found that f_s has little affect on jitter because interpolation was used with all methods. The major effect was again the occasional dip in jitter when the sampling frequency was an integer ratio of F_o (e.g. 30 kHz and 45 kHz).

Signal-To-Noise Ratio

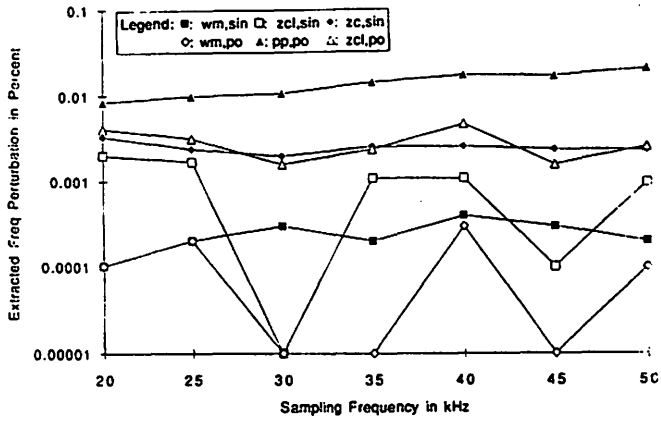
The next test was to check the effect of additive white noise on software jitter. Without modulations and with F_o fixed at 150 Hz, f_s at 20 kHz, the signal-to-noise ratio (SNR) was varied from 10 dB to 100 dB (Figure 10). In general, the SNR can be as low as 50 dB to get software jitter below 0.01% for the WM method. The PP method gives results an order of magnitude higher than the WM and ZC values, in general. It is seen that a linear relation exists between the log of PF1 and the decrease in SNR for all methods up to about 60 dB. Thereafter, the WM method continues to show a decrease, but the PP and ZC methods have reached a limit in accuracy.

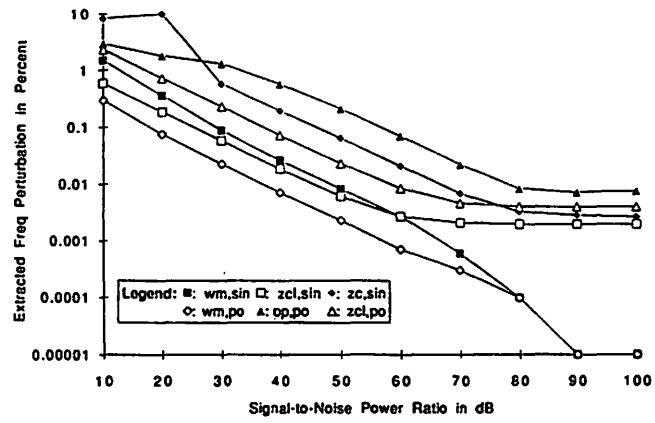
Amplitude Modulation

A test was performed to evaluate the effect of amplitude modulation on jitter. With no frequency modulation, F_o at 150 Hz, f_s at 20 kHz, and SNR at 80 dB, the modulation index k_a was fixed at 40% and the modulation frequency f_{am} was chosen according to the relation

$$f_{am} = F_o k_{apc} / (4k_a) \quad . \tag{28}$$

With this relation, the cycle-to-cycle amplitude change was regulated by k_{apc}. The value of k_{apc} was changed from 0.01% to 10% (Figure 11). Note that the 0.01% error criterion was met for all methods up to about 5% amplitude modulation per cycle. Thereafter, results deteriorate rapidly. Note also that the WM methods is much less affected by amplitude modulation that either the ZC or PP method.





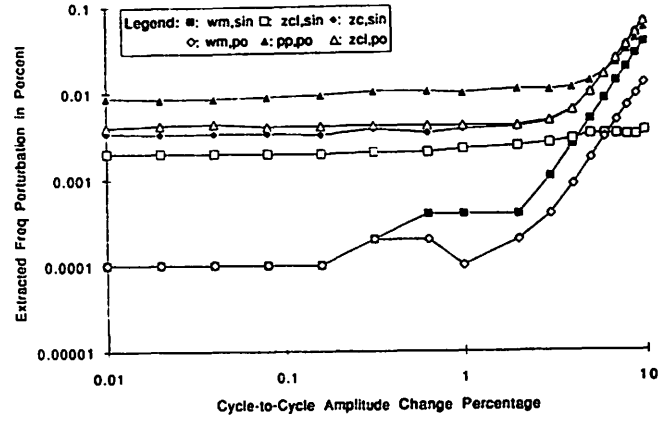


Figure 9 (upper left). Extracted F_o perturbation in percent vs. sampling frequency f_s in kHz, with F_o=150 Hz, SNR=80 dB, and no modulation. Figure 10 (upper right). Extracted F_o perturbation in percent vs. signal-to-noise ratio in dB, with F_o=150 Hz, f_s=20 kHz, and no modulation. Figure 11 (left). Extracted F_o perturbation in percent vs. amplitude modulation in percent, with F_o=150 Hz, f_s=20 kHz, SNR=80 dB, and no frequency modulation.

Frequency Modulation

The final series of experiments were designed to test imposed frequency variation against measured variation. Here the choice of extraction method became less obvious. With 1% amplitude modulation (k_a = 40%, f_{am} = 0.9375 Hz), F_o at 150 Hz, f_s at 20 kHz, and SNR at 40 dB, frequency modulation was imposed as follows. The modulation index k_f was set to 40%. The modulation frequency f_{fm} was then chosen according to

$$f_{fm} = (F_o k_{fpc}) / (4 k_f) \tag{29}$$

so that the cycle-to-cycle frequency change was regulated by k_{fpc}. The value of k_{fpc} was changed from 0.01% to 10%.

For the comparison between extracted jitter values and theoretical values, two scatter plots were made. The first plot (Figure 12, see below) shows a wide range of frequency perturbation (0-8% per cycle), whereas the second plot (Figure 13, below) shows a smaller "zoomed in" range. At a first glance, Figure 12 suggests that all the methods perform equally well, with the exception of PP on a PO analog waveform. Throughout the data presented in this paper, PP always showed the highest values and the greatest variability. Closer examination at the endpoints of the scatter plot shows, however, that there are differences between the extractions, and that these differences are waveform-dependent.

In Figure 13, the low perturbation range is explored. Best results are obtained by WM on PO analogs, followed by ZC on sinusoids, WM on sinusoids (data points are masked), ZC on filtered PO analogs, ZC on sinusoids without filtering, and finally, PP on PO analogs. For the single event methods, large errors occur for *small* perturbations. This is not the case with the waveform matching methods, which tend to get better with smaller perturbations. The reason is that small local fluctuations in the signal are de-emphasized by waveform matching, but play an important role in single event methods.

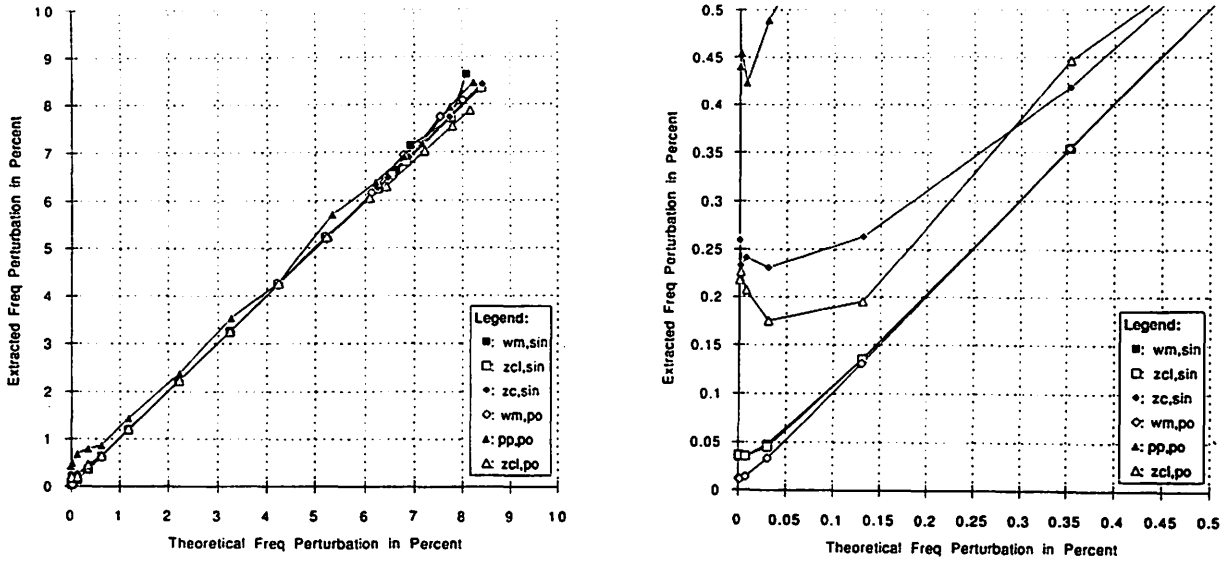


Figure 12 (left). Extracted F_o perturbation in percent vs. theoretical F_o perturbation in percent for F_o=150 Hz, f_s=20 kHz, SNR=40 dB, and 1% amplitude modulation. Figure 13 (right). Same as Figure 9, but with an expanded range of perturbation near the origin.

Summary and Conclusions

In summary, the waveform matching method meets the high precision criterion of being able to extract a 1% frequency change (per cycle) with a 1% accuracy, as long as the signal to noise ratio is greater than about 50 dB and concomitant amplitude modulations are below about 5%. This holds for sampling frequencies between 10 kHz and 50 kHz when interpolation is used. Moreover, the WM method maintains reasonable accuracy (about 10%) as perturbations get down into the 0.05 to 0.1% range. The method is relatively resistant to additive noise and small amounts of amplitude fluctuation. A primary drawback is loss of accuracy for frequency modulations above about 6%. Here, a negative return is realized by capitalizing on waveform matching because formant structure does not stretch or shrink with F_0. As the timebase expands and shrinks from cycle to cycle, adjacent cycles can no longer be matched adequately.

Peak picking and zero crossing methods do not meet the high precision criterion consistently, especially not when frequency perturbations are in the normal 0.1% to 1.0% range. Great care must be taken in the interpretation of jitter and shimmer with these single event detectors because they are not noise resistent. Thus, although they have the potential for detecting specific micro-events within the glottal cycle, one is never sure that these events are not contaminated by recording noise or by limitations on extraction algorithms, or both. Filtering of the speech waveform can reduce some of the variability in the measures, but this may eliminate the very events under investigation. For example, we have found that closely-spaced multiple peaks or noisy peaks can be smoothed out by some low pass filtering, but the results are sensitive to the type of filter (linear phase versus constant amplitude) and its parameters (cut off frequency, filter order, etc.). Unless one has some *a priori* knowledge of the pattern to be detected, the complicated interactions between source, vocal tract filter, and system noise are not separable. This was the main point of Hillenbrand's (1987) study.

In conclusion, then, until more is known about the perturbation patterns to be detected in natural voice, it makes sense to use a method that gives the best results for artificially produced patterns (modulations). For these, waveform matching is the clear choice when frequency variations are below about 6% per cycle. For higher variations, no statement about accuracy can be made at this point.

Acknowledgement

This work was supported by grant No. DC00387-05 from the National Institutes on Deafness and Other Communication Disorders for which we are grateful. The authors also express appreciation to Pamela Rios and Julie Lemke for manuscript preparation.

References

Baken, R.J. (1989). Irregularity of vocal period and amplitude: a first approach to the fractal analysis of voice.

Cox, N. (1989). Technical considerations in computations of spectral harmonics-to-noise ratio for sustained vowels. *Journal of Speech and Hearing Research,* 32(1), 203-218.

Deem, J.F., Manning, W.H., Knack, J.V., & Matesich, J.S. (1989). The automatic extraction of pitch perturbation using microcomputers: some methodological considerations. *Journal of Speech and Hearing Research,* 32, 689-697.

Doherty, E.T. & Shipp, T. (1988, September). Tape recorder effects on jitter and shimmer extraction. *Journal of Speech and Hearing Research,* 31, 485-490.

Hess, W. (1983). *Pitch determination of speech signals, algorithms and devices. Springer-Verlag Berlin Heidelberg.*

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech and Hearing Research,* 30, 448-461.

Horii, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. *Journal of Speech and Hearing Research,* 18, 192-201.

Milenkovic, P. (1987). Least mean square measures of voice perturbation. *Journal of Speech and Hearing Research,* 30(4), 529-538.

Rabiner, L.R. & Schafter, R.W. (1978). *Digital Processing of Speech Signals.* Prentice-hall Inc.

Titze, I.R., Horii, Y., & Scherer, R. (1987, June). Some technical considerations in voice perturbation measurements. *Journal of Speech and Hearing Research,* 30, 252-260.

Titze, I.R. & Winholtz, W. (in review). The effect of microphone type and placement on voice perturbation measurements. *Journal of Speech and Hearing Research.*

NCVS Status and Progress Report - 3
December 1992, 117-134

The Effect of Microphone Type and Placement on Voice Perturbation Measurements

Ingo R. Titze, Ph.D.
Recording and Research Center, The Denver Center for the Performing Arts
and
Department of Speech Pathology and Audiology, The University of Iowa
William S. Winholtz, A.A.S.
Recording and Research Center, The Denver Center for the Performing Arts

Abstract

This study was conducted to explore the effects of microphone type (dynamic versus condenser) and pattern (omnidirectional versus cardioid) on the extraction of voice perturbation measures for sustained phonation. Also of interest were the effects of distance and angle between the source and the microphone. Four professional grade and two consumer grade microphones were selected for analysis. Synthesized phonation with different amplitude and frequency modulations at fundamental frequencies of 100 Hz and 300 Hz were presented over a loudspeaker. Human phonation was also included to test the validity of loudspeaker presentations. Three microphone distances (4 cm, 30 cm, 1 m) and three angles (0°, 45°, 90°) were used for microphone placement. Among the professional grade microphones, the cardioid condenser type had the smallest effect on perturbation measures. In general, condenser types gave better results than dynamic types. Microphones with an unbalanced output did not perform as well as those with balanced outputs. Microphone sensitivity and distance had the largest effect on perturbation measures, making it difficult to resolve normal vocal jitter at anything but a few centimeters from the mouth. Angle had little effect for short distances, but a greater effect for longer distances. These conclusions are preliminary because the sampling of microphones, distances, and signal types was very coarse. The study serves only to chart the course for future work.

Introduction

Voice perturbation analysis continues to be of interest in studies of vocal fold vibration and assessment of voice disorders. It has the potential for quantifying small aperiodicities in a speech signal, which is a first step toward understanding the mechanisms by which vocal fold vibration may be disturbed.

Perturbation analysis has gone through a refining process in recent years. It has been suspected that some of the lack of consistency of perturbation measures found across laboratories and across similar subject groups (Karnell, Scherer, & Fischer, 1991) may be related to technical flaws in recording and processing of signals. Doherty and Shipp (1988) showed that some analog tape recorders can greatly inflate vocal jitter (cycle-to-cycle irregularity in fundamental frequency) and vocal shimmer (cycle-to-cycle irregularity in amplitude). Digital audio recording is now commonly used, which essentially eliminates the contamination of perturbation measures by wow and flutter of the tape drive.

Other problems in perturbation analysis relate to the appropriate length of the analysis window (Karnell, 1991), the method of extraction used (Milenkovic, 1987; Titze & Liang, in review), and the interactions between amplitude and frequency measures (Hillenbrand, 1987).

As in any high fidelity audio recording, attention also needs to be paid to the microphone, the conditioning amplifiers and filters, and the recording environment. This paper deals with two problems, microphone type and microphone placement relative to the source. Questions of interest are: (1) is the extraction of F_0 and amplitude perturbation affected more by a condenser microphone than by a dynamic microphone, (2) is directionality of the microphone an important factor, (3) how are perturbation measures affected by mouth-to-microphone distance, and (4) is off-axis placement an important factor?

Interactions between room environment, recording hardware, and extraction software can be extremely complex. For example, the interaction matrix [signal type x perturbation type x perturbation extent x microphone type x distance x angle x F_0 x room noise] can easily have several thousand items, even if only a few cases are selected for each variable. Therefore, to limit this first study in its scope, only a few typical angles and distances were used in combination with a few typical microphones. As a result, none of the questions posed above are answered unequivocally. It was felt, however, that a broad paintbrush approach was needed to determine where future efforts should be placed.

Method

Recording Environment

The experiment was conducted in an IAC isolation booth, 10.5 ft deep by 11.5 ft wide by 8 ft high. Both human speakers and a loudspeaker were used to generate a free-field acoustic signal. The locations of the microphones relative to the source are shown in Figure 1. There were three angles (0°, 45°, and 90°) for each of three distances (4 cm, 30 cm and 1 m). A distance of 1 m was chosen to encompass a wide range of research applications where the microphone may need to be far from the source. 30 cm was chosen as a mid-range distance that is commonly used in acoustic recordings, and 4 cm was chosen for possible head-mounted microphone applications.

Ambient sound pressure level (SPL) of the booth, measured with a B&K 2230 SPL meter on the linear weighing scale (20 Hz to 20 kHz), was 53 dB near the center of the booth.

To test the resonance effects of the booth, a frequency sweep of 20% around the fundamental frequency (F_o) of 300 Hz was conducted at the three distances from the source (0° angle and an SPL of 80 dB). The frequency sweep corresponded to the maximum frequency modulation imposed in later experiments. Measurements indicated that for these modulations the intensity varied as follows: 0.6 dB at 4 cm, 2.8 dB at 30 cm, and 4.4 dB at 1 m. A small portion of this amplitude variation can perhaps be attributed to nonuniform loudspeaker response, but the sizeable variation with distance suggests that amplitude perturbations result from source frequencies being selectively reinforced by room resonances. A similar effect was demonstrated with regard to the resonances of the vocal tract (Hillenbrand, 1987; Horii & Hata, 1988). Room resonances will vary with the size and type of room. As indicated by the measurements reported later, positioning the microphone nearer to the source will minimize the variations. A separate study of room effects on perturbation measures is needed to go beyond this simple rule of thumb.

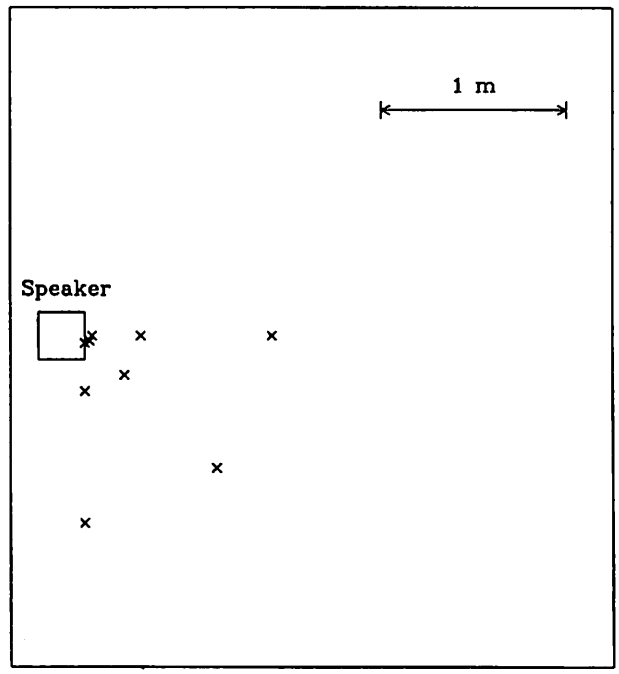


Figure 1.Diagram of loudspeaker and microphone placements in an IAC booth; drawn to scale (top view).

Microphone Type

Four professional grade microphones were chosen. They all had electrically balanced outputs and a flat frequency response (<±1.4 dB) over the ranges of frequencies used in this experiment. The microphones are identified as:

Microphone	1:	AKG 451EB CK22 condenser omnidirectional
	2:	AKG 451EB CK1 condenser cardioid
	3:	EV DO54 dynamic omnidirectional
	4:	AKG D224E dynamic cardioid

In addition, two consumer grade microphones with electrically unbalanced outputs were chosen for limited analysis and comparison to the professional grade microphones:

Microphone	5:	Realistic 33-985 dynamic omnidirectional
	6:	Realistic 33-1063 tie clip miniature condenser omnidirectional

Sensitivity of the microphones was determined by placing them 4 cm from the loudspeaker at an 0° angle, which produced a SPL of 80 dB at 200 Hz. The output of the preamplifier minus the constant gain (60 dB) was recorded. Results are shown in Table 1.

Table 1.
Sensitivity of microphones used in this study

Microphone	Type	Sensitivity	Grade
1	condenser omnidirectional	-49.56	professional
2	condenser cardioid	-48.20	professional
3	dynamic omnidirectional	-68.69	professional
4	dynamic cardioid	-70.20	professional
5	dynamic omnidirectional	-75.24	consumer
6	condenser omnidirectional	-76.03	consumer

The two professional-grade condenser microphones (1 and 2) were clearly the most sensitive, but the consumer grade condenser microphone (6) was the least sensitive. It will be seen later that microphone 2 gave consistently the best results.

Stimulus Signals

A computer synthesis program (Titze & Liang, in review) was used to generate waveforms having different types of modulation (Figure 2). On the top left of the diagram are controls for frequency modulation (FM), including the modulation index, modulation frequency, and carrier frequency (F_o). On the top right are similar inputs for amplitude modulation (AM). In the middle blocks, the instantaneous phase and the magnitude envelope of a glottal pulse are computed and noise is added. To simulate jitter or shimmer, an $F_o/2$ (subharmonic) modulation was used to vary adjacent pulse period and amplitude.

To make the waveform somewhat speech-like, a single formant was simulated by convolving a glottal pulse with a filter response (see bottom of Figure 2). The formant frequency was set to 470 Hz, a number that has no integer relation to any F_o used. This avoided any strong boost that a single harmonic might get from the formant filter. Additional formants could have been added to make the signal more speech-like, but we felt that this was a reasonable compromise between using a spectrally rich waveform and a spectrally poor waveform (which is seen in some falsetto and female productions).

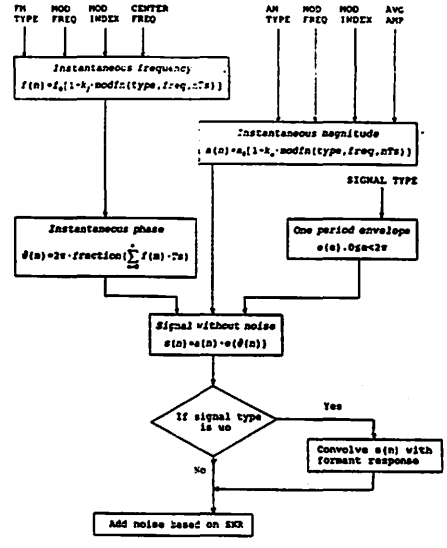


Figure 2. Flow chart of computer program to synthesize modulated speech signals.

Two groups of signals with identical modulations were generated. One group had an F_o of 100 Hz and the other an F_o of 300 Hz. Each signal was six seconds in length. Figure 3(a) shows the 100 Hz signal and Figure 3(b) shows the 300 Hz signal, both without imposed modulations.

In each F_o group there were four types of modulation: (1) AM (amplitude modulation) at 10 Hz, (2) AM at $F_o/2$, the first subharmonic, (3) FM (frequency modulation) at 10 Hz, (4) FM at $F_o/2$, the first subharmonic. All modulations were sinusoidal. The 10 Hz AM was synthesized at 50% and 5% modulation index to correspond to ranges observed for vocal tremor (Winholtz & Ramig, 1992) and normal voice, respectively. The 10 Hz FM had 20% and 2% modulation indices, realizing that FM is generally smaller than AM in speech. All 10 Hz modulations maintained at least a 10:1 ratio in carrier frequency to modulation frequency.

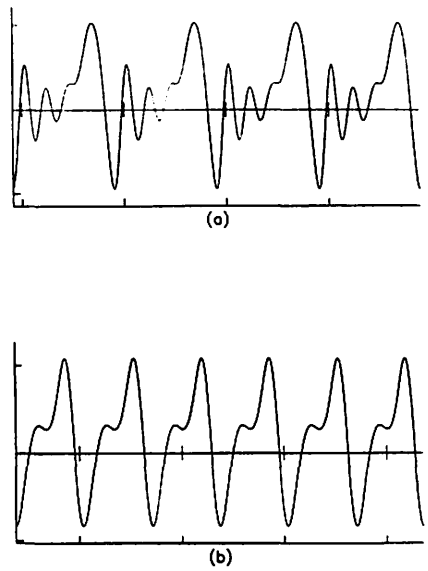


Figure 3. Stimulus signal with no imposed modulation. (a) F_o at 100 Hz and (b) at F_o at 300 Hz. The formant frequency is 470 Hz.

For the subharmonic modulations, however, the ratio was 2:1. This short-term AM and FM modulation had indices of 5% and 0.5% to simulate shimmer and jitter (Scherer, Gould, Titze, Meyers, & Sataloff, 1988). Random period fluctuations could have been used, but in our experience these are a bit more difficult to control synthetically. There are problems with discontinuities within and across cycles and there are further problems with knowing what the precise theoretical value is. The $F_o/2$ modulations give a regular stream of reversals in the F_o or amplitude contour (every cycle), but they obviously do not capture every feature of jitter and shimmer.

Included in each F_o group was one signal with no imposed modulation. This signal was used to determine the noise of the system and the errors associated with perturbation extraction. No attempt was made to synthesize "normal" perturbations because it remains questionable what such values are at this stage.

For convenience in later presentation, the signals were transferred from the computer through a DSC-200 16 bit D/A converter (at 20 kHz sampling frequency) to a Panasonic SV-3700 DAT recorder. (Direct digital transfer was not possible because of present format incompatibility).

Hardware for Stimulus Presentation

The DAT recorder was used as the signal source for a Technics SU-V303 power amplifier and a 4 inch loudspeaker (Auratone 5C). The sound pressure level (SPL) was adjusted to 80 dB at a distance of 4 cm from the loudspeaker with the 100 Hz signal (no modulation). It was not re-adjusted for the remainder of the experiment. The loudspeaker was chosen over human phonation (for most of the experiment) because it offered a wide range of control and was more consistent in presentation of the acoustic stimulus. For example, when a synthesized phonation signal with no

perturbation was presented over the loudspeaker, it produced baseline perturbation measures an order of magnitude lower than average normal human phonation. This allowed for better assessment of the influence of system noise and for better verification of the imposed signal modulations. However, to test the differences between a human speaker and a loudspeaker as a source of perturbation, some comparable measurements on humans were made. These will be described later.

Hardware for Recording

The microphone signals were preamplified (ATI M-1000), high pass filtered at 60 Hz (24 dB/oct flat phase) and sampled by a DSC-200 16 bit A/D converter at 20 kHz. To utilize the full range of the A/D converter, the gain of the conditioning amplifier (DSC-240) was adjusted for "0" VU with the unmodulated signal each time a microphone was repositioned. After the initial setting, the gain was unchanged for the remainder of the experiment for that position. A three second segment of the total six second stimulus was digitized in each case, avoiding start-up and release transients.

Analysis

Software

A software package called GLIMPES (Glottal Imaging by Processing External Signals), was used to analyze a random two second segment of the digitized signals, again avoiding the earliest and latest cycles. Since the stimulus used for presentation was highly repetitive, only the length of the token was critical for comparison in analysis. The F_o extraction algorithms are of two types (1) single event detection (peaks or zero crossings) with interpolation, and (2) waveform matching between adjacent cycles with interpolation. The accuracy of extraction of these measures is being reported (Titze & Liang, in review). Waveform matching, which determines the period on the basis of the least square error between adjacent-cycle waveforms, gives the best results for F_o extraction in the presence of additive noise and amplitude modulation. Single event detection, and in particular zero crossing detection on the low-pass filtered signal, gives better results only when F_o is modulated more than a few percent. We used the waveform matching technique in this experiment. The technique is described in detail by Milenkovic (1987) and by Titze and Liang (in review).

Amplitude and frequency perturbation measures were obtained. The measures included CV (coefficient of variation), defined as the zeroth-order RMS (Root Mean Squared) perturbation according to the Pinto and Titze (1990) nomenclature, and P1 (perturbation one), defined as the first-order MR (Mean Rectified) perturbation:

$$CV = \frac{100}{N\bar{x}} \left[\sum_{i=1}^{N} (x_i - \bar{x})^2\right]^{\frac{1}{2}} \quad (1)$$

$$P1 = \frac{100}{(N-1)\bar{x}} \sum_{i=2}^{N} |x_i - x_{i-1}| \quad (2)$$

In the above equations, the variable x can be either frequency or amplitude, i is the period index, and $\bar{x}$ is a mean value for normalization to percent.

Results

Direct analysis of the synthesized computer files (without playback, recording, and re-digitization) provided the reference for comparison with the corresponding recorded microphone signals. Table 2 illustrates the results for this direct analysis. The top half of the table is for amplitude measures and the bottom half for frequency measures. For the signal with no imposed modulation (first line in each half), amplitude perturbation ranged between 0.00 and 0.12% over both values of F_o, and frequency perturbation was undetectable to 0.01% accuracy. This is essentially the software (extraction) jitter and shimmer, which is low enough not to inflate the imposed modulations significantly.

Table 2.
Direct analysis of the synthesized signals used for stimulus. Modulation frequencies were 10 Hz, except for jitter and shimmer, which was an $F_o/2$ subharmonic modulation.

Imposed Modulation		Amplitude Measures (%)			
		F_o = 100 Hz		F_o = 300 Hz	
		CV	P1	CV	P1
None		0.00	0.00	0.08	0.12
Amplitude	50%	31.45	17.23	34.78	6.55
	5%	3.39	1.84	3.53	0.66
Frequency	20%	16.12	8.30	0.96	0.40
	2%	2.00	1.07	0.27	0.70
Shimmer	5%	1.74	3.47	2.43	4.83
	0.5%	0.18	0.35	0.26	0.49
Jitter	5%	0.52	1.01	0.85	1.67
	0.5%	0.07	0.13	0.10	0.17

		Frequency Measures (%)			
		F_o = 100 Hz		F_o = 300 Hz	
		CV	P1	CV	P1
None		0.00	0.00	0.00	0.00
Amplitude	50%	0.09	0.08	0.02	0.01
	5%	0.00	0.00	0.00	0.00
Frequency	20%	12.05	11.06	13.70	2.59
	2%	1.34	0.75	1.41	0.26
Shimmer	5%	0.05	0.11	0.02	0.04
	0.5%	0.01	0.01	0.00	0.00
Jitter	5%	1.51	3.06	1.80	3.60
	0.5%	0.15	0.30	0.18	0.36

There is no direct relation between modulation index and the CV and P1 measures. Modulation index is defined as

$$Mod = \frac{x_{max} - x_{min}}{x_{max} + x_{min}} \quad , \tag{3}$$

where x_{max} and x_{min} are maximum and minimum vlues of either frequency or amplitude. This index is used for convenience in synthesis. It is easier, mathematically, to impose a goven modulation than a given CV or P1. As long as both are calculated and measured, ther is no loss of information. Note that for 50% AM (at the 10 Hz modulation frequency), the CV for amplitude is about Table 2). Variations in P1 results from the fact that adjacent cycle differences are minimized when the ratio of modulation frequency to F_o decreases. Note that both CV and P1 scale downward appropriately by a factor of 10 when modulation is reduced by a factor of 10.

Cross-modulation between AM and FM is always a problem in speech signals (Hillenbrand, 1987). In the fourth line of the top half of the table, we see that an imposed 20% FM results in large amplitude perturbations, even though no AM was imposed. These cross-modulations result from different harmonics moving through formants at different fundamental frequencies. At 100 Hz, for example, the fifth harmonic is within 30 Hz of the 470 Hz formant. A 20% frequency variation (up and down) will sweep this harmonic through the formant, causing large amplitude fluctuations. At F_o = 300 Hz, on the other hand, a 20% frequency variation will sweep neither the fundamental nor the second harmonic through the formant frequency. The amplitude perturbations are correspondingly smaller at 300 Hz in Table 2.

