NCVS Status and Progress Report
Volume 9/April 1996

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 Editors, Julie Lemke and Julie Ostrem
Technical Editor, Martin Milder

Distribution of this report is not restricted.
However, production was limited to 700 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

The Denver Center for the Performing Arts
Wilbur James Gould Voice 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
LDS Hospital

The University of Illinois
Department of Speech and Hearing Science

NCVS Personnel

Administration

Central Office
Ingo Titze, Director
Erich Luschei, Deputy Director
Julie Ostrem, Program Associate
Julie Lemke, Secretary

Area Coordinators
Research - Ingo Titze
Training - Erich Luschei/Patricia Zebrowski
Continuing Education - Julie Ostrem
Dissemination - Cynthia Kintigh

Institutional Coordinators
University of Iowa - Ingo Titze
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.
Patricia Benjamin, B.S.
Bridget Berning, B.S.
David Berry, Ph.D.
Florence Blager, Ph.D.
Diane Bless, Ph.D.
James Brandenburg, M.D.
Gayle Brosovic, B.A.
Myrna Burt
John Canady, M.D.
Linda Carroll, M.S.
Kelly Cavanaugh, B.A.
Geron Coale, M.A.
Stefanie Countryman, M.A.
Charles Davis, Ph.D.
Jeanne Delaney, B.A.
Christopher Dromey, Ph.D.
Wendy Edwards, B.A.
Jeffrey Fields, B.M.
John Folkins, Ph.D.
Charles Ford, M.D.
Amy Furness
Steven Gray, M.D.
Carson Green
Elizabeth Hammond, M.D.
Susan Hensley, M.M.
Marilyn Hetzel, Ph.D.
Margaret Hoehn, M.D.
Henry Hoffman, M.D.
Yoshiyuki Horii, Ph.D.
Heather Hughes
Kathy Ives
Bruce Jafek, M.D.
Antonia Johnson, M.A.
Darin Johnson, B.S.
Karen Johnson, B.S.
Joel Kahane, Ph.D.
Michael Karnell, Ph.D.
Judith King, Ph.D.
Cynthia Kintigh, M.A.
David Kuehn, Ph.D.
Jennifer Lehnherr
Julie Lemke
Russel Long, M.S.
Erich Luschei, Ph.D.
Sharon Maclay, M.A.
Kathryn Maes, Ph.D.
Amy Mast, B.S.
Brian McCabe, M.D.
Arlen Meyers, M.D.
Martin Milder, B.S.
Paul Milenkovic, Ph.D.
Jerald Moon, Ph.D.
John Nichols, B.A.
Lorraine Olson Ramig, Ph.D.
Carrie Ocken, B.S.
Julie Ostrem., B.S.
Namrata Patil, M.D.
Annette Pawlas, M.A.
Kathe Perez, M.A.
Donald Robin, Ph.D.
Robin Samlan, M.S.
Ronald Scherer, Ph.D.
Suzanne Segal, Ph.D.
Elaine Smith, Ph.D.
Marshall Smith, M.D.
Nancy Solomon, Ph.D.
Brad Story, Ph.D.
Edie Swift, M.S.
Christina Taskoff, M.S.
Laetitia Thompson, Ph.D.
Sue Thompson, Ph.D.
Ingo Titze, Ph.D.
Vern Vail, B.S.
Katherine Verdolini, Ph.D.
Patricia Ward
Jennifer Weinstein, B.S.
Barbara Williams
Darrell Wong, Ph.D.
Raymond Wood, M.D.
George Woodworth, Ph.D.
Patricia Zebrowski, Ph.D.

Doctoral Students

Todd Brennan, M.S.
Eileen Finnegan, M.A.
Mark Leddy, M.S.
Phyllis Palmer, M.A.
Annie Ramos, M.S.
Nelson Roy, M.S.
Alice Smith, M.A.
Kenneth Tom, M.A.
Kristie Weaver, M.A.
Elease White, M.S.

Postdoctoral Fellows

Kristin Baker, Ph.D.
Katsuhide Inagi, M.D.
Aliaa Khidr, Ph.D.
Robert Lange, Ph.D.

Visiting Scholars

Hanspeter Herzel, Ph.D., Germany
George Shirley, B.S.

Advisory Board

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

Contents

Editorial and Distribution Information ..... *ii*
Sponsorship ..... *iii*
NCVS Personnel ..... *iv*
Forward ..... *vi*

Part I. Research papers submitted for peer review in archival journals

An Analysis of Cellular Location and Concentration in Vocal Fold Lamina Propria 1
Michael Catten, Steven Gray, Thomas Hammond, Ruixia Zhou and Elizabeth Hammond
Vocal Tract Area Functions from Magnetic Resonance Imaging 7
Brad Story, Ingo Titze and Eric Hoffman
Formants in Singers 27
Ronald Scherer
The Normal Modes of Vocal Fold Tissues 31
David Berry and Ingo Titze
Pressure-Flow Relationships During Phonation as a Function of Adduction 41
Fariborz Alipour, Ronald Scherer and Eileen Finnegan
An Experimental Study of Pulsatile Flow in Canine Larynges 47
Fariborz Alipour, Ronald Scherer and Virendra Patel
Aerodynamic Mechanisms Underlying Treatment-Related Changes in Vocal Intensity in Patients with Parkinson Disease 53
Lorraine Olson Ramig and Christopher Dromey
The Membranous Contact Quotient, A New Phonatory Parameter of Glottal Contact 63
Ronald Scherer, Fariborz Alipour, Eileen Finnegan and Chwen Geng Guo
Increased Stability of Airflow in Spasmodic Dysphonia and/or Vocal Tremor Following Botulinum Toxin Injection 71
Eileen Finnegan, Erich Luschei, James Gordon, Julie Barkmeier and Henry Hoffman
Speech Timing in Apraxia of Speech versus Conduction Aphasia 79
Samuel Seddoh, Donald Robin, Hyun-Sub Sim, Carlin Hageman, Jerald Moon and John Folkins
An Animal Model Comparing Needle and Hooked Wire Electrodes for Laryngeal EMG Recording 93
Debra Jaffe, Nancy Solomon, Robert Robinson, Henry Hoffman and Erich Luschei

Part II. Tutorial reports

Detecting Bifurcations in Voice Signals 99
Hanspeter Herzel, Joachim Holzfuss, Zbiegniew Kowalik, Bernd Pompe and Robert Reuter
Biphonation in Voice Signals 109
Hanspeter Herzel and Robert Reuter
Regulation of Fundamental Frequency with a Physiologically-Based Model of the Larynx 117
Ingo Titze
Vocal Tract Imaging: A Comparison of MRI and EBCT 125
Brad Story, Eric Hoffman and Ingo Titze
Volumetric EBCT Imaging of the Vocal Tract Applied to Male Falsetto Singing 133
Kenneth Tom, Ingo Titze, Eric Hoffman and Brad Story
Anatomy and Physiology of Normal and Disordered Velopharyngeal Function for Speech 143
Jerald Moon and David Kuehn
Behavioral Therapy for Speakers with Velopharyngeal Impairment 159
Lucrezia Tomes, David Kuehn and Sally Peterson-Falzone

Forward

In this report period, NCVS investigators not only attended the usual professional meetings in acoustics, speech motor control, speech pathology, and otolaryngology, but some were part of a special session at the Society of Photo-Optical Instrumentation Engineers (SPIE) in San Diego. This society extended invitations to Dr. Story, Dr. Berry, Kenneth Tom and myself to join a special session on imaging in speech production. We gave an overview of imaging in our field (history, current methods, and future needs) and followed this with several detailed presentations on Electron Beam Computed Tomography (EBCT) and Magnetic Resonance Imaging (MRI) of the vocal tract. Two of the papers are included in Part II of this report.

We learned much about imaging in other areas of research and hope to be able to accelerate the efforts through the NCVS.

Ingo R. Titze, Director
April, 1996

Part I

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 9
April 1996, 1-6

An Analysis of Cellular Location and Concentration in Vocal Fold Lamina Propria

Michael Catten, B.S.
Steven Gray, M.D.
Division of Otolaryngology, University of Utah
Thomas Hammond, B.S.
Ruixia Zhou, M.S.
Elizabeth Hammond, M.D.
Department of Pathology, LDS Hospital, Salt Lake City, Utah

Abstract

The lamina propria of the vocal folds is important in voice production. Understanding of the normal components of this layer and their distribution may further understanding of pathological processes. An evaluation of the distribution and ratio of fibroblasts and macrophages, the main cell types found in the lamina propria, and myofibroblasts, a repair cell, is reported. Twenty two normal human vocal folds were obtained from 11 male and 11 female patients. Using vimentin, CD68, and muscle specific actin stains, image analysis was used to determine the number and distribution of the three cell types in the lamina propria.

Three distinct areas of different cellular activity, based on cell type and distribution are identified: The first area comprises the first twenty percent of the lamina propria, the second region is found in the middle sixty percent, and the last twenty percent of the lamina propria represents the third distinct area. We found that the number of fibroblasts is higher in superficial and deep layers of the lamina with a decrease in total cell number in the middle layer. The macrophages were found almost exclusively in the first twenty percent of the lamina and the distribution of myofibroblasts was greater in the first forty percent of the lamina. No significant gender difference was noted.

These results indicate that most people (86% of our sample) with assumed normal vocal folds have reparative processes (based on the presence of myofibroblasts). In addition, some people (36% of our sample) have macrophage infiltration. The finding of three areas of different cellular activity may be useful in determining where the superficial, middle, and deep layers are located in the lamina propria and the location of associated pathological processes.

Introduction

Most benign laryngeal conditions are often either due to inflammatory causes (laryngitis, reflux, etc.), abusive injuries (nodules, polyps), or perhaps both. Cellular activity and cell type found in the lamina propria may vary depending on the disease process. Macrophage presence perhaps indicates inflammatory processes , whereas myofibroblasts may represent more chronic repair from constant injury. To evaluate this, it is necessary to characterize the cellular components contributing to the normal lamina propria. The majority of cells in the lamina propria are fibroblasts[1]. The existence of macrophages and myofibroblasts has not been well described.

In his morphological studies of the vocal folds, Hirano demonstrated distinctive layers of the vocal folds: epithelium, lamina propria, and muscularis.[2] He further subdivided the lamina propria into three distinct layers based on the density of elastin and collagen fibers.[2] The intermediate layer is abundant in elastin fibers, while the deep layer consists mainly of collagen (figure 1). In this study we compare the concentration and location of these elastin fibers with the concentration and location of fibroblasts, macrophages and myofibroblasts.

Recent research focuses on characterizing other components contributing to the normal extracellular matrix in the lamina propria[3]. Attempts to characterize these other extracellular components appear to demonstrate a similar

layering through the lamina propria that is found with collagen and elastin[4]. This layering may indicate an active role by cells in the vocal folds in the production of these extracellular components. In a study by Pawlak describing the proteoglycans of the normal vocal fold, it was found that some of the proteoglycan staining was present not in fibroblasts, but rather in macrophages. We had assumed that most of the extracellular matrix proteins were being manufactured by fibroblasts. In fact, the paper by Pawlak demonstrated that macrophages can be and are involved in the production of some extracellular matrix proteins. The investigators of this study also found that some cells were neither fibroblasts nor macrophages and through the use of histochemical staining found that these unstained or unidentified cells were myofibroblasts. This study was conducted to provide us with some general parameters as to the location and population density of these three cell types.

Fibroblast: 'Fibroblast' is a generic term for a population of cells responsible for the establishment, maintenance and repair of three-dimensional form in multicellular organisms[5] (table 1). The function of the fibroblast includes the deposition, maintenance, degradation, and rearrangement of the extracellular matrix. Fibroblasts in different tissues have distinct cytoskeletal features[6] and secrete different extracellular matrix components[7]. The phenotypic modulation of fibroblasts in different organs depends upon signals in the extracellular matrix[8]. In vivo, fibroblasts secrete the various collagen molecules[9]. They produce various proteoglycans, such as hyaluronic acid, chondroitin sulfates, heparin sulfates and dermatan sulfates, a class of macromolecules thought to participate in cellular adhesion[10]. They also synthesize fibronectin[11] and produce various proteolytic enzymes[12]. Staining of fibroblasts is done with vimentin.

Macrophage: The term macrophage (table 1) is a loose term which applies to a range of cells known more properly as "mononuclear phagocytes."[13] Although no single characteristic is unique to these cells they share a battery of common characteristics. These characteristics include: morphologic (ruffled membrane, cytoplasmic granules and mononuclear); enzymatic (lysozyme, 5'-nucleotidase, esterase, and leucine amino peptidase); membrane characteristics (Fc receptors, complement receptors and specific antigens); and functional characteristics (phagocytosis of particles). They play a key role in the inflammatory process as antigen presenting cells as well as secreting various cytokines which mediate the intensity and duration of the inflammatory response.

Myofibroblast: 'Myofibroblasts' (granulation tissue fibroblasts) are fibroblasts that develop several ultrastructural and biochemical features of smooth muscle cells, including the presence of microfilament bundles and the expression of actin (table 1). This well developed cytoplasmic actin microfilament system is not present in fibroblasts

Table 1.

Cell Type	Location	Function
Fibroblast	• All samples throughout the lamina propria	• Maintenance and degradation of the extracellular matrix • Secretion of proteoglycans and fibronectin
Macrophage	• Superficial 20% of lamina propria in 36% of samples	• Inflammatory mediator • Secretion of degradatory enzymes • Phagocytosis
Myofibroblast	• Superficial 80% of lamina propria in 86% of samples	• Fibroblasts with smooth muscle features • Found in tissues undergoing remodeling and repair • Role in wound contraction

of normal tissues[14], but is found in smooth muscle cells. Furthermore, myofibroblasts are connected by gap junctions and are connected to the extracellular matrix by the fibronexus, a transmembrane complex involving a continuity between intracellular and extracellular fibronectin fibers[15]. The origin of these cells is thought to be a gradual differentiation of local fibroblasts[16]. Cells with morphological features similar to myofibroblasts have been found in a variety of organs such as the rat intestinal villi, developing human palate mucosa, the human ovarian follicle, and the human renal glomerulus[6]. The role of these cells in normal tissue has not been fully elucidated but it has been suggested that they are found in tissues where reorganization processes take place continuously or periodically[12]. Myofibroblasts are involved in wound healing. They are poorly developed in early granulation tissue , most numerous in the phase of wound contraction and progressively disappear in the late stage of cicatrization[16]. Through the use of actin and the fibronexus, myofibroblasts are suggested to play a role in wound contraction. The secretory products of these cells have not been well described but are believed to be similar to their progenitor cell, the fibroblast. Myofibroblasts may be differentiated from fibroblasts by staining for muscle specific actin.

Materials and Methods

Tissue Preparation

Larynges were obtained from the Medical Examiner from individuals dead of traumatic causes in which the vocal folds were known not to have undergone injury or instrumentation. The selection included 11 females and 11 males between the ages of 17 and 73. All vocal folds were obtained within 24 hours of death. The vocal folds were prepared and processed for permanent sections.

Larynges were examined grossly for evidence of injury. Vocal folds were removed and cut vertically, perpendicular to the free edge of the vocal fold. One 2 mm mid section was frozen in OCT embedding compound in a Tissue tek cryostat equipped with an Instrumedics modification (Instrumedics Corps, Hackensack, NJ). A facing midsection was fixed in Carson's Fixative[18] and embedded in paraffin using routine histological methods. The staining was done on serial cuts with each specimen to assure comparison of similar vocal matrix.

Histochemical Evaluation of Histologic Sections

Four micron-thick sections of normal vocal fold were stained to detect the fibroblast content using vimentin stain. The number of MSA (muscle specific actin) staining cells (myofibroblasts) found in the successive slide was deducted from the total cell number by comparing the images taken for vimentin and MSA staining. This was done to prevent counting the myofibroblasts as fibroblasts when determining cell number. Twenty two vocal folds were examined using this method.

Macrophages were examined using cd68 monoclonal antibody stain and developed with immuno-peroxidase. Twenty two vocal folds were examined with this stain. Developing time was determined using controls of lymphoid tissue.

Myofibroblasts were stained using MSA (muscle specific actin). Developing time was determined using controls and verification of the stain could be seen on each slide by examining the muscle components of the vocal fold (arterioles, vocalis muscle).

Elastin staining had been performed previously on these same vocal folds[4] using Verhoeff's elastic tissue stain (EVG)[19].

Image Analysis

The slides were analyzed using an image analysis system specially configured for histological and cytological studies[4]. A Microimager 1400 high resolution (1280 X 1024 X 10 bit) image scanner (XILLIX Tech Corp., Vancouver, B.C. Canada) is mounted on an Olympus BH2 microscope with a DPlanApo 20UV 20X Lens and BHT 5X Photo eye piece (Olympus Corp., Lake Success, New York, U.S.A.) for the image acquisition. The illumination was carefully controlled via a DC Lamp Power Supply (Olympus Corp., Lake Success, New York, U.S.A.). The image was viewed on a high resolution color monitor to be sure that the proper image was acquired. Optimas Image Analysis Software (Bioscan Corp., Edmond, WA) developed from macros was used to analyze the images. The distribution and the intensity of the substance of interest were measured and recorded by the software.

At 40x microscopic power the lamina propria of each vocal fold was examined. The analysis started at the superficial layer of the lamina propria and continued to the beginning of the vocalis muscle. The cells were counted and recorded in each high power field. Care was taken to proceed in a linear fashion and not overlap frames. The cells were counted using the image analyzer and verified visually before proceeding to the next frame.

Data Analysis

The data was collected and organized using statistical software (SPSS). The number of high power fields required to cross the entire lamina varied between specimens depending on the thickness of the lamina propria. In order to compare these samples, each sample was normalized to one. We then determined cell distribution based on percentage of lamina propria crossed. This enabled us to compare different samples despite the thickness of the lamina propria being different between them.

We evaluated the cell content in the lamina propria in increments of 20%, which gave us five total groups (figure 1). The different cell types were compared for distribution and concentration for each 20% group using box plots. Males and females were also compared. The elastin content of outlying samples in the superficial 60% was compared with the inlying samples to detect any correlation between cell number and elastin content. Elastin content was determined in a previous experiment on the same samples (figure 2).[4]

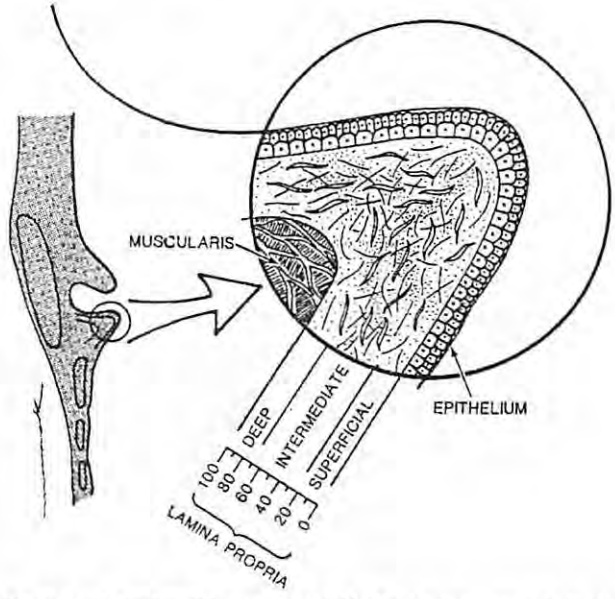


Figure 1. Illustration of human vocal fold. The lamina propria is divided into five divisions proceeding from the edge of the epidermis to the vocalis muscle. Each division represents 20% of the lamina propria. Traditional divisions by Hirano using superficial, intermediate, and deep divisions are also shown.

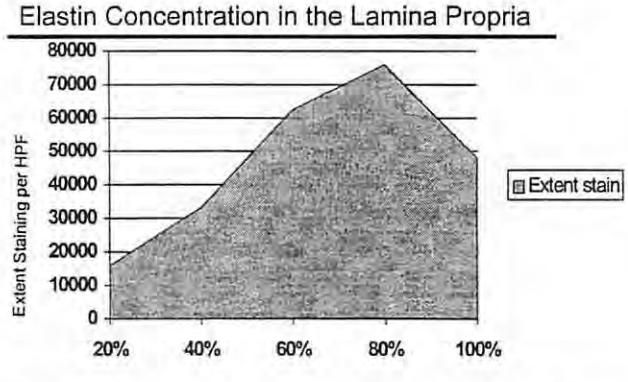
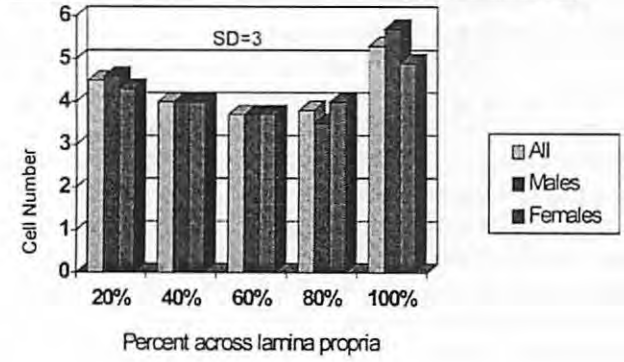
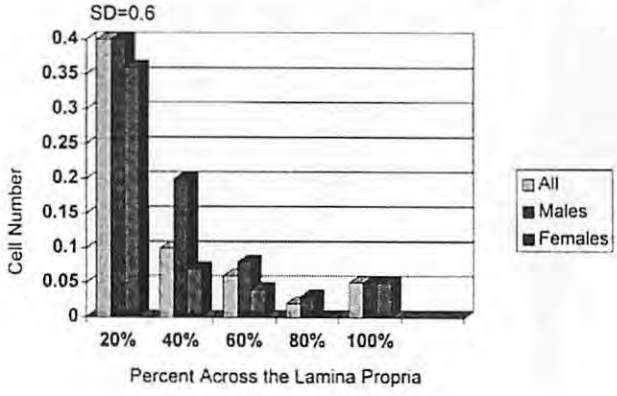


Figure 2. Area graph representing staining of elastin in the lamina propria. The y-axis measures intensity of staining as measured by image analysis. The x-axis represents each 20% division across the lamina propria.

Fibroblasts in the Lamina Propria



Macrophages in Lamina Propria



Myofibroblasts in Lamina Propria

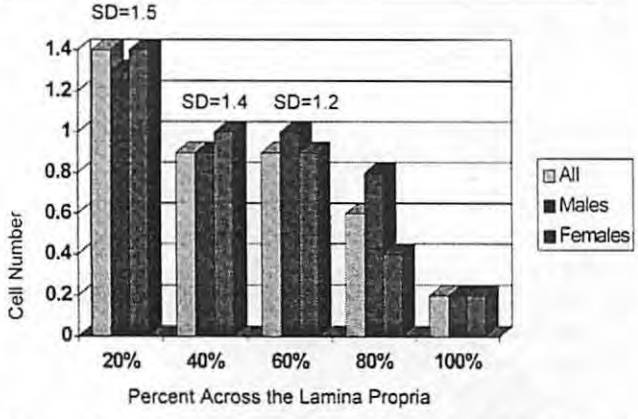


Figure 3a,b,c. Three bar graphs measuring the number of fibroblasts, macrophages, and myofibroblasts respectively found in the lamina propria. The y-axis represents the number of cells per high power field. The x-axis represents each 20% division across the lamina propria. The first bar of each set represents a total of both male and female samples. The second and third bar represent male and female samples respectively.

Results

Cellular Trends in the Lamina Propria

Three distinct areas of cellular activity were identified based on cell type and distribution: The first area comprises the first twenty percent of the lamina propria, the second region is found in the middle sixty percent, and the last twenty percent of the lamina propria represents the third distinct area. We found that the number of fibroblasts is higher in superficial and deep layers of the lamina with a slight decrease in total cell number in the middle layer. The macrophages were found almost exclusively in the first twenty percent of the lamina propria. The distribution of myofibroblasts was greater in the first sixty percent of the lamina with an almost linear reduction in population density from superficial to deep.

Nnumber and Distribution of Fibroblasts

Fibroblasts were noted in all layers of the lamina propria. Considerable variation was noted in the number of cells between samples. The average number of cells was 4 per high power field in the superficial 20% of the lamina, 3 in the middle 60% and 4.5 in the deep 20% of the lamina (figure 3a; upper). The standard deviation was 3 for all groups. No statistical difference in the number of fibroblasts was noted when comparing males to females.

Number and Distribution of Macrophages

Macrophages were observed almost exclusively in the superficial 20% of the lamina propria (figure 3b; middle). Only 36% of the total samples contained macrophages and the maximum number observed in the first 20% was 3 per high power field. The average number of cells was 0.4 for the first 20% of the lamina propria (standard deviation of 0.6). No statistical differences in the number of macrophages was noted when comparing males to females.

Number and Distribution of Myofibroblasts

Myofibroblasts were observed in 86% of the samples. Their distribution was in the superficial 80% of the lamina propria with a greater number observed on the first 20% (figure 3c; lower). The average number of cells observed in the 20%, 40%, 60% and 80% categories was 1.4 (std dev 1.5), 1 (std dev 1.4), 0.9 (std dev 1.2), and 0.6 (std dev 1.1) respectively. No significant gender difference was present.

Correlation Between Elastin Content and Cell Number

No correlation was found between statistical outlying samples (many fibroblasts, few fibroblasts, many macrophages, etc.) and the intensity of elastin protein present when using any of the three cell types in the superficial 60% of the lamina propria. The deep 40% was not evaluated.

Correlation Between Age and Cell Number

Attempts to link statistical outlying samples in cell number with any trend in age proved inconclusive.

Discussion

In this study we have described and shown evidence that three distinct cell types exist in the lamina propria and that each cell type has a unique distribution and frequency. Cell type and number give us clues as to the biologic activity occurring in the lamina propria. The location of the different cell types in the lamina propria could correlate with the extracellular properties of the vocal folds. In this study no correlation between cell type with elastin concentration was found. Even when statistical outliers were used no correlation was noted between elastin and cell location. Although this attempt was unsuccessful, each of these cells secrete a wide variety of products and in the future a comparison of these secretory products will be made.

It appears that in the vocal folds, the presence and population density of certain cells more likely reflects responses to environmental stimuli. When Hirano described the lamina propria he noted fibroblasts but not macrophages or myofibroblasts[20]. Both of these cell types are indicators of inflammation and cellular repair. In the minority, but certainly noticeable, percentage of normal vocal folds there was significant macrophage presence in the superficial 20% of the lamina propria. One has to wonder as to the purpose for the presence of these macrophages in this very superficial location. Their presence likely indicates mediation of inflammation in response to mucosal irritants. These may be infectious or chemical respiratory inhalational irritants. We do not know if these people had any smoking history. It should be mentioned that there were no epithelial abnormalities often found in smokers in these samples. Other respiratory inhalants, which can be irritating and are found either in an urban center (smog) or environmental respiratory exposure in the workplace, are not known for any of these samples. We are currently investigating the activity of these cells further.

The presence of myofibroblasts in the superficial area, but extending down to the first 60% of the lamina propria, shows a different distribution than the macrophages. Although it is possible that macrophages and the myofibroblasts are present as a result of injury and repair, it is the authors opinion that both cell types are present for different purposes. Those samples with macrophages did not necessarily have myofibroblasts, and vice versa. Furthermore, myofibroblasts occurred in a larger population of subjects than the macrophages. The presence of myofibroblasts indicates previous injury and ongoing repair. Over 80% of what we assume to be normal vocal folds had myofibroblasts present indicating ongoing reparative processes in the lamina propria. The distribution of myofibroblasts, with more cells being in the superficial layer and linearly declining, probably gives us a good working model for what part of the lamina propria is frequently injured. Certainly this distribution of myofibroblasts population density parallels our knowledge of where stress in the vocal folds is likely to occur the most. This finding further substantiates the theory that most people experience mild vocal fold injury from normal daily vocal use. The finding of some cells of repair and their consequent extracellular matrix products probably should not be viewed as pathologic.

There was a wide variation in cell number between samples. This was most prominent in the fibroblasts (figure 1). It is unknown if an increase in cell number predisposes to later pathology, is a result of increased voice use, or a variation of normal. Further studies must be undertaken to answer this question.

In past studies, it has been shown that the lamina propria in males is usually thicker than females. The thickness of the lamina propria is important because it has been shown that there is a direct relation of thickness to pliability of the fold[21]. Based on our studies this difference cannot be explained by cell number or cell type. Attempts to show a gender difference in the number or distribution of cells was inconclusive.

In summary, three distinct cell types are present. The fibroblasts are present in all, but their population density did not correlate with elastin content intensity. The presence of macrophages in the initial 20% of the lamina propria may be indicative of environmental inhalational irritants and the presence of myofibroblasts may indicate that most people have low-grade vocal injury from normal daily use. Further research will determine if correlation exists between these cells, the extracellular matrix, and vocal pathology. Benign diseases may be better studied and compared to this data on normals. We hope that a better understanding of the exact composition of the vocal folds and the interactions between these components will help with future clinical and preventative medicine.

Acknowledgments

This work was supported by NIH Grant #P60 DC00976.

References

[1]Hirano M, Kurita S, Nakashima T. Growth, development and aging of human vocal folds. In: Bless DM, Abbs JH, eds. Vocal Fold Physiology. San Diego: College-Hill Press, 1983; 22-43

[2]Hirano, M. Structure and vibratory behavior of the vocal folds. In M. Sawashima and F. S. Copper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977.

[3]Palwak A, Hammond TH, Hammond EH, Gray SD. Immunocytochemical study of proteoglycans in vocal folds. *Ann Otol Rhinol Laryngol.* 1996; 105:6-11.

[4]Hammond TH, Zhou R, Hammond E, Palwak A, Gray S. The intermediate layer: A morphological study of the elastin and hyaluronic acid constituents of the normal human vocal folds. *J of Voice*, 1996, in press.

[5]Trelstad RL, Birk D; The fibroblast in morphogenesis and fibrosis: cell topography and surface-related functions. Ciba Foundation symposium; 114: Fibrosis, 1985.

[6]Schmitt-Gräff A, Desmoulière A, Gabbiani G. Heterogeneity of myofibroblast phenotypic features: an example of fibroblastic cell plasticity. *Virchows Archiv* , 1994; 425 :3-24.

[7]Larjava H, Heino J, Krusius T, Vuorio E, Tammi M. The small dermatan sulfate proteoglycans synthesized by fibroblasts derived from skin, synovium and the gingiva show tissue-related heterogeneity. *Biocem J* , 1988; 256:35-40

[8]Juliano RL, Haskill S, Signal transduction from the extracellular matrix. *J Cell Biol*, 1993; 120:577-585

[9]Bornstein P, Sage H. Structurally distinct collagen types. Ann Rev Biocham 49:957, 1980.

[10]Iozzo RV. Proteoglycans, structure, function, and role in neoplasia. *Lab Invest*, 1985; 53:373.

[11]Gabbiani G, Rungger-Brändle E. The fibroblast. In Handbook of Inflammation. Tissue Repair and Regeneration, Vol. 3, edited by Glynn LE, pp 1-50. Amsterdam, Elsevier/North-Holland Biomedical Press, 1981.

[12]Birkedal-Hansen H, Cobb CM, Taylor RE, Fullmer HM. Synthesis and release of procollegenases by cultured fibroblasts. *J Biol Chem*, 1976; 251:3162.

[13]Dean RT, Jessup W, eds. Mononuclear phagocytes: physiology and function. Elsevier Science Publishing , 1985: Introduction.

[14]Gabbiani G, Ryan GB, Majno G . Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction. *Experientia*, 1971; 27:549-550.

[15]Singer II, Kawka DW, Kazazis DM, Clark RAF. In vivo co-distribution of fibronectin and actin fibers in granulation tissue: Immu noflorescence and the electron microscope studies of the fibronexus at the myofibroblast surface. *J Cell Biol*, 1984; 98:2091-2106.

[16]Darby I,Skalli O, Gabbiani G . alpha-Smooth muscle actin is transiently expressed by myofibroblasts during experimental wound healing. *Lab Invest*, 1990; 63:21-29

[17]Sappino AP, Schürch W, Gabbiani G. Differentiation repertoire of fibrobalstic cells : Expression of cytoskeletal proteins as a marker of phenotypic modulations. *Lab Invest*, 1990; 63:144-161

[18]Carsons FL, Martin JH, Lynn JA. Formalin fixation for electron microscopy, a reevaluation. *Am J Clin Pathol*, 1973; 59: 365-7.

[19]Carson FL. *Histotechnology. A self-Instructional Text.* Chicago: ASCP Press. 1990: 140-1.

[20]Hirano M. Structure of the human vocal fold in normal and disease states anatomical and physical studies. In: Ludlow CL, Hard MO, eds. *Proceedings of the Conference on the Assessment of Vocal Pathology.* Rockville: The American Speech-Language-Hearing Association, 1981: 11-30.

[21]Yumoto E, Katota Y, Kurokawa H. Infraglottic aspect of canine vocal fold vibration: effect of increase of mean airflow rate and lengthening of vocal fold. *J of Voice* 1993;7(4):311-8.

NCVS Status and Progress Report - 9
April 1996, 7-26

Vocal Tract Area Functions from Magnetic Resonance Imaging

Brad Story, Ph.D.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Eric Hoffman, Ph.D.
Department of Radiology, The University of Iowa College of Medicine

Abstract

There have been considerable research efforts in the area of vocal tract modeling but there is still a small body of information regarding direct 3-D measurements of the vocal tract shape. The purpose of this study was to acquire, using MRI, an inventory of speaker-specific, three-dimensional, vocal tract air space shapes that correspond to a particular set of vowels and consonants. A set of 18 shapes was obtained for one male subject who vocalized while being scanned for 12 vowels, 3 nasals, and 3 plosives. The 3-D shapes were analyzed to find the cross-sectional areas evaluated within planes always chosen to be perpendicular to the centerline extending from the glottis to the mouth to produce an "area function". This paper provides a speaker-specific catalogue of area functions for 18 vocal tract shapes. Comparisons of formant locations extracted from the natural (recorded) speech of the imaged subject and from simulations using the newly acquired area functions show reasonable similarity but suggest that the imaged vocal tract shapes may be somewhat centralized. Additionally, comparisons of the area functions reported in this study are compared with those from four previous studies and demonstrate general similarities in shape but also obvious differences that can be attributed to differences in imaging techniques, image processing methods, and anatomical differences of the imaged subjects.

Introduction

Speech simulation algorithms have been developed to provide a sophisticated representation of acoustic wave propagation through the vocal tract (Ishizaka and Flanagan (1972), Strube (1982), Maeda (1982), Sondhi and Schroeter (1987) and Liljencrants (1985)). Most speech simulation models are based on the assumption of one-dimensional wave propagation. This means that the tubular vocal tract shape can be approximated as a finite number of cylindrical elements that are "stacked" consecutively from the larynx to the mouth. A particular vocal tract shape can be imposed on a model by specifying the cross-sectional area of each cylindrical element as a function of the distance from the glottis. For modeling purposes, any vocal tract shape can be defined by its unique "area function". Hence, a necessary component for the simulation of natural sounding speech is an inventory of vocal tract area functions that correspond to the vowels and consonants used to produce human speech. The success of speech simulators has been limited, in part, by the lack of a body of morphological information about the vocal tract shape on which to base these area functions. Shape information regarding other airways such as the nasal tract and trachea is also needed. The one nearly complete set of area functions that has been used extensively as input to speech simulation models, is that published in 1960 by Fant. These area functions were indirectly acquired with sagittal x-ray projection images and plaster casts of the oral cavity and have proved to be an invaluable resource for researchers in the field of speech synthesis and simulation.

Modern imaging techniques are now being used to acquire three-dimensional shape information about the vocal tract volume and associated airways. Volumetric imaging relies on the ability of an imaging technique to acquire a series of image slices, in one or more anatomical planes,

through a desired volume of the human body. This process can be performed with either magnetic resonance imaging (MRI) or electron beam computed tomography (EBCT). Post-processing of the images typically includes segmentation of the air space from the surrounding tissue and a 3-D reconstruction of the airway shape. Once the 3-D structure has been identified, it can be measured and analyzed. During speech production, a speaker will traverse many different vocal tract shapes in a short period of time, often overlapping one shape into another (i.e. coarticulation). Unfortunately, neither MRI or EBCT can acquire a volume image set fast enough to capture the dynamically changing vocal tract shape. Thus, the use of present, commercially available, imaging techniques can be used only to study static vocal tract shapes.

MRI is an attractive imaging technique primarily because it poses no known danger to the human subject being imaged. Since human speech comprises a large number of vowel and consonant shapes it is often desirable to acquire a large number of image sets. Thus, the human subject may need to be exposed to hours of imaging, but fortunately, with no apparent risk. However, MRI has a number of distinct disadvantages. The scanning time required for acquisition of a full image set is on the order of several minutes. When this is coupled with the pauses required for subject respiration, a complete image set may require double or triple the actual scan time. Thus, the image set and subsequent reconstruction of the airway is based on a large number of repetitions of a particular shape. Also, because of various imaging artifacts associated with air/water interfaces, the air-tissue interface can be poorly defined when imaged via MRI techniques. Additionally, teeth and bone are poorly imaged by MRI because of the low hydrogen (water) concentration, making these structures appear to be the same gray scale value as air.

Baer et al. (1991) demonstrated the use of MRI to directly measure the vocal tract shape for the four point vowels with two adult male subjects. In two separate experiments, they collected combinations of image sets in three planes from which they created the first demonstration of 3-D reconstructions of the vocal tract shape using MR imaging techniques. They also reported the corresponding area functions for each vowel that were discretized along the vocal tract centerline at intervals of 0.875 cm and proposed a midsagittal width - to - area transformation based on their data. Moore (1992) performed an MRI study in which sagittal and coronal image sets of the vocal tract were obtained from five adult male subjects for three vowels and two continuants. The study investigated correlations between cavity volumes and resonance frequencies of the vocal tract, as well as other variables such as lip length and lip area. However, it did not include three-dimensional reconstructions of the vocal tract shapes. Another MRI study of the vocal tract was performed by Sulter et al. (1992). They used one male subject who was a trained singer. Their primary interest was in correlating the vocal tract length measured for an /ɛ/ vowel to resonance frequencies (formants) predicted by theory. They also imaged the vowels /i/ and /ɑ/ from which they measured cavity volumes. Greenwood et al. (1992) imaged five static vowels using one subject. They acquired an image in the midsagittal plane, and a set of images in the axial and coronal planes. Areas functions were extracted from the image slices in close accordance with the vocal tract model proposed by Mermelstein (1973). Dang et al. (1994) used MRI to acquire contiguous coronal image sets throughout the nasal tract volume and sagittal image sets of the vocal tract for nasal consonants /m/ and /n/. Using these data they produced 3-D reconstructions of the nasal tract passages and sinus cavities. The nasal tract morphology along with the vocal tract area functions were subsequently used to model the acoustic characteristics of the nasal system and production of nasal consonants. Reconstruction of 3-D vocal tract shapes for five vowels was reported for an adult male, adult female, and an eleven year old male by Yang and Kasuya (1994). Coronal and axial image sets were acquired in order to define the oral and pharyngeal cavities, respectively. Area functions that were extracted from the vocal tract reconstructions for each subject were implemented in a frequency domain model of acoustic wave propagation (Sondhi and Schroeter, 1987). Formant frequency comparisons between the model and natural recorded speech were quite close, with many cases exhibiting less than a 5 percent error. Narayanan et al.(1995) and Narayanan (1995) have reported the acquisition of coronal and axial image sets from four adult subjects (2 male and 2 female) representing the fricative consonants, /s,ʃ,f,θ,z,ʒ,v,ð/. 3-D reconstructions of both the vocal tract shape and the tongue were created from these images and a cross-sectional area analysis was performed. This study provides the most accurate information to date of the constrictions and air channels that produce the turbulence generated sound, characteristic of fricative consonants. Lakshminarayanan et al. (1991) have focused on using MRI to acquire midsagittal sections of the vocal tract from which they measured the tract widths and converted them to cross-sectional areas using an empirically derived formula from Rubin et al. (1981). Perrier et. al. (1992) and Beautemps et al. (1995) have both used x-ray techniques in an attempt to further develop transformations from the midsagittal width to area.

The intent of this experiment was to obtain a collection of vocal tract shapes (and consequent area functions) corresponding to a large number of speech sounds for *one* specific speaker. Such a collection should allow for a unique simulation of speech (i.e. of the imaged subject) in which acoustic pressures and flows generated by a speech production model can be directly compared to that of the subject. Additionally, methods of interpolating or convolving the area functions in time to form "running speech" might be

more easily studied since a direct comparison can be made between simulation and recorded natural speech of the subject. Thus, a set of area functions from one speaker will allow a more rigorous evaluation of speech modeling algorithms since the simulation can be directly compared to the real human system being modeled. Twenty-two complete vocal tract shapes, trachea, piriform sinuses, and two states of the nasal tract were scanned with MRI for this study. However, only a subset containing 12 vowels, 3 nasals, and 3 plosives will be presented in this paper.

In addition to requiring that one subject produce all of the desired vocal tract shapes it was desired that the image set representing any one particular shape be completely acquired within one imaging session; i.e. the subject would be positioned in the scanner only once for any particular vowel or consonant. This requirement was intended to avoid the potential errors associated with repositioning the subject in the scanner for separate sessions as well as the possibility that the subject might produce a particular vowel or consonant with a slightly different vocal tract shape from session to session. The volumetric imaging studies reviewed above (Baer et al., 1991; Moore, 1992; Sulter, 1992; Greenwood et al. 1992, Dang et al., 1994; Narayanan et al., 1995) all acquired multiple plane image sets. In these studies it has typically been regarded that some combination of two imaging planes is necessary for measurement of the vocal tract shape. But because of the physical demands required of a single subject to produce different vocal tract shapes for 22 full volume scans (with phonation during scanning) plus two volume scans of the nasal tract and one of the trachea, it was concluded that each image set could reasonably include only one imaging plane. Even with the reduction to one imaging plane, the subject spent between seven and eight hours over the course of three separate sessions lying in the MRI scanner. The axial plane was chosen because it would most accurately represent the small epilaryngeal region and the often constricted pharynx as well allowing a direct comparison to a limited number of image sets acquired with electron beam computed tomography (EBCT) that were part of another experiment (Story, 1995; Story et al. 1996); EBCT allows only the acquisition of axial image sets. Using only axial slices may compromise the measurement accuracy in the oral cavity, particularly for front vowels when there is little air space between the upper surface of the tongue and the hard palate. In such a case, the thickness of the MR slice may be larger than the airway passage. Ideally, multi-plane scanning would be performed for every vowel and consonant shape produced by the subject, but subject endurance as well as monetary and time constraints associated with obtaining scanning time prevented this ideal case.

This paper is primarily intended to report area function data for static vocal tract shapes. However, comparisons of formant locations produced by 12 tract shapes for both the natural speech of the subject and a simulation based on the new set of area functions will be presented.

Image Acquisition and Analysis

MRI Scanning Parameters and Image Acquisition Protocol

The MR images were acquired using a General Electric Signa 1.5 Tesla scanner. MRI produces a slightly blurred boundary between tissue and air (Baer et al., 1991). However, the acquisition mode can be chosen and the pulse sequence parameters adjusted to provide acceptable air-tissue interfaces. The parameters shown in Table 1 were empirically found to provide acceptable images of the airway.

Table 1.
MR parameters for vocal tract image acquisition.

Mode = Fast Spin-Echo, using an anterior neck coil	
TE = 13 msec	(echo delay time)
TR = 4000 msec	(repetition time)
ETL = 16 msec	(echo train length)
FOV = 24 cm	(field of view)
NEX = 2	(number of excitations)
image matrix =	256 x 256 pixels
resolution =	0.938 mm/pixel
slice thickness =	5 mm

Volumetric imaging of the vocal tract using MRI for a single subject was completed for 22 different phoneme configurations and also for the nasal tract and trachea (but as stated previously, only the analysis of 18 tract configurations will be discussed in this paper). The subject (BS) was a 29 year-old male with no history of speech or voice disorders and is a native of the midwestern United States. The subject is 5 feet 7 inches tall and weighs approximately 145 pounds. His head circumference was measured to be 57 cm and for the neck, 34 cm, measured just above the prominence of the thyroid cartilage. Height, weight, and head and neck sizes have not typically been reported in previous imaging studies, but such information may be useful to the reader when comparing the data presented here to other data sets. While the body dimensions given certainly do not define a vocal tract size, they may provide a clearer picture of the subject's body structure. The subject is also the first author of this study.

Using the parameters given in Table 1, a 26 slice series of 5 mm thick contiguous, parallel, axial sections was gathered in an interleaved acquisition. This image set, which extended from just above the hard palate down to the first tracheal ring, could be acquired with 4 minutes and 16

seconds of scan time. However, because the subject was required to phonate during image acquisition, the actual amount of time required to image one shape was approximately 10 minutes, allowing for respiration during pauses in scanning.

Prior to the imaging session, the subject spent a significant amount of time practicing phonation while holding the vocal tract shape as steady as possible. The protocol for image acquisition was as follows. The subject was first given ear plugs to attenuate the intense sound of the MR machine. The subject was then positioned in a comfortable supine position on the patient table in the MRI examination room. The subject's head was placed directly on the table (no foam or cushion) and positioned so that the table was perpendicular to the Frankfort plane. Cloth adhesive tape was then used to secure the head to the table in this position. With this approach, very little head movement was possible. An anterior neck coil was brought into position so that the desired portion of the head and neck were within its field of view. With this preparation completed, the patient table was moved such that the subject's head was in the center of the magnet. Prior to each image acquisition session, a sagittal localizer was performed to allow for identification of the appropriate field of view and scan location.

During protocol development a foot signaling method was tested to indicate to the MR technician when a breath was needed. However, the foot movement introduced motion artifacts and blurring. The subject reported that this method interrupted concentration. In the final study the technician started the image acquisition when the subject began phonation (as heard over the intercom) and then "paused" the machine after eight seconds of scanning so the subject could take a breath. As soon as the subject began phonating again, the image acquisition was continued. This method allowed the subject to be in control of the time between acquisitions and also keep the vocal tract stable. The subject was given any pertinent instructions over an intercom system. Throughout all image acquisitions a speech scientist, experienced in phonetics, was present in the control room to listen to each speech sound and halt the acquisition if the subject strayed from the desired target.

The phonemes used to create static vocal tract shapes during imaging are shown in Table 2 with the example word that was given to the subject prior to the image acquisition. For the consonants, the subject imagined that the preceding and following vowel shape was a neutral or schwa vowel (note the rather strange example for /ŋ/).

Table 2.
List of imaged phonemes using phonetic symbols, symbols, and example words.

Phonetic Symbol	Example
/i/	heed
/ɪ/	hid
/ɛ/	head
/æ/	had
/ʌ/	ton
/ɑ/	hod
/ɔ/	paw
/o/	hoe
/ʊ/	hood
/u/	who
/ɜr/	earth
/l/	lump
/m/	mum
/n/	numb
/ŋ/	ung a
/p/	puck
/t/	tuck
/k/	cut

Image Analysis

All image analysis operations were performed with a general Unix-based image display and quantitation package called VIDA™ (Volumetric Image Display and Analysis) which has been developed (and continues to be enhanced) by researchers in the Division of Physiologic Imaging at the University of Iowa (Hoffman et al., 1992). Additional information regarding VIDA can be found on the Internet by using a web browser and logging onto http://everest.radiology.uiowa.edu/.

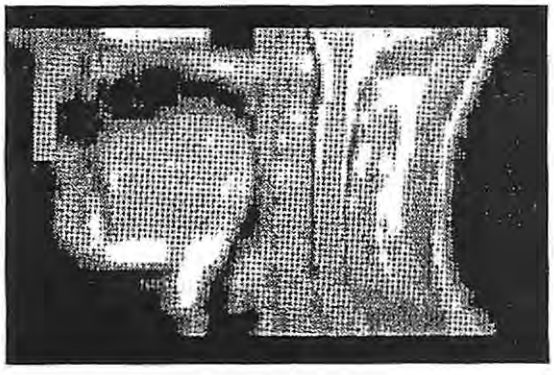
The goal of this experiment was to determine area functions for each of the phonemes listed in Table 2. To achieve this goal the image analysis process included three main steps: 1) segmentation of the airway from the surrounding tissue, 2) three-dimensional reconstruction of the airway by shape-based interpolation, and 3) determination of an airway centerline and subsequent extraction of cross-sectional areas assessed from oblique sections calculated to be locally perpendicular to the airway centerline.

Airway Segmentation

The image data sets were transferred from the GE Signa scanner to a Unix-based workstation (via magnetic tape) and translated into a file format recognized by the analysis and quantitation software (VIDA). When files are in this format they can be read into a shared memory structure, and for convenience, images were converted from 16 to 8 bit per pixel gray scale resolution. The upper gray level cutoff was determined by computing a pixel histogram on several slices sampled throughout the data set and choosing the cutoff point so that all pixel values present in the image set were below the cutoff (our version of the GE scanner downloaded the MR data sets as 8 bits of gray scale information stored in 16 bit pixel values).

The airway was segmented from the surrounding tissue by setting all voxels considered to be in the airway to a unique gray scale value. The first step in this process was to attenuate all gray scale values in an image slice by 5 percent. This ensured that no voxel in the image volume had a value of 255 (the brightest value for an 8-bit gray-scale range). Next a gray scale value that represented the threshold dividing air and tissue was determined by choosing the voxel intensity that was halfway between the darkest part of the airway and the brightest part of the tissue surrounding the airway as determined by region of interest analysis. Validation of this approach using the image data sets for a phantom study is discussed in Appendix A.

Once the threshold value had been determined the actual segmentation process could be performed. A seeded region growing algorithm was used that changed the gray scale values of all pixels below the threshold value to the brightest possible gray level (Hoffman, 1983; Udupa, 1991). This effectively set the airway to a single color that, because of the previous attenuation process, was unique with respect to the rest of the image volume. The seeded region growing process required the airway region to be non-continuous with air outside the body. For the axial slices in the pharyngeal section this was not a problem, but in the oral region where some slices have an open mouth condition, the algorithm would "leak" the brightest pixel value out of airway and into the "blackness" (air) outside the image. To correct the problem, an artificial boundary that defined the mouth termination of the vocal tract was constructed. First the image set was viewed in a reformatted midsagittal plane (extracted from the volumetric image data set) where the outline of the lips are easily seen. A mouth termination plane was then defined using the technique of Mermelstein (1973). Figure 1 shows the boundary in the midsagittal plane and the containment of the seeded region growing algorithm in the axial plane.



(a)

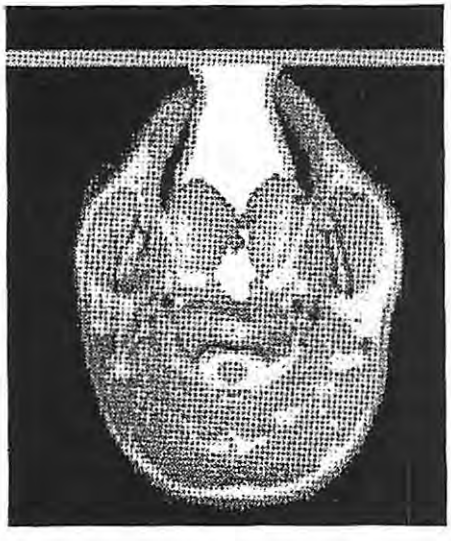


Figure 1: Method of terminating the open mouth to contain the seeded region growing. a, at top) "painted" rectangle on the midsagittal slice, b) containment "fence" in the axial plane.

A potential problem that arises in segmentation of the oral cavity is that the 5 mm slice thickness may cut through both tissue and air. For example, the upper 2.5 mm of a slice may contain a portion of the hard palate while the lower 2.5 mm contains air. The result would be a voxel density with some average value of air and tissue which could be above or below the set threshold; i.e. tissue could be included as air and vice versa. The most superior slice in a given image set normally included the lower portion of the nasal tract and also some of the velopharyngeal air space for the nasal consonants. Usually, this slice was ignored except in the case of the nasals. The next slice inferior to the first, typically contained the hard palate and extended down to include the midline raphe on the roof of the mouth. Because of the uneven surface created by the midline raphe on the roof of the mouth, this slice contained some voxels with densities made up of contributions from both air and tissue; their values often would lie below the threshold determined for segmentation. During the segmentation process, this particular slice was always segmented with the threshold-based seeded region growing method. However, the subsequent 3-D reconstruction of the tract shape using methods described in the next section, were performed with and without this particular slice. It was often the case that, when the slice was included, the 3-D reconstruction possessed some rather unnatural "fins" on its most superior portion that were most likely the result of including something other than air. Thus, in such cases, this particular slice was ignored and the interpolation, surface rendering, and area function analysis were performed without it. A similar error could occur in the inferior part of the oral cavity where portions of the tongue and adjacent airspace might be included within a slice. However, the effect may be less severe in this case because the variety of shapes assumed by the tongue in the production of different speech sounds may ensure that one slice would not usually contain the entire upper surface of the tongue. Thus, no attempt was made to alter the threshold-based segmentation in this region. With the approach taken, it is

possible that a very small volume of air was excluded from the airway which would have the most significant effect on vowels with a front constriction such as /i/, /ɪ/, and /ɛ/.

Volume Reconstruction

Interpolation to produce an isotropic (cubic voxel) image set is often used when an image set has been collected in only one image plane and one dimension has a spatial resolution equal to the thickness of an image slice. Voxel densities can be interpolated between consecutive image slices to generate a uniformly sampled image set. The interpolation is typically performed with nearest neighbor, linear, or trilinear techniques (Udupa, 1991). If interaction by the user is required to identify and segment a region or structure of interest, the cubic voxel image set greatly increases the workload. For example, the MR image sets collected for this experiment were 26 slice series of 5 mm thick slices. An interpolation to a cubic voxel image set would generate 138 slices.

Instead of the more general practice of interpolating voxel densities, a technique called shape-based interpolation (Raya and Udupa, 1990) was used reconstruct the three-dimensional shape of the vocal tract airway. The shape-based algorithm interpolates segmented image data to form an isotropic (cubic voxel) data set. Thus, user interaction can be limited to structure or region identification in the original image set. Since the image segmentation process assigned a gray-scale value to each voxel within the air space that was unique to the rest of the image (in particular the value 255), the segmented image set can be considered to represent a *binary* scene; i.e. each slice can be separated into a "patch" of air (gray level = 255) or *not* air (gray-level < 255). The technique of shape-based interpolation (Raya and Udupa, 1990) utilizes this binary representation of a structure (in this case the airway) within an image set to interpolate slices between the originals using information related to the structure's surface location above and below the new slice location to be generated. Within a given image slice, the shortest distance between each voxel and the boundary of the airway is computed and stored in a 2-D array with the same dimensions as the image slice. For example, if the image consists of a 256 x 256 pixel matrix and a shortest distance to the airway boundary is measured for the pixel located at a position of (123,50), then this measured distance is stored in the (123,50) location within the 2-D array. If the voxel was in the airway, the measured distance is considered to be positive valued, otherwise the distance is set to be negative. This process is performed for each image slice so that, when finished, a new set of "slices" (2-D arrays) containing distance measures has been created that maintains the correct spatial ordering of the original segmented image set. The interpolation between slices is performed on the shortest distance arrays, rather than actual image slices. When the interpolation is complete, the new set of shortest distance slices (2-D arrays) can be mapped back into a binary scene by assigning every positive distance value to a voxel in an image slice with a gray-level value of 255. Every negative distance value is assigned to a voxel with a gray-level value of 0. Thus, the shape-based approach interpolates only the gray scale value of the airway while all other colors are ignored, effectively "stripping" all tissue away from the airway while creating the same voxel resolution in the axial direction as the other directions; i.e. isotropic resolution.

The interpolated airway produced by shape-based interpolation was subsequently used in both a surface rendering application and a cross-sectional area analysis. The surface rendering consisted of first extracting the edges of the interpolated airway image and then creating a shaded surface display. The result was digital "cast" of the vocal tract. Surface rendering applications used effectively in many cardiac and pulmonary applications are discussed in Hoffman (1991). High quality three-dimensional representations of the airway can be rotated and magnified to show many different perspectives. It is only a qualitative tool, but nevertheless an important step in the image analysis because it shows the quality of the segmentation process and provides three-dimensional views of each vocal tract shape.

The area analysis takes advantage of the resultant isotropic, binary representation of the vocal tract to compute oblique cross-sectional areas perpendicular to the local long axis. The details of this analysis are given in the next section.

Area Function Analysis

To extract the area function from the interpolated vocal tract shape, an algorithm was used that was originally developed to analyze upper airway geometry and volume with regard to sleep disorders (Hoffman et al., 1992; Hoffman and Gefter, 1990). It computes cross-sectional areas from oblique sections calculated to be perpendicular to the local airway long axis.

A segmented and interpolated image is used from which a three dimensional centerline is computed through the airway. An iterative bisection algorithm is used to compute the centerline. A voxel is chosen by the user at the beginning and the end of the segmented image; i.e. the beginning voxel is chosen near the mouth termination and the end voxel is chosen near the glottis. A line is then drawn through three-dimensional space from the beginning voxel to the end voxel. The plane perpendicular to this line and halfway between the two specified voxels is determined, and all voxels in this plane that have the gray scale value of the airway are identified and used to calculate an average voxel location (i.e. the centroid). The new voxel location along with the original end points are now used to create two new line segments, two new planes perpendicular to the line segments, and consequently two new centroid points within the airway. The process can be repeated for any number of iterations specified by the user. When the iterations are

finished the result is a sampled version of the three-dimensional centerline through the airway. Oblique sections were calculated perpendicular to the long axis of the airway at each centerline point. A voxel counting algorithm (summing all voxels of the airway gray scale value) was applied to each oblique section yielding the cross-sectional area, anterior-posterior length, and lateral length as a function of slice location along the airway length. The distance between consecutive cross-sections can be calculated with the three-dimensional Pythagorean theorem,

$$L_1 = \sqrt{\left((x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2\right)} \qquad (1)$$

where (x_0,y_0,z_0) and (x_1,y_1,z_1) are the spatial coordinates of the airway centroid locations from consecutive cross-sections. However, because of the irregular shape of each cross-section, the centroid location often varied within the plane of the section. This will impose small displacements of the centerline out of the axis of wave propagation. Figure 2a shows the lateral (x) coordinate for the vowel /ɑ/ as a function of section number. The first section is at a point just above the glottis and the last section is at the lips. The figure shows that there is an overall left and right displacement as well as many small, sharp variations. While the variation from point to point is typically a millimeter or less, the summation of the variations can accumulate and add a centimeter or more to the total vocal tract length. For the frequency range of speech sounds, these variations are a tiny fraction of a given wavelength and it is, therefore, unlikely that a pressure wave would alter its direction of propagation in response to small off-axis variations; i.e. a propagating acoustic wave would not "see" these small variations as significant changes in the axis of propagation. To suppress their influence, an averaging filter was applied to the anterior-posterior (y) and inferior-superior (z) coordinates to smooth out sharp variations in the midsagittal plane. The lateral (x) coordinate was not used in the length calculation, hence the tract length was based on the midsagittal vocal tract profile while the true centerline was used for oblique plane and cross-sectional area calculations. Both raw and smoothed tract profiles for the vowel /ɑ/ are shown in Figure 2b.

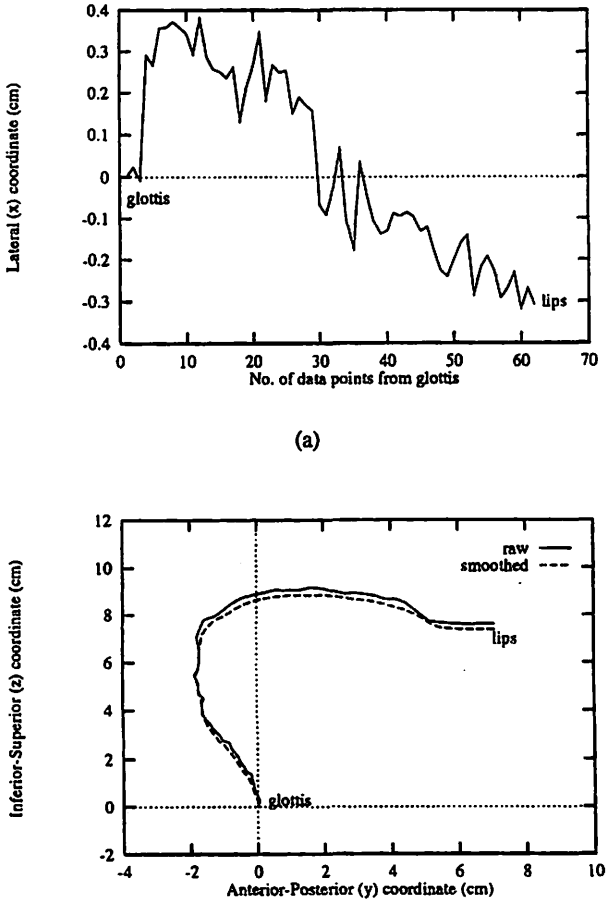


Figure 2: Off-axis variations of the vocal tract centerline, a) lateral (x) coordinate variations as a function of section number; first section is at point just above the glottis and the last point is at the lips, b) raw and smoothed vocal tract profiles in the midsagittal plane.

It was difficult to locate the seed voxels at exactly the mouth termination plane and the plane just above the glottis. Since the cross-sectional areas of the these planes are important end points for each area function, a region of interest analysis was used to measure the area of the mouth termination by viewing the interpolated airway coronally and the area just above the glottis was measured from an axial slice. Additionally, for each of these cross sections, the centroid coordinates were determined and added to the data set generated by the iterative bisection algorithm. The region of interest analysis was also used to determine the cross-sectional areas of the lateral pathways for the /l/ by measuring coronal slices through them. The iterative bisection method used for this study did not perform well for branching airways.

Each area function was generated as the set of x-y pairs that include the length coordinate (L_n) and the corresponding area (A_n),

$$\left(L_n = \sum_{k=0}^{n} L_k \ , \ . \quad A_n\right) \qquad (2)$$

The area function is assumed to begin at the glottal end of the vocal tract (i.e. L=0 is just above the glottis) and terminate at the lips.

Exclusion of the Teeth

A small concentration of hydrogen causes the teeth to be poorly imaged with MRI making these structures appear to be the same gray scale value as air. To avoid

including the teeth during the segmentation process, an estimation of their extent into the oral cavity was required. Story (1995) and Story et al. (1996) reported an imaging experiment that used electron beam computed tomography (EBCT) to acquire axial image sets for two vowels, /i/ and /ɑ/ for the same subject (BS) imaged in the present experiment. Since EBCT uses x-ray radiation, the teeth are well defined in the image slices in which they are present. Thus, the EBCT image sets provide a convenient means of measuring the dimensions of the teeth, in terms of their extent into the airway, which can then be used as an aid in segmenting the air from tissue and teeth in the MR image sets.

Using the EBCT image set for the vowel /ɑ/, the teeth were removed by segmenting them from all other tissue and air and then applying the shape-based interpolation algorithm described previously to extract only the teeth from the image set. A shaded surface rendering of the teeth extracted from the /ɑ/ vowel is shown in Figure 3. The dimensions of this "cast" of the teeth can now be measured with a general region of interest analysis program (Hoffman et al., 1992). The space occupied by the teeth within each axial MRI slice in the region of the oral cavity was estimated from these measurements and the airway segmentation was corrected by "painting out" a space approximately equivalent to the size of the teeth. In Figure 2b, the spatial contribution of the teeth have been excluded using this method; i.e. observe the black regions lateral to the airway that would have otherwise been included as air space.

A demonstration of this method of removing the contribution of teeth is shown in Figure 4 in which two perspective views of the 3-D reconstruction of the /ɑ/ vowel oral cavity using the MRI data is given. The upper pharynx is seen at the left side of the figure which bends into the oral cavity, and terminates at the lips as represented by the flat plane. Figure 4a was reconstructed assuming that the teeth did not exist (i.e. all of the space occupied by the teeth was segmented as air) while Figure 4b represents the same vowel shape but after using the dimensions of the teeth obtained from the EBCT images to estimate their extent and consequently eliminate their contribution to the MR derived air space image. Note that since the dimensions of the teeth do not change, their dimensions measured from the EBCT data can also be used in segmenting any other vowel shape acquired with MRI. Thus, one EBCT image set provides the necessary information for estimating the location and size of the teeth in any MR image set when relating the teeth dimensions to their soft tissue surroundings.