The cross-modulation is less severe when AM is imposed. Note that in the second line of the bottom half of the table, a 50% AM causes values of CV and P1 that are all less than 0.1%. There is no frequency-amplitude interaction that corresponds to the amplitude-frequency interaction of vocal tract formants.

Having assessed the baseline accuracy of the extraction software and the possible hazards due to cross-modulation, the next step was to assess the interactions between hardware and software. In particular, the effect of the six microphones and their placement were of interest.

Microphone Type

Table 3 shows the perturbation measures when the entire electroacoustic link (DAC, power amplifier, loudspeaker, microphone, preamplifier filter, DAT recorder, and ADC) were included. No modulations were imposed. Hence, these are the baseline measures for the hardware-software combination. The effect of the electroacoustic link, and particularly each microphone, can be assessed by comparing these numbers to the two rows labeled "None" in Table 2. With professional grade microphones (1-4), the overall amplitude perturbations were on the order of 0.1%-0.3% and the frequency perturbations were an order of magnitude lower. Perturbations with the dynamic microphones (3 and 4) were generally higher than with condenser types (1 and 2). With consumer grade microphones (5 and 6), baseline perturbations were about three times higher, on average, than with the professional grades. This may be caused by their electrically unbalanced outputs, which allows increased electrical noise to contaminate the signal by way of electromagnetic interference

and ground loops producing 60 Hz modulations. Much of this modulation was passed through the 60 Hz HP filter because the attenuation was only 3 dB at 60 Hz. To reduce the effects of ground loops with HP filtering, the cutoff frequency would have to be higher than 60 Hz, but this would then interfere significantly with the $F_o/2$ subharmonic at 100 Hz. Since the purpose of the HP filter was primarily to reduce low frequency acoustic noise in the IAC booth, a compromise of 60 Hz was struck.

Table 3.
Extracted perturbations from microphone signals for stimuli with no imposed modulation, 0° angle and 4 cm distance from source.

Microphone	Amplitude Measures (%)			
	F_o = 100 Hz		F_o = 300 Hz	
	CV	P1	CV	P1
1	0.19	0.22	0.15	0.20
2	0.10	0.11	0.13	0.17
3	0.34	0.34	0.25	0.34
4	0.34	0.34	0.27	0.33
5	1.53	1.67	1.28	1.69
6	0.98	1.00	0.73	0.93

	Frequency Measures (%)			
	F_o = 100 Hz		F_o = 300 Hz	
	CV	P1	CV	P1
1	0.0	0.02	0.04	0.05
2	0.01	0.01	0.02	0.03
3	0.02	0.02	0.05	0.09
4	0.02	0.03	0.03	0.05
5	0.10	0.14	0.18	0.27
6	0.05	0.07	0.10	0.14

Table 4 (see following page) presents another comparison of extracted perturbations with and without the electroacoustic link, but this time some modulations are imposed. To shorten the table, CV values are given for the 10 Hz modulations and P1 values for the $F_o/2$ modulations. These are deemed the "most appropriate" measures for the type of modulation. For the 10 Hz amplitude and frequency modulations, the values extracted with professional grade microphones were all within $\pm$2% of the values obtained from direct analysis. With the consumer grade, microphone 5 (the dynamic omnidirectional) had the effect of inflating the 10 Hz amplitude modulations slightly (first two columns), but the frequency modulations were not affected much (columns five and six). Recall from Table 3 that this microphone had the largest baseline amplitude measures when no modulations were imposed.

Table 4.
Comparison of microphone extracted modulation measures (CV for 100 Hz modulations and PF for $F_o/2$ modulations) for different types of imposed modulation, 0° angle and 4 cm distance from source.

	Imposed Modulation							
	AM(5%,10Hz)		AM(5%,$F_o/2$)		FM(2%,10Hz)		FM(5%,$F_o/2$)	
F_o	100Hz	300Hz	100Hz	300Hz	100Hz	300Hz	100Hz	300Hz
	CV		PI		CV		PI	
Direct	3.39	3.53	3.47	4.83	1.34	1.42	3.03	3.60
Mic								
1	3.42	3.56	4.77	3.68	1.36	1.42	4.52	3.38
2	3.36	3.53	2.78	4.32	1.34	1.41	3.09	3.64
3	3.43	3.55	5.37	3.56	1.37	1.41	4.23	3.40
4	3.35	3.53	3.81	4.72	1.32	1.40	2.79	3.27
5	3.68	3.72	3.90	4.09	1.32	1.40	2.44	4.19
6	3.49	3.58	5.03	4.39	1.35	1.41	2.66	2.98

For the $F_o/2$ modulations, there was greater variability. About half of the measurements were below the direct analysis values (columns two and three and columns seven and eight). This may seem strange. One does not expect electronic equipment to *deflate* perturbation values. The problem is likely to be phase distortion (discussed below), which has a strong effect on $F_o/2$ modulations. Because of the complex nature of the waveform, modulation frequencies that approach the carrier frequency change the waveshape unevenly from cycle to cycle. If the exact waveshapes are not preserved by the microphones or the amplifiers, different event locations will be detected. The waveform matching technique may smooth out some of these variations of individual events. This may be one of the drawbacks of this method of extraction, but it has many other advantages (Titze & Liang, in review). In general, the $F_o/2$ measures always appeared to be more variable across microphones than the measures for the low frequency modulations.

Phase distortion varied between microphone types as demonstrated in Figure 4. The signals were recorded simultaneously for F_o = 100 Hz, 0° angle, and 4 cm distance. The stimulus signal, for comparison, is Figure 3(a). All of the microphones distorted the signals differently. For example, compare waveforms (a) and (d), representing the omnidirectional condenser versus the cardioid dynamic microphones, respectively. Not only is there a significant phase delay in (d), as seen by the line-up at time = 0, but the largest biphasic pulse is less symmetric about the zero axis. The major downward peak is greater than the following upward peak. This difference between the major upward and downward peaks is even greater when waveforms (c) and (d) are compared. Most importantly, however, all of the recorded signals suppress the large positive peak *preceding* the major negative peak (compare with Figure 3a). This is the portion of the waveform that represents the derivative of the glottal flow pulse. Clearly, loss of low-frequency response, from a combination of the microphones and the rest of the electroacoustic link, has degraded this signal. Similar phase distortions occurred also at 300 Hz.

To verify that the phase distortion was not unique to loudspeaker presentation, experiments with human speakers were also conducted. One experiment was designed to determine any near field effects on sustained phonation. Three microphones of the same type were positioned at three angles (0°, 45°, 90°) at a distance of 4 cm from the subject. The experiment was then repeated at 15 cm. Analysis of the signals revealed no obvious microphone differences (in waveshape or perturbation measures) for these positions. In another experiment, the six microphones were positioned in a cluster at a fixed distance from the source. To determine if there were any spatial effects within the cluster, two of the microphones were reversed in position. A comparison of the resulting waveshape indicated no significant visual differences, and perturbation measures were also within baseline extraction errors.

Next, four normal subjects (two males, two females) produced sustained phonation of an [a] vowel with all six microphones positioned in a cluster (0° angle, 15 cm from the subject's mouth). The signals from the microphones were digitized simultaneously for analysis of an identical section of phonation. Table 5 presents the results of perturbation analysis on the human phonations. Note that some large outliers are found among the measures. These occur most often with microphones 3, 4, and 5, the dynamic types. Obvious examples are the amplitude CV's for subjects M2 and F1 for microphone 4. Recall that this was the microphone with the largest phase delay (part d in Figure 4).

Figure 5 shows some recorded waveshapes for subject M1, a male phonating at 121 Hz. It can be seen that the phase distortions are similar to those found for the 100 Hz loudspeaker presentations. In particular, note the phase delay with microphone 4 (part d of the figure). For the female subject F1 phonating at 246 Hz (Figure 6), a major waveshape distortion is seen in part (d). Microphone 4 has the only waveshape for which first formant ripple drifts downward instead of upward. Also, microphone 5 attenuates the formant ripple at this higher F_o.

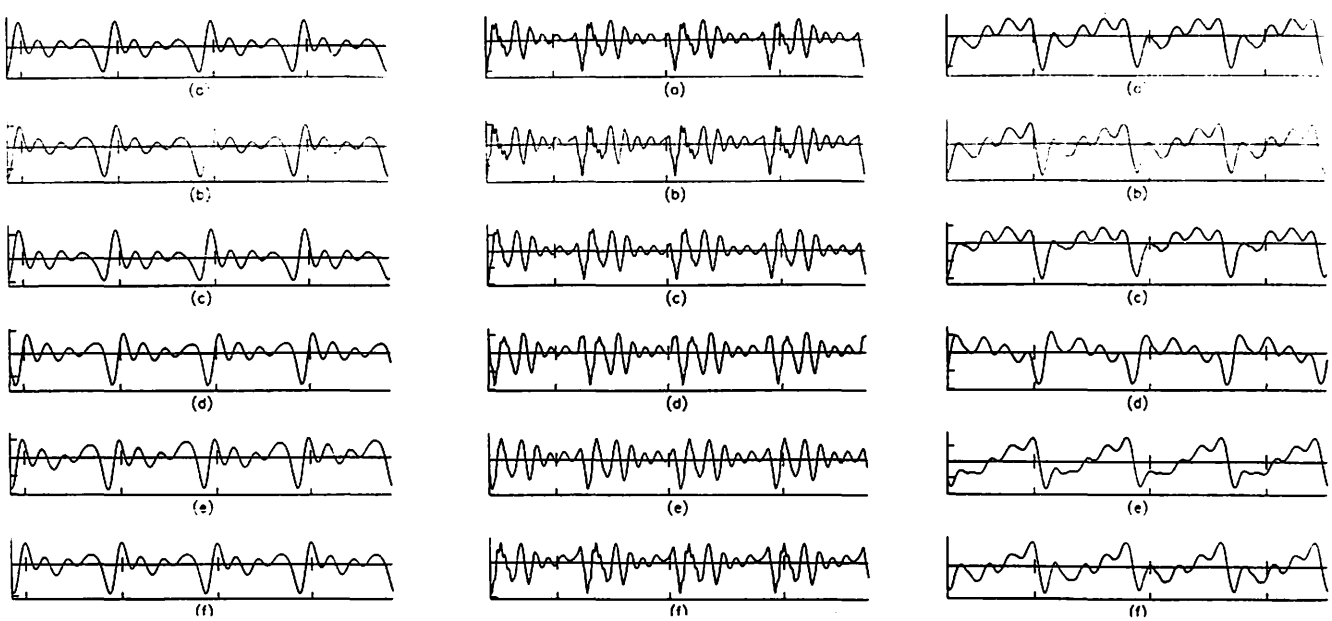


Figure 4. (left) Recorded signals from microphones for stimulus in Figure 3(a) at 0° angle and 4 cm distance. Parts (a)-(f) correspond to microphones 1-6. Figure 5. (center) Recorded signals from microphones for stimulus of male subject (M1) at 0° angle and 15 cm distance. Parts (a)-(f) correspond to microphones 1-6. Figure 6 (right). Recorded signals from microphones for stimulus of female subject (F1) at 0° angle and 15 cm distance. Parts (a)-(f) correspond to microphones 1-6.

The third formant energy associated with glottal excitation is also of interest. For this we return to Figure 5. Third formant energy is seen as a high frequency ripple on the first positive peak after the largest negative peak. It is significantly attenuated by the dynamic microphones (parts c, d, and e). Loss of this energy may effect the extraction of F_o, especially if peak detection or waveform matching is used. This may be part of the reason why the female subject showed greater variability in the perturbation measures than the male, given that little second formant energy is visible in her waveform.

A graphical comparison between the condenser microphones and the dynamic microphones for human phonations is shown in Figure 7. Scattergrams are shown for all amplitude perturbation measures against their means for each of the three microphones. Part (a) is for the three condenser microphones and part (b) is for the three dynamic microphones. It appears that the scatter is larger for the dynamic microphones than for the condenser microphones, especially at higher perturbations.

Table 5.
Perturbation measures of six microphones from identical sections of phonation over four normal subjects (0° angle, 15 cm).

			AMPLITUDE		FREQUENCY	
SUBJECT	F_o	MIC	CV	P1	CV	P1
M1	121Hz	1	4.01	0.92	0.34	0.17
		2	4.36	1.16	0.34	0.17
		3	3.99	1.00	0.34	0.17
		4	3.85	0.82	0.35	0.17
		5	3.82	0.90	0.34	0.18
		6	4.46	1.00	0.34	0.17
M2	85Hz	1	10.01	3.30	0.99	0.36
		2	9.97	3.29	0.99	0.33
		3	9.19	2.55	0.98	0.59
		4	7.60	2.19	1.00	0.34
		5	6.21	1.74	0.98	0.35
		6	9.92	2.98	0.99	0.34
F1	246Hz	1	5.82	2.66	0.54	0.59
		2	5.68	3.43	0.54	0.66
		3	5.75	2.41	0.54	0.95
		4	7.13	3.24	0.54	0.71
		5	5.58	2.66	0.54	0.64
		6	5.66	3.22	0.54	0.60
F2	164Hz	1	2.12	0.76	0.41	0.27
		2	2.12	0.84	0.41	0.27
		3	2.04	0.84	0.41	0.27
		4	2.27	0.79	0.42	0.28
		5	2.18	0.95	0.41	0.27
		6	2.20	0.74	0.41	0.26

Figure 8 shows the same comparison for synthesized perturbations and modulations. In groups of three, the condenser microphones do not outperform the dynamic types, but on an individual basis there is a substantial difference. Note how close microphone 2 is to the diagonal (asterisks in part (a) of the figure). This is by far the best microphone under the conditions we tested. Microphone 1, the omnidirectional professional grade condenser microphone, deviates more from the diagonal, especially at low perturbations. Microphone 5, the consumer grade omnidirectional condenser type, shows the worst performance at low perturbations. This microphone biases the performance of the entire condenser group. If the consumer grade microphones (5 and 6) are removed from the data set, condenser types perform collectively better than dynamic types. Also, no individual dynamic microphone outperformed the best condenser microphone (number 2).

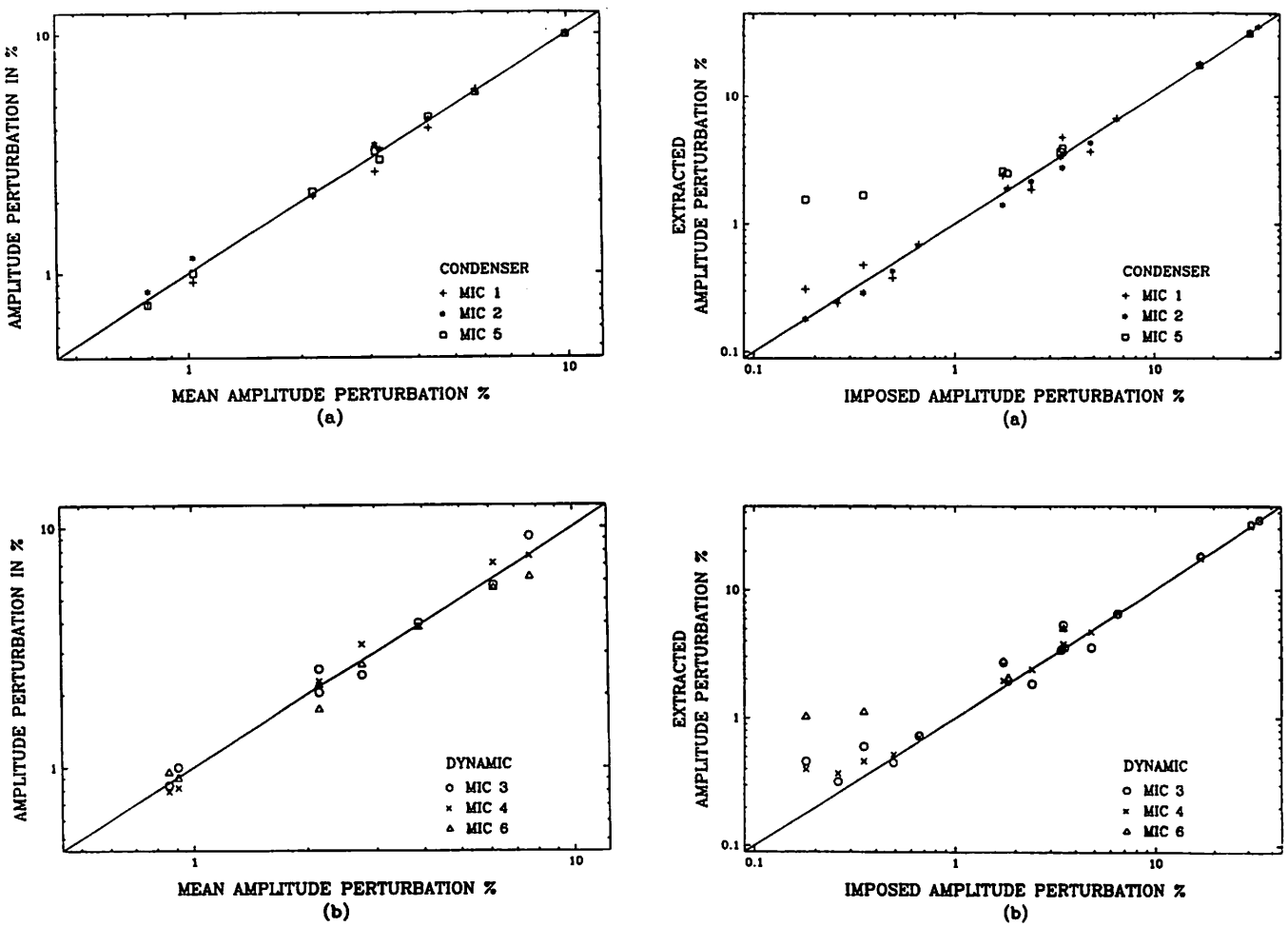


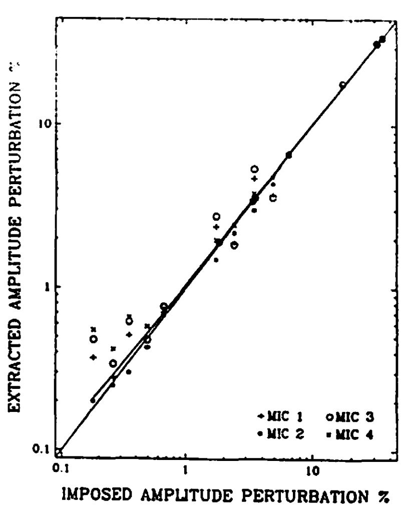
Figure 7 (left).Scatter plot for amplitude perturbations (CV and P1) extracted from four subjects (2 males and 2 females) at 15 cm and 0° angle (a) condenser microphones, and (b) dynamic microphones. Figure 8 (right).Scatter plot for amplitude perturbations (CV and P1) extracted from synthesized signals presented over a loudspeaker at 4 cm and 0° angle (a) condenser microphones, and (b) dynamic microphones.

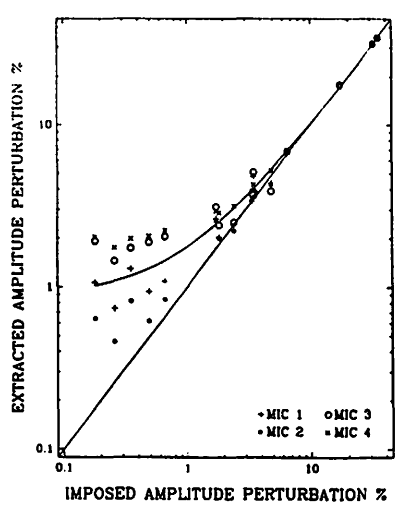
Microphone Distance

Table 6 presents an ensemble average over four professional microphones, two fundamental frequencies, and two measures (CV and P1) of perturbation to determine the baseline levels at different microphone positions. Note the general increase with distance, basically an order of magnitude from 4 cm to 1 m.

Table 6.
Perturbation measures (ensemble averages over four professional microphones, two fundamental frequencies, and two measures CV and P1) for no imposed modulation with different microphone distances and angles to the source.

Distance	Angle 0	Angle 45	Angle 90
Amplitude Measures			
4 cm	0.24	0.31	0.30
30 cm	1.33	1.75	2.09
1 m	3.96	5.03	6.72
Frequency Measures			
4 cm	0.03	0.04	0.04
30 cm	0.16	0.23	0.31
1 m	0.49	0.78	1.28





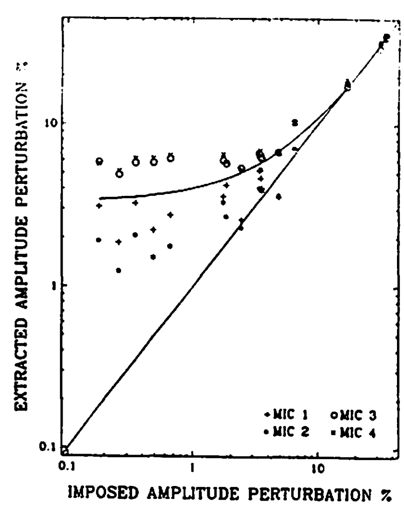


Figure 9 (left). Scatter plot for amplitude perturbations from professional-grade microphones, all modulations, F_o of 100 Hz and 300 Hz, 0° angle, and 4 cm distance. Figure 10 (center). Scatter plot for amplitude perturbations from professional-grade microphones, all modulations, F_o of 100 Hz and 300 Hz, 0° angle, and 30 cm distance. Figure 11 (right). Scatter plot for amplitude perturbations from professional-grade microphones, all modulations, F_o of 100 Hz and 300 Hz, 0° angle, and 1 m distance.

Scatter plots were made for both amplitude and frequency measures. These scatter plots relate the extracted perturbations to the imposed perturbations (direct measure). To conserve space, mainly the amplitude plots are shown because amplitude perturbation is most effected by distance and angle. Frequency plots were very similar in shape, however.

Figure 9 is a scatter plot of the amplitude perturbation measures (CV and P1) for the four professional grade microphones at 4 cm. Fundamental frequencies of 100 Hz and 300 Hz are included and the angle is 0°. The plot is basically a combination of Figures 8(a) and (b), with consumer grade microphones excluded. Note that microphone 2 (the cardioid condenser) shows the best results and microphone 3 (the omnidirectional dynamic) shows the worst results. For perturbations between 1% and 10%, the errors in extraction range between 1% and 10% at this distance. This suggests that the electroacoustic link contributes to extraction errors between 1 part in 10^4 (80 dB) and 1 part in 10^2 (40 dB), depending on the manner in which the signal is distorted. A slight upward trend from the diagonal at low perturbations (near 0.2%) is observed by fitting a second order polynomial to all the data.

Figure 10 is a similar plot at a distance of 30 cm, and Figure 11 at a distance of 1 m. In these scatter plots, the second order fit shows a major trend. There are progressively larger deviations from the diagonal for greater distances. For a commonly used 30 cm distance, a 1% amplitude perturbation is inflated by a factor of 2, and a 0.1% perturbation by a factor of 10 or more. At 1 m distance, only amplitude perturbations of 10% or more have reasonable precision.

To bracket the frequency perturbation results and to show their similarity to amplitude perturbations, Figures 12 and 13 show scatter plots for frequency measures at 4 cm and 1 m, respectively. These should be compared with the amplitude measures in Figures 9 and 11.

The microphone rankings were retained at all of the distances and with all of the measures. Best results were generally obtained with microphone 2, second best with microphone 1, and worst with microphones 3 or 4. This is consistent with previous discussions on sensitivity and phase distortion of these microphones. Phase of the recorded signals was affected by distance, as observed informally by comparing waveforms at various distances and angles. This is a topic for another discussion, however, because it involves details of sound field patterns and room acoustics.

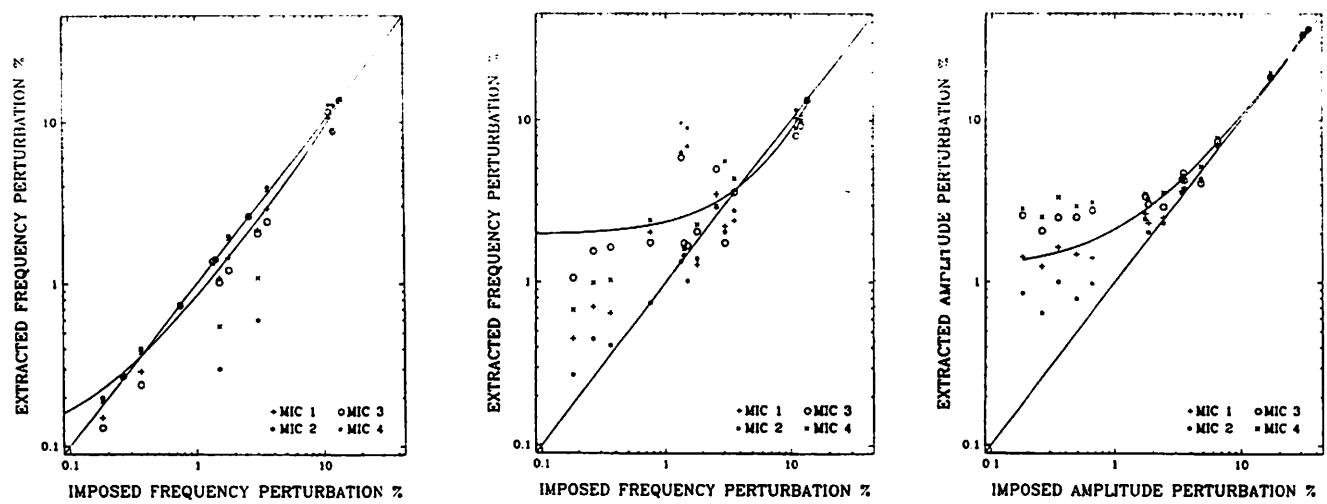


Figure 12 (left). Scatter plot for frequency perturbations from all professional grade microphones, all modulations, F_0 of 100 Hz and 300 Hz, 0° angle, and 4 cm distance. Figure 13 (center).Scatter plot for frequency perturbations from all professional grade microphones, all modulations, F_0 of 100 Hz and 300 Hz, 0° angle, and 1 m distance. Figure 14 (right).Scatter plot for amplitude perturbations from professional-grade microphones, all modulations, F_0 of 100 Hz and 300 Hz, 45° angle, and 30 cm distance.

Microphone Angle

Angular orientation of the microphone had an effect on the measures primarily at large distances (Table 6). At a distance of 4 cm, the measures were inflated by 20% - 30% for angles of 45° and 90°. At a distance of 30 cm, the measures were about 50% larger at 45° and approached 100% inflation at 90°. At 1 m distance, the inflation with angle was slightly greater than at 30 cm. A typical scatter plot for amplitude measures is shown in Figure 14 (see preceding page) for an angle of 45° and a distance of 30 cm. This plot should be compared with Figure 10 to get the graphic effect of angle.

It is interesting to ask whether microphone preference due to directionality (cardioid versus omnidirectional) changes with angular orientation. Our data showed that it does not. In every case tested [3 distances x 3 angles], the cardioid condenser microphone performed best. The sound fields from the sources and the microphones are apparently broad enough to produce essentially complete overlap.

Conclusions

Microphone sensitivity has the greatest effect on the extraction of perturbation measures from recorded signals. When sensitivity in low, the signal can be preamplified, but noise is amplified along with the signals. In this study, sensitivity was the largest variable between the professional grade microphones; with the condenser type being approximately 20 dB more sensitive than the dynamic type.

Baseline frequency perturbations with professional grade microphones and digital recording equipment were about 0.05%, and baseline amplitude perturbations are about 0.3% at a distance of 4 cm. These are the values of jitter and shimmer when there are no imposed modulations and when the entire electroacoustic link affects the signal (digital to analog converter, power amplifier, loudspeaker, microphone, preamplifier filter, DAT recorder, and analog to digital converter). Some consumer grade microphones used in conjunction with the same equipment and analysis programs inflated the frequency perturbation to a range of 0.1-0.2% and amplitude perturbation to a range of 1.0-2.0%.

When the microphone distance was changed from 4 cm to 1 m, perturbation measures increase by an order of magnitude. This increase can be explained by a loss in signal to noise ratio. In this experiment, a large increase in voltage gain (microphone preamplifier) was necessary to overcome the inverse square loss of acoustic power by distance. This makes it virtually impossible, in a normal recording environment, to resolve normal vocal jitter and shimmer at distances greater than a few cm from the mouth.

Another variable between microphone type was phase distortion. This may have significant effects on the extraction algorithms for perturbation, especially those that depend on specific events on the waveform (peaks, zero crossings, etc.), posing a problem for certain types of analyses. The interaction between distance, angle and harmonic structure of the signal complicates the transduction process, which in turn introduces variability into the perturbation analysis. Once phase distortion has altered the microphone signal, from whatever source, it may not be possible to obtain the true perturbation values.

In this study there appeared to be some evidence that the cardioid pattern is preferable to the omnidirectional pattern, probably because of its greater on-axis gain. But this is based on a very limited sampling of microphones, making generalizations impossible at this stage. At great dis-

tances and wide angles, there may be more of a pattern effect that would ultimately favor the omni-directional microphone, but this would need further study.

Under some conditions (e.g. clinical applications where the patient is immobile), microphone placement may be restricted to distances greater than a few cm, or to a relatively noisy environment. A future study is needed to determine the effects of room reverberation and ambient noise. Judging by the current results, however, severe limitations are expected.

It is also important to note that, in an attempt to simplify this study, only sustained phonation was used for the stimulus. For running speech there are considerable aerodynamic artifacts (plosives, DC drift, etc.) that will effect the acoustic signal at distances near the mouth. Therefore, a compromise in microphone placement is needed, depending upon the type of vocal tasks that are used.

If a recommendation had to be given from this initial study, such a compromise would be to place a professional grade cardioid condenser microphone a few centimeters from the mouth, at about 45°. Microphone sensitivity should be no less than -60 dB, and preconditioning filters and amplifiers should be linear phase. This recommendation applies only to perturbation analysis on sustained vowels at this stage.

Acknowledgements

This study was supported by a grant from the National Institutes of Health, grant No. R01 DC00387-04. The authors wish to thank Ron Scherer for his assistance, Chwen-Geng Guo, Larry Brown, and Mitch Wolfe for their assistance in the computer analysis and Pamela Rios for assistance in manuscript preparation.

References

Doherty, T.E., & Shipp, T. (1988). Tape recorder effects on jitter and shimmer extraction. Journal of Speech and Hearing Research,31, 485-490.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. Journal of Speech and Hearing Research, 30, 448-461.

Horii, Y. & Hata, K. (1988). A note on phase relationships between frequency and amplitude modulation in vocal vibrato. Folia Phoniatrica, 40,303-311.

Karnell, M.P. (1992). Laryngeal perturbation analysis: Minimum length of analysis window. Journal of Speech and Hearing Research,34,544-548.

Karnell, M.P., Scherer, R.S., & Fischer, L.B. (1991). Comparison of acoustic voice perturbation measures among three independent voice laboratories. Journal of Speech and Hearing Research, 34,781-790.

Milenkovic, P. (1987). Least mean square measures of voice perturbation. Journal of Speech and Hearing Research,30,529-538.

Pinto, N. & Titze, I. (1990). Unification of perturbation measures in speech analysis. Journal of the Acoustical Society of America, 87(3), 1278-1289.

Scherer, R.C., Gould, W.J., Titze, I.R., Meyers, A.D., & Sataloff, R.T. (1988). Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. Journal of Voice,3,230-244.

Titze, I.R., & Liang, H. Comparison of F_o extraction methods for high precision voice perturbation measurements. Journal of Speech and Hearing Research, (in review).

Winholtz, W.S., & Ramig, L.O., (1992). Vocal tremor analysis with the vocal demodulator. Journal of Speech and Hearing Research,35,562-573.

NCVS Status and Progress Report - 3
December 1992, 135-147

Effects of Intensive Voice Therapy on Speech Deficits Associated with Bilateral Thalamotomy in Parkinson's Disease: A Case Study

Stefanie Countryman, M.A.
Recording and Research Center, Denver Center for the Performing Arts
Lorraine Olson Ramig, Ph.D.
University of Colorado-Boulder
Recording and Research Center, Denver Center for the Performing Arts

Abstract

Stereotactic brain surgery to reduce tremor and rigidity in movement disorders is no longer a treatment option of the past. This study documents the effects of intensive voice therapy on the speech and voice characteristics in a Parkinson's diseased subject after bilateral thalamotomies. A review of the literature indicates few scientific studies on voice deficits induced by bilateral thalamotomies nor the responsiveness of these deficits to speech therapy. Using a recently-developed method of voice therapy for individuals with Parkinson's disease, the subject received 16 sessions of therapy. Acoustic analysis of variables collected pre-treatment and post-treatment indicated that the intensive voice therapy was effective. Measures of jitter, shimmer, coefficient of variation of frequency, amplitude modulation, frequency modulation, maximum duration of sustained vowel phonation, harmonics-to-noise ratio, harmonic spectral slope, and intensity all demonstrated statistically significantly improvements following the intensive voice therapy program. The subject also self-rated significant improvement in perceptual characteristics of her voice. The subject's improvement is discussed in relation to hypokinetic dysarthria versus pseudobulbar dysarthria. Further research is suggested for both effects of the surgeries and results of treatment on speech deficits created by these surgeries.

Voice Therapy in Thalamotomy and Parkinson's Disease

Stereotaxic thalamotomy is a procedure used to control rigidity and tremor in Parkinson's disease (PD) as well as in other movement disorders. Until the late 1960's,when levadopa (L-dopa) became available, it was considered to be the best treatment option for the control of tremor and rigidity in PD. After the use of L-dopa became widespread, the occurrence of thalamotomy reduced significantly (Tasker, Organ, and Hawrylyshyn, 1983). However, in the last decade, the number of thalamotomies being performed has increased as a result of drug-resistive patients suffering from severe and disabling tremors (Manen, Speelman, and Tans, 1984).

During a thalamotomy, a lesion is placed in the ventral lateral nucleus of the thalamus. It has been well-documented that unilateral and, more predominantly, bilateral lesions in the ventral lateral nucleus of the thalamus produce speech deficits (Allen, Turner, and Gadea-Ciria, 1966; Beel, 1968; Darley, Brown, and Swenson, 1975; Samra, et al., 1969; Selby, 1967; Riklan, Levita, Zimmerman, and Cooper, 1969). The specific speech symptoms that have been reported include the following: impairment of the vocal pitch and intensity range, diminished voice volume, dysarthria-impaired articulation, slurring, some changes in nasal intonation, dysphasia, and respiratory depression (Allen, et al., 1966; Darley et al., 1975; Samra, et al., 1969; Selby, 1967; Riklan, et al., 1969).

There are few well-documented studies which examine responsiveness of the resulting deficits to speech therapy. Many of the investigators mentioned above state that surgically-related speech deficits spontaneously recover. However, Selby (1967) found that bilateral lesions in the thalamus and basal ganglia were the key factor in continuing speech deterioration in thalamotomy. He also states that the patient's pre-existing speech disorder was also a contributing factor in spontaneous recovery of the surgically-induced speech and voice deficits. Riklan et al. (1969) state that 18% of patients with bilateral operations experience some speech deficit with 6% exhibiting lasting deficit. Unilateral left-sided lesions appear to result in speech deficits more often than unilateral right-sided lesions (Riklan et al., 1969). Samra et al. (1969) state that mild speech disturbances spontaneously improve in a few weeks, whereas severe cases required intensive speech rehabilitation and continuous practice before patient improved.

While speech disorders are typically associated with Parkinson's disease (Hoberman, 1958; Canter, 1965; Darley, Aronson, and Brown, 1969a, 1969b; Logemann, Fisher, Boshes, and Blonsky, 1978; Oxtoby, 1981; Streifler and Hofman, 1984), speech therapy has generally been viewed as ineffective (Allan, 1970; Aronson, 1985; Green, 1980; Rubow and Swift, 1985; Sarno, 1968). However, recently, an effective intensive voice therapy approach has been developed for patients with Parkinson's disease (Ramig, Mead, and DeSanto, 1988a; Ramig, Mead, Scherer, and Horii, 1988; Ramig, Horii, and Bonitati, 1991). This treatment approach has been designed to impact the overall functional communication skills of these patients by targeting the underlying laryngeal pathophysiology that contributes to the speech and voice deficit and focuses on increased phonatory effort. Although the results of this treatment program have been positive, it was noted that the speech of patients who had received a thalamotomy was less likely to improve than those who had not received the surgical intervention for the treatment of their Parkinson symptoms (L.O. Ramig and C. M. Bonitati, personal communication, September 26, 1991). According to Ramig and Bonitati (1991), these thalamotomy patients were stimulable for increased phonatory effort but seemed unable to maintain the day-to-day treatment effects.

Although many of the previously mentioned investigators briefly discuss improvement in speech and voice deficits after thalamotomy, there are few studies with valid, scientific data to sup-

port these conclusions. The purpose of this paper is to describe voice and speech characteristics in a bilateral thalamotomy patient with Parkinson's disease and to document the effect of intensive voice therapy on the speech and voice of that patient.

Methods

Subject

The subject was a 65 year old, right-handed female in stage III on the Hoehn and Yahr Scale for Parkinson's disease (1967). She was first diagnosed with the disease in 1985. The subject was a semi-professional opera and chorale soprano with a master's degree in voice performance. The patient's initial symptom was tremor in the right hand which eventually developed into complete right-sided tremor. At the time of diagnosis the patient reported no problems or noticeable deterioration in her speech and voice. Initially, the patient was prescribed medication to control the tremor. However, after several months without beneficial reduction of her tremor, the doctor determined she was refractory to the Parkinson medication. Consequently, the patient's neurologist referred her for a left-sided thalamotomy for relief of her right-sided tremor.

A left, radio-frequency thermothalamectomy was performed in January of 1986. Two lesions were placed in the ventralateral nucleus of the left thalamus which completely abated the patient's right-sided tremor and rigidity. Upon completion of the surgery and recovery, the attending neurologist reported that speech, voice, muscular strength, and neurologic findings were all within normal limits. The patient also reported no speech or voice deficits following surgery.

Four years post-surgery (1990), the patient developed severe left-sided tremor and rigidity and was again referred by her neurologist for a right-sided radio-frequency thermothalamectomy due to continued refraction to all medication and the debilitating effect the tremors had on the patient's lifestyle. At this time, pre-surgery, the patient and the patient's spouse reported a slight slurring in her articulation and no deterioration or changes in her vocal quality.

The second surgical procedure was completed in February 1990. Prior to the surgery, the patient was informed by the attending neurologist of a 15% chance of pseudobulbar speech dysfunction which would be temporary and a 5% chance of permanent disability which would manifest itself as speech dysarthria and/or gate disturbance.

During the second surgery, the attending neurologist reported that lesioning of the right ventralateral thalamus was terminated when a slight slowing of the patient's speech was noted interoperatively. Upon the completion of the surgery, the attending physician determined the patient to be neurologically intact with slight dysarthria. Immediately post-operatively, the patient's left-sided tremor and rigidity were completely relieved.

Within two days of the surgery, the attending neurologist reported the patient exhibited a mild to moderate dysarthria which he felt was a direct result of the surgical procedure. The patient did not exhibit any swallowing difficulties post-operatively. The attending neurologist also determined that the patient had not developed full pseudobulbar dysfunction despite her speech dysarthria. Subsequently, the patient was referred for speech therapy. The patient's self-report at that time was that her speech and voice quality had significantly deteriorated following the second surgical procedure. She reported reduced loudness, slurring of speech, a thickening of her tongue, reduced timbre and resonance, and inability to project her voice while speaking and singing.

The patient's initial post-operative speech evaluation occurred in February of 1990. At that time, the evaluating speech pathologist reported a thin, weak voice quality, reduced respiratory

capacity during phonatory tasks, mildly impaired articulation, normal pitch, and no pitch breaks. An oral peripheral exam revealed facial and oral speech musculature within normal limits. The speech pathologist diagnosed a mild hypokinetic dysarthria and moderate voice disorder.

The patient began biweekly speech therapy in March of 1990 and continued until the end of April. Weekly therapy was reinstated in June until October of the same year. The speech pathologist reported therapy focused on improved respiratory support, the use of pausing and phrasing techniques, improved articulatory precision, improved inflection and intonation, and reducing a wet and raspy vocal quality. The speech pathologist noted some improvements at that time, particularly in inflection and articulatory precision. However, these improvements were only rated subjectively and no measurable data was obtained pre- to post-treatment.

An otorhinolaryngological evaluation in June of 1990 reported a raspy and hoarse voice with mucous observed stringing between the cords and a moderate loss of substance on the left true cord. No lesions were noted, but irregularity of the vocal cord surface was seen. Senile changes of the true vocal cords were noted as well as a potential of voice abuse or misuse.

In July 1991, the patient enrolled in the current voice therapy clinical research project. Perceptual speech and voice characteristics included reduced loudness, hoarseness, abnormally low pitch for the patient's sex and age, intermittent monotonicity, articulatory imprecision, vocal fry, intermittent harsh raspy voice, mild intermittent vocal straining, reduced respiratory support, a slight slowness of speech, and intermittent vocal tremor. In addition, the patient reported that her singing voice had lost its resonance and timbre and sounded flat. All of these characteristics combined to create a moderate to severe voice and speech disorder. The patient rated her speech intelligibility at 50 percent. The speech pathologist rated speech intelligibility at 80 percent.

An oral peripheral exam revealed facial, mandibular, and tongue musculature to be within normal limits. Palatopharyngeal musculature also appeared to be within normal limits.

Extensive neuropsychological and neurological evaluations were completed just prior to beginning voice treatment. The neuropsychological examination revealed a very mild inefficiency in learning new non-verbal information (Learning Component of Figure Memory Test) and a mild to moderate impairment of sustained attention and concentration (Digit Vigilance Test). All other cognitive tests yielded results within normal limits. Psychosocial testing revealed normal functioning. The neurological examination confirmed the patient's diagnosis of PD with no unusual secondary symptoms or other medical problems that would contribute to her speech deficits. The neurologist described the patient's speech as almost pseudobulbar quality. The otolaryngological examination revealed no significant changes since her previous examination.

Treatment Program

The patient participated in 16 sessions of voice therapy (4 times a week for one hour). The treatment approach for Parkinson's disease designed by Ramig et al. (1988a, 1988) was used for this patient. Treatment techniques were designed to improve perceptual characteristics of voice by targeting the hypothesized underlying laryngeal pathophysiology and maximizing phonatory and respiratory effort. The elements for the development of this treatment approach are summarized in Table 1. The program was intensive and specifically designed to maximize patient motivation, effort, and carry-over. Specific techniques included maximizing vocal fold closure through vocal isometric exercises, maximizing frequency range of phonation through vocal range exercises, and improving respiratory support by breathing deeply and frequently while speaking as well as thinking loud and speaking on top of the breath. A typical session included the following activities: repeated maximum

sustained vowel phonations (/a/) while utilizing optimum respiratory support, performing the highest and lowest possible fundamental frequencies the patient could obtain, breathing exercises with visual biofeedback (NIMS Respigraph System PN SY03), and speech tasks utilizing the same phonatory and respiratory techniques that were used in sustained phonation. A tape recorder was used to provide vocal quality feedback to the patient during each task.

Table 1

Framework and rationale for perceptual characteristics of speech; hypothesized laryngeal and/or respiratory pathophysiology; acoustic/physiologic variables measured; perceptual variables measured.

Perceptual characteristics of speech	Hypothesized laryngeal and/or respiratory pathophysiology measured	Acoustic, physiologic variables	Perceptual variables measured
Reduced loudness breathy, weak voice (Logemann, et al., 1978; Aronson, 1985)	**Bowed vocal folds** (Hansen et al., 1984), **rigidity, hypokinesia in laryngeal and/or respiratory muscles; reduced adduction; reduced inspiratory, expiratory volumes** (Critchley, 1981)	**Maximum duration of sustained vowel phonation (sec)**	**Loudness; Breathiness; Intelligibility**
Reduced pitch variability; monopitch (Logemann, et al., 1978; Aronson, 1985).	**Rigidity cricothyroid muscle** (Aronson, 1985).	**Maximum range of fundamental frequency (st)**	
		Variability of fundamental frequency in reading (stsd)	**Monotone; Intelligibility**

Data Collection

Experimental data were collected on the following variables pre- and post-treatment: maximum duration of sustained /a/ vowel phonation, maximum fundamental frequency range, fundamental frequency variation in reading, vocal fold adduction, amplitude modulation, frequency modulation, and intensity. Pre- and post-treatment perceptual ratings of the the subject's speech were made by the subject and two speech pathologists.

The subject was seated in an IAC sound-treated booth with a headset microphone (AKG 410) positioned 8 cm in front of the lips. After preamplification through an ATI-1000, the microphone signal was recorded onto a Sony Digital PC-108M (DAT) eight-channel recorder. A Bruel and Kjaer Type 2230 sound level meter was placed in the booth 1/2 meter from the subject's mouth. Sound

Type 2230 sound level meter was placed in the booth 1/2 meter from the subject's mouth. Sound pressure level was recorded onto the Sony DAT recorder. The electrodes of a Synchrovoice Inc. electroglottograph (EGG) were placed on the subject's neck at the thyroid lamina. After amplification (Tektronix Amplifier 502 TM 506), the EGG signal was recorded onto the eight-channel DAT recorder.

In order to obtain measures of maximum duration of sustained vowel phonation and both amplitude and frequency modulations, the subject was asked to take a deep breath and sustain phonation of the vowel /a/ for as long as possible. A timer with a second hand was within the subject's view to encourage her to monitor her performance and sustain phonation longer with each repetition. Encouragement was given to motivate performance after each phonation. Six maximally sustained phonations were collected during the recording. In order to obtain measures of fundamental frequency variation in reading, the subject was asked to read aloud the phonetically-balanced Rainbow Passage (Fairbanks, 1960) at a comfortable rate and loudness.

The subject was asked to generate her maximum fundamental frequency range including falsetto and vocal fry (Hollien, Dew, and Philips, 1971) to collect measures of maximum phonation range. The experimenter used both the step and gliding (glissando) methods (Reich, Frederickson, Mason, and Schlauch, 1990) until both the experimenter and subject were satisfied that performance limitations had been met. This task was repeated six to nine times within each session.

In order to evaluate the subject's impression of the treatment effectiveness as well as the impact of the treatment on the subject's daily living, perceptual rating scales were completed by the subject at the beginning of each voice data collection session. The subject was asked to rate perceptual variables of loudness, monotonicity, and intelligibility in her speech.

The subject used the same visual analog scale (Kempster, 1984) pre- and post-treatment for perceptual ratings. The scale, which targets perceptual characteristics related to treatment goals for the subject, was designed to be a clinically-feasible tool for both the patient and professional. Visual analog scales were selected for all perceptual ratings due to their increased sensitivity when compared with other forms of scaling such as equal-appearing intervals (Kempster, 1984; Schiffman, Reynolds, and Young, 1984).

Because of potential instabilities in measures of maximum performance (R. D. Kent, J. F. Kent, and Rosenbek, 1987) as well as the variability associated with Parkinson's patients' performance, two pre-treatment voice recordings and two post-treatment voice recordings were completed. The same experimenter collected pre- and post-treatment data. The experimenter was careful to elicit maximum performance as determined by her clinical decision. In addition to the data collection sessions, clinical treatment data were collected daily on the measures of maximum duration of sustained vowel phonation and maximum fundamental frequency range using a stop watch and the Vocal Demodulator MVD-100 (Winholtz and Ramig, 1992) respectively. It was expected that the daily clinical measures would serve as a general indicator of consistency in pre- to post-treatment changes across experimental and clinical settings.

Speech Data Analysis

To obtain duration measures of maximally sustained vowel phonation, each phonation was input into a Data Precision 6000 Model 611 digital oscilloscope at a sampling rate of 20 samples per second. Cursors were hand positioned to mark the monitored-displayed zero crossing preceding the first negative-going peak at the onset and the zero crossing following the final positive-going peak at the offset of each vowel.

To obtain measures of maximum fundamental frequency range, each attempt at maximum high phonation was high-pass filtered and maximum low phonation was low-pass filtered below and above the predicted fundamental frequency range respectively. The signal passed into a Hewlett Packard 54503A digitizing oscilloscope and frequency in Hertz (Hz) was obtained through cursor marking. For this procedure, sampling rate varied from 5,000 to 10,000 samples per second, depending upon fundamental frequency. The maximum high and low frequencies were then converted to express the maximum range in semitones (ST). Measures of mean fundamental frequency and fundamental frequency variability during reading (semitone standard deviation; STSD) were obtained by digitizing the reading of the Rainbow Passage using the application C-Speech (Milenkovic, 1987).on a 386 computer.

Amplitude and frequency modulation were obtained by passing the signal into the Vocal Demodulator (WVD-100) and reading the percentages from the demodulator at the midvowel point. Using the analysis/synthesis software GLIMPES (Glottal Imaging by Processing External Signals) (Titze, 1984), amplitude and frequency perturbation and EGGW measures were obtained. EGGW is a measure of glottal adduction and is defined as the width of the EGG signal at 25%, 50%, and 75% of the wave's amplitude divided by the period (Scherer and Vail, 1988).

Intensity measures were obtained by digitizing at 20,000 samples per second the mic signal into the VAX system computer through a 16 bit resolution DSC-200 A/D converter. The signal was calibrated using a tone of known loudness. Two points were marked on the signal and the average intensity level between the points was calculated.

Standard procedures for analysis of visual analogue scales (Boeckstyns and Backer, 1989) were used to obtain perceptual data.

Table 2.

Median, lower hinge and upper hinge of the acoustic variables collected pre- and post- treatment.

	PRE 1			PRE 2			POST 1			POST 2		
Variable	Med	LH	UH	Med	LH	UH	Med	LH	UH	Med	LH	UH
Max Duration (sec)	21.30	20.45	21.45	21.80	21.75	22.45	25.0	24.1	27.5	26.4	24.35	26.85
Max Dur (daily sec)	17.50	16.00	19.00	20.00	18.00	22.00	22.00	21.00	26.00	23.00	21.00	26.00
Semitone Range	28.65	****	****	****	****	****	33.24	****	****	****	****	****
Jitter (%)	0.38	0.29	0.48	0.47	0.45	0.55	0.19	0.17	0.19	0.18	0.17	0.22
Shimmer (%)	1.89	1.80	1.91	1.70	1.62	2.65	0.85	0.54	1.03	0.93	0.78	1.02
HNR (dB)	19.93	19.90	20.72	19.18	16.40	19.34	23.86	23.86	24.93	23.79	23.09	23.82
HSS	-9.87	-10.21	-8.56	-8.88	-9.08	-8.60	-11.51	-11.72	-10.47	-12.64	-12.81	-11.78
CVA	5.43	4.38	7.19	4.68	4.31	12.61	2.99	2.92	3.17	4.94	4.42	5.10
CVF	0.95	0.74	1.22	1.28	0.77	2.81	0.66	0.52	0.77	0.67	0.61	0.81
Amp Mod (%)	5.90	5.80	6.20	5.55	5.10	7.05	2.80	2.10	4.50	2.70	2.70	3.70
Freq Mod (%)	1.90	1.30	2.00	1.20	0.85	2.35	0.50	0.40	0.60	0.70	0.60	0.80
EGGW25	.6504	.6264	.6637	.7044	.6835	.7044	.6917	.6899	.6917	.6468	.6412	.6960
EGGW50	.5384	.5128	.5522	.6012	.5787	.6139	.5830	.5653	.5923	.5433	.5431	.5866
EGGW75	.4042	.3923	.4147	.4462	.4253	.4774	.4484	.4342	.4622	.4268	.4203	.4623
Intensity (dB)	70.82	70.39	70.95	72.06	70.39	73.11	77.35	74.90	77.56	77.56	76.80	78.60

Statistical Analysis

Observed medians and upper and lower hinges are presented for all pre-treatment and post-treatment measurements. Wilcoxon ranked-sum tests are used to compare the medians of the pre-treatment and post-treatment variables. Wilcoxon signed-ranked tests are performed to determine change versus no change. A statistical significance of $p < .05$ was used to determine significance.

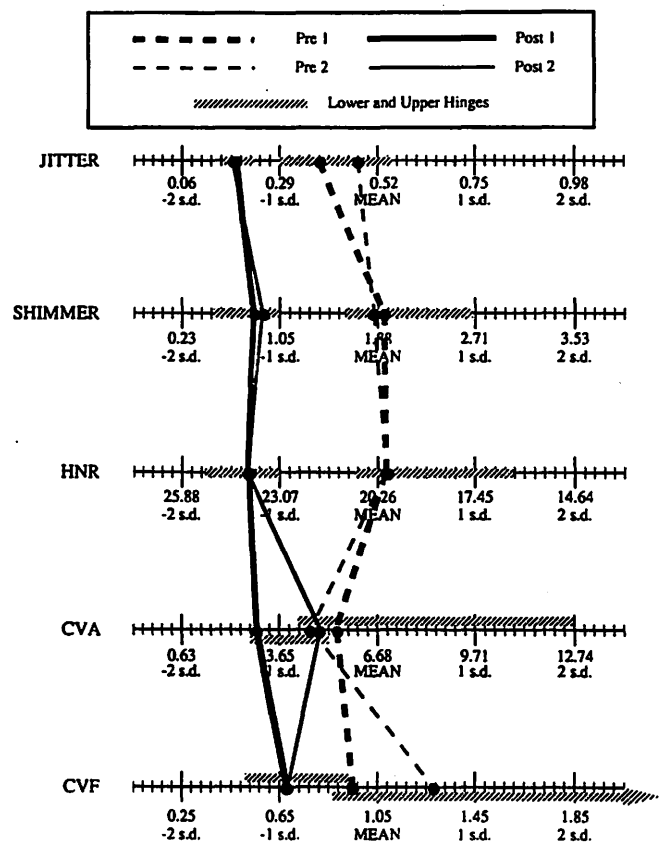
Twenty percent of the data for each variable was re-analyzed (interexaminer) for reliability. All repeated measures ranged between 0.88 and 0.99 using Pearson correlation coefficients.

Results

Acoustic Variables

The median, lower hinge and upper hinge of all acoustic variables analyzed are presented in Table 2. When the two pre-treatment conditions were not significantly different and the two post-treatment conditions are not significantly different they were combined to determine percent change pre- to post-treatment. Figure 1 is a graph of the medians, lower hinges and upper hinges for both pre-treatment and post-treatment perturbation measures.

Figure 1: Perturbation Measures



Maximum duration of sustained vowel phonation (seconds)

The median maximum duration of sustained vowel phonation for the two pre-treatment data collection sessions were 21.3 (lower hinge = 20.45, upper hinge = 21.45) seconds and 21.8 (21.75, 22.45) seconds and the two post-treatment data-collection sessions were 25.0 (24.1, 27.5) seconds and 26.4 (24.35, 26.85) seconds This pre- to post-treatment difference was statistically significant. The median pre- to post-treatment change was 19.26 percent. Daily clinical measures of maximum duration of sustained vowel phonation supported the experimental data collection measures and were statistically significant as well pre- to post-treatment. The subject evidenced a significant pre- to post-treatment median percent change of 31.43 percent for daily clinical measures.

Maximum fundamental frequency range (ST)

The maximum fundamental frequency range (semitones) for the first pre-treatment data collection session based on the highest high frequency and the lowest low frequency was 28.6469 ST. A valid ST range for the second pre-treatment recording was unable to be obtained due to the amount of vocal fry in the patient's low phonation. The maximum fundamental frequency range for the post-treatment data collection session (range was only collected in one of the two data collection sessions) based on the highest high and the lowest low was 33.2365 ST. This pre- to post-treatment difference was not statistically significant. Daily clinical variables for ST range supported these findings.

Fundamental frequency variability (Hertz standard deviation)

The subject's fundamental frequency standard deviation during reading was 22.41 Hz pre-treatment and 22.39 Hz post-treatment. This pre- to post-treatment difference was not statistically significant. This analysis was only completed on one data collection session pre-treatment and post-

significant. This analysis was only completed on one data collection session pre-treatment and post-treatment.

Intensity

The change of intensity level of sustained vowel phonation was found to be significant pre-treatment to post-treatment. The median values of the pre-treatment conditions were 70.82 (70.39, 70.95) dB and 72.06 (70.39, 73.11) dB and for the post-treatment conditions were 77.35 (74.90, 77.56) dB and 77.56 (76.80, 78.60) dB. This pre- to post-treatment difference was statistically significant. The median pre- to post-treatment change was 8.42 percent.

Perturbation

Jitter. The median value for the pre-treatment conditions were 0.38 (0.29, 0.48) percent and 0.47 (0.45, 0.55) percent. The median values for post-treatment conditions were 0.19 (0.17, 0.19) percent and 0.18 (0.17, 0.22) percent. This pre- to post-treatment difference was statistically significant. The median pre- to post-treatment change was 56.48 percent.

Shimmer. The median value for the pre-treatment conditions were 1.89 (1.80, 1.91) percent and 1.70 (1.62, 2.65) percent and for the post-treatment conditions were 0.85 (0.54, 1.03) percent and 0.93 (0.78, 1.02) percent. This pre- to post-treatment difference was statistically significant. The median pre- to post-treatment change was 50.42 percent.

CVA. The median values for the pre-treatment conditions were 5.43 (4.38, 7.19) and 4.68 (4.31, 12.61) and for the post-treatment conditions were 2.99 (2.92, 3.17) and 4.94 (4.42, 5.10). The pre- to post-treatment difference was statistically significant for pre 1 compared to post 1. The median pre- to post-treatment change was 44.94 percent for the first condition. Differences were not statistically significant for the second pre- to post-treatment conditions.

CVF. The median values for the pre-treatment conditions were 0.95 (0.74, 1.22) and 1.28 (0.77, 2.81) and for the post-treatment were 0.66 (0.52, 0.77) and 0.67 (0.61, 0.81). This pre- to post-treatment difference was statistically significant. The median pre- to post-treatment change was 40.36 percent.

Amplitude Modulation. The median values for the amplitude of amplitude modulation pre-treatment were 5.90 (5.80, 6.20) percent and 5.55 (5.10, 7.05) percent and for the post-treatment conditions the values were 2.80 (2.10, 4.50) percent and 2.70 (2.70, 3.70) percent. This pre- to post-treatment difference was statistically significant. The median pre to post-treatment change was 51.97 percent.

Frequency modulation. The median values for the amplitude of frequency modulation pre-treatment were 1.90 (1.30, 2.00) percent and 1.20 (0.85, 2.35) percent and for the post-treatment condition values were 0.50 (0.40, 0.60) percent and 0.70 (0.60, 0.80) percent. This pre- to post-treatment difference was statistically significant. The median pre- to post-treatment change was 61.29 percent.

HNR. The median values for the pre-treatment conditions were 19.93 (19.90, 20.72) dB and 19.18 (16.40, 19.34) dB and for post-treatment 23.86 (23.86, 24.93) dB and 23.79 (23.09, 23.82) dB. This pre- to post-treatment difference was statistically significant. The median pre- to post-treatment change 21.83 percent.

Spectral Slope. The median values for the pre-treatment conditions were -9.87 (-10.21, -8.56) and -8.88 (-9.08, -8.60) and the post-treatment were -11.51 (-11.72, -10.47) and -12.64 (-12.81, -11.78). This pre- to post-treatment difference was statistically significant. The median pre- to post-treatment change was 28.8 percent.

Parameterization

EGGW 25. The median values for the pre-treatment conditions were 0.6504 (.6264, .6637) and 0.7044 (.6835, .7044) and for the post-treatment were 0.6917 (.6899, .6917) and 0.6468 (.6412, .6960). This pre- to post-treatment difference was not statistically significant.

EGGW 50. The median values for the pre-treatment conditions were 0.5384 (.5128, .5522) and 0.6012 (.5787, .6139) and for the post-treatment were 0.5830 (.5653, .5923) and 0.5433 (.5431, .5866). This pre- to post-treatment difference was not statistically significant.

EGGW 75. The median values for the pre-treatment conditions were 0.4042 (.3923, .4147) and 0.4462 (.4253, .4774) and for the post-treatment were 0.4484 (.4342, .4622) and 0.4268 (.4203, .4623). This pre- to post-treatment difference was not statistically significant.

Perceptual Ratings. For the variable -loud enough, the subject self-rated an increase in loudness of 20 to 60 percent across the two pre- and the two post-treatment data collection sessions. For the variable shaky voice, the subject self-rated a 7 to 11 percent reduction in shakiness in her voice. For the variable mumbling, the subject rated a 15 to 38 percent reduction in mumbling. For the variable intelligibility, the subject rated a 12 to 41 percent improvement in her overall intelligibility.

Discussion

The data presented in this study support the hypothesis that intensive voice therapy can have a positive effect on speech deficits associated with bilateral thalamotomy surgery in a patient with Parkinson's disease. Statistically significant differences were measured between pre- and post-treatment on the variables maximum duration of sustained vowel phonation (both in data collection and daily clinical measures), jitter, shimmer, HNR, HSS, CVF, intensity, amplitude modulation, and frequency modulation. It is interesting to note that the subject's CVA measure was not statistically significant although the amplitude modulation measure was significant. However, if the CVA measures are examined closely, the measure on her first post-treatment recording does improve although the measure on the second recording does not indicate improvement. Although the measurement itself does not improve on the second voice recording, much less variability is documented in her post-treatment CVA measure. As demonstrated in Figure 1, these data suggest movement towards greater phonatory stability, both long-term and short-term (Scherer, Gould, Titze, Meyers, & Sataloff, 1988). In pathophysiological terms, increased vocal fold adduction and respiratory support contributed to increased phonatory stability and intensity.

Although this subject demonstrated improvement, this type of speech deficit (thalamotomy induced) responded less enthusiastically to the intensive voice therapy program than the typical idiopathic Parkinson's disease speech deficit. It is pertinent to address the fact that the deficit was similar to pseudobulbar speech dysarthria associated with basal ganglia dysfunction. However, inappropriate crying and laughing were not observed in this patient. Although they are not always present (Aronson, 1985), these symptoms are important indicators of pseudobulbar dysarthria. Another important distinction between pseudobulbar dysarthria and hypokinetic dysarthria is a strained and/or strangled voice quality. This quality was observed in this patient, although not consistently. It is generally accepted that a small percentage of Parkinson's disease patients demonstrate hyperfunction of the larynx. Therefore, it is reasonable to conclude that this subject demonstrated both hypokinetic dysarthria and pseudobulbar dysarthria. Nevertheless, the therapy was successful for this subject.

Another area that requires further research is speech deficits created by unilateral and bilateral basal ganglia surgery in individuals with Parkinson's disease as well as other movement disorders and the effects of intensive voice therapy on these deficits. The current study clearly documents deficits that were a direct result of her surgery. However, due to the fact the patient entered this current research program two years post-surgery, it is difficult to determine what percentage of the deficits were created by the surgery and what percentage are a result of the natural progression of her Parkinson's disease.