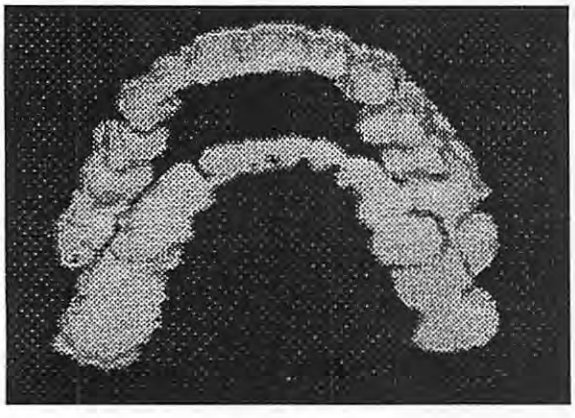
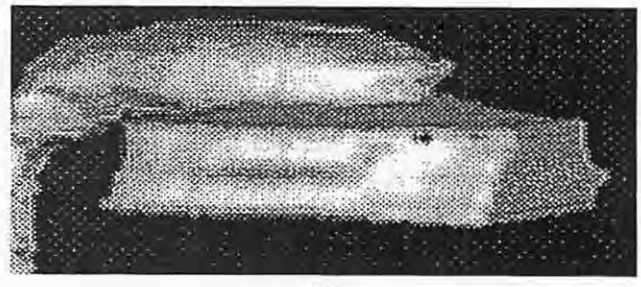


Figure 3: A "digital cast" of the teeth using the EBCT image set for the vowel /ɑ/ (inferior, frontal view; i.e. looking up toward the teeth from below).

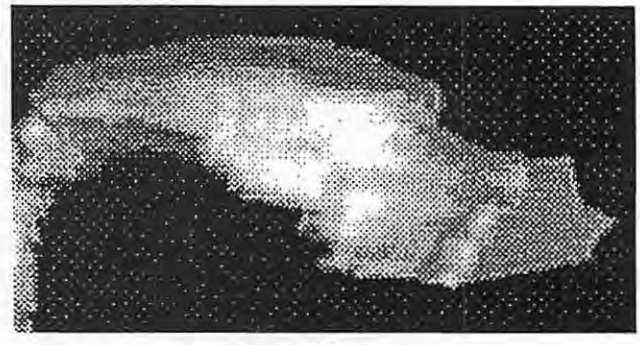
Speech Simulation and Acoustic Recording

Simulation Model

A wave-reflection analog vocal tract model (Story, 1995) was used to simulate the vowel sounds based on the area functions measured from the MR image sets. Energy losses due to the yielding properties of the vocal tract walls, fluid viscosity, and radiation from the mouth have been incorporated into the model. An acoustic side branch representing the piriform sinuses was also implemented. The model was sampled at a frequency of 44100 Hz and each finite section of the area function represented a tube length of 0.396 cm.



(a)



(b)

Figure 4: Shaded surface displays of the vowel /ɑ/ where the flat surface in the upper right portion of the picture is the mouth termination plane: a) perspective view including *the teeth - the "fins" that are visible on right side of the oral region (they also exist on the left side) are due to the included presence of the teeth, b) perspective view* excluding *the teeth.*

The simulation of each vowel sound was performed by injecting a parameterized glottal flow waveform (Titze et al., 1994) into the glottal end of the vocal tract to serve as the voice source. A parameterized source allows precise control of fundamental frequency (pitch) and spectral content (glottal waveshape) so that all of the simulations can be produced with exactly the same voice.

Acoustic Recordings of Natural Speech

In order to compare simulations to the natural speech of the subject, a high quality audio recording was made in which the subject produced speech sounds that corresponded to the static shapes that were acquired with MRI. An attempt was made to simulate, as closely as possible, the conditions experienced during the MRI sessions. The subject was set in a supine position on the suspended floor of an anechoic chamber. The Frankfort plane was perpendicular to the floor. No attempt was made to replicate the acoustic signal produced by the MR scanner but the subject did use ear plugs to create similar acoustic feedback conditions in terms of perceived vocal intensity. The speech sounds were recorded three separate times.

Analysis by Linear Predictive Coding (LPC)

The natural and simulated speech samples were compared in terms of linear prediction spectra. The LPC algorithm was a 50 pole autocorrelation method (Markel and Gray, 1976). For each recorded vowel sound, a sample approximately 0.10 seconds long was extracted from near the beginning of the total recorded production. An LPC analysis was then performed on this sample. Once the LPC spectrum had been computed, a peak picking algorithm based on a parabolic interpolation method (Titze et al., 1987) was used to find the first three resonance frequencies of the vocal tract. The simulated sounds were subjected to the same analysis as the natural speech.

Results and Discussion

Vowels

Sagittal views of the surface rendered airways, along with their "raw" area functions (i.e. points are not necessarily spaced in equal increments), are shown in Figures 5,6,7, and 8 for the vowels produced by subject BS. Since the imaging was performed on static vocal tract shapes, the /ɜ/ and /l/ are included in the group of "vowels"; i.e. any open tract shape was considered to be a vowel. A rotated and tilted view for the /l/ is shown so that the lateral pathways are visible. Additionally, five coronal cross-sections spaced at intervals of 0.188 cm show the shape of the lateral pathways. For each surface rendered airway, the most inferior point of the 3-D shape begins with the uppermost section of the trachea. Above the trachea, the airway becomes small in the region of the glottis and then widens, more or less depending on the vowel or consonant, into the lower pharyngeal section. The finger-like extensions that hang down below the pharynx are the piriform sinuses. As the vocal tract bends into the horizontally oriented oral cavity the airway becomes narrow or wide and terminates with the mouth opening. For all area functions, the glottis is considered to be at the 0 cm point. The figures are grouped to show a progression of shapes that first exhibit a constricted oral section, then a constricted pharyngeal region, and finally a constricted mid-section of the tract.

A general observation with regard to all of the vocal tract shapes is that they show a widening of the tract above the glottis that starts at 2 to 3 cm and narrows again at approximately 4 to 5 cm. This is primarily due to the piriform sinuses merging with the main vocal tract tube. The point above the glottis at which the widening begins is nearly the same for all vowels, suggesting that the analysis of each vowel was consistent in terms of defining the glottal termination.

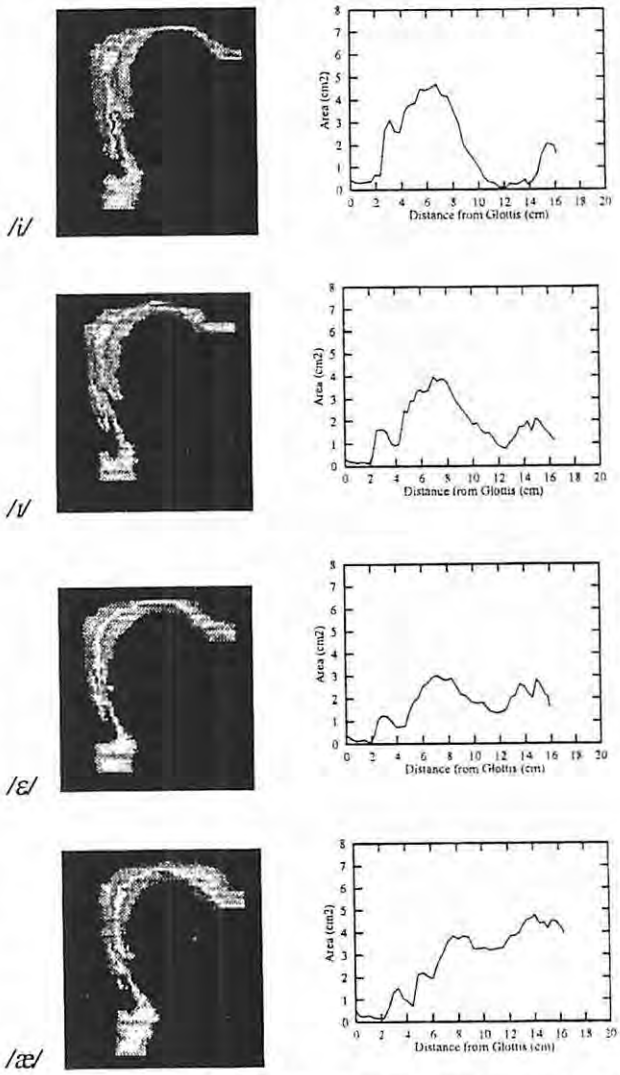


Figure 5: Surface rendered airways and "raw" area functions for /i/, /l/, /ɛ/, and /æ/ (subject BS). Shaded surface display, left; area function, right.

The first four airway surface renderings and area functions in Figure 5 (/i/, /ɪ/, /ɛ/, and /æ/), show that the region just above the glottis is nearly constant up to 2 cm above the glottis. Beyond this point there is an abrupt increase in area, with the /i/ vowel achieving the greatest area of 4.2 cm^2. The other two "front" vowels, /ɪ/ and /ɛ/ have successively lower areas in this region (peak areas of 3.5 cm^2 and 2.9 cm^2, respectively). In the front half of the area function the cross-sectional area of the /i/ vowel drops to values on the order of 0.2 cm^2, far below the/ɪ/ and /ɛ/ areas which are in the range of 1 cm^2. Thus, the /i/ vowel defines the extreme areas in both the front and back vocal tract cavities. The vowel /ɪ/ reaches less extreme areas than the /i/ but more extreme than /ɛ/. The area function for the /æ/ vowel is similar to /i/, /ɪ/, and /ɛ/ in the pharyngeal region and to /ʌ/, /ɑ/, and /ɔ/ (shown in Figure 6) in the oral cavity. It appears to be a transition vowel between the front and back categories.

Surface renderings of the airways for /ʌ/, /ɑ/, and /ɔ/ (Figure 6) all show a similar constricted pharynx and widened oral cavity, except that the /ɑ/ clearly has a larger mouth opening. The area functions for these "back" vowels are all very similar in the region between the glottis and about 4.5 cm above the glottis at which point the /ʌ/ remains at approximately 0.8 cm^2 and the /ɑ/ and /ɔ/ both drop to about 0.25 cm^2. All three vowels steadily increase in area from about the 7 cm point up to the 13.5 cm point with /ɔ/ achieving the greatest area of 7 cm^2. From the point of peak area out to the mouth termination (at approximately 17.5 cm) the area functions for /ʌ/ and /ɔ/ steadily decrease to a final mouth opening area of about 1.0 cm^2. The area function for /ɑ/ initially exhibits a similar decrease, but at 15.5 cm the area begins to rise and continues out to a mouth termination of 5.0 cm^2.

The airway images for /o/, /ʊ/, /u/ and /ɝ/ in Figure 7 clearly show the profile of the tongue and especially the tongue tip. The area functions demonstrate how each of these vowels are more or less divided into two distinct chambers by a tight constriction. The /o/ and /ʊ/ both have large cavities in the front of the tract and smaller cavities in

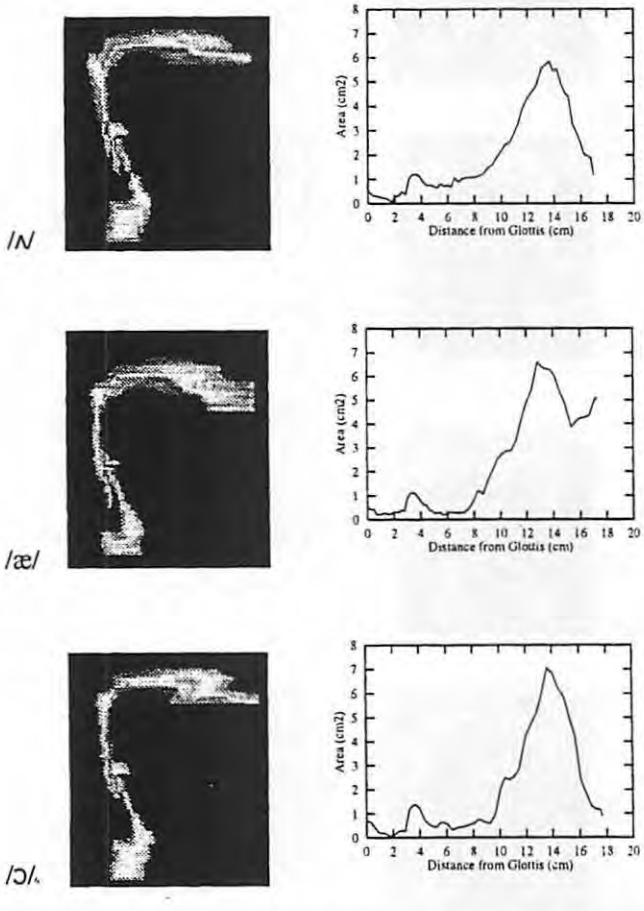


Figure 6: Surface rendered airways and "raw" area functions for /ʌ/, /ɑ/, and /ɔ/ (subject BS). Shaded surface display, left; area function, right.

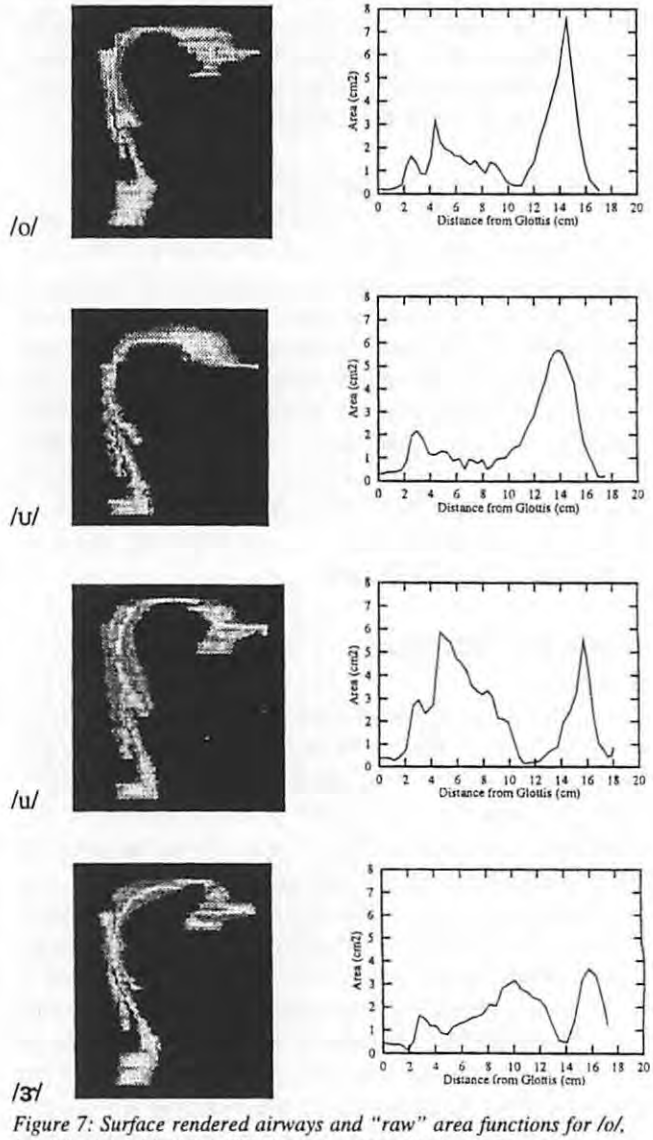


Figure 7: Surface rendered airways and "raw" area functions for /o/, /ʊ/, /u/, and /ɝ/ (subject BS).Shaded surface display, left; area function, right.

the back. However, the constriction in the /o/ is 2 cm closer to the mouth and more well-defined than for the /u/. The front and back cavities of /u/ both have cross-sectional areas that reach about 5.5 cm^2, providing a more equal front-back distribution of cavity size than the other three vowels. The /ɜ/ shape also exhibits a nearly equal front-back cavity distribution, but the largest cross-sectional areas are on the order of 3 cm^2. In addition, the point of constriction occurs only 2.5 cm posterior to the mouth, which is closer than for any of the three vowels /o/, /ʊ/, and /u/.

The /l/ is shown in Figure 8 from a rotated and tilted perspective to give an indication of the lateral air flow (and acoustic wave) path around the tongue. The gap between the oral section and the lip section is due to the presence of the tongue contacting the hard palate. Ideally a lateral path would exist on both sides of the oral cavity but the subject imaged in this experiment closed off the right pathway (right refers to the subject's right). The area function for the /l/ shows a generally uniform area in the range of 2-3 cm^2, throughout the pharyngeal region, but a rapid area increase up to 7 cm^2 just prior to the lateral pathway constriction. Immediately following the constriction the area again increases rapidly with a final mouth termination area of 3.7 cm^2.

To gain a better view of the lateral pathways, a series of 5 coronal slices from the interpolated vocal tract shape is shown in Figure 9. The slices are spaced at intervals of 0.188 cm. The series begins with a section located approximately 1 cm posterior to the mouth termination plane (Figure 9a) and the view is from the perspective of looking into the vocal tract from the anterior end (mouth); i.e. the right side of each figure is the subject's left. At this point the airway is a contiguous region. Figure 9b shows the airway split into three regions; the right and left region define the lateral air spaces while the region in the lower center represents the small air space below the tongue tip. A fourth region appears in the superior portion of Figure 9c which is part of the main vocal tract airway. Figure 9d shows the pathway on the right side (subject's left) split into two regions. Figure 9e shows the merging of the right pathway with the main airway. Note that the lower part of the pathway on the right side of the figure, the entire pathway on the left side, and the air space under the tongue never reconnect with the main airway. Thus, the constricted portion of the /l/ area function at approximately 17 cm from the glottis was based on measurements of only the lateral pathway that did reconnect with the main airway.

Two of the area functions shown in this section, /ɛ/ and /ɔ/, contain an extremely small area, less than 0.1 cm^2, located at about 2 cm from the glottis (see Table 3 on the following page for numerical values). It is interesting that the location of this small area is almost exactly the same for the two vowels. A close check of the original image slices indicated that the airway was, in fact, very constricted in this region of the vocal tract. However, for an area as small as these, only a few voxels define the airway; e.g. an area of 0.06 cm^2 would be comprised of only 7 voxels (0.06 cm^2 / (.0938

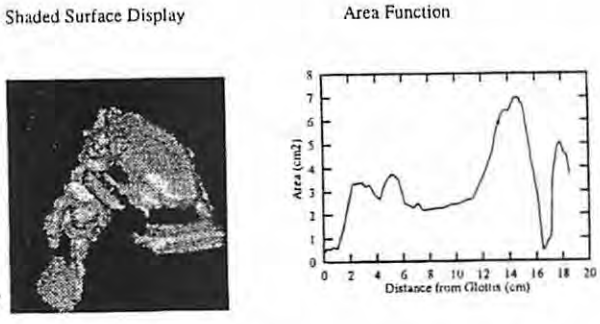


Figure 8:Surface rendered airway and "raw" area function for /l/ (subject BS).

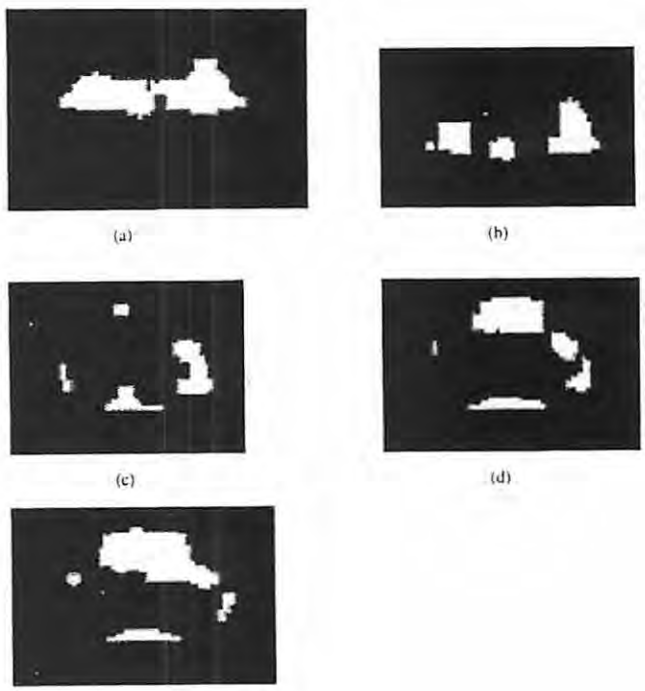


Figure 9: A series of five coronal slices through a portion of the oral region of the interpolated vocal tract shape for /l/. The slices show the shape of the lateral air space. The interval between slices is 0.188 cm and the slices are shown from a view looking into the vocal tract from the mouth: a) slice located approximately 1cm posterior to the mouth termination plane, b) airway is split into three regions; the right and left region define the lateral air spaces while the region in the lower center represents the small air space below the tongue tip, c) a fourth region appears which is part of the main vocal tract airway, d) pathway on the right side (subject's left) splits into two regions, e) right pathway merges with the main airway - the lower part of the pathway on the right side of the figure, the entire pathway on the left side, and the air space under the tongue never reconnect with the main airway.

$cm)^2 \approx 7$). A slight increase in the threshold value for the seeded region growing algorithm (see section 2.4) could have easily included a few more voxels, which would increase the measured cross-sectional area. However, it was decided not to adjust the threshold value for segmentation of these small areas, but maintain a consistent threshold throughout the vocal tract. A potential problem with imaging these vowels is that they are typically not sustained for long periods of time in normal speech (i.e. short vowels), and as a result may have been more susceptible to movement artifact and blurring than say the point vowels /i, æ, ɑ, u/ which are often sustained. While such small areas may be due to measurement error or movement artifact, it is shown later in section 4.6 that these two vowels produced very close acoustic matches to natural speech in terms of locations for formants F1 and F2.

Nasals and Plosives

Nasal and plosive tract shapes are included in the same section because they both possess occlusions of the vocal tract. In fact, the occlusions should be analogous for /m/ and /p/, /n/ and /t/, and /ŋ/ and /k/. Since the imaging protocol allows only static shapes to be acquired, no distinction between voiced and unvoiced plosives can be made; i.e. the shape for a /p/ is considered to be equivalently useable as a /b/, /t/ or /d/, etc. The airway surface renderings and area functions for each of the three nasal and plosive consonants are shown in Figures 10 and 11, respectively. The significant feature in each of these shapes is the location of the main vocal tract occlusion, and in the case of the nasals, the presence of the open velopharyngeal port through which air flow and acoustic waves are diverted. The tract occlusion for the /m/ and /p/ occurs at the lips so that a "break" in the vocal tract shape is not observed other than an apparent shortening

Table 3.
Equal interval (0.396825 cm) area functions for 18 vocal tract shapes.
Section 1 is the glottal end of the vocal tract and "n.c." denotes the nasal coupling.

Sect. No.	/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ɜ˞/	/l/	/m/	/n/	/ŋ/	/p/	/t/	/k/
1	0.33	0.20	0.21	0.22	0.33	0.45	0.61	0.18	0.32	0.40	0.41	0..55	0.57	0.26	0.51	0.31	0.38	0.34
2	0.30	0.17	0.13	0.26	0.28	0.20	0..28	0.17	0.39	0.38	0.38	0.63	0.57	0.24	0.47	0.39	0.50	0.35
3	0.36	0.18	0.16	0.26	0.23	0.26	0.19	0.23	0.39	0.28	0.40	0.75	0.21	0.17	0.45	0.42	0.40	0.49
4	0.34	0.18	0.14	0.16	0.15	0.21	0.10	0.28	0.43	0.43	0.29	1.80	0.58	0.21	0.30	0.71	1.07	0.78
5	0.68	0.10	0.06	0.13	0.17	0.32	0.07	0.59	0.56	0.55	0.13	2.98	2.18	0.15	0.48	1.28	1.38	1.31
6	0.50	1.08	0.78	0.21	0.33	0.30	0.30	1.46	1.46	1.72	0.53	3.56	3.15	0.36	0.67	1.80	1.65	1.34
7	2.43	1.66	1.25	0.83	0.39	0.33	0.18	1.60	2.20	2.91	1.58	3.45	2.96	1.37	0.83	1.70	1.29	1.19
8	3.15	1.64	1.24	1.50	1.02	1.05	1.13	1.11	2.06	2.88	1.56	3.22	2.89	1.66	0.96	1.43	1.01	0.94
9	2.66	1.19	0.99	1.35	1.22	1.12	1.42	0.82	1.58	2.37	1.22	3.20	3.70	1.35	1.43	1.25	0.92	0.69
10	2.49	0.92	0.72	0.99	1.14	0.85	1.21	1.01	1.11	2.10	1.19	2.67	4.21	0.90	1.14	0.90	0.86	0.92
11	3.39	1.13	0.73	0.69	0.82	0.63	0.69	2.72	1.11	3.63	1.00	3.02	3.57	0.71	0.84	2.06	1.03	1.45
12	3.80	2.48	1.06	1.35	0.76	0.39	0.51	2.71	1.26	5.86	0.77	3.55	3.59	0.93	0.69	2.77	1.60	1.73
13	3.78	2.76	1.77	2.32	0.66	0.26	0.43	1.96	1.30	5.63	0.92	3.76	2.97	1.41	0.82	2.19	2.46	1.67
14	4.35	2.97	1.97	2.13	0.80	0.28	0.66	1.92	0.98	5.43	1.19	3.53	3.17	2.07	0.86	2.35	2.24	2.13
15	4.50	3.43	2.46	1.94	0.72	0.23	0.57	1.70	0.93	4.80	1.27	2.62	3.25	2.12	0.57	2.67	2.47	1.61
16	4.43	3.32	2.70	2.17	0.66	0.32	0.32	1.66	0.83	4.56	1.35	2.40	2.58	2.04	0.81	2.17	2.86	1.56
17	4.68	3.48	2.92	2.85	1.08	0.29	0.43	1.52	0.61	4.29	1.48	2.32	2.74	2.16	1.00	1.77	2.74	1.54
18	4.52	3.96	3.03	3.26	0.91	0.28	0.45	1.28	0.97	3.63	1.56	2.43	2.77	2.36	0.66	2.09	3.32	1.18
19	4.15	3.79	2.84	3.73	1.09	0.40	0.53	1.44	0.75	3.37	1.61	2.13	2.49	2.52	0.80	2.16	3.83	1.44
20	4.09	3.88	2.84	3.80	1.06	0.66	0.60	1.28	0.93	3.16	1.87	2.27	2.93	2.88	0.97	2.26	3.97	1.12
21	3.51	3.47	2.83	3.69	1.09	1.20	0.77	0.89	0.53	3.31	2.10	2.28	3.33	2.30	0.78	2.26	4.16	0.76
22	2.95	2.98	2.36	3.87	1.17	1.05	0.65	1.25	0.65	3.22	2.01	2.26	2.27	1.93	0.58	2.29	4.41	0.96
23	2.03	2.62	2.14	3.68	1.39	1.62	0.58	1.38	0.95	2.33	2.62	2.33	2.57	1.77	0.46	2.17	4.11	1.09
24	1.66	2.37	2.00	3.20	1.55	2.09	0.94	1.09	0.99	2.07	2.96	2.43	2.17	0.96	0.44	2.13	3.95	0.79
25	1.38	1.99	1.78	3.26	1.89	2.56	2.02	0.71	1.07	2.07	3.07	2.44	1.84	0.89	0.47	2.64	3.64	0.25
26	1.05	1.90	1.81	3.29	2.17	2.78	2.50	0.46	1.39	1.52	3.11	2.54	1.98	1.22	0.41	2.65	3.37	0.00
27	0.60	1.70	1.79	3.19	2.46	2.86	2.41	0.39	1.47	0.74	2.77	2.64	1.73	1.30	0.11	2.30	2.89	0.06
28	0.35	1.44	1.50	3.23	2.65	3.02	2.62	0.32	1.79	0.23	2.67	2.67	1.43	1.30	0.00	2.12	2.61	0.03
29	0.32	1.45	1.37	3.23	3.13	3.75	3.29	0.57	2.34	0.15	2.47	3.16	1.73	1.14	0.00	1.67	2.69	0.09
30	0.12	1.06	1.36	3.40	3.81	4.60	4.34	1.06	2.68	0.22	2.34	3.68	2.08	0.77	0.00	1.44	2.32	0.10
31	0.10	0.87	1.43	3.78	4.30	5.09	4.78	1.38	3.36	0.22	2.25	4.30	2.32	0.34	0.00	1.16	2.04	0.06
32	0.16	0.75	1.83	3.84	4.57	6.02	5.24	2.29	3.98	0.37	1.90	5.14	2.84	0.15	0.00	1.51	1.64	0.03
33	0.25	1.06	2.08	3.98	4.94	6.55	6.07	2.99	4.74	0.60	1.32	5.83	3.51	0.22	0.00	1.76	1.39	0.48
34	0.24	1.29	2.59	4.41	5.58	6.29	7.08	3.74	5.48	0.76	0.76	6.44	4.25	0.21	0.00	1.93	1.26	1.27
35	0.38	1.78	2.54	4.56	5.79	6.27	6.81	4.39	5.69	0.86	0.44	6.54	4.79	0.00	2.18	1.98	0.87	2.28
36	0.28	1.83	2.11	4.79	5.51	5.94	6.20	5.38	5.57	1.82	0.45	6.91	4.61	0.00	4.72	2.21	0.60	2.35
37	0.36	1.70	2.34	4.39	5.49	5.28	5.89	7.25	4.99	2.35	0.92	6.72	4.07	0.00	6.86	2.35	0.10	2.40
38	0.65	1.97	2.74	4.42	4.69	4.70	5.04	7.00	4.48	2.55	2.05	5.61	3.64	0.00	8.58	2.45	0.00	2.41
39	1.58	1.92	2.19	4.23	4.50	3.87	4.29	4.57	3.07	3.73	3.25	4.08	2.84	0.00	8.76	2.37	0.00	4.21
40	2.05	1.62	1.60	4.56	3.21	4.13	2.49	2.75	1.67	5.47	3.63	2.73	1.42	0.00	7.21	2.47	0.13	3.37
41	2.01	1.36		4.31	2.79	4.25	1.84	1.48	1.13	4.46	3.59	0.45	0.29	1.59	4.92	1.75	0.18	2.46
42	1.58	1.18		3.94	2.11	4.27	1.33	0.68	0.64	2.39	3.07	0.90	0.00	1.60	3.21	1.09	1.48	2.46
43					1.98	4.69	1.19	0.39	0.15	1.10	2.25	3.92	0.00	1.69	2.17	0.70	1.60	2.14
44					1.17	5.03	0.88	0.14	0.22	0.77	1.20	4.99	0.00	1.17	1.41	0.00	1.43	1.50
45										0.41		4.57						
46										0.86		3.70						
n. c.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.04	1.26	1.09	0.00	0.00	0.00
VT length	16.67	16.67	15.88	16.67	17.46	17.46	17.46	17.46	17.46	18.25	17.46	18:25	17.46	17.46	17.46	17.46	17.46	17.46

relative to the previously shown vowel shapes. Since the mouth is closed, the only outlet for sound during production of /m/ is through the nasal cavity, which means that the entire oral cavity becomes a side branch resonator. Area functions for the /m/ and /p/ demonstrate a generally uniform vocal tract shape with cross-sectional areas in the range of 2-3 cm^2 in the pharyngeal region, followed by a slight narrowing to about 1.5 cm^2 and subsequent expansion in the oral section to 4.5 cm^2 for /m/ and 2.5 cm^2 for /p/ before closing down to zero area. This particular shape is probably due to the requirement that the subject produce the /m/ while imagining that the preceding and succeeding vowels would be a schwa (neutral) vowel. Had an /ɑ/ or /i/ been used as the imagined "target", the area function might have shown a closer resemblance to those vowels. Simulation of dynamic speech segments may require different area functions for the /m/ and /p/, depending upon the surrounding vowel shapes. The most important information to take from the /m/ images is the area and location of the nasal coupling port. The port area was measured to be 1.04 cm^2 and is located approximately 8 cm from the glottis.

The /n/ and /t/ shapes exhibit a break in the vocal tract where the tongue contacts the hard palate and buccal walls. For the /n/, a small side branch cavity is created by the airspace between the location of the nasal coupling (8 cm from the glottis) and the constriction. The cavity of air near the mouth termination was included only to demonstrate the "break" in the tract. Acoustically speaking, this volume of air is inactive for this particular shape. Area functions for the /n/ and /t/ both show a widened upper pharyngeal region with an abrupt closure in the oral cavity, with the /t/ requiring more volume than the /n/. The vocal tract is also about a centimeter longer for the /t/ than for the /n/, which places the point of constriction at a more anterior location. The nasal coupling area was determined to be 1.09 cm^2 and located 8 cm from the glottis.

Surface rendered airways for the /ŋ/ and /k/ are nearly identical except for the open velopharyngeal passage in the /ŋ/. The occlusion of the tract for /ŋ/ occurs at a point that is posterior enough that there is just a very small side branch cavity created. The 3-D image shows an almost straight tract from the glottis up into the velopharyngeal tube. The area function for /ŋ/ verifies this notion, where the indicated branch point is only 2 cm posterior to the occlusion. The coupling area was found to be 1.26 cm^2 and again located 8 cm from the glottis. Like the airway renderings, the area functions show the similarity of the two consonants. The cavities in front of the occlusion are quite different, but since

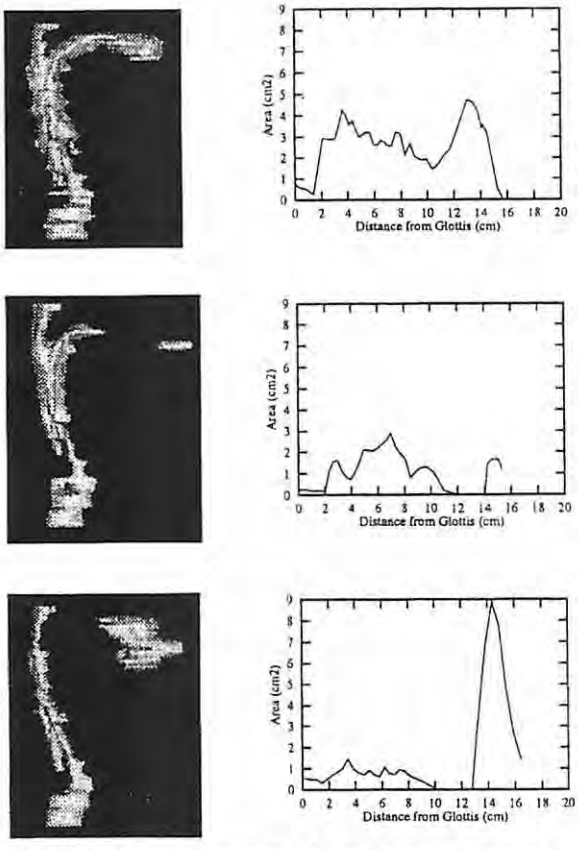


Figure 10: Surface rendered shapes and area functions for nasals (subject BS). Shaded surface display, left; area function, right.

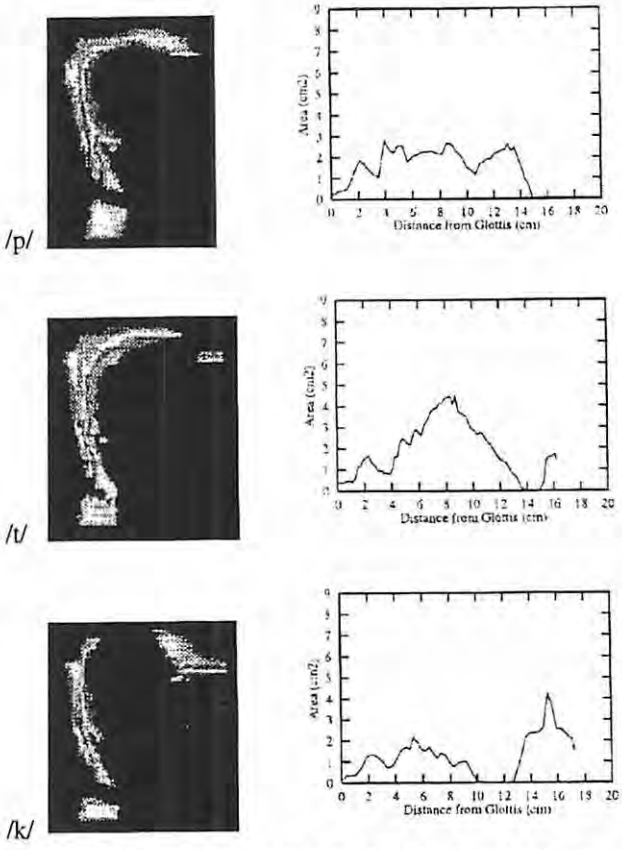


Figure 11: Surface rendered shapes and area functions for plosives (subject BS). Shaded surface display, left; area function, right.

this section of the tract is acoustically inactive, it is not a significant feature.

Numerical Area Functions

The linear distance between consecutive oblique cross-sections determined by equation (1) is not necessarily constant throughout the vocal tract. The area functions produced by equation (2) contained 50 to 80 data points, depending on the number of iterations that were performed during the analysis and were spaced at intervals that ranged from 0.2 cm to 0.4 cm. The wave-reflection algorithm discussed in Section 3.1 requires that each area function be discretized at equal length intervals and the length of the final area function must be an even integer multiple of the length interval. To transform the "raw" area functions determined by equation (2) into a usable form for the speech simulation, they were first normalized to a discrete length that was close to their measured length. A wave-reflection type model dictates that the length of each finite section of the area function be equal to the speed of sound divided by two times the sampling frequency. For this study, the length interval was chosen to be 0.396825 cm which results from using a sampling frequency of 44.1kHz and a speed of sound equal to 350 m/s. If the measured length of the tract was 17.2 cm, a 44 section area function would be chosen to represent it since 17.46 cm is the closest discrete length. The raw area function would first be normalized to 17.46 cm and then transformed to an equal interval function by fitting it with a cubic spline and then sampling the resulting curve at equally spaced intervals. If desired, all of the raw area functions could be normalized to one particular length so that all of the discretized functions would have, for example, 44 sections. The effect would be a slight compression of the longer area functions and an expansion of the shorter area functions.

Equal interval (0.396825 cm) area functions are shown in Table 3 and should be read by assuming that the glottal end of the tract is represented by section 1 while the last section for each phoneme represents the mouth termination. The second row from the bottom of the table, labeled "n.c.", is the nasal coupling area, which is zero for all shapes except the nasal consonants. The last row indicates the vocal tract length for each area function.

Comparison of Natural and Simulated Vowel Sounds

A comparison of the first three formant locations extracted from LPC spectra for the natural and simulated versions of each vowel is shown in Figure 12. The natural speech sounds were recorded three separate times, so that each diamond shaped symbol represents the mean formant values of the three sets while the error bars indicate the range. The dashed line passes through the simulated values.

The first formants extracted from the simulations show, eight out of the twelve vowels falling within the range of the natural speech, while the remaining four vowels were positioned outside the range. For five of the vowels, the second formants from the simulations fell within the natural speech range while the other seven fell outside the range. Similarly, the third formants show five vowels positioned within the range of the natural speech, but significant deviations from natural speech are observed for /ɔ/, /o/, /ɜ/, and /l/. The large range shown for the third formant of /ɜ/ was an artifact caused by the second and third formants merging in one of the recorded cases so that formant peak picking algorithm skipped to the next formant and reported it as the third. Simply by chance, the mean value of the third formant for /l/ deviated from the simulated version by only 11 percent. The true deviation in this case was much greater.

The data presented in Figure 12 is tabulated in Table 4, along with the percentage error for the simulations relative to the mean formant values of the natural cases.

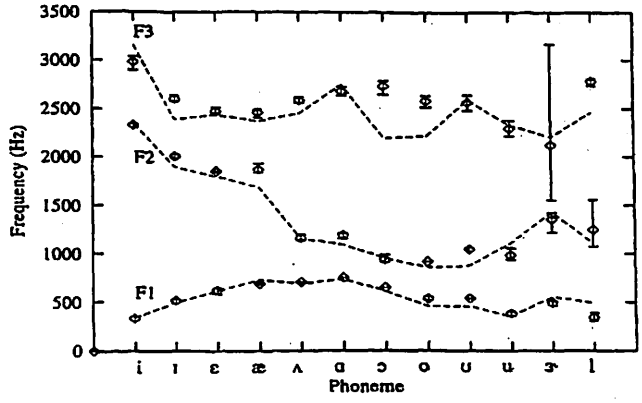


Figure 12: Comparison of formant locations for F1, F2, and F3 extracted from LPC spectra for both natural and simulated speech. The mean value for three separate recording sessions of natural speech is depicted by the diamond symbols and the error bars indicate the range. The dashed line represents the simulation.

Table 4.

First three formants from natural recorded speech and simulated speech based on the area functions given in Table 3. The superscript "N" denotes the natural speech and "S" the simulated version. The Δ's represent the percent error of the formants from simulated speech relative to the mean value of the natural speech formants.

	i	ɪ	ɛ	æ	ʌ	ɑ	ɔ	ʊ	u	ɜ	l
$F1^N$	333	518	624	692	707	754	654	540	541	389	500	348
$F2^N$	2332	2004	1853	1873	1161	1195	944	922	1045	987	1357	1250
$F3^N$	2986	2605	2475	2463	2591	2685	2739	2584	2568	2299	2124	2785
$F1^S$	337	499	621	732	689	738	618	461	461	356	559	500
$F2^S$	2340	1894	1795	1689	1159	1093	958	861	877	1108	1431	1127
$F3^S$	3158	2388	2436	2370	2454	2757	2195	2217	2596	2334	2206	2574
Δ1	1.3	3.7	0.4	5.8	2.5	2.1	5.5	14.6	14.7	8.5	11.7	43.6
Δ2	0.4	5.5	3.1	9.8	0.2	8.5	1.4	6.6	16.2	12.3	5.5	9.8
Δ3	5.76	8.3	1.6	3.8	5.3	2.7	19.9	14.2	1.1	1.5	3.9	7.6

Across all vowels and formants, percentage errors range from 0.37 percent to 43.6 percent with the majority below 10%. The largest error of 43.6 percent occurred in the first formant for /l/. This is not surprising considering that the oral cavity for /l/, with the lateral air spaces, was far more complex than any of the other shapes, thus more susceptible to measurement error. In fact, one could argue that an /l/ cannot be adequately described by a simple area function implemented in a one-dimensional acoustic model. However, the second and third formants both were in error by less than ten percent so that the measured area function should not be entirely disregarded. The simulation of the two vowels, /ɛ/ and /ɔ/, both produced formant locations for F1 and F2 which deviated from the natural speech by less that 6% . The presence of a very small cross-sectional area in these two area functions at approximately 2 cm from the glottis was discussed in section 4.2.

A study of Figure 12 reveals that, for many of the simulated vowels, the second and third formants tend to be displaced from the natural speech in the direction of a neutral (schwa) vowel formant structure (i.e. F1, F2, F3≈ 500, 1500, 2500 Hz). For example, the second formant for the simulated /æ/ is displaced from the natural speech mean value of 1873 Hz down to 1689 Hz; movement toward a more neutral 1500 Hz. In addition, the third formants for /i/, /ɛ/, /æ/, /ʌ/, /ɔ/,/o/, and /l/ are all displaced from natural speech in the direction of 2500 Hz. This suggests that the subject (BS) tended to centralize the production of vowels during the MR imaging sessions. The vowel centralization could be due to fatigue of the articulatory musculature as well as listening fatigue of the aural system. Fatigue effects are quite possible, since the MRI protocol required the subject to produce many repetitions of a given vowel (approx. 30 repetitions for each vocal tract shape). However, another observation with respect to Figure 12 is that neither the simulated /i/ or /ɑ/ vowels show centralizing effects. In fact, the third formant of /i/ and the second formant of /ɑ/ are displaced in the opposite direction of a neutral vowel. It is possible that the intermediate (short) vowels may require more precise muscular control to maintain the appropriate formant structure than the extreme /i/ and /ɑ/ shapes; /i/ and /ɑ/ are "asymptotic" positions that have physiologically imposed constraints or boundaries. In general, the MRI based area functions may be somewhat centralized and should be considered to be an average shape over many productions of the same vowel.

Comparisons of Area Functions with Previous Imaging Studies

At this point it is of interest to compare the area functions reported in the present study with those given in several previous publications. In the following discussion, whenever reference is made to "subject BS" the reader should assume this to mean the area function(s) from the present study.

Figure 13 shows a graphical representation of the discretized area functions (given in Table 3) for the vowels /i, æ, ɑ, u/ superimposed with the area functions reported by Baer et al. (1991) for the same vowels (two subjects). To be analogous to the area functions in the present study, the Baer et al. area functions have been plotted *without* the contributions of piriform sinuses as indicated in Table I of their paper (p. 811). The area functions for each vowel show the same

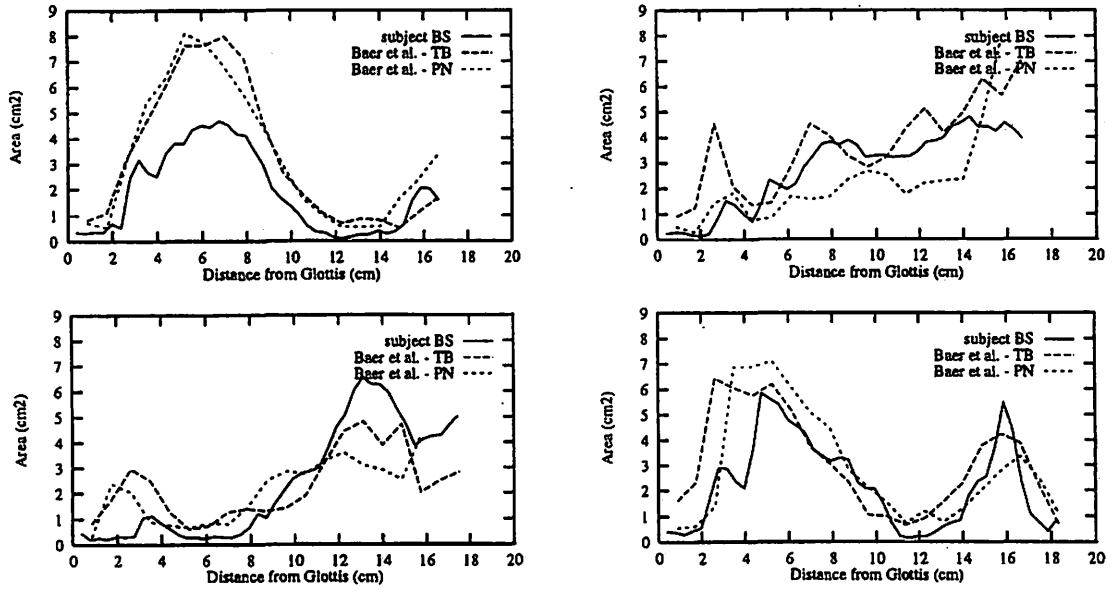


Figure 13: Comparison of the area functions for subject BS (solid) with Baer et al. (1991) area functions for subjects TB(long dashed) and PN (short dashed): a, upper left) /i/; b, upper right) /æ/; c, lower left) /ɑ/; d, lower right) /u/ .

general shapes but demonstrate obvious individual differences. The /i/ vowel in Figure 13a for subject BS is smaller than the Baer et al. area functions in every region of the vocal tract except close to mouth termination where it is slightly larger than that for subject TB. This characteristic might suggest that the areas in the present study were underestimated. But the /æ/ area function (Figure 13b) is larger throughout nearly all of the tract than that for subject PN and mostly smaller than the same vowel for subject TB. In the /ɑ/ area functions (Figure 13c) the BS version is smaller than both TB and PN in the pharyngeal region but larger than both in the oral cavity and at the mouth termination. The /u/ (Figure 13d) is similar for all three subjects except that the BS version has a tighter constriction in the middle part of the tract. In general, across all four vowels, the constricted parts of vocal tract are smaller in the BS area functions. In particular, the region just above the larynx (0 to 3 cm) seems to always have smaller cross-sectional area than the Baer et al. area functions. The reasons for this could be an underestimation of the areas due to the image processing techniques that were used or it could simply be an individual anatomical difference. However, it should be noted that, based on the height and weight information given in section 2.1, the subject BS is not a large person. Thus, small cross-sectional areas in the vocal tract may not be entirely unexpected.

Figure 14 shows the discretized area functions for the vowels /i, ɑ, u/ along with the area functions given in Yang and Kasuya (1994) for an adult male subject. Again the general shape for each vowel is similar, as would be expected, but individual differences are apparent. The glottal end (0 cm) of each of the Yang and Kasuya area functions has an area 0.5 cm^2 to 1.0 cm^2 larger than the corresponding BS version. The major constrictions are again smaller for the BS area functions but interestingly the front and back chambers for the /u/ vowel are on the order of 2 cm^2 larger than for the Yang and Kasuya version. This was not the case when comparisons were made with the Baer et al. area functions. One interesting aspect of the Yang and Kasuya data is that the /i/ vowel had a longer vocal tract length than either the /ɑ/ or the /u/. In the present study and the Baer et al. paper, the /u/ vowel has the longest vocal tract length.

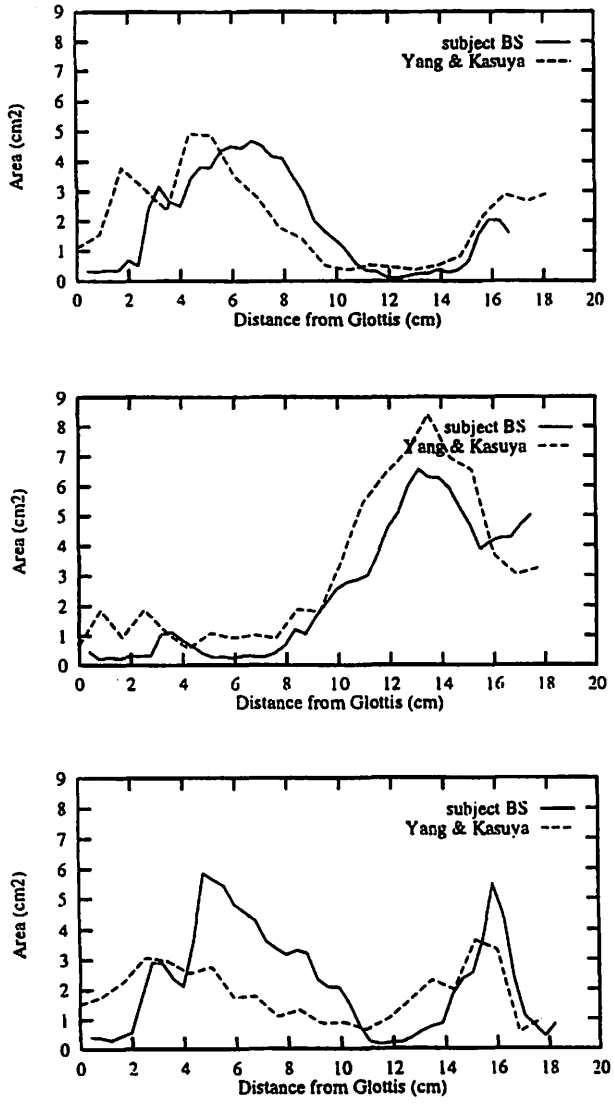


Figure 14: Comparison of the area functions for subject BS (solid) with Yang and Kasuya (1994) (dashed): a), top /i/; b, center) /ɑ/; c, lower) /u/.

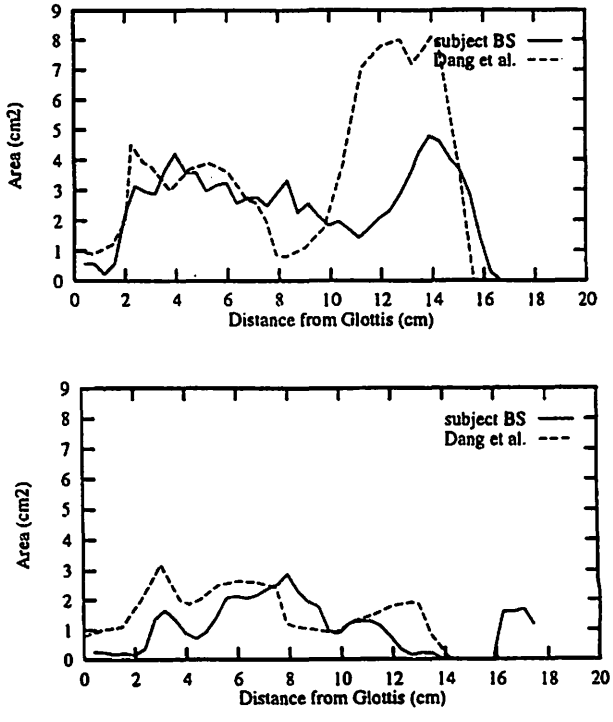


Figure 15: Comparison of the area functions for subject BS (solid) with Dang et al. (1994) (dashed): a, upper) /m/; b, lower) /n/.

In addition to an extensive analysis of the nasal tract, Dang et al. 1994 also present vocal tract area functions for the nasal consonants /m/ and /n/. The numerical values for these area functions have been estimated from their Figure 13 (p. 2097) and are shown with the /m/ and /n/ area functions for subject BS in Figure 15. The /m/ (Figure 15a) for both subject BS and the Dang et al. study exhibit a back chamber in the pharynx with a cross-sectional area of about 3.5 cm^2. This region extends from a point approximately 2 cm from the glottis out to the 8 cm point in the Dang et al. area function while the same region would be roughly defined to be between 2 cm and 10.5 cm from the glottis for subject BS. Both area functions also show an expanded oral cavity which reaches a cross-sectional area of 8 cm^2 in the Dang et al. version but only 4.8 cm^2 for subject BS. The point of tract closure occurs 1 cm farther from the glottis for subject BS than for the Dang et al. area function. In Figure 15b, the overall shapes of the /n/ area functions are quite similar. The areas vary between about 1 cm^2 and 3 cm^2 in the region from 3 to 11 cm above the glottis. The point of closure is approximately 0.2 cm farther from the glottis for the Dang et al. area function than for subject BS.

Finally, comparisons of the vowels /i, ɑ, u/ from Fant (1960) with those of subject BS are shown in Figure 16. For all of the vowels, the areas given by Fant seem to be quite large in comparison to those for subject BS and are also large when compared to the Baer et al. and Yang and Kasuya area functions. The areas in the pharyngeal region for the vowel /i/ are approximately 5 cm^2 larger for the Fant version than for the previous study. The Fant area function for /ɑ/ is almost uniformly larger across the entire vocal tract while the /u/ shows exceptionally large front and back chambers that are on the order of 3 to 7 cm^2 larger than the BS version. The Fant /u/ is also more than a centimeter longer than that for subject BS.

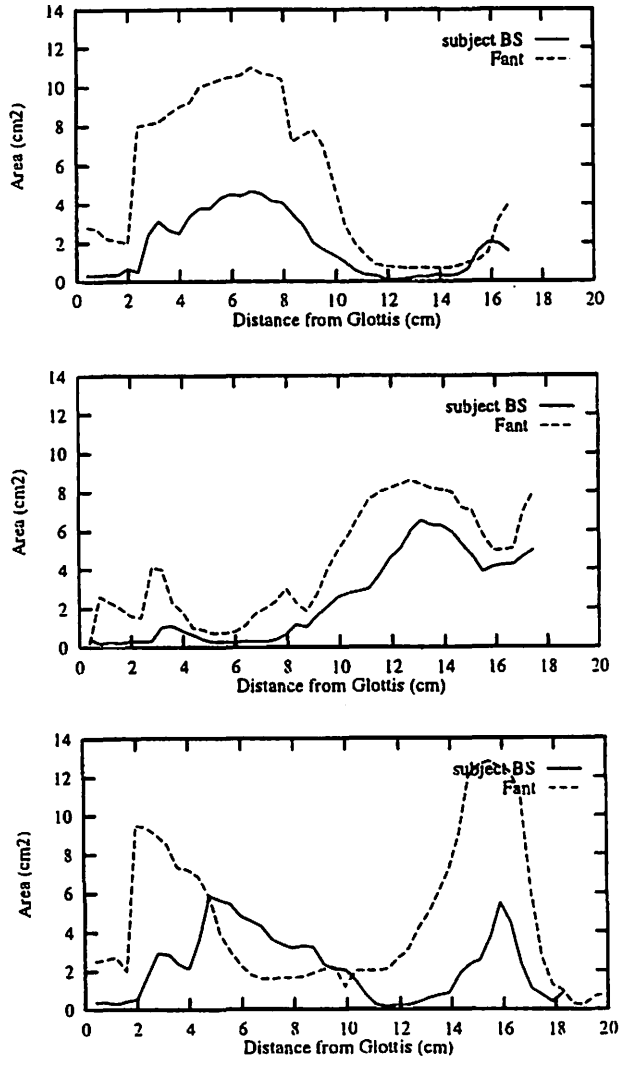


Figure 16: Comparison of the area functions for subject BS (solid) with Fant (1960) (dashed): a, upper) /i/; b, center) /ɑ/; c, lower) /u/.

Conclusion

MRI has been used to volumetrically image the vocal tract airway of one male subject for 12 vowels, 3 nasals, and 3 plosives. The 3-D image sets were segmented to extract the airways which in turn were analyzed to find the cross-sectional area as a function of the distance from the glottis (along the long axis of the vocal tract). This experiment has provided qualitative and quantitative information about the vocal tract shape. The 3-D surface renderings of the various vowel configurations are a helpful aid in visualizing the effect of articulator positioning on resultant vocal tract shape. The numerical area functions provide a speaker-specific inventory of vocal tract configurations that can be used as input for a speech simulation system. However, since the acquisition of the image set representing each vocal tract shape required many repetitions (approx. 30 repetitions), the area functions need to be considered as an "average" shape for a particular vowel or consonant. Due to the large number of repetitions, some fatigue of the articulatory musculature would be expected and as a result, the tract shapes may be somewhat centralized; i.e. fatigue effects may tend to slightly move the vocal tract toward a more neutral or schwa-like shape.

The extracted area functions were used as input to a computer model of one-dimensional acoustic wave propagation in the vocal tract. The simulated speech sounds were compared, in terms of formant locations, to recorded natural speech of the subject was imaged. Results indicated that the formant locations were reasonably well represented but some of centralizing effects mentioned above were observed.

The area functions for several vowels and nasal consonants obtained in this study were compared to those given in four previous imaging studies of vocal tract shape. The variability in the area functions observed across these studies are likely due to differences in imaging techniques and procedures, image processing and analysis, and, maybe most importantly, anatomical and physiological differences of the subjects who were imaged. The general shape of each vowel is similar while the details maintain uniqueness. This means, for example, that the /ɑ/ vowel area functions compared in section 4.7 could all be used to generate an /ɑ/ sound but each would possess a unique vowel quality or formant structure.

Measurements of the nasal tract, trachea, and fricative consonants will be presented in the future to augment the data set presented in this paper. A more extensive presentation of speech modeling using these area functions is also planned. Additional work is needed to collect vocal tract shape inventories for more subjects. In particular, there exists little morphological information regarding the female vocal tract and some future studies should be directed specifically at acquiring an area function inventory for female subjects.

Acknowledgements

The authors would like to thank Carol Gracco for providing the initial scanning parameters from which we developed those shown in Table 1 and three anonymous reviewers for their helpful suggestions and comments. This study was supported in part by grant No. P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders.

Appendix A - Phantom Study

To test the accuracy of the image acquisition and analysis process, a tubular phantom of known dimensions was imaged in the MR scanner. The phantom consisted of a three-section system of hollow (air-filled) plastic tubes connected in a stair-step fashion. The tube system was sealed at both ends and mounted in a water-filled plastic chamber with a 6 inch outer diameter (Figure A-1). The middle section of the tube system had a cross-sectional area of 1.14 cm^2 if only the air was considered and 1.98 cm^2 if both the tube walls and air were included. Both of the end pieces had a 2.45 cm^2 cross-section for air only and a 3.87 cm^2 cross-section for both tube and air. Because of the phantom geometry and its positioning in the scanner, accurate cross sectional area could not be obtained from area measurements in a single slice. True area could only be obtained if the analysis software adequately extracted appropriate oblique sections from the volumetric image data set.

The scans consisted of a series of contiguous, parallel, axial slices that included the entire phantom volume. An axial slice refers to the image plane perpendicular to the axis of the cylindrical phantom chamber. The slice thickness was 5 mm.

A.1 Phantom Measurements

The image sets for the phantom were segmented, interpolated, and analyzed using the methods described in section 2. The set of images does not show the plastic tubing or the air, since hydrogen is not present in either substance.

Table A-1 shows the known cross sections of the large and small tubes (tube 1 and tube 2, respectively) that make up the phantom along with the cross-sections that were measured using the image analysis methods. The measurement overestimated the cross-sectional area of the larger tube by 1.5 percent. However, measurement of cross-sectional area of the small tube was underestimated by nearly 10 percent. Since the boundary of a region of interest cannot be determined any closer than ± 1/2 a voxel around its perimeter, the resolution of the MR image sets would produce a larger uncertainty in measurement of the region area than would a finer voxel resolution. For a large tube, the voxels on a region boundary would contain a small fraction of the total region area, contributing a small percentage error in area measurement. However, the error in measuring the area would be expected to grow progressively larger as the cross-sectional area becomes smaller; i.e. the voxels on the boundary of a region of interest will contain a significant

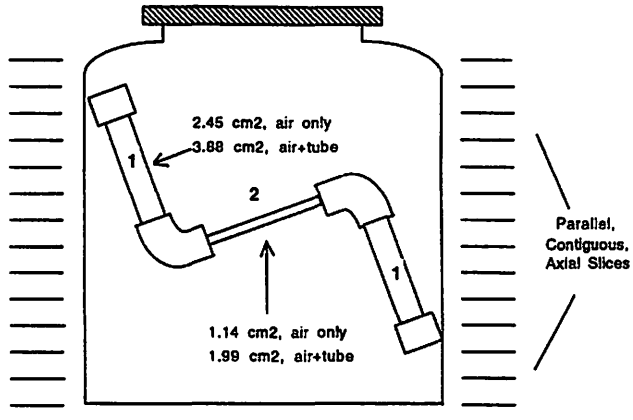


Figure A-1. Schematic diagram of tubular phantom.

Table A-1.
Known and measured cross-sectional areas of phantom sections.

MRI	Tube 1 (cm^2)	Tube 2 (cm^2)
Air & Tube Wall, known	3.88	1.99
Measured MRI Image	3.94	1.8
% error	1.5 %	9.5 %

fraction of the total area, contributing increasing uncertainty in the total area.

A rough estimate of the expected error for the area measurement of an arbitrarily shaped cross-section can be computed by representing the cross-section as an equivalent circle. Using the area, the circumference of an equivalent circle can be computed, which can then be divided by the voxel dimension. This gives an estimate of the number of voxels on the boundary of the cross-section (this process can also be performed by considering an equivalent area square; similar answers will result). Since the uncertainty for the voxels on the boundary is ±1/2 voxel, the uncertainty in the measured area will be the total number of voxels consumed by the cross-section plus or minus one-half of the number of edge voxels. For example, the error produced in the measurement of a 1.99 cm^2 cross-section using MR images would be calculated as follows. The circumference of the equivalent circle would be computed to be 5.0 cm, which, when divided by a voxel dimension of 0.0938 cm, is equivalent to about 53 edge voxels. The total area consumes 226 voxels (i.e. $1.99/(.0938)^2$), and the error associated with it should be ±26.5 voxels (i.e. 53/2). This produces a percentage error of 11.7 percent, which is similar to the measured error for the small diameter tube given in Table A-1.

References

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W., "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," JASA, 90, 799-828, 1991.

Beautemps, D., Badin, P., and Laboissiere, R., "Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data," Speech Communication, 16, 27-47, 1995.

Boyd, D. P., and Lipton, M. J., "Cardiac computed tomography," Proceedings of the IEEE, 71: 298-307, 1983.

Dang, J., Honda, K., and Suzuki, H., "Morphological and acoustical analysis of the nasal and the paranasal cavities," JASA, 96(4), 2088-2100, 1994.

Fant, G., The Acoustic Theory of Speech Production, Mouton, The Hague, 1960.

Greenwood, A. R., Goodyear, C. C., and Martin, P. A., "Measurements of vocal tract shapes using magnetic resonance imaging," IEEE Proceedings-I, 139(6), 553-560, 1992.

Hoffman, E. A., Gnanaprakasam, D., Gupta, K. B., Hoford, J. D., Kugelmass, S. D., and Kulawiec, R. S., "VIDA: An enviroment for multidimensional image display and analysis," SPIE Proc. Biomed. Image Proc. and 3-D Microscopy, 1660, San Jose, CA, 10-13 Feb., 1992.

Hoffman, E. A., and Gefter, W. B., "Multimodality imaging of the upper airway: MRI, MR spectroscopy, and ultrafast x-ray CT," Sleep and Respiration, Eds: F.G. Issa, P.M. Suratt, and J.E. Remmers, 291-301, Wiley-Liss, Inc., New York, NY, 1990.

Hoffman, E. A., Sinak, L. J., Robb, R. A., and Ritman, E. L., "Noninvasive quantitative imaging of shape and volume of lungs," American Physiological Society, 1414-1421, 1983.

Hoffman, E. A., "An historic perspective of heart and lung imaging," in 3D Imaging in Medicine, J. K. Udupa and G. T. Herman (Eds.), 285-311, CRC Press, Lewis Publishing, 1991.

Ishizaka, K., and Flanagan, J. L., "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J., 51, 1233-1268, 1972.

Kiritani, S., Tateno, Y., Iinuma, T., and Sawashima, M., "Computer tomography of the vocal tract," in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (Eds), University of Tokyo Press, Tokyo, 203-206, 1977.

Lakshminarayanan, A. V., Lee, S., and McCutcheon, M. J., "MR imaging of the vocal tract during vowel production," J. Mag. Res. Imag., 1(1), 71-76, 1991.

Liljencrants, J., "Speech Synthesis with a Reflection-Type Line Analog," DS Dissertation, Dept. of Speech Comm. and Music Acous., Royal Inst. of Tech., Stockholm, Sweden, 1985.

Maeda, S., "A digital simulation method of the vocal-tract system," Speech Comm., 1, 199-229, 1982.

Markel, J. D., and Gray, A. H., Linear Prediction of Speech, Springer-Verlag, New York, 1976.

Mermelstein, P., "Articulatory model for the study of speech production," JASA, 53(4), 1070-1082, 1973.

Moore, C. A., "The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images," JSHR, 35, 1009-1023, 1992.

Narayanan, S. S., "Fricative consonants: An articulatory, acoustic, and systems study," Ph. D. thesis, UCLA, Dept. of Electrical Engineering, Los Angeles, CA, 1995.

Narayanan, S. S., Alwan, A. A., and Haker, K., "An articulatory study of fricative consonants using magnetic resonance imaging," JASA, 98(3), 1325-1347, 1995.

Perrier, P., Boe, L-J, and Sock, R., "Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: Modeling the transition with two sets of coefficients," J. Speech and Hearing Research, 35, 53-67, 1992.

Raya, S. P., and Udupa, J. K., "Shape-based interpolation of multidimensional objects," IEEE Trans. Med. Imag., 9, 32-42, 1990.

Rubin P., Baer, T., and Mermelstein, P., "An articulatory synthesizer for perceptual research," JASA, 50, 1180-1192, 1981.

Strube, H. W., "Time-varying wave digital filters for modeling analog systems," IEEE Trans. ASSP, ASSP-30(6), 864-868, 1982.

Sondhi, M. M., and Schroeter, J., "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. ASSP, ASSP-35(7), 1987.

Story, B. H., "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph. D. Dissertation, University of Iowa, 1995.

Story, B. H., Hoffman, E. A., and Titze, I. R., "Speech simulation based on MR images of the vocal tract," SPIE Proc. Physiology and Function from Multidimensional Images, 2433, San Diego, CA, 26 Feb - 2 Mar., 1995.

Story, B. H., Hoffman, E. A., and Titze, I. R., "Vocal tract imaging: A comparison of MRI and EBCT," SPIE Proc. Physiology and Function from Multidimensional Images, 2709, Newport Beach, CA, 10-15 Feb., 1996.

Sulter, A. M., Miller, D. G., Wolf, R. F., Schutte, H. K., Wit, H. P., and Mooyaart, E. L., "On the relation between the dimensions and resonance characteristics of the vocal tract: A study with MRI," Mag. Res. Imag., 10, 365-373, 1992.

Titze, I. R., Horii, Y., and Scherer, R. C., "Some technical considerations in voice perturbation measurements," JSHR, 30, 252-260, 1987.

Titze, I. R., Principles of Voice Production, Prentice-Hall, 1994.

Titze, I. R., Mapes, S., and Story, B., "Acoustics of the tenor high voice," J. Acoust. Soc. Am., 95(2), 1133-1142, 1994.

Udupa, J. K., "Computer aspects of 3D imaging in medicine: A tutorial," in 3D Imaging in Medicine, J. K. Udupa and G. T. Herman (Eds.), CRC Press, 1-64, 1991.

Yang, C-S, and Kasuya, H., "Accurate measurement of vocal tract shapes from magnetic resonance images of child, female, and male subjects," Proc. of ICSLP 94, 623-626, Yokohama, Japan, 1994.

NCVS Status and Progress Report - 9
April 1996, 27-30

Formants in Singers

Ronald Scherer, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts

Abstract

This overview paper discusses certain topics of resonance in performers with an emphasis on classical singing. Topics include general acoustic "rules", formant tuning, the Singer's Formant, voice classification, and vocal covering.

Quality and intensity in vocal performance are more than an aspect of entertainment. Behind performance sound lie secrets of efficient physiology and effective acoustics, as well as cultural and personal artistic creativity. The study of performance sound leads to reconsiderations of what target vocal and vocal tract behaviors might be adopted for the most effective communication.

This report is a selective overview of performance resonance. Topics to be covered include formant tuning, the singer's formant, voice classification, and vocal covering.

The source of sound in phonation is considered to be the glottal volume velocity or its derivative. It is produced by the flow-induced mechanical action of the larynx and has some dependence on the acoustic pressures of the airways. The sound source as a time or frequency signal sent into the vocal tract can be highly varied. Partials or components of the source must exist if we want any resonance structure of the vocal tract to radiate those frequencies from the mouth, if we assume the usual acoustic theory.

Vocal Tract Length and Constriction

Discussions of vocal tract resonance usually begin with the open-closed uniform tube. This is helpful because the glottis end of the vocal tract is nearly closed. As a uniform tube shortens, the formant frequencies increase. The formants themselves act to allow partials from the source to pass through the vocal tract if the partials are near the formant center frequencies. Thus a person will have formant frequencies inherent to his or her vocal tract length. Also, vocal tract length can be voluntarily altered by raising or lowering the larynx and protruding or spreading the lips, thus providing some limited control of formant frequencies through vocal tract length adjustment.

The formant values are highly dependent on the shape of the vocal tract. Both sung and speech sounds must conform to rules governing formant frequency values: (1) Vocal tract lengthening lowers all formant values. (2) A pharyngeal constriction raises the first formant and lowers the second formant, as in the production of /a/. (3) A constriction near the front of the mouth lowers F1 and raises F2, as in /i/. (4) Formant frequencies decrease with lip rounding and increase with lip spreading, similar to the first rule, as is /u/. (5) The first formant increases as jaw opening increases, to be discussed later. Ingo Titze [1] gives physical acoustic reasoning for these effects.

Formant Tuning

Formant tuning occurs when a partial of the glottal source has the same or nearly the same frequency as one of the vocal tract formant center frequencies. When this occurs, that source partial will be radiated with relatively greater intensity. If the fundamental frequency is above the first formant value, however, there may be a clear mismatch of partials and formants, resulting in poor vowel integrity and relatively low intensity. As Sundberg [2,3] has shown, female singers can remedy this problem by modifying the first formant frequency by raising it with a greater jaw opening to bring the first formant in line with the fundamental frequency. Notice that in sung music, the fundamental is prescribed, but the formants can be manipulated. Data from [3] show that jaw opening for a soprano singing different vowels at different pitches increases by approximately 10-15 mm over the range of 250 to 700 Hz for most vowels (but not for /a/). The first formant and fundamental frequency match preserves intensity and a desired (acceptable) quality in classical singing.

Does formant tuning take place in performance voices other than the soprano's? Raphael and Scherer [4] studied the training techniques named "Call" by Arthur Lessac. This technique emphasizes quality and physical sensation, and de-emphasizes vowel specificity. The primary difference between the Speech and Call techniques is the amplitude of the second partial, which differ by about 16 dB in their most trained subject.. The implication is that the subject tunes or matches the first formant with a source partial more for the Call than for the Speech techniques. This experiment suggests that formant tuning is at least a part of the training of some actors, if not an intent of quality enhancement during performance.

Sopranos utilize F1 formant tuning, and so may actors. Do tenors use formant tuning for high pitch singing? Data from Titze and colleagues [5] suggest that at high pitches, tenors will raise F1, but not tune it to a voicing partial. For some vowels, however, they may tune F2 to the second or third partial. For example, for the vowel /i/, F1 increased and F2 decreased when the spoken formants were compared to the sung formants. F1 did not match with any partial, and the F1 movement even appeared to avoid the partials by going midway between them for /i, e, a, u/. But there was a movement to match F2 with the second partial for /u/ and with the third partial for /i/ and /e/. It is curious that F2 for /a/ and /o/ moved midway between partials 2 and 3. These results may suggest that tenors attempt F2 tuning for high vowels. Since F2 is strongly related to vowel identification, tenors may attempt to preserve vowel identity as pitch rises, as well as a characteristic quality that would be destroyed by close formant tuning to F1, the strongest vocal tract formant. Spectra for male classical singers often are seen to be flat with approximately equal intensity in F1, F2 and the F3 region.

Carlsson and Sundberg [6] studied an extreme form of formant tuning. They played synthesized chromatic scales to conservatory students where the scales included F1 and F2 formant matches to partials for each note. These scales with formant-partial matches were compared to a scale with constant F1 and F2 values. The scales with constant F1 and F2 values by far were preferred. This study suggests that formant tuning in general is not a preferred device. Any gain in overall intensity would be overwhelmed by undesirable timbre discontinuities.

Singer's Formant

Formant tuning concerns F1 and F2 primarily. The singer's formant concerns a significant boost in the F3 and F4 frequency region between about 2500 and 3500 Hz. As shown by Sundberg [3], the Singer's Formant for a male classical singer can have a spectral value of approximately 15 dB above the corresponding orchestra spectral value. The creation of the Singer's Formant can arise from the movement of F4 closer to F3, thereby boosting the vocal tract spectrum in the region, relative to F1 and F2 amplitudes.

We studied the spectra of voice modification of a mezzo-soprano [7]. She sang a scale without changing her vocal tract shape to preserve the vowel integrity, and then with changing her vocal tract shape to preserve the desired classical voice quality and intensity. In the unmodified version at the highest pitch of the scale, 740 Hz, the first and second partial of the voice for her intended /^/ vowel were strong and very close to the expected speaking formant values. There was a de-emphasis of energy between 2,000 and 4,000 Hz, and a prominence between 4,000 and 6,000. In contrast, in the modified production, the first partial or the fundamental frequency was relatively strong, and was accompanied by a relatively broad spectral area with an intense peak at approximately 3,400 Hz. There was a significant de-emphasis of the second formant. Thus, the modified production included both the strong representation of the fundamental and the Singer's Formant, and that was about all.

Do actors use the Singer's Formant, that is, an enhancement of the third formant region? It would seem evident from listening to the stentorian sounds of some actors in their loud voicings. We asked whether students in our National Theatre Conservatory increased the presence of their third formant region if their vocal quality were rated higher [8]. In the preliminary study, 32 students were recorded twice within an academic year producing 10 /a/'s in their best stage voice. Their voice qualities were rated on a 1-10 scale with 3 being "normal" and 10 being "exceptionally good". Mutual agreement was reached by both a voice coach and a voice scientist. One male and one female were randomly chosen from each rating category from 3 to 9. The results indicated a general trend for F3 to increase as the voice quality rating increases. The men produced an F1-F3 intensity difference of about 20 dB for the low quality ratings, but about 10 dB or less for the higher ratings. The relation was stronger for the males. An interesting finding was that, for the men across the 10 /a/ tokens, the standard deviation of the F3 level values decreased strongly as the voice quality rose. These results suggest that the presence of F3 for male actors may tend to increase with better voice quality, and is more consistent from a technique point of view.

What is the source of the Singer's Formant? This is a highly significant question because its implied association with the vocal tract, carrying power of the voice, and often a greater clarity and quality acceptance of the voice, suggests that it provides desirable communication enhancement and highly efficient production without added laryngeal and respiratory efforts.

Sundberg proposed [3,9] that larynx lowering to widen the hypopharynx, and laryngeal ventricle widening,

might relate to the enhancement of the Singer's Formant on the basis of a laryngeal acoustic tube orientation. The primary logic was the acoustic isolation of the larynx tube so that its own resonance could ring through, regardless of the vowel. If the larynx tube is modeled as a cylinder with a length of 2 to 3 cm, the approximate length of the larynx tube, the independent resonance is in the Singer's Formant region. The opening of the larynx tube was prescribed to be smaller than the laryngopharynx by approximately a factor of 6. There have been a number of video examples of the anterior-posterior motion of the epiglottis and arytenoid cartilages to create the relatively smaller larynx tube exit (e.g., Yanagisawa et al. [10]).

Wilson's 1972 dissertation [11] at Indiana University studied laryngeal and hypopharyngeal airways in singers using radiography. Different vowels, pitches and intensities were sung. The suggestion by Yanagisawa et al. that there is more constriction for louder sounds was generally supported. This subject showed reduced A-P distance for the loud productions, in general. These characteristics did not hold, however, for some subjects.

Further consideration of the generality of the early criteria suggestions by Sundberg has come from a study by Detweiler [12]. She studied a tenor and 2 baritones using magnetic resonance imaging and video laryngoscopy. The subjects sang in vocal fry, wherein the ventricular folds touched the true folds, obliterating the ventricle, and found that the Singer's Formant was as strong as in their normal classical singing. She also tested the 6:1 area ratio concept and found that her singers produced about a 3:1 area ratio while producing a strong Singer's Formant. We can conclude that further refinement of our knowledge of the origin of the Singer's Formant is needed.

The acoustic isolation by larynx tube manipulation may refer to a highly refined control of three specific regions within a few millimeters of each other - the vocal folds and intrinsic muscles, the aryepiglottic structures, and the hypopharynx.

Vocal Classification

Classifying the singing voice into bass, baritone, and so forth, as well as creating useful subclassifications, is important for practical performance literature purposes as well as for vocal health purposes. Cleveland [13] found that the voice classifications of 8 professional singers were highly related to the center of each subject's vocal range, thus relating voice classification for men to the physiologic properties of the vocal folds - the higher the placement of the singing range, the higher the voice classification. Cleveland also showed that the center of the pitch range corresponded to the average of the first four formant frequency values, thus indicating a relation between laryngeal pitch properties and vocal tract resonance properties. In addition, he showed that the voice classifications were highly correlated with the value of the specific formant frequencies, especially F3 and F4.

Dmitriev and Kiselev [14] suggested that the specific classification of voice from bass to high soprano correspond to two aspects: the frequency of the high prominent formant area, which we will consider to be the Singer's Formant, as well as to the length of the vocal tract. This suggests the first acoustic rule discussed - the longer the vocal tract, the lower the formants, and here then, the lower the voice classification.

Berdtsson and Sundberg [15] performed an interesting classification study using singing synthesis. A panel of 6 singing teachers heard descending triads. There were two important results. First, as the Singer's Formant frequency rose, the teachers rated the voice with a higher voice classification, even when F1 and F2 were held constant. And secondly, if the triad was raised in its octave but without changing the Singer's Formant, the teachers tended also to raise the voice classification.

Thus there is evidence to support the notions that voice classification into primary categories depends upon pitch range, average formant values, and even the frequency of the Singer's Formant alone.

Vocal Covering

Lastly we will mention the most recent study on vocal covering, a topic with a long history. Hertegard et al. [16] studied eleven professional male singers (bass, baritone and tenor) who sang both normal open and covered productions on the vowel /ae/ on a passagio pitch. The first formant was lowered consistently in the covered sound compared to the open sound. Also, for the limited available data, the second formant was higher in covered than in open singing. The effect for both was 100 Hz or more. The interesting associated findings with these formant results were that there were tendencies to raise the soft palate, widen the pharynx, widen the laryngeal ventricles, and increase the transglottal peak-to-peak flow in the covered compared to open singing. Of significance is the anecdotal statement by the singers that this manipulation was necessary so as not to strain the voice above the passagio.

We can now add additional "rules" and observations regarding singing and vocal tract formants. (6) F3 is dependent on the size of the cavity immediately behind the incisors (F3 is lower when the cavity increases). Please refer to [3]. (7) F4 may be strongly dependent on the shape of the laryngeal tube. (8) Narrowing the frequency distance between F3 and F4 boosts the Singer's Formant amplitude. (9) Pitch-dependent formant tuning above 500 Hz for females preserves intensity and classical performance quality. (10) Pitch-dependent formant tuning may aid the intensity and quality (or the training) of the actor's voice. (11) Formant

tuning of F1 and F2 in singing is not a general goal in Western classical training and performance, partly because of quality preference, although this needs further exploration, especially regarding tuning the second formant. (12) The Singer's Formant can be present in both male and female singers and actors, but is less consistently apparent in sopranos. (13) The Singer's Formant may be strongly associated with larynx tube configuration, but further modelling is necessary. (14) Voice classification (bass to soprano) depends upon the person's pitch range and formant frequency values, and can be estimated based on the Singer's Formant frequency itself, and (15) Covered singing appears to involve a lowered F1 and potentially a raised F2, and in general a more elevated soft palate, a wider pharynx and laryngeal ventricles, and an increased peak-to-peak AC glottal flow.

If vocal efficiency of source and vocal tract filter is to be well established, the study of trained performers is obligatory. The selected studies reviewed here (ref. also, e.g., [17]) help us to better conceptualize potential efficiency enhancement for the non-performer through observation of resonance behaviors in the trained vocal performer.

Acknowledgement

Gratitude is extended to Chwen-Geng Guo, Christopher Linker and Paula Saraceno for their help in preparing this manuscript and the manuscript for the corresponding talk at the 1st World Voice Congress in Oporto, Portugal. Copies of that talk, with figures, can be obtained by writing the author at 1245 Champa St., Denver, CO, USA, 80204. This paper and its presentation were supported by Grant P60 DC00976 from the National Institutes of Health.

References

1. Titze IR. Prinicples of Voice Production. Englewood Cliffs:Prentice Hall, 1994

2. Sundberg J. The acoustics of the singing voice. Scientific Am 1977;March:82-91.

3. Sundberg J. The Science of the Singing Voice. Dekalb,IL:The Northern Illinois University Press, 1987.

4. Raphael BN, Scherer RC. Voice modifications of stage actors: Acoustic analyses. J Voice 1987;1:83-87.

5. Titze IR, Mapes S, Story B. Acoustics of the tenor high voice. J Acoust Soc Am 1994;95:1133-1142.

6. Carlsson G, Sundberg J. Formant frequency tuning in singing. J Voice 1992;6:256-250.

7. Scherer RC, Church T, Root RS. Contrasting speech and singing - performance demonstration. J Acoust Soc of Am 1993;94(3) Pt. 2:1871J(A).

8. Scherer RC, Maes K, Horii Y. Voice quality of the actor in training: acoustic correlates (in preparation).

9. Sundberg J. Articulatory interpretation of the "singing formant". J Acoust Soc of Am 1974;55:838-844.

10. Yanagisawa E, Estill J, Kmucha S. Contribution of aryepiglottic constriction to "ringing" voice quality. Video tape, Yale Unviersity School of Medicine, New Haven.

11. Wilson JG. Variations of the laryngo-pharynx in singing. NATS Bull 1976;Dec:22-24; based on the 1972 dissertation from Indiana University School of Music.

12. Detweiler RF. An investigation of the larngeal system as the resonance source of the singer's formant. J Voice 1994;8:303-313.

13. Cleveland TF. Acoustic properties of voice timbre types and their influence on voice classification. J Acout Soc Am 1977; 61:1622-1629.

14. Dmitriev L, Kiselev A. Relationship between the formant structure of different types of singing voices and the dimension of the superglottal cavaties. Folia Phoneatrica 1979;31:238-241

15. Berndtsson G, Sundberg J. Perceptual significance of the center frequency of singer's formant. STL-QPSR 1994;4:95-105.

16. Hertegard S, Gauffin J, Sundberg J. Open and covered singing as studies by means of fiberoptics, inverse filtering, and spectral analysis. J Voice 1990 4: 220-230.

17. Bloothooft G, Plomp R. The sound level of the singer's formant in professional singing. J Acoust Soc Am 1986;79:2028-2033.

NCVS Status and Progress Report - 9
April 1996, 31-40

The Normal Modes of Vocal Fold Tissues

David Berry, Ph.D.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

The Ritz method is used to calculate the eigenmodes and eigenfrequencies of a 3D vocal fold-like continuum. The investigation represents a rectification and extension of previous studies, emphasizing the indispensability of utilizing natural boundary conditions when computing the characteristic modes of a system. Concurring with previous assertions, two of the eigenmodes are theorized to play a major role in facilitating self-oscillation of the folds. For the case of incompressible tissues, the investigation suggests that the eigenfrequencies associated with these two eigenmodes are nearly identical across the entire parameter space considered in the study. If truly indicative of mechanisms governing vocal fold dynamics, it may help explain why these two modes entrain so easily for conditions of self-oscillation.

Introduction

Much of the theoretical groundwork for treating vocal fold vibrations as viscoelastic waves in a continuum can be found in Titze and Strong (1975) and Titze (1976). Both of these papers are based on small-amplitude vibrations where linearization is assumed to be valid. The assumption of linearity allows natural frequencies of oscillation and normal modes to be calculated. Although there are many nonlinearities associated with vocal fold dynamics (Titze, Baken, and Herzel, 1993), variations of the lowest-order normal modes describe some of the most commonly observed vibration patterns of the folds (Moore and Von Leden, 1958). Similarly, despite explicit nonlinearities that have been incorporated into both low and high-dimensional models, simulations of vocal fold movement suggest that typical vibration patterns may be approximated as superpositions of a few of the low-order eigenmodes (see Titze, 1976; Berry, Herzel, Titze, and Krischer, 1994). Furthermore, the correspondence between the phonation frequency and the natural frequencies of oscillation has been noted in both experimental and theoretical studies (Titze, 1976; Kaneko, Masuda, Shimada, Suzuki, Hayasaki & Komatsu, 1986; Ishizaka, 1988).

Given the assumption of completely compressible tissue, analytic solutions for the eigenmodes and eigenfrequencies of a vocal fold-like continuum were presented in Titze and Strong (1975). However, because free boundary conditions were not utilized in determining the solutions, they did not correspond to the true eigenmodes and eigenfrequencies of the system. Specifically, because tangential stresses were not required to vanish on the free surfaces, the modes and frequencies presented were not necessarily the characteristic ones of the system, i.e., they may have been influenced by the external stresses.

Although the mechanics of vocal fold vibration for a 3D continuum were extended in Titze (1976) to incorporate the incompressible and viscous nature of the folds, calculation of eigenfrequencies and eigenmodes was not attempted in this more general development. Instead, glottal impedances were investigated for conditions of a laterally-driven glottis. Although this condition may approximate intraglottal driving pressures during self-oscillation, it does not exploit the normal mode concept, which suggests that each linear system has its own characteristic vibration patterns and frequencies, independent of externally applied forces.

Thus, while for several decades the eigenmode concept has shown promise for interpreting various aspects of vocal fold dynamics, the specific consequences of eigentheory analysis for continuum models of vocal fold vibration remain relatively unexplored. The purpose of the present investigation is to help remedy this situation.

The Equations of Motion

As in former treatments (Titze and Strong, 1975; Titze, 1976), several assumptions are made in the derivation and solution of the equations of motion for the 3D continuum. First of all, vocal fold tissues are assumed to be elastic with transverse isotropy about the y-axis, which is positioned along the anterior-posterior length of the folds. Because tissues are generally stiffer in the direction of muscle fibers, transverse isotropy is probably the simplest realistic assumption that can be made for vocal fold tissues. To a first approximation, the direction of the muscle fibers is assumed to lie along the y-axis. Isotropy, then, exists only in the x-z plane, i.e., the plane transverse to the muscle fibers.

The folds are also assumed to take the shape of a rectangular parallelepiped, which is presumed to be a reasonable first approximation. In general, simple geometrical shapes are required in order to obtain analytic solutions for normal modes. Fixed boundary conditions are imposed at the lateral, anterior and posterior surfaces, while free boundary conditions are dictated at the medial, superior, and inferior surfaces (see Fig. 1). For small oscillations about this equilibrium configuration, two additional assumptions are made: (1) the system is assumed to be linear, and (2) all y-displacements are assumed to be negligible, or zero. This latter assumption is based on observations of trajectories of vocal fold fleshpoints during self-oscillation (Baer, 1981; Saito et. al, 1985).

For infinitesimal displacements in an elastic continuum, Newton's second law of motion may be expressed as:

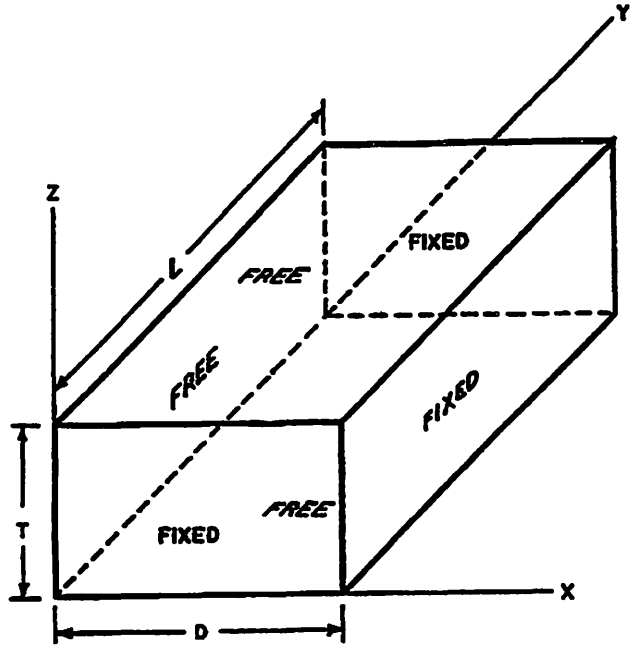


Figure 1. The rectangular parallelepiped geometry and associated boundary conditions used to approximate a vocal fold-like continuum. (After Titze and Strong, 1975).

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{xz}}{\partial z} + \bar{X} = \rho \frac{\partial^2 \xi}{\partial t^2}, \qquad \text{(1a)}$$

$$\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{yz}}{\partial z} + \bar{Y} = \rho \frac{\partial^2 \psi}{\partial t^2}, \qquad \text{(1b)}$$

$$\frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \sigma_z}{\partial z} + \bar{Z} = \rho \frac{\partial^2 \zeta}{\partial t^2}. \qquad \text{(1c)}$$

where ξ, ψ, and ζ are displacements in the x, y, and z directions; the σ_i's are normal stresses to the exterior of the infinitesimal volume element; the τ_{ij}'s are the corresponding tangential stresses; and X, Y, and Z are the body forces, i.e., forces which act over the entire volume such as gravity or inertia. In this development, Eq. 1b is not relevant because of the constraint on y-displacements for small vibrations. Consequently, the general tissue displacement vector for small oscillations about the equilibrium position may be written as:

$$\vec{\Psi}(x,y,z,t) = \xi(x,y,z,t)\,\mathbf{i} + \underbrace{\psi(x,y,z,t)}_{=0}\,\mathbf{j} + \zeta(x,y,z,t)\,\mathbf{k} \qquad \text{(2)}$$

Note that Eq. 1 is a mixed equation, i.e., the equation contains both stresses and displacements. In the solution of these equations, one usually expresses the equations in terms of stresses only, or displacements only. In this investigation, it is more appropriate to work with displacements, since eigenmode shapes are expressed in terms of displacements. Stresses may be removed from the equations of motion using Hooke's law:

$$\begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{xz} \end{pmatrix} = C\,\epsilon, \quad \textit{where } \epsilon = \begin{pmatrix} \partial\xi/\partial x \\ \partial\psi/\partial y \\ \partial\zeta/\partial z \\ \partial\psi/\partial x + \partial\xi/\partial y \\ \partial\zeta/\partial y + \partial\psi/\partial z \\ \partial\xi/\partial z + \partial\zeta/\partial x \end{pmatrix} \qquad \text{(3)}$$

where C is the stiffness matrix and ε is the strain vector, as defined above. In general, C is a 6 x 6 symmetric matrix with 21 independent constants. For the case of transverse isotropy about the y-axis, the inverse of C may be expressed in terms of just 5 independent mechanical constants (as adapted from Lekhnitskii, 1963):

$$C^{-1} = \begin{pmatrix} \frac{1}{E} & -\frac{\nu'}{E'} & -\frac{\nu}{E} & 0 & 0 & 0 \\ -\frac{\nu'}{E'} & \frac{1}{E'} & -\frac{\nu'}{E'} & 0 & 0 & 0 \\ -\frac{\nu}{E} & -\frac{\nu'}{E'} & \frac{1}{E} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu' & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu' & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix} \qquad \text{(4)}$$

where E is the Young's modulus in the transverse plane, E' is the Young's modulus along the y-axis, μ' is the shear modulus along the y-axis, ν is the Poisson's ratio in the transverse plane, and ν' is the Poisson's ratio along the y-axis. Although μ is another common engineering constant (the shear modulus in the transverse plane), it is not independent of the other constants, but may be expressed as:

$$\mu = \frac{E}{2(1+\nu)} \tag{5}$$

The corresponding stiffness matrix, C, is:

$$C = \begin{pmatrix} \frac{(-E'+\nu'^2 E)E}{(1+\nu)k} & -\frac{EE'\nu'}{k} & -\frac{(\nu'^2 E+\nu E')E}{(1+\nu)k} & 0 & 0 & 0 \\ -\frac{EE'\nu'}{k} & \frac{E'^2(\nu-1)}{k} & -\frac{EE'\nu'}{k} & 0 & 0 & 0 \\ -\frac{(\nu'^2 E+\nu E')E}{(1+\nu)k} & -\frac{EE'\nu'}{k} & \frac{(-E'+\nu'^2 E)E}{(1+\nu)k} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mu'} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\mu'} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\mu} \end{pmatrix} \tag{6}$$

$$\text{where } k = -E'(1-\nu) + 2\nu'^2 E$$

Values of the Mechanical Constants

A brief rationale for the mechanical constants used in study will now be given. First, consider the role that the Poisson's ratio ν' plays in scaling the longitudinal strain, $\partial\psi/\partial y$, induced by the lateral strain, $\partial\xi/\partial x$:

$$\frac{\partial\psi}{\partial y} = -\nu' \frac{\partial\xi}{\partial x} \tag{7}$$

Because of the constraint on y-displacements throughout the entire parallelepiped (for small oscillations about the equilibrium position), ψ and all of its spatial derivatives must be zero, including the longitudinal strain in Eq. 7. Because the longitudinal strain remains zero for any arbitrary lateral strain, the constant ν' must be identically zero. This result considerably simplifies the stiffness matrix:

$$C = \begin{pmatrix} \frac{E}{(1+\nu)(1-\nu)} & 0 & \frac{\nu E}{(1+\nu)(1-\nu)} & 0 & 0 & 0 \\ 0 & E' & 0 & 0 & 0 & 0 \\ \frac{\nu E}{(1+\nu)(1-\nu)} & 0 & \frac{E}{(1+\nu)(1-\nu)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mu'} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\mu'} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\mu} \end{pmatrix} \tag{8}$$

Analogous to Eq. 7, the function of the Poisson's ratio, ν, is illustrated in the following expression, showing the resultant vertical strain $\partial\zeta/\partial z$ induced by the lateral strain, $\partial\xi/\partial x$:

$$\frac{\partial\zeta}{\partial z} = -\nu \frac{\partial\xi}{\partial x} \tag{9}$$

In the case of a completely compressible material (ν equals zero), Eq. 9 predicts that a given lateral strain will induce no vertical strain. On the other hand in the case of a completely incompressible material (ν equals one), a given lateral strain must induce an equal and opposite vertical strain (e.g., lateral compression must result in vertical elongation). This can be substantiated with the well-known kinematic expression for incompressibility, which states that the divergence of the displacement vector equals zero:

$$\nabla \cdot \vec{\Psi} = \frac{\partial\xi}{\partial x} + \frac{\partial\zeta}{\partial z} = 0 \tag{10}$$

Through comparison of Eqs. 9 and 10, one sees that ν must equal one for the case of incompressibility. Although many texts on elastic theory state that the upper bound on the Poisson's ratio is 0.5, such a bound is only relevant for isotropic materials. As demonstrated by Lempriere (1968), the upper bound on ν for transversely isotropic materials is given by:

$$\nu \leq 1 - 2\nu'^2 E/E' \tag{11}$$

Because ν' equals zero in this problem, the upper bound on ν is one, which corresponds to the condition of incompressibility.

A series of experiments have been conducted to quantify E' for the body, cover, and ligament of the folds (Alipour and Titze, 1991, Min et. al, 1995). Unfortunately, the results of those investigations do not directly impact the present study, because E' does not enter the equations of motion for the condition of no y-displacements. However, a useful observation that may still be gleaned from these studies is that for low strains, the values of E' for the body, cover, and ligament are all on the order of 10^5 dynes/cm^2. In the absence of measurements for mechanical constants in the transverse plane, we make the assumption that for low-strains, both E and μ' equal approximately 10^5 dynes/cm^2 throughout all the layers of vocal fold tissues. Because of the uncertainty surrounding the values of E and μ', one of the final topics investigated in this paper will be the influence of these parameters on the resultant eigenfrequencies. This will be accomplished through an examination of iso-frequency contours generated in the E-μ' plane.

Solution of the Equations of Motion

By employing Hooke's law (Eq. 3) and the stiffness matrix given in Eq. 8, the equations of motion may be expressed as:

$$\Box_2\xi + \frac{1+\nu}{1-\nu}\frac{\partial}{\partial x}\left(\frac{\partial\xi}{\partial x} + \frac{\partial\zeta}{\partial z}\right) = 0, \tag{12a}$$

$$\Box_2\zeta + \frac{1+\nu}{1-\nu}\frac{\partial}{\partial z}\left(\frac{\partial\xi}{\partial x} + \frac{\partial\zeta}{\partial z}\right) = 0, \tag{12b}$$

$$\text{where } \Box_i = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \frac{\mu'}{\rho c_i^2}\frac{\partial^2}{\partial y^2} - \frac{1}{c_i^2}\frac{\partial^2}{\partial t^2}, \quad i = 1,2 \tag{12c}$$

$$c_1 = \sqrt{\frac{1}{1-\nu^2}\frac{E}{\rho}}, \qquad c_2 = \sqrt{\frac{1}{2(1+\nu)}\frac{E}{\rho}}. \tag{12d}$$

Following a similar development given by Teodorescu (1972), an effective method for de-coupling the lateral and vertical displacements, ξ and ζ, may be realized by expressing Eq. 12a in the following form:

$$D_1\xi = D_2\zeta \tag{13}$$

where D_1 and D_2 are differential operators. Assuming that the differential operators are interchangeable, ξ and ζ can be expressed in terms of a common displacement function, F:

$$\xi = D_2 F, \qquad \zeta = D_1 F \tag{14}$$

Substitution of Eq. 14 into Eq. 12b yields a fourth-order, biharmonic wave equation:

$$\Box_1\Box_2 F = 0 \tag{15}$$

Without loss of generality, the function F may be expressed in terms of two components satisfying distinct second-order wave equations:

$$\Box_1 F_1 = 0, \qquad \Box_2 F_2 = 0, \qquad \textit{where} \quad F = F_1 + F_2. \tag{16}$$

where ξ and ζ may be obtained from the displacement functions using the following relations:

$$\xi = \xi_1 + \xi_2, \tag{17a}$$

$$\zeta = \zeta_1 + \zeta_2, \tag{17b}$$

$$\xi_1 = -\frac{1+\nu}{1-\nu}\frac{\partial^2 F_1}{\partial x \partial z}, \tag{17c}$$

$$\xi_2 = -\frac{1+\nu}{1-\nu}\frac{\partial^2 F_2}{\partial x \partial z}, \tag{17d}$$

$$\zeta_1 = -\frac{1+\nu}{1-\nu}\frac{\partial^2 F_1}{\partial z^2}, \tag{17e}$$

$$\zeta_2 = \frac{1+\nu}{1-\nu}\frac{\partial^2 F_2}{\partial x^2}. \tag{17f}$$

At this point, it is possible to assign a physical interpretation to the two distinct components of the eigenmode solutions. For example, using Eq. 17, one can verify that the first component is an irrotational wave about the y-axis:

$$\left(\nabla \times \vec{\Psi}_1\right) \cdot \mathbf{j} = \frac{\partial \zeta_1}{\partial x} - \frac{\partial \xi_1}{\partial z} = 0 \tag{18}$$

Similarly, the second component is an incompressible wave:

$$\nabla \cdot \vec{\Psi}_2 = \frac{\partial \xi_2}{\partial x} + \frac{\partial \zeta_2}{\partial z} = 0 \tag{19}$$

In the limit of incompressibility, the irrotational component vanishes. However, in the general case, eigenmodes consist of both irrotational and incompressible components.

The Free Boundary Conditions

As already mentioned, fixed boundaries are imposed at the lateral, anterior, and posterior surfaces; and free boundaries at the medial, superior, and inferior surfaces. But what exactly is a free boundary? In the literature, free or natural boundary conditions correspond to those conditions which require the first variation in the potential energy to vanish (McFarland, 1972; Saada, 1974). For the second-order equations in question, this condition is obtained by requiring the normal and tangential stresses to vanish on the free surfaces (Teodorescu, 1972; Leissa and Zhang, 1983). However, because the y-displacements, ψ, have already been forced to vanish, it is not possible to simultaneously demand that the tangential stresses in the y-direction also vanish. Consequently, only the tangential stresses normal to the y-axis are required to vanish on the free surfaces.

By imposing the fixed boundary conditions at the anterior and posterior surfaces, the method of separation of variables may be used to obtain an expression for the y-dependence of ξ and ζ:

$$F_n = F_{1n} + F_{2n} = \sin\left(\frac{n\pi y}{L}\right)\left[f_1(x,z) + f_2(x,z)\right] \tag{20}$$

From the equations of motion (Eq. 16), it also appears that the method of separation of variables may be used to de-couple the x and z-dependence of ξ and ζ. However, the complex boundary conditions do not permit such a de-coupling. In fact, the boundary conditions may only be satisfied exactly through means of an infinite series solution. However, through appropriate truncation of the series, approximations may be obtained to desired degree of accuracy. For example, the method of superposition has been used to compute eigenmodes and eigenfrequencies of similar problems utilizing a Fourier series approximation (Gorman, 1982).

The Potential Energy and the Ritz Method

A simpler approach for calculating the eigenmodes and eigenfrequencies involves an implicit, rather than explicit, application of the boundary conditions. Recall that the free boundary conditions are deduced by requiring the first variation of the potential energy to vanish. Based on this principle, the Ritz method is implemented by assuming a general form of the solution. The general solution is inserted into the potential energy function, and the free parameters of the solution are determined by requiring the first variation in the potential energy to vanish. Leissa and Zhang (1983) have shown this to be an efficient method for determining the eigenfrequencies of an isotropic cantilevered rectangular parallelepiped. A cantilevered parallelepiped differs from the present problem in that it has only one fixed surface, rather than three. The isotropic assumption also represents a simplification of the present study.

Washizu (1968) has shown that the total potential energy for an elastic continuum in free vibration may be written as:

$$\Pi = I_1 - \omega^2 I_2, \tag{21a}$$

$$\text{where} \quad I_1 = \frac{1}{2}\int_{x=0}^{x=T}\int_{y=0}^{y=L}\int_{z=0}^{z=D} \epsilon^T C\, \epsilon \, dx\, dy\, dz, \tag{21b}$$

$$I_2 = \frac{1}{2}(\Psi, \rho\Psi) = \frac{\rho}{2}\int_{x=0}^{x=T}\int_{y=0}^{y=L}\int_{z=0}^{z=D} (\xi^2 + \zeta^2)\, dx\, dy\, dz. \tag{21c}$$

As in Eq. 3, ε refers to the strain vector.

Using the results from Eq. 20, a general solution for a few of the lowest-order eigenmodes may be written as:

$$\xi(x,y,z,t) = \sin\omega t \, \sin\frac{\pi y}{L} \sum_{i=1}^{I}\sum_{j=0}^{J} A_{ij} x^i z^j \tag{22a}$$

$$\zeta(x,y,z,t) = \sin\omega t \, \sin\frac{\pi y}{L} \sum_{m=1}^{M}\sum_{n=0}^{N} B_{mn} x^m z^n \tag{22b}$$

Notice that the general solution automatically incorporates all the fixed boundary conditions. In order to simultaneously satisfy the free boundary conditions, the first variation in the potential energy must vanish, as expressed in the following equations:

$$\frac{\partial \Pi}{\partial A_{ij}} = 0 \quad (i = 1,\ldots,I;\ j = 1,\ldots,J) \tag{23a}$$

$$\frac{\partial \Pi}{\partial B_{mn}} = 0 \quad (m = 1,\ldots,M;\ n = 1,\ldots,N) \tag{23b}$$

The general form of the potential energy allows the foregoing equations to be expressed in a standard matrix formulation of the eigenvalue problem:

$$\mathbf{Q}\,\mathbf{v} = \omega^2\,\mathbf{R}\,\mathbf{v} \tag{24}$$

where

$$\mathbf{v}^T = [A_{10},\ldots,A_{IJ},B_{10},\ldots,B_{MN}]^T \tag{25a}$$

$$\mathbf{Q} = \begin{pmatrix} \frac{\partial I_1}{\partial v_1^2} & \frac{\partial I_1}{\partial v_1 \partial v_2} & \frac{\partial I_1}{\partial v_1 \partial v_3} & \cdots \\ \frac{\partial I_1}{\partial v_2 \partial v_1} & \frac{\partial I_1}{\partial v_2^2} & \frac{\partial I_1}{\partial v_2 \partial v_3} & \cdots \\ \frac{\partial I_1}{\partial v_3 \partial v_1} & \frac{\partial I_1}{\partial v_3 \partial v_2} & \frac{\partial I_1}{\partial v_3^2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{25b}$$

$$\mathbf{R} = \begin{pmatrix} \frac{\partial I_2}{\partial v_1^2} & \frac{\partial I_2}{\partial v_1 \partial v_2} & \frac{\partial I_2}{\partial v_1 \partial v_3} & \cdots \\ \frac{\partial I_2}{\partial v_2 \partial v_1} & \frac{\partial I_2}{\partial v_2^2} & \frac{\partial I_2}{\partial v_2 \partial v_3} & \cdots \\ \frac{\partial I_2}{\partial v_3 \partial v_1} & \frac{\partial I_2}{\partial v_3 \partial v_2} & \frac{\partial I_2}{\partial v_3^2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{25c}$$

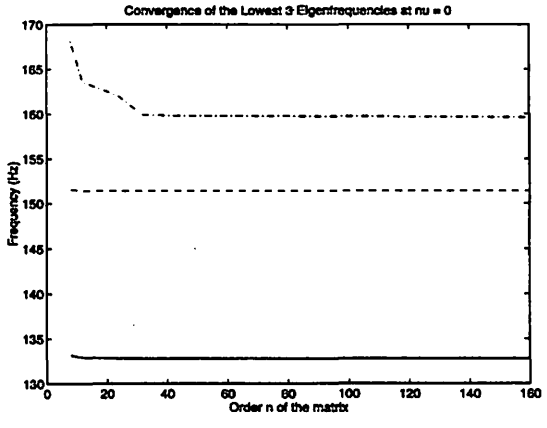
Utilizing MAPLE V, exact differentiations and integrations were performed on all terms in the equations, yielding analytic expressions for each element in the Q and R matrices. Only in the final step of the analysis were numerical values utilized to compute the eigenmodes and eigenfrequencies. Results were also verified using ABAQUS, a finite element software package.

Results and Discussion

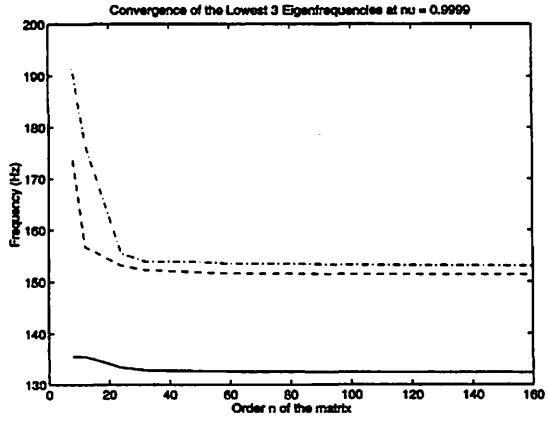
Table I shows the parameter values used in computing the eigenmodes and eigenfrequencies. In order to compare our findings with earlier studies (Titze & Strong,

Table I.
Parameter Values Used in Computing Eigenmodes and Eigenfrequencies

Lateral depth, D	1.0 cm
Longitudinal (anterior-posterior) length, L	1.2 cm
Vertical thickness, T	0.7 cm
Tissue density, ρ	1.03 g/cm³
Transverse Young's modulus, E	10^5 dynes/cm²
Longitudinal shear modulus, μ'	10^5 dynes/cm²
Transverse Poisson's ratio, ν	0/0.9999
Longitudinal Poisson's ratio, ν'	0



(a)



(b)

Figure 2.Convergence of the three lowest eigenfrequencies as a function of the order n *of the Q and R matrices for (a) completely compressible tissue, and (b) nearly incompressible tissue.*

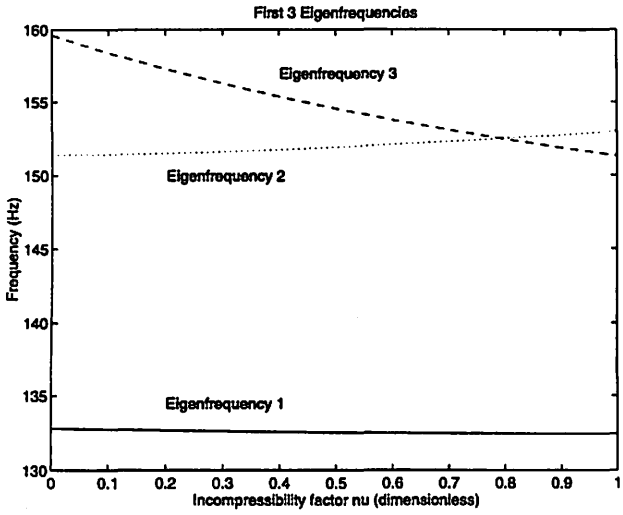


Figure 3. The three lowest eigenfrequencies of the vocal folds as a function of the incompressibility of the tissue.

1975; Titze, 1976), the two following cases are examined: (1) the case of complete compressibility ($\nu = 0$), and (2) the case of near incompressibility ($\nu = 0.9999$). Fig. 2 shows how the eigenfrequencies converge as the order n of the Q and R matrices increases. For both cases, three significant digit accuracy is obtained as the order of the matrices approaches 60. Also, in both instances, the lowest eigenfrequency is approximately 133 Hz. The second and third eigenfrequencies range between 150 and 160 Hz, and closely approximate each other for the case of near incompressibility.

Fig. 3 illustrates the dependence of the eigenfrequencies on the degree of incompressibility. While the first eigenfrequency appears relatively insensitive to variations in incompressibility, the influence of incompressibility on the next two eigenfrequencies is clear: the second eigenfrequency increases with incompressibility, while the third eigenfrequency decreases. This effect is sufficiently pronounced that the ordering of the eigenfrequencies changes for $\nu \geq 0.8$.

Fig. 4 depicts a few of the lower-order eigenmodes for the case of completely compressible tissues. Because the y-dependence of the eigenmodes is already known (see Eq. 20), the eigenmodes are illustrated as 2D coronal cross-sections in the x-z plane. While eigenmode 1 (see Fig. 4a) is similar to the z-10 mode presented in Titze and Strong (1975), this eigenmode contradicts a previous assertion that eigenmodes for completely compressible tissues are either purely lateral or purely vertical, with no x-z coupling. Eigenmode 1 clearly contains both lateral and vertical components. However, because it is predominantly a vertical mode of vibration, it probably does not play a major role in phonatory dynamics. In contrast, eigenmode 2 (see Fig. 4b) is a purely lateral mode of vibration, with no vertical

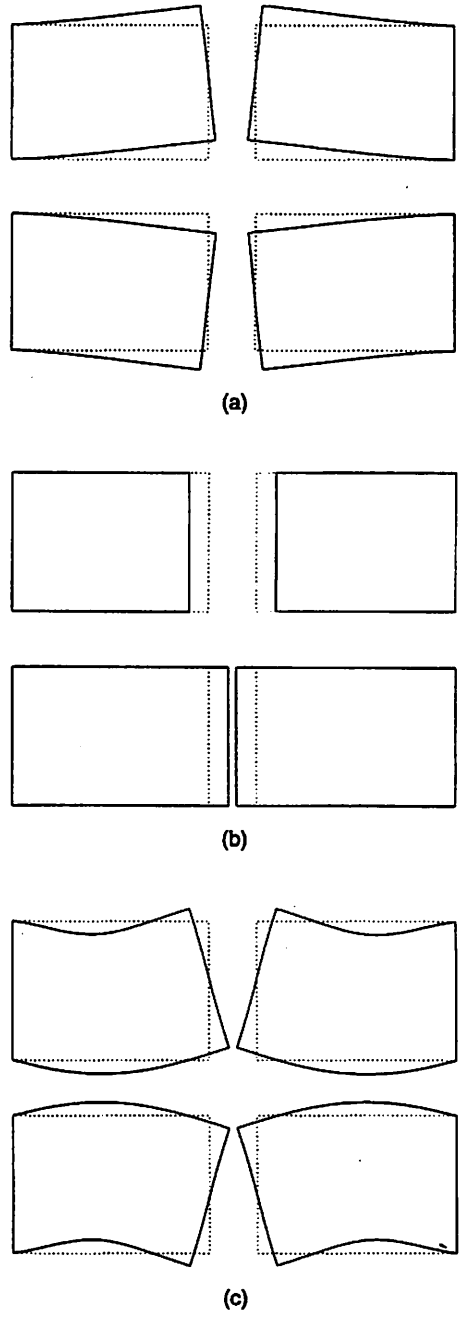


Figure 4. The three lowest-order eigenmodes for completely compressible tissue: (a) eigenmode 1, analogous to the z-10 mode; (b) eigenmode 2, identical to the x-10 mode; (c) eigenmode 3, analogous to the x-11 mode. Left and right folds shown to give indication of glottal geometry produced by corresponding mode. Upper and lower frames indicate extreme positions of eigenmodes, spaced 180° apart in a vibratory cycle. Dotted lines indicate equilibrium positions.

components. With all lateral displacements moving in-phase with each other, this eigenmode is identical to the x-10 of Titze and Strong (1975). It is the only lower-order eigenmode which can be expressed in terms of a simple trigonometric function and still meet the free boundary

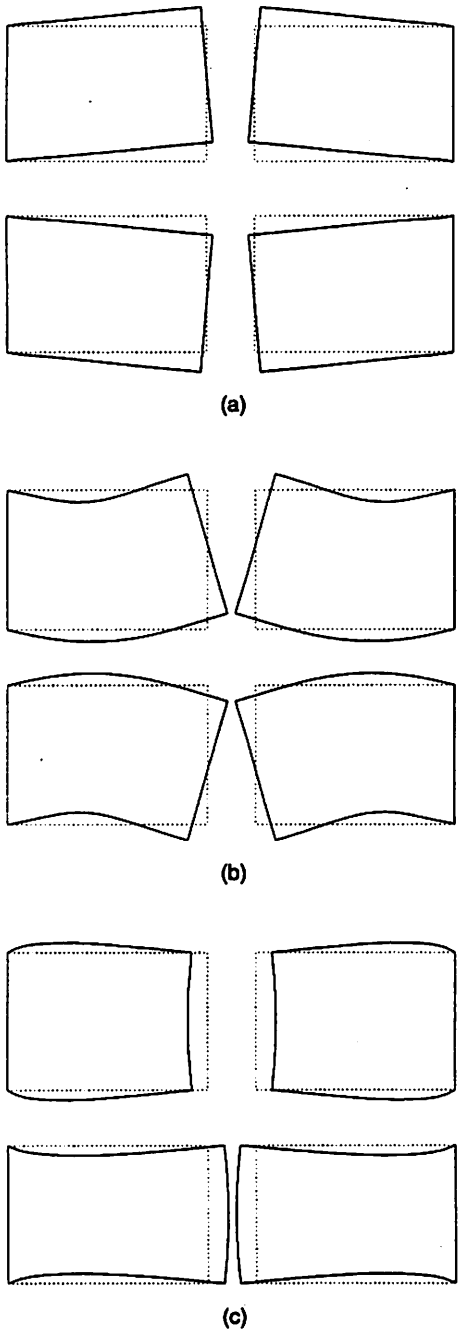


Figure 5. The three lowest-order eigenmodes for nearly incompressible tissue: (a) eigenmode 1, analogous to the z-10 mode; (b) eigenmode 2, analogous to the x-11 mode; (c) eigenmode 3, analogous to the x-10 mode.

conditions. Next, eigenmode 3 (see Fig. 4c) exhibits strong lateral and vertical components. Nevertheless, it is qualitatively similar to the x-11 mode in the role that it plays in glottal shaping. During oscillation, this eigenmode causes the glottal airway to alternate between convergent and divergent shapes, as portrayed in the upper (divergent) and lower (convergent) frames of Fig. 4c. In fact, because the medial edge of this eigenmode is characterized by a straight line, it has a simpler, more direct control over glottal convergence/divergence than the x-11 mode, which is characterized by a sine function along this medial edge.

As opposed to the case of complete compressibility, the condition of near incompressibility is believed to yield eigenmodes and eigenfrequencies which better characterize the actual vocal fold tissues. This is because vocal fold tissues are composed mostly of water, and are thus essentially incompressible. A few of these lower-order eigenmodes are shown in Fig. 5. With one possible exception, these eigenmodes have not changed remarkably from the previous case. For example, using visual inspection it is difficult to detect any difference between the eigenmodes in Figs. 4a and 5a, or in Figs. 4c and 5b. Dot products of the eigenmodes, presented in Table II, are a further indication that the eigenmodes are very similar, although not identical. Furthermore, although visual differences are apparent in the eigenmodes shown in Figs. 4b and 5c, qualitative similarities are still present. For example, along the medial edge of the folds, the tissue movement is predominantly lateral and in-phase, just as in Fig. 4b. Even numerical comparisons demonstrate the correspondence between these two eigenmodes, their dot product yielding 0.94.

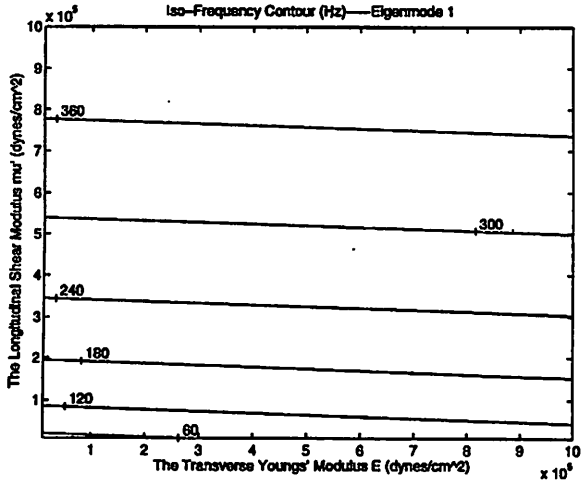
Table II.
Dot Product of Corresponding Compressible and Incompressible Eigenmodes

$\psi_1^{compr}, \psi_1^{inc}$	0.996
$\psi_2^{compr}, \psi_3^{inc}$	0.987
$\psi_3^{compr}, \psi_2^{inc}$	0.940

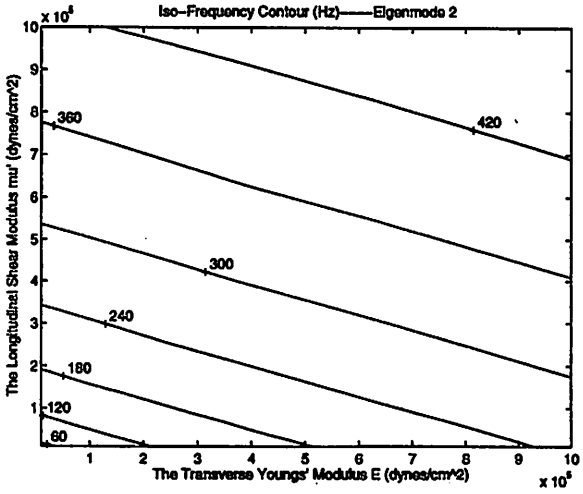
Earlier in the manuscript, it was noted that all eigenmodes may consist of both irrotational and incompressible components. Table III indicates the relative dominance of these components for each of the eigenmodes. Fig. 4b is an example of an eigenmode that is purely irrotational, i.e., it has no curvature. The corresponding eigenmode in Fig. 5c, however, is purely incompressible.

Table III.
Irrotational and Incompressible Components of Eigenmodes (in %)

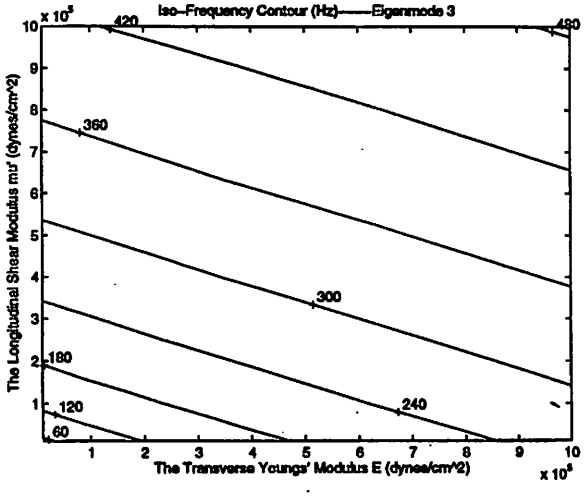
	ψ_i^{irrot})	$\psi_i^{incompr}$)
(ψ_1^{compr},	17.3%	82.7%
(ψ_2^{compr},	100.0%	0.0%
(ψ_3^{compr},	39.6%	60.4%
(ψ_1^{inc},	0.0%	100.0%
(ψ_2^{inc},	0.0%	100.0%
(ψ_3^{inc},	0.0%	100.0%



(a)



(b)



(c)

Figure 6. Iso-frequency contours in the E-μ' parameter plane associated with the eigenfrequencies for (a) eigenmode 1, (b) eigenmode 2, and (c) eigenmode 3. Computed for the condition of nearly incompressible tissue.

Note that the other two "compressible" eigenmodes in Figs. 4a and 4c are predominantly "incompressible". As substantiated by Eqs. 18 and 19, even completely compressible solutions may have both irrotational and incompressible components (however, irrotational components do vanish in the limit of incompressibility). Thus, it is not surprising that these two eigenmodes which are largely incompressible, even in the case of complete compressibility, change very little as the incompressibility condition is mandated.

As already stated, the z-10 like mode probably does not plays an important in vocal fold dynamics. However, it has long been asserted that the other two eigenmodes are essential for self-oscillation (Titze, 1988). The x-11 like eigenmode has already been shown to produce alternating divergent and convergent shapes in the glottis during oscillation (Fig. 5b). Based on physical considerations, Titze (1988) has argued that this eigenmode is roughly in-phase with the intraglottal pressure. The x-10 like eigenmode, on the other hand, yields a rough description of net lateral movement of the folds.

The basic condition for self-oscillation is that the glottal airflow must impart sufficient energy to the folds to overcome the energy losses associated with vocal fold vibration (i.e., viscous damping, collision, etc.). In order to satisfy this condition and maximize the energy imparted to the folds, the tissue velocity (controlled predominantly by the x-10 mode) must be in-phase with the intraglottal pressure (controlled predominantly by the x-11 mode). A unique phase relationship between the x-10 and x-11 like eigenmodes ensures that this condition is satisfied. This relationship has been observed and carefully investigated in finite element simulations of vocal fold movement (Berry et. al, 1994).

The two-mass model of Ishizaka and Flanagan (1972), upon linearization, also has two eigenmodes which are conceptually equivalent to the x-10 and x-11 modes. For example, the mode with both masses in-phase corresponds to the x-10 mode and the mode with both masses out-of-phase corresponds to the x-11 mode. Because of this correspondence, throughout the remainder of this discussion, we refer to the eigenmodes of the two-mass model as x-10 and x-11 modes. Much of the success of the two-mass model might be attributed to its ability to capture these two eigenmodes which facilitate self-oscillation.

It has been observed that during self-oscillation the fundamental frequency of the two-mass model corresponds closely to the eigenfrequency of the x-10 mode, which is the lowest-order eigenmode of the model. In the present investigation of eigenfrequencies in a continuum, it has been observed that the x-10 mode has a lower eigenfrequency than x-11 mode only for the case of compressible tissues. For the more realistic case of near incompressibility, it is the x-11 mode, the mode most closely related to intraglottal air pressure, which has the lower eigenfrequency. However, because these two eigenfrequencies are so closely spaced in

the continuum model, it may not be meaningful to unduly emphasize this issue.

Indeed, a major point of departure between the two-mass model and the continuum study concerns the relative spacing of the eigenfrequencies of the two dominant eigenmodes. For typical Ishizaka and Flanagan (1972) parameters, the eigenfrequencies of the x-10 and x-11 modes are approximately 120 Hz and 201 Hz, respectively. However, for the parameters used in this study, the corresponding eigenfrequencies are 153 Hz and 151 Hz for the case of near incompressibility.

Because of the relative uncertainty concerning the parameter values used for E and μ' in the continuum model, the eigenfrequencies for the 3 lowest-order eigenmodes were computed across the entire E-μ' plane for the case of near incompressibility, as shown in Fig. 6. Across the entire plane, the x-11 mode always had a lower eigenfrequency than the x-10 mode. However, the spacing of these two eigenfrequencies (Figs. 6b and 6c) never differed by more than several Hz. If the close spacing of these two dominant eigenfrequencies is truly indicative of mechanisms governing vocal fold dynamics, it may explain why these modes entrain so easily for conditions of self-oscillation.

In regards to Fig. 6, the influence of the parameters E and μ' on the eigenfrequencies also merits discussion. For low values of E, the eigenfrequencies for all three eigenmodes are nearly identical. Increasing the value of μ' boosts the eigenfrequencies of all the modes. Increasing of the value of E also boosts the eigenfrequencies of eigenmodes 2 and 3, however, it has little influence on eigenmode 1. Thus, the eigenfrequencies of the lowest-order eigenmode are mostly a function of μ', while the two higher-order modes have a strong dependence on both E and μ'. One implication of this observation is that for large values of E, a greater spacing exists between the lowest eigenfrequency and those of the two higher-order modes. As already mentioned, the two higher eigenfrequencies retain their close spacing throughout the entire E-μ' plane.

Summary and Directions for Future Research

The solution of eigenmodes and eigenfrequencies for a vocal fold-like continuum has been performed using standard analytic and numerical approximation techniques from elastic theory. In particular, the Ritz method has proven useful in calculating the eigenfunctions.

The investigation was both a rectification and extension of previous treatments. One novel outcome of this study was that the eigenfrequencies of the x-10 and x-11 like eigenmodes of nearly incompressible tissues were nearly equal (i.e., never differed by more than a few Hz) throughout the entire parameter space considered in the study. This result merits further investigation. If truly indicative of mechanisms of vocal fold dynamics, it help may explain why these modes entrain so easily for conditions of self-oscillation.