Unlike Ramig et al.'s study (1988), speech and voice deficits that were not directly addressed during therapy did not improve. For example, in Ramig' s study, the subjects' demonstrated an increase in their fundamental frequency standard deviation range during reading while the subject in this study did not. This may be a result of the pseudobulbar-like speech characteristics in this subject. If her speech deficits were not from bradykinesia, the theory of increased effort enhancing speech/ voice production may not apply to this disorder; or an even greater amount of effort is needed to not only over-ride bradykinesia but the actual brain damage as well. However, due to the fact the subject did improve, it could be concluded that physiological changes of the respiratory and phonatory system were accomplished whether by increased effort or another reason, such as stimulating the formation of new neural pathways for speech production.

It is recognized that basal ganglia, thalamic, cerebellar, and cortical systems contribute to the motor apparatus mediating vocalization and articulation (Riklan, Levita, Zimmerman, Cooper 1969). Therefore, it can be hypothesized that lesions in these areas will create speech deficits. Further research is needed in this area to determine how the deficits relate to certain neurological illness and the speech/voice deficits created by the illness. Finally, further research and documentation is needed to determine how the deficits created by the lesioning respond to speech and voice therapy.

Acknowledgements

Deep appreciation is extended to the patient who participated in this study for her contributions and continuous challenges. The authors gratefully acknowledge the contributions of Dr. Larry Brown, William Winholtz, Dr. Chris O' Brien, Dr. Marshall Smith, Dr. Patricia Thompson, Mary Serkowski, Jeanne Delaney, Patricia Benjamin, Wendy Savoy, Sharon MacClay, John Girardeau, and Susan Hensley.

The research presented here was supported in part by OE-NIDRR grant #H133G00079, NIH grant R01 DC01150-01, and Center Grant University of Iowa NIH#1 P60 DC00976.

References

Allan, C. M. (1970). Treatment of non-fluent speech resulting from neurological disease: Treatment of dysarthria. British Journal of Disordered Communication, 5, 3-5.

Allen, C. M., Turner, J. W., & Gadea-Ciria, M. (1966). Investigations into speech disturbances following stereotaxic surgery for Parkinsonism. British Journal of Speech Communication Disorders, 1, 55-59.

Aronson, A. (1985). Clinical Voice Disorders. New York: Thieme-Stratton.

Bell, D. S. (1968). Speech functions of the thalamus inferred from the effects of thalamotomy. Brain, 91, 619-636.

Boeckstyns, M. E., & Backer, M. (1989). Reliability and validity of the evaluation of pain in patients with total knee replacement. Pain, 38(1), 29-33.

Canter, G. J. (1965). Speech characteristics of patients with Parkinson's disease:-II. Physiology support for speech. Journal of Speech and Hearing Disorders, 30, 44-49.

Critchley, E. M. R. (1981). Speech disorders of Parkinsonism: a review. Journal of Neurology, Neurosurgery and Psychiatry, 44, 751-758.

Darley, F. L., Aronson, A., & Brown, J. (1969a). Differential patterns of dysarthria. Journal of Speech and Hearing Research, 12, 246-249.

Darley, F. L., Aronson, A., & Brown, J. (1969b). Clusters of deviant speech dimensions in the dysarthrias. Journal of Speech and Hearing Research, 12, 462-496.

Darley, F.L., Brown, J. R., & Swenson, W. M. (1975). Language changes after neurosurgery for Parkinsonism. Brain and Language, 2, 65-69.

Fairbanks, G. (1960). Voice and articulation drill book. New York:Harper and Brothers.

Green, H. C. (1980). The Voice and Its Disorders. London: Pitman Medical.

Hansen, D. G., Gerratt, B. R., & Ward, P. H. (1984). Cinegraphic observations of laryngeal func tion in Parkinson's disease. Laryngoscope, 94, 348-353.

Hoberman, S. G. (1958). Speech techniques in aphasia and Parkinsonism. Journal of Michigan State Medical Society, 57, 1720-1723.

Hoehn, M., & Yahr, M. (1967). Parkinsonism: onset, progression and mortality. Neurology, 17, 427.

Hollien, H., Dew, D., & Philips, P. (1971). Phonational frequency ranges of adults. Journal of Speech and Hearing Research, 14, 755-760.

Kempster, G. B. (1984). A multidimensional analysis of vocal quality in two dysphonic groups. An unpublished doctoral dissertation. Northwestern University, Evanston.

Kent, R. D., Kent, J. F., & Rosenbek, J. C. (1987). Maximum performance tests of speech production. Journal of Speech and Hearing Disorders, 52, 367-387.

Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and occurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson's patients. Journal of Speech and Hearing Disorders, 42, 47-57.

Manen, J., van Speelman, J. D., & Tans, R. J. (1984). Indications for surgical treatment of Parkinson's disease after levodopa therapy. Clinical Neurology and Neurosurgery, 86, 207-212.

Oxtoby, M., (1981). Parkinson's Disease Patients and Their Social Needs. London: Parkinson's Disease Society.

Ramig, L. A., Mead, C., & Horii, Y. (1988a). Voice therapy and Parkinson's disease. Journal of the American Speech and Hearing Association, 30(10), 128.

Ramig, L. O., Mead, C., Scherer, R., Horii, Y., Larson, K., & Koehler, D. (1988, February). Voice therapy and Parkinson's disease: a longitudinal study of efficacy. A paper presented at the Clinical Dysarthria Conference, San Diego, CA.

Ramig, L.O, Horii, Y., & Bonitati, C. M. (1991). The efficacy of voice therapy for patients with Parkinson's disease. National Center for Voice and Speech Status and Progress Report, 1, 61-86.

Reich, A. R., Frederickson, R. R., Mason, J. A., & Schlauch, R. S. (1990). Methodological variables affecting phonation frequency range in adults. Journal of Speech and Hearing Disor ders, 55, 124-131.

Riklan, M., Levita, E., Zimmerman, J., & Cooper, I. S. (1969). Thalamic correlates of language and speech. Journal of Neurological Science, 8, 307-328.

Rubow, R., & Strand, E. (1985). A microcomputer-based wearable biofeedback device to improve transfer of treatment in Parkinsonian dysarthria. Journal of Speech and Hearing Disorders, 50, 178-185.

Samra, K., Riklan, M., Levita, E., Zimmerman, J., Waltz, J. M., Bergmann, L., & Cooper, I. S. (1969). Language and speech correlates of anatomically verified lesions on thalamic surgery for Parkinsonism. Journal of Speech and Hearing Research, 12, 510-540.

Sarno, M. T. (1968). Speech impairment in Parkinson's disease. Journal of Speech and Hearing Disorders, 49, 269-275.

Scherer, R. C., Gould, W. J., Titze, I. R., Meyers, A. D., & Sataloff, R. T. (1988). Preliminary evaluation of selected acoustic and glottographic measures of clinical phonatory function analysis. Journal of Voice, 2(3), 230-244.

Schiffman, S., Reynolds, M. L., & Young, F. W. (1984). Introduction to Multidimensional Scaling: Theory, Methods and Applications. New York: Academic Press.

Selby, G. (1967). Stereotactic surgery for the relief of Parkinson's disease: 2. An analysis of the results in a series of 303 patients (413 operations). Journal of Neurological Science, 5, 343-375.

Streifler, M., & Hofman, S. (1984). Disorders of verbal expression in Parkinsonism. In R.G. Hassler and J.F. Christ (Eds.). Advances in Neurology, 40, 385-393.

Tasker, R., Organ, L. W., & Hawrylyshyn, P. A. (1982). The Thalamus and Midbrain of Man. Springfield, IL: Thomas.

Titze, I. (1984). Parameterization of the glottal area, glottal flow and vocal fold contact area. Journal of the Acoustical Society of America, 75, 570-580.

Weiner, W. J., & Singer, C. (1989). Parkinson's disease and nonpharmacological treatment programs. Journal of the American Geriatric Society, 37, 359-363.

Winholtz, W., & Ramig, L. O. (1992). Vocal tremor analysis with the Vocal Demodulator. Journal of Speech and Hearing Research, 35(3), 562-573.

NCVS Status and Progress Report - 3
December 1992, 149-155

Preliminary Observations on the Diffusion of Botulinum Toxin from the Site of Injection in Laryngeal Muscles

K. Linnea Peterson, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
National Center for Voice and Speech
Erich S. Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
National Center for Voice and Speech
Mark T. Madsen, Ph.D.
Department of Radiology, Division of Nuclear Medicine, The University of Iowa

Introduction

Botulinum neurotoxin A is a very potent neurotoxin which causes a flaccid paralysis upon ingestion or injection. Although it has been a feared food poisoning for years, clinical uses in medicine have been developed and refined so that localized injection is now an accepted form of treatment of various neuromuscular disorders. The clinical use of botulinum toxin was pioneered by Scott[1], who developed it as a treatment modality for strabismus and blepharospasm. It has been used in ophthalmology since about 1980. The implications for otolaryngology and neurology have also been noted and it has since been used for such disorders as hemifacial spasm, torticollis, and spasmodic dysphonia[2,3,4,5,6,7]. Despite the known systemic effects from ingestion of the toxin, very few systemic effects from local injection have been reported. Ingestion is known to cause symptoms which suggest CNS involvement (e.g. dizziness, lethargy, general locomotor dysfunction), however the toxin has not been shown unequivocally to cross the blood brain barrier[8]. Antibody production has been noted in patients who received the larger doses (200-300 Units) used in the treatment of torticollis[3]. In addition, a patient with parkinsonism was noted in one study[3] to develop fatigue after two series of injections (81.7 total units). Those studies that have explored the systemic effects physiologically have involved single fiber EMG studies at sites distant from the injection[9,10]. These studies have shown increased "jitter" in distant muscles, but no clinically apparent effects. A recent study

looking at spread of the toxin after intramuscular injection looked at whether or not the toxin crossed fascial planes[11]. The results showed that botulinum toxin passed through muscle fascia even at subclinical doses, with fascia reducing the spread of toxin by only 23%.

It has seemed possible to us that a significant fraction of the botulinum toxin injected into a muscle for treatment of muscle spasm is picked up in the blood circulation at the injection site before it is bound to cholinergic nerve terminals. Whether or not this occurs must depend upon the ability of toxin molecules, in areas of high concentration, to enter the muscle's microcirculation. The purpose of our study was to determine whether this occurs in laryngeal muscles, and if so, to estimate the degree and time course of this path of diffusion from the site of injection. To do this, we have injected radiolabelled botulinum toxin into the cricothyroid muscles of anesthetized dogs and quantitatively measured the amount of radioactivity in this region at frequent intervals. If toxin molecules do not enter the systemic circulation, then the amount of radioactivity in the area of the larynx should not change substantially over time. On the other hand, if they do enter the circulation, we would expect the amount of radioactivity to decrease substantially over the first few hours of observation.

Materials and Methods

Radiolabelling

Botulinum Neurotoxin A (Sigma Chemical Company, St. Louis) was labelled with ^{131}I (New England Nuclear) according to the Chloramine T method as described by Williams[12], with reference to Greenwood, et al.[13] ^{131}Iodine was chosen due to its energy level, which allows both *in vivo* imaging with a standard gamma camera as well as scintillation counting *in vitro*. It has a half life of 8.5 days, and a high energy decay which contributes to protein degradation.

Once the toxin had been labelled, it was important to verify what has been labelled. Since radiolabelling could result in breakdown of the protein, it was critical to know what was being measured in order to be able to draw valid conclusions. Thus a purification step was used to eliminate components of varying activity, including free iodine. We used SDS-PAGE as described by Tse et al.[14], in order to evaluate the labelling efficacy.

Experimental Procedures

Three mongrel dogs, obtained from the Animal Care Unit at The University of Iowa, were initially anesthetized with an intramuscular injection of ketamine and zylazine. The femoral vein was then cannulated for infusion of supplemental anesthetic and normal saline, and for withdrawal of blood samples. Nembutal was given in a dose of approximately 25 mg/kg, or as necessary to produce and maintain deep surgical anesthesia, and then the animal's head was placed in a head-holding frame in a supine position. A ventral incision in the neck region over the larynx was used, with minimal dissection, to expose the cricothyroid muscles. During the course of the experiment, the depth of the anesthesia was checked frequently and supplemental nembutal given as needed.

Serial images of the radiolabelled botulinum toxin were acquired by a gamma camera interfaced to a computer. After the gamma camera was centered over at the larynx of the dog, at a distance of about 25 cm, 10 C (experiment 1) or 5 C (experiment 2 and 3) of purified radiolabelled botulinum toxin was injected into the cricothyroid muscle on each side. The radiolabelled toxin had been diluted to obtain a solution of approximately 10 C/0.1 ml. The actual injection required about 15 seconds. The syringe containing the labelled toxin was immediately withdrawal from the field of the camera, and records of the radioactivity were immediately started. Scanning took place at time

intervals of every five seconds for the first minute, then approximately every minute for the next five minutes, approximately every five minutes out to thirty minutes, then once every ten minutes. The clearance of the radioactivity was monitored from the counts in each computer image in a region of interest covering the injected muscle (Figure 1). The collimation used with this camera was not very precise, but was sufficient to ensure that radiolabelled material located in the animal's body would not register in the region that was analyzed.

In addition to imaging the site of injection, serum radioactivity was measured by drawing one milliliter blood samples at various intervals corresponding to the imaging protocol (except during the first minute following injection). Once the blood samples were all collected for a given experiment, they were analyzed by running them through a scintillation counter, using a ten minute counting period.

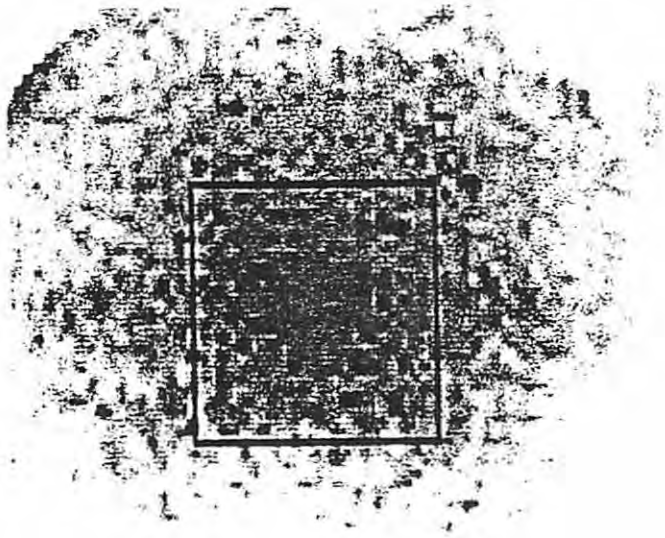


Figure 1. Typical computer-generated image of the region covered by the gamma camera (experiment 2). The computer provided total counts in the area defined by the square. Two slightly brighter "peaks" can be seen in the middle of the bright area; they correspond to the two cricothyroid muscles.

Additional Studies of the Radiolabelled Botulinum Toxin

A Lowry protein assay was done to correlate the radioactivity of the labelled toxin with its protein concentration. Also, an LD-50 (amount of toxin which would kill 50 percent of a group of mice injected with this same dose, within four days) was conducted using Swiss-Webster mice as described by Williams, et. al., (12) with reference to Reed and Muench (15). This procedure was necessary to correlate the biological activity of the labelled toxin with its specific radioactivity.

Results

Results of the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions showed two bands consistent with the results of other investigators[14], one at 97 kilo Daltons, and one at 55 kilo Daltons. These two bands are consistently labelled with the ^{131}I

(Figure 2). In addition there are two bands of much smaller molecular weight, 29 kilo Daltons and 34 kilo Daltons, which are also labelled.

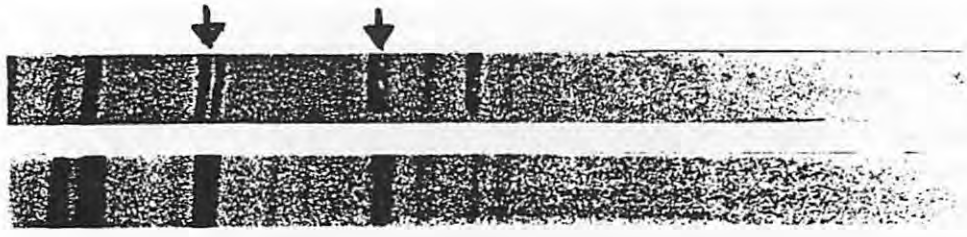


Figure 2. SDS-PAGE gel of radiolabelled botulinum toxin. The stained gel (top) shows prominent bands corresponding to 55 kilo Daltons (left arrow) and 97 kilo Daltons (right arrow). These bands are radioactive, as shown on the film exposure of the gel (below). Two bands corresponding to lighter-weight fractions, are also radioactive.

The Lowry protein assay of each batch of radiolabelled toxin showed that the amount of toxin recovered from the labelling procedure was quite variable. The protein concentration of the labelled product ranged from 0.02 mg/ml to 0.09 mg/ml, as compared with the unlabelled toxin concentration of 1.0 mg/ml to 1.2 mg/ml. Obviously some of the toxin was lost in the labelling and purification steps, with dilution all playing a role. This observation draws attention to the need for assay of all batches of labelled toxin.

The results of the experiment to establish the biological activity of the labelled toxin were very disappointing. Only the first batch of labelled material resulted in death of the mice injected with higher doses. In all the subsequent experiments, none of the mice died even though the doses given were large enough to compensate for the fact that the protein concentrations of these latter batches was lower than the first batch. Nevertheless, the first batch did exhibit biological activity, which represented about 300 ng per mouse Unit.

The time course of the radioactivity in the neck region of the three dogs during the first two hours after injection is shown in Figure 3. The results are reasonably consistent in showing a rapid initial decay followed by a much slower loss of radioactivity. The most rapid decay is within the first twenty minutes for each trial. The higher levels of radioactivity for the first experiment are a result of using twice the amount of radioactivity (20 C) as compared with subsequent trials (10 C). Despite the different scale, the pattern is the same. Expressed as a percentage, the loss of radioactivity was about 25%, 46% and 43% in experiments 1-3 respectively.

None of the blood samples collected from the dogs were found to contain significant levels of radioactivity. Radiation was detected from control samples of diluted labelled toxin, however, suggesting there was nothing fundamentally wrong with our procedure. Calculations taking into account the sensitivity of the scintillation counter and dilution within the blood volume of the dog suggest that our procedure should have been able to detect the diffusion of less than 1% of our injected dose. Such a calculation makes the assumption, however, that the toxin lost from the injection site remains in the blood. Our failure to detect what appears to be much larger losses of radioactivity from the injection site may reflect the fact that toxin molecules entering the blood system at the injection site quickly leave the circulation in areas of lower toxin concentration, i.e. other muscles in the body, and become bound to receptors.

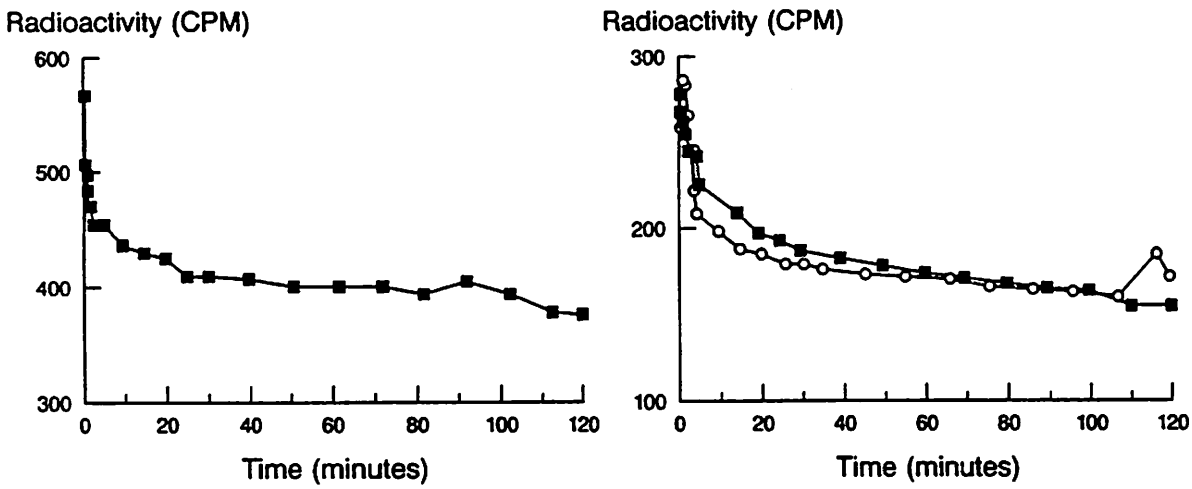


Figure 3. Change in the radioactivity in the neck region of the dogs over two hours. Results of experiment 1 are shown on the left, and experiments 2 (solid squares) and experiment 3 (open circles) on the right.

Discussion

Our results suggest, however tentatively, that a substantial fraction of a dose of botulinum toxin injected into a well-perfused muscle disappears from that site within the first two hours. This result is not too surprising if one supposes that the toxin molecule can cross the epithelium of muscle capillaries, which it must be able to do if ingestion of the toxin is able to paralyze muscles throughout the body. Yet our experiment is seriously flawed in several respects, so even this "common sense" interpretation must be regarded with suspicion.

Determining the systemic distribution of botulinum toxin emerges as a more difficult problem than we had initially supposed it to be. Part of the problem is the toxin itself; it is very potent, having a median lethal dose of 5-50 ng/kg body weight[16], thus making detection very difficult. There are various labelling techniques that can be considered, each with advantages and disadvantages. We chose to use ^{131}I for several reasons. Radioiodination of botulinum toxin has previously been described in detail[12], with good preservation of biological activity. Previous descriptions of the technique have described the use of ^{125}I. ^{131}Iodine, however, has a higher energy level than ^{125}I, making dynamic detection more feasible, and standard equipment used for clinical imaging of radioisotopes is scaled to the energy levels of ^{131}I. ^{131}I has drawbacks of a shorter half-life and greater risk of protein degradation, however.

The SDS-PAGE gel appears to be a necessary check on labelling procedure since free label present in the injected solution would obviously give false results. Our gels showed the label attached to heavy fragments of the toxin, but smaller fragments were also labelled. This fact needs to be kept in mind when interpreting our findings.

Our experience also indicates that it is necessary to determine the biological activity of the labelled material. Inactive toxin may fail to bind to receptors locally, and thus falsely suggest diffusion from the injection site. Unless the biological activity is known with some precision the results will never be suitable for extrapolating back to the doses that are typical of clinical use.

In presenting these very preliminary results we hope to draw attention to what we regard as an important question, and to address a few of the problems that one may encounter in trying to use radiolabelling methods to answer the question. It is important to refine a labelling technique for ^{131}I until it provides a reasonably high and consistent concentration of protein, has the label attached to the known heavy fragments of botulinum toxin, and has biological activity suggesting that most of the labelled toxin molecules are intact. These factors must be controlled in order to draw valid conclusions.

Acknowledgements

This research was supported by grant P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders.

References

1. A.B. Scott, Botulinum toxin injection of eye muscles to correct strabismus, *Trans. Am. Ophthalmol. Soc.* 79:734-770 (1981).
2. C.L. Ludlow, R.F. Naunton, S.E. Sedory, G.M. Schultz, and M. Hallett, Effects of botulinum toxin injections on speech in adductor spasmodic dysphonia, *Neurology* 38:1220-1225 (1988).
3. M.F. Brin, S. Fahn, C. Moskowitz, A. Friedman, H.M. Shale, P.E. Greene, A. Blitzer, T. List, D. Lange, R.E. Lovelace, and D. McMahon, Localized Injections of botulinum toxin for the treatment of focal dystonia and hemifacial spasm, *Movement Disorders* 2:237-254 (1987).
4. A. Blitzer, M.F. Brin, P.E. Greene, and S. Fahn, Botulinum toxin injection for the treatment of oromandibular dystonia, *Ann. of Otol., Rhin. and Laryngol* 98: 93-97 (1989).
5. J. Jankovic and J. Orman, Botulinum A toxin for cranial-cervical dystonia: A double-blind, placebo-controlled study, *Neurology* 37:616-623 (1987).
6. C.L. Ludlow, R.F. Naunton, S. Terada, and B.J. Anderson, Successful treatment of selected cases of abductor spasmodic dysphonia using botulinum toxin injection, *Otolaryngol. Head Neck Surg.* 104: 849-855 (1991).
7. C.L. Ludlow, M. Hallett, S.E. Sedory, M. Fujita, and R.F. Naunton, The pathophysiology of spasmodic dysphonia and its modification by botulinum toxin, *in*: "Motor Disturbances II," A. Berardelli, R. Benecke, M. Manfredi, and C.D. Marsden, eds., Academic Press Limited, London (1989).
8. D.A. Boroff, and G.S. Chen, On the question of permeability of the blood-brain barrier to botulinum toxin, *Int. Archs. Allergy Appl. Immun.* 48:495-504 (1975).
9. H. Lundh, H.H. Schiller, and D. Elmqvist, Correlation between single fibre jitter and endplate potentials studied in mild experimental botulinum poisoning, *Acta Neurol. Scandinav.* 56:141-152 (1977).
10. D.B. Sanders, E.W. Massey, aand E.G. Buckley, Botulinum toxin for blepharospasm: Single-fiber EMG studies, *Neurology* 36:545-547 (1986).
11. C.M. Shaari, E. George, B-L Wu, H.F. Biller, and I. Sanders, Quantifying the spread of botulinum toxin through muscle fascia, *Laryngoscope* 101: 960-964 (1991).

12. R.S. Williams, C-K Tse, J.O. Dolly, P. Hambleton, and J. Melling, Radioiodination of botulinum neurotoxin type A with retention of biological activity and its binding to brain synaptosomes, *Eur. J. Biochem.* 131:437-445 (1983).
13. F.C. Greenwood, W.M. Hunter, and J.S. Glover, The preparation of ^{131}I-labeled human growth hormone of high specific radioactivity, *Biochem J.* 89:114-123 (1963).
14. C-K Tse, J.O. Dolly, P. Hambleton D. Wray, and J. Melling, Preparation and characterisation of homogenous neurotoxin type A from Clostridium botulinum, *Eur. J. Biochem.* 122:493-500(1982).
15. L.J. Reed and H. Muench, A simple method of estimating fifty percent endpoints, *Amer. J. Hygiene.* 27:493-497 (1938).
16. L.C. Sellin, The action of botulinum toxin at the neuromuscular junction, *Med. Biol.* 59:11-20 (1981).

NCVS Status and Progress Report - 3
December 1992, 157-173

Non-specific Laryngeal Granuloma: A Literature Review and a Case Study involving a Professional Singer

Katherine Verdolini-Marston, Ph.D.,
Department of Speech Pathology and Audiology, The University of Iowa
Henry Hoffman, M.D.
Department of Otolaryngology, The University of Iowa
Scott McCoy, D.M.A.
Department of Music, The University of Iowa

Abstract

This paper first provides a literature review on functional aspects of non-specific laryngeal granulomas. According to some authors, with these lesions voice may be abnormal and the functional response to surgery may be poor. However, well-controlled quantitative data are lacking. The second part of the paper is a case report on a professional singer with granulomas, who underwent surgery following failed response to conservative treatments. Two primary findings emerged. First, although a sizeable lesion was present initially, pre-surgical voice measures were largely normal or superior. Second, post-surgical voice functions were entirely normal or superior. Further, the patient successfully resumed professional singing within a few weeks following surgery, without recurrence. Thus, the findings are inconsistent with previous descriptions of obvious voice abnormalities in patients with granulomas, and of poor functional response to surgery.

Laryngeal granulomas are a set of benign pathologies including proliferative granulations of the vocal processes of the arytenoids, unilaterally or bilaterally, that may be considered along a continuum with simpler ulcerative lesions or contact ulcers. Granulomas can arise from intubation trauma (1-3), or specific disease processes including tuberculosis, Wegener's disease, syphilis (4-6), or even AIDS (7). Probably the more commonly encountered clinical case, however, involves non-specific lesions characterized by arytenoid inflammation and accumulation of grainy, vascular, and

sometimes epithelialized tissue. Pathological reports noted by Miko (8) described necrosis involving the vocal process in some cases, mounted by new capillary formations basally, collagen fibers, radially organized fibroblasts, and fibrinous exudate with tissue debris. In the same reports, neutrophils were pervasive and the lesions were sometimes capped by pseudoepitheliomatous hyperplasia.

The goal of clinical management is sometimes to rule out neoplastic involvement and, in the case of exceptionally prominent lesions, to ensure airway patency. However, more typically the goal is symptomatic relief, where symptoms include hoarseness, local discomfort or pain, the impression of a foreign body, secondary iterative coughing and throat clearing, referred otalgia (6,8-12), and in rare cases, stridor and dyspnea (9).

Based on etiologic hypotheses, conservative treatment approaches may include anti-reflux and/or voice therapy. In the case of failed response to conservative approaches or for the purpose of biopsy or airway maintenance, laryngeal microsurgery may be undertaken (3).

Despite a general consensus about the appearance of characteristic symptoms and about appropriate treatment approaches, little quantitative information is available about the actual symptoms and signs associated with non-specific granulomas, and about the functional outcomes of different treatment approaches. This information is important for patients with granulomas in general. However, the information is especially important in cases involving professional voice users, who in particular need to know the possible effect of treatment on voice and anticipated time frames for functional recovery. The purpose of this report is (a) to review existing literature on the effects of non-specific laryngeal granulomas on voice and on the functional outcome of various treatments, to point out what is already known and to emphasize that information is still needed, and (b) to quantitatively describe the voice of a professional singer with granuloma, as well as the functional outcome of laryngeal microsurgery following failed response to conservative measures.

Review of the Literature

Different treatment approaches to laryngeal granuloma and rationale

As noted, clinical approaches to laryngeal granulomas may be surgical and/or non-surgical. Speech (voice) therapy is provided in some cases, based on the speculation that mechanical factors play an etiologic role. In fact, numerous authors propose a mechanical contribution of arytenoid trauma from coughing, throat clearing, and from a phonation mode that we describe as "pressed voice"[1] (6,10,14-20). Related, some authors also propose a mechanical contribution of chronic low pitch during phonation, because of a tendency for pronounced arytenoid rocking at low pitches [Baer, in discussion following Hirano et al. (21)] and potentially, tissue abrasion (20,22,23). Thus, voice therapy is typically aimed at behavioral changes in phonation mode (decreased arytenoid

[1] The term we use here, "pressed voice," is a subset of a more global concept described by most authors as "vocal abuse," which includes any behaviors assumed to result in laryngeal tissue injury. We oppose the term "vocal abuse" on several grounds, foremost of which is that it carries a moral connotation that may be detrimental to the therapeutic process. In the place of this term, we prefer descriptive language such as coughing, throat clearing, and "pressed voice." This approach has the advantage of avoiding moral connotations, while at the same time providing possible physiological descriptors that may be etiologically relevant. For example, "pressed voice" refers to the perception of a relative arytenoid pressing during phonation, generally confirmed on videoscopic examination. This term is particularly appealing because recent experimental evidence suggests that arytenoid "pressing" during phonation results in increased laryngeal contact stress in excised preparations (13).

"pressing" and increase in average speaking pitch) and in vegetative functions (decreased non-productive coughing and throat clearing). Increasingly, conservative treatment approaches involve anti-reflux therapy as an adjunct to or instead of voice therapy. The rationale is that esophageal reflux appears to be etiologically relevant in a preponderance of cases, with a spill-over of stomach acids onto the posterior glottis, resulting in primary ulceration (6,8,9,11,19,24-26).

Combined conservative treatment approaches may be provided, including both voice therapy and anti-reflux therapy, based on the assumption that mechanical and medical factors may interact in pathogenesis (6,19,27).