Although analytic treatments are generally only appropriate in modeling certain idealized physical systems, they provide a solid foundation from which more realistic physical models may be developed. In particular, the finite element method might be used to model many complexities of the vocal fold structure, including the complex geometry of the folds, the layered tissue structure, the nonlinear elastic properties of the tissues, aerodynamic forces, collision forces, and viscous losses. However, in the development of such complex models, it is helpful if simpler analytic models also exist, to which its solutions should converge in certain limiting cases. In the case of the vocal fold dynamics, even low-order models have benefitted from the application of eigentheory. It is hoped that this study will be just as useful in furthering the development of higher-order models of vocal fold oscillation.

Acknowledgments

This work was supported by Grant No. P60 00976 from the National Institute on Deafness and Other Communication Disorders. The ICAEN computing center at the University of Iowa generously allocated computer time and software to allow the results of this investigation to be verified with the finite element method. The authors also thank Dr. B. Kent Harrison of Brigham Young University for advice concerning the analytic solution of this problem.

References

Alipour-Haghighi, F. and Titze, I.R. (1985). Simulation of particle trajectories of vocal fold tissue during phonation, in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, edited by I.R. Titze and R.C. Scherer (Denver Center for the Performing Arts, CO).

Alipour-Haghighi, F. and Titze, I.R. (1991). Elastic models of vocal fold tissues. J. Acoust. Soc. Am. 90, 1326-1331.

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

Berry, D.A., Herzel, H., Titze, I.R., and Krischer, K. (1994). Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. J. Acoust. Soc. Am. 95, 3595-3604.

Gorman, D.J. (1982). Free vibration analysis of rectangular plates. (Elsevier, New York).

Ishizaka, K. and Flanagan, J.L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords, Bell Syst. Tech. J. 51, 1233-1267.

Ishizaka, K. (1988). Significance of Kaneko's measurement of natural frequencies of the vocal folds, in *Voice Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimara (Raven Press, Ltd., New York), pp. 181-190.

Kaneko, T., T. Masuda, A. Shimada, H. Suzuki, K. Hayasaki, K. Komatsu (1986). Resonance characteristics of the human vocal folds in vivo and in vitro by an impulse excitation, in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaki, K. Harris, (Little, Brown and Company, Boston), pp. 349-377.

Leissa, A., and Zhang, Z.D. (1983). On the three-dimensional vibrations of the cantilevered rectangular parallelepiped. J. Acoust. Soc. Am. 73, 2013-2021.

Lekhnitskii, S.G. (1963). Theory of Elasticity of an Anisotropic Body (Holden-Day, San Francisco), p. 25.

Lempriere, B.M. (1968). Poisson's Ratio in Orthotropic Materials. American Institute of Aeronautics and Astronautics Journal 6, 2226-2227.

McFarland, D., Smith, B.L., and Bernhart, W.D. (1972). Analysis of plates. (Spartan Books, New York), p. 22.

Min, Y.B., Titze, I.R., and Alipour-Haghighi, F. (1995). Stress-strain response of the human vocal ligament. Annals of Otology, Rhinology, and Laryngology 104, 563-569.

Moore, G., and Von Leden, H. (1958). Dynamic variations of the vibratory pattern in the normal larynx, Folia Phoniatr. 10, 205-238.

Saada, A.S. (1974). Elasticity: theory and applications. (Pergamon Press, Inc., New York)., p. 452.

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

Teodorescu, P.P. (1972). Dinamica corpurilor liniar elastice (Editura Academiei Republicii Socialiste România). English translation: Dynamics of linear elastic bodies, (1975, Abacus Press, England), Chapter 3.

Titze, I.R. and W. Strong (1975). Normal modes in vocal cord tissues, J. Acoust. Soc. Am. 57, 736-744.

Titze, I.R. (1976). On the mechanics of vocal-fold vibration, J. Acoust. Soc. Am. 60, 1366-1380.

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

Titze, I.R., Baken, R., and Herzel, H. (1993). Evidence of chaos in vocal fold vibration, in *Vocal Fold Physiology: Frontiers in Basic Science*, edited by I.R. Titze (Singular Publishing Group, San Diego).

Washizu, K. (1968). Variational methods in elasticity and plasticity. (Pergamon Press, New York), pp. 43-44.

NCVS Status and Progress Report - 9
April 1996, 41-46

Pressure-Flow Relationships During Phonation as a Function of Adduction

Fariborz Alipour, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Ronald Scherer, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Eileen Finnegan, M.A.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Pressure-flow relationships were obtained for five excised canine larynges. Simultaneous recordings were made of average subglottal pressure, average air flow, and the electroglottograph at various levels of adduction and vocal fold lengths. The level of adduction was controlled by positioning the arytenoid cartilages via laterally imbedded three-prong attachments and by the use of intra-arytenoid shims. Adduction was quantified by measuring the vocal process gap. Results indicated a linear pressure-flow relationship within the experimental range of phonation for each level of adduction. Differential glottal resistance increased as the vocal process gap was reduced. A model is presented for the differential resistance as a hyperbolic function of vocal process gap. The pressure-flow relationship and the model can be used in computer simulations of speech production and for clinical insight into the aerodynamic function of the human larynx.

Introduction

Glottal flow resistance, an aerodynamic aspect of phonation, is the ratio of the time-averaged transglottal pressure and glottal flow, and is dependent upon the glottal adductory mechanism and vocal fold tension (1,2). It is considered an important aerodynamic parameter in describing voice disorders. Glottal flow in human subjects can be obtained from pneumotachograph methods of transducing the oral plus nasal flow. Although direct measurement of subglottal pressure can be made invasively, oral pressure techniques to estimate subglottal pressure are often used (e.g. ref. 2-5).

Glottal flow resistance has been found to vary with gender (6-9), age and development (6,9-11), and intensity of voice production (7,8,12). Glottal flow resistance has also been a measure in the evaluation of laryngeal hyperfunction (13), spasmodic dysphonia (14,15), Parkinson disease (16), surgical procedures for arytenoid adduction (17), thyroid cartilage fractures (18), rheumatoid arthritis (19), mucosal stiffening (20), velopharyngeal incompetence (21,22), cancer (23), cochlear implants (24), and closed head injury (25).

The issue addressed in the research of this report is the relation between the transglottal pressure and the glottal flow for specific levels of glottal adduction. Does the flow increase significantly as pressure increases? Is the relationship between pressure and flow linear or nonlinear? If linear, is the slope steep or shallow? The answers determine the theoretical approach used in thinking about flow resistance. The answers also depend on what kind of laryngeal model is used, whether there is vocal fold vibration, and what range of flows and pressures are used. Static physical models (e.g., 26) and human glottal configurations with non-vibrating vocal folds (27) yield nonlinear pressure-flow relationships with significant changes in both flow and pressure. If phonation is induced using excised canine larynges, both the pressures and flows for a relatively constant level of arytenoid cartilage adduction increase significantly and can operate in a similar pressure-flow range as the human (20). The *in vivo* canine studies of Tanaka and Tanabe (28), which involved stimulation of the thyroarytenoid muscles, also showed pressures and flows that varied significantly together for approximately constant levels of adduction. These results were contradicted by a study by Berke et al. (29) in which their *in vivo* procedure with laryngeal nerve stimulation produced constant subglottal pressure as glottal flow increased at constant levels of adduction. Their methodology was used in other canine studies (30-32) in which the relation between flow and pressure varied, some cases

showing mutual increases, some with constant pressure with increasing flow. These latter studies differ from the research reported here in that they used relatively large values of subglottal pressure, typically out of the range of normal human phonation, and used relatively tight adductions of the arytenoid cartilages.

The complete aerodynamic evaluation of a larynx depends upon the interdependence of subglottal pressure, airflow, arytenoid adduction, and vocal fold length and tension, and can be described by subglottal pressure as a function of airflow and the control variables. It is the purpose of this study to investigate the pressure-flow relationships of excised canine larynges as a function of adduction, the results of which may lend insight into the role of glottal flow resistance in the assessment of human phonation.

Methodology

Excised canine larynges were used in accordance with the procedures described by Alipour and Scherer (33). Air was humidified to approximately 100% humidity and heated to 37°C using a Concha Therm III Servo Control Heater unit (RCI Laboratories) (air came from the building's compressed and filtered air supply). The larynx (with tracheal tissue) was mounted on a pseudo-tracheal rigid tube so that the glottis was easily viewed by a camera and accessible by the equipment (Fig. 1). Control of adduction was established by two methods. In one method, three-pronged needles pressed against the arytenoid cartilages. The position of the needles was adjusted via a micrometer traverse (Fig. 2). In the second method, metal shims of various thickness (0.1 - 1.0 mm) were inserted between the arytenoid cartilages posterior to the vocal processes, and the three-pronged needle system was used to keep the shims in place. To control length changes, another micrometer traverse was attached directly to the thyroid prominence above the anterior commissure.

Although the shims and micrometer traverse were used to adjust the adduction, the shim size itself was not used as the adduction level because tightness of fit altered the actual position of the vocal processes. In this paper the level of adduction was taken as the actual distance between the vocal processes (the vocal process gap, VPG).

The mean pressure 7 cm below the glottis was monitored with a wall mounted water manometer (Dwyer No. 1230-8). The mean flow rate was monitored with an in-line flowmeter (Gilmont rotameter model J197). The time-varying subglottal pressure was transduced by a piezoresistive pressure sensor (Microswitch 136PC01G1) at the same location as the manometer tap. The bandwidth for the pressure transducer was approximately 0-800 Hz.

Two electrodes from an electroglottograph (EGG) made by Synchrovoice Inc. were attached to the exterior of the thyroid cartilage with push pins. The third (reference) electrode was placed on the posterior surface of the larynx. The EGG and subglottal pressure signals were monitored on a digital oscilloscope (DATA 6000, Data Precision of Analogic Corp.). These signals and the sound of the vibrating larynx were used to judge the perceptual stability of the phonation.

During the experiment, analog signals from the EGG and the pressure transducer were recorded on DC channels of a Sony model PC-108M Digital Audio Tape (DAT) recorder. These data were digitized later with an analog/digital converter and analyzed on VAX-station computers. The multiple channels of data were digitized at 10 kHz per channel. Mean airflow measures were recorded

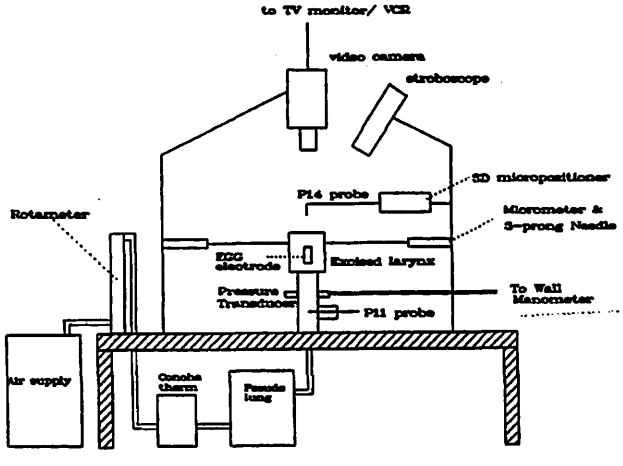


Figure 1. Schematic of the excised larynx setup.

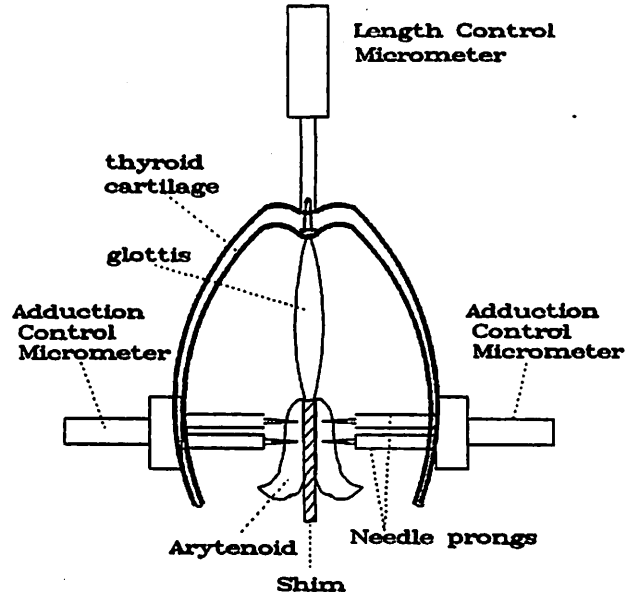


Figure 2. Schematic of the controls for glottal configuration.

during the experiment by announcing and recording the rotameter ball positions. During the experiment, the top view of the larynx and vocal folds was recorded on video tape for later analysis of the images. For stroboscopic images, a Pioneer DS-303ST stroboscope was used to manually control fundamental frequency.

Experiments were conducted for five canine larynges. For each larynx, pressure-flow experiments were carried out at various levels of adduction. In these experiments, flow rate was varied within a range of stable phonation. The frequency of oscillation was extracted from the EGG signal. The vocal process gap was measured from the image of the glottis on the monitor of the VCR in freeze mode. Prior to data collection, a piece of ruler placed on the vocal folds was video recorded to provide a means for calibrating distance measures. The resolution of distance measures was approximately ± 0.03 mm based on a 15 to 1 enlargement on the video recordings.

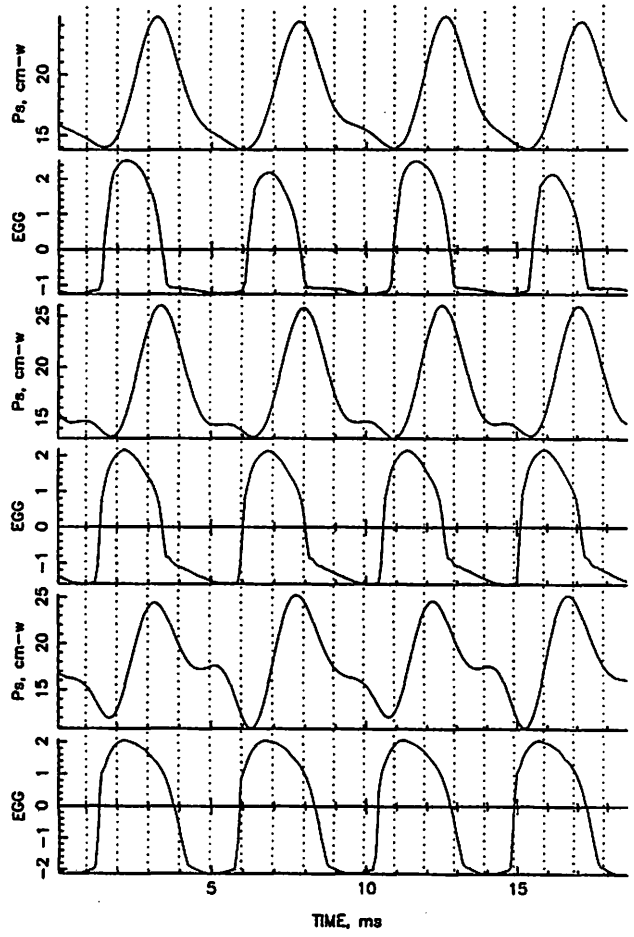


Figure 3. EGG and pressure waveform pairs for three levels of adduction, increasing from top to bottom. Excised larynx 6. As adduction increases, the EGG pulse width increases and a harmonic ripple seen in the pressure signal becomes more prominent.

Results and Discussion

Figure 3 shows typical subglottal pressure waveforms for three levels of adduction with the same mean subglottal pressure of 18 cm-H_2O. The accompanying EGG tracings provide an orientation to the timing of these pressure changes relative to vocal fold oscillation. These data were obtained from an excised larynx of a 27 kg male dog. The vocal process gap (VPG) was 2.00, 1.73, and 1.53 mm for the three waveform pairs from top to bottom. Decreased VPG indicates increased adduction. As adduction increased and mean subglottal pressure was held constant, the mean flow rate decreased (737, 619, 529 ml/s for the three levels of adduction, respectively), giving rise to increasing flow resistance. This is consistent with prior literature (e.g., 28,31). The widened pulse shape of the EGG signal reflects a decrease in the open quotient as adduction increased (34). The peak-to-peak height of the EGG signal also increased with adduction (with constant subglottal pressure), suggesting an increased amount of medial tissue contact. The fundamental frequency of phonation increased only slightly (214, 219, 223 Hz for the three levels of adduction, respectively).

Figure 4 shows typical pressure-flow findings for excised larynx phonations for two larynges. A linear relation was observed between mean subglottal pressure and mean flow rate at the various levels of adduction and for the pressure ranges used in this study. As the adduction increased (i.e., as VPG decreased), the regression lines through the data rotated counterclockwise, indicating an increased slope of the pressure-flow relation. Each slope is defined here as the differential flow resistance, R_d, for each level of adduction. This is in contrast to the simple flow resistance, R_s, the ratio of any specific pressure value and flow value.

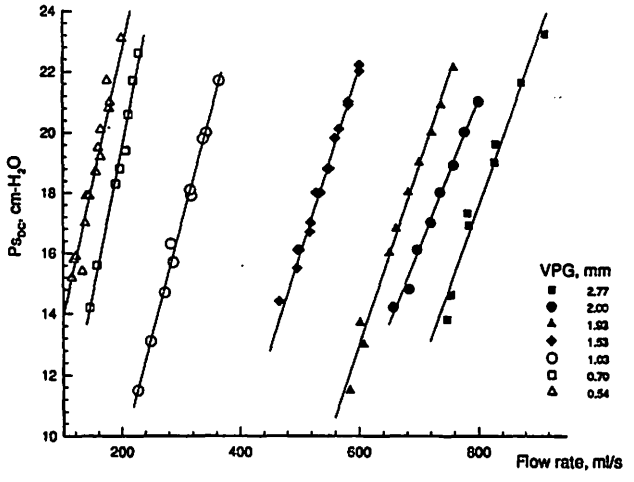


Figure 4. Typical pressure-flow as a function of level of adduction for two excised larynges. (The filled symbol data correspond to one larynx, the open symbol data to the other).

The differential flow resistance is theoretically approximately twice as much as simple flow resistance (35), a suggestion supported by these data sets for a given adduction. It is noted that the R_s values do vary for a given adduction, whereas the corresponding R_d remains constant (due to the apparent linearity). R_s values for data of a particular adduction level increase as flow increases, indicating subglottic pressure increases at a proportionately greater rate than flow. This is in contrast to the *in vivo* data of the UCLA lab (30,31). Their data show a decreasing value of R_s with flow increase. This finding may be due to the use of relatively high adduction and subglottal pressures, a region of phonation where flow may increase relatively more for less pressure rise. The pressure-flow relationship in the study by Muta and Fukuda (20) was slightly non-linear for normal cases. Since they adducted the vocal folds by suturing the vocal processes with silk thread, the level of their adduction might have decreased as subglottal pressure increased, which would create a slight non-linearity of the pressure-flow data.

The differential glottal flow resistance (R_d) was calculated using linear regressions for the data for the 5 canine excised larynges for each level of adduction. Figure 5 shows the differential glottal flow resistance in cm-H_2O/LPS as a function of VPG. The dynamic flow resistance increased as VPG decreased. The differential glottal flow resistance ranged between 27 to 109 cm-H_2O/LPS, with an average of 62 and standard deviation of 19.2. The range of simple flow resistance for these larynges was approximately three times that for R_d, between 11 to 270 cm-H_2O/LPS, with an average of 58 and standard deviation of 43.9. The differential flow resistance data points were fitted to a hyperbolic function using a parameter optimization method. The following expression was obtained:

$$R_d = 17.26 + \frac{107.7}{VPG + 0.9075} \qquad (1)$$

where R_d is in cm-H_2O/LPS. The model predicts the resistance with a correlation coefficient of 0.875. Figure 5 indicates that the pressure-flow ratio for smaller vocal process gaps should increase, making the larynx aerodynamically highly sensitive to small gap changes. Equation (1) can be used in computer simulations of speech, where one needs to estimate mean pressure-flow behavior for a particular adduction configuration. For example, an initial estimate of the flow rate from knowing the initial vocal process gap and subglottal pressure could be a good start in the calculation of laryngeal flow. In addition, equation (1) expresses a relationship between flow resistance and laryngeal adduction that may lend itself to quantitative clinical assessment potential determined in future studies.

Although the mean subglottal pressure was adjusted to remain the same in Figure 2, there was an increase in the peak-to-peak pressure (from 13 to 15.4 to 16.2 cm-H_2O) with the increase in adduction. The increased subglottal AC pressure values and more prominent harmonic structure (the second peak in the pressure cycle) suggest less damping and greater subglottal acoustic excitation with greater adduction.

Figure 6 displays typical AC pressure as a function of flow rate at various levels of adduction in two excised larynges (ref. Fig. 4). The AC pressures were obtained from the time-varying pressure waveforms by dividing their peak-to-peak values by two. Again a linear relation appears

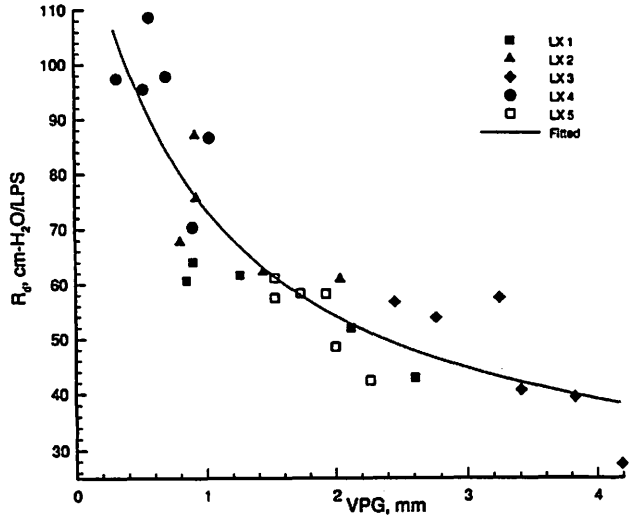


Figure 5. Differential glottal flow resistance as a function of vocal process gap for 5 excised larynges.

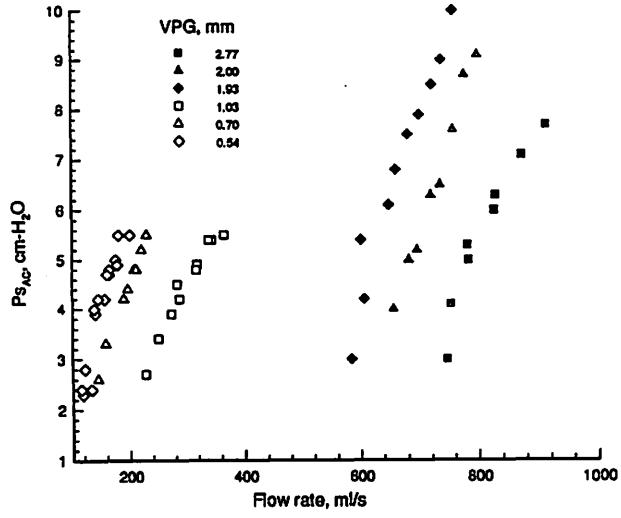


Figure 6. Subglottal AC pressure variations with mean flow rate and adduction for the two larynges of Figure 4.

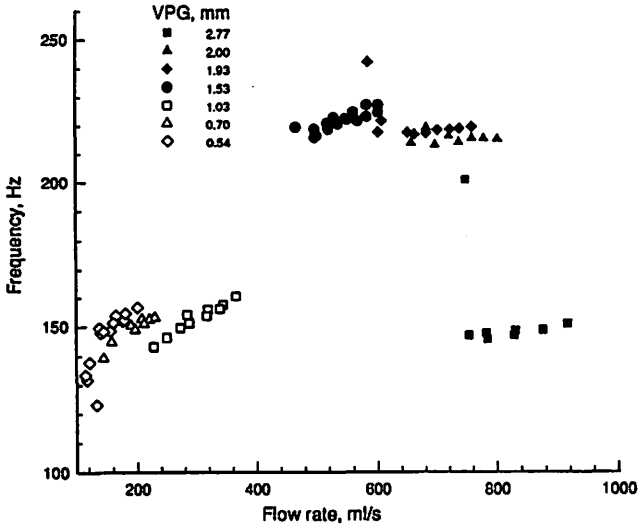


Figure 7. Effect of flow rate and adduction on fundamental frequency for the two larynges of Figure 4.

appropriate for these data. As adduction increases, higher AC pressures for the same mean flow rate can be observed. The two larynges were operated in two separate ranges of pressure and flow and are distinguished with solid and hollow symbols. The extent of the AC pressure changes may be different from those that would be obtained in the human subglottal system because of the lack of match to the human subglottal impedance in the present experiment. The data here, and the linear relation with flow at given adduction levels, can serve as hypotheses for human studies.

Figure 7 represents the variation of fundamental frequency with the flow rate at various levels of adduction for two larynges (the same as the Fig. 6). It appears that the fundamental frequency has an increasing trend with the flow rate. This may be due to the increased subglottal pressure as flow increases (36,37), and is stronger at greater values of adduction. If flow rate is held constant, increasing adduction corresponds to increased fundamental frequency. The frequency and flow rate also appear to relate in a linear fashion. This is expected from the finding that frequency of oscillation has a linear relationship with pressure (38), and pressure and flow are also linearly related.

Summary and Conclusions

Pressure-flow relationships were obtained for excised canine larynges during phonation. Adduction was used as a primary control parameter. The results of the study lead to the following conclusions:

1. For specific (non-zero) values of glottal adduction, the mean subglottal pressure and mean flow are linearly related during phonation of excised canine larynges for the subglottal pressure ranges studied (11-24 cm H_2O).
2. The slope of the pressure-flow lines (the differential flow resistance) tends to increase with level of adduction.
3. The subglottal AC pressure also is linearly related to the flow rate for given values of adduction, and increases with increased adduction.

Results of this study may provide insight into the aerodynamics of phonation related to the configurational competency of the human larynx. The empirical model relating differential flow resistance to vocal process gap may be used in computer simulations of speech where the production of aerodynamics is dependent upon the relationship among mean subglottal pressure, mean glottal airflow, and glottal adduction.

Acknowledgments

The authors would like to thank Susan Samson for her assistance in data collection. This work was supported by research grant number 5 R01 DC00831-04 from the National Institute on Deafness and Other Communication Disorders, National Institute of Health.

References

1. Isshiki N. Vocal intensity and air flow rate. *Folia Phoniatrica* 1965;17:92-104.

2. Smitheran JR, Hixon TJ. A clinical method for estimating laryngeal airway resistance during vowel production. *Journal of Speech & Hearing Disorders* 1981;46(2):138-146.

3. Lofqvist A, Carlborg B, Kitzing P. Initial validation of an indirect measure of subglottal pressure during vowels. *Journal of the Acoustical Society of America* 1982;72(2):633-635.

4. Rothenberg M. Interpolating subglottal pressure from oral pressure. *Journal of Speech & Hearing Disorders* 1982;47:219-220.

5. Hertegard S, Gauffin J. Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography, and electroglottography. *Journal of Speech & Hearing Research* 1995;38(1):85-100.

6. Hoit JD, Hixon TJ. Age and laryngeal airway resistance during vowel production in women. *Journal of Speech & Hearing Research* 1992;35(2):309-313.

7. Stathopoulos ET, Sapienza C. Respiratory and laryngeal function of women and men during vocal intensity variation. *Journal of Speech & Hearing Research* 1993;36(1):64-75.

8. Holmes LC, Leeper HA, Nicholson IR. Laryngeal airway resistance of older men and women as a function of vocal sound pressure level. *Journal of Speech & Hearing Research* 1994;37(4):789-799.

9. Netsell R, Lotz WK, Peters JE, Schulte L. Developmental patterns of laryngeal and respiratory function for speech production. *Journal of Voice* 1994;8(2):123-131.

10. Melcon MC, Hoit JD, Hixon TJ. Age and laryngeal airway resistance during vowel production. *Journal of Speech & Hearing Disorders* 1989;54(2):282-286.

11. Sodersten M, Hertgard S, Hammarberg, B. Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *Journal of Voice* 1995;9(2):182-197.

12. Leeper HA, Graves DK. Consistency of laryngeal airway resistance in adult women. *Journal of Communication Disorders* 1984;17(3):153-163.

13. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: an experimental framework and initial results. *Journal of Speech & Hearing Research* 1989;32(2):373-392.

14. Witsell DL, Weissler MC, Donovan MK, Howard JF, Martinkosky SJ. Measurement of laryngeal resistance in the evaluation of botulinum toxin injection for treatment of focal laryngeal dystonia. *Laryngoscope* 1994;104:8-11.

15. Finnegan EM, Luschei ES, Barkmeier JM, Hoffman HT. Sources of error in estimation of laryngeal airway resistance in persons with spasmodic dysphonia. *Journal of Speech & Hearing Research* in press.

16. Dromey C, Ramig LO, Johnson, AB. Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: a case study. *Journal of Speech and Hearing Research* 1995;38:751-764.

17. Neuman TR, Hengesteg A, Lepage RP, Kaufman KR, Woodson GE. Three-dimensional motion of the arytenoid adduction procedure in cadaver larynges. *Annals of Otology, Rhinology & Laryngology* 1994;103:265-270.

18. Stanley RB, Cooper DS, Florman SH. Phonatory effects of thyroid cartilage fractures. *Annals of Otology, Rhinology & Laryngology* 1987;96(5):493-496.

19. Blosser JL, DePompei R. A proactive model for treating communication disorders in children and adolescents with traumatic brain injury. *Clinics in Communication Disorders* 1992;2(2):52-65.

20. Muta H, Fukuda H. Pressure-flow relationships in the experimental phonation of excised canine larynges. In: Fujimura O, ed. *Vocal Physiology: Voice Production, Mechanisms, and Functions* Raven Press, Ltd., New York 1988;239-247.

21. Lewis JR, Andreassen ML, Leeper HA, Macrae DL, Thomas J. Vocal characteristics of children with cleft lip/palate and associated velopharyngeal incompetence. *Journal of Otolaryngology* 1993;22(2):113-117.

22. Zajac DJ. Laryngeal airway resistance in children with cleft palate and adequate velopharyngeal function. *Cleft Palate-Craniofacial Journal* 1995;32(2):138-44.

23. McGuirt WF, Blalock D, Koufman JA, Feehs RS. Voice analysis of patients with endoscopically treated early laryngeal carcinoma. *Annals of Otology, Rhinology & Laryngology* 1992;101(2 Pt 1):142-146.

24. Leeper HA, Gagne JP, Parnes LS, Vidas S. Aerodynamic assessment of the speech of adults undergoing multichannel cochlear implantation. *Annals of Otology, Rhinology & Laryngology* 1993;102(4 Pt 1):294-302.

25. Theodoros DG, Murdoch BE. Laryngeal dysfunction in dysarthric speakers following severe closed-head injury. *Brain Injury* 1994;8(8):667-684.

26. Scherer RC, Titze IR, Curtis JF. Pressure-flow relationships in two models of the larynx having rectangular glottal shapes. *Journal of the Acoustical Society of America* 1983;73:668-676.

27. Tully A, Brancatisano A, Loring SH, Engel LA. Relationship between thyroarytenoid activity and laryngeal resistance. *Journal of Applied Physiology* 1990;68(5):1988-1996.

28. Tanaka S, Tanabe M. Glottal adjustment for regulating vocal intensity. An experimental study. *Acta Oto-Laryngologica* 1986;102(3-4):315-324.

29. Berke GS, Hanson DG, Gerratt BR, Trapp TK, Macagba C, Natividad M. The effect of air flow and medial adductory compression on vocal efficiency and glottal vibration. *Otolaryngology - Head & Neck Surgery* 1990;102:212-218.

30. Smith ME, Green DC, Berke GS. Pressure-flow relationships during phonation in the canine larynx. *Journal of Voice* 1991;5(1):10-17.

31. Bielamowicz S, Berke GS, Kreiman J, Sercarz JA, Green DC, Gerratt BR. Effect of tension, stiffness, and airflow on laryngeal resistance in the in vivo canine model. *Annals of Otology, Rhinology & Laryngology* 1993;102:761-768.

32. Sercarz JA, Berke GS, Bielamowicz S, Kreiman J, Ye M, Green DC. Changes in glottal area associated with increasing airflow. *Annals of Otology, Rhinology & Laryngology* 1994;103:139-144.

33. Alipour F, Scherer RC. Pulsatile airflow during phonation: an excised larynx model. *Journal of the Acoustical Society of America* 1995;97(2):1241-1248.

34. Scherer RC, Vail VJ, Rockwell B. Examination of the laryngeal adduction measure EGGW. In: Bell-Berti F, Raphael LJ, eds. *Producing Speech: Contemporary Issues: for Katherine Saford Harris.* American Institute of Physics 1995:269-290.

35. van den Berg Jw, Zantema JT, Doornenbal PJr. On the air and the Bernoulli effect of the human larynx. *Journal of the Acoustical Society of America* 1957;29(5):626-631.

36. Titze IR. On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America* 1989;85(2):901-906.

37. Hsiao TY, Solomon NP, Luschei ES, Titze IR. Effect of subglottic pressure on fundamental frequency of the canine larynx with active muscle tensions. *Annals of Otology, Rhinology & Laryngology* 1994;103:817-821.

38. Titze IR, Durham P. Passive mechanisms influencing fundamental frequency control. In: Baer T, Sasaki C, Harris K, eds. *Laryngeal Function in Phonation and Respiration* College Hill, San Diego, 1987:304-319.

NCVS Status and Progress Report - 9
April 1996, 47-52

An Experimental Study of Pulsatile Flow in Canine Larynges

Fariborz Alipour, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Ronald Scherer, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Virendra Patel, Ph.D.
Department of Mechanical Engineering & Iowa Institute of Hydraulic Research, The University of Iowa

Abstract

Pulsatile flow in an excised canine larynx was investigated with simultaneous recordings of air velocity, subglottal pressure, volume flow rate, and the signal from an electro-glottograph (EGG) for various conditions of phonation. Canine larynges were mounted on a pseudotrachea and sustained oscillations were established and maintained with sutures attached to cartilages to mimic the function of laryngeal muscles. The pitch and amplitude of the oscillations were controlled by varying the airflow, and by adjusting glottal adduction and vocal-fold elongation. Measurements with hot-wire probes suggest that subglottal inlet flow to the larynx is pulsatile but mostly laminar, while the exiting jet is non-uniform and turbulent. In the typical ranges of flow rate, subglottal pressure, and oscillation frequencies, the Reynolds number based on the mean glottal velocity and glottal hydraulic diameter varied between 1600 to 7000, the Strouhal number based on the same parameters varied between 0.002 and 0.032, and the Womersley number ranged from 2.6 to 15.9. These results help define the conditions required for computational models of laryngeal flow.

Introduction

Human speech sounds are created as a result of the conversion of aerodynamic power to acoustic power through oscillatory tissue dynamics. The mechanism of this conversion process to air flow through the glottis is poorly understood. The glottis is a three dimensional space between the vocal folds that changes dynamically during phonation. As a dynamic valve, it creates pulsatile flow which becomes the source of sound in phonation. In addition, from a biomechanical point of view, the calculation of the driving forces on the vocal folds requires the knowledge of both the pressure distribution and the velocity pattern within the glottis (Alipour and Titze, 1988).

The pulsatile flow within the glottis has not been studied in detail except with simple one-dimensional, experimental and computational models such as collapsible-tube model vocal folds (Conrad, 1980), which was based on the theory developed by Shapiro (1977). However, unsteady flows in constricted channels are studied in other applications. For example, Thornburg et al. (1992) studied steady and unsteady laminar flows through an axisymmetric rigid cosine-shaped constriction. They found that the flow became unsteady at a Reynolds number of 2000 and a constriction area ratio of 0.75. Rosenfeld (1993) simulated pulsatile laminar flow in a two-dimensional constricted channel and validated his model with the flow-visualization results of Park (1989). The constriction ratio in Rosenfeld's simulation was no more than 0.75 and the Reynolds number was less than 140. In contrast, laryngeal flow, for which the constriction ratios exceed 0.9 and the Reynolds numbers frequently exceed 1500, is both unsteady and turbulent. Einav and Sokolov (1993) discuss pulsatile pipe flow with ranges of transitional Reynolds number and Stokes parameter similar to those of the present study (also see Nakamura, Sugiyama and Haruna, 1993; Ahmed and Giddens, 1984), although their application is more appropriate to stenoses rather than flow-induced dynamic valving as in the larynx.

There has been much experimental work on laryngeal fluid mechanics, but few studies have measured air velocities. The first attempt to quantify the glottal velocity profile appears to have been made by Berke et al. (1989). Using constant-temperature hot-wire anemometry, they

measured the velocity waveform in an in-vivo canine larynx at five locations along the midline of the glottis. Although the data were limited in scope, they demonstrated that the canine larynx is a feasible model for the study of pulsatile flow during phonation. The results can be reasonably extrapolated to human phonation due to the close similarity between the canine and human glottographic waveforms (Slavit et al., 1990).

Experimental and computational techniques for diagnostics of pulsatile flow in the larynx complement each other. Experimental investigations with excised larynx models provide realistic laryngeal flow information that is useful for validation of computational models. In particular, experiments guide establishment of the inlet and outlet flow conditions, and the range of Reynolds and Strouhal numbers that occur in typical operating conditions. The present study was undertaken to quantify the laryngeal flow below and above the vocal folds, to establish the procedures for making measurements in such a pulsatile flow, and to obtain the range of Womersley number, and the Reynolds and Strouhal numbers under typical oscillation conditions. The results provide an understanding of the flow in the glottis, and could be used to define the conditions required in computational models of laryngeal flow.

Experimental Arrangement

Fresh excised canine larynges were acquired from other laboratories at the University of Iowa Hospitals and Clinics. The larynges were kept in saline solution prior to their use in the experiments. Figure 1 shows the overall experimental arrangement. After establishing an air supply with 100% relative humidity and a temperature of about 37°C using a Concha-therm heater unit (RCI Laboratories), the larynx was mounted on the rigid vertical tube of a pseudo-lung mechanism. The glottis was easily viewed by a camera and was accessible to other equipment. Adduction (the mechanism of bringing the vocal folds together) and tension controls were established by connecting the cartilages to micrometers with sutures (see Figure 2). Vocal fold length was controlled by sutures attached to the thyroid cartilage anteriorly and the cricoid cartilages posteriorly (the latter are not shown in Figure 2). A suture was sewn to each arytenoid cartilage and passed laterally through the other to control adduction.

Electrodes from an electroglottograph (EGG) made by Synchrovoice Inc. were attached to the laminae of the thyroid cartilage with push pins to pick up the EGG signal during oscillations. The EGG signal represents the change in impedance through the neck via a 5 MHz electric field, and has been related to vocal fold contact area (Childers, 1985; Scherer et al., 1988). It can be used to identify the approximate opening and closing phases of the glottis (exemplified below). The EGG signal was also used in the data analysis as a trigger signal to process the phase-averaged waveform from the instantaneous air velocity signal.

The mean pressure in the subglottal region (measured about 10 cm below the vocal folds) was monitored with a wall mounted well-type water manometer with an accuracy of ± 10 Pa (Dwyer No. 1230-8). The mean flow rate was monitored with an in-line flowmeter (Gilmont rotameter model J197; accuracy: ± 3%) as shown in Figure 1. The time varying subglottal pressure was measured with a piezoresistive pressure sensor (Microswitch 136PC01G1). This pressure (accuracy: ± 2.3%) was measured at the same location as the manometer tap (Figure 1) for ease of calibration.

The velocity was measured with a constant-temperature hot-wire anemometer system (Dantec 56C01). A straight miniature probe (Dantec 55P11) was positioned in the subglottal tube about 12 cm below the vocal folds

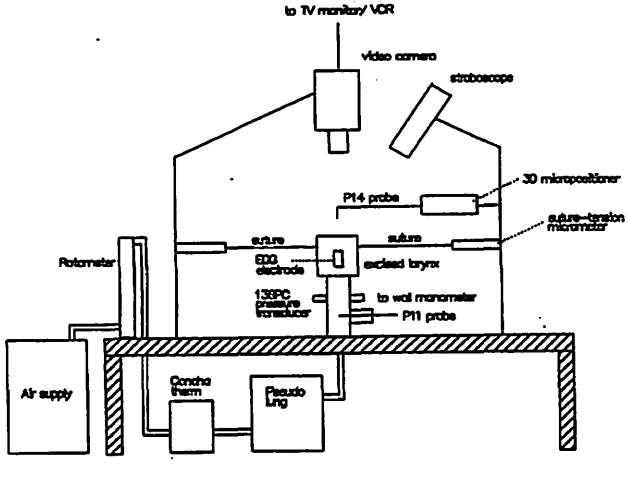


Figure 1. Experimental arrangement (not to scale).

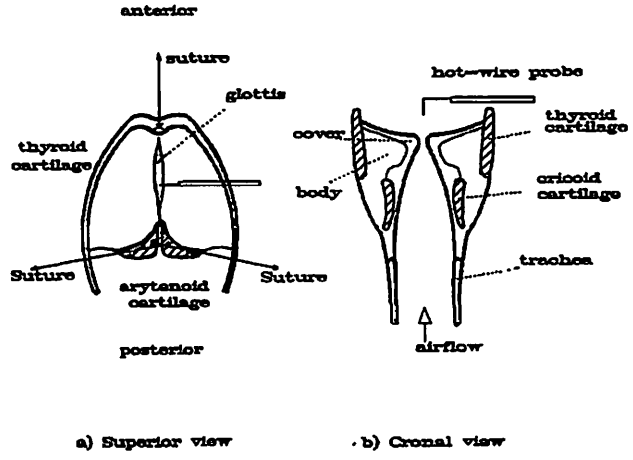


Figure 2. Schematic of the superior (top) view and coronal (vertical sectional) cut of the excised larynx. The top view shows the glottal slit frozen at maximum excursion.

(Figure 1) to measure velocity upstream of the glottis. The velocity in the jet above the glottis was measured at various locations with a miniature probe (with right-angled prongs, Dantec P14) with the sensor's direction perpendicular to the flow axis. The probe was mounted on a three-dimensional micropositioner (Deadal model 393M) access any location in the glottal jet (Figure 2). Experiments were conducted with several frequencies and mean subglottal pressures.

The subglottal pressure signal was amplified with a bridge amplifier (Bioamp model 205) and calibrated against the manometer at regular intervals. The hot-wire signal was calibrated against a pitot tube placed over the center of an air jet exiting a 0.6 cm diameter uniform tube. Humidified air at 35-37°C was used during both the hot-wire calibration and the larynx experiments. A best fit to the calibration was obtained with a least-square polynomial and used in subsequent data reduction. The velocity measurement accuracy was estimated to be no worse than ± 4% in the experimental range.

During the experiment, analog data from the hot-wire probes, EGG, and pressure transducer were monitored on a digital oscilloscope (Data 6000 Universal Waveform Analyzer, Data Precision) and simultaneously recorded on a Sony model PC-108M Digital Audio Tape (DAT) recorder. Portions of these recordings were digitized later with a 16-bit analog/digital convertor and analyzed on VAX-station computers. The multiple channels of signals were DC-coupled and digitized at the rate of 20 kHz (to ensure full waveform specification) for at least one second per channel (between 1.1 to 1.5 seconds). During the experiment, the top view of the larynx and vocal folds was recorded onto video tape for later analysis of the glottal images during oscillations.

To separate the turbulence from the periodic component of the velocity signal, the technique of phase shift averaging was employed. In this method, the EGG signal was used as the trigger signal to obtain cycle information. Every instantaneous velocity measurement on consecutive cycles corresponding to a phase point on the EGG trigger cycle was averaged over 100 to 300 cycles for the range of frequencies encountered in the experiments. This averaging process provided the so-called deterministic component of the oscillatory flow (the ensemble average). This average was then cross correlated to a shifted raw cycle to obtain the maximum correlation coefficient (see Kitney and Giddens, 1984, for details). This process was iterated until a reasonable convergence was achieved. This phase-averaged velocity shows the deterministic or pulsatile nature of the flow. When this signal is subtracted from the instantaneous velocity, the turbulent component is obtained. However, as noted by Walburn et al. (1983), when there is a cycle-to-cycle variation of oscillation period, using the phase averaged velocity will over predict the turbulent component. Thus, a nonrecursive digital low-pass filter (Hamming, 1973) was first used to smooth the velocity, so that the turbulent velocity fluctuations could be obtained independently of the phase shift averaged velocity. The turbulence intensity was calculated by taking the root mean square of the turbulent velocity fluctuation signal.

The recorded stroboscopic superior view of the larynx showed an essentially elliptical glottal opening that changed from a minimum to a maximum during each cycle. To obtain estimates of the oscillation amplitude, the strobe light was externally triggered with the EGG signal to freeze the image. By controlling the delay between the EGG and the trigger signal, the glottal view was set to its maximum area. The selected frame of glottal area was captured by a frame grabber and saved for digital image analysis. This image (Figure 3) was analyzed by the WinTrace software (developed at the Image Analysis Facilities of the University of Iowa). The calibration procedure for length and area was performed using a known distance on the image (a digitized millimeter grid) within the software itself. The maximum glottal area value (A_g) and glottal perimeter were measure by tracing the glottis contour with accuracy of ± 2% for area and ± 5% for perimeter. The mean glottal velocity (U_m; accuracy: ± 5%) was calculated by dividing the mean volume flow rate obtained from the rotameter by this maximum glottal area. The maximum glottal area, together with the corresponding measured glottal perimeter (G_p; accuracy: ± 5%), was used to obtain the hydraulic diameter ($D_h=4A_g/G_p$; accuracy: ± 7%) of the glottis during oscillations. The mean Reynolds number was calculated based on this length scale and mean glottal velocity ($Re=U_mD_h/\nu$; accuracy: ± 12%). The Strouhal number was also calculated based on these parameters and the frequency of oscillation ($St=fD_h/U_m$; accuracy: ± 8%). The Womersley number, defined on the basis of hydraulic diameter, frequency and viscosity ($\alpha=0.5D_h(2\pi f/\nu)^{0.5}$; accuracy: ± 8%), was also calculated.

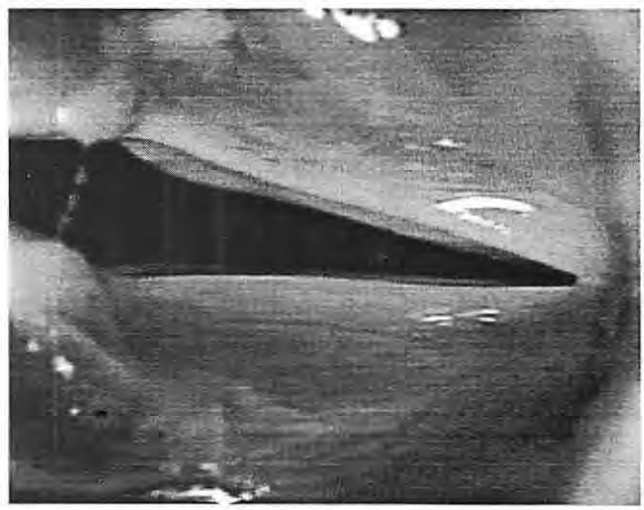


Figure 3. A digitized image of the glottis.

Results and Discussion

The experiments performed with four excised larynges: (1) male, weight of 22 kg, in situ vocal fold length of 14.0 mm; (2) male, 27 kg, 18.0 mm; (3) male, 25 kg, 17.5 mm; (4) female, 27 kg, 18.4 mm. For each larynx there was a range of pressure and flow rate for which the oscillation was audibly stable. For each larynx, one to three cases of oscillation with different adductory configurations were considered. The pressure and flow rate were varied within the stable range for each larynx and signals were recorded while the phonation was stable. Figure 4 shows the operation of the four larynges with twelve cases of pressure and flow variations. Every point on the figure represents a stable working condition of a larynx, and each symbol corresponds to one particular larynx. The overall data indicate that flow at the exit was mostly turbulent with a Reynolds number ranging between 1600 to 7000. The Strouhal number ranged between 0.002 and 0.03 with an average of 0.016. The corresponding range of the unsteadiness parameter was 2.6 to 15.9 with an average of 9.1. The data of Figure 4, obtained for realistic ranges of glottal pressure and flow, give values of Re and St which may be used in numerical simulations of pulsatile flow in the larynx.

Figure 5 shows a sample of the data from one particular larynx indicating, from the bottom, the EGG signal, the subglottal pressure (Ps), the subglottal velocity measured in the trachea (Vs), the supraglottal jet velocity (V_j), the phase-averaged jet velocity (<V>), and the turbulent component of the jet velocity (V'), respectively. There are four points indicated on the EGG waveform. Points 1 and 4 correspond to the approximate start of the glottal opening (Anastaplo and Karnell, 1988), point 2 is the approximate location of the beginning of glottal closing, and point 3 is where glottal closing is completed. In this example, the glottis was open longer than it was closed. This finding can be seen also from the velocity signals.

The subglottal pressure increased during the glottal closing portion of the cycle (interpreted from the EGG signal), continued to build up until the vocal folds began to open, and then decreased to a minimum. At lower frequencies, resonance in the subglottal system causes the subglottal pressure to contain more than one spectral component.

The subglottal velocity was pulsatile and appeared to be laminar (that is, had no apparent turbulence component). The pulsatile flow fell to a non-negligible minimum, indicating that the vocal folds did not close completely, or there was subglottal tracheal expansion near the time of glottal closure and maximum subglottal pressure. The peak of this signal fell in the region of the fully-open glottis (between points 1 and 2). This velocity may also have been affected by the resonance in the subglottal system.

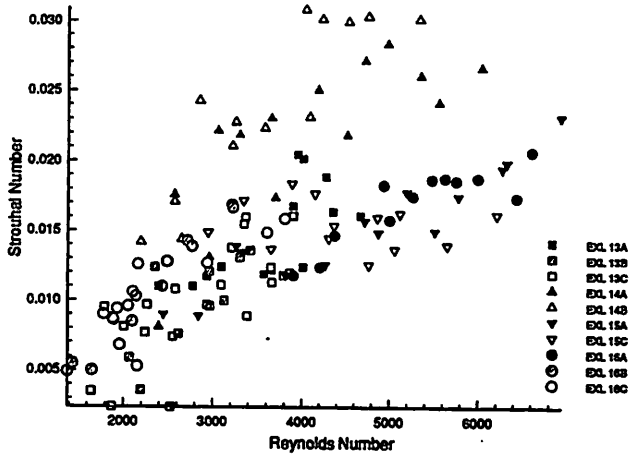


Figure 4. Strouhal and Reynolds numbers in tests with four excised larynges. Each larynx represented by different symbol. For these cases, the mean subglottal pressure ranged between 0.7 and 3.0 kPa, and mean flowrate ranged between 92 and 970 ml/s. (Experimental uncertainty in Reynolds number = ± 12% and in Strouhal number = ± 8%).

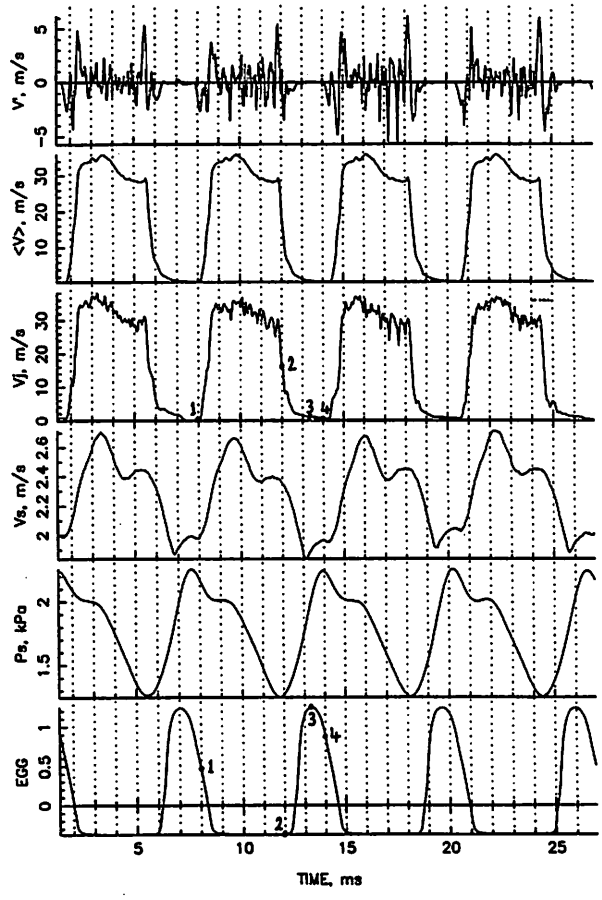


Figure 5. Waveforms of flow in typical larynx. The signals are, from the bottom, electroglottograph (EGG), subglottal pressure (Ps), subglottal velocity (Vs), jet velocity 10 mm above the vocal folds (Vj), phase-averaged velocity (<V>), and the turbulent velocity fluctuation (V'). (Experimental uncertainty in pressure = ± 2% and in velocity = ± 4%).

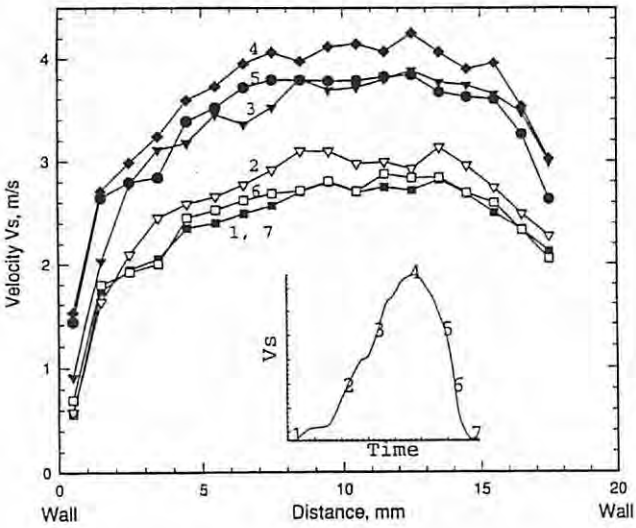


Figure 6. Velocity profiles upstream of the glottis (in the subglottal section). Profiles are selected at different phase positions in the oscillation cycle. (Experimental uncertainty in distance = ± 0.1% and in velocity = ± 4%).

In Figure 5, the hot-wire probe (P14) was positioned over the midline of the glottis and 10 mm above the vocal folds. At the instants of glottal opening and closing, the instantaneous velocity signal, as well as the phase-averaged velocity signal, show very large gradients, and these large gradients appear to be responsible for the spikes in the turbulent velocity fluctuations.

Figure 6 shows six velocity profiles in the subglottal section of the larynx. The probe was traversed in a direction parallel to the glottal edge (anterior-posterior). The profiles are foreshortened on the right of the figure because the hot-wire probe was kept away from the wall on that side to protect it from damage. These velocity profiles were selected at different phase positions of the oscillation cycle as indicated by the numbers on the single cycle of the subglottal velocity (insert). The same larynx was used for all 6 profiles. Ten cycles of each subglottal velocity waveform were averaged at 18 transverse locations and at 6 phases of the cycles. The resulting instantaneous velocity profiles are plotted in Figure 6. The first marker was selected at the peak of the EGG signal which, from Figure 5, is seen to be near maximum glottal closure. These subglottal velocity profiles changed in shape with time and were nearly parabolic in shape. The overall averaged subglottal velocity waveform (Figure 6 insert) was skewed to the right (asymmetric).

To quantify the shape of the jet exiting the glottis at various phases of the oscillation cycle, the following experiment was conducted. First a stable phonation was obtained. Then, by locking the strobe light with the EGG signal and adjusting the delay, the maximum opening was observed on the video in a stationary (freeze) mode. The area of the opening was divided into a 9x13 point grid with 0.25 mm

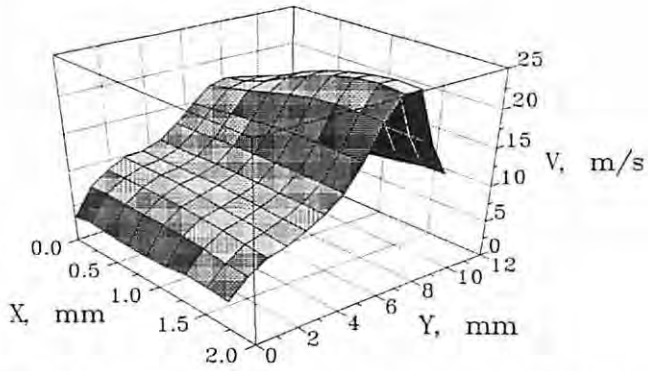


Figure 7. Time-averaged (mean) velocity 10 mm above the glottal cross-section.

spacing laterally (in the X direction, across the glottis) and 1 mm spacing longitudinally (in the Y direction, parallel to the glottis). The hot-wire probe (P14) was located at every grid point for about 5 seconds and moved to the next position using the X-Y micropositioner. The data were digitized later and processed on a VAX-station computer.

Figure 7 shows a velocity surface in the cross section of the glottal jet time-averaged over a duration of one second. For this case, the mean flow rate was approximately 230 ml/s and the subglottal pressure was approximately 1.4 kPa. The maximum glottal opening was about 1.25 mm. The measurement grid included a 0.25 mm extension on each side. Figure 7 clearly shows a highly nonuniform average velocity distribution in the glottal jet. There are large variations in velocity especially in the Y direction. The maximum velocity of approximately 25 m/s occurred at a location of about 3 mm from the anterior glottal commissure, whereas maximum glottal opening (as measured from the video tape recording) was approximately 5 mm from the anterior commissure. The location of the point of maximum velocity in the anterior glottis, anterior to the maximum vocal fold displacement, is similar to that reported by Berke et al. (1989) who measured a mean velocity as large as 40 m/s. The dynamic glottal configuration and flow impedance need further study to explain these data.

Conclusions

This study was concerned with the pulsatile flow from the glottis during phonation. Air velocities in excised canine larynges were measured using hot-wire anemometry, and were accompanied by simultaneous measurement of subglottal pressure, subglottal air velocity, and the electroglottographic signal. Phase-averaging of the velocity signals was performed to obtain the deterministic velocity. Subtracting the digital low-pass filtered velocity from the instantaneous velocity gave the turbulent velocity fluctuations. Recordings of velocity were made over the glottal midline as well as over a measurement grid above the glottis

to map the velocity field in the three-dimensional jet exiting the vocal folds.

Results of this study suggest the following:

1. The flow regime above the glottis in the excised canine larynx is turbulent over most of the open glottis portion of the vibratory cycle. The subglottal velocity did not show turbulence.
2. The Reynolds number ranged from 1600 to 7000; the corresponding Strouhal number range was 0.002 to 0.03, the Womersley number range was 2.6 to 15.9. The turbulence intensity ranged between 1.1 to 5.7 m/s.
3. The largest velocity values occurred through the anterior glottal region, anterior to the location of the cyclic maximum glottal excursion, in the example given.

These results together provide guidance to development of a theory or model of phonatory aeroacoustics.

Acknowledgments

The authors would like to thank Jim Knowles, Daryl Lorel and Katie Sacket for their assistance in larynx preparation and data collection. This work was supported by National Institute on Deafness and Other Communication Disorders, Grant No. DC00831-03.

References

Ahmed, S.A. and Giddens, D.P. (1984). "Pulsatile poststenotic flow studies with laser Doppler anemometry," Journal of Biomechanics, Vol. 17, pp. 695-705.

Alipour, F. and Titze, I.R. (1988). "A finite-element simulation of vocal fold vibration," in the Proceedings of the Fourteenth Annual Northeast Bioengineering Conference, Durham, N.H., IEEE publication #88-CH2666-6, pp. 186-189.

Anastaplo, S. and Karnell, M.P. (1988). "Synchronized videostroboscopic and electroglottographic examination of glottal opening," Journal of the Acoustical Society of America, Vol. 83, 1883-98.

Berke, G.S., Moore, D.M., Monkewitz, P.A., Hanson, D.G., and Gerratt, B.R. (1989). "A preliminary study of particle velocity during phonation in an *in-vivo* canine model," Journal of Voice, Vol. 3, No. 4, pp. 306-313.

Childers, D.G. (1985). "A critical review of electroglottography" CRC Critical Reviews in Biomedical engineering, Vol. 12, issue 2, pp. 131-161.

Conrad, W.A. (1980). "A new model of the vocal cords based on a collapsible tube analogy," Medical Research Engineering Vol. 13, pp. 7-10.

Einav, S. and Sokolov, M. (1993). "An experimental study of pulsatile pipe flow in the transition range," Journal of Biomechanical Engineering, Vol. 115, pp. 404-411.

Hamming, R.W. (1973). Numerical Methods for Scientists and Engineers, Second Edition, McGraw-Hill Book Company, New York.

Kitney, R.I. and Giddens, D.P. (1983). "Analysis of blood velocity waveforms by phase shift averaging and autoregressive spectral estimation," Journal of Biomechanical Engineering, Vol. 106, pp. 398-401.

Nakamura, M., Sugiyama, W. and Haruna, M. (1993). "An experiment on the pulsatile flow at transitional Reynolds numbers - the fluid dynamical meaning of the blood flow parameters in the aorta," Journal of Biomechanical Engineering, Vol. 115, pp. 412-417.

Park, D.K. (1989). A biofluid mechanics study of arterial stenoses, M.Sc. Thesis, Lehigh University, Bethlehem, PA.

Rosenfeld, M. (1993). "Validation of numerical simulation of incompressible pulsatile flow in a constricted channel," Computers and Fluids, Vol. 22, No. 2/3, pp. 139-156.

Scherer, R.C., Druker, D.G., Titze, I.R. (1988). "Electroglottography and direct measurement of vocal fold contact area", in O. Fujimura (Ed.) Vocal Physiology: Voice Production, Mechanisms, and Functions, (pp. 279-291), Raven Press, Ltd., New York.

Shapiro, A.H. (1977). "Steady flow in collapsible tubes," Trans. ASME, Journal of Biomechanical Engineering, August 1977, pp. 126-147.

Slavit, D.H., Lipton, R.J., and McCaffrey, T.V. (1990). "Glottographic analysis of phonation in the excised canine larynx," Annals of Otology, Rhinology & Laryngology, Vol. 99, No. 5, pp. 396-402.

Thornburg, H.J., Ghia, U., Osswald, G.A. and Ghia, K.N. (1992). "Efficient computation of unsteady vortical flow using flow-adaptive time-dependent grids," Fluid Dynamics Research, Vol. 10, pp. 371-397.

Tu, C., Deville, M., Dheur, L., and Vanderschuren, L. (1992). "Finite element simulation of pulsatile flow through arterial stenosis," Journal of Biomechanics, Vol. 25, No. 10, pp. 1141-1152.

Tutty, O.R. (1992). "Pulsatile flow in a constricted channel," Journal of Biomechanical Engineering, Vol. 114, pp. 50-54.

Walburn, F.J., Sabbah, H.N., and Stein, P.D. (1983). "An experimental evaluation of the use of an ensemble average for the calculation of turbulence in pulsatile flow," Annals of Biomedical Engineering, Vol. 11, pp. 385-399.

NCVS Status and Progress Report - 9
April 1996, 53-62

Aerodynamic Mechanisms UnderlyingTreatment-related Changes in Vocal Intensity in Patients with Parkinson Disease

Lorraine Olson Ramig, Ph.D., CCC-SP
Christopher Dromey, Ph.D., CCC-SP
Department of Communication Disorders and Speech Science, University of Colorado at Boulder
Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts

Abstract

The purpose of this study was to document changes in aerodynamic and glottographic aspects of vocal function in patients with Parkinson disease who received two forms of high effort treatment. Previous reports (Ramig, Countryman, Thompson, & Horii, 1995) have documented increased sound pressure level (SPL) following treatment which trained phonation and respiration (Lee Silverman Voice Treatment: LSVT), but not for treatment which trained respiration only (R). In order to examine the mechanisms underlying these differences, measures of maximum flow declination rate (MFDR) and estimated subglottal pressure (Psub) were made before and after treatment. A measure of relative vocal fold adduction (EGGW) was made from the electroglottographic signal during sustained vowel phonation. Sound pressure level data from syllable repetition, sustained vowel phonation, reading and monologue tasks were also analyzed to allow a more detailed understanding of treatment-related change in several contexts. Consistent with increases in SPL, significant increases in MFDR, estimated Psub and EGGW were measured post-treatment in patients who received the LSVT. Similar changes were not observed following R treatment. These findings suggest that the combination of increased vocal fold adduction and subglottal pressure is a key in generating post-treatment increases in vocal intensity in idiopathic Parkinson disease (IPD).