Finally, laryngeal microsurgery may be undertaken in the case of failed response to conservative measures, for functional reasons, for biopsy, or to improve the airway. Despite a large number of published reports indicating a variety of preferred surgical treatment methods (3,10,28), little useful data are available to direct the surgical approach to lesions refractory to medical management.

Cryotherapy and laser vaporization of vocal process granulomas have been used. However, the less traumatic approach with resection employing microlaryngeal forceps and scissors is preferred by many laryngologists. The noted response of laryngeal granulomas to systemic steroid administration and the high recurrence rate observed following surgical removal have prompted some surgeons to employ regional infiltration of steroids to the base of the granuloma as an adjunct to its excision.

Descriptions of voice characteristics associated with granulomas

A limited number of reports have provided important preliminary, quantitative information about the phonatory characteristics of laryngeal granulomas or contact ulcers. In one report by Hillman and colleagues (22), acoustic and aerodynamic characteristics of voice in patients with contact ulcers, and apparently also granulomas, were compared to the same characteristics in normal subjects for a simple speech task. Consistent findings for the patients with contact ulcers included low fundamental frequency, high estimated peak glottal flows, and high estimated AC glottal flows, as compared to norms. The results from this study are very important for a variety of reasons, including methodological ones. However, with specific reference to granulomas and contact ulcers, further quantitative information is needed because the number of participants was small, and limited aspects of voice functioning were assessed.

In another report by Hirano (29), a series of voice characteristics was described for a large number of patients with laryngeal granulomas (N = 40) as well as for patients with other voice-related diagnoses (total N = 1563). For patients with granulomas, data about maximum phonation time, mean flow rate, phonation quotient, habitual, lowest, and highest frequency and frequency range, and habitual, softest, and loudest intensity and intensity range were provided. However, specific comparisons to laryngeally healthy subjects were not included, and available North American norms cannot be applied with confidence for comparison purposes because the patient population was Japanese. Further, detailed information about measurement procedures was not provided in all cases, making cross-study comparisons difficult. Thus, although this report is vitally important in providing a substantial corpus of data about the voice characteristics of various pathological conditions and also for other reasons, voice profiles that distinguish patients with granulomas from laryngeally normal subjects did not necessarily emerge.

Information about functional outcomes of treatments

Limited information is available about voice characteristics following different treatment approaches to laryngeal granulomas, and about functional recovery rates. At least two reports imply favorable outcomes with behavioral approaches, and one of the same reports simultaneously implies an unfavorable functional outcome following surgery. For example, in a report by Ward and colleagues (6), an opera singer was described whose granuloma resolved five weeks following discontinued throat clearing, without speech therapy. The patient's voice symptoms (hoarseness) presumably improved as well within the same time frame. However, no information about post-treatment voice was actually provided. A second report about voice functioning following behavioral and in some cases adjunctive surgical treatment was provided by Peacher (30). This report, which involved a retrospective study of 70 patients with contact ulcers, indicated an average laryngeal healing time of 10.05 weeks (two-and-a-half months) in 36 patients who received voice therapy only, compared to a longer average healing time of 26.52 weeks (six-and-a-half months) in 34 patients who underwent surgery before or during voice therapy. Further, post-treatment symptoms, including hoarseness and vocal fatigue, appeared more prevalent in surgical than in non-surgical patients. Thus, the overall functional outcome appeared favorable with voice therapy only, and relatively poorer for surgical cases. Unfortunately, the study by Peacher presents a series of interpretative problems. Most importantly, "patient group" (surgical versus non-surgical) appears entirely confounded with pathological profile in this study. Surgery was apparently provided only in more severe (recalcitrant) cases. In fact, the author emphasized surgery only in the event of "massive ulcers in the presence of large granulomas" (p. 47). Thus, the study provides virtually no information about the effectiveness of behavioral or surgical treatments for voice, assuming proper experimental controls. Further, well-described, quantitative information about post-treatment voice functioning was not provided.

Other descriptions of surgical outcomes emphasize a tendency for post-treatment recurrences (9,10,12,31). In such cases, final or interim functional results may be expected to be poor. However, except for the problematic report by Peacher (30) already described, we have failed to find actual descriptions of voice following surgery. Presumably, if such information is available, it is rare.

Finally, a limited number of reports describing anti-reflux therapy have indicated a resolution of lesions as early as two weeks following initiation of treatment (8) and as late as one to four months (11) in successful cases. Again, however, these reports do not provide quantitative information about the level of post-treatment voice functioning.

In summary, some quantitative information is available about the voice characteristics associated with laryngeal granulomas (22,29). However, more information is needed both in terms of number of observations and relation of observations to normative data. Further, limited reports provide specific information about the functional response to different treatment approaches, and none of the information that we have found includes adequate descriptions of post-treatment voice. The case study that follows provides information in this regard by quantitatively describing the voice characteristics of a professional singer with granuloma, his post-surgery voice, and his rate of post-surgical voice recovery.

Methods

Patient description and early clinical course

The patient was a 36 year old male with a negative history for smoking. He was a professional tenor and a faculty member in the department of music at our university. He described himself as a lyric tenor, with 8 years of formal voice study at major institutions. His performing career focused on opera, oratorio, concert, and recital work with major regional symphonies and opera companies in the United States.

Prior to his initial clinical contact with us, we had interacted with the patient frequently and one of us had even performed with him in an opera. Thus, we had a considerable sampling of his pre-clinical voice-use patterns. Before his clinical presentation, we specifically noted a distinct absence of "pressed voice" in his speech and singing; i.e., we noted the absence of a phonation mode that videoscopic examination in our center and elsewhere repeatedly confirms involves a relative pressing of the arytenoids and general supraglottic constriction. The consistent pre-clinical perception was instead of slight arytenoid abduction, as described by Titze (32), particularly for loud singing. In fact, the patient later indicated that a cornerstone of his technique was intended arytenoid abduction with increasing loudness.[2]

During a singing tour several months prior to our initial examination, the patient developed what he described as a cold and consulted an otolaryngologist elsewhere. According to the patient, that evaluation indicated "swollen vocal folds" and incidentally, unilateral granuloma involving the left vocal process. The patient recovered from his cold and was further asymptomatic. Several months later, we performed a videoscopic evaluation of his larynx in the context of a pedagogical exercise, and noted the presence of a sizeable granuloma (see Figure 6 on center-bound plate) on the anterior left vocal process, without apparent contralateral findings. On questioning, the patient confirmed heartburn symptoms which resolved with an anti-reflux treatment program which he initiated at our recommendation (described below).

Conservative treatment and response

For two months following our initial laryngeal evaluation, the patient followed an anti-reflux regime. He discontinued alcohol and evening repasts, successfully engaged in a weight loss program, dropping from 210 to 185 lbs., slept with torso elevation, and took 150 mg of Zantac before reposing. Because of what we perceived as a pervasive absence of arytenoid "pressing" in his speech and singing, and because iterative coughing and throat-clearing were not noted and were specifically denied, no behavioral therapy was provided initially. After the initial 2-month treatment period, the granuloma appeared stable or marginally improved compared to the previous examination, although a small contralateral concavity was noted on the right arytenoid process on close fiberscopic examination. However, the patient reported the onset of a series of symptoms in singing, including "pitch-splitting" during warm-ups, losses in dynamic control, and rapid phonatory fatigue. As a treatment approach and also to rule out a malignancy, microscopic laryngoscopy was performed.

[2] This patient had previously taken a voice science course at our university, and was thus able to describe his voice technique in physiological terms, for example with reference to intended arytenoid adduction.

Surgical treatment and medical follow-up

To minimize injury to the vocal tract, microdirect laryngoscopy was performed under general anesthesia during apnea without intubation or the use of jet ventilation. Employing this technique, adequate operating time was permitted to perform laryngeal suspension and to accomplish an excisional biopsy of the lesion under magnification with microlaryngeal scissors and forceps. Following excision, ventilation was resumed by mask application. A second and final suspension of the laryngoscope was performed to reevaluate the adequacy of excision and to apply Triamcinolone Acetonite Suspension (E.R. Squibb & Sons, Inc.), 10 mg/cc, to the resection site. Following surgery, the patient was referred for a gastrointestinal medical consult, and he continued with anti-reflux therapy.

Peri-surgical behavioral intervention

Minimal behavioral intervention (voice therapy) was provided to this patient before and after surgery. At the time of the pre-surgical voice evaluation, we counselled him regarding post-operative voice use. Specifically, we recommended gradual resumption of phonation starting 3-4 days post-surgically, and emphasized "falsetto" and slightly aspirate phonation to minimize putative arytenoid pressing and mechanical trauma. In fact, the patient's pre-surgical trials with the type of phonation we described differed little with respect to his habitual, pre-clinical phonation mode. Subsequently, we saw the patient 8 days post-surgically to monitor his progress. At that time, we documented stable maximum phonation time on /a/ and /z/ compared to pre-surgical status, grossly indicating a good preliminary functional outcome. After the 8-day post-surgical visit, we communicated with the patient intermittently by telephone to assess voice use and to monitor recovery informally. Seven weeks following surgery, we repeated pre-surgical voice measures (described below) to formally assess the surgical outcome from a functional perspective.[3]

Pre and post-surgical measures and equipment

The larynx was evaluated several times pre-surgically with standard mirror and endoscopic examinations, and again about 2 weeks and 2 1/2 months post-surgically. Still photographs of the larynx were made just prior to surgery and at the 2-week follow-up.

In addition to the laryngeal evaluations, voice measures were made 4 days prior to surgery and again 7 weeks following surgery, and were compared to norms. Measures equal to or exceeding $\pm$ one standard deviation from the norm for vocally healthy subjects ($z \geq /1/$) would be considered an indication of poor performance or, in some cases, superior performance, depending on the test and the sign of the z-score. In cases for which standard deviations could not be calculated due to the unavailability of sufficient normative data or due to the nature of the scaling procedure (ordinal), judgments about "normal" or "superior" status versus "mild, moderate, moderately-severe, and severe" impairment were made based on clinical experience. The voice measures are described in the following sections.

[3] With similar pathological profiles in other patients, our bias is generally to provide more behavioral therapy than was provided to this patient. Limited therapy was involved in the present case because of our perception of a low-trauma approach to both speech and singing, an ability to consistently produce the phonatory behaviors we described without training, and a high level of motivation and compliance.

Measure of phonatory effort. Phonatory effort was measured based on an approach originally described by Colton and colleagues (33-35), and more recently used in a modified version by Verdolini-Marston and colleagues (36-38). The patient was instructed to rate his general effort level separately during speech and during singing on a ratio scale on which 100 represented a comfortable amount of effort, 200 represented twice as much effort ("very effortful"), and 50 represented half as much effort ("very easy").

Measures of dysphonia. Dysphonia was measured with auditory-perceptual and with acoustic measures. For the auditory-perceptual measures, the patient's voice was tape-recorded during reading and during conversational speech. Tapings were conducted in a quiet room using an AKG SA40 condenser microphone, with a microphone-to-mouth distance of approximately 6 inches, and a Panasonic Digital Audio Tape Deck, SV-3500. An approximately constant recording intensity was maintained across sessions by monitoring the VU meter of the recording apparatus, and recordings were stored on digital tape. After both pre- and post-surgical tape recordings were made, two speech pathologists with 9 and 10 years of clinical experience respectively, and who reported normal hearing, independently listened to pre- and post-surgical taped samples at a standard distance and intensity in freefield. Neither of the listeners knew the patient nor were familiar with his case. The order of the pre- and post-surgical samples was counterbalanced across the two judges. Prior to stimulus presentation, the listeners were simply informed that they would hear one pre-and one post-surgical sample from the same patient, each sample including both reading and conversational segments. The instruction was to rate each sample on a scale from 1 (normal voice) to 5 (severe dysphonia), using personal criteria. Not only integer scores, but also scores between integers were encouraged. Average scores across listeners are reported.

Acoustic measures of dysphonia included jitter (cycle-to-cycle variations in the period of the signal, in ms), shimmer (cycle-to-cycle variations in the amplitude of the signal, in %), and signal-to-noise (SN) ratio (in dB). For these measures, the patient first sustained the vowels /i/, /a/, and /u/ for 2 seconds each, at a comfortable conversational pitch. Input was obtained with the condenser microphone, and was routed to a Gateway 2000 386 computer and digitized at 50,000 Hz using Hypersignal software (39). After the signals were collected, CSpeech software (40) was used to extract average jitter (ms), shimmer (%), and SN ratio (dB) for a 300 ms segment selected from the middle portion of each vowel. Care was taken to select a segment for analysis with the least apparent intensity variation, based on visual inspection of the signal. Because post-surgical measures indicated high analysis error for the vowel /i/, pre- and post-surgical comparisons were not possible for this vowel. Thus, only the values for /a/ and /u/ are reported in this document. For both pre- and post-surgical analyses, average measures were collapsed across the two vowels.

Measure of respiratory volume and management. Respiratory volume and management were estimated with a standard prolonged /s/ task (41). That is, the patient was instructed to sustain an /s/ for as long as possible on a single trial. A stopwatch was used to time his performance.

Measures of vocal fold closure during phonation. The adequacy of laryngeal valving during phonation was assessed with three, independent measures. The first measure was maximum phonation time on /a/ (42); again, a single trial was involved, and a stopwatch was used. The second measure was the s:z ratio, that is, the ratio of maximum prolongation of /s/, in seconds, divided by the maximum prolongation of /z/, in seconds (41). One trial was performed for each phoneme, and durations were again timed with a stopwatch.[4] Third, the patient sang the first two phrases of

[4] All maximum prolongation trials were spaced so that several minutes separated successive trials.

"Happy Birthday" as quietly as possible ("Happy Birthday to you, Happy Birthday to you"), starting with fo = 329.63 Hz (E4), and emphasizing falsetto register. Performance was rated by one of the authors (KVM) on a scale from 1-10, on which 1 indicated severely delayed onset of phonation with markedly discontinuous phonation, and 10 indicated immediate onset of quiet phonation, and continuous phonation, as described by previous authors (43).

Fundamental frequency and intensity during speech. The approximate average fundamental frequency (fo) during speech was estimated from a rote speech task. For this measure, the patient counted from 1-5. The examiner, who had good pitch-matching abilities, extracted the fundamental frequency for the steady-state of the vowel on the word "three" by matching the pitch to a Casio keyboard, and converting to fo.

The approximate average intensity during speech was estimated by engaging the patient in conversational speech, and observing the approximate average dB from a portable Realistic Sound Level Meter set on the A-weighted scale and the slow response mode. Microphone-to-mouth distance was approximately 3 feet.

Plot of intensity of frequency range (phonetogram). A plot of minimum and maximum intensities over the frequency range, or phonetogram (44), was obtained by first identifying minimum and maximum frequencies. The patient initiated phonation on /a/ at an arbitrarily selected intermediate pitch, and then descended in pitch by scaler degrees or by semitones, until both he and the examiner were convinced by repeated attempts that the minimum pitch had been obtained. A similar procedure was used to determine the highest pitch, including falsetto. For actual phonetogram trials, the patient produced the quietest voice possible, and subsequently the loudest voice possible, on the vowel /a/, on six different notes arbitrarily selected by the examiner, across the total pitch range including falsetto. The first trial involved an intermediate pitch. Then, progressively lower pitches were produced, and finally, progressively higher pitches. For each production, target pitches were provided with a Yamaha studio piano, and the duration of each trial was at least 2 seconds. The approximate average intensity for each trial was noted using the Realistic Sound Level Meter set in the A-weight, slow response mode. For all trials, the patient was seated and the microphone-to-mouth distance was approximately 3 feet. Relevant for the results, the dB meter used for these trials registered intensities no lower than 50 dB.

The phonetogram procedure just described was the standard procedure used in our clinic at the time of data collection, and provided a pre-surgical as well as a post-surgical phonetogram at 7-week follow-up. The results could be grossly compared with available norms for frequency range, and independently, for maximum and minimum intensities without consideration of frequency, measured with a microphone-to-mouth distance of about 1 foot. Subsequent to these data collections, we updated the routine phonetogram procedures in our clinic in a way that allowed for specific comparisons with actual phonetograms described by Klingholz (45) for trained singers including tenors, as opposed to comparing our phonetograms to norms for frequency and intensity ranges independently. We extracted a repeated phonetogram for our patient 15 weeks following surgery according to the new protocol. Compared to the previous methodology, the changes in the updated protocol were: (a) the data were collected in a sound-isolated room (ambient noise was approximately 25 dB); (b) the patient stood for all trials; (c) falsetto trials were excluded; (d) the microphone-to-mouth distance was approximately 1 foot (as compared to the previous 3 feet); (e) only intensities that were sustained for a minimum of 2 seconds were recorded; and (f) a Brüel and Kjaer Type 16131 Impulse Precision Sound Level Meter set in the A-weighted mode was used to measure phonation intensities (note that this instrument measures intensities as low as 0 dB SPL). Inciden-

tally, pitches were provided with a portable Casio keyboard instead of the Yamaha studio piano, for logistical reasons.

Pragmatic measures of voice function and recovery. Critical measures of voice recovery were provided by the results of singing engagements that the patient had accepted before he was aware of his pathology. That is, soon after his surgery, the patient performed as a soloist with two different symphonies. Two-and-a-half weeks following surgery, he performed as a tenor soloist with a professional symphony singing Mozart's Requiem, and five-and-a-half weeks following surgery, he performed with a major symphony singing the entire tenor role in Handel's Messiah. Measures of voice proficiency were provided both by the patient's impressions of his performances, and by a publicized review of the second performance.

Results and Discussion

Laryngeal measures

As already noted, the pre-surgical lesion, shown in Figure 6 on plate, involved the left vocal process. Contralateral involvement was noted only on later, close videoscopic examination. Histological results from surgery indicated ulceration and granulation tissue. Figure 7 (see center-bound plate) is the view at 2-week post-surgical follow-up. This figure reveals small, persistent tissue irregularities at the site of biopsy. Follow-up examination at 2 1/2 months indicated no significant change compared to the previous exam.

Voice measures

Effort measures. Table 1 (see following page) shows that measures of phonatory effort were normal ("comfortable") for speech before surgery, but reflected moderately-severe or severe impairment during singing. Post-surgically, the effort measure remained "comfortable" for speech and improved to "comfortable" for singing.

Measures of dysphonia. Table 1 also displays auditory-perceptual measures of voice. Both of two blind judges who listened to taped samples of reading and conversation rated the patient's voice as normal before surgery (both judges' scores = 1), and both also rated his voice as normal after surgery (both scores = 1).

Also shown in Table 1, acoustic measures of dysphonia were also normal both pre- and post-operatively. Our interpretation of numerically worse acoustic measures post-surgically as compared to before surgery is that purposefully decreased arytenoid adduction post-surgically led to a slight increase in unmodulated airflow, and thus elevations in noise. However, as noted, post-surgical measures remained well within normal limits.

Respiratory volume and/or management. Table 1 shows that respiratory volume and/or management, estimated by maximally prolonged /s/, was superior both pre- and post-surgically.

Measures reflecting adequacy of laryngeal closure during phonation. Also indicated in Table 1, three different measures reflecting adequacy of laryngeal valving indicated good vocal fold closure during phonation both pre-and post-surgically. That is, maximum phonation time on /a/, the s:z ratio, and performance on a high-frequency, low-intensity singing task were all within normal limits.

Average fundamental frequency and intensity during speech task. Approximate average speaking fundamental frequency, extracted during a rote speech task, was normal both pre- and post-operatively, and was certainly not lower than the expected average in either case. Pre-operatively,

the average fo during speech was virtually the same as the norm, and post-operatively, the average fo was 13.47 Hz (approximately 2 semitones) higher than the norm.[5] Both pre-and post-surgical values fell well within one standard deviation of the norm. Thus, descriptions of low fundamental frequency in patients with laryngeal contact ulcers or granuloma, reported by other authors (22) were not confirmed for this patient.

The approximate average intensity of conversational speech was normal or somewhat loud prior to surgery (70 dB) based on norms at three feet, but was normal or even somewhat quiet post-surgically (60-65 dB). Overall, however, both pre- and post-operative conversational intensities were grossly consistent with the observation of normal "comfortable" intensities in patients with granuloma, reported by Hillman et al. (22).

Frequency range, intensity range, and phonetogram. Pre- and post-operative frequency ranges including falsetto are displayed in Table 2 and in Figure 1. As indicated, total frequency range was superior to general norms both pre-and post-operatively. Also shown, pre-surgical minimum and maximum intensities, measured at about 3 feet and including falsetto trials, appeared approximately normal, allowing for slight discrepancies due to differences in measurement distance as compared to normative procedures. More importantly, repeating the same data collection procedure as pre-surgically, the patient's dynamic range was clearly improved at 7-week follow-up compared to pre-operative status. This result is also shown in Table 2 and in Figure 1; post-surgically the patient gained 9 dB or more in minimum intensity in the lower dynamic range, obtaining minimum intensities of 50 dB or less following surgery.[6]

As noted in the Methods section, phonetogram measures were repeated with an updated protocol at 15-week follow-up. This protocol allowed for specific comparisons with phonetograms, and moreover, phonetograms for trained tenors (45). The most important methodological differences were that falsetto trials were excluded, and intensities were measured at about one foot with a sensitive sound level meter. The results of this later phonetogram are displayed in Figure 2, where they are compared to results for trained tenors. This figure shows a slight shift in our patient's phonetogram performance towards lower frequencies, compared to the tenors described by Klingholz (45), but overall a superior frequency range and also a superior dynamic range, both in terms of minimum and maximum intensities. We interpret our patient's performance as an indication of a somewhat "heavier" voice category than the tenors described by Klingholz, but generally indicative of superior voice functioning at 15-week follow-up, even compared to trained tenors.

Pragmatic measures. The patient's impressions of his professional performances 2 1/2 and 5 1/2 weeks following surgery were good. In fact, he thought that the latter performance series in particular was among the best in his professional career. A published review of the later performance was consistent with the patient's impression of a good result.[7] Thus, based on the patient's own impressions as well as a published review, post-surgical voice was excellent as early as 2 1/2 weeks following surgery, and certainly by 5 1/2 weeks.

[5] After he performed the counting task post-surgically, the patient commented that he thought his speaking pitch was probably higher than usual because he had just finished singing. Based on this comment, the slight increase in speaking pitch post-surgically should not be attributed to a surgical result.

[6] Recall that for these phonetogram trials, the intensity measuring device that we used did not indicate values lower than 50 dB.

[7] In order to preserve the patient's anonymity, the contents of the review cannot be reported here.

Table 1

Normative Data and Standard Deviations (SD), and Patient's Pre- and Post-Operative Performance on Measures of Phonatory Effort, Dysphonia, Respiratory Volume/Management, Glottal Closure during Phonation, Average Fundamental Frequency (Fo) and Intensity (I) during Speech, Frequency Range, and Summary of Phonetogram Task. Specific Tasks and Measurement Units Indicated in Table, Patient's Z-scores Provided where Possible, and Summary Evaluative Statements also Provided.

			PRE-OPERATIVE			POST-OPERATIVE		
	Norm	SD	Raw-score	Z-score	Summary	Raw-score	Z-score	Summary
Measure/Task								
Effort[8]								
Speech	100	NA	100	NA	Normal	100	NA	Normal
Singing	100	NA	200	NA	Mod-Severe or Severe	100	NA	Normal
Dysphonia								
Auditory-Perceptual[9]								
Reading	1	NA	1	NA	Normal	1	NA	Normal
Conversation	1	NA	1	NA	Normal	1	NA	Normal
Acoustic[10]								
Jitter (ms)	≤ .04	NA	0.01	NA	Normal	0.01	NA	Normal
Shimmer (%)	≤3.40	NA	1.10	NA	Normal	2.90	NA	Normal
SN Ratio (dB)	≥18.20	NA	30.89	NA	Normal	24.97	NA	Normal
Respiratory[11]								
/s/ (s)	17.73	7.65	46.00	3.70	Superior	40.00	2.91	Superior
Glottal closure during phonation								
/a/ (s)[12]	25.89	7.41	29.00	0.42	Normal	31.00	0.69	Normal
s:z ratio[13]	0.99	0.36	1.06	0.19	Normal	1.03	0.11	Normal
singing task[14]	8-10	NA	8-9	NA	Normal	9	NA	Normal
Average fundamental frequency (fo) and intensity (I) in speech								
fo (Hz)[15]	112	2.5[16] semitones	110	≤ /1/	Normal	123.5	≤ /1/	Normal
I (dB at 3 ft.)[17]	65-70	NA	70	NA	Normal-Loud	60-65	NA	Normal-Quiet
Frequency range (semitones)[18]								
	37.9	NA	43	1.02	Superior	43	1.02	Superior
Phonetogram								
(see Table 2, Figures 1 and 2)					Superior			Superior

[8]Informal norms and Summary (Evaluative) Statements from Clinical Voice Lab, The University of Iowa.

[9]Normal performance assumed to correspond to rating of "1", on an ordinal scale ranging from 1 (normal) to 5 (severely dysphonic).

[10]Informal norms based on data presented in CSpeech manual (for normal subjects, jitter values were 0.37 ms or less, shimmer values were 3.40% or less, and SN ratios were 18.2 dB or greater.

[11]Norms based on Eckel and Boone [41].

[12] Norms summarized by Colton and Casper [42], based on review by Kent, Kent, and Rosenbek [47] of numerous studies.

[13]Norms based on Eckel and Boone [41].

[14] Norms based on Bastian, Keidar, & Verdolini-Marston (43).

[15] Norm based on Hollien and Shipp (48).

[16] The standard deviation was not indicated by Hollien and Shipp (48). However, this value can be estimated roughly from other values indicated by Colton and Casper (42) as about 2.5 semitones.

[17] Norms based on Hirsh (49), who indicated 70 dB as normal intensity in conversational speech, without specifying measurement distance, and Hirsh (personal communication), who indicated 65 dB as normal intensity in conversational speech, measured at 3 feet.

[18] Norms from Hollien, Dew, and Phillips (50), including falsetto.

Other. At 7-week post-surgical follow-up, the only persisting symptom that the patient reported was a dull laryngeal pain that occurred about 20 minutes after rigorous singing. By the 15-week follow-up date, this remaining symptom was entirely resolved.

Table 2.
Results of Phonetogram Trials Pre-Surgically, at 7-Week Post-Surgical Follow-Up, and at 15-Week Post-Surgical Follow-Up. Frequencies Indicated in Hz; Intensities Indicated in dB. (Recall that Trials at 7-Week Test Included Falsetto, and Measurements Were Made at 2 feet. At 15-Week Test Trials Excluded Falsetto, and Measurements Were Made at 1 Foot.) Norms Indicated in Figure 1.

Pre-Surgical

Frequency	Pitch	Minimum Intensity	Maximum Intensity
82.4	E2	61	68
116.5	Bb2	59	88
220.0	A3	64	95
329.6	E4	69	105
587.3[19]	D5	70	103
987.8	B5	89	103

Post-Surgical (7-Week Follow-Up)

Frequency	Pitch	Minimum Intensity	Maximum Intensity
82.4	E2	NA	NA
87.3	F2	50 or less	70
146.8	D3	50 or less	90
246.9	B3	55	100
349.2[20]	F4	65	106
587.3	D5	69	105
987.8	B5	90	105

Post-Surgical (15-Week Follow-Up)

Frequency	Pitch	Minimum Intensity	Maximum Intensity
73.4	D2	55	66
110.0	A2	51	85
146.8	D3	47	92
196.0	G3	52	100
293.7	D4	64	102
440.0	A4	81	106
523.3	C5	85	104
544.4	C#5	NA	NA

[19]587.3 Hz and 987.8 Hz were produced in falsetto.
[20]Trials at 349.2 Hz and higher were produced in falsetto.

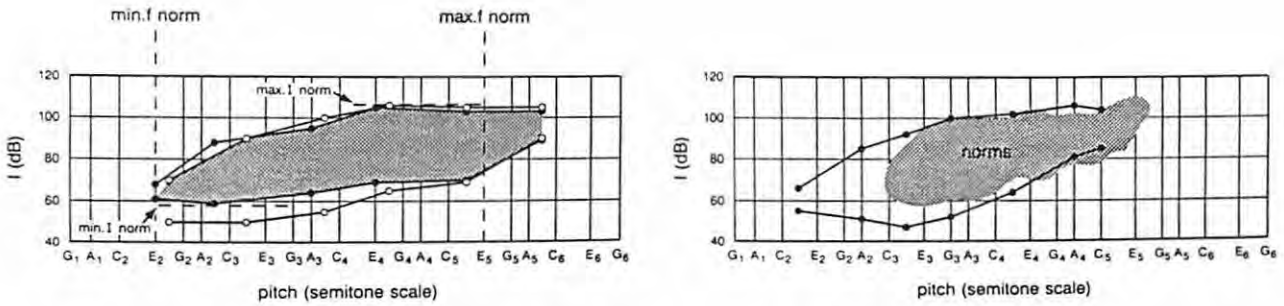


Figure 1 (left):Phonetogram results (intensity in dB over pitch in semitones) for pre-surgical trials, and for trials at 7-week post-surgical follow-up. Pre-surgical status is indicated by darkened space. Post-surgical status is indicated by solid lines. For both tests, intensities were measured at about 3 feet, and falsetto trials were included. The norm for maximum intensity, expected to occur in the higher frequency range and measured at about 1 foot (46), is indicated by the higher dashed horizontal line. The norm for minimum intensity, expected to occur in lower frequency range and again measured at 1 foot, is indicated by the lower dashed horizontal line, and is based on visual inspection of data reported by Klingholz (45, pp. 19,21), for trained tenors. The norms for minimum and maximum frequencies are indicated by dashed vertical lines [Colton and Casper (42) pp. 312, who cite Hollien et al. (50)]. Figure 2 (right):Phonetogram results (intensity in dB over pitch in semitones) for 15-week post-surgical trials, as compared to results for trained tenors (45). Measurements were made at about 1 foot, and falsetto trials were excluded. The patient's performance is indicated by solid lines. The darkened space indicates the performance of trained tenors.

General Discussion

The results from the case study can be summarized and discussed in five points. First, an important finding is that pre-surgical voice measures indicated mostly normal and even superior voice functioning, despite the presence of a sizeable laryngeal granuloma, and despite the patient's comments about subtle symptoms. The primary exception to the impression of normal or superior voice pre-surgically was a moderately-severe to severe elevation in perceived effort during singing. The largely normal voice profile might be surprising, considering the prominence of the pre-surgical lesion, and also considering frequent descriptions in the literature of hoarseness associated with laryngeal granulomas (6,10). However, in our patient the integrity of the vocal folds themselves was not affected, as is often the case with granulomas. With this factor in mind, in fact the more surprising observation would have been hoarseness or other notable voice dysfunctions, rather than normal voice. In addition to an absence of hoarseness, approximate average fundamental frequency during speech was entirely normal in our patient both pre- and post-surgically. This result is inconsistent with previous descriptions of low fos in patients with similar lesions (22). The important point is that although granulomas may sometimes be associated with hoarse voice, low fos, and other phonatory abnormalities, such abnormalities do not necessarily accompany these pathologies.

A second and related point that is actually incidental to the primary purpose of the case report (the description of pre- and post-surgical voice status and post-surgical recovery rate), is that contrary to speculations about a mechanical contribution to granulomas' etiology in general, mechanical factors are not strongly implied in this particular case. Pre-clinical auditory-perceptual observations failed to reveal perceived arytenoid "pressing" during speech and singing, and similarly, iterative coughing and throat clearing appeared absent. Further, as already indicated,

fundamental frequency during speech was entirely normal both pre-and post-surgically, and the singing tessitura was high (tenor). Therefore, causal mechanical factors do not seem central in this case, not only with regard to "pressed" voice or coughing and throat clearing, but also with regard to low fundamental frequencies. Of course, some degree of trauma related to this patient's profession as a classical singer cannot be excluded. However, at this point our interpretation is that if traumatic contributions were present, the etiologic impact was secondary compared to the presence of other factors such as gastric reflux. This interpretation does not exclude a role of mechanical factors in other cases or even a secondary role in the present case. The point is rather that mechanical factors may not be the predominant ones in all cases.