Introduction

The voices of patients with Parkinson disease are characterized by low vocal intensity, monotonicity, and hoarseness (Aronson, 1990; Logemann, Fisher, Boshes & Blonsky, 1978), which may contribute to a reduction in speech intelligibility (Ramig, 1992, Maclay, 1992). These characteristics have been associated with reduced vocal fold adduction (Hanson, Gerratt, & Ward, 1984; Perez, Ramig, Smith, & Dromey, in press; Smith, Ramig, Dromey, Perez, & Samandari, 1995) and impaired respiratory function (Critchley, 1981; Murdoch, Chenery & Bowler, 1989; Solomon & Hixon, 1993). Recently, post-treatment increases in vocal intensity (measured as the sound pressure level of a subject's speech) and fundamental frequency variability have been reported in patients with Parkinson disease following one of two forms of intensive speech treatment (Ramig, Countryman, Thompson, & Horii, 1995). One type of treatment focused on increasing inspiratory and expiratory respiratory volumes for speech in order to increase subglottal air pressure, since greater passive recoil forces at high lung volumes can contribute to pressure increases (Hixon, 1973). This approach was termed the respiratory effort treatment (R). The other treatment focused on increasing both respiratory effort and vocal fold adduction, and is known as the Lee Silverman Voice Treatment (LSVT). Post-treatment findings documented clinically and statistically significant increases in sound pressure level primarily for the subjects who received treatment designed to increase vocal fold adduction as well as respiratory volumes (Ramig et al., 1995). For example, the LSVT subjects increased on average 12 dB for sustained vowels, 6 dB for a reading passage, and 4 dB for a monologue task. On the other hand, subjects who received the respiratory treatment alone decreased on average 2 dB for sustained vowel phonation, and increased 2 dB for reading and 1 dB for the monologue task (Ramig et al., 1995).

It was speculated that the combination of increased subglottal air pressure and vocal fold adduction were necessary to optimize the aerodynamic mechanism of intensity control in patients with Parkinson disease. This speculation was based upon reports of vocal intensity control described in healthy subjects (Gauffin & Sundberg, 1989; Titze & Sundberg, 1992). For healthy speakers, increases in vocal intensity have been associated with increases in respiratory volumes (Hixon, 1973; Russell & Stathopoulos, 1988), subglottal air pressure (Isshiki, 1964) and vocal fold adduction (Scherer, 1991; Sundberg, Titze & Scherer, 1993).

The purpose of this study was to evaluate the aerodynamic and adductory changes associated with intensity control in patients with Parkinson disease following one of two forms of intensive speech treatment. It was hypothesized that subjects whose treatment focused on increased vocal fold adduction and respiratory volumes would show changes in aerodynamic measures which would correspond to mechanisms of intensity change in healthy speakers (Dromey, Ramig & Johnson, 1995). It was hypothesized that those subjects whose treatment emphasized increased respiratory volumes would have limited changes in aerodynamic and glottographic measures accompanying any SPL changes.

Method

Subjects

Aerodynamic data were collected from a group of 45 patients with idiopathic Parkinson disease (IPD) in an investigation comparing two forms of high effort speech treatment (Ramig et al., 1995). Because of extraneous movements (tremor, dyskinesia) and/or inability to perform the necessary syllable repetition task, data from 28 of these subjects were eliminated. In a number of cases, poor lip closure resulted in a visibly aberrant pressure waveform and thus an invalid estimate of subglottal pressure during /p/ closure. While care was taken to optimize the seal between the mask and the subject's face, subsequent analysis of the data suggested that leaks might have occurred in some

Table 1.
Selected mean (standard deviation in parentheses), minimum and maximum measures on demographic and speech variables, and treatment-related changes in SPL, for the subjects in the present study and for the larger group (Ramig et al., 1995) of which they were a subset.

Variable	LSVT Subset mean (sd)	Range min/max	LSVT Main Group[7]	Range min/max	Respiration Subset	Range min/max	Respiration Main Group[7]	Range min/max
Group n	10		26		7		19	
Age	62.7 (10.5)	49/79	63.5 (11.5)	32/79	63.7 (7.7)	52/71	65.6 (8.9)	51/83
PD Stage[1]	2.4 (.52)	2/3	2.7 (.68)	1.5/4	2.4 (.85)	1/3.5	2.3 (.80)	1/3.5
Years since Dx	5.3 (5.5)	0/18	7.0 (5.9)	0/20	6.6 (6.2)	1/18	5.6 (4.6)	0/18
UPDRS[2]	22.9 (11.1)	4/40.5	26.4 (13.4)	1.5/48	28.2 (13.6)	3/43	25.6 (14.4)	1/43
Depression[3]	6.9 (5.9)	1/20	9.0 (5.6)	0/20	10.6 (6.1)	1/22	9.9 (6.3)	0/25
Speech severity[4]	2.3 (.82)	1/3	2.8 (1.0)	1/5	2.7 (1.3)	1/4	2.7 (1.1)	1/4
/a/ SPL[5]	67.7 (4.2)	61.1/76.0	68.0 (4.7)	61.1/76.7	67.6 (4.0)	62.1/72.1	69.3 (5.1)	61.6/75.9
Reading SPL[6]	67.5 (3.4)	60.5/72.4	66.3 (4.0)	59.4/76.0	65.0 (1.9)	62.8/67.6	65.7 (2.7)	61.1/70.0
Pre/post /a/ SPL difference	14.1 (4.3)	7.1/21.7	13.2 (5.3)	4.6/22.1	-2.3 (3.6)	-7.6/2.1	-1.3 (5.3)	-11.7/10.3
Pre/post reading SPL difference	6.7 (4.0)	1.37/12.6	8.0 (5.9)	0.4/22.5	1.9 (2.3)	-1.4/4.5	2.5 (3.4)	-1.4/12.5

1 Hoehn and Yahr (1967)
2 Motor exam score
3 Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961)
4 Perceptual rating of impairment; 1 = mild, 5 = severe
5 dB SPL at 50 cm; mean of 6 vowels
6 dB SPL at 50 cm; mean over the entire Rainbow Passage
7 Ramig, Countryman, Thompson and Horii (1995)
Subjects in the present study were numbered 3, 6, 7, 9, 12, 15, 17, 18, 20, 22, 27, 35, 36, 38, 39, 43, 45 in the Ramig et al. (1995) study.

subjects, which resulted in their exclusion from the present study. In some instances, these mask leaks led to distortion of the flow signal, rendering it unusable for inverse filtering. A criterion for the inclusion or exclusion of a subject's data was selected based on the method described by Higgins and Saxman (1991). In their study, subjects whose minimum flow offset went below -.05 L/s were considered to have had mask leaks serious enough to jeopardize the validity of their aerodynamic data. Since the present investigation dealt with subjects who were both elderly and dysarthric, a slightly less strict level of -.075 L/s was chosen. Subjects whose data were poor for either flow or pressure were excluded from consideration in the present study. As shown in Table 1, the subgroup of 17 subjects who remained after the application of these criteria is comparable in terms of Parkinson disease characteristics and the severity of speech disorder to the group of 45 subjects presented previously (Ramig et al., 1995). The treatment-related changes in SPL were also comparable (see Table 1).

Statistical testing was undertaken to compare the pre-treatment characteristics of the two groups reported in the present study, who were a subset of the randomly assigned participants in the larger study of 45 subjects. Of those who remained after the removal of subjects with unusable aerodynamic data, 10 were from the LSVT group and 7 from the R group. There was one female in each treatment group. A one-way analysis of variance revealed no pre-treatment differences between these two treatment groups prior to the study on the variables of age ($F[1,15] = .0476$, $p = .830$), disease stage ($F[1,15] = .0168$, $p = .899$), score on the motor section of the UPDRS ($F[1,14] = .7282$, $p = .408$), time since diagnosis ($F[1,15] = .1982$, $p = .663$), depression ($F[1,15] = 1.7787$, $p = .202$), or severity of speech disorder ($F[1,15] = .6827$ $p = .422$). Additionally, there were no pre-treatment differences between these groups in vocal intensity for sustained vowel phonation ($F[1,15] = .0078$, $p = .931$), reading ($F[1,15] = 2.882$, $p = .110$) or monologue ($F[1,11] = .3701$, $p = .555$).

Instrumentation

A sound level meter (Bruel and Kjaer Type 2230) was positioned at 50 cm and a head-mounted microphone (AKG C410) at 8 cm from the subject's lips. A Synchrovoice Research Electroglottograph (EGG) was used to obtain the electroglottographic signal. A Rothenberg circumferentially vented pneumotachograph mask (Glottal Enterprises MS 100-A2 with transducer PTL-1) was held on the subject's face to collect the oral air flow signal. An intraoral pressure tube leading to a pressure transducer (Glottal Enterprises PTW-1) mounted on the pneumotachograph mask rested in the center of the oral cavity to allow the estimation of subglottal air pressure during /p/ closure (Smitheran & Hixon, 1981). The orientation of the tube was perpendicular to the flow of air. The output of the pressure transducer was monitored on an oscilloscope to ensure proper task performance through monitoring of the shape of the pressure peak.

The pneumotachograph transducer was calibrated against a flow meter across a range of values which exceeded the range of flows produced by subjects in the study. The pressure transducer was similarly calibrated across a range of pressures using a U-tube water manometer and syringe system. Vital capacity was measured using a Collins wet spirometer (Model 2785).

The analog sound pressure level, acoustic, EGG, air flow and oral pressure signals were stored on a Sony PC-108M 8 channel digital audio tape (DAT) recorder, which allowed signals up to 5 kHz bandwidth to be reproduced. In addition, speech acoustic and EGG signals were recorded on a Panasonic SV 3700 2 channel DAT recorder, which allowed higher bandwidth storage because of its 44 kHz sampling rate. Signals played back from the recorders were re-digitized into computer files using a 16 bit Digital Sound Corporation A/D converter to a VAX 4000/200 computer.

Procedure

Experimental data were collected during the week preceding treatment and during the week following treatment. The following procedure was carried out for each data collection session. The subject was seated in a medical examining chair in an IAC sound-treated booth. To limit extraneous movement, the subject's arms and legs were secured to arm and foot rests using three inch wide Velcro bands. After two minutes of tidal breathing, forced vital capacity (FVC) was measured. The subject was asked to take his or her deepest breath and blow out the air into a Collins wet spirometer "as hard and fast and long as you can". This task was repeated three times at the beginning of the session and twice at the end. The best performance was taken as the FVC. Care was taken to ensure that the subject's lips were tightly sealed on the mouthpiece; to prevent any nasal air flow, nose clips were used for this procedure.

For the collection of air flow and intraoral air pressure data, the subject repeated a series of seven /pae/ syllables with the Rothenberg mask held firmly in place by the experimenter, and the air pressure tube in the middle of the oral cavity. Syllables were produced at normal pitch and loudness and flat intonation at a rate of 1.5 syllables per second, as modeled by the experimenter (LR). The /pae/ syllable was selected because the distance between the fundamental frequency and first formant of the vowel facilitates the inverse filtering procedure (Rothenberg, 1973). When F0 and F1 are closely spaced, it can be difficult to adequately filter out the effects of vocal tract resonances in order to obtain a glottal flow waveform (Rothenberg, 1977).

During the data collection session, subjects also read the Rainbow Passage and produced a 20-30 second monologue, from which SPL and fundamental frequency data were extracted (Ramig et al., 1995).

Treatment

The subjects participated in one of two types of treatment. Both forms of treatment were designed to be intensive (16 1-hour sessions within 4 weeks) and high effort. Subjects were strongly encouraged to employ maximum effort in all therapy tasks. The first half of each session consisted of high effort drills, while the second half was used to carry over the increased effort to speech tasks. The implementation of high effort, intensive treatment is based on neurology and physical therapy practices (England & Schwab, 1959; Hallet & Khosbin, 1980; McDowell, Lee & Sweet, 1986) that suggest when pushed to higher effort levels, patients with Parkinson disease can compensate for bradykinesia and improve task performance. The daily training of high effort to increase magnitude of speech motor output is based upon the potentially similar problems in scaling the magnitude of output (e.g., stride length in walking, letter stroke in writing) observed in patients with Parkinson disease (Brooks, 1986; Muller & Stelmach, 1991; Stelmach, 1991). We speculate that intensive high effort speech treatment teaches patients to rescale the magnitude of speech motor output (Ramig, 1995; Ramig, Bonitati, Lemke & Horii, 1994), and that daily treatment with intensive practice (Schmidt, 1975; 1982) and feedback (knowledge of results) (Adams, 1971; 1986) facilitates this training.

The respiratory (R) treatment was designed to increase the activity of the respiratory musculature in order to generate increased volumes and subglottal air pressure for speech (Netsell & Daniel, 1979; Yorkston, Beukelman & Bell, 1988). Treatment tasks included those designed to increase ventilation as well as train expiratory muscles (Leith & Bradley, 1976; McKenzie & Gandevia, 1986): maximum inhalation and exhalation (Hardy, 1983; Netsell & Rosenbek, 1986), maximum duration of voiceless fricatives (/s/ and /f/), maximum duration of counting on one breath and sustaining oral air pressure for as long as possible using the IOPI (Iowa Oral Performance Instrument; Robin, Goel, Somodi & Luschei, 1992) both with and without the leak tube. This device, which was originally designed to measure tongue strength, had been modified to allow subjects to view the level of air pressure they were producing during therapy tasks. Subjects were encouraged to take frequent, deep breaths during speech, (e.g., think "breathe") and they were given visual feedback (NIMS Respigraph PN SY03) concerning chest wall movements during the maximum sustained fricative productions, reading and speaking tasks. The visual feedback allowed the subjects to view the relative magnitude of their respiratory maneuvers as they implemented the therapy techniques. The impact of the respiratory effort treatment on maximum duration vowel phonation and utterance and pause duration have been reported in a study by Ramig, Countryman, Thompson and Horii (1995).

The Lee Silverman Voice Treatment (LSVT) targeted increased vocal intensity through improved vocal fold adduction and increased respiratory effort (Ramig, Pawlas & Countryman, 1995). Therapy drills included tasks to increase respiratory/phonatory effort and coordination: maximum duration sustained vowel phonation and maximum fundamental frequency range. The tasks which were used to increase vocal fold adduction included stimulation to generate "loud" phonation or pushing the hands together, pulling or pushing on the arms of a chair while phonating a sustained vowel (Aronson, 1990; Froeschels, Kastein & Weiss, 1955). Feedback regarding intensity was provided with a voice light (a device which illuminates more LEDs as intensity increases) and a tape recorder. Subjects were encouraged to "think loud" and to use increased vocal intensity in speech tasks during the second half of each treatment session. The impact of the LSVT on vocal intensity and fundamental frequency, as well as several neuropsychological variables has been reported in a study by Ramig et al. (1995). Changes in pre- to post-treatment videolaryngostroboscopic ratings have been reported previously by Smith et al. (1995).

It is important to point out that the treatment goal of increasing vocal fold adduction in Parkinson disease is designed to maximize the efficiency of the phonatory source. It is never the goal of treatment to increase vocal fold adduction so that the voice becomes pressed or hyperadducted. The goal is a voice with sufficient loudness, generated with maximum phonatory efficiency (Ramig, 1995).

The treatment intensiveness, daily homework, daily quantification of treatment variables and carryover were all stimulated equally in both treatment groups. No direct attention was given to improving intonation, articulation or rate in either treatment group. Details of the daily treatment sessions, patient progress and clinician training have been summarized in Ramig et al. (1995), and Ramig, Pawlas & Countryman (1995).

Data analysis

From the /pae/ syllable train task, simultaneous recordings of air flow, intraoral air pressure and SPL were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz to allow detailed waveform preservation for the analysis of maximum flow declination rate. The signals were subsequently analyzed with custom software which interpolated between the pressure peaks during /p/ closure to allow the estimation of mid-vowel subglottal pressure (Smitheran & Hixon, 1981). Flow and SPL values for their respective time-aligned channels were also obtained in this way, so that values for these measures represented the temporal midpoint of the vowel in each syllable. These analyses were performed on a VAX 4000/200 computer, using the middle three syllables from each of three syllable trains from each recording session. Thus, nine tokens contributed to the mean for each subject on each occasion.

Maximum flow declination rate (MFDR), which reflects the interaction of subglottal air pressure and vocal

fold adduction, and is an index of the speed of glottal flow "shut-off," correlates highly with vocal intensity (Holmberg, Hillman, & Perkell, 1988; Titze & Sundberg, 1992). To measure maximum flow declination rate, the air flow signal was transferred to a 486 PC and was inverse filtered with CSpeech 4.0. The maximum flow declination rate was measured as the magnitude of the downgoing peak from the derivative of the glottal flow signal (see Figure 1a) for 10 successive cycles at the vowel midpoint. Again, nine tokens were examined to derive a mean value for each subject on each recording date.

Open quotient (OQ), defined as the time during which the vocal folds are open, divided by the period of the vibratory cycle, decreases as intensity increases (Dromey, Stathopoulos & Sapienza, 1992; Timcke, von Leden and Moore, 1958) in healthy speakers. Open quotient was measured using a custom software program (Stathopoulos & Sapienza, 1993a, 1993b) at a 20% AC flow criterion level. Thus, the time of glottal opening was operationally defined as the point at which the inverse-filtered flow reached 20% of the height of the waveform between the minimum flow offset (DC offset) and the peak flow. The beginning of the closed period was defined as the time at which the flow again decreased to the 20% level (see Figure 1b). Between 30 and 50 consecutive cycles were measured from the glottal flow waveform at the vowel midpoint during the /pae/ task. The mean values from three vowels from each of three trials for each subject for each session were calculated.

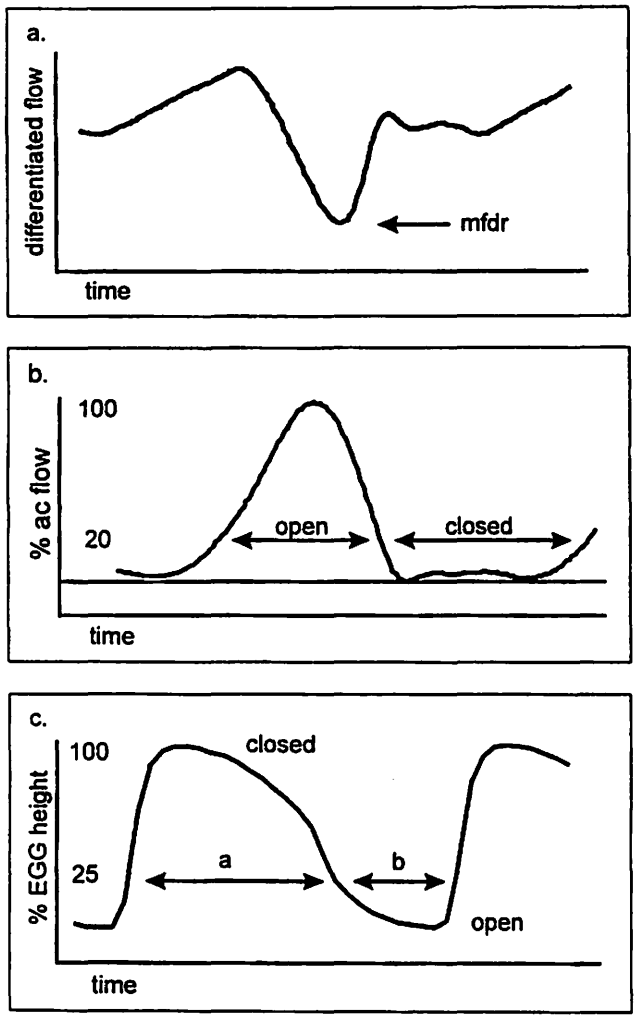


Figure 1. a) Measurement of maximum flow declination rate (MFDR) from the derivative of the glottal flow signal. MFDR represents the most rapid decrease in flow during vocal fold closure. b) Measurement of open quotient (OQ) at 20% of the height of the glottal flow pulse. OQ = open duration / open + closed durations. c) Measurement of EGGW-25 - the relative width of the electroglottographic waveform at 25% of its height. EGGW-25 = a / a + b.

EGGW-25, which is calculated as the relative width of the electroglottographic (EGG) waveform at 25% of its height (Scherer & Vail, 1988; Scherer, Vail & Rockwell, in press) has been found to increase with vocal fold adduction (Brosovic, 1994). Ten consecutive cycles of the electroglottographic (EGG) signal from the temporal midpoint from each of six maximum sustained vowel productions for each subject on each date were analyzed. This analysis was performed with an in-house software program on a VAX 4000/200 computer following digitization of the 2 channel DAT recording at 20 kHz. Figure 1c shows how this measure was derived.

SPL analysis was performed for six vowels and one reading passage for each subject on each recording date. The analog output of the sound level meter, which had been stored in digital format on the 8 channel DAT recorder, was re-digitized at 1 kHz into computer files. The mean SPL was calculated using a custom software program which allowed a SPL baseline to be set by the operator. By setting the baseline at an intensity corresponding to the lowest level during a phrase, intensity points falling below this level during inhalation were excluded from analysis, in order to measure the mean SPL during speech, rather than during speech and pauses combined.

Measurement Reliability

Twenty percent of the data on each measure were reanalyzed after several months by the same investigator to assess measurement reliability. A paired t-test revealed no significant differences between original and repeated analyses. Pearson correlation coefficients ranged from .995 to 1.0 between the original and the reanalyzed data. On the measures of estimated subglottal pressure and mid-vowel SPL, there was no measurement error, since these variables were automatically extracted from the waveforms. Mean measurement errors on the other variables were in the order of 0.5%.

To assess test-retest intrasubject reliability, 9 (53%) of the subjects had a second pre-treatment recording session,

and 5 (29%) of the subjects had a second post-treatment recording. Paired t-tests revealed that there were no significant differences between pre-treatment recording sessions for any variables except the measure of forced vital capacity, which was smaller (0.12 L) for the second pre-treatment recording ($t[9] = 2.45$, $p = .037$) than for the first. There were no differences on any measures between the first and second post-treatment recordings.

Results

Means and standard deviations for each measure are presented in Table 2. A two factor time (pre-to-post) by treatment group (LSVT vs. R) repeated measures analysis of variance (ANOVA) was performed on each of the dependent measures to examine changes which occurred following treatment, as well as differences in these pre/post changes according to the type of treatment that was given. The F-ratios and p-values for these ANOVAs are reported in Table 3.

SPL

For sustained vowel phonation, there was a significant pre/post by treatment group interaction ($p < .001$). The LSVT subjects increased on average by 14 dB (sd 4.3). The R group subjects decreased 2.3 dB (sd 3.5). For the mid-vowel intensity in the syllable repetition task there was a significant pre/post by treatment group interaction of ($p = .042$). The LSVT subjects increased on average 6.5 dB (sd 7.4) and the R subjects decreased 0.1 dB (sd 2.8) during the syllable repetition task (see Figure 2a) from pre- to post-treatment. For the reading passage, there was a significant pre/post by treatment group interaction ($p = .013$). The mean increase for the LSVT subjects was 6.7 dB (sd 4.2), and for R group, it was 1.9 dB (sd 2.3). In the monologue task, the pre/post main effect was significant ($p = .005$), but there was not a significant interaction with treatment group at the $p < .05$ level. Nevertheless, there was a tendency for the LSVT subjects to show larger SPL increases (5.5 dB, sd 3.1) than the R group subjects (2.2 dB, sd 3.8).

The treatment-related SPL changes reported for the 17 subjects here are similar to those for the larger group of 45 subjects of which they were a subset (Ramig, Countryman, Thompson & Horii, 1995 — see Table 1).

For estimated subglottal pressure, there was a significant pre/post by treatment group interaction ($p = .003$). LSVT subjects increased on average 2 cmH2O (sd 1.4), and the R group decreased 0.3 cmH2O (sd 1.3) (see Figure 2b). There was a significant pre/post by treatment group interaction ($p = .001$) for maximum flow declination rate. The mean increase for the LSVT subjects was 215 L/s/s (sd 119), and the mean decrease for the R group was 2 L/s/s (sd 103) (see Figure 2c).

The changes in open quotient following treatment were not statistically significant for main or interaction effects. The LSVT group decreased by .042 (sd .056), while the R group increased by .014 (sd .086). A lower value on this measure is associated with greater vocal fold adduction. There was a significant pre/post by treatment group interaction ($p = .003$) for EGGW-25. The LSVT group increased on average by .073 (sd .050) following treatment, whereas the R group decreased by .035 (sd .063) (see Figure 2d). A higher value on this measure is associated with greater vocal fold adduction (Scherer et al., in press).

Table 2.
Means (and standard deviations) for each measure before and after treatment as a function of treatment type.

	LSVT				Respiration			
Variable	Pre-Tx	sd	Post-Tx	sd	Pre-Tx	sd	Post-Tx	sd
SPL for /a/	67.7	(4.2)	81.7	(3.1)	67.5	(4.0)	65.2	(4.2)
SPL for /pae/	68.5	(4.2)	75.0	(5.8)	66.7	(2.9)	66.6	(1.7)
SPL for reading	67.5	(3.4)	74.2	(3.5)	65.0	(1.9)	66.9	(3.9)
SPL for monologue	64.5	(2.9)	70.0	(3.5)	64.9	(2.4)	67.1	(2.9)
Psub	6.6	(2.5)	8.6	(2.9)	6.6	(2.3)	6.3	(1.9)
MFDR	488	(243)	703	(275)	350	(65)	348	(88)
OQ	.427	(.082)	.385	(.090)	.443	(.160)	.449	(.192)
EGGW-25	.545	(.051)	.619	(.051)	.575	(.051)	.550	(.073)
FVC	3.57	(.760)	3.53	(.824)	3.29	(1.09)	3.21	(1.04)

SPL measures are dB SPL at 50 cm; mean for 6 vowels, 9 /pae/ syllables or entire passage or monologue
Psub = estimated subglottal pressure (cmH2O); mean for 9 /pae/ syllables
MFDR = maximum flow declination rate (L/s/s); mean for 9 /pae/ syllables
OQ = open quotient using 20% AC flow criterion; mean for 9 /pae/ syllables
EGGW-25 = EGG pulse width adduction measure using 25% height criterion; mean for 6 vowels
FVC = forced vital capacity (liters); maximum value of three attempts

Table 3.
F-ratios and p-values for the repeated measures ANOVAs on all variables.

	Pre/post effect		Interaction with treatment type		
Variable	F-ratio	p-value	F-ratio	p-value	df
SPL for /a/	34.68	< .001	67.62	< .001	1,15
SPL for /pae/	4.71	.048	5.02	.042	1,14
SPL for reading	25.97	< .001	8.05	.013	1,15
SPL for monologue	13.43	.005	2.42	.154	1,9
Psub	3.72	.020	12.21	.003	1,15
MFDR	14.70	.002	15.15	.001	1,15
Open quotient	1.44	.249	2.59	.128	1,15
EGGW-25	3.20	.095	13.40	.003	1,14
FVC	1.67	.215	.13	.720	1,15

The changes following treatment were not statistically significant for main or interaction effects for forced vital capacity.

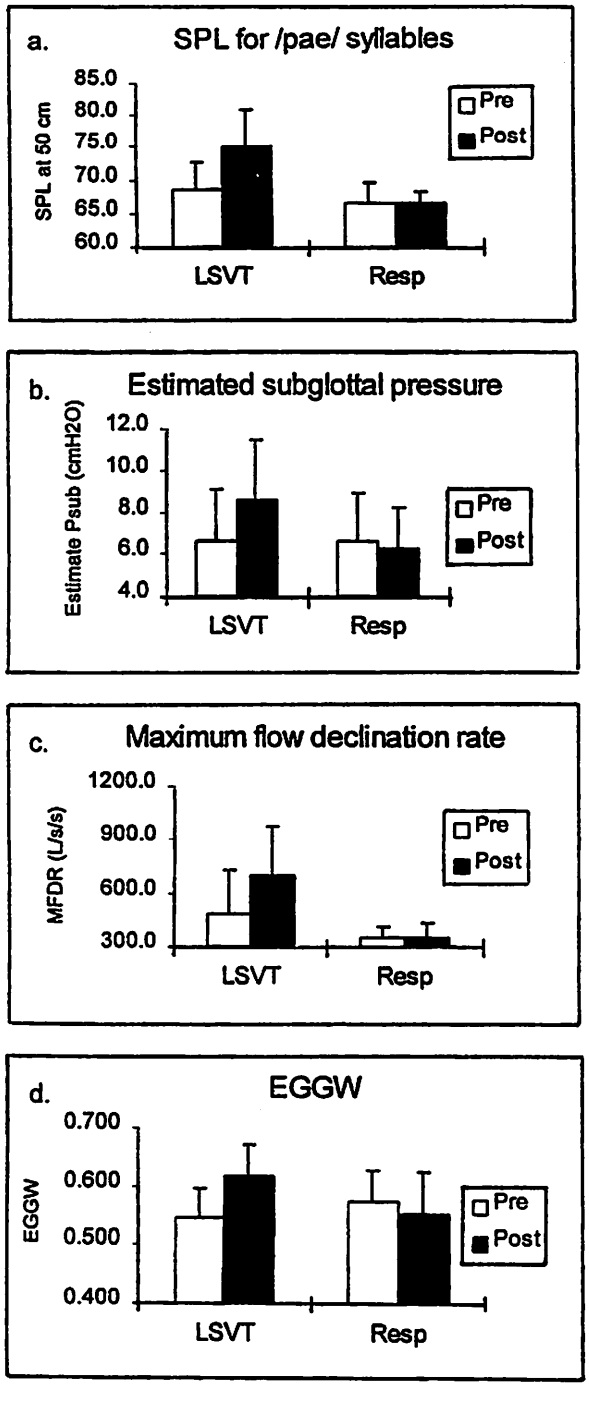


Figure 2a) Mean and standard deviation sound pressure level (dB SPL at 50 cm) for /pae/ syllable repetition before and after speech treatment. b) Mean and standard deviation estimated subglottal pressure (cmH2O) before and after speech treatment. c) Mean and standard deviation maximum flow declination rate (L/s/s) before and after speech treatment. d) Mean and standard deviation EGGW (25% of AC waveform criterion) before and after speech treatment.

Discussion

The present study was undertaken to examine selected aerodynamic and glottographic measures of vocal function in patients with Parkinson disease. A primary objective was to document changes in these measures accompanying two different forms of speech treatment.

Since reduced vocal intensity is one of the primary speech disorders of this patient population, and contributes to reductions in speech intelligibility, a treatment program which improves vocal intensity is of prime interest. The data for four speech tasks - syllable repetition, sustained vowel phonation, reading and monologue - all show that patients receiving the LSVT were able to increase their SPL following treatment. The fact that patients whose treatment focused solely on increased respiratory volumes and subglottal air pressure did not make similar SPL gains suggests that it is necessary to compensate for the documented deficiencies in vocal fold adduction in this population in order to achieve the goal of increased SPL (Ramig, Countryman, Thompson & Horii, 1995). While increasing subglottal pressure through respiratory volume improvements leads to increased intensity in the healthy larynx (Stathopoulos & Sapienza, 1993a), greater pulmonary effort alone did not lead to improvements in SPL in the glottally incompetent patients in the present study. The treatment tasks of the LSVT included exercises to increase vocal fold adduction, and as the EGGW data and the videostroboscopic findings (Smith et al., 1995) suggest, post-treatment increases in vocal fold adduction were realized. This adductory improvement allowed the LSVT subjects to increase their vocal intensity. These findings are consistent with data reported by Berke, Hanson, Gerratt, Trapp, Macagba, and Natividad (1990), who found that higher SPL was achieved more efficiently by means of medial adductory compression than by higher air flow levels in an in vivo canine model. The post-treatment SPL for /pae/ syllables in the LSVT group is comparable to values reported by Stathopoulos and Sapienza (1993a) for healthy males at a comfortable intensity level.

It is of particular interest that all subjects receiving respiratory effort treatment did not experience substantial improvements in estimated subglottal pressure; only three of the seven subjects increased subglottal air pressure pre- to post-treatment. Observations of statistically significant post-treatment increases in pause durations during connected speech and increases in maximum duration of sustained vowel phonation (Ramig et al., 1995) both suggest that the R group did implement the target therapy techniques to increase inspiratory volumes prior to speech. The lack of corresponding increases in estimated subglottal pressure and SPL allow the speculation that for individuals with an incompetent glottal valving mechanism, merely increasing respiratory force may not be sufficient to lead to improvements in subglottal pressure for speech. An alternative explanation is that the activities used in treatment to increase respiratory

function did not adequately train subglottal pressure or generalize the use of expiratory effort to the syllable repetition task examined in the present study.

It is possible that treatment approaches which focus solely on respiration may even be counter-productive in this population. The decreases in vocal fold adduction reflected in EGGW-25 data, together with previous findings by Smith et al. (1995) suggest that glottal incompetence actually became greater in certain subjects who increased respiratory effort without simultaneously improving vocal fold adduction. While not demonstrating such a clear trend, the open quotient data also suggest a slight increase in glottal incompetence for the R group, while the LSVT group appeared to increase adduction based upon this measure. Berke at al. (1990) found that higher flows in the absence of adductory changes led to increased open quotient in their canine preparation — a similar pattern to that seen in the R treatment group in the present study.

Glottal incompetence may not preclude generation of subglottal air pressure in individuals with a non-neurologically disordered respiratory system. However, for individuals with Parkinson disease, it may be necessary to stimulate respiration and phonation simultaneously in order to maximize treatment effectiveness to generate SPL increases.

It might be speculated that even though patients with Parkinson disease in the R group were trained to inhale to higher lung volumes before speaking, they counteracted the unusually high respiratory recoil forces by using inspiratory muscular effort (Hixon, 1973). It is conceivable that the R group subjects might have checked the increases in subglottal pressure to allow speech production to continue in an intensity range to which they were accustomed. This may be similar to the technique employed by singers, who have been reported to inhale to high lung volume levels and then use inspiratory muscles to counteract excess pulmonary recoil forces when singing softer passages (Watson & Hixon, 1985).

The pre-treatment values for estimated subglottal pressure (Psub) fell within the range reported by Gracco, Gracco, Lofqvist and Marek (1994) for untreated patients with Parkinson disease. Prior to treatment, Psub values were lower than those reported by Higgins and Saxman (1991) for healthy elderly males. For the LSVT group only, post-treatment estimated Psub values were closer to the normative data. This suggests that the treatment might have brought the subglottal pressure of these subjects into a more normal range for their age, whereas the R treatment did not have this effect. However, the lack of SPL data in the reports of these previous authors precludes a more detailed comparison of the healthy and pathological subject groups.

The increase in SPL for the LSVT group was accompanied by a similar increase in MFDR. This can be taken as evidence that the subjects in this group increased their sound pressure level by the same means documented in the healthy larynx (Dromey et al., 1995), in that glottal flow shut-off occurred more rapidly, allowing a greater excitation of the vocal tract (Gauffin & Sundberg, 1989). These changes could be interpreted as an indication that the LSVT therapy was successful in helping subjects achieve a more normal mode of phonation. The R group, on the other hand, did not increase in SPL, and therefore continued with the weak phonation which is characteristic of patients with Parkinson disease.

The open quotient (OQ) data from the inverse filtered airflow signal, which reflect changes in vocal fold adduction, did not reach statistical significance in any pre/post treatment analysis. However, EGGW-25 values, which also change with vocal fold adduction, showed a significant time by treatment group interaction. A recent study of patients with Parkinson disease which compared EGGW measures with stroboscopic examination reported that the two measures correlated (Brosovic, 1994), in that visible improvements in vocal fold closure coincided with higher EGGW values. The differences in the strength of the trends for the two types of measure reported here could be due to the nature of the respective speech tasks. The EGGW measure was derived from the midpoint of a sustained vowel, whereas the open quotient measure was obtained from a syllable repetition task. It is possible that the rapid changes in laryngeal configuration in the dynamic speech task contributed to greater variability in the aerodynamic measure, thus precluding statistically significant results. It is also possible that the EGGW measure offers greater validity than airflow open quotient as a measure of laryngeal adduction.

The fact that over 60% of the aerodynamic data collected were eliminated from further analysis in this study has important implications for researchers investigating disordered speech production. Many of the subjects were unable to perform the /pae/ syllable repetition task in a way that allowed valid data to be obtained (Hertegard, Gauffin, & Lindestad, 1995). Some were unable to produce syllables with constant effort in a single breath group, or even relatively stable intensity within a single syllable. Others did not achieve a sufficient lip seal around the oral pressure tube, which precluded valid pressure estimates. These difficulties limit the applicability to disordered speakers of protocols or techniques which have been developed using healthy individuals as subjects (Finnegan, Luschei, Barkmeier & Hoffman, 1996). This is unfortunate, given the value of physiologic data in assessing disordered speech production in order to plan treatment or monitor progress.

In summary, this study was undertaken to document the aerodynamic mechanisms of change in two groups of patients with Parkinson disease, one of which was successful in increasing vocal intensity. Subjects who received the LSVT were able to achieve increases in SPL through improved vocal fold adduction and increases in subglottal

pressure. Sound pressure level did not consistently increase pre to post-treatment for subjects who received only respiratory training.

Acknowledgements

This research was supported in part by grants NIH-NIDCD #R01 DC01150 and #P60 DC00976. We are grateful for the suggestions of three anonymous reviewers based on an earlier version of this manuscript.

References

Adams, J.A. (1971). A closed-loop theory of motor learning. Journal of Motor Behavior, 3, 111-150.

Adams, J.A. (1986). Use of the model's knowledge of the results to increase observer's performance. Journal of Human Movement Studies, 12, 89-98.

Aronson, A.E. (1990). Clinical voice disorders. New York: Thieme-Stratton.

Beck, A.T., Ward, C.H., Mendelson, M, Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561-571.

Berke, G.S., Hanson, D.G., Gerratt, B.R., Trapp, T.K., Macagba, C., & Natividad, M. (1990). The effect of air flow and medial adductory compression on vocal efficiency and glottal vibration. Otolaryngology Head and Neck Surgery, 102, 212-218.

Brookes, V.B. (1986). The neurological basis of motor control. New York: Oxford University Press.

Brosovic, G. B. (1994). Voice therapy and Parkinson disease: Measures of vocal fold adduction. Unpublished thesis - University of Colorado at Boulder.

Critchley, E.M.R. (1981). Speech disorders of Parkinsonism. A Review Journal of Neurology, Neurosurgery and Psychiatry, 44, 751-758.

Dromey, C., Ramig, L.O., & Johnson, A.B. (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. Journal of Speech and Hearing Research, 38, 751-764.

Dromey, C., Stathopoulos, E.T., & Sapienza, C.M. (1992). Glottal airflow and electroglottographic measures of vocal function at multiple intensities. Journal of Voice, 4, 44-54.

England, A.C., & Schwab, R.S. (1959). The management of Parkinson's disease. AMA Archives of Internal Medicine, 104, 439-460.

Finnegan, E.M., Luschei, E.S., Barkmeier, J.M., & Hoffman, H.T. (1996). Sources of error in estimation of laryngeal airway resistance in persons with spasmodic dysphonia. Journal of Speech and Hearing Research, 39, 105-113.

Froeschels, E., Kastein, S., Weiss, D. (1955). A method of therapy for paralytic conditions of the mechanisms of phonation, respiration and glutination. Journal of Speech and Hearing Disorders, 20, 364-370.

Gauffin, J., & Sundberg, J. (1989). Spectral correlates of glottal voice source waveform characteristics. Journal of Speech and Hearing Research, 32, 556-565.

Gracco, L.C., Gracco, V.L., Lofqvist, A., & Marek, K.P. (1994). Aerodynamic evaluation of Parkinsonian dysarthria: Laryngeal and supralaryngeal manifestations. In J.A. Till, K.M. Yorkston, & D. Beukelman (Eds.), Motor speech disorders: Advances in assessment and treatment (pp. 65-79). Baltimore: Paul H. Brookes.

Hallet, M. & Khosbin, S. (1980). A psychological mechanism of bradykinesia. Brain, 103, 301-314.

Hanson, D.G., Gerratt, B.R., & Ward, P.H. (1984). Cinegraphic observations of laryngeal function in Parkinson's Disease. Laryngoscope, 94, 348-353.

Hardy, J. (1983). Cerebral palsy. Englewood Cliffs, NJ: Prentice Hall. Hertegard, S., Gauffin, J., & Lindestad, P. (1995). A comparison of subglottal and intraoral pressure measurements during phonation. Journal of Voice, 9, 149-155.

Higgins, M.B., & Saxman, J.H. (1991). A comparison of selected phonatory behaviors of healthy aged and young adults. Journal of Speech and Hearing Research, 34, 1000-1010.

Hixon, T.J. (1973). Respiratory function in speech. In F. Minifie, T. Hixon, & F. Williams (Eds.), Normal aspects of speech, hearing and language (pp. 73-125). Englewood Cliffs, NJ: Prentice-Hall.

Hoehn, M., & Yahr, M. (1967). Parkinsonism: Onset, progression and mortality. Neurology, 17, 427-442.

Holmberg, E.B., Hillman, R.E., & Perkell, J.S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. Journal of the Acoustical Society of America, 84, 511-529.

Isshiki, N. (1964). Regulatory mechanism of voice intensity variations. Journal of Speech and Hearing Research, 7, 17-29.

Leith, D. & Bradley, M. (1976). Ventilatory muscle strength and endurance training. Journal of Applied Physiology, 41(4), 508-516.

Logemann, J.A., Fisher, H.B., Boshes, B., & Blonsky, E.R. (1978). Frequency and concurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and Hearing Disorders, 42, 47-57.

Maclay, S. (1992). The effectiveness of intensive voice therapy on speech intelligibility in patients with Parkinson disease. Unpublished master's thesis; University of Colorado-Boulder.

McDowell, F.H., Lee, J.E., & Sweet, R.D. (1986). Extrapyramidal disease. In A.B. Baker & R.J. Joynt (Eds.) Clinical neurology (pp.24-26). Philadelphia: Harper and Row.

McKenzie, D., & Gandevia, S. (1986). Strength and endurance of inspiratory, expiratory, and limb muscles in asthma. American Review of Respiratory Disease, 134, 999-1004.

Muller, F., & Stelmach, G.E. (1991). Scaling problems in Parkinson's disease. In Requin & G.E. Stelmach (Eds.) Tutorials in motor neuroscience (pp.161-174). Netherlands: Kluwer Academic Publishers.

Murdoch, B.E., Chenery, H.J., & Bowler, S. (1989). Respiratory function in Parkinson's subjects exhibiting a perceptible speech deficit: A kinematic and spirometric analysis. Journal of Speech and Hearing Disorders, 54, 610-626.

Netsell, R., & Daniel, B. (1979). Dysarthria in adults: Physiologic approach to rehabilitation. Archives of Physical Medicine and Rehabilitation, 60, 502-508.

Netsell, R., & Rosenbek, J. (1986). Treating the dysarthrias. In R. Netsell, A Neurobiologic view of speech production and the dysarthrias. San Diego: College Hill Press (pp. 123-152).
Perez, K.S., Ramig, L.O., Smith, M.E., & Dromey, C.D. (in press). The Parkinson larynx: Tremor and videostroboscopic findings. Journal of Voice.

Ramig, L.O. (1992). The role of phonation in speech intelligibility: A review and preliminary data from patients with Parkinson's disease. In R. Kent (Ed.), Intelligibility in speech disorders: Theory, measurement, and management (pp. 119-155). Amsterdam: John Benjamin.

Ramig, L.O. (1995). Speech therapy for Parkinson's disease. In W. Koller & G. Paulson (Eds.), Therapy of Parkinson's disease (pp. 539-548). New York: Marcel Dekker.

Ramig, L.O., Bonitati, C.M., Lemke, J.H., & Horii, Y. (1994). Voice therapy for patients with Parkinson's disease: Development of an approach and preliminary efficacy data. Journal of Medical Speech-Language Pathology, 2, 191-210.

Ramig, L.O., Countryman, S., Thompson, L.L., & Horii, Y. (1995). A comparison of two forms of intensive speech treatment for Parkinson disease. Journal of Speech and Hearing Research, 38, 1232-1251.

Ramig, L.O., Pawlas, A.A., & Countryman, C. (1995). The Lee Silverman Voice Treatment (LSVT): A practical guide to treating the voice and speech disorders in Parkinson disease. Iowa City: National Center for Voice and Speech.

Robin, D.A., Goel, A., Somodi, L.B., & Luschei, E.S. (1992). Tongue strength and endurance: Relation to highly skilled movements. Journal of Speech and Hearing Research, 35, 1239-1245.

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. Journal of the Acoustical Society of America, 53, 1632-1645.

Rothenberg, M. (1977). Measurements of airflow in speech. Journal of Speech and Hearing Research, 20, 155-176.

Russell, N.K., & Stathopoulos, E.T. (1988). Lung volume changes in children and adults during speech production. Journal of Speech and Hearing Research, 31, 146-155.

Scherer, R.C. (1991). Physiology of phonation: A review of basic mechanics. In C.N. Ford and D.M. Bless (Eds.), Phonosurgery: Assessment and surgical management of voice disorders (pp. 77-93). New York: Raven Press Ltd.

Scherer, R.C., & Vail, V.J. (1988). Measures of laryngeal adduction. Journal of the Acoustical Society of America, suppl. 84, (1) S81 (A).

Scherer, R.C., Vail, V.J., & Rockwell, B. (in press). Examination of the laryngeal adduction measure EGGW. In K. S. Harris, F. Bell-Berti, & L.J. Raphael (Eds.), Producing speech: Contemporary Issues (pp. 533-555). American Institute of Physics.

Schmidt, R.A. (1975). A schema theory of discrete motor skill learning. Psychological Review, 82, 225-260.

Schmidt, R.A. (1982). Motor control and learning. Champaign, IL: Human Kinetic Publishers.

Smith, M.E., Ramig, L.O., Dromey, C., Perez, K.E., & Samandari, R. (1995). Intensive voice treatment in Parkinson's disease: Laryngostroboscopic findings. Journal of Voice, 9, 453-459.

Smitheran, J.R., & Hixon, T.J., (1981). A clinical method for estimating laryngeal airway resistance during vowel production. Journal of Speech and Hearing Disorders, 46, 138-146.

Solomon, N.P., & Hixon, T.J., (1993). Speech breathing in Parkinson disease. Journal of Speech and Hearing Research, 36, 294-.

Stathopoulos, E.T., & Sapienza, C.M. (1993a). Respiratory and laryngeal function of women and men during vocal intensity variation. Journal of Speech and Hearing Research, 36, 64-75.

Stathopoulos, E.T., & Sapienza, C.M. (1993b). Respiratory and laryngeal measures of children during vocal intensity variation. Journal of the Acoustical Society of America, 94, 2531-2543.

Stelmach, G.E. (1991). Basal ganglia impairment and force control. In J. Requin & G.E. Stelmach (Eds.), Tutorials in motor neuroscience (pp.137-148). Netherlands: Kluwer Academic Publishers.

Sundberg, J., Titze, I.R., & Scherer, R.C. (1993). 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, 7, 15-29

Timcke, R., von Leden, H., Moore, P. (1958). Laryngeal vibrations: Measurements of the glottic wave. AMA Archives of Otolaryngology, 68, 1-19.

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

Watson, P.J., & Hixon, T.J. (1985). Respiratory kinematics in classical (opera) singers. Journal of Speech and Hearing Research, 28, 104-122.

Yorkston, K.M., Beukelman, D.R., & Bell, K.R. (1988). Clinical management of dysarthric speakers. Boston: Little, Brown and Co.

NCVS Status and Progress Report - 9
April 1996, 63-69

The Membranous Contact Quotient, A New Phonatory Parameter of Glottal Contact

Ronald Scherer, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Fariborz Alipour, Ph.D.
Eileen Finnegan, M.S.
Department of Speech Pathology and Audiology, The University of Iowa
Chwen Geng Guo, M.S.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts

Abstract

The Membranous Contact Quotient, MCQ, is introduced as a measure of anterior-posterior (A-P) glottal contact and glottal competence. It is defined as the ratio of the membranous contact glottis (the A-P length of the contact between the two membranous vocal folds) and the membranous vocal fold length. An elliptical approximation to the vocal fold contour during phonation was used to predict MCQ values as a function of vocal process gap (adduction), maximum glottal width, and membranous glottal length. MCQ is highly dependent on vocal process gap and maximum glottal width, but not on vocal fold length. Excised larynges were used to obtain MCQ data for a wide range of vocal process gap and maximum glottal width. Predicted and measured MCQ values had a correlation of 0.93. The model is better at higher values of MCQ. The model of MCQ is also expressed as a function of vocal process gap and subglottal pressure to suggest production control potential. The MCQ parameter is obtainable with the use of stroboscopy and appears to be a potentially useful clinical measure, especially for cases requiring a measure of glottal incompetence.

There are a number of vocal fold motion parameters that are nondimensional and have kinematic, acoustic, and aerodynamic importance in phonatory description and diagnostics. For example, the open quotient corresponds to how long the glottis is open divided by the length of the glottal period. The speed quotient for glottal motion is the time it takes the vocal folds to travel from glottal midline to the maximum excursion divided by the time from maximum excursion back to the midline. These two quotients can be obtained using stroboscopy and frame counting of the slow motion glottal movement (1). In this paper we introduce another nondimensional quotient which can be obtained using stroboscopy. It partially fills the need to describe glottal competency, and is especially applicable to breathy voice and cases of glottal obstruction and bowing.

The new phonatory parameter is a measure of the degree of glottal contact and is called the Membranous Contact Quotient, MCQ. During phonation, the membranous vocal folds may not come into complete anterior-posterior (A-P) contact. The MCQ parameter quantifies the portion of the membranous vocal fold length in contact during maximum closure within a cycle of phonation.

The definition of MCQ is the length of the maximum A-P contact of the vocal folds (the membranous contact glottis, MCG, see Figure 1) during a cycle of vibration, divided by the linear distance between the anterior commissure and the vocal processes (the membranous glottal length, L). Typically the MCQ parameter would be obtained by viewing stroboscopic slow motion imaging of vocal fold vibration, and measuring the relative lengths of the longest membranous contact glottis and the membranous glottal length.

MCQ allows a more complete description of the dynamic aspects of glottal motion during phonation. It can be included with other characteristics of glottal vibration

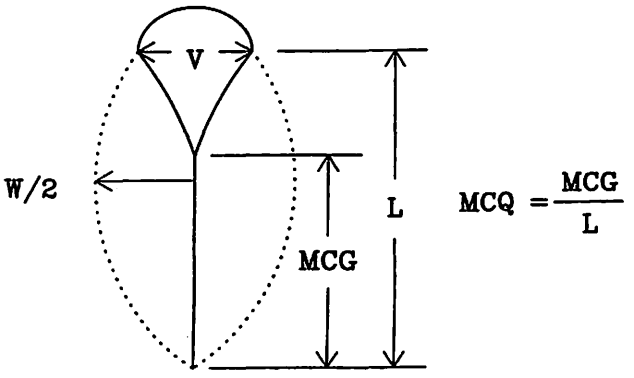


Figure 1. Superior view of the glottal configuration for purposes of obtaining MCQ. V refers to the vocal process gap. L refers to the membranous glottal length. MCG is the length of the membranous contact glottis obtained at the maximum glottal closure during a phonatory cycle. W/2 is half of the amplitude of vibration of the vocal folds. MCQ is the Membranous Contact Quotient, defined by the ratio of MCG and L.

obtained with stroboscopy, such as symmetry of motion of the two vocal folds, shape of the glottis during maximum closure, and the open quotient mentioned above.

MCQ will be shown to be related to the distance between the vocal processes (that is, the level of adduction). If adduction is greater, that is, if the vocal process gap is smaller, the vocal fold tissues will come together more completely in the A-P direction, and MCQ will be larger. Also, MCQ will be shown to be related to the maximum lateral vocal fold excursion during a vibratory period. Larger excursions may result in greater medial compression and an increase in the relative A-P contact.

A similar measure of the A-P portion of glottal closure was presented by Södersten, Hertegård and Hammarberg (2), but in their study, subjects viewed recordings and psychophysically estimated the degree of A-P closure on a vertical line. Their "percentage glottal closure" (GLC) significantly increased with loudness for their female subjects (male subjects were not studied). Their range of GLC was 43.5% to 100%. Soft utterances averaged 71.4%, Normal 87.2%, and Loud 94.0%. This suggests that a wide range of MCQ values is expected for healthy subjects.

Methods

Five excised canine larynxes were used in this study. Excised larynxes were chosen in order to control the level of adduction and measure dynamic aspects of glottal configuration. Three dogs were female and two were male, ranging in weight from 25 to 27 kg. They all had normal-appearing laryngeal tissue. The larynxes were acquired from the University of Iowa Hospitals and Clinics and were kept in saline solution prior to the experiments. The larynxes were trimmed of nonessential tissue and the false vocal folds were removed to completely expose the true vocal folds. Air was supplied to the vocal folds by means of a Concha-Therm III Servo Control Heater unit (RCI Laboratories). The air had approximately 100% humidity and a temperature of 37° C. The tracheal portion of the tissue was mounted on a pseudo-tracheal rigid tube (i.d. of 17.5 mm, o.d. of 20.0 mm). A camera was situated above the larynx for video recording. Adduction controls were created by connecting three-pronged holders to the lateral portions of the arytenoid cartilages and adjusting their position with a micrometer travers. When this was not used, shims were placed between the arytenoid cartilages. Glottal length was changed by adjusting a suture line attached to the thyroid notch and tied to a metal bar anterior to the thryoid cartilage. Mean subglottal pressure via a subglottal pressure tap was obtained by viewing a wall mounted u-tube manometer (Dwyer 1230-8). Mean flow was measured via an in-line flowmeter (Gilmont Rotometer Model J 197) upstream of the Concha-Therm unit. For a more complete description of the experimental set up, refer to Alipour, Scherer and Knowles (3).

Figure 1 illustrates the vocal process gap, V, the distance between the anterior portion of the two vocal processes; half of the maximum glottal width, W, the maximum distance between the two vocal folds at the time of maximum vibratory amplitude as seen using stroboscopy; and the membranous glottal length, L.

There were three to six different cases of preset adduction for each larynx. For each case, subglottal pressure was raised and then lowered in six to nine steps. The distance measures L, V, and W were obtained by pausing the video tape and taking measures on the monitor for each subglottal pressure condition. A ruler segment was placed on the tissue and video-recorded prior to phonation to allow conversion of measurements to absolute distance. Measurement resolution on the monitor screen was 0.1 cm, and the screen magnification was 15 to 1. The estimated accuracy in real measurement was therefore approximately +/- 0.006 cm.

A model of the relationship among MCQ and the vocal process gap, glottal width, and glottal length would help to consider how MCQ may be affected by changes in these production characteristics of a patient or subject. A rather simple model can be used, and is discussed next.

Model

MCQ can be modeled with a relatively straightforward geometric rendering of vocal fold contact. It is first noted that when the two vocal folds come into contact, the contact is normally along the midline of the glottis. For modeling purposes, let the vocal fold contour take the shape of an ellipse for both outward and inward motion. Figure 2 shows an elliptical shape of the vocal fold contour intersecting a line. The intersection point is essentially where the two vocal folds would meet given the conditions of amplitude and vocal process gap as shown. It is assumed in this model

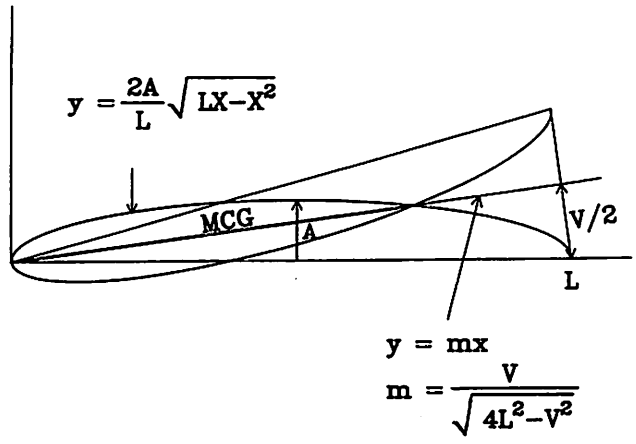


Figure 2. Geometric model for predicting MCQ. The vocal fold contour is modeled elliptically. Contact between the two vocal folds in the A-P direction is modeled linearly along the glottal midline. A is the amplitude of the vocal fold motion, V/2 is half of the vocal process gap, and L is the membranous glottal length.

that an approximation of the degree of A-P contact between the two vocal folds can be modeled by an "overlapping" of the two vocal folds, rather than the more realistic A-P deformation at contact due to tissue incompressablility. The object of the model is to solve the intersection of the elliptical and linear equations as a function of amplitude A, membranous glottal length L, and vocal process gap V. The solution leads to Eqn. 1 which shows MCQ as a function of the three variables just mentioned.

$$MCQ = \frac{8A^2L(4L^2-V^2)^{0.5}}{V^2L^2+4A^2(4L^2-V^2)} \quad (1)$$

where A is the amplitude of motion of the glottis, L is the membranous glottal length, and V is the vocal process gap. For simplicity and practicality, the amplitude of motion in Eqn. 1 can be replaced by one-half of the maximum glottal width (as shown in Figure 1), which leads to Eqn 2, viz.,

$$MCQ = \frac{2W^2L(4L^2-V^2)^{0.5}}{V^2L^2+W^2(4L^2-V^2)} \quad (2)$$

The sensitivity of the theoretical MCQ to changes in maximum glottal width, glottal length, and vocal process gap will now be explored. Figure 3 shows the sensitivity of MCQ to changes in the vocal process gap for three values of the maximum glottal width. The vocal process gap ranges between 0 and 0.7 cm (which encompasses the realistic range), and the parametric vibratory widths are .1, .2, and .3 cm (values that are not atypical). MCQ is shown to be highly sensitive to the value of the vocal process gap, especially for smaller amplitudes of motion. For example, for W = 0.1 cm, as the vocal process gap becomes larger than 0, MCQ quickly drops. This would be explained by the reasoning that for small amplitudes, as the vocal processes separate, the vocal folds will not be able to come into contact, leading to smaller values of MCQ.

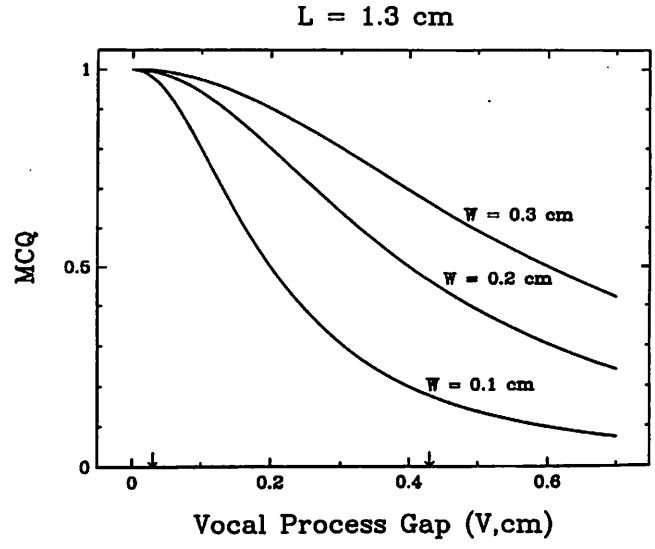


Figure 3. Sensitivity of MCQ to changes in vocal process gap. Eqn. 2 was used. The membranous glottal length was held constant at 1.3 cm. Three values of the maximum glottal width were used parametrically. The small arrows on the abscissa indicate the range of the vocal process gap used in the excised canine larynx studies of this report.

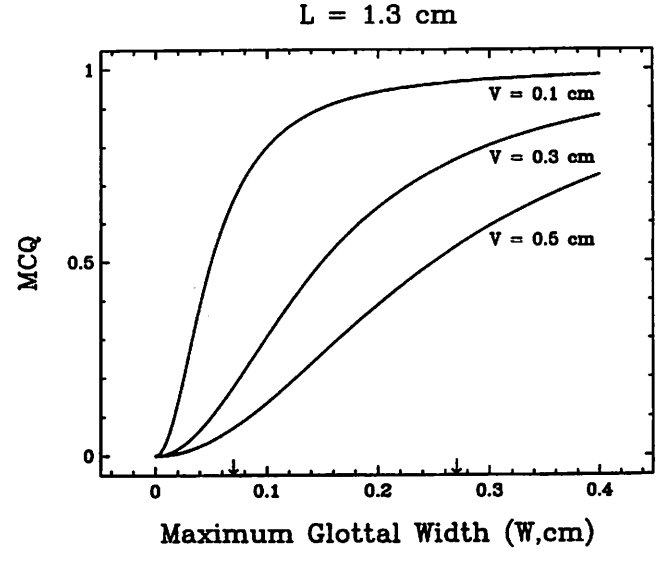


Figure 4. Sensitivity of MCQ to changes in the maximum glottal width. Eqn. 2 was used. The membranous glottal length was held constant at 1.3 cm. Three values of the vocal process gap were used parametrically. The arrows on the abscissa indicate the range of the maximum glottal width used in the excised canine larynx studies of this report.

Figure 4 shows the sensitivity of MCQ to the maximum glottal width for three values of the vocal process gap. MCQ is theoretically highly sensitive to changes in the maximum excursion of the vocal folds, especially when the vocal process gap is narrow. That is, if the gap were narrow, then relatively small increments of the maximum excursion

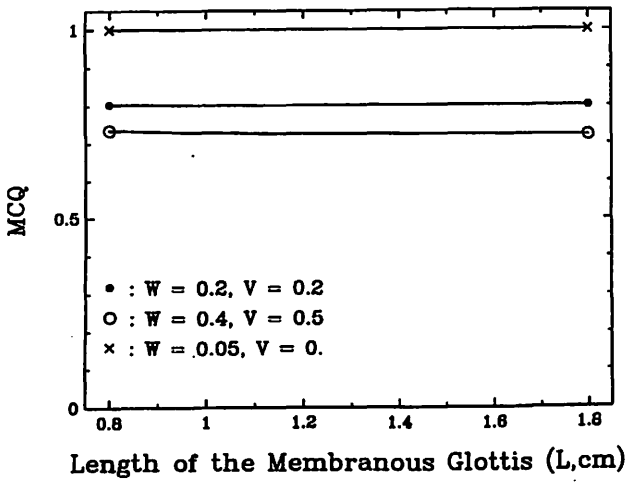


Figure 5. Sensitivity of MCQ to changes in the membranous glottal length, L. Eqn. 2 was used. The figure indicates that MCQ is insensitive to changes in L for essentially any value of vocal process gap and maximum glottal width.

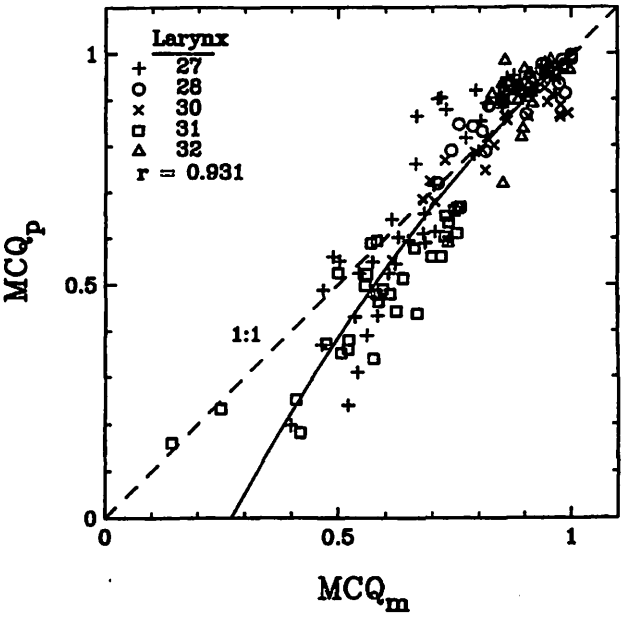


Figure 6. Predicted vs. measured values of MCQ. Each symbol refers to the data for a different canine larynx. The predicted MCQ values are based on Eqn. 3.

of the vocal folds will result in significant increases in MCQ because the folds come together more completely in the A-P direction.

MCQ is, however, theoretically not sensitive to the length of the membranous glottis (or equivalently, the length of the vocal folds), as shown in Figure 5. For a variety of values for W and V, MCQ is nearly constant as L is changed (changing L effectively changes both the numerator and denominator of Eqn. 2 by the same amount). This finding allows the replacement of L in Eqn. 2 with a value of 1 cm in order to simplify the equation, giving Eqn. 3.

$$MCQ = \frac{2W^2(4-V^2)^{0.5}}{V^2+W^2(4-V^2)} \qquad (3)$$

Results

Figure 6 shows how well the predicted MCQ values (using Eqn. 3) compare with the actual measurements of MCQ for the canine data. The prediction of MCQ is best for values above 0.8. Most of the data for the MCQ values are predicted to be lower than the measured values for values less than approximately 0.7. Figure 6 shows that the measured MCQ values are well predicted best for 3 of the 5 larynxes (larynx 28, 30 and 32). For larynx 27, some data fall above and some fall below the 1:1 line. For larynx 31, almost all the data are below the 1:1 line, indicating that the prediction values were too small. The overall correlation between the measured and predicted MCQ values was 0.931.

Figure 7 shows how well the excised canine data match the predictions from Eqn. 3 by displaying MCQ as a function of maximum width W and the vocal process gap V (as was shown in Figure 4). The parameter V has six values (0.05, 0.075, 0.10, 0.15, 0.20, and 0.25 cm), and the data displayed were taken from the data corpus such that they varied from the six nominal values by no more than +/- 0.01 cm. The data indicate the same general trend as the theory of Eqn. 3. The overall absolute difference between the theoretical and measured MCQ values of Figure 7 was 7.4% (sd = 8.9%).

Discussion

For this study using normal canine larynges, glottal closure took place without open gaps along the closed glottis during maximum glottal closure. For abnormal cases, as in pathological conditions of local vocal fold swelling or vocal fold bowing, the effective membranous contact glottis can also be obtained. The maximum length of contact between the vocal folds would be measured by adding the separate contact lengths of the membranous glottis. Dividing this number by the membranous glottal length would constitute a practical measure of MCQ, which would presumably change with intervention procedures. The object of many intervention procedures would be to create MCQ values closer to one. For example, medialization procedures for vocal fold paralysis should result in closer vocal folds and larger, more normal, values of MCQ. These procedures and the resultant increase in MCQ may correspond to decreased levels of DC flows and decreased levels of breathiness.

It is known that the maximum amplitude of glottal vibration is dependent upon the value of subglottal pressure or glottal flow used (4,5). Titze and Durham (6) quantified

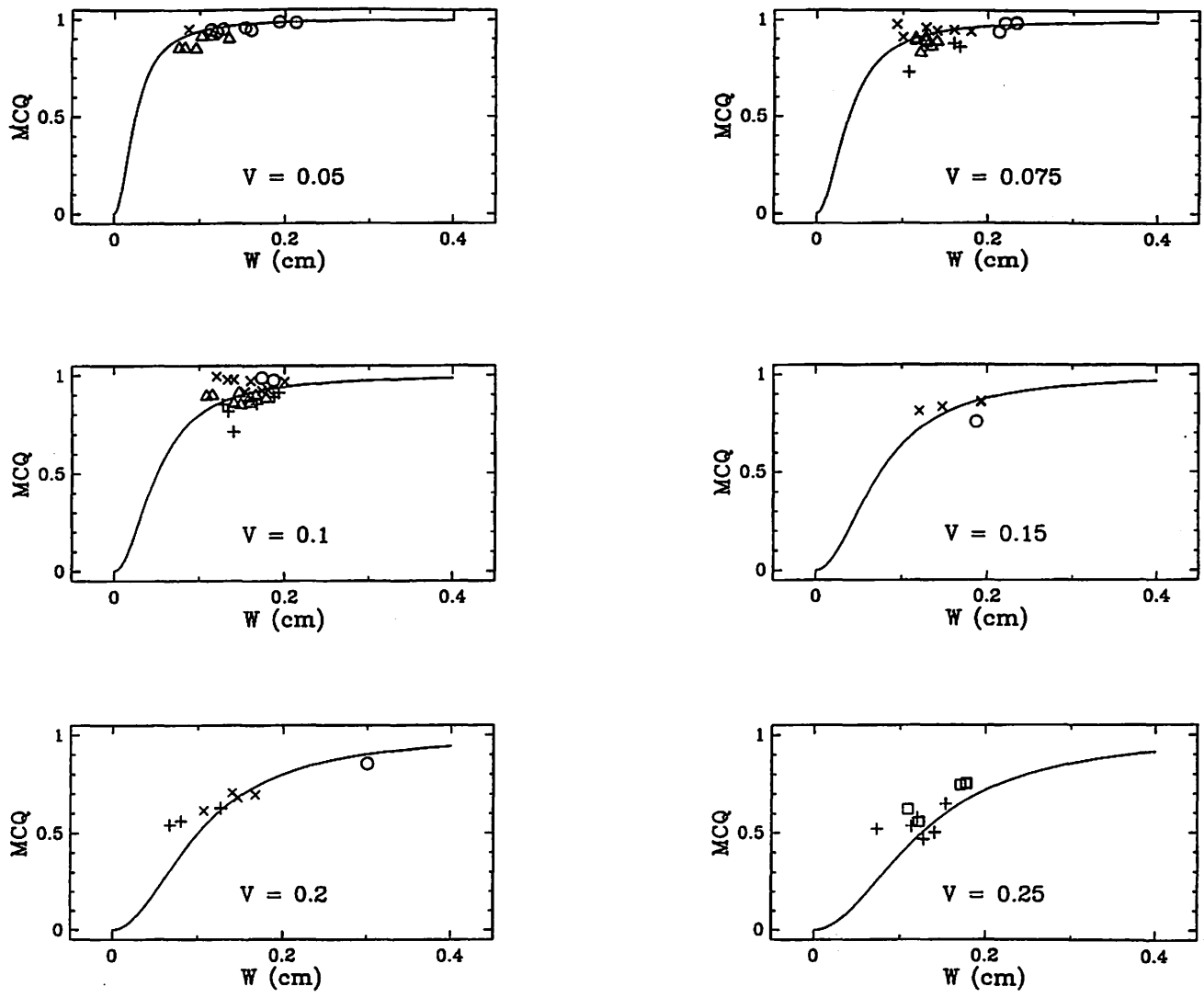


Figure 7. MCQ values from the excised data corpus compared to the MCQ predictions. Each figure shows the predicted MCQ curve over a range of maximum glottal width from zero to 0.4 cm. The figures differ by the indicated vocal process gap value (0.05, 0.075, 0.1, 0.15, 0.2, and 0.25 cm). The excised larynx data were matched to each vocal process gap value within the range of +/- 0.01 cm.

the relationship between maximum glottal width and subglottal pressure as a function of vocal fold strain for a number of excised canine larynxes. Their findings indicated that the maximum width was related to the square root of subglottal pressure. We examined the data of the present experiment for the relationship between maximum glottal width, subglottal pressure, and length of the vocal folds. We found a significant relationship between subglottal pressure and maximum glottal width, but the relationship did not differentiate across the values of glottal length. We therefore obtained the relationship between the maximum glottal width and the square root of the subglottal pressure,

$$W = k_i \times P_s^{0.5} \qquad (4)$$

using individual k_i coefficients for each of the five larynxes. Figure 8 shows the fit between the expression and data for one of the larynxes.

With the replacement of the subglottal pressure expression for the maximum glottal width, Eqn. 3 becomes Eqn. 5, where MCQ is shown as a function of subglottal pressure and the vocal process gap,

$$MCQ = \frac{2k_i^2 P_s (4 - V^2)^{0.5}}{V^2 + k_i^2 P_s (4 - V^2)} \qquad (5)$$

and each k_i value corresponds to the data for each experimental larynx. Comparing the predictions of Eqn. 5 to the measured MCQ data results in a figure similar to Figure 6.

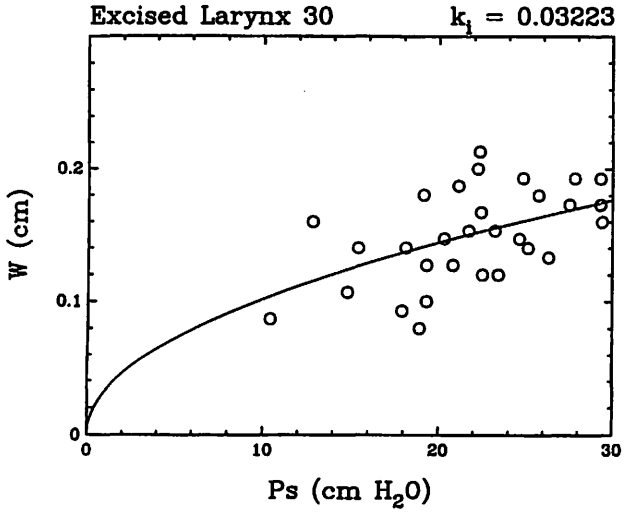


Figure 8. Relation between maximum glottal width and subglottal pressure. The data shown are for excised larynx 30 for which k_i = 0.03223 in Eqn. 4.

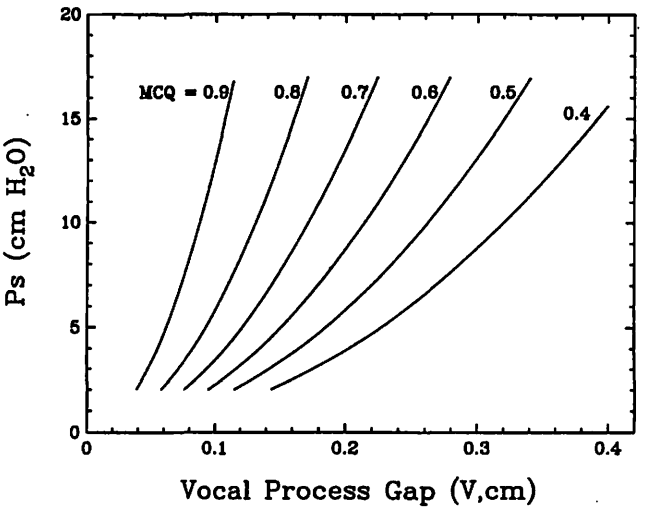


Figure 9. Tradeoff of adduction and subglottal pressure for maintaining MCQ values. Each curve represents a particular value of MCQ, as derived from Eqn. 5. Each line indicates that maintaining MCQ values for an increase of abduction requires a nonlinear increase in subglottal pressure.

The model of Eqn. 5, MCQ as a function of subglottal pressure and vocal process gap, is helpful in showing how MCQ might be held constant for a variety of values of pressure and adduction. Figure 9 shows this relation. Essentially, in order to keep MCQ constant, if adduction decreases (larger vocal process gap), which would lower MCQ, subglottal pressure would have to increase to raise MCQ. For adequate voicing behavior, MCQ should not be too low (certainly not beyond the configuration corresponding to the phonatory adductory threshold, see Scherer (7)). Figure 9 can be used to suggest the adductory and subglottal pressure demands for the maintenance of specific levels of MCQ, or the control of changes in MCQ.

As Figures 6 and 7 showed, the prediction of MCQ did not give perfect matches to the measured MCQ values. Tissue incompressibility, rather than the tissue overlap provided by the model, should give larger values of measured MCQ than predicted, as was the case for lower values of MCQ. The elliptical model may correspond best to the dynamics of the upper edge of the vocal fold, whereas in reality the vocal folds come together first at a position inferior to the upper edge, thus leading to larger MCQ values than predicted. The model would be improved by incorporating a three-dimensional closure. Despite some limitations with the model, the essentials of the model appear to be helpful in conceptualizing the dependence that MCQ has on adduction, pressure, and maximum glottal width.

A practical limitation of this new parameter is that scoping procedures do not always allow the visualization of the necessary laryngeal tissue. The anterior commissure may not be viewable, and the location of the vocal processes may be obscured by the corniculate cartilages. The MCQ parameter requires an adequate procedure to obtain both the membranous contact glottis and the membranous glottal length.

MCQ can be thought of as a measure of glottal incompetence because the measure is related to the vocal process gap. As such, it supplements visual ratings of adduction (e.g., ref. 1, 8), glottal configuration designations (e.g., 9, 10, 11, 12, 13), and electroglottographic measures of adduction and closed quotient (e.g., 1, 8, 14, 15). In addition, it is dynamic in that it is obtained using the maximum A-P contact length during a cycle, and thus should be related to the glottal volume velocity waveform and the corresponding acoustics - topic areas for further research.

Conclusions

This study introduces a new parameter of phonation called the Membranous Contact Quotient, MCQ, defined as the ratio of the membranous contact glottis at maximum closure during a phonatory cycle divided by the membranous glottal length between the anterior commissure and the vocal processes. MCQ should be a useful diagnostic parameter as one indication of the degree of dynamic glottal competency. It is a fairly easy parameter to obtain when the stroboscopic view of the larynx shows the entire length of the membranous glottis.

An elliptical model of the glottal contour during vocal fold motion was used to create a theoretical relation between MCQ and the vocal process gap, maximum glottal width, and the membranous glottal length. MCQ was theoretically highly sensitive to both adduction and maximum glottal width, especially at low values for each, but not

sensitive to changes in the membranous glottal length. Also it was theoretically shown that the maintenance of a particular level of MCQ was possible through coordinated control of subglottal pressure and adduction. The model predicted MCQ values well for three of the five canine excised larynxes and for higher MCQ values. The model would be improved, but more complicated, by taking tissue incompressibility and the three-dimensional glottis into account.

Describing MCQ empirically and theoretically for human subjects and patients should help to further explore the importance of MCQ as a dynamic glottal competency parameter related to adduction, maximum glottal excursion, subglottal pressure, and abnormal vocal fold tissue conditions, as well as to other important signals such as the glottal volume velocity and the electroglottograph waveform.

The relation between MCQ and vocal process gap, maximum cyclic glottal width, subglottal pressure, and levels of breathiness should be explored for humans. Also, knowing the values of MCQ for speakers judged to use near optimal phonation may provide important clinical targets for the biomechanical behavior of the larynx.

Acknowledgements

We would like to thank Paula Saraceno for secretarial assistance. This research has been supported by grants DC 00831 and P60 DC00976 from the National Institutes of Health.

References

1. Scherer RC, Vail VJ and Rockwell B. Examination of the laryngeal adduction measure EGGW. In: Bell-Berti F, Raphael LJ, eds. *Producing speech: Contemporary issues: for Katherine Safford Harris*. American Institute of Physics, 1995:269-290.

2. Södersten M, Hertegård S, Hammarberg B. Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *J Voice*, 1995;9(2):182-197.

3. Alipour F, Scherer R and Knowles J. Velocity distributions in glottal models. *J Voice* (in press).

4. Berke GS, Hanson DG, Gerratt BR, Trapp TK, Macagba C and Natividad, M. The effect of air flow and medial adductory compression on vocal efficiency and glottal vibration. *Otol - Head Neck Surg* 1990;102(3):212-218.

5. Sercarz JA, Berke GS, Bielamowica S, Kreiman J, Ye M & Green DC. Changes in glottal area associated with increasing airflow. *Ann Otol Rhinol Laryngol* 1994;103(2):139-144.

6. Titze IR, Durham PL. Passive mechanisms influencing fundamental frequency control. In: Baer T, Sasaki C, Harris K, eds. *Laryngeal function in phonation and respiration.* Boston: A College-Hill Publication, 1987:304-319.

7. Scherer RC. Laryngeal function during phonation. In: Rubin JS, Korovin G, Sataloff RT, Gould WJ, eds. *Diagnosis and treatment of voice disorders*. New York: Igaku-Shoin Medical Publishers, Inc., 1995:86-104.

8. Peterson KL, Verdonini-Marston K, Barmeier JM, Hoffman HT. Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. *Ann Otol Rhinol Laryngol* 1994;103:335-346.

9. Hirano M, Bless DM. *Videostroboscopic examination of the larynx*. San Diego: Singular Publshing Group, Inc., 1993.

10. Linville SE. Glottal gap configurations in two age groups of women. *J Speech Hear Res*, 1992;35:1209-1215.

11. Biever DM, Bless DM. Vibratory characteristics of the vocal folds in young adult and geriatric women. *J Voice*, 1989;3(2):120-131.

12. Linville SE, Skarin BD, Fornatto E. The interrelationship of measures related to vocal function, speech rate, and laryngel appearance in elderly women. *J Speech Hear Res*, 1989;32:323-330.

13. Sodersten M, Lindestad P.-A. A comparison of vocal fold closure in rigid telescopic and flexible fiberoptic laryngostroboscopy. *Acta Otolaryngol (Stockh)*, 1992;112:144-150.

14. Rothenberg M, Mahshie J. Monitoring vocal fold abduction through vocal fold contact area. *J speech Hear Res*, 1988;31:338-351.

15. Scherer R, Gould WJ, Titze I, Meyers A Sataloff R. Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. *J Voice*, 1988;2:230244.

NCVS Status and Progress Report - 9
April 1996, 71-78

Increased Stability of Airflow in Spasmodic Dysphonia and/or Vocal Tremor Following Botulinum Toxin Injection

Eileen Finnegan, M.A.
Erich Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Department of Otolaryngology, The University of Iowa Hospitals and Clinics
James Gordon, M.D.
Department of Otolaryngology, The University of Iowa Hospitals and Clinics
Julie Barkmeier, Ph.D.
National Institute on Deafness and Other Communication Disorders, Bethesda, Maryland
Henry Hoffman, M.D.
Department of Otolaryngology, The University of Iowa Hospitals and Clinics

Abstract

The purpose of this study was to determine if stability of airflow, as well as mean airflow, increased following botulinum toxin injection to laryngeal muscles in persons with spasmodic dysphonia &/or vocal tremor (SD/VT). Aerodynamic data were collected from 5 subjects before, and at 2 weeks, 4 weeks, and 8 weeks after treatment via each of two injection protocols, one involving injection of intrinsic laryngeal muscles only, the other involving injection of extrinsic and intrinsic laryngeal muscles. Measures of mean airflow and coefficient of variation (COV) of airflow during phonation were obtained. A decrease in the COV of airflow would indicate increased stability of phonatory airflow. Prior to injection, subjects with SD/VT exhibited mean airflows that were similar to controls, and COV of airflow that ranged from similar to controls to substantially elevated Following botulinum injection, mean airflow was typically increased and COV of airflow decreased. This finding suggests there is a change in the type, as well as the level, of activity in the muscles of speech production following botulinum toxin injection. Increase in stability could be due to increased stability of the laryngeal system and possibly of the respiratory system as well.

Researchers have used aerodynamic measures as a means of studying persons with spasmodic dysphonia (SD) to better understand the pathophysiology of the disorder and the effects of treatment. Zwirner, Murry, Swenson, and Woodson (1992) suggest "in SD, a condition associated with excessive vocal fold closure, a low airflow suggesting high glottal resistance is expected" (p. 404). However, investigations of mean airflow in persons with SD have provided somewhat conflicting results (see Table 1). Some researchers found mean airflow was reduced in subjects with SD in comparison to controls (Briant, Blair, Cole, and Singer, 1983; Zwirner et al., 1992; Woo, Colton, Casper, & Brewer, 1992). Others report mean airflow within normal range (Hirano, Koike, & von Leden, 1968), or generally within normal limits, aside from a small number of subjects with low airflow (Davis, Boone, Carroll, Darveniza, & Harrison, 1988). Elevated mean airflow at one week following injection of botulinum toxin has been reported, with near-normal values one month after the injection (Zwirner et al., 1992; Woo et al., 1992). Zwirner et al. (1992) attributed the increased airflow "to the paralytic state of one vocal fold" (p. 401). Woo et al. (1992) state "[botulinum toxin] injection results in partial chemodenervation of the injected muscle [the TA]", so "the vocal folds continue to abduct and adduct" (p. 345), and that as a result, at 4 weeks post-injection "nearly normal phonatory airflow values were achieved" (p. 349).

Table 1.
Summary of Findings of Previous Studies of Mean Airflow in Persons with SD

Study	Task	Mean Airflow (LPS)		Conclusion
		Spasmodic Dysphonia	Controls	
Briant et al., 1983	mean of sustained /a/	.029	.116 to .160	SD mean flow lower than controls
Hirano et al., 1968	mean of maximum phonation	.063 to .186	.043 to .222	SD mean flow within same range as controls
Davis et al., 1988	sustained /a/ analyzed 2 second segment, 2 seconds into phonation	females: .091 +/- .015 males: .077 +/- .023	*	Most SD subjects with in same range as controls, some had low flows
Woo et al., 1992	sustained /i/	pre-inj: .072 4 week post: .138		low flows pre-tx normal flows post-tx
Zwirner et al., 1992	sustained /a/ first 1 second and 1 second from midpoint	pre inj: .107 +/- .046 1 week post unilat inj of TA (5-30 units): .358 +/- .169 1 month post: .231 +/- .082	.180 +/- .060	SD subjects had low flows

* compared to Hirano et al. 1968 control subjects

Variable airflow, as indicated by a high coefficient of variation of airflow, was observed by Davis et al. (1988) in 15 of their 20 subjects with SD. Other investigators have also reported elevated COV of airflow in some subjects with SD (Finnegan, Luschei, Barkmeier, & Hoffman, 1996). Davis et al. also noted that one subject exhibited decreased COV of airflow with improvement in voicing, and another subject showed increased COV of airflow with worsening of voice symptoms. These were anecdotal observations, because change in airflow following treatment was not the focus of the Davis et al. study. An increase in stability of airflow, as well as mean airflow, following botulinum toxin injection is of interest because it suggests that there is a change in the type, as well as the level, of activity in the injected muscle. As noted above, change in mean airflow has been attributed to chemical denervation at the neuromuscular junction. Increased stability of airflow could reflect an alteration in the neural input to the laryngeal muscle (i.e., an increase in stability of the neural signal from the laryngeal motor neuron pool).

The purpose of our study was to systematically investigate change in stability of airflow in persons with spasmodic dysphonia &/or vocal tremor (SD/VT) following treatment with two separate approaches to injection of botulinum toxin. We measured changes in airflow following injection to both intrinsic and extrinsic laryngeal muscles, as well as changes in airflow following injection of intrinsic muscles only.[1] We hypothesized that stability of airflow would be increased based on previous reports of reduction in hyperfunction of intrinsic laryngeal muscles, and to a lesser extent, tremor and extrinsic laryngeal muscle hyperfunction following botulinum toxin injection (Woodson, Zwirner, Murry, & Swenson, 1991; Zwirner et al., 1992). The coefficient of variation (COV) of airflow was used to provide a means of quantifying stability of airflow and tracking change over time.

Method

Subjects

Five subjects (1 male, 4 female, mean age: 64.6 yr.) were included in the study, one subject diagnosed with adductor spasmodic dysphonia (S1), and four with adductor spasmodic dysphonia and vocal tremor (S2-S5). The diagnosis was based on a consensus between two speech pathologists (E.F. and J.B.) and an otolaryngologist (H.H.) subsequent to review of videoendoscopic, laryngeal electromyography, and medical and social history information. In addition, all subjects were assessed by a neurologist to

[1] Comparison of the treatment protocols was not the focus of this study and will be reproted separately at another time.

rule out other neurologic diseases. The subjects participated in 8 one-hour sessions of speech therapy to identify and reduce any compensatory patterns and to allow opportunity for further diagnostic therapy prior to a final decision regarding inclusion in the study. Previously reported aerodynamic data (Finnegan et al., 1996) from ten control subjects with no history of voice disorders (1 male, 9 females, mean age 45.7 years) were included to provide comparison data. The signals collected from the control subjects were reprocessed so they would be comparable to the subject data in this study.

Treatment was provided following a double-blinded, randomized, crossover design. Subjects and speech pathologists were blinded to treatment. Subjects were randomly assigned to an initial treatment protocol. Protocol A required injection of 2.5 units of botulinum toxin in the thyroarytenoid muscles (TAs) with injection of saline to the sternothyroid muscles (STs) and thyrohyoid muscles (THs) bilaterally for a total of 5.0 units of botulinum toxin. Protocol B consisted of injection of 2.5 units of botulinum toxin in the TAs, and 7.5 units to the STs and THs bilaterally for a total of 35 units of botulinum toxin. Injections were accomplished percutaneously using electromyographic (EMG) needle guidance. Needle placement was verified by the presence of increased EMG activity during phonatory and positional maneuvers (Hirano & Ohala, 1969). Following a return of dysphonia the subject received the second injection protocol.

Procedure

Simultaneous recordings of intraoral pressure and airflow were obtained using the procedures delineated by Smitheran and Hixon (1981). Airflow measurements were made by placing an anesthesia mask over the subject's mouth and nose to direct airflow through a pneumotachometer (Hans Rudolph model 4719) connected to a differential pressure transducer (Honeywell Microswitch 162PC01D). The airflow signal was amplified and recorded on a digital instrumentation tape recorder (Sony PC108-M). Oral pressure was transduced by a 7-inch polyethylene tube with a 1.67 mm inner diameter (Intra-medic PE #240) that was inserted through a small tight-fitting hole in the front wall of the anesthesia mask. The tube was positioned in the subject's oral cavity to sample intraoral pressure without interfering with lip and/or tongue movement. The oral pressure signal was recorded on another channel on the digital recorder. The sampling rate on the tape recorder was 10 kHz per channel. Aerodynamic data were collected prior to botulinum toxin injection and at 2 weeks, 4 weeks, and 8 weeks post-injection for each treatment protocol.

Calibration of the airflow and pressure instrumentation was performed on each day of data collection. Two flow levels were generated and monitored (0 and 1000 cc/s) using an airtank and a rotometer. Air pressure was calibrated at two levels (0 and 10 cm H_2O) using a water-filled U-tube manometer. Signals from the pneumotachometer and differential pressure transducer were adjusted so that 1000 cc/s of airflow equaled 1 Volt. These signals were recorded onto the digital recorder and used to calibrate experimental signal measures.

Subjects were instructed to repeat the syllable /pi/ in a continuous manner, at comfortable pitch and loudness, as modeled by the investigator at 1.5 syllables/second using a metronome. Subjects were directed to place the mask firmly against their face at the start of each trial and to maintain that placement during production of all tokens in that trial. Subjects were provided with practice and recordings were repeated until they produced three series of repetitions at an appropriate rate, maintaining adequate face-mask seal. The data were low-pass filtered at 300 Hz, then digitized at 3 kHz per channel onto the software system used for data analysis (CODAS, Dataq Inc., Akron, OH). A moving average with a 33 msec. window was applied twice consecutively, first to the original airflow signal and then to the smoothed signal.

Measures

Measures of mean airflow and COV of airflow were obtained from data collected during the syllable repetition task. Figure 1 illustrates the method for determining these measures: (a) onset and offset of the vowel were manually selected by marking the period from 40 msec following the sharp drop in oral pressure to 40 msec prior to the sharp increase in oral pressure (i.e., from release of the plosive to lip closure for the /p/), (b) mean and standard deviation of airflow during the vowel were calculated by CODAS, and (c) the coefficient of variation (COV) for each token was calculated by dividing standard deviation by mean flow.

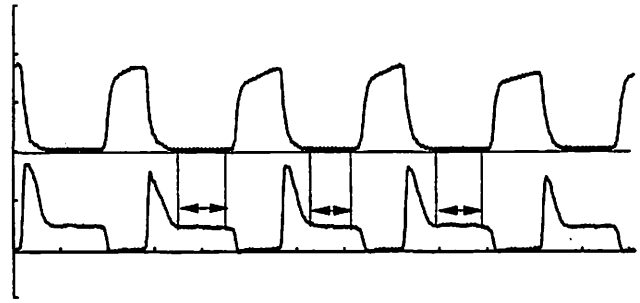


Figure 1. Example of measures of mean and COV of airflow (see text for details). Time is displayed on the horizontal axis (200 msec per mark). On the vertical axis, amplitude of oral pressure (5.0 cm H_2O per mark) is indicated by the top trace and amplitude of airflow (.25 LPS per mark) by the bottom trace.

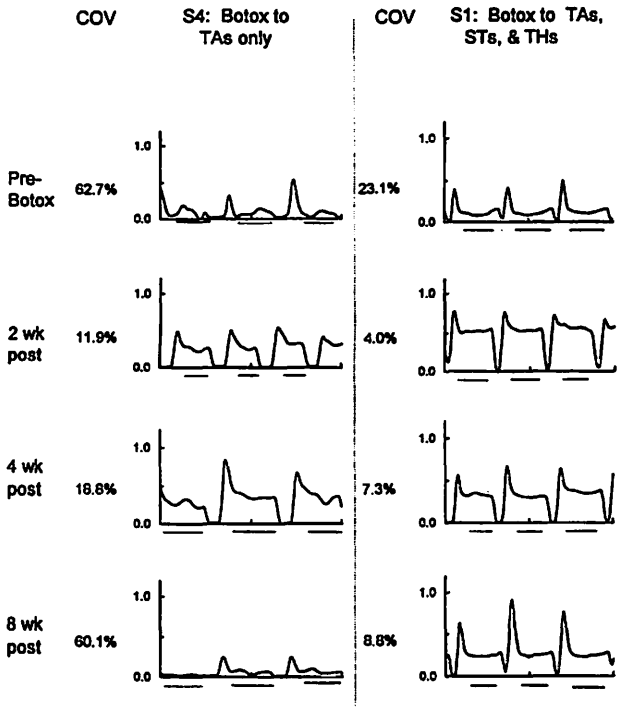


Figure 2. Examples of increased mean airflow and stability of airflow following botulinum toxin injection in subject S4 (injection to the TAs only) and for subject S1 (injection to the TAs, STs, and THs). Time is displayed on the horizontal axis (1 sec per mark); amplitude of airflow is on the vertical axis (.500 LPS per mark).

Results

Samples of airflow data from two subjects are presented in Figure 2. The data samples are organized in chronological order from top to bottom. The results of injection of botulinum toxin to the TAs only of subject 4 are on the left. The results of injection of botulinum toxin to the TAs, STs, and THs of subject 1 are on the right. The bars beneath each token indicate approximately the portion of data over which each measure of mean airflow and COV of airflow was calculated. The mean COV of airflow, listed to the left side of each data sample, is an average of the COVs calculated for each of five central tokens from the second set of repetitions of /pi/. The airflow tracings displayed are from tokens 3-5 of the second trial of repetition of /pi/.

Mean Airflow

Prior to treatment, mean airflows for subjects with SD/VT were similar to the range of airflow values exhibited by the control subjects (Figure 3). The mean airflows for the 5 subjects with SD/VT and the 10 control subjects are presented in Tables 3 and 4. Following both injection protocols, the group mean airflow (Table 2) was increased at 2 weeks post-injection, with gradual decrease at 4 and 8 weeks post-treatment. This pattern of change in mean airflow is seen in the sample data presented in Figure 2. As seen in Figure 4, the individual mean airflows followed this pattern of an initial increase at 2 weeks in 8 of 10 cases and a subsequent decrease at 8 weeks in 7 of those 8 cases. In one

Table 2.
Summary of Mean Airflow Measures for Each Subject Before and After Each Botulinum Toxin Injection.
Means are Based on Measures from 5 Central Tokens of the Second Trial of /pi/.

Mean Airflow (LPS)

	Treatment A: Botox to TAs only				Treatment B: Botox to TAs, STs, &THs			
	Pre B/S	2wk	4wk	8wk	Pre B/B	2wk	4 wk	8 wk
S	.104 (.010)	.481 (.026)		.175 (.024)	.121 (.021)	.557 (.033)	.326 (.027)	.248 (.020)
S2	.029 (.020)	.354 (.040)	.471 (.157)	.290 (.075)	.161 (.053)	.616 (.048)	.293 (.072)	.100 (.020)
S3	.282 (.064)	.477 (.041)	.380 (.028)	.209 (.040)	.245 (.039)	.391(.047)	.276(.053)	.340(.083)
S4	.076 (.023)	.396 (.147)	.306 (.077)	.045 (.037)	.087 (.032)	.025 (.011)	.044 (.006)	.087(.008)
S5	.052 (.016)	.054 (.024)	.046 (.011)	*.201 (.030)	.067 (.023)	.139 (.059)	.014 (.008)	.128(.033)
Mean	.109 (.101)	.352 (.175)	.301 (.183)	.180 (.089)	.136 (.070)	.346 (.257)	.191 (.149)	.181(.109)

* Following booster injection of 2.5 units bilaterally to the TAs (5.0 units total), 14 days prior to this measurement. This measure is not included in the group mean.

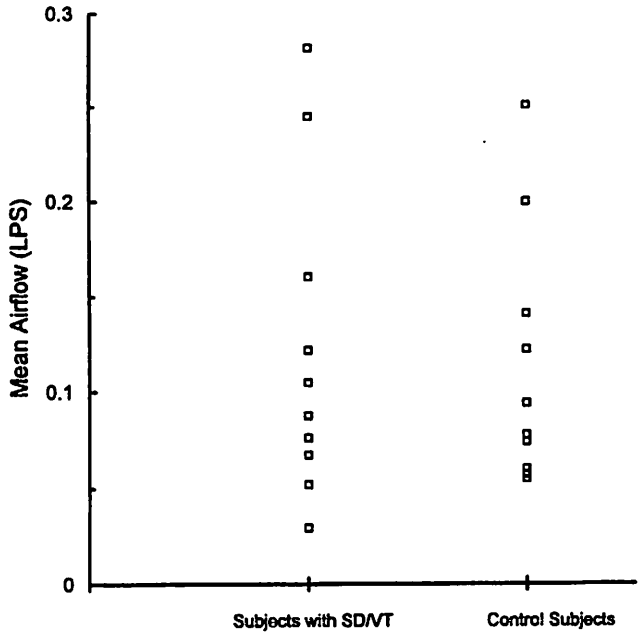


Figure 3. Mean COV of airflow for subjects with SD/VT prior to each treatment (5 subjects x 2 treatments = 10 data points) in comparison to 10 control subjects.

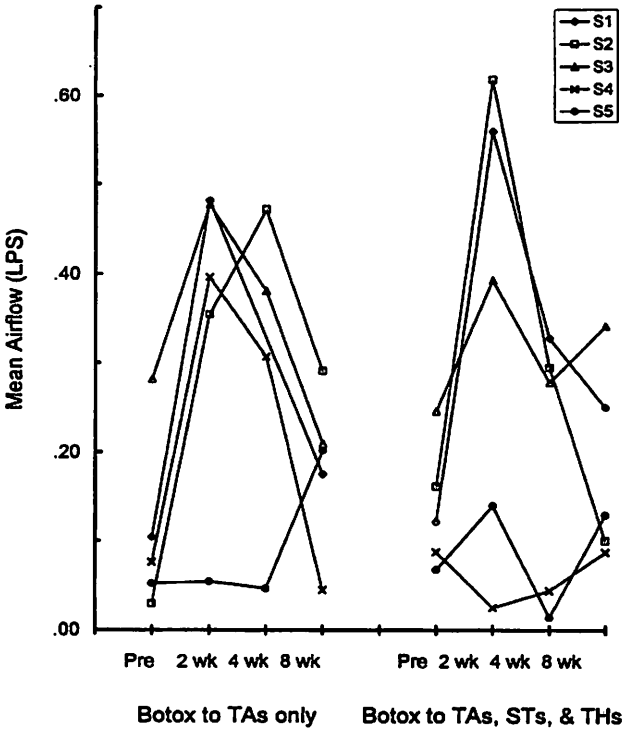


Figure 4. Change in COV of airflow following botulinum toxin injection. Note: S5 receieved a booster injection of 2.5 units bilaterally to the TAs (5.0 units total), 14 days prior to the 8 week post (TA only) measurement.

Table 3.
Summary of Airflow Measures for 10 Control Subjects. Means and Standard Deviation are Based on Nine Measures, Three Tokens from Each of Three Trials.

Control Subject	COV of Airflow: Mean +/- SD	Mean Airflow (LPS)
C1	8.8% +/- 3.3%	.060 +/- .008
C2	6.9% +/- 3.2%	.074 +/- .006
C3	7.3% +/- 4.7%	.094 +/- .003
C4	11.6% +/- 7.9%	.055 +/- .014
C5	10.2% +/- 2.5%	.077 +/- .008
C6	22.9% +/- 14.7%	.140 +/- .035
C7	22.2% +/- 15.3%	.057 +/- .014
C8	8.5% +/- 2.2	.122 +/- .022
C9	10.7% +/- 6.1%	.200 +/- .007
C10	14.0% +/- 5.4%	.251 +/- .015
Group Mean (SD)	15.7% (8.7%)	.154 (.019)

case, (S5), voicing was not substantially changed following injection to the TAs only, so a booster injection was provided. Injection of an additional 2.5 units of botulinum toxin in the TAs bilaterally resulted in an increase in mean airflow.

COV of Airflow

Prior to treatment the subjects with SD/VT exhibited a wider range of COVs of airflow than the controls. One subject (S1) produced steady airflow during the pre-treatment recordings and had COVs of airflow values which were similar to controls. The other subjects (S2-5) exhibited fluctuating flow pre-treatment, with flow dropping to zero during vowel production in some cases (S2, S4, S5). These subjects (S2, S4, S5) had COVs of airflow that were elevated in comparison to the control subjects (Figure 5; Tables 3 & 4). Following botulinum toxin injection, mean COV of airflow was decreased, indicating increased stability of airflow during phonation, in 8 of 10 cases (Figure 6). Decrease in the COV of airflow persisted for 4 weeks (in 3 of 8 cases) to 8 weeks (in 5 of 8 cases). Although COV of airflow was decreased following botulinum toxin injection, it was not always reduced to within the range exhibited by control subjects. In some cases, airflow became fairly steady post-injection. In others, it was improved but still varying.

Table 4.
Summary of Stability of Airflow Measures for all Subjects.
Each Mean COV of Airflow is Based on Measures from Five Central Tokens of the Second Trial.

Coefficient of variation (COV) of Airflow: Mean and Standard Deviation

Subject	Treatment A: Botox to TAs only				Treatement B: Botox to TAs, STs, &THs			
with SD	Pre B/S	2wk	4wk	8wk	Pre B/B	2wk	4 wk	8 wk
S1	16.5% (2.0%)	3.3% (2.6%)	——	8.6% (3.3%)	23.1% (6.9%)	4.0% (0.8%)	7.3% (3.2%)	8.8% (4.1%)
S2	111.1% (56.4%)	18.2% (9.8%)	22.3% (17.4%)	31.4% (12.0%)	25.3% (9.0%)	6.6% (5.2%)	22.6% (7.6%)	35.5% (16.1%)
S3	17.3% (5.2%)	5.7% (1.9%)	9.1% (3.1%)	25.3% (5.0%)	29.6% (16.5%)	8.5% (3.9%)	10.3% (5.5%)	14.4% (5.2%)
S4	62.7% (14.9%)	11.9% (4.1%)	18.8% (13.8%)	60.1% (30.9%)	67.8% (26.6%)	69.5% (19.8%)	55.4% (14.5%)	44.0% (14.3%)
S5	109.0% (39.9%)	114.4% (20.1%)	102.5% (16.4%)	*22.9% (8.0%)	146.4% (49.5%)	37.3% (24.3%)	79.8% (56.9%)	85.6% (20.6%)
Mean	63.3% (46.6%)	30.7% (47.2%)	38.2% (43.2%)	31.3% (19.0%)	58.5% (52.4%)	25.2% (28.2%)	35.1% (31.4%)	37.7% (30.5%)

* Following booster injection of 2.5 units bilaterally to the TAs (5.0 units total), 14 days prior to this measurement. This measure is not included in the group mean.

Discussion

The mean airflow data reported in our study support previous reports of mean airflow within normal range prior to treatment (Davis et al., 1988; Hirano et al., 1968). Following botulinum toxin injection, group mean airflow was increased, similar to the findings of Zwirner et al. (1992). Our finding that, prior to treatment, the COV of airflow for subjects with SD/VT ranged from 16.5% to 146.4%, was consistent with the range reported by Davis et al. (1988) of 21.5% to 141% . The main finding of our study was that there was an increase in stability of airflow, in addition to the increase in mean airflow, following botulinum toxin injection to the laryngeal muscles. These results support the hypothesis that the type, as well as the level, of muscle activity is effected by injection of botulinum toxin into the laryngeal muscles. One explanation for the change in type of muscle activity is that injection of botulinum toxin to laryngeal muscles effects the neural signal from the motorneurons to the muscles of speech production.

The mechanism of action of botulinum injection is not entirely understood. Investigators have studied the activity of botulinum toxin at the neuromuscular junction (see review by Simpson,1981 & 1992). Using autoradiographic techniques, Black and Dolly (1986) identified the process by which botulinum toxin enters the motor axon terminal. The botulinum toxin, which is injected into the muscle, binds with mobile acceptors (Dolly, Black, Williams, & Melling, 1984). These acceptors tend to be localized to unmyelinated areas of the motor nerve terminal membrane at the neuromuscular junction (Black & Dolly, 1986a). The toxin enters the cell through a process of endocytosis, in which the mobile acceptors in the membrane, called coated pits, invaginate to form vesicles that surround the botulinum toxin and move into the cytoplasm, then fuse with one another to form endosomes (Black & Dolly 1986b). After the initial binding of the toxin, the next step is internalization of the toxin. Simpson, 1992 suggests that it is probable that "a proton pump in the endosome membrane causes pH of the endosome to become acidic. The toxin molecule has a "pH sensor" that detects this fall in pH, and it induces the toxin to undergo a conformational change. This allows the toxin to penetrate the endosome membrane and reach the cytosol". This suggestion is supported by findings that (a) botulinum toxin is capable of forming the ion-conducting channels needed to pass through the endosome membrane, and (b) this ability is pH dependent (Donovan & Middlebrook, 1986). The final step is intercellular poisoning and this is the aspect that is least understood. It is "a prevailing belief that [once the toxin is released into the cytosol, it] catallytically modifies a substrate essential for exocytosis", the process by which the transmitter acytelcholine is released (Simpson, 1992 p. 360). Although the intricacies of mechanism of action for botulinum toxin is not entirely understood, it is clear that one effect of botulinum toxin is the paralysis of muscle fibers associated with effected neuromuscular junctions (Burgen, Dickens, & Zalman, 1949).

However, there is evidence that suggests the effect of botulinum toxin not limited to the neuromuscular junc-

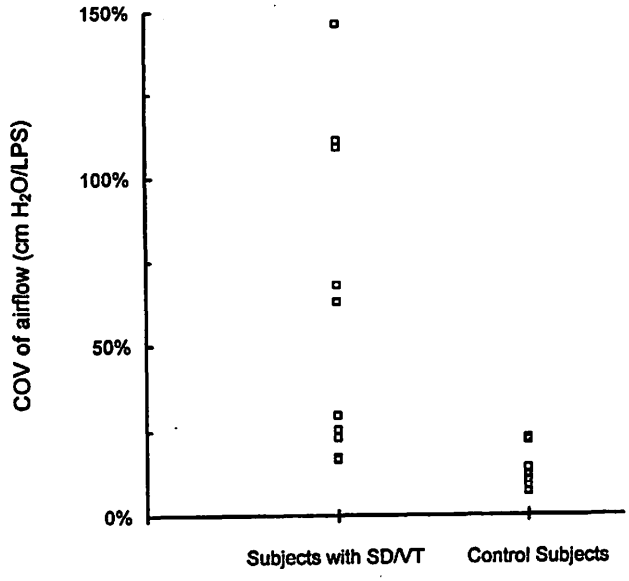


Figure 5. Mean phonatory airflow in subjects with SD/VT prior to each treatment (5 subjects x 2 treatments = 10 data points) in comparison to control subjects.

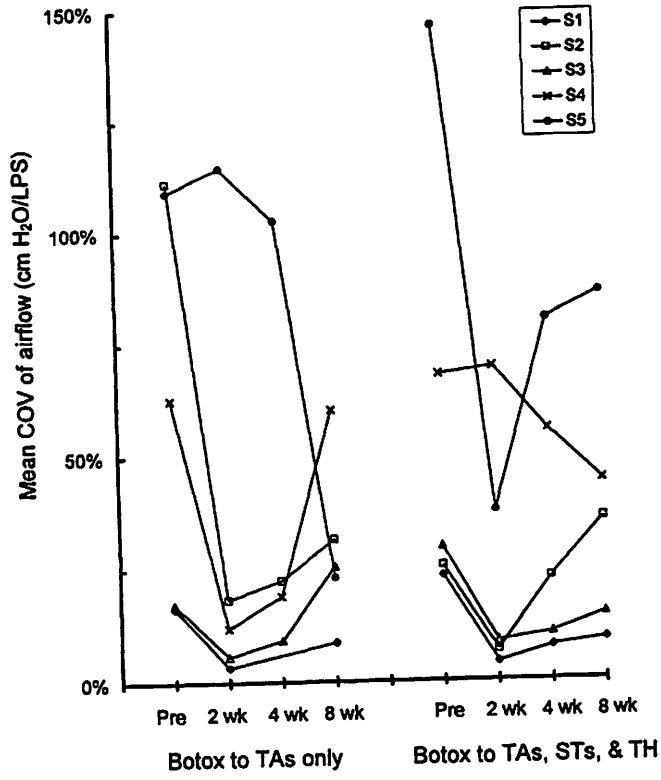


Figure 6. Change in mean phonatory airflow following botulinum toxin injection. Note: S5 receieved a booster injection of 2.5 units bilaterally to the TAs (5.0 units total), 14 days prior to the 8 week post (TA only) measurement.

tion. Ludlow, Naunton, Fijita, and Sedory (1990) observed that the EMG activity of some non-injected muscles was changed following treatment with botulinum toxin. They assessed the response of SD subjects with symptom recurrence following recurrent laryngeal nerve (RLN) surgery to injection of botulinum toxin. Comparison of bilateral EMG recordings from TA and CT muscles, obtained prior to and following treatment, revealed activity was diminished in the noninjected TA and CT muscles, as well as the injected TA. Zwirner et al. rated videoendoscopic data collected pre and post injection of botulinum toxin. They also observed "intrinsic [laryngeal muscle] hyperfunction and tremor were reduced bilaterally with unilateral injection" (p.406). Some evidence exists to suggest that the effects of the chemical denervation caused by botulinum toxin could extend beyond the intrinsic laryngeal muscles to muscles of respiration. Shipp, Izdebski, Schutte, and Morrissey (1988) measured subglottal pressure directly in one subject with SD, following recurrent laryngeal nerve block, and found a reduction in low frequency perturbations in subglottal pressure, suggesting an increase in stability of subglottal pressure.

A number of mechanisms have been suggested to explain how injection of botulinum toxin into the muscle might effect neural signals to that muscle. Ludlow, Hallett, Sedory, Fujita, and Naunton (1990) suggested that reduced muscle activity following botulinum toxin injection could result in reduced sensory feedback to the motorneuron pools. A second possibility is retrograde transport of the botulinum toxin. Simpson (1992) states "small quantities of the toxin can undergo retrograde axonal transport to reach the CNS... In healthy individuals the toxin cannot penetrate directly across the blood-brain barrier to affect the CNS" (p. 359). However, according to Malmgren (1992) "the neuromuscular junction present opportunities for the selective introduction of .. agents ...{that} are taken up into axon terminals at the neuromuscular junction and then transported to the neuronal cell body in the brainstem." (p.305). Malmgren, Olsson, Olsson, and Kristensson (1978) states "retrograde axonal transport of biologically active macromolecules to the cell body can sometimes result in large alterations in the metabolism of the neuron" (p.478). A third possibility is an effect on other cholergenergic synapses in addition to the neuromuscular junction. Although botulinum toxin has the highest affinity to nerve endings of motor fibers, it binds to post ganglionic parasympathetic cells and ganglionic cells as well (Simpson, 1981; Simpson, 1992).

An alternative explanation for the change in type as well as level of activity could be that, due to paralysis, the muscles ability to respond to the "spasmodic" neural signal is reduced, thus stability of airflow is minimized due to failure of muscle fibers to contract (or to contract as strongly). Our data show the changes in COV and mean airflow were closely related. In most cases, an decrease in COV of airflow was associated with an increase in mean flow. Our results would be consistent with either hypothesis (i.e. increase in stability is due to paresis alone or to paresis in combination with a change in the neural signal from the motorneurons in the brainstem) because the effect of the botulinum toxin on the laryngeal motor neuron pool could decrease at the same time as the effect at the neuromuscular junction..

A recently published article by Adams, Durkin, Irish, Wong, and Hunt (1996) measured the mean and COV of airflow in 30 patients with adductor SD following botulinum toxin injection and reported results similar to ours. Although they used the same task for data collection, they reported mean airflow values which tended to higher than ours and COV of airflow which were lower. This may have been due to differences in the method used for making these measures, since their measures were based on a 400 msec segment during vowel production. However, despite some differences in measurement technique, Adams et al. (1996) reported the same pattern found in our study of increased mean airflow and decreased COV of airflow at 2 weeks with subsequent decrease in mean airflow and increase in COV of airflow at 10 weeks post-injection. The agreement between the studies indicates that this pattern of change in airflow following botulinum injection is a robust finding.

Conclusion

Examination of changes in airflow following botulinum toxin injection may provide clues as to the nature of SD and the physiologic effect of laryngeal injection of botulinum toxin. Coefficient of variation (COV) of airflow, a measure of instability of flow, typically decreased following botulinum toxin injection. One can speculate whether decreased COV of airflow reflects the effect of botulinum toxin on the neuromuscular junction or on the laryngeal motorneuron pool as well, and if the laryngeal motorneuron pool is effected whether the effect is exerted on the respiratory as well as the laryngeal muscles, however further studies are needed. Direct measurement of subglottal pressure would be useful in accessing change in respiratory drive during phonation.

Acknowledgment

This research was supported by Grant No. P60 DC00976 from the National Institutes of Deafness and Other Communication Disorders.

References

Adams, S.G., Durkin, L.C., Irish, J.C., Wong, D.L.H., & Hunt, E.J. (1996). Effects of botulinum toxin type A injections on aerodynamic measures of spasmodic dysphonia. Laryngoscope, 106, 296-300.

Black & Dolly (1986a). Interaction of ^{125}I-labeled botulinum neurotoxins with nerve terminals. I. Ultastructural autoradiographic localization and quantitation of distinct membrane acceptors for types A and B on motor nerves. Journal of Cell Biology, 103, 521-34.

Black & Dolly (1986b). Interaction of ^{125}I-labeled botulinum neurotoxins with nerve terminals. II. Autoradiographic evidence for its uptake into motor nerves by acceptor-mediated endocytosis. Journal of Cell Biology, 103, 535-44.

Briant, T.D.R., Blair, R.L., Cole, P., & Singer, L. (1983). Laboratory investigation of abnormal voice. Journal of Otolaryngology, 12, 285-90.

Burgen, A.S., Dickens, F., & Zalman, L.J. (1949). The action of botulinum toxin on the neuromuscular junction. Journal of Physiology (London), 109, 10-24.

Davis, P. J., Boone, D. R., Carroll, R. L., Darveniza, P., & Harrison, G. A. (1988). Adductor spastic dysphonia: Heterogeneity of physiologic and phonatory characteristics. Annals of Otology, Rhinology & Laryngology, 97, 179-185.

Dolly, J.O., Black, J., Williams, R.S., & Melling, J. (1984). Acceptors for botulinum neurotoxin reside on motor nerve terminal and mediate its internalization. Nature (London), 307, 457-460.

Donovan, J.J. & Middlebrook, J.L. (1986). Ion-conducting channels produced by botulinum toxin in planear lipid membranes. Biochem, 25, 2872-76.

Finnegan, E.M., Luschei, E.S., Barkmeier, J.M., & Hoffman, H.T. (1996). Sources of error in estimation of laryngeal airway resistance in persons with spasmodic dysphonia. Journal of Speech and Hearing Disorders, 39, 105-113.

Hirano, M., Koike, Y., & von Leden, H. (1968). Maximum phonation time and air usage during phonation. Folia Phoniatri (Basel), 20, 185-201.

Hirano, M., & Ohala, J. (1969). Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. Journal of Speech and Hearing Research, 12, 362-73.

Ludlow, C. L., Hallett, M., Sedory, E.S., Fujita, M., & Naunton, R.F. (1990). The pathophysiology of spasmodic dysphonia and its modification by botulinum toxin. In: A Beradelli, R. Benecke, M Manfredi, & CD Marsden (eds) Motor Disturbances II. Academic Press, London. pp. 274-288.

Ludlow, C.L., Naunton, R.F., Fujita, M., & Sedory, S.E. (1990). Spasmodic dysphonia: botulinum toxin injection after recurrent nerve surgery. Otolaryngology-Head and Neck Surgery, 102, 122-131.

Malmgren, L.T. (1992). Anatomy and cell biology of the laryngeal motor neurons: Implications for the treatment of spasmodic dysphonia. Journal of Voice, 6, 298-305.

Malmgren, L., Olsson, Y., Olsson, T., & Kristensson, K. (1978). Uptake and retrograde axonal transport of various exogenous macromolecules in normal and crushed hypoglossal nerves. Brain Research, 153, 477-493.

Shipp, T., Izdebski, K., Schutte, H., & Morrissey, P. (1988). Subglottal air pressure in spastic dysphonia speech. Folia Phoniatrica, 40, 105-110.

Simpson, LL. (1981) The origin structure and pharmacological activity of botulimun toxin. Pharmacol Rev, 33, 155-88.

Simpson, L.L. (1992). Clinically relevant aspects of the mechanism of action of botulinum neurotoxin. Journal of Voice, 6, 358-364.

Smitheran, J.R., & Hixon, T.J. (1981). A clinical method for estimating laryngeal airway resistance during vowel production. Journal of Speech and Hearing Disorders, 46, 138-146.

Woo, P., Colton, R., Casper, J., & Brewer, D. (1992). Analysis of spasmodic dysphonia by aerodynamic and laryngostroboscopic measurements. Journal of Voice, 6, 344-351.

Woodson, G.E., Zwirner, P., Murry, T., & Swenson, M. (1991). Use of flexible fiberoptic laryngoscopy to assess patients with spasmodic dysphonia. Journal of Voice, 5, 85-91.

Zwirner, P., Murry, T., Swenson, M., & Woodson, G.(1992).Effects of botulinum toxin therapy in patients with adductor spasmodic dysphonia: Acoustic, aerodynamic, and videoendoscopic findings. Laryngoscope, 102, 400-40.

NCVS Status and Progress Report - 9
April 1996, 79-92

Speech Timing in Apraxia of Speech Versus Conduction Aphasia

Samuel Seddoh, M. Phil.
Donald Robin, Ph.D.
Hyun-Sub Sim, M.S.
Department of Speech Pathology and Audiology, The University of Iowa
Carlin Hageman, Ph.D.
Department of Communicative Disorders, The University of Northern Iowa
Jerald Moon, Ph.D.
John Folkins, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

This study examined temporal parameters of speech in subjects with apraxia of speech, conduction aphasia and individuals with normal speech. Participants were asked to repeat target words in a carrier phrase 10 times. Acoustic analyses involved measurement of stop gap duration, voice onset time, vowel nucleus duration, and consonant-vowel (CV) duration. Speakers with apraxia of speech had longer and more variable stop gap, vowel, and CV durations than did subjects with aphasia or normal speech. Speakers with conduction aphasia had longer vowel durations and CV durations than subjects with normal speech. Also, subjects with apraxia of speech showed greater token-to-token variability than the other subject groups. The variability shown by subjects with apraxia of speech was significantly correlated with perceptual judgments of their speech. The significance of these results is discussed in the context of motoric and phonological explanations for apraxia of speech and conduction aphasia.

The past two decades have witnessed an intensive debate on the nature of the mechanism underlying apraxia of speech (AOS). Whereas one school of thought contended that the disorder is primarily motoric, that is, without the involvement of the linguistic system (e.g., Wertz, LaPointe & Rosenbek, 1984), other researchers argued for an underlying linguistic disturbance (e.g., Martin, 1974). Recent investigations supporting the motoric interpretation have shown that speakers with apraxia exhibit impairment not only during speech movements, but also during nonspeech movements of the articulators (Hageman, Robin, Moon, & Folkins, 1994, 1996; Itoh, Sasanuma, & Ushijima, 1979; McNeil & Kent, 1990; McNeil, Weismer, Adams, & Mulligan, 1990).

One characteristic of AOS that also gives credence to the motoric view is temporal discoordination (Kelso & Tuller, 1981; Kent & Rosenbek, 1983). Different acoustic (Collins, Rosenbek, & Wertz, 1983; Kent & Rosenbek, 1983; McNeil, Liss, Tseng, & Kent, 1990), kinematic (McNeil & Adams, 1991) and perceptual (Odell, McNeil, Rosenbek, & Hunter, 1990, 1991) investigations have demonstrated that temporal parameters of speech are abnormal in speakers with apraxia. Some specific temporal abnormalities reported include overlapping ranges of voice onset time (VOT) for voiced and voiceless stop consonants (Freeman, Sands, & Harris, 1978; Itoh, Sasanuma, Tatsumi, & Kobayashi, 1979), variable ranges of VOT across subjects (Kent & Rosenbek, 1983; Itoh, Sasanuma, Tatsumi, & Kobayashi, 1979), longer than normal durations for vowels (Collins et al.; Freeman et al.), and longer than normal durations for consonants (Kent & Rosenbek). Because these temporal parameters were believed not to subserve phonemic functions in the stimuli used in these studies, abnormalities in the speech of AOS patients associated with them have been assumed to have origins in the motoric domain.

Whereas motoric impairment seems to contribute to the misarticulations of patients with AOS, sound errors made by patients with conduction aphasia (CA) have often been

accounted for in terms of breakdowns in phonological encoding (Beland, Caplan, & Nespoulous, 1990; Brown, 1975; Caplan, Vanier, & Baker, 1986; Friedrich, Glenn, & Marin, 1984; Gandour, Akamanon, Dechongkit, Khunardorn, & Boonklam, 1994a; Goodglass & Kaplan, 1983; Kohn, 1984; McCarthy & Warrington, 1984; Yamadori & Ikumura, 1975). For example, Friedrich et al. suggested that patients with CA are impaired in the ability to generate and maintain abstract phonological codes. A similar view was expressed recently by Gandour and colleagues (1994a) in a study of phonemic approximation in a Thai-speaking subject with CA. Gandour et al. (1994a) found that although the subject with CA seemed to know what he wanted to say, "in the process of producing the form of the word, he somehow mixes up the phonemes" (p. 89). Observations such as these led Gandour et al. (1994a) to the conjecture that the locus of the underlying deficit in CA might be related to a process in which lexical phonological representation is converted into an appropriate form for articulatory implementation.

Recently, however, some investigators have presented data challenging linguistic only interpretations for the deficit in CA (Kent & McNeil, 1987; McNeil & Adams, 1991; McNeil, Liss, et al., 1990). These researchers showed that their subjects with CA produced abnormal durations related to segmental, intersegmental and word level timing, relative to performances of normal speaking subjects. On the basis of these observations, it has been argued that motoric impairments contribute to CA errors. For example, Kent and McNeil (1987) contended that temporal abnormalities such as those associated with segment duration and VOT "are best interpreted as meaning that motoric planning, or execution, or both, are disrupted" (p. 213) in CA.

One factor that might support the view that the deficit underlying sound errors in CA has motoric components has to do with performance variability. Although abnormal variability is thought to be characteristic of AOS (e.g., Kent & Rosenbek, 1983; McNeil, Caligiuri, & Rosenbek, 1989; Seddoh et al., in press; Odell et al., 1990), a number of investigators have shown that speech performance in CA is also more variable than normal. For example, Kent and McNeil (1987) reported that VOT values and second formant (F2) trajectories were more variable for patients with AOS and CA, compared to results for subjects with normal speech.

Increased variability on speech tasks might occur from different sources (Nolan, 1982; Sharkey & Folkins, 1985; Smith & Kenney, 1994). However, many investigators have extrapolated from acoustic and physiological data to suggest that greater than normal variability in speech production may be an indication of instability in the speech motor control system (DiSimoni, 1974a, 1974b; Janssen & Wieneke, 1987; Kent & Forner, 1980; Tingley & Allen, 1975; Smith, 1992, 1994; Smith & Kenney, 1994; Wieneke & Janssen, 1987).

Performance of speakers with CA on nonspeech motor control tasks also lends support to the view that some of their speech errors may be motoric in nature. McNeil, Weismer, et al. (1990) reported that the instability of isometric force control of subjects with CA was not different from the performance of subjects with apraxic or normal speech. As in the Kent and McNeil (1987) study, the variability was more pronounced for the subjects with AOS than the subjects with CA. By contrast, Hageman et al., (1996) found that subjects with CA had normal motor control as assessed by a visuomotor tracking task. Thus, it is not known if speakers with CA have normal or abnormal motor systems and if motor control problems might explain some of their sound production errors.

CA, however, is known as a linguistic disorder, and phonology in particular is believed to be disturbed in these patients (e.g., Beland et al., 1990; Brown, 1975; Caplan et al., 1986; Gandour et al, 1994a). The phonologic level of language often contains time-based information (e.g., Abramson, 1974; Bolinger, 1976; Chomsky & Halle, 1968; Clements & Keyser, 1985; Cooper, Lapointe, & Paccia, 1977; Klatt, 1973, 1976; Lehiste, 1972; Lisker & Abramson, 1964; Nolan, 1982; Peterson & Lehiste, 1960). For example, some languages make phonemic distinctions between vowel length (e.g., Abramson, 1974), and there are also languages that have syllable timing rules that shorten or lengthen the durations of syllables occurring in different environments (e.g., Bolinger; Cooper et al.; Klatt, 1976). In English, the duration of an unstressed syllable is lengthened when the syllable is immediately followed by a stressed syllable (Bolinger, but see also Cooper et al.; Klatt, 1976; Lehiste, 1972)[1]. Also, the duration of an English stem vowel decreases (Lehiste, 1970) with increase in the length of an utterance (House, 1961).

Several phonologic theories, such as cyclic phonology (Rubach, 1984), lexical phonology (e.g., Luelsdorff, 1975) and nonlinear phonology (Bernhardt & Stoel-Gammon, 1994; Schwartz, 1992) have been developed that contain explicit time-based rules that are believed to drive the speech production system. For example, autosegmental theory (Goldsmith, 1990) and other nonlinear phonologies (Bernhardt, 1992) posit a skeletal tier that contains representations that include a formal mechanism for timing as well as the ordering of speech segments. Moreover, Browman and Goldstein (1989) have posited that phonologic systems may be organized around articulatory gestures. Furthermore, they suggest that problems in speech production may be reflective of misordering of phonologic rules. Because

[1] Cooper, Lapointe, and Paccia (1977), however, posited the same rule in the reverse direction as follows: "The duration of a stressed syllable that is immediately followed by an unstressed syllable (thereby forming a bisyllabic trochee) is normally shortened relative to its duration as a monosyllable" (p. 1314).

For the purposes of the present discussion it is the existence rather than the direction of the rule that is the relevant factor.

speech production is linked to the linguistic system, temporal abnormalities in speech might originate either from disturbance in the linguistic system or from motoric factors or both. Given that patients with CA have an impaired phonological system, it is possible that the timing errors of these patients might arise from defective phonological processing, and not defective motor control processes.

Although temporal control has been extensively studied in subjects with normal speech (e.g., Lehiste, 1972; Lofqvist & Yoshioka, 1984) and communicatively-impaired subjects such as people who stutter (e.g., Jancke, 1994; Wieneke & Janssen, 1991), relatively few data exist on subjects with AOS or CA (e.g., Gandour, Dechonkit, Ponglorpisit, & Khunadorn, 1994b; Kent & McNeil, 1987; McNeil, Liss, et al., 1990). Consequently, the nature of temporal control in the speech of these populations is poorly understood.