Third, because mechanical factors did not seem to play a role, and also because we perceived this patient as able to control his voice functions in prescribed directions without outside training, minimal voice therapy was provided and an excellent functional outcome was obtained anyway (discussed next). This observation does not imply that voice therapy is generally irrelevant in the treatment of granulomas. Rather, the point is that voice therapy may be irrelevant in a subset of cases that involve a limited mechanical contribution.

Fourth, another important finding is that despite the persistence of a small posterior glottal finding post-surgically, post-operative voice functioning was entirely normal or superior at 7-week follow-up and, for the repeat phonetogram, also at 15-week follow-up, when compared to trained tenors. The principal pre-surgical finding of increased effort during singing resolved post-surgically. Also, pre-surgical symptoms of "pitch-splitting" and loss of medium dynamic range were gone following surgery, and complaints of fatigue with singing, described at 7-week follow-up, were resolved 15 weeks after surgery or earlier. Thus, for this patient the surgical outcome for voice can be considered excellent, despite previous reports implying poor functional outcomes following surgery (30).

Fifth and finally, post-surgical functional recovery was rapid. The patient performed as a soloist with two different major symphonies 2 1/2 and 5 1/2 weeks following his surgery, respectively. His own impressions, corroborated by an artistic review, were of successful performances. Further, the performances did not appear to provoke secondary injury at these early post-surgical dates, as indicated by normal or superior voice functioning at 7-week follow-up and stable laryngeal appearance at 2 1/2 month follow-up. This observation again supports the point that the functional outcome of surgery can be good in some cases, both in terms of level of post-surgical voice functioning and also in terms of functional recovery time.

Having made these points, caution about generalization is warranted for a series of reasons. Beyond the obvious but important caveat related to the limited series (N = 1), a restriction to generalization is that in our judgment, this patient had an uncommon ability to control his voice. Based on auditory-perceptual impressions, he specifically appeared able to control degree of arytenoid "pressing" consistently during both speech and singing, without outside training. We attribute to this factor his ability to perform well soon after his surgery without incurring injury. We emphasize this point because we do not take the present results as evidence that patients should necessarily resume aggressive and professionally demanding phonation two weeks following surgery for removal of granuloma, particularly without voice therapy. Rather, we interpret the present results as the description of a "best-case scenario" in terms of recovery rate in particular. Our bias is that most patients with similar profiles may need voice therapy for a good functional outcome, and probably not all patients should engage in taxing voice performances soon after surgery. A final caution is

that although voice functioning was good for our patient at follow-up, minor tissue irregularities persisted posteriorly as did symptoms of reflux. Therefore, the possibility of recurrence cannot be excluded.

Acknowledgements

The authors acknowledge Dr. Robert Bastian for a discussion with KVM about rehabilitation expectations prior to the patient's surgery. The following persons are also thanked: Ms. Sharon Lindsey, for technical assistance and for comments on an earlier version of this paper; Mrs. Linnie Southard for secretarial support; Ms. Kay Klein and Med Labs at The University of Iowa for assistance with preparing figures for publication.

References

1.Bishop MJ, Weymuller EA, & Fink R. Laryngeal effects of prolonged intubation. Anesth Analg, 1984;63:335-42.
2.Hefter E. Das intubationsgranulom. Der Anaesthetist, 1969;18:194-6.
3.McIlwain J. Clinical aspects of the posterior glottis. J of Otolaryngol, 1991;20(2):74-87.
4.Caldarelli DD, Friedberg SA, Harris AA. Medical and surgical aspects of the granulomatous diseases of the larynx. Otolaryngol Clin North Am, 1979;12(4):767-81.
5.Terent A, Wibell L, Lindholm C-E, Wilbrand H. Laryngeal granuloma in the early stages of Wegener's granulomatosis. ORL, 1980;42:258-65.
6.Ward PH, Zwitman D, Hanson D, Berci G.a Contact ulcers and granulomas of the larynx: New insights into their etiology as a basis for more rational treatment. Otolaryngol Head Neck Surg, 1980;88:262-69.
7.Desai SD, Rajratnam K. Laryngeal granuloma - an unusual presentation of AIDS. J Laryngol Otol, 1988;102:372-73.
8.Miko TL. Peptic (contact ulcer) granuloma of the larynx. J Clin Pathol, 1989;42:800-4.
9.Goldberg M, Noyek AM, Pritzker KPH. Laryngeal granuloma secondary to gastro-esophageal reflux. J Otolaryngol, 1978;7(3):196-202.
10.Jaroma M, Pakarinen L, Nuutinen J. Treatment of vocal cord granuloma. Acta Otolaryngol, 1989;107:296-99.
11.Ward PH, Berci G. Observations on the pathogenesis of chronic non-specific pharyngitis and laryngitis. Laryngoscope, 1982;92(12):1377-82.
12.Wenig BM, Heffner DK. Contact ulcers of the larynx. Arch Pathol Lab Med, 1990;114:825-28.
13.Jiang J. A methodological study of hemilaryngeal phonation and the measurement of vocal cord intraglottal pressure and impact stress. [Dissertation]. The University of Iowa, 1991.
14.Hirano M, Kurita S, Matsuo K, Nagata K. Laryngeal tissue reaction to stress. Transcripts of the Ninth Symposium: Care of the Professional Voice, Part II. NY: The Voice Foundation, 1980;10-20.
15.Jackson C. Contact ulcer of the larynx. Ann Otol Rhinol Otolaryngol, 1928;37:228-30.
16.Jackson C, Jackson CL. Diseases of the nose, throat and ear. Philadelphia: Saunders, 1959.
17.Jensen R, Lauerman LH, Braddy PM, Horton DP, Flack DE, Cox MF, Einertson N, Miller GK, Rehfeld CE. Laryngeal contact ulcers in feedlot cattle. Vet Path, 1980;17:667-71.
18.Moore GP. Organic voice disorders. Englewood Cliffs, N.J.: Prentice-Hall Inc, 1971.
19.Svensson G, Schalén L, Fex S. Pathogenesis of idiopathic contact granuloma of the larynx. Results of a prospective clinical study. Acta Otolaryngol, 1988;449:123-25.

20.von Leden H, Moore P. The Larynx and Voice: Laryngeal Physiology Under Daily Stress (sound film). Chicago: The William & Harriet Gould Foundation, 1958.
21.Hirano M, Yoshida T, Kurita S, Kiyokawa K, Sato K, Tateishi O. Anatomy and behavior of the vocal process. In: Baer T, Sasaki C, Harris C, eds. Laryngeal function in phonation and respiration. Boston:College-Hill Little Brown and Co Inc, 1987:3-13.
22.Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughn C. Objective assessment of vocal hyperfunction: An experimental framework and initial results. J Speech Hear Res, 1989;32(2):373-92.
23.von Leden H, Moore P. Contact ulcer of the larynx: Experimental observations. Arch Otolaryngol, 1960;72:746-52.
24.Cherry J, Margulies SI. Contact ulcer of the larynx. Laryngoscope, 1968;78:1937-40.
25.Delahunty JE, Cherry J. Experimentally produced vocal cord granulomas. Laryngoscope, 1968;78:1941-47.
26.Chodosh PL. Gastro-esophago-pharyngeal reflux. Laryngoscope, 1977;87:1418-27.
27.Öhman L, Tibbling L, Olofsson J, Ericsson G. Esophageal dysfunction in patients with contact ulcer of the larynx. Ann Otol Rhinol Laryngol, 1983;92:228-30.
28.Jaroma M, Pakarinen L, Rouhiainen P, Nuutinen J. How to handle vocal fold granuloma. Acta Otolaryngol, 1988;449:29.
29.Hirano M. Objective evaluation of the human voice: Clinical aspects. Folia Phoniatrica, 1989;41:89-144.
30.Peacher G. Vocal therapy for contact ulcer of the larynx. A follow-up of 70 patients. Laryngoscope, 1961;71:37-47.
31.Bastian RW. Benign mucosal disorders, saccular disorders, and neoplasms. In: Cummings CW, Fredrickson JM, Harker LA, Krause CJ, Schuller DE, eds. Otolaryngology - head and neck surgery. St. Louis: The C.V. Mosby Company, 1986:1965-87.
32.Titze IR. Messa di voce. NATS J, 1992;48(3):24-25.
33.Wright HN, Colton RH. Some parameters of autophonia level. Presented at the American Speech and Hearing Association, November 1972(a).
34.Wright HN, Colton RH. Some parameters of vocal effort. J Acoust Soc Am, 1972b;51:141.
35.Colton RH, Brown WS. Some relationships between vocal effort and intra-oral air pressure. Presented at the 84th meeting of the Acoustical Society of America: Miami FL, November 1972.
36.Verdolini-Marston K, Titze IR, Fennell A. Update on the effectiveness of hydration on phonation threshold pressure. Presentation at the NCVS Annual Conference: Iowa City IA, 1992.
37.Verdolini-Marston K, Sandage M. Effect of hydration treatments on laryngeal nodules and polyps and related measures. Manuscript submitted, 1992.
38.Verdolini-Marston K, Burke MK, Lessac A, Glaze L, Caldwell E. Manuscript in preparation.
39.Hypersignal Users Manual. Dallas, Texas, March 1989.
40.Milenkovic PH. CSpeech Core User's Manual: Cspeech Version 3. Madison, Wisconsin.
41.Eckel FC, Boone DR. The s:z ratio as indicator of laryngeal pathology. J Speech Hear Dis, 1981;46:147-49.
42.Colton RH, Casper JK. Understanding voice problems: A physiological perspective for diagnosis and treatment. Baltimore: Williams & Wilkins, 1990.
43.Bastian RW, Keidar AK, Verdolini-Marston K. Simple vocal tasks for detecting vocal fold swelling. J Voice, 1990;4(2): 172-83.

44.Gramming, P. The phonetogram: An experimental and clinical study. Malmö, Sweden. The Department of Otolaryngology, University of Lund, Malmö General Hospital, 1988.
45.Klingholz, F. Das Stimmfeld: Eine praktische Anleitung zur Messung und Auswertung. München: Verlag J. Peperny, 1990
46.Ptacek PH, Sander EK, Maloney WH, Jackson CCR. Phonatory and related changes with advanced age. J Speech Hear Res, 1966;9:353-60.
47.Kent RD, Kent J, Rosenbek J. Maximum performance tests of speech production. J Speech Hear Res, 1987;52:367-87.
48.Hollien H, Shipp T. Speaking fundamental frequency and chronologic age in males. J Speech Hear Res, 1972;15:155-59.
49.Hirsh IJ. The measurement of hearing. N.Y.: McGraw-Hill Book Company Inc, 1952:150.
50.Hollien H, Dew D, Philips P. Phonational frequency ranges of adults. J Speech Hear Res, 1971;14:755-60.

NCVS Status and Progress Report - 3
December 1992, 175-199

Effect of Hydration Treatments on Laryngeal Nodules and Polyps and Related Voice Measures

Katherine Verdolini-Marston, Ph.D.
Department of Speech Pathology and Audiology and Department of Music, The University of Iowa
Mary Sandage, M.A.
Saint Luke's Hospital, Cedar Rapids, Iowa
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology and Department of Music, The University of Iowa

Abstract

In this study, a double-blind, placebo-controlled approach was used to assess the effectiveness of hydration treatments in the clinical management of selected voice disorders. Six adult females with laryngeal nodules or polyps each received five consecutive days of a hydration treatment and five consecutive days of a placebo/control treatment. The combined results indicated improvements in voice and in laryngeal appearance following both placebo/control and hydration treatments, compared to baseline. However, the greatest improvements were obtained following the hydration treatment. Although caution about generalization of the effects to the typical clinical situation is emphasized, the study provides preliminary evidence of a therapeutic benefit from hydration treatments in patients with nodules or polyps. Based on previous theoretical work, hydration effects may be related to reductions in vocal fold tissue viscosity, although other explanations are also possible.

Hydration treatments are often adopted in the clinical management of voice disorders. Typical recommendations include ambient humidification (1, 2, 3), direct steam inhalations (2, 4), elevated intake of water or other hydrating fluids (2, 3, 4), mucolytic drug administration (4) and nose breathing (1, 3, 5). Sometimes the use of lozenges (2), propylene glycol solutions (4), lemon

juice and herbal teas (5), or comfrey (1) is recommended for laryngeal lubrication. Corollary recommendations include the avoidance of drying conditions and agents such as dry environments, smoke (5, 6), alcohol, caffeine, and other diuretics (1, 3, 5, 6), antihistamines (1, 5, 7), drying nasal sprays and decongestants (3, 7), and mouth breathing (1).

Various rationale for hydration treatments are given. For example, Sataloff (5) describes improvements in "scratchy voice" and "tickling cough" with nose breathing. Lawrence (2) implies that phonatory effort may be improved with hydration treatments, and that the risk of upper respiratory tract viral infection by droplet spread may also be reduced. Verdolini-Marston (3, 8, 9) and other clinicians claim that hydration treatments may be useful for the prevention and treatment of specific laryngeal pathologies such as nodules and polyps.

Despite the widespread clinical use of hydration treatments, little experimental evidence is available about their actual impact on voice, and no formal data are available about their clinical utility. Similarly, limited work has been done to explain the physiological or biomechanical mechanisms that may mediate hydration effects. The study reported here is part of an ongoing investigation assessing the effect of hydration treatments on voice in normal and in clinical populations, and underlying mechanisms. In the present study, the general question posed is a pragmatic one that focuses on the clinical population: Can any evidence be obtained of a specific benefit from hydration treatments for voice and for laryngeal status, beyond placebo/control effects, in subjects with laryngeal nodules or polyps?

In addition to this general question, more specific questions are asked about the effects of hydration treatments, based on previous theoretical and empirical work. Before introducing the specific questions, it is important to review the previous work to clarify the basis for the questions and to point out the need for further information.

Review of previous theoretical and empirical work

Some theoretically-based predictions can be made about the specific impact of hydration treatments on the larynx and consequently, on voice, by assuming that these treatments reduce vocal fold tissue viscosity, which in turn decreases energy losses during vocal fold vibration. The viscosity of pure water is 0.01 poise (P). The viscosity of vocal fold tissue is generally higher, ranging from approximately 1-10 P (10), and the viscosity of mucus and sputum normally varies from about 0.1-1.0 P (11)[1]. Although the exact mechanisms are not known, hydration treatments may decrease vocal fold tissue viscosity by infusing the tissue with low viscosity fluids, superficially or systemically. For example, mucolytic drugs, which are commonly included in hydration treatments, may decrease the viscosity of oral and respiratory secretions (12), as may water intake. Together with ambient humidification, the result could be a net reduction in vocal fold tissue surface viscosity. Assuming that such effects or similar ones do occur, energy losses during phonation should be decreased because of a direct relation between energy losses and tissue viscosity (13). Two different types of effects on voice and on the larynx should follow. First, the minimum lung pressure required to initiate and sustain vocal fold oscillation, the *phonation threshold pressure* (PTP), should be reduced (14). Second, the likelihood of edematous tissue reactions should also be minimized (13). The theoretical basis for both effects and supporting data are discussed next.

[1]The viscosity of human surface tissues including skin, muscle, and fat is about 100 P (10).

Possible reductions in PTP with hydration treatments. According to a theoretical framework proposed by Titze (14), PTP is regulated by the following parameters:

$$PTP = (2k/T)\,(Bc)\,(w/2)$$

where k is a transglottal pressure coefficient (which has a dimensionless value of about 1.1), T is vocal fold thickness, B is a damping coefficient that is proportional to vocal fold viscosity, and $w/2$ is pre-phonatory glottal half-width. The formula applies to a rectangular prephonatory glottis, but similar relations can be derived for other configurations. Relevant for the present discussion, PTP is theoretically proportional to vocal fold tissue viscosity which, as already noted, may depend inversely on hydration level.

The results from an experiment by Finkelhor, Titze and Durham (15) are consistent with the proposed relation between PTP and vocal fold tissue viscosity. In that experiment, four excised canine larynges were successively bathed in solutions with different osmotic pressures, that is, with more or less saline concentration. Water was transported through the mucosa by the osmotic gradient (the difference in saline concentration across the mucosal barrier). Subsequent to the bathing, PTPs were measured over a range of vocal fold elongations. The result was that the greatest pressures were required to produce phonation following baths with the least water transportation, and the lowest pressures were required following baths involving the greatest water transportation. Thus, PTPs depended inversely on hydration level and, by extension, appeared to vary directly with vocal fold tissue viscosity, as predicted. However, no direct measures of tissue viscosity were made.

Another experiment, by Verdolini-Marston, Titze, and Druker (16), produced qualitatively similar results. The primary methodological differences with respect to the previous study were that human subjects were used, and hydration and dehydration treatments were administered systemically and by altering the relative humidity of inspired air, under the assumption that the treatments respectively lowered and raised vocal fold tissue viscosity. Each of 6 subjects underwent a 4-hr hydration treatment, which involved exposure to high ambient humidity, elevated water intake, and administration of a mucolytic drug. The same subjects also underwent a 4-hr dehydration treatment, which involved exposure to low ambient humidity, liquid deprivation, and administration of a decongestant drug. About 10 min following both treatments, PTPs were estimated from oral pressures during consonant-vowel strings at low, conversational, and high pitches, and were compared to estimated PTPs in a no-treatment control condition. The outcome was that PTPs were highest following the dehydration condition and lowest following the hydration condition. The reduction in PTPs following hydration was particularly marked at high pitches. Although no direct measures of vocal fold tissue viscosity were possible, reductions in PTP with the hydration treatment were attributed to decreases in tissue viscosity, as predicted theoretically.

Based on the results from these studies, it is tempting to conclude that hydration treatments reduce pulmonary effort during phonation and, by extension, perceived phonatory effort. However, further work is required to substantiate these claims, for several reasons. First, vocal fold viscosity was not explicitly measured in either of the experiments, thus the role of tissue viscosity requires further, more direct investigation. Second, perceived phonatory effort was also not measured. Therefore, the possibility that hydration treatments reduce perceived phonatory effort in addition to PTPs remains entirely speculative at this point. Third, in the study by Verdolini-Marston and colleagues using human subjects (16), 4 of the 6 subjects were not naive to the experimental

hypotheses, and the experimenter who elicited PTPs was not blind to subjects' conditions. Therefore, unwanted subject and/or experimenter biases could have affected the results. Finally, in the study using humans, only laryngeally healthy subjects were examined. Therefore, no information was provided about the effect of hydration treatments on PTPs (or perceived phonatory effort) for the typical clinical population, for whom such information seems particularly relevant.

Possible reduced risk of laryngeal injury with hydration treatments. According to another theoretical framework proposed by Titze (13), reductions in vocal fold tissue viscosity may also minimize the risk of edema-based laryngeal injury. The fundamental hypothesis is that internal friction generates heat during tissue deformation and that heat varies directly with tissue viscosity. The additional piece of information is that a physiological response to high levels of heat loss may be local edema, as a "cooling" effect. Low viscosity tissue is anticipated to result in relatively small heat losses and as a consequence, the likelihood of edema may be minimized. Therefore, assuming that hydration treatments reduce vocal fold tissue viscosity, such treatments may be useful in the prevention and treatment of local injury. These predictions remain altogether hypothetical, however, because no empirical reports are available.

Summary of previous work and specific questions asked in the present study

In summary, a general hypothesis is that hydration treatments may reduce energy dissipation in vocal fold tissue during phonation, by decreasing tissue viscosity. Based on previous theoretical work, two specific effects are anticipated. First, PTPs and perceived phonatory effort should be reduced. Second, the risk of edema-based injury should be minimized. Based on these predictions, the specific questions in the present study are: (a) Do hydration treatments reduce perceived phonatory effort and PTP more than placebo/control treatments, in subjects with laryngeal nodules or polyps? and (b) Do hydration treatments reduce edema-related laryngeal injuries (nodules or polyps), more than placebo/control treatments? Related to this, do hydration treatments improve auditory-perceptual and acoustic measures of voice?

Methods

Subjects. Six adult females with laryngeal nodules or polyps participated as volunteers. Subjects' ages ranged from 18 - 33 yr ($\underline{M}$ = 22 yr). According to subjects' reports, the average time since onset of symptoms was 2 yr 4 mo (6 mo - 5 yr 8 mo), and the average time since the initial diagnosis was 1 yr 1 mo (0 - 5 yr 8 mo). Subjects denied active illnesses prior to participation, and specifically denied a history of risk factors related to experimental treatments, that is kidney dysfunction, glaucoma, diabetes, and hypoglycemia. Pure-tone hearing screening indicated normal thresholds (15 dB or lower in the better ear) at 250, 500, 1000, 2000, 4000, and 8000 Hz.

All of the subjects were college students. Two subjects were recruited from university sororities. Neither of these subjects had received prior voice training or voice therapy. The other four subjects were inducted from the clinical voice therapy program at our university, and all four of these subjects were singers with varying amounts of prior voice training, ranging from 3 mo to 5 yr 6 mo ($\underline{M}$ = 1 yr 11 mo). Three of the singer-subjects had not yet initiated voice therapy, nor had they received prior voice therapy elsewhere. The remaining singer-subject had received approximately 5 months of voice therapy in our clinic before beginning the protocol. None of the subjects received voice therapy during participation in the experiment.

One subject normally smoked about one cigarette a day, and had a cumulative smoking history of 0.2 pack yr (a standard unit in otolaryngologic practice, that is the number of packs smoked per day times the number of years smoked). Another subject had a cumulative history of 12.5 pack yr, but had discontinued smoking about 10 yr prior to her participation. All other subjects had negative smoking histories. Subjects typically drank alcohol very little or moderately, mostly on weekends. The average number of "drink years" (number of drinks per day times the number of years) was 2.88 yr.

Treatments. As an overview, each subject received a hydration treatment and a "placebo/control" treatment. The purpose of the placebo/control treatment was to account for all changes in voice and laryngeal measures caused by factors that were not of direct interest in this study, that is placebo factors, general voice conservation and hygiene factors, and maturational factors, discussed shortly.

Both the hydration and the placebo/control treatments were administered for five consecutive days during consecutive weeks. For the hydration treatment, subjects were instructed to drink 8 or more glasses of water per day (plastic 16-oz glasses were provided), to take one teaspoon of a "hydration medication" (Robitussin expectorant, a mucolytic) three times per day at about 6-hr intervals, and were exposed to high humidity environments in our clinic (90-100% relative humidity, monitored at approximately half-hour intervals) for two hours each day. For the placebo/control treatment, subjects were instructed to perform 8 or more sets of 20 bilateral forefinger flexions per day, to take one teaspoon of an "herbal medication" (actually cherry syrup) three times per day at about 6-hr intervals, and were exposed to a room with commercial airfilters and scented candles for two hours daily in our clinic. An airfiltered, candle-lit environment was used for daily exposures in the placebo/control treatment because pilot work indicated that this particular situation created a "suggestive" environment, about as suggestive as the humidity environment used in the hydration treatment. During the placebo/control treatment, supposed "air purity" was monitored and recorded at approximately half-hour intervals. The (unmanipulated) humidity level in the "airfilter/candle" environment ranged from approximately 30-40%.

In addition to these specific aspects of the treatments, for both weeks of treatment subjects were instructed to observe general voice conservation and hygiene measures by limiting heavy voice use, in particular yelling and screaming, as well as alcohol and caffeine intake and exposure to smoke. The rationale was that yelling in particular might obscure treatment effects by causing exacerbations of the existing laryngeal pathologies. Additionally, caffeine, alcohol, and smoke might counteract hydration effects. Because we wished to make extra-clinical behaviors as similar as possible across treatment periods, we instructed subjects to restrict exposure to these agents during both hydration and placebo/control treatments.[2] To indicate their adherence to the aspects of the protocol carried out away from the clinic, that is all aspects of the protocol except for the daily exposures, which we monitored directly, subjects filled out logs that we provided. During the hydration treatment, in addition to indicating amount of water intake, subjects also recorded the

[2]Of course, by restricting dehydrating agents such as caffeine, alcohol, and smoke during the placebo/control treatment as well as during the hydration treatment, hydration effects might be underestimated in this experiment, where hydration effects are defined as relative benefits from hydration treatments as compared to benefits from placebo/control treatments. However, if the results were confounded by this manipulation, the confound would be in the direction a conservative interpretation, which is generally acceptable.

amount of other non-dehydrating fluids consumed. Information about water and other non-dehydrating fluid intake was not requested during the placebo/control treatment, because we did not wish to inadvertently encourage the consumption of such fluids during this treatment. However, subjects did retrospectively estimate this information after the experiment was completed.

Pilot work indicated that voices and laryngeal status tended to vary with day of the week in the population from which we were drawing, that is, female college students. Specifically, voice and laryngeal status tended to deteriorate markedly during weekends, because of party activity. From this observation it became clear that the treatment effects we were attempting to detect in the experiment would be obscured if treatments were delivered over weekends. In order to administer treatments over several consecutive days but to avoid treatment delivery over weekends, pre-treatment measures for all subjects were made on a Sunday, treatments were initiated the same day, and all post-treatment measures were made on a Friday. Obviously, by using this design a considerable proportion of any treatment effects that we found would be due to a maturational factor, that is day-of-the-week. However, as already noted, the placebo/control group provided an experimental check on this factor, as it controlled for all treatment factors beyond those of specific interest (hydration factors).

Finally, to include observations across different seasons, some subjects participated in the experiment in late winter (March; Subjects 1-4), and some participated in early summer (June; Subjects 5-6).

Voice and laryngeal measures. A series of voice and laryngeal measures were made immediately prior to initiation of the protocol, and one day following the termination of each treatment. These included measures of phonatory effort, PTP, laryngeal status, auditory-perceptual status and acoustic status of voice.

Measures of phonatory effort were similar (but not identical) to a magnitude estimation approach described by Wright and Colton (18, 19) and Colton and Brown (20). This approach was appealing because of a direct relation with estimated pulmonary pressures reported in one of the earlier studies (20). Subjects were provided a reference of "100" to describe "a comfortable amount of effort during phonation." The instruction was to rate phonatory effort over the preceding 24 hr, on a scale on which 200 indicated "twice as much effort as comfortable", that is very effortful, and 50 indicated "half as much effort as comfortable", that is very easy.

Also following methodologies described by previous authors (16, 21, 22, 23, 24), PTPs were estimated from oral pressures during stop consonants in consonant-vowel strings. The syllables /pae pae pae pae pae/ were produced at a rate of 92 beats per minute, indicated by a metronome. During each of three measurement sessions (pre-treatment, post-hydration, and post-placebo/control), subjects performed five trials of the consonant-vowel string at a conversational pitch, five trials at a high pitch, and finally five trials at a low pitch. The different pitches were individually established for each subject during the pre-treatment measurement session, and the same pitches were used for that subject for all subsequent testing in the experiment. The conversational pitch was extracted during a rote speech task. For this task, subjects counted out loud slowly from one to five and the experimenter matched the sustained pitch on the vowel /i/ in the word "three" to a keyboard. The high pitch was calculated as the 80th percentile pitch in the overall semitone pitch range on the vowel /a/, determined from descending and ascending pitch trials that were repeated until both the experimenter and the subject were convinced that the lowest and highest notes had been produced. The low pitch was the 10th percentile pitch in the overall semitone range on /a/. For all PTP trials, subjects were instructed to phonate as quietly and as evenly as possible. Prior to each trial, the

experimenter provided the target pitch both vocally and with the keyboard. Trials that the experimenter perceived as differing from the target pitch by more than one-quarter tone were discarded and repeated. Calibration of pressure signals was monitored regularly on a voltmeter, with the requirement that zero pressure inputs correspond to a display of less than /1.0/ mV on the meter. Calibration was further ensured following each trial by inspecting the pressure signal on the waveform display function of the software that we used for data collection. Any trials for which the minimum y-axis offset on the display was perceptibly displaced from zero were rejected and recollected. Signals whose magnitude exceeded the output capacity of our data collection system (clipped signals) were recollected with 10 dB of attenuation, and signals that were anticipated to be clipped were collected with 10 dB of attenuation straightaway. Also trials for which the magnitude of successive pressure peaks appeared highly irregular were excluded and repeated. Across sessions, the total percent of repeated trials was approximately 10%. Pressure signals were digitized at a rate of 2,000 samples per second and stored for later analysis. Analyses indicated the output (in arbitrary units generated by the data collection software) between adjacent pressure peaks pairs, for peaks 2-5 of the repeated string /pae pae pae pae pae/, and also the average output across the three peaks as well as the standard deviation. All results were then converted to pressure, in cm H20, following calibration procedures.

Visual-perceptual measures of the larynx were based on videoscopic views. For each subject, approximately five seconds of video segments from each of three measurement sessions (pre-treatment, post-placebo/control, and post-hydration) were presented without the sound track, successively and in random order, twice, to four different judges individually. The judges included two otolaryngology residents at our university and two speech pathologists with extensive experience in videoscopic evaluation of the larynx. None of the judges were involved in any other phase of the experiment and all were uninformed about subjects' specific conditions at the time of viewing. None had prior experience with the particular rating scale we used. The instruction was to rate each video segment on a 5-point scale, on which 1 = "healthy vocal folds", 2 = "mild nodules/polyps or related", 3 = "moderate", 4 = "moderately-severe", and 5 = "severe nodules/polyps or related". ("Related" was intended to refer to visually appreciable tissue characteristics that may accompany nodules and polyps, for example erythema.) Thus, for each segment the raters provided a single, global severity rating.

Auditory-perceptual measures were based on tape recordings of a reading passage, "A Man and his Boat". Recordings were made in a quiet room with a microphone-to-mouth distance of approximately 7 in, and were monitored for a constant average intensity within and across subjects, as indicated by the VU meter of the recording apparatus. After all subjects had completed the protocol, three audio passages for each subject (pre-treatment, post-placebo/control, and post-hydration passages) were presented successively, in random order, twice, to four different listeners individually in free-field in a quiet room. Two of the listener-judges were speech pathologists with extensive experience in voice therapy, and two were voice teachers (faculty) in The University of Iowa Department of Music. All of the listener-judges were blind to subjects' specific conditions at the time of the ratings. None were involved in any other phase of the experiment, and none had prior experience with the auditory-perceptual rating scale that was used. The instruction was to rate each audio passage on a 5-point scale on which 1 = "healthy voice", 2 = "mildly impaired voice", 3 = "moderately impaired voice", 4 = "moderate-severely impaired voice", 5 = "severely impaired voice". Thus, as for the visual-perceptual ratings, a single, global severity rating was provided for each segment. We preferred this approach to multidimensional scaling procedures because in this

study, for the auditory-perceptual rating we were interested in the bottom-line clinical question: "How good/bad do the subjects sound?"

For acoustic measures, subjects sustained the vowels /i/, /a/, and /u/ for three seconds each, at a comfortable (conversational) pitch. Signals were digitized at 50,000 samples per second and were stored for later analysis. For analysis, for each vowel three different 300-ms samples were selected from an earlier, an intermediate, and a later portion of the vowel. For each sample, the analysis indicated jitter (average cycle-to-cycle deviation in the period of the waveform) in ms[3], shimmer (average cycle-to-cycle deviation in the amplitude of the wave) in percent, and signal-to-noise (SN) ratio in dB, for each of three adjacent 100-ms tokens. Tokens indicating error messages greater than 10 were discarded. For reasons that are unclear at this point, it turned out that a large number of tokens for /i/ and /u/ were discarded because of large error messages, 57/162 and 58/162, respectively. The result was a biased data set for these vowels, because signals with poorly identifiable long-term frequencies and therefore potentially high jitter appeared to generate the highest error messages. For this reason, only the results for the vowel /a/, for which comparatively few analyses were discarded (34/162), were retained and reported.