The purpose of the present investigation was to explore the nature of temporal control during speech in subjects with AOS or CA. Specifically, we compared the temporal patterns exhibited by these two patient groups to determine whether there were similarities or differences in the patterns. Similar patterns of temporal errors for subjects with AOS and subjects with CA and might indicate that the deficit contributing to the timing disturbance in the two populations is the same. On the other hand, significant differences in the patterns of temporal dissolutions for the two patient groups might suggest that the underlying deficits are different: motoric mechanisms responsible for the speech problems of individuals with AOS and linguistic mechanisms responsible for the speech problems of individuals with CA.

Method

Participants

Two experimental groups participated in the study (Table 1). They were comprised of five subjects with apraxia (AOS) and four subjects with conduction aphasia (CA). Three of the subjects with AOS were males and two were females. All of the four subjects with CA were males. The ages of the subjects with AOS and the subjects with CA ranged from 35 to 72 (mean = 59.2) years, and 21 to 66 (mean = 41.3) years, respectively. Because there were wide margins of age differences between the participants of the two groups ($p < .05$), a separate group of subjects with normal speech was recruited for comparison with each experimental group. The participants with normal speech were age and gender matched with the subjects in the experimental groups (AOS-C group and CA-C group). All participants including the experimental ones were in good physical health at the time of testing, by self report. For the experimental subjects, review of the chart reports of medical examinations were also used to confirm health status.

Each participant was right handed and had at least 12 years of education. Each participant in the two patient groups had unilateral left hemisphere lesions. The diagnostic characteristics of the subjects with AOS were consistent with the selection criteria outlined by Kent and Rosenbek (1983). In addition to these criteria, language tests were also administered and the results revealed no evidence of aphasia for the subjects with AOS.

Table 1.
Brain-Damaged Patients' Demographic Information

Particip. Job History	Age	Gender	Handedness*	Education
AOS1 Homemaker	35	Female	+100	12
AOS2 Farmer	55	Male	+100	12
AOS3 Electrical Engineer	66	Male	+100	16
AOS4 Homemaker	72	Female	+100	12
AOS5 Retired Teacher	68	Male	+100	16
CA1 Manufacturing Supervisor	49	Male	+100	12
CA2 College Student	29	Male	+100	13
CA3 Custodian	21	Male	+100	12
CA4 Retired Postal Worker	66	Male	+100	12

* Based on modified Oldfield Questionnaire (Tranel, 1992)
+100 = full right handedness, -100 = full left handedness

Inclusion criteria for both AOS and CA are identical to those used by McNeil and colleagues (Kent & McNeil, 1987; McNeil, Liss et al., 1990; McNeil, Weismer et al., 1990). These criteria included perceptual judgment by an ASHA-certified speech-language pathologist and the administration of a battery of standardized speech, language, and cognitive tests. Tests conducted included the Wechsler Adult Intelligence Scale, Revised (Weschler, 1981) for the evaluation of verbal ability and Intelligent Quotient (IQ); the Weschler Memory Scale (Russel, 1975) for the assessment of memory function in verbal and nonverbal domains; the Rey Auditory-Verbal Learning Test (Rey, 1941) for verbal memory; and the Complex Figure Delayed Recall Test (Osterrieth, 1944) for visuospatial memory. Visual perception was tested using the Revised Visual Retention Test (Benton, 1974), the Facial Recognition Test (Benton, Hamsher, & Varney, 1983) and the Judgment of Line Orientation Test (Benton & Hamsher, 1989; Ixenton et al., 1983). Constructional abilities were tested with the Complex Figure Test-Copy (Osterrieth, 1944). All neuropsychological tests were obtained from, and interpreted by, a neuropsychologist. Language was assessed using the Multilingual Aphasia Examination (Bénton & Hamsher, 1989), the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983), and the Token Test (De Renzi & Vignolo, 1962).

Table 2.
Language Information For Individuals with Apraxia (AOS) and Conduction Aphasia (CA)

(1 = Normal; 2 = Mild Impairment; 3 = Moderate Impairment; 4 = Severe Impairment).*

Particip.	Spontaneous Speech	Auditory Comprehension	Repetition	Naming	Reading Comprehension
AOS1	Nonfluent	1	3	1	1
AOS2	Nonfluent	1	3	1	1
AOS3	Nonfluent	1	4	1	1
AOS4	Nonfluent	2	3	2	1
AOS5	Nonfluent	1	4	NA	1
CA1	Fluent	3	4	3	3
CA2	Fluent	3	4	3	3
CA3	Fluent	2	3	3	2
CA4	Fluent	2	3	3	2

* Severity judgments were made by a speech-language pathologist (second author) and a neuropsychologist based on results from the Multilingual Aphasia Examination (Benton & Hamsher, 1989), clinical interview and chart review.

NA = Not tested.

Table 3.
Lesion Information
(All Brain-Damaged Patients Had a Single Left Hemispheric Stroke).

Particip.	Time Post Onset (Yrs.)	Lesion*
AOS1	8	Left inferior sector of precentral gyrus and posterior sectors of inferior parietal gyrus, superior temporal gyrus and inferior parietal lobule.
AOS2	7	Left inferior sector of precentral gyrus, posterior sector of inferior frontal gyrus and insula.
AOS3	7	Left inferior sector of precentral gyrus and posterior sector of inferior frontal gyrus.
AOS4	6	Left sensory motor cortex, posterior sector of inferior frontal gyrus, insula and basal ganglia.
AOS5	13	Left basal ganglia and insula.
CA1	5	Left inferior segment of inferior parietal lobule and insula.
CA2	5	Left inferior segment of inferior parietal lobule and posterior segment of superior temporal gyrus.
CA3	3	Left basal ganglia and insula.
CA4	2	Left inferior half of precentral gyrus and superior half of inferior frontal gyrus.

* Lesion data reported by Dr. Hanna Damasio, using a standard plotting analysis system (Damasio & Damasio, 1991).

The subjects with AOS exhibited normal performance on the various tests related to verbal performance, memory and vision. The diagnosis of CA was based on results of the Multilingual Aphasia Examination (Benton & Hamsher, 1989) and by the criteria described by Goodglass and Kaplan (1983). The hallmark of conduction aphasia is the inordinate impairment of repetition and the prevalence of phonemic paraphasias. Results of all nonverbal tests were within normal limits for all experimental subjects. Language data are shown in Table 2.

Neuroimaging data for lesion localization were obtained from MRI examinations using a standard lesion plotting method (Damasio & Damasio, 1991). Brain images were taken at least three months poststroke and analyzed by a standard plotting method. Lesion information is provided in Table 3. All subjects with brain damage suffered a single, left cerebral vascular accident.

Speech Materials

A set of four target words ("pop", "pea", "Bob", and "bee"), embedded in the carrier phrase, "That's a _____ a day," were used as the stimuli. One consideration that influenced the choice of these words was the presence of an initial voiced or voiceless stop consonant. This characteristic of the stimulus was deemed important in order to be able to assess participants' performances on temporal variables such as VOT and stop gap duration (i.e., the period of silence preceding the release for a stop consonant). The interest in these temporal parameters arose partly from the fact that voicing errors have been demonstrated to be a common problem for patients with apraxia or aphasia (Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980; Freeman et al., 1978; Gandour & Dardarananda, 1984; Itoh, Sasanuma, Tatsumi, Murakami, Fukusako, & Suzuki, 1982; Kent & Rosenbek, 1983). Moreover, voice onset time (VOT) has been shown to be a good index of the assessment of the temporal coordination of the larynx and the oral articulators (Blumstein et al., 1980; Blumstein, Cooper, Zurif, & Caramazza, 1977; Gandour & Dardarananda, 1984; Itoh et al., 1982; Lisker & Abramson, 1964).

Procedure

Participants were given an auditory model of the test phrases and were than asked to repeat each phrase 10 times at a comfortable rate. This yielded 10 tokens of each target word for each participant. The number of repetitions were limited to ten due to constraints of the time participants were available as well as possible fatigue. Participants spoke into a high quality microphone (with a flat frequency response of 10 Hz - 18 kHz) attached to a headset at a fixed distance of 15 cm. The utterances were recorded on a Sony 8-Channel Digital Instrumental Cassette Recorder (PC-108M) and were later low-pass filtered at 11 kHz and digitized at 22 kHz sampling rate using an IBM compatible 486-computer with an IBM M-Audio Capture and Playback Adapter (International Business Machines Corporation, 1989). Only perceptually accurate tokens, as judged by broad transcription, were included in the study (less than 1% of the tokens were omitted). The perceptual judgment was undertaken by two ASHA-certified speech-language pathologists who were asked to transcribe all sounds in the target word. Judgments were made independently and then the transcriptions were compared. There was 100% agreement between the two judges.

Temporal measures were made from the digitized speech samples using the CSpeech program (Milenkovic, 1990). Wide band spectrograms (bandwidth=300 Hz, frequency range 0-8000 Hz) were used to measure stop-gap duration (SGD), VOT, second formant transition duration for a vowel (F2D), steady-state vowel duration (VD), and CV duration (CVD), that is, the duration of the initial consonant and the following vowel in each target word. These measures were made from spectrographic displays of each target word as produced by a normal speaker (Figure 1), a CA speaker (Figure 2) and an AOS speaker (Figure 3). Note, as shown in the examples of speech tokens in the figures, all subjects ceased voicing before the target word.

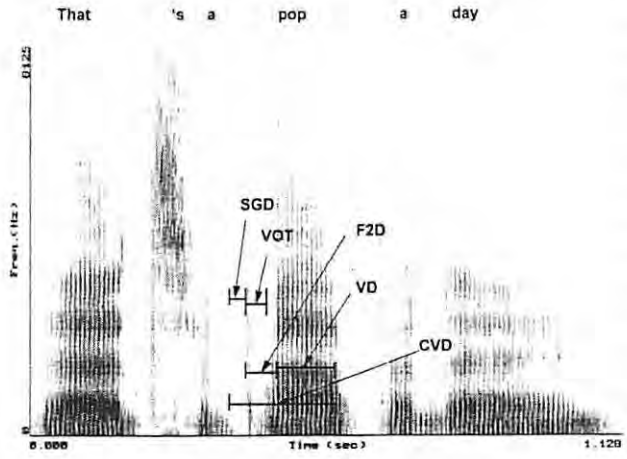


Figure 1. Spectrographic record of the phrase, "That's a pop a day," produced by a normal speaker (SGD=Stop Gap Duration, VOT=Voice Onset Time, VD=Vowel Nucleus Duration, and CVD=CV Duration).

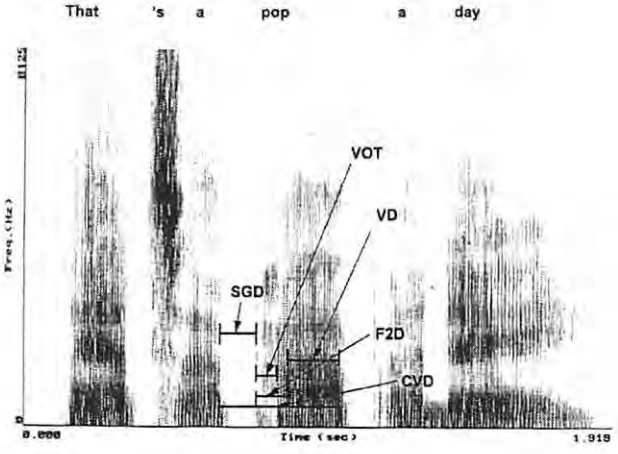


Figure 2. Spectrographic record of the phrase, "That's a pop a day", produced by a speaker with conduction aphasia.

Variability was examined using the standard deviations from the group data and by examination of each individual participant's standard deviations (see McNeil, Weismer, et al., 1990; Seddoh et al., in press). Analysis of data for individual subjects was deemed important in order to determine if the variability arose from the group data or from the individual participants' performances.

Perceptual analyses involving judgments of subjects' elicited spontaneous speech and reading by six ASHA-certified speech-language pathologists were also conducted. The speech samples for the perceptual analyses were gathered from presentation of sequence picture cards (developed in our laboratory) for the subjects with brain damage and subjects with normal speech to describe. In addition, subjects were also presented a passage from Readers' Digest (May, 1991, pp. 55-56) to read (see Appendix). The passage was printed boldly on two 8/12 x 11 sheets of white paper. The judges of this task, each of whom had at least 5 years experience as a full time clinician, rated the speech samples in three areas: overall speech defectiveness (OSD), overall articulatory imprecision (OAI), and intelligibility. The judgments were done using a 10 point equal-appearing interval scale with 1 being normal and 10 being severely defective.

Reliability

Acoustic Measures: The task of measuring the temporal variables was shared between the first and the third authors. Intraobserver reliability was determined by having each author reanalyze 10% of his own allotted portion of the data. Following this procedure each author also reanalyzed 10% of the other author's data to determine the interobserver reliability. Intra- and inter-observer reliability procedures,

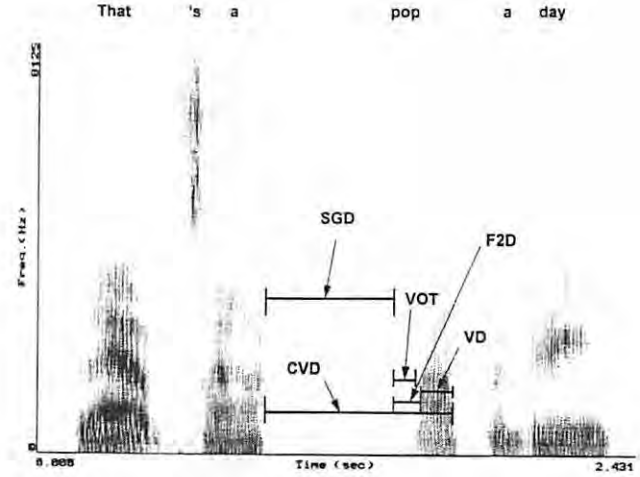


Figure 3. Spectrographic record of the phrase, "That's a pop a day," produced by a speaker with apraxia of speech. Note the long durations in comparison to those produced by the normal speaker (Figure 1) and the speaker with conduction aphasia (Figure 2).

agreements were 99% and 98%, respectively. Differences in judgments between the two observers of +/- 3 ms and less were deemed acceptable for the acoustic measures.

Perceptual Measures: Intrajudge reliability on the perceptual task wasalso based on a repetition of 10% of the samples from all participants. It was 99%. Interjudge reliability was determined by computing Pearson Product Moment correlations. Average correlations were 0.84 for OSD, 0.88 for OAI, and 0.81 for intelligibility.

Experimental Design: Separate two way analyses of variance (ANOVAs), Group (4 levels) x Word (4 levels) with repeated measures on the second variable were computed for each temporal variable, SGD, VOT, F2D, VD, and CVD. Tukey Honestly Significant Difference (HSD) post-hoc tests were performed at the .05 level of significance for main effects or interactions.

The token-to-token variability data did not meet the normal distribution assumption of parametric tests. Therefore, the Kruskal-Wallis test was performed on the token-to-token variability data.

Results

Mean Group Duration Results

Stop Gap Duration: ANOVA revealed a significant Group main effect for SGD, $F(3, 707) = 15.06$, $p = .0001$. The Word main effect was not significant, $F(3, 707) = 0.73$, $p = .5351$, nor was the two-way interaction, $F(9, 707) = 1.63$, $p = .1040$. The Group main effect is plotted in Figure 4. Posthoc Tukey HSD comparisons showed that the AOS group had a significantly longer mean SGD in comparison to the mean SGD for any of the other groups (AOS-C, CA, CA-C) who did not differ significantly. The AOS group demonstrated the largest SDs for SGD.

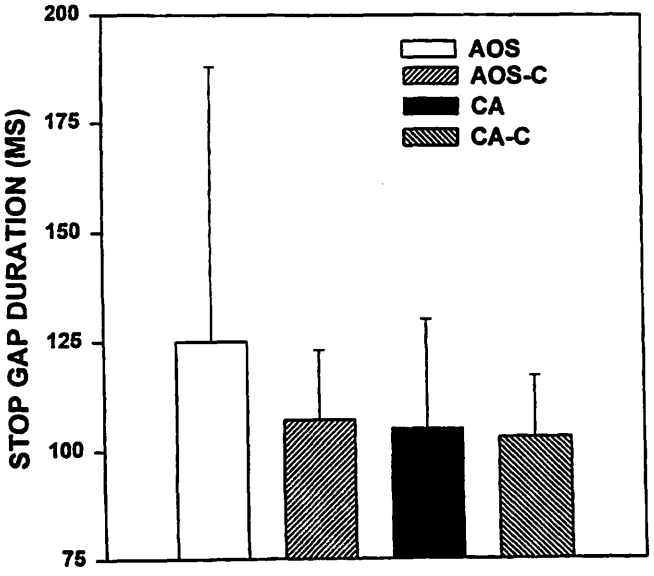


Figure 4. Mean stop gap duration for each participating group. Error bars indicate one standard deviation.

Voice Onset Time

For VOT, ANOVA revealed significant main effects for Group, $F(3, 709) = 22.50$, $p = .0001$, and Word $F(3, 709) = 325.75$, $p = .0001$, which were subsumed by a significant two-way interaction $F(9, 709) = 4.92$, $p = .0001$. The Group x Word interaction is shown in Figure 5. Posthoc Tukey HSD comparisons revealed that for the word, "pop," the apraxic group produced the longest mean VOT (65 ms). They were followed by the AOS-C group, who had a significantly longer mean VOT (54 ms) than the CA and the CA-C groups, who did not differ significantly. For the word "pea," the CA group produced a significantly shorter mean VOT (46 ms) than all the other groups, whose performances did not differ significantly. For "Bob," the AOS group had the longest mean VOT (18 ms) and the CA group the shortest mean VOT (6 ms), although the CA group's mean VOT did not differ significantly from the mean VOT (8 ms) of the CA-C group. For the word "bee," the CA group had a significantly shorter mean VOT (7 ms) than all other groups (AOS, AOS-C, CA-C) who were comparable. The AOS group demonstrated the largest SDs for most words.

Second Formant Transition Duration

ANOVA results for F2D showed significant main effects for Group, $F(3, 707) = 16.85$, $p = .0001$, and Word, $F(3,707) = 118.81$, $p = .0001$, which were subsumed by a significant two-way interaction, $F(9, 707) = 2.10$, $p = .0274$. The Group x Word interaction is displayed in Figure 6. Posthoc Tukey HSD comparisons revealed that the CA-C group produced the shortest mean F2D for the word "pop." The mean F2Ds of all the other groups (AOS, AOS-C, CA) for "pop" were comparable. For the word "pea," the CA

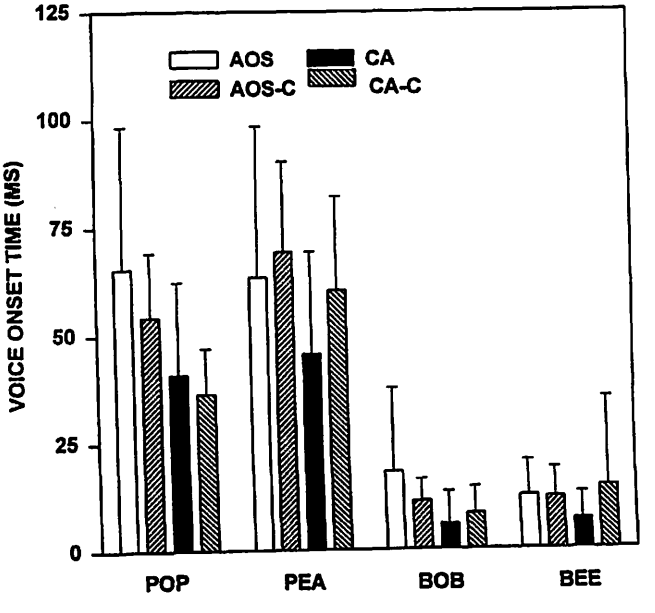


Figure 5. Mean voice onset time for each participating group by each target word.

group had the longest mean F2D and the CA-C group the shortest. There were no significant differences for the word "pea" for the F2D for the AOS and the AOS-C groups. For the word "Bob", the CA group had the longest mean F2D, and the other groups did not differ significantly. For the word "bee", the mean F2D for the AOS and CA groups did not differ significantly. The CA group had a significantly longer mean F2D than the CA-C group. By contrast, the mean F2D of the AOS group was significantly shorter than that of the AOS-C group.

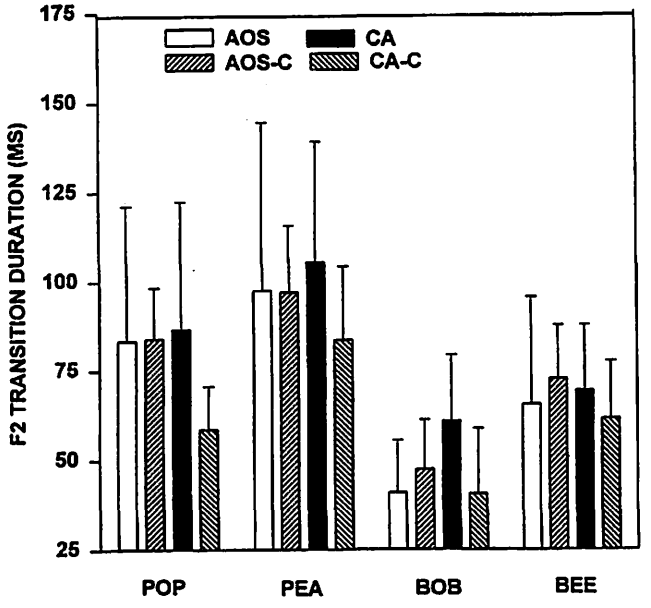


Figure 6. Mean second formant transition duration for each participating group by each target word.

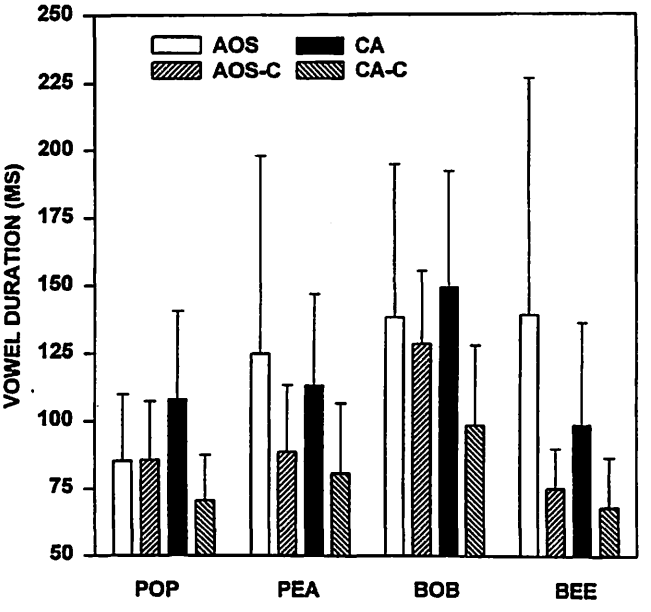


Figure 7. Mean vowel nucleus duration for each participating group for each target word.

The AOS and the CA groups exhibited SDs that were about twice the magnitude exhibited by the AOS-C and CA-C groups for the word "pop". The SD for the AOS group was largest for the word "pea" and was followed in magnitude by the SD for the CA group. The SDs for all four groups for "Bob" were comparable. The SD for the AOS group for the word "bee" was greater than those exhibited by all the other groups (AOS-C, CA, CA-C), which were comparable in magnitude.

Vowel Steady State Duration

For VD, significant main effects were found for Group, $F(3, 707) = 39.13$, $p = .0001$, and Word, $F(3, 707) = 33.33$, $p = .0001$, which were subsumed by a significant two-way interaction, $F(9, 707) = 5.35$, $p = .0001$. Figure 7 displays the Group x Word interaction. Follow-up Tukey HSD tests revealed that for the word "pop," the CA group had the longest mean VD and the CA-C the shortest. Unlike the CA and CA-C groups who differed significantly, the AOS group and the AOS-C group did not differ significantly for the mean VD for the word "pop." For the word "pea," the AOS group had the longest mean VD followed by the CA group, although these two groups did not differ significantly. The CA group had a significantly longer mean VD than the AOS-C and CA-C groups. For the word "Bob", the AOS, CA, and AOS-C groups' mean VDs did not differ significantly, but were significantly longer than the mean VD of the CA-C group. For the word "bee," the AOS group had a significantly longer mean VD than the CA group, who had a mean VD that was significantly longer than the AOS-C and CA-C groups.

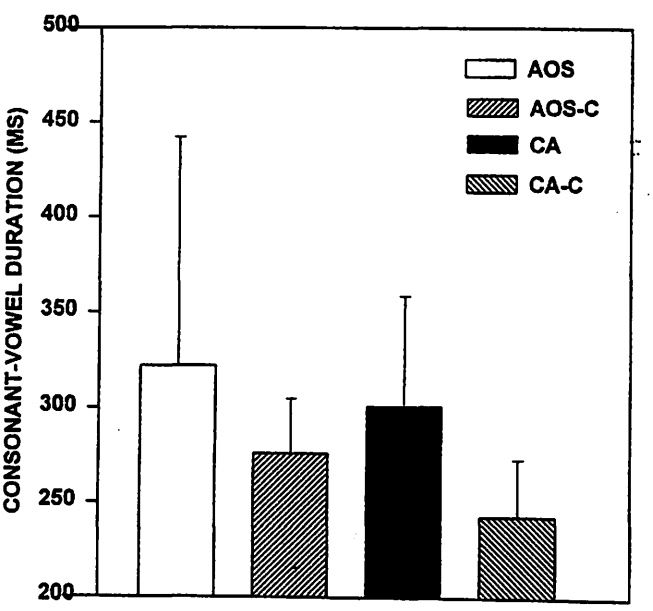


Figure 8. Mean CV duration for each participating group.

The AOS group had the largest SD for the words "pea", "Bob" and "bee" followed by the CA group. For the word "pop" the CA group had the largest SD.

Consonant-Vowel Duration

CVD analyses indicated a significant main effect for Group (Figure 8), $F(3, 705) = 35.84, p < .0001$ and Word, $F(3, 705) = 9.45, p < .0001$. The two-way interaction was not significant, $F(9, 705) = 1.38, p = .1926$. Posthoc Tukey HSD comparisons indicated that the AOS group had the longest mean CVD, followed by the CA, AOS-C, and the CA-C groups, respectively. The AOS group had the largest SD, followed by the CA group.

Summary of Group Durational Results

In summary, the AOS group had the longest (and the most variable) mean durations followed by the CA group, particularly with respect to SGD, VD and CVD. VOT measures showed an inconsistent pattern of results, particularly for the AOS group. Although the AOS group had significantly longer mean VOTs for "pop" and "Bob" than the CA group, in general, the mean VOTs of the AOS group were comparable to the performance of the AOS-C group. The CA group had longer mean durations for VD and CVD, in comparison to CA-C. The CA group also had shorter mean VOTs than any of the other groups (AOS, AOS-C, CA-C).

Table 4.
Standard Deviations of Individual Tokens for Individual Participants for Stop-Gap Duration (SGD), Voice Onest Time (VOT), Second Formant Transition Duration (F2D), Vowel Duration (VD), and CV Duration (CVD) for Apraxic (AOS), Apraxic Control (AOS-C), Aphasic (CA), and Aphasic Control (CA-C) Groups. Figures in Parenthesis are Group Standard Deviations.

Groups	SGD	VOT	F2D	VD	CVD
AOS					
AOS1	51.4	13.7	13.0	16.0	60.0
AOS2	30.1	9.3	16.2	19.5	35.5
AOS3	11.6	5.5	7.7	10.0	17.1
AOS4	83.6	16.9	18.8	52.7	102.8
AOS5	7.8	7.0	13.3	38.5	35.4
Mean	36.9(31.3)	10.5(4.7)	13.8(4.15)	27.3(17.7)	50.1(33.1)
AOS-C					
AOS-C1	10.5	7.7	10.8	17.9	24.1
AOS-C2	9.0	8.6	11.5	14.2	15.5
AOS-C3	11.2	11.4	13.8	13.9	19.5
AOS-C4	15.8	10.3	13.3	13.5	27.8
AOS-C5	14.0	10.8	15.0	11.9	19.6
Mean	12.1(2.8)	9.8(1.6)	12.9(1.7)	14.2(2.2)	21.3(4.8)
CA					
CA1	19.1	10.8	23.5	19.8	26.5
CA2	9.6	7.0	11.9	30.3	42.4
CA3	7.6	5.3	14.4	14.3	15.9
CA4	17.6	8.2	15.2	12.9	22.9
Mean	13.5(5.7)	7.8(2.3)	16.3(5.0)	19.3(7.9)	26.9(11.2)
CA-C					
CA-C1	14.1	15.1	11.2	9.6	17.3
CA-C2	7.8	6.8	16.8	14.6	20.7
CA-C3	8.3	6.2	13.0	14.2	22.3
CA-C4	7.0	5.2	10.2	8.1	13.9
Mean	9.3(3.2)	8.3(4.6)	12.8(2.9)	11.6(3.3)	18.5(3.7)

Token-to-Token Variability

Generally, more speakers with AOS exhibited greater token-to-token variability than did the speakers in all the other groups. SDs based on the individual token-to-token variability measures are shown in Table 4. Unlike the case for the CA group whose token-to-token SDs were generally within the range of the subjects with normal speech, a number of SDs from the subjects with AOS fell outside the normal range for SGD, VD and CVD.

Kruskal-Wallis Tests were performed on the mean token-to-token SDs to evaluate differences in the magnitude of variability for different groups. The tests revealed significant Group main effects for SGD ($p < .0040$), VD ($p < .0241$) and CVD ($p < .0101$). No other significant main effects or interactions were revealed. Posthoc comparisons indicated that for SGD the subjects with AOS had greater token-to-token SDs than any of the other groups (AOS-C, CA, CA-C), who did not differ significantly from one another. For VD, the subjects with AOS had significantly greater token-to-token SDs than the subjects in the AOS-C and CA-C groups. The subjects with CA had token-to-token SDs that were not significantly different from those for the subjects in the AOS group or the subjects in the AOS-C and CA-C groups. For CVD, the AOS group exhibited significantly greater token-to-token SDs than subjects in the CA and the AOS-C group, but they did not differ significantly from CA-C groups. The subjects in the CA group had greater token-to-token SDs than the subjects in the CA-C group.

Perceptual analysis

Table 5 shows the means and standard deviations from the results of the perceptual judgments of speech for all groups. The AOS group was significantly different ($p < .05$) from the other groups in all three perceptual variables, namely OAI, OSD, and intelligibility. Thus, the speech

Table 5.
Mean Perceptual Judgments of Speech (and Standard Deviations) for Individuals With Apraxia (AOS), Apraxic Control Participants (AOS-C), Individuals with Aphasia (CA), and Aphasic Control (CA-C) Participants for Overall Articulatory Imprecision (OAI), Overall Speech Defectiveness (OSD), and Intelligibility.

	OAI	OSD	Intelligibility
AOS	7.33 (2.13)*	7.17 (2.52)*	7.92 (2.07)*
AOS-C	9.11 (0.53)	9.47 (0.38)	9.92 (0.12)
CA	8.63 (0.46)	9.06 (0.23)	9.50 (0.18)
CA-C	8.96 (0.46)	8.95 (0.72)	9.50 (0.47)

(* = $p < .05$)

Table 6.
Correlation of Individual Variability With Perceptual Judgment of Speech for Individuals With Apraxia.

	OAI	OSD	Intelligibility
SGD	-0.14	-0.72*	-0.80*
VOT	-0.01	-0.61	-0.68*
VD	-0.14	-0.84*	-0.64*
CVD	-0.28	-0.79*	-0.78*

(* = p < .05)

from the individuals with apraxia was considered less precise, more defective, and less intelligible than the speech of the other individuals.

For individual participants in the AOS group, token-to-token variability was correlated with the perceptual judgments of their speech for OAI, OSD, and intelligibility. This analysis was performed on the AOS group only, because the CA group was not different from controls in terms of the perceptual judgments of their speech and the CA group had a fairly restricted range of variability. Correlations for the AOS group for each dependent variable are shown in Table 6. Strong negative correlations were found for OSD and Intelligibility, but not for OAI. These correlations were particularly robust for SGD, VD, and CVD.

Discussion

Results of the present study suggest that patients with AOS or CA have abnormal temporal control during speech. These results are consistent with those of previous investigations that have demonstrated that subjects with AOS or CA have a tendency to produce longer than normal durations (e.g., Kent & McNeil, 1987; McNeil, Liss, et al., 1990). In addition, the findings also indicate that individuals with AOS exhibit greater variability in repeated productions of the same utterance, compared to subjects with CA or normal speech.

Recent acoustic and physiological investigations have suggested that in communicatively impaired individuals like those with apraxia or fluency problems, temporal abnormalities such as those observed in the present study, are reflections of a disrupted motor control system (Janssen & Wieneke, 1987; Kent & McNeil, 1987; Kent & Rosenbek, 1983; McNeil & Adams, 1991; McNeil, Liss, et al., 1990; Wieneke & Janssen, 1987). Given these previous interpretations of temporal control in speech, the findings in the present study suggest that patients with AOS have a defective speech motor control system. While the four subjects with CA showed abnormal durations for VD and CVD, they did not demonstrate abnormal token-to-token variability. This suggests that any motor control problems in this group of speakers with CA is subtle and that a linguistic mechanism likely contributes in a larger way to their errors than a motoric one. Thus, these data stand in contrast to Kent and McNeil (1987) who explicitly stated that with respect to the source of temporal disturbance, "conduction aphasia is similar to apraxia of speech" (p. 213).

Our results do not support the view that the primary underlying deficit contributing to the temporal abnormalities in both AOS and CA is the same. Although it is possible that motoric consequences on timing could have affected the performances of participants of either group, this conclusion is not particularly compelling when one takes into account the fact that the patterns of temporal dissolution exhibited by the two groups of subjects take different shapes, both qualitatively and quantitatively. Recall that, in general, the abnormal duration and variability measures were significantly more pronounced for the AOS than the CA group. Visual inspection of Figures 4-8 also reveals patterns of performances for the two groups that are not similar. For example, whereas the CA group produced the longest mean VD for the word "pop" (108 ms), see Figure 7, the AOS group exhibited the longest mean VD for "bee" (139 ms). Also, although the mean VD for the AOS group for the word "pea" (125 ms) did not differ significantly from that produced by the CA group for the same word (113 ms), the mean token-to-token SD exhibited by the AOS group for this word (30.8 ms) was significantly greater than that exhibited by the CA group (20 ms) who did not differ significantly from their control (CA-C) group (15 ms). Furthermore, unlike the case for measures of VD, the subjects with AOS exhibited significantly longer mean VOTs than subjects with CA (see Figure 5) across all target words. If the underlying source of the deficit is the same for both groups, then it would be difficult to account for why they do not exhibit approximately similar patterns of output in their performances.

As noted above, our findings on AOS appear to implicate motoric factors as the impairment underlying this group's abnormal temporal pattern. Analyses of the SGD data indicated that the AOS group had a significantly longer mean duration than all the other groups. On the other hand, the CA group's performance was comparable to the mean durational performance of the AOS-C and CA-C groups (see Figures 2, 3, and 4). In addition, the AOS group also exhibited the largest token-to-token SD. A similar finding was reported by Kent and Rosenbek (1983) who noted that although their apraxic patients lengthened nearly all segments, "two that stand out are the nasal murmur for [n] and the stop gap for [d]" (p.238). On the basis of this and other findings, they characterized apraxic speech as slow. Thus, the abnormal SGD exhibited by the AOS patients in the present study may be indicative of the involvement of factors related to the motor system. By contrast, the relatively normal performance of the subjects with CA on the SGD measure points to the absence or minimal involvement, if at all, of primarily motoric factors in the underlying deficit for this group.

The patterns of performance of the subjects with AOS and the subjects with CA on other measures further implicate motoric factors for the AOS group. On measures of VD, AOS patients had comparable mean durations as the AOS-C group for the words "pop" and "Bob." On the other hand, the subjects with AOS had mean VDs that were abnormally long for "pea" (about two times longer than normal) and "bee" (about three times longer than normal). In addition, their SDs were also largest for these two words. By contrast, subjects with CA exhibited about two times the VD exhibited by their control counterparts, and relatively normal SDs, for all target words.

The finding that the subjects with AOS exhibited greater difficulty (especially in the form of greater SDs) for high front unrounded than for low back vowels is counter to what previous investigators have reported for this group. In their perceptual study of vowel and prosodic production in subjects with AOS, CA or dysarthria, Odell and colleagues (1991) found that their subjects with AOS exhibited over 20% more errors on back vowels than front vowels. On the other hand, as was also the case for the four subjects with CA of the present study, these phonetic characteristics were found to be less influential in the performances of Odell's subjects with CA or dysarthria. The difference in the two studies with respect to the results for subjects with AOS may be partly due to methodological factors. Odell et al.(1991) used perceptual analyses, whereas the current investigation used an acoustic methodology, and the results might have been influenced by the difference in methods. Indeed, it is noteworthy that the acoustic abnormalities observed in the current study occurred despite the fact that all the tokens included in the study were judged to be perceptually accurate. Secondly, the stimuli used by Odell et al. (1991) were words of different lengths (mono-, di- and tri-syllabic words) spoken in isolation; the stimuli used in the current investigation were monosyllabic target words embedded in a carrier phrase. Because our stimuli, unlike those used by Odell et al. (1991), were sentences, the strategies used by the subjects in the current investigation during might have been different from those employed by the participants in the Odell et al. (1991) study.

The performance of the subjects with AOS in the present study may also have been influenced by other factors such as the phonological status of the vowels in the two groups of words ("pop"/"Bob" vs. "pea"/"bee"). Phonologically the /i/ in "pea" and "bee" is [+ tense], and is distinguishable from the /ɔ/ (or /a/)[2] in "pop" and "Bob" in terms of the feature, [-tense] (or lax) (Chomsky & Halle, 1968). Tense sounds are longer in duration than their lax counterparts (Chomsky & Halle, 1968). The finding of longer than normal mean VDs for the tense vowels by the AOS group therefore suggests that these patients might have been exaggerating the tense-lax phonological distinction between the two vowel types. On the other hand, the abnormal SDs from the subjects with AOS point to articulatory constraints, presumably arising from the motoric demands on the articulators for the production of tense vowels. This motoric view implies that whereas the capacity for phonological processing might be preserved in AOS, articulation is disturbed, consistent with the characterization of this disorder as a motor level impairment.

Unlike the case for AOS, the CA group findings indicate a predominant problem due to a phonological mechanism. The magnitudes of the mean VDs produced by the subjects with CA were approximately uniform across all target words except "Bob" for which a significantly higher VD was exhibited. The SDs exhibited for all target words were comparable to the SDs for the CA-C group. Generally, the SDs for the CA group for "pea" and "bee" (the target words with the tense vowel, [i]) were about two times less than the SDs exhibited by the AOS group for these words. In summary, the data from the current study suggest the possibility that the source of the abnormal timing in CA might lie in the phonological system due to the low SDs exhibited by the CA group.

Another factor that also seems to point at phonological involvement in the subjects with CA has to do with stress production. It is known that stressed syllables in English are longer relative to their durations as unstressed syllables (Borden & Harris, 1980; Fry, 1955). The target words used in the present study occurred in stressed positions in the carrier phrase. This means that their appropriate production required the application of stress placement rule (Chomsky & Halle, 1968). If the subjects with CA did not apply this rule correctly, it is possible that the durations of the nuclei (vowels) of the target words (or other syllables in the utterance), were inappropriately produced. The consistently higher and approximately uniform magnitude of VD shown by the CA group suggests that these patients might have been inappropriately applying the stress placement rule (although it is possible that the subjects with CA did use other acoustic cues to mark stress). The abnormal VD in the CA group suggests that although the capacity for phonological processing may not be lost in this population, the processing of phonological rules might be inefficient.

The finding that subjects with CA had longer F2D than their matched subjects with normal speech, but similar F2Ds as the AOS and the AOS-C groups warrants discussion. This finding suggests that the subjects with CA may have some difficulty controlling timing in the motoric domain. However, their token-to-token variability for the F2D measure was no different from normal, making a motoric

[2] Although most participants produced the vowel in "Bob" and "pop" as [ɔ], a few others produced it as [a]. They differ from each other phonetically, but these two vowels behave phonologically the same in the English language (e.g., they are both [-tense]). Partly for this reason, and also for the sake of simplicity, only [ɔ] will be used in the discussion.

explanation somewhat difficult. It may be that the increased duration in this case is a subtle indication of motoric involvement, but data from the current study are unable to address this issue further.

Similar comments apply to the token-to-token variability data for VD for the subjects with CA. The results showed that the subjects with CA were not different from the subjects with AOS; but they were also not different from the two normal control groups. Such an ambiguous result was also reported for the subjects with CA studied by McNeil, Weismer, et al. (1990) for non-speech motor control tasks. This result might suggest a subtle motoric involvement in the subjects with CA in the present study. However, a motoric explanation as the underlying factor in CA on the basis of this result alone is inappropriate at this time. Exploration of these results must await further investigation.

Results for CVD for both subjects with AOS and subjects with CA may be indicative of cumulative effects of mistiming of different temporal components of the segment or the syllable. The AOS group had the longest mean CVD, followed by the CA group, consistent with previous findings (e.g., Kent & McNeil, 1987). This result is what one would expect given the fact that the subjects with AOS exhibited longer mean durations than the subjects with CA on most other measures (all of which are constituent components of CVD).

It is not quite clear why the mean VOTs exhibited by the subjects with AOS were normal in the target words with open syllables ("pea" and "bee") but abnormal in the target words with closed syllables ("pop" and "Bob"). One possible explanation, however, points to phonetic planning difficulties. Target words with the closed syllables may be more difficult to produce than those with open syllables. If this is the case, it might be expected that motor programming of closed syllables is more prone to errors in comparison to the programming for open syllables. This view, however, is subject to empirical verification.

Variability

There is a growing consensus that increased variability in speech timing may point to the involvement of the speech motor control system (DiSimoni, 1974a, 1974b; Janssen & Wieneke, 1987; Smith, 1992, 1994; Smith & Kenney, 1994; Wieneke & Janssen, 1987). The data from the subjects with AOS in the present study on both the mean durational control measures and their associated SDs are consistent with the view that the underlying impairment in this group is primarily motoric. In support of this view, many of the subjects with AOS exhibited considerably greater group and token-to-token variability (SDs), particularly for SGD, VD and CVD, compared to other groups. Differences between the performances of AOS and the subjects with normal speech on measures of VOT, however, were less marked, suggesting that some aspects of temporal control may be preserved in AOS despite the motoric impairment.

In contrast to the pattern of token-to-token variability exhibited by the subjects with AOS, the data from the subjects with CA were comparable to those of the subjects with normal speech. Previous studies have also reported that patients with CA exhibit less variability in speech compared to speakers with apraxia (e.g., Kent & McNeil, 1987; McNeil et al., 1990). It was suggested above that the data from the subjects with CA on the mean durational control data were consistent with phonological impairment. The relatively normal token-to-token variability exhibited by the subjects with CA further supports this view, or at least demonstrates that a motoric deficit may be unlikely as the only source of the abnormal timing in this population.

Finally, the relation between token-to-token variability and perceptual judgments of the speech of the subjects with AOS warrants comment. Significant correlations between judgments of intelligibility and overall speech deficit and token-to-token variability were found for the speakers with AOS in this study. The perceptual measures were obtained from a picture description and paragraph reading tasks and not the experimental task (because we included only accurate tokens from the experimental task and less than 1% of the experimental tokens had to be removed from the analysis due to perceptual inaccuracy). Thus, high token-to-token variability on the production of target words in carrier sentences was associated with poorer speech in reading and elicited spontaneous speech tasks. If temporal variability provides information about the stability of the motor control system, then one can interpret these data to suggest that those subjects who had greater variability were motorically less stable, and therefore were more prone to speech breakdown. It is the case that only perceptually accurate tokens were acoustically analyzed in this study. One interpretation is that, although the AOS patients had motor impairment, they also had the flexibility to compensate for the variable performance on the experimental task. It is also possible that the production of the target words in a carrier phrase is an easier task than reading or elicited descriptions of action pictures.

In summary, the subjects with AOS in the present study were impaired in terms of durational control. This impairment was manifest in longer than normal mean durations of stop gap, vowel and consonant-vowel productions as well as increased token-to-token variability associated with these measures. These data support the view that AOS is a motor speech disorder. Although the subjects with CA had longer than normal vowel and CV durations, their productions were not more variable than normal. These data suggest that the underlying factors in timing deficits in CA may be phonological rather than motoric.

References

Abramson, A. S. (1974). Experimental phonetics in phonology: Vowel duration in Thai. Pasaa, 4 (1), 71-90.

Beland, R., Caplan, D., & Nespoulous, J. L. (1990). The role of abstract phonological representation in word production: Evidence from phonemic paraphasias. Journal of Neurolinguistics, 5, 125-164.

Benton, A. L. (1974). The Revised Visual Retention Test, (4th ed.). New York: Psychological Corporation.

Benton, A. L., & Hamsher, K. (1989). Multilingual Aphasia Examination (2nd. ed.). Iowa City, IA: AJA Associates.

Benton, A. L., Hamsher, K., & Varney N. R. (1983). Contributions to Neuropsychological Assessment. New York: Oxford University Press.

Bernhardt, B. (1992). Developmental implications of nonlinear phonological theory. Clinical Linguistics and Phonetics, 6, 259-281.

Bernhardt, B., & Stoel-Gammon, C. (1994). Nonlinear phonology: Introduction and clinical application: Tutorial. Journal of Speech and Hearing Research, 37, 123-143.

Blumstein, S., Cooper, W., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice onset time analysis. Brain and Language, 9, 153-170.

Blumstein, S., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. Neuropsychologia, 15, 371-383.

Bolinger, D. L. (1976). Length, vowel, juncture. Bilingual Review, 3, 43-61.

Borden, G.J., & Harris, K.S. (1980). Speech science primer: Physiology, acoustics, and perception of speech. Baltimore: Williams & Wilkins.

Browman, C.P. & Goldstein, L. (1989). Articulatory gestures as phonological units. Phonology, 6, 201-251.

Brown, J. W. (1975). The problem of repetition: A study of "conduction" aphasia and the "isolation" syndrome. Cortex, 11, 37-52.

Caplan, D., Vanier, M., & Baker, C. (1986). A case study of reproduction conduction aphasia: I. Word production. Cognitive Neuropsychology, 3, 99-128.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Clements, G. N, & Keyser, S. J. (1985). CV Phonology: A Generative Theory of the Syllable. Cambridge, MA: MIT Press.

Collins, M. J., Rosenbek, J. C., & Wertz, R. T. (1983). Spectrographic analysis of vowel and word duration in apraxia of speech. Journal of Speech & Hearing Research, 26, 224-230.

Cooper, W. E., Lapointe, S. G., & Paccia, J. M. (1977). Syntactic blocking of phonological rules in speech production. Journal of the Acoustical Society of America, 61(5), 1314-1320.

Damasio, H., & Damasio, A. R. (1991). Lesion analysis in neuropsychology. New York: Oxford University Press.

De Renzi E., & Vignolo, L. A. (1962). The Token Test: A sensitive test to detect disturbances in aphasics. Brain, 85, 665-678.

DiSimoni, F. G. (1974a). Some preliminary observations on temporal compensation in the speech of children. Journal of the Acoustical Society of America, 56, 697-699.

DiSimoni, F. G. (1974b). Influence of vowel environment on the duration of consonants in the speech of three-, six-and nine-year old children. Journal of the Acoustical Society of America, 55, 360-361.

Freeman, F. J., Sands, E. S., & Harris, K. S. (1978). Temporal coordination of phonation and articulation in a case of verbal apraxia: A voice onset time study. Brain and Language, 6, 106-111.

Friedrich, F. J., Glenn, C. G., & Marin, O. S. M. (1984). Interruption of phonological coding in conduction aphasia. Brain and Language, 22, 266-291.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America, 27, 765-768.

Gandour, J., Akamanon, C., Dechongkit, S., Khunadorn, F., & Boonklam, R. (1994a). Sequences of phonemic approximations in a Thai conduction aphasic. Brain and Language, 46, 69-95.

Gandour, J., & Dardarananda, R. (1984). Voice onset time in aphasia: Thai. II. Production. Brain and Language, 23, 177-205.

Gandour, J., Dechongkit, S., Ponglorpisit, S., & Khunadorn, F. (1994b). Speech timing at the sentence level in Thai after unilateral brain damage. Brain and Language, 46, 419-438.

Goldsmith, J. (1990). Autosegmental and metrical phonology. Cambridge, MA: Basil Blackwell.

Goodglass, H., & Kaplan, E. (1983). The assessment of aphasia and related disorders (2nd. Ed.). Philadelphia, PA: Lea Febiger.

Hageman, C., Robin, D. A., Moon, J. B., & Folkins, J. W. (1994). Oral motor tracking in normal and apraxic speakers. In M. L. Lemme (Ed.), Clinical Aphasiology, Vol. 22 (pp. 219-229).

Hageman, C., Robin, D. A., Moon, J. B., & Folkins, J. W. (1995). Visuomotor tracking abilities of apraxic, conduction aphasic, and normal speakers. Manuscript submitted for publication.

House, A. S. (1961). On vowel duration in English. Journal of the Acoustical Society of America, 33, 1174-1178.

International Business Machines Corp. (1989). M-Audio Capture and Playback Adapter.

Itoh, M., Sasanuma, S., Tatsumi, I., & Kobayashi, Y. (1979). Voice onset time characteristics of apraxia of speech. Annual Bulletin No. 13 (pp. 123-132). Tokyo: Research Institute of Logopedics and Phoniatrics, University of Tokyo.

Itoh, M., Sasanuma, S., Tatsumi, I., Murakami, S., Fukusako, Y., & Suzuki, T. (1982). Voice onset time characteristics in apraxia of speech. Brain and Language, 17, 193-210.

Itoh, M., Sasanuma, S., & Ushijima, T. (1979). Velar movements during speech in a patient with apraxia of speech. Brain and Language, 7, 227-239.

Jancke, L. (1994). Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults. Journal of Fluency Disorders, 19, 21-37.

Janssen, P., & Wieneke, G. (1987). The effects of fluency inducing conditions on the variability in the duration of laryngeal movements during stutterers' fluent speech. In H. F. M. Peters & W. Hulstijn (Eds.), Speech Motor Dynamics in Stuttering (pp. 337-344). Wien: Springer-Verlag.

Kelso, J. A. S., & Tuller, B. (1981). Toward a theory of apractic syndromes. Brain and Language, 12, 224-245.

Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. Journal of Phonetics, 8, 157-168.

Kent, R. D., & McNeil, M. R. (1987). Relative timing of sentence repetition in apraxia of speech. In J. Ryalls (Ed.), Phonetic approaches to speech production in aphasia and related disorders (pp. 181-220). San Diego: College-Hill Press.

Kent, R. D., & Rosenbek, J. C. (1983). Acoustic patterns of apraxia of speech. Journal of Speech & Hearing Research, 26, 231-249.

Klatt, D. (1973). Interaction between two factors that influence vowel duration. Journal of the Acoustical Society of America, 54, 1102-1104.

Klatt, D. (1976). Linguistic uses of segment duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 59, 1208-1221.

Kohn, S. E. (1984). The nature of phonological disorder in conduction aphasia. Brain and Language, 23, 97-115.

Lehiste, I. (1970). Suprasegmentals. Cambridge, MA: MIT Press.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. Journal of the Acoustical Society of America, 51, 2018-2024.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops. Word, 20, 384-422.

Lofqvist, A., & Yoshioka, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. Speech Communication, 3, 279-289.

Luelsdorff, P. (1975). A segmental phonology of black English. The Hague: Mouton.

Martin, A. D. (1974). Some objections to the term apraxia of speech. Journal of Speech and Hearing Disorders, 39, 53-64.

McCarthy, R., & Warrington, E. K. (1984). A two-route model of speech production: Evidence from aphasia. Brain, 107, 463-485.

McNeil, M. R., & Adams, S. (1991). A comparison of speech kinematics among apraxic, conduction aphasic, ataxic dysarthric, and normal geriatric speakers. Clinical Aphasiology, 19, 279-294.

McNeil, M. R., Caligiuri, M., & Rosenbek, J. C. (1989). A comparison of labio-mandibular kinematic durations, displacement, velocities and dysmetrias in apraxic and normal adults. Clinical Aphasiology, 18, 173-193.

McNeil, M. R., & Kent, R. D. (1990). Motoric characteristics of adult aphasic and apraxic speakers. In G. R. Hammond (Ed.), Advances in psychology: Cerebral control of speech and limb movements (pp. 349-386). New York: Elsevier/North Holland.

McNeil, M. R., Liss, J. M., Tseng, C-H., & Kent, R. D. (1990). Effects of speech rate on the absolute and relative timing of apraxic and conduction aphasic sentence production. Brain & Language, 38(1), 135-158.

McNeil, M. R., Weismer, G., Adams, S., & Mulligan, M. (1990). Oral structure nonspeech control in normal, dysarthric, aphasic and apraxic speakers: Isometric force and static position control. Journal of Speech & Hearing Research, 33, 255-268.

Milenkovic, P. (1990). Cspeech (computer program). Madison, WI: University of Wisconsin.

Mohanan, K.P. (1986). The theory of lexical phonology. Dordrecht: D. Reidel Publishing Co.

Nolan, F. J. (1982). The role of Action Theory in the description of speech production. Linguistics, 20, 287-308.

Odell, K., McNeil, M. R., Rosenbek, J. C., & Hunter, L. (1990). Perceptual characteristics of consonant production by apraxic speakers. Journal of Speech & Hearing Disorders, 55, 345-359.

Odell, K., McNeil, M. R., Rosenbek, J. C., & Hunter, L. (1991). Perceptual characteristics of vowel and prosody production in apraxic, aphasic, and dysarthric speakers. Journal of Speech & Hearing Research, 34, 67-80.

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe. Archives de Psychologie, 30, 206-356.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32, 673-703.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. Archives de Psychologie, 28 (112), 286-340.

Rubach, J. (1984). Cyclic and lexical phonology: The structure of Polish. Dordrecht: Foris Publications.

Russel, E. W. (1975). A multiple scoring method for the assessment of complex memory functions. Journal of Consulting and Clinical Psychology, 43, 800-809.

Schwartz, R.G. (1992). Nonlinear phonology as a framework for phonological acquisition. In R.S. Chapman (Ed.), Processes in language acquisition and disorders (pp. 108-124). St. Louis: Mosby-Year Book, Inc.

Seddoh, S. K., Robin, D. A., Hageman, C., Sim, H-S., Moon, J. B., & Folkins, J. W. (in press). Temporal control in apraxia of speech: An acoustic investigation of token-to-token variability. Clinical Aphasiology.

Sharkey, S. G., & Folkins, J. W. (1985). Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control. Journal of Speech and Hearing Research, 28, 8-15.

Smith, B. L. (1992). Relationships between duration and temporal variability in children's speech. Journal of the Acoustical Society of America, 91, 2165-2174.

Smith, B. L. (1994). Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adults' speech. Journal of Phonetics, 22, 155-175.

Smith, B. L., & Kenney, M. K. (1994). Variability control in speech production tasks performed by adults and children. Journal of the Acoustical Society of America, 96(2), 699-705.

Tingley, B. M., & Allen, G. D. (1975). Development of speech timing control in children. Child Development, 46, 186-194.

Tranel, D. (1992). Neuropsychological Assessment. Psychiatric Clinics of North America, 15, 283-299.

Wertz, R. T., LaPointe, L. L., & Rosenbek, J. C. (1984). Apraxia of speech in adults: The disorder and its management. Orlando, FL: Grune & Stratton.

Weschler, D. (1981). Manual for the Weschler Adult Intelligence Scale-Revised. New York: Psychological Corporation.

Wieneke, G., & Janssen, P. (1987). Duration variations in fluent speech of stutterers and nonstutterers. In H. F. M. Peters, & W. H. Hulstijn (Eds.), Speech motor dynamics in stuttering (pp. 345-352). Wien: Springer-Verlag.

Wieneke, G., & Janssen, P. (1991). Effect of speaking rate on speech timing variability. In H. F. M. Peters, W. H. Hulstijn, & C. W. Starkweather (Eds.), Speech motor and stuttering (pp. 325-331). Amsterdam: Elsevier.

Yamadori, A., & Ikumura, G. (1975). Central (or conduction) aphasia in a Japanese patient. Cortex, 11, 73-82.

Acknowledgments

The research reported in this paper was supported by NIH (NIDCD) grant #DC90076 (NCVS) and by NIH (NINDS) grant #NS19632. Portions of this manuscript were presented at the 24th Clinical Aphasiology Conference in Traverse City, Michigan (June, 1994). The token-to-token variability data from the individuals with apraxia have been presented in a different form with different analyses and for different theoretical reasons in Seddoh et al. (in press). We wish to thank the participants at the conference for their insightful questions, particularly Mick McNeil, Joe Duffy, Terry Wertz, Patrick Doyle and Jay Rosenbek. We appreciate the excellent comments on an earlier version of the manuscript by two anonymous reviewers. We also appreciate the secretarial and logistic support of Mary Jo Yotty.

Appendix

Perceptual Analysis: Paragraph Reading Task

SWEPT OFF MOUNT McKINLEY

Winds gusted to gale force, creating a near whiteout on the upper reaches of the highest mountain in North America. Guide Dave Staley, 35, and his party of eight climbers camped for the night of May 25, 1989. They made camp at 16,400 feet on a small icy shelf. The next day they would press on in their attempt to reach the 20,300-foot summit.

The men set their three dome-shaped nylon tents about a foot into the shelf's surface, then built a two-foot wall around each one with chunks of snow. Staley, who had climbed Mt. McKinley 16 times, preferred higher walls, but the ice was like concrete. He compensated with metal stakes at all guy lines and anchor flukes driven hard through the floor loops - 12 to a tent. We're bombproof, he thought as he zipped into his sleeping bag for the night.

By morning, the storm had not abated. So all that day - and the next - the group stayed put, listening to the nylon snap in the wind.

At about 9 p.m. that Saturday, John Richard, 39, a Kentucky lawyer and an ardent mountaineer, checked the anchors one last time before entering his tent. His companions, experienced climbers Howard Tuthill, 31, and Jim Johnson, 41, were reading in the long Alaskan daylight.

Richard suddenly raised his head. Coming through the whistle of the wind was a roar, louder and louder. He felt a blast of it against the tent roof, then popping sounds as wind thundered down the sides. "We're moving!" Tuthill shouted. The tent pivoted, heeled over, then began sliding. Richard dived for the zippered door but fell short, somersaulting as the tent spun over the side of the ridge.

In the same instant, Johnson, slashing at the nylon wall with a knife, was slammed onto his back.

As the tent gathered speed down the slope, men and equipment tumbling inside, Richard's hands touched snow through the gap opened by the knife. Clawing, he tried to drag his body clear. A sound of ripping nylon and he was outside, on his face, skidding downhill.

When he came to rest, deep in snow, he was alone. The tent with Tuthill and Johnson inside had rocketed off into the milky gloom toward Peters Glacier, 1500 feet below.

---(Readers' Digest, May 1991)

NCVS Status and Progress Report - 9
April 1996, 93-98

An Animal Model Comparing Needle and Hooked Wire Electrodes for Laryngeal EMG Recordings

Debra Jaffe, M.D.
Department of Otolaryngology Head and Neck Surgery, The University of Iowa College of Medicine
Nancy Solomon, Ph.D.
Department of Communication Disorders, The University of Minnesota
Robert Robinson, M.D., Ph.D.
Department of Pathology, The University of Iowa College of Medicine
Henry Hoffman, M.D.
Department of Otolaryngology Head and Neck Surgery, The University of Iowa College of Medicine
Erich Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

The use of a specific electrode type in laryngeal electromyography (EMG) has not been standardized. Laryngeal EMG is usually performed with hooked wire electrodes or concentric needle electrodes. The hooked wire electrode is thought to be more stable during physiologic laryngeal movement. Signal consistency over time and induced tissue injury of the two electrode types were compared. Electrodes were percutaneously placed into the thyroarytenoid muscles of 10 adult dogs. Amplitude of EMG activity was measured and compared during vagal stimulation preceding and following large scale laryngeal movements. Animals were euthanized and vocal fold injury was graded and compared histologically.

Waveform morphology did not consistently differ between electrode types. The variability of EMG amplitude was greater for the hooked wire electrode ($p < .05$) while the mean amplitude measures preceding and following large scale laryngeal movements did not differ ($p > .05$). Inflammatory responses and hematoma formation were also similar.

This study demonstrates that there is no difference between the hooked wire electrode and the concentric needle electrode in terms of vocal fold injury or electrode stability in the TA muscle following large scale laryngeal movements.

Introduction/Background Information

Laryngeal electromyography (EMG) is a tool whose utility is well documented in studying laryngeal motor activity and facilitating treatment of voice and swallowing disorders. There are many types of electrodes utilized in performing laryngeal EMG. Surface electrodes are placed on the overlying skin or mucosa of the larynx and do not insert into muscle. These electrodes are used infrequently for laryngeal EMG recordings as individual muscles are cluttered with signals from adjacent muscles (Miller & Rosenfield, 1984; Kuna et al., 1990). In addition, the high frequency components of the EMG signal are not captured with surface recordings (Kimura, 1989).

The vast majority of laryngeal EMG recording is conducted with intramuscular electrodes. One type of intramuscular electrode is the needle electrode for which there are several varieties. The standard or concentric type of needle electrode utilizes a wire in the center of the needle shaft. Differential electrical activity is recorded between the exposed end of the wire at the tip of the needle and the uninsulated shaft of the needle. A separate electrode is used as a reference electrode. The bipolar concentric needle electrode employs two fine stainless steel or platinum wires in a fixed location that serve as electrodes in the core of the needle. The EMG signal is generated by the electrical

activity of the small number of muscle fibers between the wires, while the shaft of the needle serves as the reference electrode (Miller & Rosenfield, 1984; Kimura, 1989).

The second basic type of the intramuscular electrode is made by a pair of very fine insulated wires with a bared "hook" on the end (the "hooked wire" electrode). Hooked wire electrodes for laryngeal EMG were first described by Hirano and Ohala in 1969. Thin metal wires are threaded through a hypodermic needle which is used to insert the wires into the muscle. The tips of the wires are stripped of insulation and are reflected back until almost parallel with the shaft of the hypodermic needle. The needle and wires are inserted into the muscle and the needle is slowly withdrawn, allowing the ends of the wire to catch within the tissue (Hirano & Ohala, 1969). The EMG activity of the muscle fibers between the two wires is recorded.

Hirano and Ohala (1969) suggested that the primary advantage of the hooked wire electrode over the concentric needle electrode is its propensity to remain in place, particularly during movement of the larynx. Once positioned, subjects experience minimal discomfort and relatively uninhibited speech (Hirano & Ohala, 1969). The advantages of the hooked wire electrodes are not accepted unanimously and remain to be demonstrated empirically. Many physicians use concentric needle electrodes in their clinical practices as some believe hooked wire electrodes lack the rigid standardization required for the quantitative measures of the amplitude, latency and waveform of action potentials (Miller & Rosenfield, 1984; Stalberg & Trontelj, 1979).

It is possible that large movements of laryngeal muscles with rigid needle electrodes in place may produce intramuscular damage. Similarly, it is conceivable that hooked wire electrodes may damage muscles by shearing fibers as the electrodes are removed (by pulling them out). Histologic examination of muscles in which the concentric needle and the hooked wire are utilized have yet to be conducted.

The purpose of this study was to compare the performance of concentric needle electrodes with hooked wire electrodes. Every attempt was made to model the typical use of these electrodes in the clinical evaluation of the human larynx. The thyroarytenoid (TA) muscle is frequently studied in man, and was therefore the muscle evaluated. Specific objectives included:

1) To determine whether bipolar hooked wire electrodes permit more stable recordings than (standard) concentric needle electrodes at rest and following large movements of the larynx and surrounding structures.

2) To test whether bipolar hooked wire electrodes provide comparable information regarding waveform and amplitude of action potential as the standard concentric needle electrodes.