Equipment. A Casio model portable keyboard was used in the estimation of average speaking pitch and semitone pitch range, which were relevant for PTP trials. Figure 1 shows a schematic of the remaining equipment used to extract voice measures. PTPs were collected using a 15-gage translabial stainless steel blunt needle attached to a pressure transducer embedded in a circumferentially vented face mask. The pressure transducer was connected to a Glottal Enterprises (GE) MS100 - A2 pressure/flow measurement system. Pressure signals from the main output of the GE device were routed to a Hewlett Packard 3466A Digital Multimeter for on-line calibration purposes, and in parallel, to a Hewlett Packard 350D attenuator used to attenuate signals that exceeded the output capacity of the data collection system. Signals from the attenuator were conducted to a Gateway2000 386 computer, where they were digitized using Hypersignal software (27). Subsequent analyses of the signals were conducted using a program that we developed, already described under "Voice and laryngeal measures".

Audio recordings of spoken voice were collected using an AKG C 460 B condenser microphone, powered by a Symetrix SX202 Dual Mic pre-amplifier. Microphone signals were connected through the pre-amplifier to a Realistic STA-785 Digital Synthesized Am/Fm Stereo Receiver/Amplifier and a Panasonic Digital Audio Tape Deck SV-3500. Audio signals were played back through Realistic Minimus 7 speakers. Sustained vowels used for acoustic analyses (jitter, shimmer, and signal-to-noise ratio) were collected using the condenser microphone, re-routed from the pre-amplifier to the Gateway computer. Vowels were digitized using Hypersignal software (27) and were later analyzed with CSpeech (28).

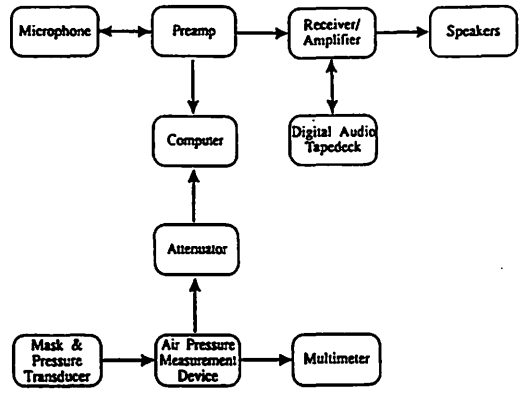


Fig. 1 Basic equipment set-up for voice measures.

[3]According to some authors, jitter is most properly reported as a proportion or a percent of the average oscillatory period, as a normalizing procedure (see for example review by Baken, 1987) (25). However, recent evidence argues against such approaches for female voices (26). Therefore, in the present report, jitter is simply reported in ms.

Videoscopic images of subjects' larynges were obtained with an R. Wolf 4450.47 90 degree rigid videoscope, a Karl Storz 9000 Mini Solid State CCD Video Camera, a Brüel and Kjaer Rhino-Larynx Stroboscope light source, Type 4914, and a NEC autocolor PM-1971A color television monitor. After the experimental protocol was completed, video segments for each subject were arranged as already described under "Voice and laryngeal measures", and were viewed by judges on a SONY CVM 1750 Trinitron video monitor.

A Panasonic WV-3060 Video Camera was used to videotape instructions that subjects received about the experimental protocol, described shortly. Instructions were later played back to subjects on a Sony PVM-1340 Trinitron color video television monitor.

Several commercial humidifiers and airfilter machines were employed to humidify and to "airfilter" experimental environments. Humidifiers included hot water vaporizers and cool-mist apparatuses. Scented candles ("wildflower") purchased at a local convenience store were also used in the placebo/control treatment.

For the hydration treatment, environmental humidity was monitored with an Abbeon Cal Cras Certified Hygrometer and Temperature Indicator (Model HTAB-176). For the placebo/control treatment, a Taylor Humidiguide hygrometer was altered so that it appeared to measure "air purity" in parts per million.

Procedure. Subjects first underwent screening procedures to ensure that the criteria for participation were met. That is, the presence of laryngeal nodules or polyps was established, risk factors were excluded, and normal hearing was confirmed. No later than 4 days following the screening, subjects initiated the experimental protocol.

For the actual protocol, subjects first viewed videotaped instructions made by a professional actress who was naive to the experimental hypotheses, and who later confirmed that she did not suspect the involvement of placebo manipulations. The videotaped instructions began with general information about the two types of treatment that subjects would receive, presented as a "hydration treatment" and a "nerve stimulation" treatment, respectively. Both were described as treatments commonly used in voice therapy, the hydration treatment in Western cultures, and the nerve stimulation treatment in Oriental cultures. An important aspect of the instructions was the assertion that both types of treatments were anticipated to make voice "easier", the hydration treatment by "moistening" vocal fold tissue and the nerve stimulation treatment by stimulating vocal nerves. The instructions ended with precise directives about the treatment that was about to be initiated, the hydration treatment for one-half of subjects and the placebo/control treatment for the other half of subjects. Videotaped instructions about initial hydration and placebo/control treatments were equivalent in terms of number of words, syntax, and ordering of ideas.

Following the delivery of videotaped instructions, subjects underwent pre-treatment testing of voice and of the larynx. That is, data were collected that would be used for measurements of phonatory effort, auditory-perceptual ratings of voice, acoustic measures of voice (jitter, shimmer, and SN ratio), phonation threshold pressure, and finally, visual-perceptual ratings of the larynx. The experimenter who elicited all voice measures was uninformed about which treatments subjects were about to initiate.

Immediately following the initial (pre-treatment) testing session, subjects initiated their first treatment in the clinic. This treatment involved a 2-hr exposure to a high humidity environment for the hydration treatment, or a 2-hr exposure to an "airfiltered/candle" environment for the placebo/control treatment. Two-hr daily exposures in the clinic as well as extra-clinical aspects of treatment continued for a total of 5 consecutive days. As already described, for the hydration

treatment the extra-clinical manipulations included elevated water intake and self-administration of a mucolytic drug. For the placebo/control treatment, extra-clinical manipulations included bilateral forefinger flexions and self-administration of a placebo drug. Across both types of treatment, subjects were instructed to restrict yelling, alcohol, smoke, and caffeine. Subjects indicated their adherence to extra-clinical aspects of the protocol on a log that we provided. The day following the last day of treatment, subjects underwent the same testing procedures as previously. Post-treatment tests of voice were conducted by an experimenter who did not know what treatment subjects had just received.

Following the first set of post-treatment tests, subjects had one day off and then returned for the next week of treatment. Again, videotaped instructions similar to the previous ones were provided about the second treatment (hydration for one-half of the subjects and placebo/control for the other half of subjects), and subjects initiated the second 5-day treatment. Again, subjects were required to observe voice conservation and hygiene measures as previously, and to log their adherence to extra-clinical aspects of the protocol. The day following termination of the second treatment, voice and laryngeal testing were conducted one last time. Again, post-treatment voice tests were conducted by an experimenter who did not know what treatments subjects had just received.

After each week of treatment, subjects filled out a questionnaire that required them to rate, on 5-point scales, the perceived effect of the preceding treatment. That is, subjects rated the extent to which they thought that the treatment affected their voice in general, and affected ease and clarity of phonation in particular (1 = "not at all"; 2 = "very little"; 3 = "somewhat"; 4 = "quite a bit"; 5 = "a lot"). The results of this questionnaire provided a check on the degree to which subjects "bought" the treatments as valid. At the end of the entire protocol, subjects filled out a final questionnaire to compare the perceived efficacy of both treatments, and to indicate suspicions about anything "fishy" in either of the treatments. The purpose of this final question was to determine if subjects frankly suspected placebo manipulations. Finally, subjects were debriefed.

Design and statistical analyses. The experiment involved a double-blind within-subject, cross-over design. The double-blind aspect of the design refers to the fact that experimenters as well as subjects were uninformed about critical aspects of the protocol: the experimenter who provided the instructions about the protocol was unaware of the experimental hypotheses, the experimenter who conducted all voice tests did not know what treatments subjects were about to receive or had just received, the experimenters who rated subjects' voices and larynges were uninformed about the ordering of subjects' treatment conditions relative to the segments presented, and subjects were unaware that placebo/control manipulations were involved. The within-subjects aspect of the design refers to the fact that all subjects received both hydration and placebo/control treatments. The cross-over aspect of the design refers to the fact that the order of treatments (hydration and placebo/control) was counterbalanced across subjects to control for order effects.

To address the general question of whether any evidence could be obtained of a specific benefit from hydration treatments, three separate binomial tests were conducted. One test assessed the likelihood of the proportion of measures that improved numerically, relative to baseline, following the placebo/control treatment, compared to a proportion of .50 (chance). A second test assessed the likelihood of the proportion of measures that improved numerically, relative to baseline, following the hydration treatment, again compared to chance ($p = .50$), and a third and final test assessed the likelihood of the proportion of measures that was numerically better following the hydration treatment as compared to following the placebo/control treatment, again compared to chance. Evidence of a benefit from hydration treatments would be provided by a significant proportion of

measures that improved following the hydration treatment, together with a significant proportion of measures with superior performance following the hydration treatment, as compared to the placebo/control treatment.

To address the more specific questions posed, separate Analyses of Variance (ANOVAs) were conducted for each of the laryngeal and voice measures. For measures of perceived phonatory effort and acoustic noise (jitter, shimmer, and SN ratio), mixed-model ANOVAs were used, with three subjects nested with each treatment order (hydration-placebo/control versus placebo/control-hydration) as a random factor, and with measurement condition (pre-treatment, hydration treatment, and placebo/control treatment conditions) and treatment order as fixed factors. For PTP, the same analysis was conducted with the exception that the initial approach included three fixed factors (pitch, measurement condition, and treatment order). A subsequent pre-planned analysis was conducted for the high pitch alone, on the basis of previous studies indicating a particular dependence of PTPs at high pitches (15, 16). Finally, for auditory-perceptual and visual-perceptual ratings, analyses involved mixed-model ANOVAs as for phonatory effort and acoustic measures, with an additional random effect of judge (x 4) in both cases. For all statistical tests, the significance level was set at $p < .05$.

Results and Discussion

The results are displayed in Tables 1-5 and in Figure 2. Tables 1, 2, and 5a-e display individual data. Table 3 provides summary information across all laryngeal and voice measures in the experiment, and Table 4 and Figure 2 indicate specific group data for the same measures.

Note that treatment order was varied across subjects as a counterbalancing procedure, to control for order effects, but was not itself a conceptual focus in the current study. For this reason and for the sake of parsimony, order effects and interactions of treatment order and measurement condition will not be reported here. Specific information about these effects can be obtained from the authors.

Adherence to protocol. According to subjects' logs, shown in Table 1, subjects generally adhered to the extra-clinical aspects of the protocol. Most subjects apparently did the required number of finger flexion exercises or more, they drank the prescribed amount of water of more, they took the recommended amounts of hydration and placebo drugs, they consumed little or no alcohol, they did not smoke, they drank less than one cup of a caffeinated beverage per day, and they generally limited loud voice use. The primary exception was Subject 4, who performed fewer than the required number of finger flexion exercises, who drank less water than the recommended amount, and who took less of the placebo and hydration drugs than recommended. Also Subject 6 took less of the placebo drug than recommended, and Subject 2 did fewer finger flexions and drank slightly less than recommended[4].

[4]Recommended quantities of treatments indicated in Table 1 take into account that because subjects initiated both placebo/control and hydration treatments in the afternoon or evening of Day 1 of participation, actual recommended quantities of finger exercises, water intake, and drug intake were less for Day 1 than for Days 2-5. Specifically, for the placebo/control treatment, a minimum of 4 sets of finger flexions was recommended for Day 1, and a minimum of 8 sets per day was recommended for Days 2-5. Thus, the average recommended amount of finger flexion sets and glasses of water across the 5 treatment days was 7.2. Similarly, 2 doses of placebo medication and 2 doses of hydration medication were recommended for Day 1, as compared to 3 doses daily recommended for Days 2-5. Therefore, the average recommended drug dosages across treatment days was 2.8 doses per day.

Of interest, Table 1 also shows that during the hydration treatment, in addition to water intake, subjects also indicated an average consumption of 0.74 glasses of other non-dehydrating fluids per day (assuming 16-ounce glasses). Adding this amount to the average amount of water, 8.3

Table 1.
Adherence to placebo and hydration protocols, based on sjubjects' logs.

PLACEBO/CONTROL TREATMENT							
TREATMENT ASPECTS	FINGER FLEXIONS	GLASSES OF NON-DEHYDRATING FLUID INTAKE, INCLUDING WATER[5] (ASSUMING 16-OZ GLASSES)	DRUG DOSES	ALCOHOL	SMOKE	CAFFEINE	LOUD VOICE
Recommended Amounts[6]	7.2 sets/day	NA	2.8 doses/day	ø drinks/ day	ø ciga-rettes/day	ø cups/day	ø min/day
Logged Amounts							
SUBJECT							
1	11.2	2.0	2.8	0	0	0.2	0
2	6.8	0.25	2.8	0.2	0	0.4	0
3	9.6	1.00	2.8	0	0	0.6	0
4	5.8	2.25	2.8	0	0	0.7	1
5	10.0	2.50	3.0	0	0	0.0	0
6	8.4	4.00	2.4	0.2	0	0.2	0
AVERAGE	8.6	2.0	2.8	0.1	0	0.4	0.2

HYDRATION TREATMENT							
TREATMENT ASPECTS	GLASSES OF WATER	GLASSES OF NON-DEHYDRATING FLUID INTAKE, IN ADDITION TO WATER (ASSUMING 16-OZ GLASSES)	DRUG DOSES	ALCOHOL	SMOKE	CAFFEINE	LOUD VOICE
Recommended Amounts[7]	7.2 16-oz glasses/ day	NA	2.8 doses/day	ø drinks/ day	ø ciga-rettes/day	ø cups/day	ø min/day
Logged Amounts							
SUBJECT							
1	11.0	1.20	2.8	0	0	0	2
2	7.0	0.45	2.8	0.6	0	0	3
3	8.4	0.90	2.8	0	0	0.4	0
4	6.2	0.00	1.8	0	0	0.4	0
5	7.4	0.90	3.0	0	0	0	0
6	10.0	1.00	3.0	0.2	0	0	0
AVERAGE	8.3	0.74	2.7	0.1	0	0.1	0.8

[5]Based on retrospective estimates.
[6]See Results section of text for explanations regarding recommended treatment quantities.

glasses per day, the total average amount of non-dehydrating fluid intake during the hydration treatment appeared to be 9.04 16-ounce glasses/day. As already indicated, during the placebo/control treatment subjects did not indicate their fluid intake because we did not wish to inadvertently encourage it during that treatment. However, retrospective estimates indicated that subjects drank a total of about 2.0 16-ounce glasses of water and other non-dehydrating fluids per day during the placebo/control treatment. Thus the average total amount of daily non-dehydrating fluid intake appeared considerably greater during the hydration treatment as compared to during the placebo/control treatment.

Subjects' impressions about the effectiveness of treatments. Table 2 displays the results of post-hoc questionnaires assessing subjects' impressions about the treatments' effectiveness. This table shows that on average, subjects thought that the placebo/control treatment improved voice symptoms "somewhat", and made voice "somewhat" clearer and easier. Subjects tended to rate the hydration treatment similarly, albeit slightly better. On average, subjects thought that the hydration treatment improved voice symptoms "somewhat", and made voice "somewhat" to "quite a bit" clearer and easier. Also indicated in Table 2, after the experiment was completed but before debriefing, 4 of the 6 subjects indicated a preference for the hydration treatment in terms of its perceived global effectiveness for voice. Conversely, one subject (Subject 4) thought that both treatments were about equally effective, and another subject (Subject 3) thought that the placebo/control treatment was more effective than the hydration treatment. Finally, not indicated in Table 2, only one subject reported a suspicion of "anything fishy" about the treatments; Subject 5 indicated she thought something was "fishy" about the placebo/control treatment. However, she denied a suspicion of frank placebo manipulations when directly questioned. Further, all subjects indicated surprise when informed that the "nerve stimulation" treatment involved placebo manipulations. Based on these observations, it appears that subjects were generally convinced that both treatments were valid. Thus the potential for unwanted subject biases seemed minimal.

Combined results across measures. Table 3 provides information that is relevant for the general question posed in this experiment, "Can any evidence be obtained of a specific hydration effect, beyond placebo/control effects, in subjects with nodules and polyps?" For each measure that was later submitted to independent statistical analysis and for both hydration and placebo/control treatments, a "+" in Table 3 indicates a numerical improvement in the average status compared to baseline, and a "++" indicates not only a numerical improvement in average status, but the best group performance across all conditions. The specific values are displayed in Tables 4 and will be discussed in detail below. However, inspection of Table 3 reveals that for all of the eight measures that were submitted to statistical analysis, the average performance was numerically better following the placebo/control condition as compared to baseline performance. Also for all measures, the best performance was obtained following the hydration treatment. Although the planned binomial tests were superfluous, their results confirmed a significant overall benefit from the placebo/control treatment, a significant overall benefit from the hydration treatment, and a significantly superior hydration effect ($p = .0000$ for each of three independent binomial tests). Therefore, at this general level, clear evidence was provided of both placebo/control and hydration effects, but a superior hydration effect.

Table 2.
Subjects' responses to post-hoc questions about treatments' effectiveness. (For first set of questions, 1 = not at all, 2 = very little, 3 = somewhat, 4 = quite a bit, 5 = a lot.)

Questions	"To what extent did this treatment improve your voice symptoms?"	"To what extent did this treatment make your voice clearer?"	"To what extent did this treatment make it easier to talk and/or sing?"
Subject	**PLACEBO/CONTROL TREATMENT**		
1	3	3	3
2	1	1	2
3	4	5	4
4	3	4	3
5	3	3	3
6	3	3	4
Average	2.8	3.2	3.2
Subject	**HYDRATION TREATMENT**		
1	4	4	4
2	3	3	5
3	3	3	3
4	4	4	4
5	3	4	3
6	4	4	4
Average	3.5	3.7	3.8

Question: "Which of the two treatments do you think helped your voice more?"

Subject	
1	Hydration
2	Hydration
3	"Nerve stimulation" (Placebo/Control)
4	Hydration - "Nerve stimulation" (Placebo/Control)
5	Hydration
6	Hydration

Measures of phonatory effort. Figure 2 and Table 4 show that prior to treatment, on average subjects rated phonatory effort as relatively high. Group effort measures improved following the placebo/control treatment, but were best following the hydration treatment. A mixed-model ANOVA indicated a reliable main effect of measurement condition, $F(2,8) = 7.55$, $p = .01$. Post-hoc Scheffe' paired comparisons confirmed that improvements from the pre-treatment condition to the post-hydration treatment condition were reliable, but improvements from the pre-treatment condi-

tion to the post-placebo/control condition were not. That is, there was a statistically significant improvement with the hydration treatment but not with the placebo/control treatment. However, an interpretational problem is posed because paradoxically, the paired comparison between post-hydration and post-placebo/control conditions also failed to confirm a significant difference.

Table 3.
Combined Results. ("+" Inidcates Improvement in Average Performance Compared to Baseline; "++" Indicates Best Overall Performance across Baseline Post-Placebo/Control and Post-Hydration Conditions)

MEASURE	POST-PLACEBO/CONTROL	POST-HYDRATION
Effort	+	++
PTP: High pitch	+	++
All pitches	+	++
Visual Rating	+	++
Auditory Rating	+	++
Acoustic Analyses:		
Jitter	+	++
Shimmer	+	++
SN Ratio	+	++

Although the group data pose some interpretational problems, the individual data displayed in Table 5a provide support for a specific hydration effect, and weaker evidence of a placebo/control effect. This table shows that hydration effects, superior to placebo/control effects, were consistently obtained across subjects. For all subjects, including the subject who thought that the placebo/control treatment was generally more effective (Subject 3) and the subject who thought that both treatments were equally effective (Subject 4), measures of phonatory effort were best following the hydration treatment. Effort measures were also better following the placebo/control treatment as compared to the pre-treatment condition for all subjects except one. For this subject (Subject 5), the effort measure was worse following the placebo/control treatment as compared to baseline. Considering group and individual data together, the results indicate a benefit from hydration treatments and more weakly, from placebo/control treatments.

Phonation threshold pressures. Assuming that perceived phonatory effort is at least partially regulated by phonatory subglottal pressures, qualitatively similar results should be obtained for PTPs as for measures of perceived effort. That is, PTPs should be highest in the pre-treatment condition, and lowest following the hydration condition. In fact, as indicated in Table 4, group data collapsed across pitches showed a trend in this direction.[7] Also the average data for the high pitch,

[7]Because in many cases PTPs were unstable across trials, average PTPs were calculated by weighing PTPs from single trials by the reciprocal of the interatrial variance. In this way, more stable performances were weighted relatively more in the calculation of the average performance, allowing for better estimates of true PTPs than unweighted procedures.

which in previous experiments was most sensitive to changes in PTP with hydration treatments (15, 16), showed the same trend (Table 4 and Figure 2). However, independent two mixed-model ANOVAs failed to confirm a main effect of treatment for either all pitches collapsed together,

Table 4.
Average performance before treatment, and following placebo/control and hydration treatments.

	CONDITION				
MEASURE	Pre-treatment	Placebo/Control	Hydration	p[1]	Norm
Effort	143.33 (26.58)	126.67 (17.51)	94.17 (29.23)	.01	100[2]
PTP(cm H_2O)				.12[3]	
High pitch	11.92 (3.13)	10.91 (6.51)	8.64 (3.30)	.09[4]	6.73[5]
Conversational pitch	3.28 (7.48)	3.88 (3.53)	4.01 (3.62)		3.34[5]
Low pitch	4.40 (6.80)	2.93 (4.85)	4.14 (3.29)		3.52[5]
Average	6.53	5.90	5.60		
Visual-perceptual rating	2.64 (0.79)	2.26 (0.46)	2.03 (0.80)	.0001	1.0[6]
Auditory-perceptual rating	2.78 (1.06)	2.65 (0.87)	2.50 (0.82)	.10	1.0[6]
Acoustic noise /a/					
Jitter (ms)	0.016 (0.008)	0.014 (0.007)	0.010 (0.004)	.001	-[7]
Shimmer (%)	3.33 (1.46)	2.85 (1.22)	2.78 (1.15)	.001	
SN ratio (db)	19.54 (3.65)	22.89 (3.95)	22.93 (4.00)	.001	

[1]p-values reflect the main effect of measurement condition.
[2]Norm assumed to be 100 ("comfortable"), based on informal data from Clinical Voice Lab, The Univ. of Iowa.
[3]p-value for PTP collapsed across pitches.
[4]p-value for PTP at high pitch only.
[5]Norms for PTP based on Verdolini-Marston et al., 1990 (15) for adult female subjects in a "no-treatment" condition.
[6]Norm assumed to be 1 ("healthy voice" or "healthy vocal folds").
[7]Formalized norms for our equipment and software are unavailable.

$F(2,32) = 2.27$, $p = .12$, or for the high pitch alone, $F(2,8) = 3.222$, $p = .09$. Similarly, inspection of the data for individual subjects (Table 5c) reveals inconsistent patterns of relative benefits from hydration and placebo/control treatments across subjects, for all pitches.

Although an effect of treatment condition was not clearly confirmed, in the analysis assessing all pitches collapsed together the main effect of pitch was reliable, $F(2,32) = 64.19$, $p = .0001$. The greatest PTPs occurred at the high pitch, as reported in previous studies (15, 16).

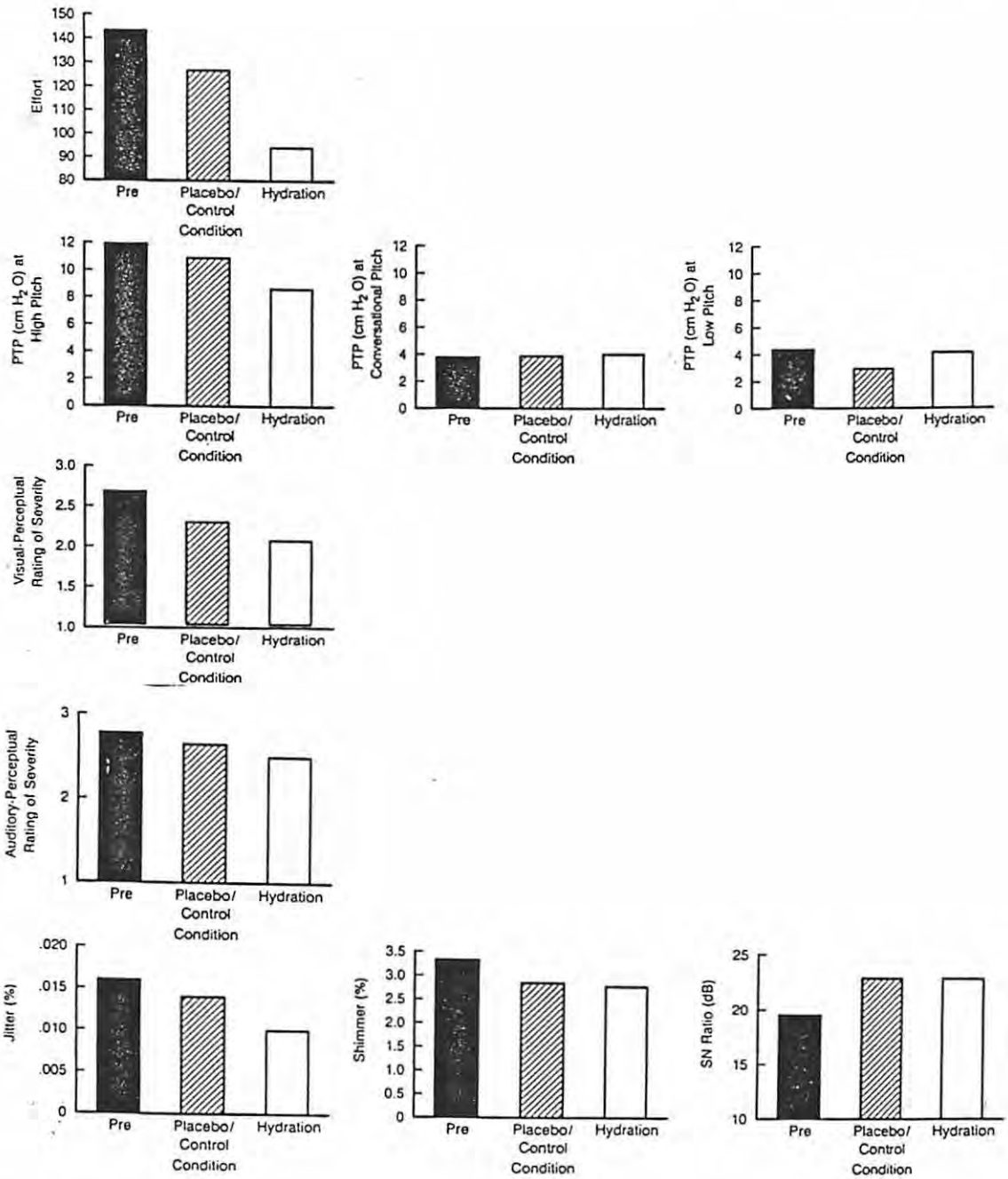


Figure 2. Average scores for phonatory effort, PTP at high, conversational, and low pitches, visual-perceptual ratings of the larynx, auditory-perceptual ratings of voice, and acoustic analyses (jitter, shimmer, and SN ratio).

Table 5.
Individual Data
5a: Perceived Phonatory Effort (100 = "comfortable effort")

SUBJECT	CONDITION		
	PRE-TREATMENT	POST PLACEBO/CONTROL	POST HYDRATION
1	150	130	120
2	150	120	50
3	150	100	85
4	180	140	125
5	100	150	75
6	130	120	110

5b: Phonation Threshold Pressure (cm H2O)

SUBJECT	PITCH	CONDITION		
		PRE-TREATMENT	POST PLACEBO/CONTROL	POST HYDRATION
1	High	9.20	10.90	9.20
	Conversational	2.03	3.33	2.77
	Low	2.52	2.91	2.54
2	High	17.90	8.51	9.28
	Conversational	12.10	5.02	5.32
	Low	7.31	3.43	4.39
3	High	16.52	11.33	11.51
	Conversational	6.57	3.72	3.72
	Low	7.50	6.22	4.87
4	High	10.75	6.72	7.50
	Conversational	3.83	3.77	4.45
	Low	3.27	2.33	4.59
5	High	11.43	12.86	11.23
	Conversational	4.07	4.78	4.26
	Low	4.03	4.18	4.00
6	High	12.98	11.39	9.34
	Conversational	7.19	5.88	4.40
	Low	6.70	5.08	4.32

5c: Visual-Perceptual Rating (5 = "severe")

SUBJECT	CONDITION		
	PRE-TREATMENT	POST PLACEBO/CONTROL	POST HYDRATION
1	2.0	2.4	1.5
2	3.2	2.9	3.4
3	3.4	2.3	2.1
4	3.5	1.5	1.3
5	1.7	2.1	1.3
6	2.2	2.3	2.5

5d: Auditory-Perceptual Rating (5 = "severe")

SUBJECT	CONDITION		
	PRE-TREATMENT	POST PLACEBO/CONTROL	POST HYDRATION
1	2.7	2.8	1.7
2	3.9	3.9	3.9
3	3.3	3.1	2.7
4	3.8	2.8	2.8
5	1.5	1.8	1.9
6	1.6	1.6	2.0

5e: Acoustic Measures (Jitter in ms; Shimmer in %; SN Ratio in dB

SUBJECT	MEASURE	CONDITION		
		PRE-TREATMENT	POST PLACEBO/CONTROL	POST HYDRATION
1	Jitter	0.0111	0.0079	0.0101
	Shimmer	3.0844	1.6667	2.9833
	SN Ratio	20.4811	25.6967	21.8500
2	Jitter		0.0180	
	Shimmer	*	4.1900	*
	SN Ratio		17.3800	
3	Jitter	0.0213	0.0156	0.0122
	Shimmer	5.1122	3.1244	2.4717
	SN Ratio	17.1978	20.7189	22.1567
4	Jitter	0.0213	0.0250	0.0124
	Shimmer	3.0189	4.6513	3.5556
	SN Ratio	16.6311	17.7125	19.4256
5	Jitter			0.0130
	Shimmer	*	*	3.4800
	SN Ratio			20.9750
6	Jitter	0.0072	0.0076	0.0054
	Shimmer	1.5150	2.5356	1.3967
	SN Ratio	26.0050	28.0167	29.2056

* Data not available due to high error messages in analyses.

Note that as indicated in Table 4, at conversational and low pitches average PTPs were generally similar, or only somewhat larger across measurement conditions for the pathological subjects in the present study, as compared to average PTPs for vocally normal female adults in a "no-treatment" condition described by Verdolini-Marston and colleagues (16). Conversely, mean PTPs were considerably larger for our subjects at high pitches, as compared to normals. As importantly, already indicated by inspection of individual data, PTPs were remarkably unstable across our pathological subjects. Standard deviations calculated from average PTPs for each subject in each

measurement time by pitch condition ranged from 3.12 - 7.48 cm H2O in the present study, as compared to standard deviations that ranged from 0.24 - 1.13 cm H2O for the healthy adult females in the previous study (16).

In summary, although a specific benefit from hydration treatments and more weakly, for placebo/control treatments was shown for measures of perceived phonatory effort, and despite numeric trends for similar effects for PTPs, statistical analyses fell short of confirming the trend for PTPs. This result points to the possibility that the perceptual effort measures we used may be more sensitive than aerodynamic measures for detecting long-term physiological changes in vocal fold tissue with hydration treatments. The poor sensitivity of PTP measures in this experiment appears at least partly attributable to extremely large inter-subject variability in PTPs across our pathological subjects.

Visual-perceptual ratings. Average visual-perceptual ratings of the larynx improved following both placebo/control and hydration treatments compared to baseline, but the greatest improvements were obtained following the hydration treatment. Figure 2 and Table 4 show this result. On a scale on which 1 indicated "healthy vocal folds", 2 = "mild nodules/polyps or related", 3 = "moderate", 4 = "moderately-severe", and 5 = "severe nodules/polyps or related", the average rating of larynges before treatment was "mild to moderate". Following placebo/control treatment, the average rating was somewhat better, and following hydration treatment, the average rating was the best ("mild"). In a mixed-model ANOVA, the main effect of measurement condition was statistically significant, $F(2,116) = 14.68$, $p = .0001$, but the effect of judge was not, $F(3,15) = 0.061$, $p = .98$. Thus, laryngeal appearance was affected by treatment, and laryngeal ratings were consistent across judges. Post-hoc Scheffe' paired comparisons confirmed that post-hydration ratings were significantly better than pre-treatment ratings, but post-placebo/control ratings were not. This result implies a benefit from hydration treatments, but not from placebo/control treatments. However, also as for measures of perceived phonatory effort, an interpretational problem is posed because post-hydration and post-placebo/control ratings did not reliably differ. Again, some support in favor of a specific hydration effect is provided by the examination of individual data, in Table 5c. These data indicate that for the majority of subjects, larynges appeared the best following the hydration treatment (4/6), whereas the best laryngeal rating was obtained following the placebo/control treatment for only one subject. Therefore, combining group and individual data, some evidence of a specific benefit from hydration treatments was obtained for visual-perceptual ratings.

Auditory-perceptual ratings. The findings for visual-perceptual measures leads to the expectation of qualitatively similar results for auditory-perceptual ratings, which presumably depend on laryngeal status. Figure 2 and Table 4 indicate a gross confirmation of this expectation, at least at a numeric level. On a scale on which 1 = "healthy voice", 2 = "mildly impaired voice", 3 = "moderately impaired voice", 4 = "moderate-severely impaired voice", and = "severely impaired voice", the average pre-treatment rating was "moderate". Following the placebo/control treatment the average rating was somewhat improved, and following the hydration treatment the average rating was the best ("mild to moderate"). Despite this trend, in a mixed-model ANOVA the main effect of measurement condition fell short of significance, $F(2,116) = 2.31$, $p = .10$. In fact, inspection of individual data in Table 5d reveals inconsistent results. As importantly, and probably related, the random effect of judge was statistically significant, $F(3,15) = 7.3$, $p = .003$, meaning that auditory-perceptual ratings differed across judges.

Thus, auditory-perceptual ratings showed a trend for improvements in voice quality following placebo/control and hydration treatments compared to baseline, and the greatest improvements

following hydration treatments, parallel to the results for the visual-perceptual measures. However, the trend for the auditory-perceptual data fell short of statistical significance at least partly due to differences in ratings across judges, and also due to inconsistent results across subjects.

Acoustic measures. Average acoustic measures for the vowel /a/ are shown in Figure 2 and in Table 4. Trends for this vowel were consistent with patterns obtained for most other measures in the experiment. Average jitter, shimmer, and signal-to-noise (SN) ratio were all worst before treatment, they improved somewhat following the placebo/control treatment, and they were best following the hydration treatment. Mixed-model ANOVAs indicated reliable main effects of measurement condition for all three measures, $F(2,100) = 7.27$, $p = .001$ for jitter, $F(2,100) = 7.71$, $p = .001$ for shimmer, and $F(2,100) = 70.78$, $p = .0001$ for the SN ratio. Post-hoc paired comparisons confirmed that for jitter, all of the two-way comparisons corresponded to significant differences. For shimmer, post-hydration measures were reliably different from pre-treatment measures, but none of the other comparisons indicated reliable differences. Again, this result is a paradoxical one that is problematic to interpret. For the SN ratio, post-hydration and post-placebo/control measures were both significantly different from pre-treatment measures, but were not different from each other. This result indicates an effect of treatment in general, but not a specific hydration effect.

Examination of individual data was possible for 4 subjects (for the remaining 2 subjects, so many tokens were discarded because of high error messages that examination of their data across experimental conditions was precluded). These data are displayed in Table 5e. For 3 of the 4 subjects, the best jitter values were obtained following the hydration treatment. Only 2 subjects obtained the best shimmer values following hydration, and 3 subjects had the best SN ratios following hydration.

In summary, some evidence of hydration effects emerged from acoustic analyses of the vowel /a/. Jitter clearly improved with both placebo/control and hydration treatments, but the greatest improvements were obtained with the hydration treatment. For this measure a reliable hydration effect, superior to a placebo/control effect, was supported by both group and individual data. A similar trend was seen for shimmer, although a statistically ambiguous result was obtained and a specific benefit from hydration treatments was not supported by individual data. Finally, SN ratios improved with both placebo/control and hydration treatments, but there was little evidence of a specific hydration effect for this measure.

General Discussion

The general question asked in the present experiment was: Can any evidence be obtained of a specific benefit from hydration treatments for voice and for laryngeal status, beyond placebo/control effects, in subjects with laryngeal nodules or polyps? The results from the present study indicate an affirmative answer to the general question. Considering the combined results, a series of analyses confirmed that general benefits were obtained from both the placebo/control and hydration treatments, but overall, significantly greater benefits were obtained from the hydration treatment.

In addition to the general question, the specific, theoretically-motivated questions in this experiment were: (a) Do hydration treatments reduce phonatory effort, and related, subglottal pressures required to initiate and sustain phonation (PTPs)? and (b) Do hydration treatments promote laryngeal healing, and related, improvements in auditory-perceptual and acoustic evaluations of dysphonia? The results from this study indicate partially, but not unequivocally positive answers to these questions.

At a numeric level, measures of perceived phonatory effort, phonation threshold pressures, laryngeal appearance, auditory-perceptual status, and acoustic status for the vowel /a/ all improved following both placebo/control and hydration treatments, but improved the most following the hydration treatment. For one measure (jitter), statistical analyses unambiguously confirmed both a benefit from the placebo/control treatment and a superior benefit from the hydration treatment. For three additional measures (perceived phonatory effort, laryngeal appearance, and shimmer), the results of statistical evaluations were paradoxical and thus posed interpretational problems. However, for two of these same measures, (phonatory effort and laryngeal appearance), the impression of a benefit from hydration treatments was supported by superior performance in the hydration condition for all, or almost all, subjects. Finally, for one last measure (SN ratio), statistical analyses indicated a significant effect of treatment, but not a specific hydration effect.

Discussing the specific results in greater detail, it is somewhat perplexing that measures of perceived phonatory effort more clearly revealed a benefit from hydration treatments than PTPs did. It is interesting to ask why there should be a discrepancy in the sensitivity of these measures, which intuitively seem related. This question is particularly pertinent because clear benefits from hydration treatments were obtained for PTP measures in a previous study by Verdolini-Marston and colleagues using human subjects (16). The question is also pertinent because measures of pulmonary effort were strongly related to perceptual measures of effort in previous work by Colton and Brown (20).

With regard to the failure to clearly replicate previous hydration effects for PTP, reported by Verdolini-Marston et al. (16), the most obvious explanation is that intersubject variability in PTPs was extremely large for the pathological subjects in the present study, as compared to small variability in PTPs in the previous study using laryngeally healthy subjects. The large variability in PTPs in the current study clearly obscured the detection of possible treatment effects. Further, the apparent sensitivity of PTPs to treatment effects in the previous study may have been artefactually enhanced by unintended experimenter and/or subject biases in that study.

A series of other factors may have also contributed to the poor sensitivity of PTP measures to treatment in the present experiment, but relatively better sensitivity of perceptual effort measures, despite reports of good agreement between similar measures described by Colton and Brown (20). In the study by Colton and Brown, perceptual and physiological measures reflected perceived effort and pulmonary effort at the same moment in time, across a wide range of scaler values (minimum to maximum). In the present study, perceptual and pulmonary measures reflected effort across different time-windows, and for different scaler values. Perceptual measures reflected average perceived effort over a 24-hr period, including the time-period surrounding the final treatment, whereas PTPs reflected minimum pulmonary effort during a limited time-period, one day following treatment, when treatment effects may have partially worn off. Perhaps these perceptual and physiological measures clearly covary over a wide range of scaler values (minimum to maximum) when they reflect the same time-frame, as in the study by Colton and Brown, but the measures fail to covary when they reflect different time-frames and different scaler values.

To summarize, the question of why perceptual measures of effort were more sensitive to hydration effects than PTP measures has a series of possible explanations. However, we consider the large variability in PTPs for the pathological subjects the primary, if not the unique reason.

Despite some evidence of therapeutic benefits from hydration treatments in the present study, from the foregoing results several caveats about clinical applicability should be emphasized. First, although the study as a whole indicates benefits from hydration treatments in subjects with nodules

and polyps, further investigations are required before confident conclusions can be drawn about the specific benefits that may be obtained. In particular some of the measures (PTPs, auditory-perceptual measures, and shimmer) indicated inconsistent reactions to hydration treatments across subjects. Therefore, the ways in which hydration effects may be manifested and the extent to which they are manifested appear at least partly regulated by individual subject variables. This result is not a surprising one, and further systematic investigation of subject variables is important for clinical reasons.

Second, because the study was designed to maximize the likelihood of detecting hydration effects, the hydration treatments provided were intensive and were not typical of the average clinical treatment, particularly with regard to the humidity exposure. Subjects were exposed to extremely high humidities (90%-100%) for relatively brief durations (2 hr/day). Clinically, humidity exposures generally involve lower humidification levels distributed over longer periods. This difference between the experimental procedure and common clinical practice, or even feasible clinical practice, requires us to be cautious about generalization of any aspect of the results, even the more general ones, to the typical clinical situation, at this point.

Third, the fact that hydration effects were obtained in the present experiment, despite counterbalancing of treatment orders, indicates a relative wearing off of hydration benefits following the placebo/control treatment, when the hydration treatment was administered first. The implication is that based on the present results, benefits that may be obtained with hydration treatments may persist only so long as hydration treatments continue to be administered.

Fourth, and quite importantly, although benefits in voice and in laryngeal status were obtained following hydration treatments in particular, for no subject did laryngeal lesions and related voice problems entirely resolve with treatments, at least not over the limited time-period of the experimental protocol. A reasonable conclusion that most clinicians would support is that although some evidence of a hydration benefit was provided by this study, hydration treatments should be viewed as supportive, but not necessarily primary and certainly not independent, in the treatment of nodules and polyps.

Although we tentatively ascribe the apparent hydration effects detected in this experiment to reductions in vocal fold tissue viscosity, other explanations are possible. In particular, when lesioned cells are present as with nodules and polyps, fluid extravasions can occur into interstitial space (see for example Robbins, Cotran, and Kumar, 29). It cannot be excluded that in the present study, extravasion increased with fluid intake, in the hydration condition. In this case, a relative "planing" of laryngeal tissue could have occurred, reducing the geometric discrepancy between the lesions and surrounding vocal fold tissue. This possible response could explain the improvements obtained with the hydration treatment, if not wholly, at least in addition to the proposed changes in vocal fold tissue viscosity. However no speculations can be made about the mechanisms underlying hydration effects on the basis of the present results, because the study was not designed with this purpose in mind.

In conclusion, the results from this study provide preliminary evidence that intensive hydration treatments may provide some benefits in the treatment of vocal nodules and polyps. Further work is needed to confirm the present results in general, to clarify which particular benefits may occur and to what degree they may occur particularly considering subject variables, to determine the critical aspects of treatment and the critical dosages for therapeutic effects, and to investigate the physiological and biomechanical mechanisms.

Acknowledgments

This research was partially supported by Grant No. P60 DC00976-02 from the National Institute on Deafness and Other Communication Disorders. The following persons are also acknowledged: Kice Brown, Jon Lemke, Frank Karamagianis, Martin Milder, Hughlett Morris, Jerry Moon, Kelvin Lehrman, Ann Fennell, Darla Max, Young Min, Linnie Southard, Colleen Gardner, and in particular, Becky Blum.

References

(1)Bradley M. Prevention and correction of vocal disorders in singers. NATS Bull 1980;36(5):38-41.
(2)Lawrence VL. Handy household hints: To sing or not to sing. NATS Bull 1981;37(3):23-25.
(3)Verdolini K. Practice good vocal health and prevent those voice disorders. Choristers Guild LETTERS 1988;(2):40-44.
(4)Sataloff RT. The professional voice: Part III. Common diagnoses and treatments. J Voice 1987;1(3):283-292.
(5)Sataloff RT. The professional voice: Part I. Anatomy, function, and general health. J Voice 1987;1(1):92-104.
(6)Bastian RW, Lawrence VL. Hoarseness in singers. NATS Bull 1984;40(3):26-27.
(7)Sataloff RT, Baron BC, Brodnitz FS, Lawrence VL, Rubin W, Spiegel J, Woodson G. Discussion: Acute medical problems of the voice. J Voice 1988;2(4):345-353.
(8)Verdolini-Marston K. Evaluation and treatment of voice disorders. Paper presented at the Conference of The Art and Science of the Professional Voice, St. Louis, Missouri, May 1990.
(9)Verdolini-Marston K. The most common vocal injuries: Nodes and polyps and their causes. Voice Speech Train Assoc Inc. 1991;5(1):1,7.
(10)Titze IR, Talkin DT. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. JASA 1979;66(1):60-74.
(11)Fung YC. Biomechanics: Mechanical properties of living tissues. NY: Springer-Verlag, 1981.
(12)Physicians' Desk Reference, Edition 41. Oradell, NJ: Publisher Edward R. Barnhart, 1987.
(13)Titze IR. Heat generation in the vocal folds and its possible effect on vocal endurance. In: Lawrence VL, ed. Transcripts of the Tenth Symposium: Care of the Professional Voice. Part I: Instrumentation in Voice Research. NY: The Voice Foundation, 1981:52-65.
(14)Titze IR. The physics of small-amplitude oscillation of the vocal folds. JASA 1988;83(4):1536-1552.
(15)Finkelhor BK, Titze IR, Druker DG. The effect of viscosity changes in the vocal folds on the range of oscillation. J Voice 1988;1(4):320-325.
(16)Verdolini-Marston K, Titze IR, Druker DG. Changes in phonation threshold pressure with induced conditions of hydration. J Voice 1990;4(2):142-151.
(17)Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughn C. Objective assessment of vocal hyperfunction: An experimental framework and initial results. JSHR 1989;32(2):373-392.
(18)Wright HN, Colton RH. Some parameters of vocal effort. JASA 1972;51:141.
(19)Wright HN, Colton RH. Some parameters of autophonic level. Presented at the American Speech and Hearing Association, November 1972.

(20)Colton RH, Brown WS. Some relationships between vocal effort and intra-oral air pressure. Presented at the 84th meeting of the Acoustical Society of America, Miami, Florida, November 1972.
(21)Rothenberg MR. Measurements of airflow in speech. JSHR 1977;20(1):155-176.
(22)Smitheran JR, Hixon TJ. A clinical method for estimating laryngeal airway resistance during vowel production. JSHD 1981;46(2):138-146.
(23)Löfqvist A, Carlborg B, Kitzing P. Initial validation of an indirect measure of subglottal pressure during vowels. JASA 1982;72(2):633-634.
(24)Holmberg E, Perkell J, Hillman R. Methods for using a non-invasive technique for estimating glottal functions from oral measurements. Cambridge, Massachusetts: Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1987:47-48.
(25)Baken RJ. Clinical measurement of speech and voice. Boston: Little, Brown, 1987.
(26)Orlikoff RF, Baken RJ. Consideration of the relationship between the fundamental frequency of phonation and vocal jitter. Folia Phoniatr 1990;42(1):31-40.
(27)Hypersignal Users Manual. Dallas, Texas, March 1989.
(28)Milenkovic PH. CSpeech Core User's Manual: CSpeech Version 3. Madison, Wisconsin, 1990.
(29)Pathologic basis of disease, Third Edition. SL Robbins, RS Cotran, & W. Kumar (Eds.), Philadelphia: WB Saunders Company, 1984.

NCVS Status and Progress Report - 3
December 1992, 201-207

Targeted Problems Treated by Vocal Fold Injection of Collagen

Charles N. Ford, M.D.
University of Wisconsin-Madison
Department of Surgery, Division of Otolaryngology and Middleton VA Hospital
Diane M. Bless, Ph.D.
University of Wisconsin-Madison
Department of Communicative Disorders, The Waisman Center, and Department of Surgery, Division of Otolaryngology

Introduction

Injectable cross-linked bovine collagen has been shown to be effective in the management of glottic insufficiency. Previous studies have demonstrated that collagen injection is uniquely helpful in the management of atrophic vocal folds, minimal glottic insufficiency, scarred vocal folds, and focal defects or adynamic vocal fold segments[1]. With the proliferation of new treatment options for glottic insufficiency, it seemed appropriate to target a specific population who were not ideal candidates for other forms of treatment. Since collagen injection had been shown safe and effective for vocal fold augmentation and yet remained an off-label application, this National Center for Voice and Speech study was conducted at the University of Wisconsin Clinical Science Center under an investigator device exemption and supported largely by the Federal Drug Administration Orphaned Product Division.

The primary aim of this project was to assess a treatment option for patients with glottic insufficiency for whom other forms of treatment were considered inadequate or contraindicated. Unlike earlier collagen injection studies[2,3], we excluded patients who were suitable candidates for other forms of treatment. The study was designed to allow rigorous voice assessment so that efficacy or failure of treatment could be objectively documented. A related question to be addressed was the relative efficacy of collagen in the treatment of several different vocal fold pathologic conditions. Although safety of collagen injection has be previously reported[1,4,5] another aim of this study was to further document the safe use of this implant material in the human larynx.

Materials and Methods

Patients were invited to participate in this study who presented to the University of Wisconsin Otolaryngology and Voice Clinic with glottic insufficiency that was not likely to recover spontaneously and who were not ideally suited for other, more conventional, forms of therapy. The spectrum of pathology included presbylaryngis, vocal fold paralysis or paresis associated with atrophy, and scarred vocal folds secondary to trauma, surgery, or Teflon granuloma. A total of 45 patients were treated; there were 25 females and 20 males. The mean age at the time of initial treatment was 57.5 years, with a range of 26 to 87 years. The mean length of follow-up was 12.0 months. All patients underwent a preliminary voice and otolaryngologic examination. Suitable subjects were enrolled after they reviewed the protocol and signed an informed consent that indicated the risks and need for careful follow-up assessment. A skin test was placed on the forearm using 0.1 cm^3 Zyderm Collagen (Collagen Corporation, Palo Alto, California) and read over a six-week period. A second voice evaluation was done just prior to treatment and used as a baseline for postoperative comparison. Zyplast Collagen (Collagen Corporation, Palo Alto, California) was used for injection through the Oro-tracheal Injector Device (Xomed-Treace, Jacksonville, Florida) as previously described[2,6]. There were a total of 70 injections performed on 45 patients. Eighteen patients received multiple injections. The patients that required repeated injections included several patients with bilateral pathology such as presbylaryngis where the initial injection was deliberately minimal to assess response and to determine if there were any adverse effects on the voice. The strongest indication for multiple injections was scarring of the vocal fold. This was due to the fact that it is difficult to inject a substantial amount of any substance into scar tissue. Since injectable collagen softens scar tissue over time[7], we often injected initially to achieve some filling and to initiate the softening process. Repeated injections were then made easier and it was often possible to improve glottic closure.

The patients presented with a variety of pathologic vocal fold problems. In attempting to categorize subjects based on the histopathology we divided the group into 3 pathological groups:

(1) SCARRED. This included all patients with scarred larynges from trauma, ablation, Teflon fibrosis, etc. regardless of whatever else they had wrong.

(2) BILATERAL VOCAL FOLD PATHOLOGY. This included patients with Parkinson's disease, presbylaryngis, spasmodic dysphonia, and so forth, even if they had concurrent unilateral paralysis.

(3) UNILATERAL VOCAL FOLD PATHOLOGY. This group included only those patients with one vocal fold involved and no scarring present.

Assessment

Assessment and documentation was accomplished using multiple parameters. Four general categories of voice assessment were used to measure treatment success: 1) Subjective patient self-assessment of voice quality and effort of phonation, 2) perceptual judgments of voice recordings by an impartial panel of experts, 3) laboratory data based on acoustic signal analysis and aerodynamic measures during voice production, and 4) studies of the appearance of vocal fold vibration by videostroboscopy.

Audio tape recordings were made in a sound-controlled booth using a Nagra tape recorder, as previously described[2]. The test sample recorded included sustained vowels /a/ and /i/ produced at normal pitch and loudness and the sentence "The blue spot is on the key" produced at conversational level. These samples were used for both the acoustic and perceptual analysis.

Perceptual Judgements. Recordings of the speech samples were coded in pairs with preoperative and postoperative samples scrambled for the listener. Judges were asked to determine which sample was the better voice or if voices were the same. Samples were repeated and 10% were foils to test reliability. Foils were used to determine validity of the perceptual ratings. The ratings were based on perceptual qualities but there was no attempt to individually quantify parameters of breathiness, roughness, weakness, or straining.

Acoustic and Aerodynamic Measures. Samples were digitized on an IBM-PC computer using a locally developed software program (C-Speech). Jitter, shimmer, and signal-to-noise ratio (SNR) were derived using the least mean square fit of a waveform model to the speech waveform. Vocal function measures were recorded using a face mask attached to the Nagashima PS-77 Phonatory Function Analyzer to obtain maximal intensity (dB) and mean airflow (ml/sec). Transglottic airflow was assessed during production of /a/ and /i/ during a variety of phonatory tasks. High airflows were considered an indicator of poor laryngeal closure, as might occur in glottic insufficiency. Low airflows generally signified a competent glottis but were noted to occur also with hyperfunction of laryngeal structures, obstruction, low patient effort, or poor respiratory support.

Videostroboscopy. Vocal fold vibration was observed using a Bruel and Kjaer laryngeal videostroboscope coupled to a rigid 90° angled Wolfe endoscope and videorecorder. Recordings were used to assess ongoing treatment and to study individual cases over time. Laryngeal videostroboscopy was useful in assessing and localizing the pathology, particularly in some of the scarred larynges. These examinations were also helpful in determining how patients responded to treatment and in particular, the effect of injected collagen on vocal fold motion and closure. The images were used for videoprint documentation and were helpful in assessing the need for additional therapy. Stroboscopic recording were not helpful for objective assessment of results because it was impossible to do a blinded study; the injected vocal fold was usually evident due to displacement and discoloration from the implant, so the judge could usually tell which sample had undergone treatment.

Data Management. In cases where pre-injection data were available from more than one date, the date closest to injection was selected. Post-injection data were selected from the latest date the patient had been evaluated following the previous injection. When collagen injection was done after thyroplasty, Teflon removal, or reinnervation surgery, baseline measures were taken after the previous surgery and before collagen injection.

Efficacy of collagen injections was addressed by asking two questions: 1) Was there a significant change in any of the variables post-injection, and 2) was there any difference between pathologic groups in response to the collagen treatment. To analyze for pre-treatment and post-treatment changes in the acoustic and aerodynamic parameters, we used paired t-tests to determine whether or not mean change was equal to zero. Differences between diagnosis groups were studied using analysis of variance. For measures with skeweddistributions (airflow, jitter, shimmer), the log of the measure was used. These transformed measures had distributions close to normal, thus making the transformation changes nearly equivalent to analyzing percentage changes rather than absolute changes.

Results

Efficacy was measured by comparing the second pretreatment assessment to the most recent follow-up assessment. By combining the patient self-assessment with the perceptual judgments, acoustic analysis, and aerodynamic measures, four general patterns of response to treatment were observed:

(1) Marked improvement: Improvement was noted immediately or within one to three months following injection.

(2) Moderate improvement: Improvement noted initially might deteriorate slightly during first three months but then voice stabilized at improved level.

(3) Slight improvement: This pattern was characterized by the patient having achieved a satisfactory voice result for a variable period (months to years) followed by rapid or gradual deterioration. Such a pattern was invariably associated with some incidental viral or systemic illness.

(4) No change: These patients achieved no clear benefit from the procedure. No patient was worse after treatment although one patient treated for a severely scarred larynx seemed worse for several weeks before improvement was noted.

An attempt was made to perform a complete voice evaluation on all patients. Unfortunately some of the tasks could not be completed by some patients because of the nature or severity of their voice disorder. One patient had severe spasmodic dysphonia in addition to glottic insufficiency so she was unable to sustain a vowel sound during pretreatment assessment; three other patients had such severe glottic insufficiency pre-treatment that they were aphonic. Under these circumstances, it was impossible to obtain a speech sample sufficient tocomplete all of the acoustic measures. For this reason, not all measures were completed on all 45 subjects; the N range for all measures was 40-45.

Perceptual ratings were completed on 44 subjects by three judges, as described. A value of -1 was given where the pre-treatment voice seemed better, 0 when there was very little perceived difference, and 1 if the post-treatment voice was clearly better. The range of ratings was therefore -3 (if all judges found the pre-treatment voice better) to 3 (if the post-treatment voice was rated better). In summary, the scores showed:

SCORE	N
-3	1
-2	3
-1	5
0	17
1	6
2	7
3	5

Perceptual Judgments

The mean score was 0.477 with a standard deviation of 1.486. P- value for mean not equal zero is 0.039.

Transglottic airflow was reduced overall for the entire group (N = 43, log mean change = -0.26707, SD 1.00701, p= 0.0893). This included some subjects with preoperative supraventricular hyperfunction.

Acoustic measures revealed favorable changes in all parameters studied:

I_{max}	N= 43	mean change = 2.9069	SD= 8.2572	p= 0.0260
SNR	N= 43	mean change = 4.2447	SD= 8.4156	p= 0.0019
I_{range}	N= 42	mean change = 2.4762	SD= 8.7073	p= 0.0726
$I_{habitual}$	N= 44	mean change = 0.8636	SD= 5.7043	p= 0.3209
F_{range}	N= 40	mean change = 3.025	SD= 9.0099	p= 0.0401
$F_{habitual}$	N= 42	mean change = 0.2227	SD= 0.3479	p= 0.0002
Jitter	N= 43	mean change =-0.7730	SD= 1.5376	p= 0.0020
Shimmer	N= 42	mean change =-0.6493	SD= 1.2672	p= 0.0019

Although there was improvement in mean values for the entire group, there was considerable individual variation. There were overall differences between pathological groups in their response to treatment. These differences are apparent by contrasting the clinical patterns of response to the underlying vocal fold pathology:

CLINICAL PATTERN OF RESPONSE =	1	2	3	4
VOCAL FOLD PATHOLOGY				
SCARRED N=12	0	7	3	2
BILATERAL N=22	4	10	4	4
UNILATERAL N=11	3	5	3	0

Analysis of each measured parameter failed to demonstrate significant differences between the pathologic groups in their response to treatment.

	Bilateral Pathology			**Scarred Vocal Fold**			**Unilateral Pathology**		
	N	MEAN	SD	N	MEAN	SD	N	MEAN	SD
AF(log)	21	-0.60	1.10	11	0.24	0.75	11	-0.14	0.85
I_{max}	21	2.43	7.97	11	-0.27	8.49	11	7.00	7.56
SNR	20	3.80	3.31	12	3.31	9.98	11	6.07	8.12
Jitter	20	-0.93	1.55	12	-0.70	1.76	11	0.56	1.35
Shimmer	20	-0.76	1.16	12	-0.44	1.31	10	-0.68	1.51

The above measures are all comparisons of pre-treatment to most recent post- treatment values and represent changes ascribed to collagen injection. The measures include transglottic airflow (AF), maximum intensity (I_{max}), signal-to-noise ratio (SNR), jitter, and shimmer.

Discussion

Treatment was provided for a group of 45 symptomatic patients with glottic insufficiency for whom conventional treatment modalities were either unsuitable or had failed. Some of these patients had failed multiple treatment attempts of various types prior to entry into this study. In most instances patients were helped with collagen injection in situations where there were no other good options for management. In some cases the improvement was marginal but documented using a variety of measures. There were no major complications andno patient experienced hypersensitivity reactions or airway obstruction as a result of collagen injection in the vocal folds.

Patients often felt better about their voice even though the measures and perceptual judgments did not reflect improvement. In such instances, patients often stated that it was easier to produce voice following treatment. Patients generally equated a louder voice with improvement, although loudness was not a substantial factor in perceptual judgments. We initially tried to do perceptual judgments using a 5 point scale and looked at very subtle perceptual differences but the data was too variable, especially in distinguishing slight improvement from slight decrement or no change. Using a simplified scale, judges found it easier to distinguish those subjects that were substantially better or worse. Most patients exhibited improved voices but there were 17 whose voices were not appreciably changed by perceptual ratings. It is important to put this in perspective. Perceptual judgments do not reflect ease of phonation. In this treatment group of 45 patients, 33 had scarred vocal folds or contralateral pathology; the potential for perceptual improvement was therefore limited by the underlying pathology.

In prior studies[1,3] transglottic airflow rate was found to be a useful aerodynamic indicator of success in the correction of glottic insufficiency problems. In general, a decrease in airflow has indicated improved glottic closure and improved efficiency. The changes in airflow were substantial for the overall group of patients treated in this study ($p= 0.0893$) but these results were not significant largely because of compensatory activity; several of patients exhibited considerable supraglottic valving preoperatively. With correction of their glottic insufficiency, they often had some increase in airflow but improved glottic closure and function.

Prior studies on glottic insufficiency management have shown that maximum intensity and signal-to-noise ratio increased with improved glottic efficiency[1-3]. In addition to these measures, this study looked at other acoustic parameters. There were significant changes noted in percent jitter and shimmer. Favorable changes were also documented in intensity range, habitual intensity, frequency range and habitual intensity.

In summary, this study demonstrates the usefulness of collagen injection in the treatment of a specific group of patients with glottic insufficiency who were not good candidates for other forms of treatment. The absence of any complications supplements the literature on vocal fold injection results indicating no real danger of serious complications. Objective analysis of results indicates that collagen injection was successful in improving function, as measured by voice analysis, in this difficult-to-treat patient population. Further clinical investigation of this important option for management of glottic insufficiency problems seems warranted by these results.

References

1. Ford CN, Bless DM, Loftus JM: Role of injectable collagen in the treatment of glottic insufficiency: a study of 119 patients. Ann Otol Rhinol Laryngol 1992;101:237-247.
2. Ford CN, Bless DM: A preliminary study of injectable collagen in human vocal fold augmentation. Otolaryngol Head Neck Surg 1986;94:104-112.
3. Ford CN, Bless DM: Clinical experience with injectable collagen for vocal fold augmentation. Laryngoscope 1986;96:863-869.
4. Remacle M, Marbaix E, Hamoir M, Bertrand B, Van den Eeckhaut J: Correction of glottic insufficiency by collagen injection. Ann Otol Rhinol Laryngol 1990;99:438-444.
5. Okamoto K, Kawamura Y, Yumoto E, Okamura H: The application of injectable collagen in vocal rehabilitation. Proc IALP 1986;1:484-485.
6. Ford CN: A multipurpose laryngeal injector device. Otolaryngology-Head and Neck Surgery 1990;103:135-137.
7. Ford CN, Bless DM: Collagen injection in the scarred vocal fold. J Voice 1987;116-118.