3) To compare the damage resulting from placement and removal of the two types of electrodes by histologic evaluation of the vocal folds.

Methods

All techniques were approved by the University of Iowa Animal Care Facility and were in accordance with the Guidelines for the Care and Use of Laboratory Animals.

Ten adult mongrel dogs of either sex weighing approximately 20 kg were utilized. Each dog was initially sedated with intramuscular ketamine (7.5 ml/kg) with xylazine (2.5 mg/kg). Intravenous access was established for hydration and anesthesia (Nembutal; 25 mg/kg).

The head was stabilized in a stereotaxic instrument that was fixed to a table. The animal was placed in the supine position and a brace was inserted into the mouth to provide direct visualization of the larynx.

With meticulous attention to preserving the skin and soft tissue over the larynx, two vertical incisions 3 cm in length were created bilaterally along the lateral edges of the trachea at the level of the fifth tracheal ring. The vagus nerves were carefully dissected and cuff electrodes were placed around the nerve for electrical stimulation of axons going to the thyroarytenoid muscle.

A bipolar hooked-wire stainless-steel electrode (constructed using .002 inch bifiler wire by California Fine Wire company, and a 25 gauge needle) was percutaneously placed into one thyroarytenoid muscle through the intact skin and soft tissue overlying the cricothyroid membrane. On the contralateral side, a standard disposable 26 gauge concentric needle electrode (Dantec) was similarly placed into the thyroarytenoid muscle. Confirmation of location of electrode within the TA muscle was determined by noting movement of the vocal fold under direct laryngoscopy during electrode placement and appropriate EMG activity upon vagal stimulation. The electrode was successfully placed within the TA muscle on the first attempt in all instances but one. A reference needle electrode was placed in the subcutaneous tissue of the axilla. Placement of the TA electrodes was confirmed by visualizing vocal fold and arytenoid adduction upon vagal stimulation (0.2 ms pulses, 2 pps, 0.3 ma nerve parameters). Once threshold parameters were obtained, supra maximal levels of stimulation were determined (usually 0.2 ms pulses, 2 pps, and 1 ma).

Each vagus nerve was stimulated independently ten times with single supramaximal 0.2 ms pulses ten times, evoking TA activation without appreciable movement of the laryngeal framework visualized by direct laryngoscopy. After a five minute rest period, the sequence was repeated. At least three sets of ten successive stimuli to the vagus nerve bilaterally were administered in each study. The EMG responses from each TA muscle were recorded as baseline measurements.

A second set of hooked wire electrodes was placed through the intact skin and soft tissue into the thyrohyoid muscle. The thyrohyoid muscle was stimulated (2 s, 2 ms pulses, 0.2-0.4 ma, 50 Hz), producing a large scale movement of the larynx and surrounding structures, simulating laryngeal elevation during a swallow. In each instance, laryngeal elevation was noted to be at least 2 cm.

Each vagus nerve was once again independently supramaximally stimulated (0.2 ms pulses, 2 pps, 1 ma) ten times evoking TA activation. Five cycles of the following procedure were performed after baseline measurements were obtained:

1. (m1) Stimulation of the thyrohyoid muscle resulting in laryngeal elevation.
2. (a) Stimulation of each vagus nerve 10 times while recording EMG responses.
3. A five minute rest interval (no stimulation).
4. (b) Stimulation of each vagus nerve 10 times while recording EMG responses.

The electrodes were removed. The animal was euthanized by Nembutal (80 mg/kg/kg). The larynx was then excised and fixed in 10% formalin. The larynx was sectioned coronally through the TA muscles. Laryngeal sections were embedded in paraffin, cut on a microtome and stained with hematoxylin and eosin. Specimens in which the electrode was placed successfully into the TA upon the initial attempt were examined for damage to muscle fibers, hemorrhage and inflammation.

Histologic analysis was conducted on 18 specimens (2 TA muscles for each of 9 dogs). One TA muscle was eliminated from analysis because more than one attempt had been required to place the electrode accurately. The TA muscle from the contralateral side was eliminated to avoid an incomplete block design. Specimens were examined for damage to muscle fibers, hemorrhage and inflammation. Each specimen was rated by two examiners (DMJ and RAR) on a scale of 0 (normal) to 4 (extravasation of red blood cells, with dissection along muscle fascicles and infiltration of the region with polymorphonuclear leukocytes). Percent agreement within one point on the scale between the two examiners were found to be 83%. Both investigators who performed the histological analysis are physicians; one (RR) is a board certified pathologist with more than 15 years of experience.

Data Analysis

Because of technical difficulties with the digitized signal, EMG data were obtained and analyzed from only 8 of the 10 dogs. EMG signals registered by the electrodes were amplified and filtered (using an eight-pole elliptical filter with a cut off frequency of 2500 Hz). During data collection, the EMG signals were simultaneously transmitted to an audio speaker for auditory monitoring and a computer monitor for on-line visual monitoring. EMG data were digitized into a desktop computer using data acquisition and analysis software (DATAQ Instruments, Inc.; Akron, OH), at a sampling rate of 5000 Hz per channel. Three signals were digitized: the stimulus signal, EMG signal from the TA muscle with the hooked wire electrode, and EMG from the TA muscle with the needle electrode. Each EMG response from vagal stimulation was identified and marked for further analysis. The data acquisition software was used to calculate peak to peak amplitude and the sum of squares for each responses. Means and standard deviations of peak to peak amplitudes of EMG waveforms of each set of responses (10 responses per set) were computed for further statistical analysis.

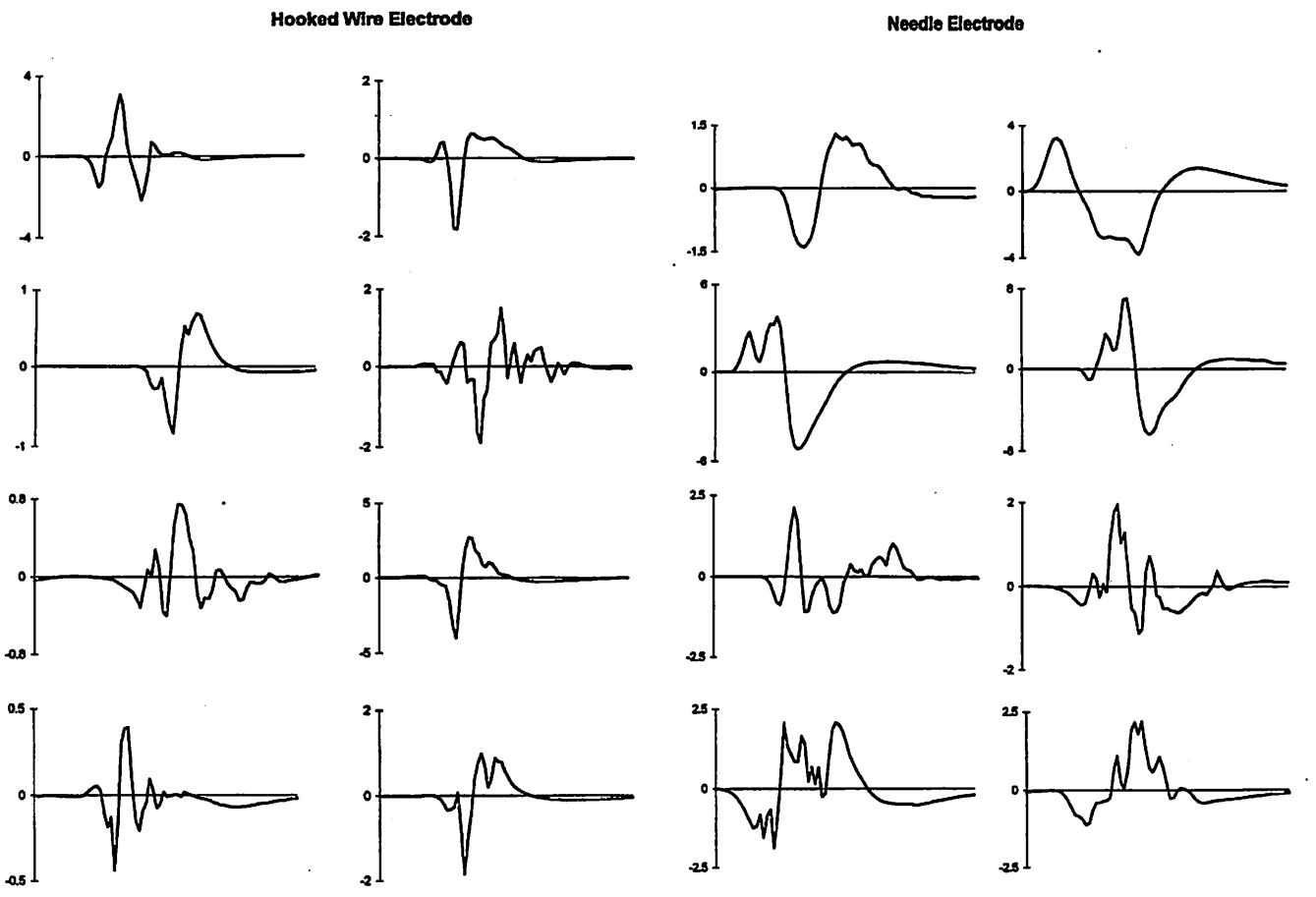
Results

Typical waveforms from each of the two types of electrodes for each dog are illustrated in Figures 1 and 2 (following pages). The morphology of waveforms from both the hooked wire electrode and the standard concentric needle electrode appear to be complex multiphasic action potentials. Sequential single motor unit action potentials cannot be delineate in these waveforms because the EMG activity recorded is gross EMG resulting from vagal nerve stimulation.

The standard deviation of peak-to-peak amplitude (in mV) was examined from each set of 10 vagal stimulations. An analysis of variance model (ANOVA) revealed a significant larger variability of peak to peak amplitude recorded from the hooked wire electrode ($p < .05$), whereas the variability of peak to peak amplitude recorded from the concentric needle electrode was smaller and not statistically significant ($p > .05$). Further support of this finding is provided in Table 1, which demonstrates a larger variability in peak to peak amplitude across all dogs and all conditions for the hooked-wire electrode.

Table 1.
Average Standard Deviation of Amplitude (mV)
This table demonstrates the averaged standard deviation in mV of all vagally induced contraction of the TA muscles for each dog and each electrode. All of the standard deviation measures were then averaged, further demonstrating the larger variability of peak to peak amplitude of the hooked wire in comparison to the concentric needle electrode.

Canine	Hooked Wire	Concentric Needle
1	0.189440	0.141418
2	0.304400	0.192401
3	0.314618	0.545357
4	0.359500	0.201956
5	0.092517	0.201307
6	0.702649	0.397095
7	0.347699	0.325995
8	0.405077	0.286285
Average	0.339488	0.286477



Figures 1 and 2. These figures demonstrate typical waveform morphologies of EMG signals registered from the hooked wire electrode (Figure 1 - left) and the needle electrode (Figure 2 - right) from each dog in which data was collected. Each X axis represents 150 ms. Each Y axis demonstrates peak to peak amplitude in mV.

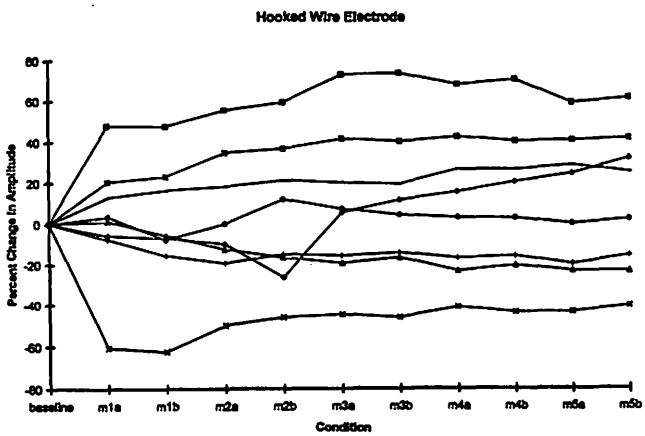
Peak-to peak amplitude of the EMG waveforms were averaged for each set of 10 vagal stimulations. The three or four sets of responses obtained before electrically evoked laryngeal elevation were compared to the 5 sets of responses obtained after laryngeal elevation. The percent change in amplitude was calculated from the sets of responses following laryngeal movement. These results are plotted for each TA muscle studied in Figure 3 for the hooked-wire electrode, and Figure 4 for the needle electrode. Amplitude of the EMG waveform decreased, increased or stayed approximately the same for several TA muscles for both types of electrodes. Statistical analysis revealed no significant tendency in amplitude to increase of decrease after laryngeal elevation for either type of electrode (by ANOVA $p<.05$ for both the hooked wire electrode and the concentric needle electrode). It is obvious, however, that large changes occurred in some studies following laryngeal movements. Such changes, when they occurred appeared to be larger with the hooked wire electrode.

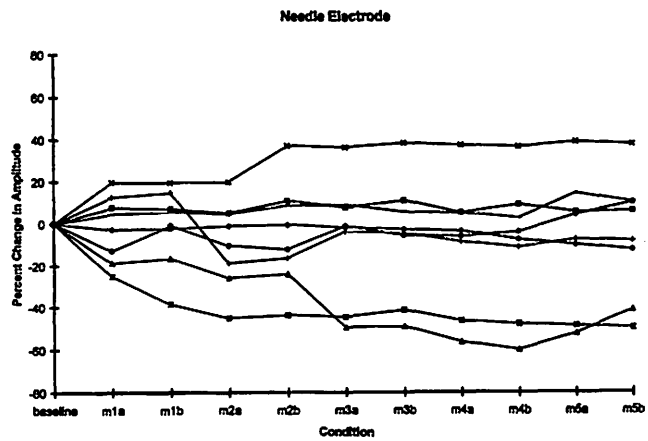
The average peak-to-peak amplitudes measures from the two sets of movement (movement 1a and movement 1b) obtained after electrically evoked laryngeal elevation can be compared. A five minute interval between the two sets of ten successive vagally induced TA excitation did not result in statistically significant differences in amplitude measures of peak to peak amplitude for both the hooked wire electrode and the concentric needle electrode ($p=.42$ for the hooked wire electrode and $p=.64$ for the needle electrode by ANOVA).

Histologic Results

Ratings of histologic preparations of the TA muscles, averaged across the 2 judges are listed in Table 2.

A Wilcoxon signed-ranks test revealed no statistically significant difference between the ratings of the histologic specimens from the hooked wire electrode and the concentric needle electrode ($v=6$, $p=.6858$, 2-tailed). Four of the TA muscles which utilized the concentric needle electrode, but only one of the specimens from the hooked-wire electrode was rated as normal.





Figures 3 and 4. These figures demonstrate the change in amplitude over time for all eight dogs in which data were collected for the hooked wire electrode (Figure 3 - upper) and the concentric needle electrode (figure 4 - lower). The average of all baseline measures was computed. The baseline measures were then normalized to zero. Each Y axis demonstrates the percentage change in amplitude. Each X axis illustrates the conditions for in which measures were made. See methods section for explanation of sequence of events represented by these figures. Note, although the specific averaged points are connected with a line the data obtained was not continuous, as there was a five minute rest interval between the two measures obtained after each electrically evoked laryngeal movement. The line is used for clarity.

Discussion

This study was conducted in an attempt to determine whether electromyographic recordings from hooked wire electrodes or concentric needle electrodes are more stable in the TA muscle after electrically evoked laryngeal elevation. An in vivo canine model was used. This study was designed to model the clinical use of these electrodes in the human larynx.

The finding that the morphology of the waveforms recorded from the two types of electrodes was similar (i.e. both appeared to be complex multiphasic action potential) indicates that both electrodes can represent EMG activity in a similar fashion despite the differences in physical characteristics of electrodes. The bipolar hooked wire electrode is composed of two small electrodes with the differential between the two being registered. The EMG activity recorded is a result of electrical activity occurring in a small region of myofibrils between the two electrodes. The waveform recorded from the concentric needle electrode is the result of the EMG activity being registered from a much larger region. The differential between the tip of the needle and the shaft is measured and recorded (Kimura, 1989). Both electrodes allow for a large range of recordings after nerve induced muscle contraction.

Table 2.
Histologic Results
This table demonstrates the ratings of histologic preparations of the TA muscles, averaged across the 2 judges.

Canine	Hooked Wire	Concentric Needle
1	2.5	1.5
2	1.5	1.5
3	1.5	2.5
4	0	0
5	2.0	0
6	1.5	3.0
7	0.5	2.5
8	1.0	0
9	3.0	0
Average	1.5	1.2

The peak-to-peak amplitudes of the EMG signals recorded by the hooked-wire electrode were more variable when averaged across all dogs and all conditions. This finding for the data from the hooked-wire electrode may be attributed to differences in impedance from tissue properties secondary to red blood cell extravasation or tissue edema. However, it may also be due to the more circumscribed area of recording used by the bipolar hooked-wire configuration. That is to say, the amplitude of waveforms may have been just as variable in the muscle with the needle electrode, but the electrode may not detect such subtle differences (see Table I).

Although there were changes in amplitude measures as large as 60-80% for EMG data recorded from both the hooked wire electrode and the concentric needle electrodes after laryngeal elevation, the differences were not statistically significant. Upon repeated testing, variable increases and decreases in amplitude relative to the averaged baseline measures were observed for recordings employing both types of electrodes. The amplitude of EMG activity appeared to stabilize after the five electrically evoked contraction of the thyrohyoid muscle regardless of electrode type. We have altered our practice of EMG recordings based on this finding. Now, utilizing bipolar hooked wire elec-

trodes, prior to obtaining final EMG recordings, patients are instructed to swallow five times so as to allow for the electrode to settle into a stable position.

Use of either hooked wire or concentric needle electrodes resulted in a similar degrees of tissue injury as assessed through microscopic evaluation of the TA muscles. Concentric needle electrodes in the TA muscles resulted in uninjured muscle more often (4/9 in our study) in comparison to hooked wire electrodes (1/9). This finding must be interpreted cautiously as the sample size in this study was small. Yet when injury to the TA muscle in which a concentric needle electrode was used, it tended to be more severe in comparison to the injury detected in TA muscles in which the hooked wire electrode was used. The most severe injury observed involved hematoma formation with a small amount of dissection along muscle fascicles. This degree of injury was noted in one specimen for each electrode type. Of importance, none of the injuries observed were believed to be unrecoverable in live tissue or result in functional impairment of the muscle.

One issue that was not addressed in this study but is clinically relevant is the patient's comfort during the use of intramuscular laryngeal electrodes. From our clinical as well as personal experiences, the hooked wire electrode allows laryngeal movement with ease and comfort, as the wires are pliable, whereas laryngeal movement with rigid needles in place is significantly less comfortable. If, as this study suggests, differences in the objective characteristics and quality of EMG recordings and damage to muscle tissue are not markedly different depending on electrode type, then the otolaryngologist may choose to select hooked-wire electrodes for evaluation of intrinsic laryngeal muscle function to optimize patient comfort.

Similarly, the technical aspect of placing the electrode accurately in muscle was not addressed in this study. With a hooked wire electrode the clinician has only one attempt in placing the electrode. If the electrode is inaccurately placed, it is removed and placement is attempted again. In contrast, a needle electrode may be advanced, withdrawn or redirected in an attempt to place the electrode appropriately.

Conclusion

Waveform morphology of EMG signals registered from hooked wire electrodes and concentric needle electrodes resulted in similar complex action potentials when tested in canine TA muscles. No difference between the hooked wire electrode and the concentric needle electrode was detected in terms of electrode stability or vocal fold tissue injury following large scale laryngeal movements. The highly variable amplitude of gross EMG registered from the hooked wire electrode is likely to have resulted from the smaller electrical field sampled in comparison to that recorded by the concentric needle electrode. Based on these results of this study, either hooked wire or concentric needle electrodes appear to provide adequate and accurate indications of laryngeal muscle activity.

Acknowledgments

This research was supported by Grant No. P60-DC00976 from the National Institute on Deafness and Other Communication Disorders and Grant No. DC 00040 from the National Institutes of Health Research Training Program for Otolaryngology.

References

Blair, R. L. (1989). Laryngeal electromyography. *Archives of Otorhinolaryngology*, 246: 395-296.

Hirano, M. and Ohala, J. (1969). Use of hooked - wire electrodes for electromyography of the intrinsic laryngeal muscles. *Journal of Speech and Hearing Research*. 12: 362-373.

Kimura, J. *Electrodiagnosis in Diseases of Nerves and Muscles: Principles and Practice*. Philadelphia, Pennsylvania: A. Cavis Company (1989).

Kuna, S. T., Smickley, J. S., Insalaco, G., Woodson, G. E. (1990) Intramuscular and esophageal electrode recordings of posterior cricoarytenoid activity in normal subjects. *Journal of Applied Physiology*. 68 (4):1739-45.

Miller, Robert H. and David B. Rosenfield. (1984). The role of electromyography in clinical laryngology. *Otolaryngology - Head and Neck Surgery*. 92: 287-291.

Stalberg, E., and Trontelj, J.V. *Single Fibre Electromyography*. Old Woking, Surrey, UK: The Miravalle Press Limited (1979).

Part II

Tutorial Reports

NCVS Status and Progress Report - 9
April 1996, 99-107

Detecting Bifurcations in Voice Signals

Hanspeter Herzel, Ph.D.
Institute of Theoretical Physics, Technical University, Berlin, Germany
Joachim Holzfuss, Ph.D.
Institut fur Angewandte Physik, Technische Hochschule, Darmstadt, Germany
Zbiegniew Kowalik, Ph.D.
Institute of Experimental Audiology, University of Munster, Germany
Bernd Pompe, Ph.D.
Institute of Physics, E.-M.Arndt—University, Greifswald, Germany
Robert Reuter, M.S.
Institute of Electronics, Technical University, Berlin, Germany

The Voice —A Highly Nonlinear Oscillator

The composites of speech production are respiration, phonation, and articulation. The respiratory airflow, typically from the lungs via the bronchi and trachea, serves as the main driving force. Articulation is controlled by the vocal tract which can be regarded as an approximately linear filter [1]. The resonance frequencies (termed formants) are governed by the oral and nasal cavities.

In this paper we focus our attention on phonation — the generation of the primary voiced sound within the larynx. The vocal folds are set into vibration by the combined effect of subglottal pressure, the visco-elastic properties of the folds, and the Bernoulli effect [2,3]. The effective length, mass and tension of the vocal folds are determined by muscle action, and in this way the fundamental frequency ("pitch") and the waveform of the glottal pulses can be controlled.

For sustained vowels the driving lung pressure and the muscle tensions can be regarded as slowly varying parameters since they change on time scales of a few hundred milliseconds whereas vocal fold vibration cycles have periods of only a few milliseconds. More details about the mechanisms of phonation are provided in section 3 in connection with aerodynamic-biomechanical modeling.

Limit cycle oscillations can be considered as a reasonable model of a normal healthy voice, although some small perturbations (in the order of a percent) are always present. However, under certain circumstances much larger irregularities are observed in vocalizations. These are often associated with the term roughness. Also a healthy vocal apparatus can generate a rough voice quality under extreme conditions. Examples are newborn cries [4,5], simulated creaky voice [6,7,8], or Russian lament [9]. Bifurcations to subharmonic regimes, toroidal oscillations, and chaos have been reported in these extreme vocalizations.

Of particular interest are vocal instabilities due to organic, neurological, and functional diseases [10,11,12,13]. It has been shown that subharmonics, low-frequency modulations, and biphonation (two independent pitches) are often symptomatic in voice pathology [14,15,16]. An understanding of the underlying physiological and physical mechanisms will certainly be helpful for diagnosis and treatment of voice disorders.

Our paper is organized as follows: In the next section five techniques of sliding signal analysis are introduced that are applied afterwards. These methods include conventional tools of voice research (spectrograms, pitch contours) as well as novel nonlinear techniques (generalized mutual information, dimensions, and Lyapunov exponents). Sections 3 and 4 are devoted to modeling: W e introduce the two-mass model — the most simple but physiologically reasonable model — and discuss the most sophisticated continuum models. Computer simulations of both models are analyzed with our sliding analysis techniques. In section 5, bifurcations in an excised larynx experiment are discussed. Finally, the methods are applied to vocalizations of

patients with certain voice disorders (papilloma and laryngitis).

It turns out that the applied methods provide some insight into transition phenomena in computer-generated and natural voice signals. The results of the different techniques are consistent and complement each other.

Sliding Analysis Techniques

In this section we describe signal analysis techniques that are applied in the remainder of this paper. The spectrogram technique — a sliding power spectral analysis — is of widespread use in speech research. It allows the simultaneous characterization of the local (in time) signal properties, e.g. the pitch and the formants, and the long-term changes during utterances. For slowly varying external parameters as subglottal pressure and muscle tension, spectrograms can be regarded as spectral bifurcation diagrams. Such diagrams have been applied also to bifurcation analysis in acoustic cavitation [17], heterogeneous catalysis [18], and laser models [19].

In order to detect bifurcations in sustained phonation, narrow band spectrograms are required. Typically we use windows of about 200 ms giving a spectral resolution of 5 Hz. Then the (dis)appearance of spectral peaks due to bifurcations of the underlying dynamical system can be monitored. In this way subharmonics (related to period-doubling or -tripling), side bands (a manifestation of toroidal modulations), and independent frequencies have been identified by spectrograms in voice signals [5,16,20,21].

Instead of using power spectra one can study also correlation measures. For example, short-term autocorrelation functions contain the same information as the corresponding spectra. However, correlation functions detect only linear statistical dependences. Therefore in nonlinear systems the mutual information has been applied frequently [22,23,24]. This function vanishes, if and only if, signals are statistically independent [25,26]. However, the estimation of the mutual information from finite realizations is problematic [27]. Recently a generalized mutual information has been introduced [28] which is based on correlation integrals. Therefore, it can be estimated much more easily and with moderate computational effort. Consequently, also a sliding analysis of signals can be performed which has been termed "miogram" [29]. Some more details about this technique are given in figure captions and in another chapter of this volume [30].

It has been argued in the preceding section that sustained phonation of normal healthy voices leads to nearly periodic signals. Deviations from periodicity are often symptomatic of voice pathologies. Therefore various "perturbation measures" have been developed to quantify irregularities. Many of these measures are based on "pitch contours", i.e. sequences of consecutive estimations of the local pitch period. Since the pitch originates usually from the larynx, diseases of the voice generating system are effectively studied by pitch contours. In a sense, these contours are comparable to RR-interval series in cardiology. Period-doubling is easily recognized in contours as alternating values, and low-frequency modulations lead to oscillating pitch contours [7,16,31].

Since many voice signals reflect low-dimensional attractors it is natural to estimate dimensions and Lyapunov exponents. This has been done in Refs. [5,14] for some vocalizations. However, the well-known non-stationarities on time-scales of a few hundred milliseconds prevent precise estimations. In this paper we apply nevertheless related techniques to our data. In order to allow comparison to spectrograms and miograms we calculate local (in time) estimations of dimensions and Lyapunov exponents. We are convinced that these estimates are very crude estimators of the actual invariants due to the short windows and non-stationarities. However, they exploit quite different signal properties than spectrograms since they are based on phase space reconstruction and, hence, they reflect attractor properties instead of temporal correlations.

Estimates of the correlation dimension are obtained by averaging local densities (see [32,33,34,35] for details). In the resulting log-log plot (correlation integral versus distance) scaling regions are detected automatically [33]. In this way a fast sliding dimension analysis of also rather long voice signals are possible (a signal of 200,000 data points can be processed in a few minutes).

Lyapunov exponents measure the rate of separation of nearby states and are related, therefore, to the predictability of the system. There are several methods to estimate Lyapunov exponents from long stationary time-series [36,37,38,39].

It has been pointed out [40,41] that often also local stability measures [42,43] provide useful information about the dynamics. Deterministic mechanisms of physiological processes allow us to state that at least in the short term, the processes exhibit quasistationary character. The observation of dynamical measures in short time-slices will then characterize their momentary state, and the observation of the temporal evolution of such measures provide information about the development of the physiological system. Combining the local divergence measure with a windowing technique should give us a very useful tool for measurement of global system dynamics.

It will be shown that in view of the intrinsic non-stationarity of voice signals sliding analysis techniques as introduced above are appropriate tools. We demonstrate below that non-stationarities also have an advantage. Slowly varying parameters may in duce several bifurcations within a single vocalization. It is not seldom that limit cycles, subharmonic regimes, tori, and chaotic episodes can be detected in a single newborn cry [4,5] or in a sustained vowel [16,21,31].

A Two-Mass Model of the Vocal Folds

Any phonation model has to include aerodynamical and biomechanical components. In principle, the Navier-Stokes equations with time-varying boundaries together with nonlinear and inhomogeneous visco-elastic equations should be solved. However, much insight has been gained with the aid of simplified models, such as a two-mass model [13,44,45,46,47] or a 16-mass model [48]. Recently we have reduced the intensively studied Ishizaka-Flanagan model to its very basic features [7,49]. Such drastic simplifications al low extensive bifurcation analysis.

In the following we present first the symmetric model version in order to demonstrate the underlying physical mechanisms of phonation. Then a time-series from the asymmetric version will be analyzed.

In two-mass models both folds are modeled by a lower and upper mass m_1 and m_2, respectively. Their elongations x_i (i=1,2) are governed by the usual mechanical equations with spring constants k_i and c_i, damping r_i and coupling k_c:

$$m_i\ddot{x}_i + r_i\dot{x}_i + k_i x_i + \Theta(-a_i)c_i\left(\frac{a_i}{2l}\right) + k_c(x_i - x_j) = F_i(x_1, x_2). \quad (1)$$

The θ-function is related to an additional restoring force during closure of the glottis ($a_i=a_{i0}+2lx_i < 0$; a_{i0}: rest area; l: length of the glottis). The driving forces F_i can be derived as follows: We assume constant pressure below the glottis (termed subglottal pressure P_s) and above the glottis (vocal tract input pressure $P_i=0$). Moreover we assume that at the point of minimum area a_{min} a jet is formed which induces an immediate pressure decay to zero. Consequently, the driving force of the upper mass F_2 is identically zero for all glottal configurations. $F_1 = ld_1 P_1$ (d_1: thickness of the lower mass) is the force exerted by the pressure P_1 on the lower part of the glottis. The corresponding pressure P_1 can be obtained from the Bernoulli law:

$$P_s = \frac{\rho}{2}\left(\frac{U}{a_{min}}\right)^2 = P_1 + \frac{\rho}{2}\left(\frac{U}{a_1}\right)^2 \quad (2)$$

Here ρ and U denote the air density and the glottal volume flow, respectively. Using

$$U = \sqrt{\frac{2}{\rho}\frac{P_s}{a_{min}^2}} \quad (3)$$

one obtains

$$P_1 = P_s\left[1 - \Theta(a_{min})\left(\frac{a_{min}}{a_1}\right)^2\right]\Theta(a_1). \quad (4)$$

It can be seen that for a convergent glottis ($a_1>a_{min}=a_2$) the pressure is reduced by the Bernoulli term. This points to the essential mechanism of vocal fold vibrations: Normally the lower mass opens first and, hence, the full subglottal pressure pulls apart the vocal folds. After some delay (about 60 degrees phase shift) the upper mass pair opens leading to an increasing flow. Consequently, during the closing phase the pressure P_1 is reduced according to Eq.(4). This pressure asymmetry between opening and closing phase constitutes the main driving force of vocal fold vibration in the chest register. The described interaction between flow and geometry allows the energy transfer from the vertical air flow to the vocal fold motion. The described wave-like motion of the vocal folds is found in deed in stroboscopic and high-speed observations. A more detailed justification of the various simplifying assumptions (laminar flow, no viscous losses, constant sub- and supraglottal pressures, linear restoring force, ...) can be found elsewhere [49]. A representative set of parameters is given in the Appendix.

Bifurcations in the symmetric model version have been studied in [7,49]. As an attempt to model unilateral paralysis we analyzed in [20,50] an asymmetric version of the two-mass model. For this purpose the masses and elastic constants of one fold have been scaled as follows using an asymmetry parameter Q:

$$\begin{aligned} k_i^{right} &= Q\, k_i^{left} \\ c_i^{right} &= Q\, c_i^{left} \\ m_i^{right} &= m_i^{left}/Q \\ 0 < Q &< 1 \end{aligned} \quad (5)$$

In this way the eigenfrequency of the affected fold is reduced by the factor Q. Instabilities have been found for Q<0.6 and subglottal pressure above P_s=0.013 [50]. Typically, at the borderline of normal phonation abrupt jumps to subharmonic regimes are observed [50]. Figs. 1 and 2 show a time-series and a miogram for a stepwise decrease of Q from 0.56 to 0.5. It can be seen that the pitch period increases with decreasing Q.

At around 40,000 data points suddenly subharmonics appear at one-half of the pitch. At 46,000 another complicated subharmonic regime is reached. Inspection of the peaks of the elongations x_1^{right} and x_1^{left} reveals that during one cycle (about 5 times the original pitch period) five maxima of x_1^{right} and eight maxima of x_1^{left} occur, i.e. we can interpret this regime as a 5:8 resonance. From 50,000 subharmonics of one third of the original pitch ("period-tripling") dominate corresponding to a 3:5 resonance in the above sense. The sequence of decreasing ratios from 2:3 to 5:8 to 3:5 is reminiscent of "Arnold tongues" in bifurcation diagrams of coupled oscillators. Comprehensive bifurcation diagrams in the P_s-Q parameter plane also reveal the appearance of toroidal and chaotic oscillations [50].

At the beginning of the miogram in Fig. 2 a pitch period T_0 of about 120 data points (≈ 8ms) can be detected. Note that the dark band at about 60 points corresponds to one half of the period, i.e. to the minimum of the autocorrelation

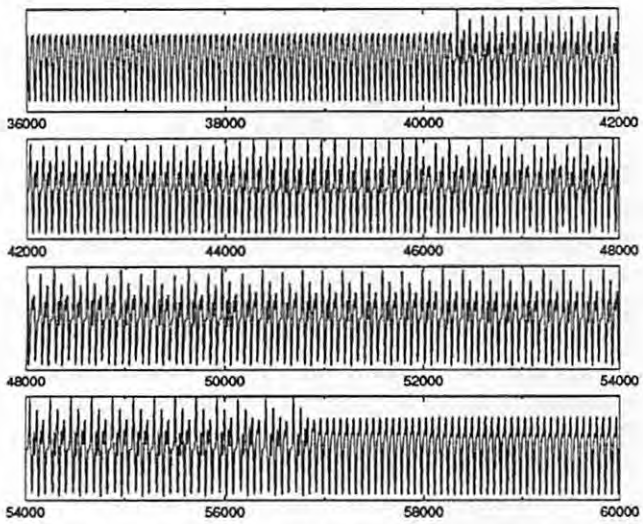


Figure 1. Smoothed time-series dU/dt for the asymmetric two-mass model. Every 2000 data points (400 ms) the asymmetry parameter Q was decreased by 0.005. In this way bifurcations to subharmonic regimes were induced.

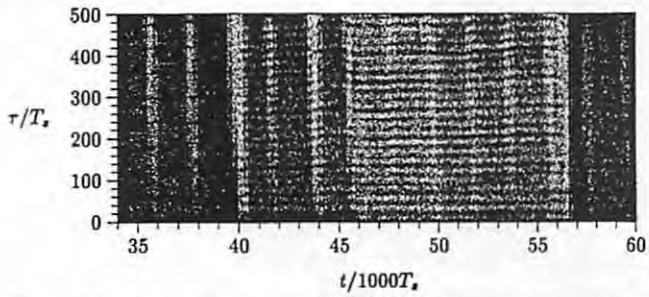


Figure 2. Miogram of the signal in Fig. 1 (window length: 1000; window shift: 250. For each window the generalized mutual information was estimated with a coarse graining of 5%. The amplitudes are encoded by a grey scale and, hence, dark horizontal bands correspond to peaks of the mutual information and indicate, therefore, periodicities.

function. This reflects the fact that the mutual information is related to the squared autocorrelations [26,28] and, hence, also negative peaks induce peaks in the mutual information function. Around 40,000 a jump to approximately the double period is visible (note the dark stripes around 130, 260, and 390 points). Then period five and period three can be detected in the miogram. Similar information can be obtained from the corresponding spectrogram (Fig. 5 in Ref. [20]) and the pitch contour [31]).

Despite the qualitative resemblance of our simulations to some observations in unilateral paralysis [13,12,20], we have to keep in mind that a 2-mass model is only a crude approximation of the real vocal folds. However, in simulations of a three-dimensional model based on partial differential equations, similar bifurcations to subharmonic regimes and chaos have been found [15,51]. Moreover, the calculation of empirical orthogonal functions from continuum models reveal that the dynamics is often governed by only

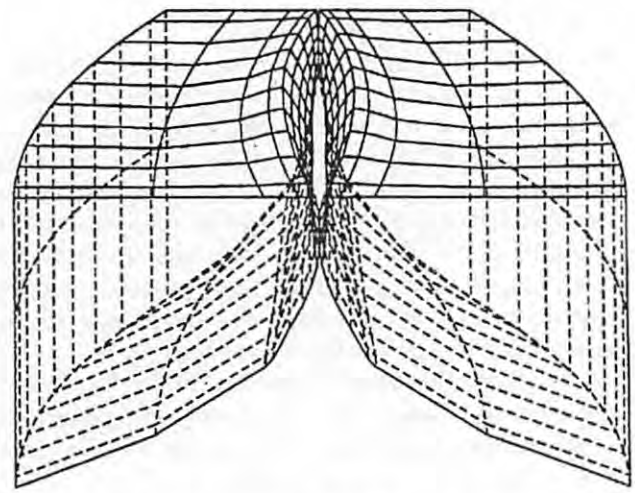


Figure 3. A view of the continuum model immediately before glottal closure with the posterior edges of the folds in the foreground.

a few dominant modes. This can be regarded as a justification to study specific aspects of vocal fold vibrations such as left-right asymmetry with appropriate low-order models.

A Continuum Model

The most sophisticated models of the voice have been developed in the last decade at the National Center for Voice and Speech of the USA in Iowa-City [15,51,52,53]. These models are based on finite element simulations of vocal fold tissue. They incorporate detailed knowledge of the geometry of the vocal folds and allow the modeling of the layered structure (mucosa, ligament, and muscle vocalis). The most recent version also includes the numerical solution of the Navier-Stokes equations [54]. However, these simulations are extremely time-consuming and we discuss, therefore, simulations based on the laminar flow hypothesis, i.e., the Bernoulli equation is used as above in the two-mass model. However, the visco-elastic equations are treated in much detail (414 nodal points). These finite element simulations allow a rather realistic modeling of the vibrations. In fact, visualizations of the computer-generated pattern resemble high-speed video observations very closely.

Fig. 3 gives an impression of the simulated vocal folds. The model contains nine longitudinal layers. Each layer consists of 32 triangular finite elements. Altogether, there are 207 nodes per fold free to oscillate, the others are placed on fixed boundaries. Details about the simulations of this continuum model can be found in [51,55]. In these papers we have studied bifurcations due to varying subglottal pressure and muscle tension. It is shown using empirical orthogonal functions that only a few of the 14 potential modes contain most of the variance of the vibrations. More precisely, for normal phonation two modes cover 98% of the variance, and also in the cases of subharmonic or chaotic

vibrations the dominant four modes contain more than 90% of the variance. It turned out that the empirical modes were quite similar to the modes of the linearized model. This indicates that despite the highly nonlinear excitation via aerodynamical forces the excited modes are governed by the linear visco-elastic properties. The complex vibratory patterns due to decreasing stiffness [51] can be considered as a model of vocal fry phonation [7,8].

In the following we study the transition to chaos for increasing subglottal pressure. Fig. 4 shows a spectrogram

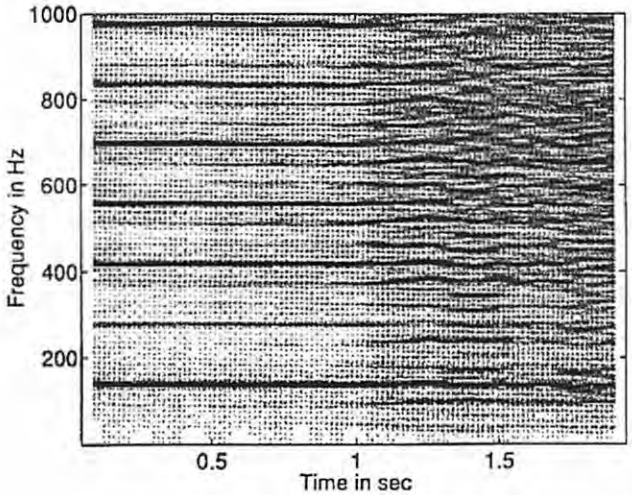


Figure 4. Spectrogram from the output of the continuum model. The driving subglottal pressure was increased every 8,000 data points (400 ms) from 11.6 cm H_2O to 12.6 cm H_2O. In between the regular oscillations and the highly irregular dynamics some additional spectral peaks indicate subharmonic or quasiperiodic components.

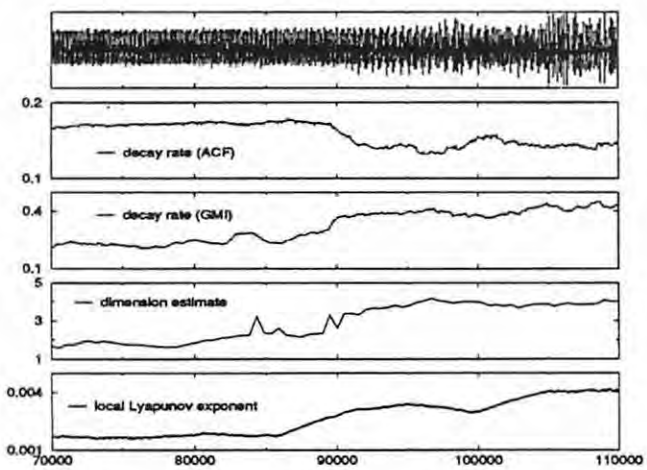


Figure 5. 1st graph: Mouth sound pressure of the continuum model for increasing subglottal pressure (compare Fig. 4). 2nd graph: Initial decay of the autocorrelation function. 3rd graph: Decay rate of the mutual information function. 4th graph: Correlation dimension estimated for segments of 4096 data points (embedding dimension: 10; delay-time: 20 points; shift: 512 points). 5th graph: Estimations of the maximum Lyapunov exponent for segments of 1024 data points. The resulting curve has been smoothed by a running average over 512 subsequent estimates.

of these simulations displaying a transition from regular behavior to irregular oscillations.

In the upper graph of Fig. 5 the corresponding time-series shows the discussed transition to irregular vibrations. The second and third graph allow a comparison of autocorrelations and mutual information. For each window of 1000 points the mean initial slope (1-5 sampling points) was calculated. Both graphs reflect the transition around 90,000 points, but the decay rate of the correlation function decreases whereas the rate for the mutual information increases.

The other graphs show dimensions and Lyapunov exponents of the signal. The estimations of these invariants are consistent with the visual interpretation of the time-series and the spectrogram: There is an overall increase of both quantities from left to right and, moreover, there is a stepwise increase of the Lyapunov exponent which is consistent with the increasing irregularity displayed by the spectrogram. Since the decay of the mutual information is closely related to the Lyapunov exponent [22,24] the rather similar behavior of the decay rate (3rd graph) and the Lyapunov exponent is understandable.

Of particular interest is the saturation of the dimension estimate around four. This may indicate low-dimensional dynamics and provides, therefore, additional information which cannot deduced from the spectrogram. A low dimensionality is also consistent with the dominance of a few empirical orthogonal functions as discussed above (see Refs. [51,55] for details).

Excised Larynx Experiments

Experiments with human or animal larynges serve as a link between the human voice source in vivo and computer models [2,45,56]. They allow controlled and systematic parameter variations and easy observation of vibratory patterns.

We have examined five larynges from large (about 25 kg) mongrel dogs coming from coronary research units at The University of Iowa. The dissected larynges were mounted on an apparatus described in detail elsewhere [56,57]. Heated and humidified air was supplied from below as the driving force of the oscillations. The device was attached to several micrometers to control the adduction and the elongation of the vocal folds. To facilitate observation of vocal fold movement, a strobe light with adjustable frequency was placed above the glottis. The data were recorded on a color video system and afterwards digitized with 16 bit resolution and a sampling rate of 20 kHz.

In our experiments instabilities have been studied for varying subglottal pressure and for asymmetric adduction and elongation of the vocal folds. Two-parameter bifurcation diagrams can be found elsewhere [57]. Here we only summarize briefly the various dynamic regimes which

have been observed for overcritical asymmetry and pressure:

- symmetric periodic phonation in head-like and chest-like registers
- whistle-like sound
- periodic vibrations of the lax fold only
- subharmonics, modulations, and irregular vibrations with both folds involved
- aphonia, i.e. vibrations ceased for very strong asymmetric tension

Typically, the parameter ranges of these regimes overlap, i.e. hysteresis is observed. Sometimes spontaneous transitions between different dynamical regimes appeared without external parameter changes. The bifurcations discussed below have been induced by slowly increasing the driving pressure. Fig. 6 shows the resulting transition from nearly periodic vibrations via period-doubling to irregular oscillations with a strong subharmonic component. The smooth period-doubling and the sudden frequency jump can be seen very clearly in the corresponding pitch contour. A spectrogram of these bifurcations can be found elsewhere [57]. Despite the turbulent noise also the estimations of dimensions and Lyapunov exponents exhibit the expected increase in the transition region.

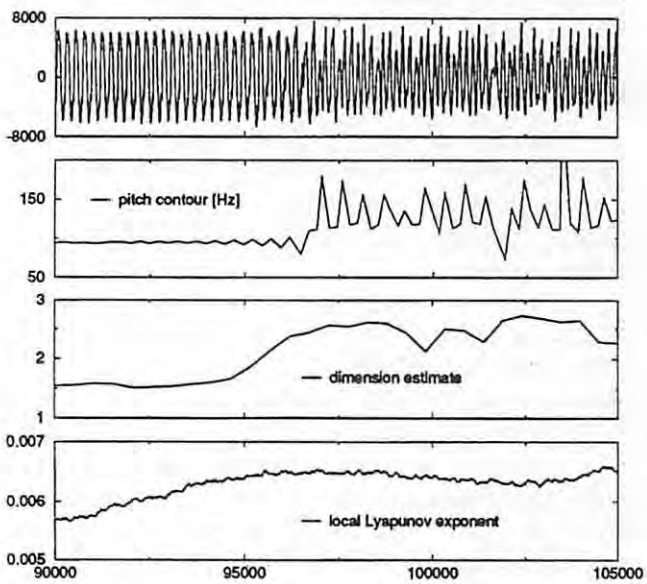


Figure 6. Bifurcations in an excised larynx experiment for apparently symmetric folds at large pressure (about 19 cm H_2O). 1st graph: Time-series smoothed with a running average over 20 points to reduce the turbulent noise. 2nd graph: Pitch contour displaying period-doubling, frequency jump, and subharmonic components (period three). 3rd graph: Estimations of the correlation dimension from segments of 4096 data points (embedding dimension: 10; delay: 10; shift: 512 points). 4th graph: Smoothed estimates of Lyapunov exponents from segments of 1024 points.

Voice Disorders

In the last years bifurcations have been described in vocalizations of patients with nodules, polyps, papilloma, edema, hypo- and hyperfunctional dysphonia [16], chore a Huntington [58], spasmodic dysphonia [15], and unilateral paralysis [20]. To understand the underlying mechanisms of vocal instabilities is the main motivation of computer modeling [46,49,51], excised larynx experiments [57], and high-speed glottography [59].

In this section we study two examples of vocalizations with a rich variety of transition phenomena. Fig. 7 shows a miogram of a sustained "i" from a female patient with highly asymmetric papillomas. During most of the time the voice sounds breathy and has a pitch of about 300 Hz. However, intermittently transitions to regimes with another fundamental frequency (about 180 Hz) are found. Around 9000 this frequency dominates whereas around 12,000 and

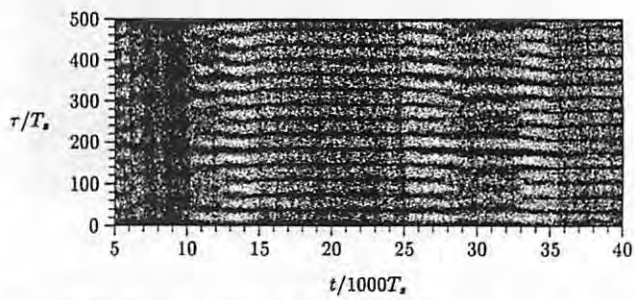


Figure 7. Miogram of a vowel "i" with frequency jumps and toroidal oscillations (see text and Fig. 8).

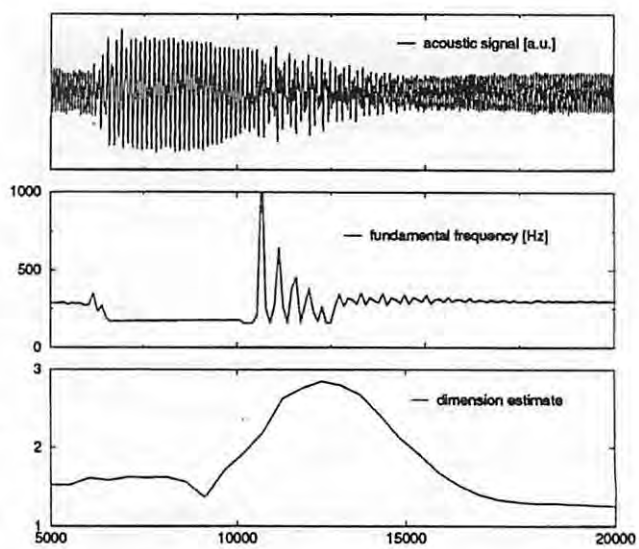


Figure 8. Upper graph: Initial part of the signal analyzed in Fig. 7. Middle graph: Pitch contour displaying frequency jumps and toroidal oscillations. Lower graph: Correlation dimensions (window length: 4096; shift: 512; embedding dimension: 10; delay: 10).

between 25,000 and 35,000 both frequencies are observed. Consequently, the whole signal can be characterized roughly by the following sequence of attractors: limit cycle 1 — limit cycle 2 — decaying toroidal oscillations — limit cycle 1 — torus — limit cycle 1. We note that due to turbulent noise and the obvious non-stationarity only traces of these attractors can be found. However, the analysis of the segment in Fig. 8 is consistent with the above characterization. The time-series and the pitch contour reveal these episodes very clearly. Since 180 Hz are 3/5 of the other frequency around 12,000 a period five is indicated by the pitch contour. As one would expect there is an increase of the dimension during the irregular segment. It is very likely that the instabilities of this patient are due to the desynchronization of the left and right vocal fold. However, a detailed characterization of the eigenfrequencies and the vibratory modes requires high-speed glottography which was not available during examination of this patient.

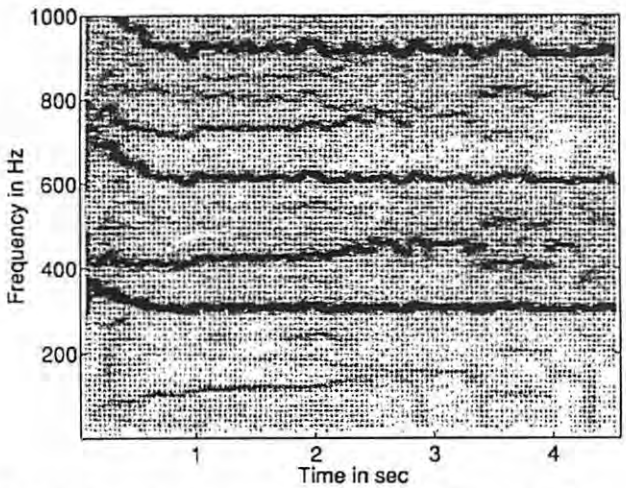


Figure 9. Narrow-band spectrogram (window size: 4096 points; Hamming window; shift: 512) for a rough sounding voice due to acute laryngitis.

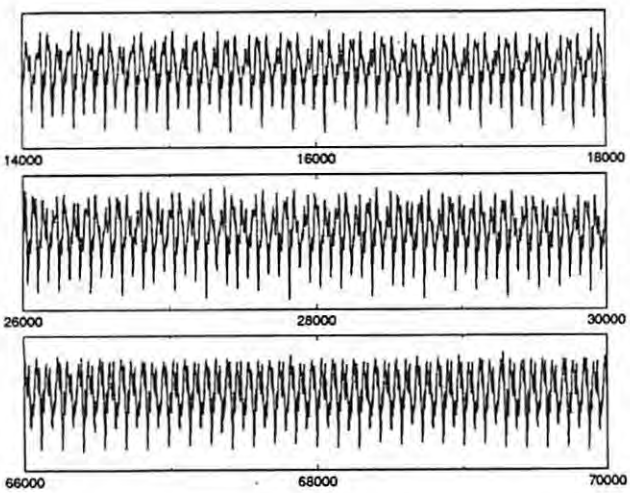


Figure 10. Representative segments of the signal studied in Fig. 9. Upper graph: 1:3 entrainment around 0.7 seconds. Middle graph: Toroidal oscillations around 1.3 seconds. Lower graph: 1:2 entrainment around 3.1 seconds.

The other example is a sustained vowel "a" of a male person with acute laryngitis. Again the spectrogram in Fig. 9 displays a variety of dynamic regimes. First of all, there are two independent frequencies involved. This can be seen clearly between 1 s and 2 s. If we denote the basic frequencies by $f_1 \approx 100$ Hz (increasing in that range) and $f_2 \approx 300$ Hz (almost constant) one can detect the following linear combinations: $f_1 + f_2$, $2f_2$, $2f_2 + f_1$, $3f_2 - f_1$, $2f_2 + 2f_1$, $3f_2$. Moreover, 1:2 and 1:3 frequency-locking can be observed at the beginning and near the end of the signal. In Fig. 10 we present time-series of selected segments displaying toroidal oscillations and frequency-locking.

Summary and Discussion

As demonstrated in earlier papers [5,7,14,16] and by our examples in his paper, bifurcations and low-dimensional attractors are frequently observed in voice signals. There is no doubt that the theory of nonlinear dynamics provides the appropriate framework for the analysis of various voice instabilities. A hierarchy of aerodynamic-biomechanical computer models exist exhibiting bifurcations as found in voice signals.

In this paper we have analyzed data from a rather simple model and a more sophisticated continuum model of the vocal folds. In addition we studied acoustic signals from an excised larynx experiment and vocalizations of voice patients.

The central idea of this paper was the comparative application of five analysis techniques to all these data sets. Since natural voice data are characterized by slow variations o f external parameters we have varied slowly the parameters in our computer models as well. In view of these non-stationarities the analysis techniques have been applied using sliding windows of typically 200 ms. In the following we comment on the insight obtained by the different methods.

Interestingly, conventional methods such as spectrograms and pitch contours proved to be very useful in the detection and characterization of bifurcations. Compared with global linear models (as, e.g., autoregressive processes) spectrograms allow a much more detailed analysis of signals. Local spectra are in a sense state-dependent quantities and, hence, spectrograms can be regarded as a first step towards intrinsically phase-space based techniques. The other conventional tool in voice analysis — the pitch contour — is also closely related to phase space analysis. The detection of events which define the borderlines of pitch periods (maxima, zero-crossing, ...) are just special versions of Poincare sections. Consequently, contours are intimately related to Poincare maps of continuous systems.

The mutual information is a more powerful indicator of statistical dependences than correlation functions and spectra. In our examples, on long time scales spectrograms reveal essentially the same in formation about the signals as miograms. However, as demonstrated in Fig. 5 linear correlations measures such as the autocorrelation decay may behave quite differently. Consequently, in our examples, on short time scales miograms contain more information comparable to Lyapunov exponents or metric entropies (sum of all positive Lyapunov exponents) [22,23,24].

It is now widely realized that the estimation of dimensions and Lyapunov exponents from physiological data is quite complicated. However, these quantities provide information about t he underlying attractors which cannot be obtained by spectral analysis. One should not expect precise estimations from relatively short segments of the data but a sliding analysis allows the comparison of the estimates for different dynamic regimes. For the simulations of the continuum model and the excised larynx data there was indeed the expected increase of dimensions and Lyapunov exponents due to bifurcations. Although the absolute values of dimension estimates from only 4096 data points should be interpreted with caution, the calculated dimensions are consistent with expectations: For noisy limit cycles the estimates are between 1 and 2 and more irregular segments lead to values between 2 and 4.

In summary, the analysis of voice signals has to be adapted to the specific characteristics of the voice source. For example, pitch contours remove effects of the vocal tract filter and constitute, therefore, a valuable variant of a Poincare section. Known parameter variations on time-scales of a few hundred milliseconds require a sliding application of any measure based on stationarity assumptions. However, slowly varying parameters have the advantage that bifurcations of the underlying dynamical system can be detected in single voice signals.

Appendix: Parameters of the Two-Mass Model

Following mostly Ishizaka and Flanagan we choose the following parameters to model a normal voice.

m_1=0.125	c_1=3k_1	r_1=0.02	d_1=0.25	a_{01}=0.05
m_2=0.025	c_2=3k_2	r_2=0.02	d_2=0.05	a_{02}=0.05

k_1=0.08	P_s=0.008	($\approx$ 8cm H_2O)
k_2=0.008	ρ=0.00113	
k_c=0.025	l=1.4	

All units are given in cm, g, ms and their corresponding combinations.

Acknowledgments

Many of the analyzed data are the output of a fruitful cooperation with colleagues from Iowa-City (Ingo Titze, David Berry, and Brad Story) and Berlin (Jurgen Wendler and Ina Steinecke). Support was provided by the Deutsche Forschungsgemeinschaft.

References

[1] G. Fant. Acoustic theory of speech production. Mouton, The Hague, 1960.

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

[3] I. R.Titze and F. Alipour—Haghighi. Myoelastic Aerodynamic Theory of Phonation. (forthcoming book).

[4] J. Lind (Ed.). Newborn infant cry. Almquist and Wiksells Boktrycken, Uppsala, 1965.

[5] Mende, H. Herzel, and K. Wermke. Bifurcations and chaos in newborn cries. Phys. Lett. A, 145, 418—424 (1990).

[6] L. Dolansky and P. Tjernlund. On certain irregularities of voiced-speech waveforms. IEEE Trans., AU—16, 51—56 (1968).

[7] H. Herzel. Bifurcations and chaos in voice signals. Appl. Mech. Rev. 46, 399—413 (1993).

[8] R. Scherer. Physiology of creaky voice and vocal fry. J. Acoust. Soc. Am., 86 (S1), S25(A) (1989).

[9] M. Mazo. Lament made visible: a study of paramusical elements in russian lament. In Themes and Variations edited by B. Yung and J. S. C. Lam, Harvard University, Cambridge, 1994, pp. 164—211.

[10] W. Kelman. Vibratory pattern of the vocal folds. Folia phoniat., 33, 73—99 (1981).

[11] M. Hirano. Objective evaluation of the human voice: clinical aspects. Folia phoniat., 41, 89—144 (1989).

[12] B. Hammarberg, B. Fritzell, J. Gauffin, and J. Sundberg. Acoustic and perceptual analysis of vocal dysfunction. J. of Phonetics, 14, 533—547 (1986).

[13] M. E. Smith, G. S. Berke, B. R. Gerrat, and J. Kreimann. Laryngeal paralyses: theoretical considerations and effects on laryngeal vibration. J. Speech Hear. Res., 35, 545—554 (1992).

[14] H. Herzel and J. Wendler. Evidence of chaos in phonatory samples. In EUROSPEECH, ESCA, Genova, 1991, pp. 263—266.

[15] I. R. Titze, R. Baken, and H. Herzel. Evidence of Chaos in Vocal Fold Vibration. In Vocal Fold Physiology: Frontiers in Basic Science edited by I. R. Titze, Singular Publishing Group, San Diego 1993, pp.143—188.

[16] H. Herzel, D. A. Berry, I. R. Titze and M. Saleh. Analysis of Vocal Disorders with Methods from Nonlinear Dynamics. J. Speech Hearing Res., 37, 1008—1019 (1994).

[17] W. Lauterborn and E. Cramer. Subharmonic routes to chaos observed in acoustics. Phys. Rev. Lett., 47, 1445—1448 (1981).

[18] M. A. Liauw, K. Koblitz, N. I. Jaeger, and P. Plath. Periodic perturbation of a drifting heterogeneous catalytic system. J. Phys. Chem., 97, 11724—11730 (1993).

[19] Merbach, O. Hess, H. Herzel, and E. Scholl. Injection-induced bifurcations of transverse spatio-temporal patterns in semiconductor laser arrays. Phys. Rev. E, 52, 1571—1578 (1995).

[20] H. Herzel, D. Berry, I. Titze, and I. Steinecke. Nonlinear Dynamics of the Voice: Signal Analysis and Biomechanical Modeling. CHAOS, 5, 30—34 (1995).

[21] H. Herzel and R. Reuter. Biphonation in Voice Signals. In Nonlinear, Chaotic, and Advanced Signal Processing Methods For Engineers and Scientists. edited by R. A. Katz, T. W. Frison, J. B. Kadtke, and A. R. Bulsara, American Institute of Physics, Woodbury, 1996.

[22] H. Herzel and W. Ebeling. The decay of correlations in chaotic maps. Phys. Lett. A, 111, 1—4 (1985).

[23] A. M. Fraser and H. L. Swinney. Using Mutual Information to Find Independent Coordinates for Strange Attractors. Phys. Rev. A, 33, 1134—1140 (1986).

[24] B. Pompe and R. W. Leven. Transinformation of Chaotic Systems. Physica Scripta, 34, 8—13 (1986).

[25] A. Renyi. Probability Theory. North-Holland, Amsterdam, 1970.

[26] H. Herzel and I. Grosse. Measuring correlations in symbolic sequences. Physica A, 216, 518—542 (1995).

[27] H. Herzel, A. Schmitt, and W. Ebeling. Finite sample effects in sequence analysis. Chaos, Solitons, and Fractals, 4, 97—113 (1994).

[28] B. Pompe. Measuring Statistical Dependencies in a Time Series. J. Stat. Phys., 73, 587—610 (1993).

[29] B. Pompe and M. Heilfort. Miogram for Short-Time Signal Analysis with an Application to Speech. In Proc. 6. Joint EPS-APS Int. Conf. on Physics Computing edited by R. Gruber and M. Tomassini, Lugano, Switzerland, 22—26 August 1994.

[30] B. Pompe. Ranking in Non-linear Time Series Analysis with an Application to Entropy Estimations. (this volume).

[31] H. Herzel. Possible Mechanisms of Vocal Instabilities. Proceed. Vocal Fold Physiology: Controlling Complexity & Chaos, Singular Publ. Group, San Diego, in press.

[32] P. Grassberger and I. Procaccia. Measuring the Strangeness of Strange Attraktors. Physica D, 9, 189—208 (1983).

[33] J. Holzfuss and G. Mayer-Kress, An approach to error-estimation in the application of dimension algorithms. In Dimensions and entropies in chaotic systems edited by G. Mayer-Kress, Springer, Berlin, 1986, pp. 114—122.

[34] W. Lauterborn and J. Holzfuss. Evidence for a low-dimensional strange attractor in acoustic turbulence. Phys. Lett. A, 115, No. 8, 369—372 (1986).

[35] J. Kurths and H. Herzel. An attractor in a solar time series. Physica D, 25, 165—172 (1987).

[36] A. Wolf, J. Swift, H. Swinney and J. Vastano. Determining Lyapunov exponents from a time series. Physica D, 16, 285—317 (1985).

[37] J.-P. Eckmann, S. Kamphorst, D. Ruelle and S. Ciliberto. Liapunov exponents from time series. Phys. Rev. A, 34(6), 4971—4979 (1986).

[38] M. Sano and Y. Sawada. Measurement of the Lyapunov spectrum from a chaotic time series. Phys. Rev. Lett., 55(10), 1082—1085 (1985).

[39] J. Holzfuss and U. Parlitz. Lyapunov exponents from time series. In Lyapunov Exponents edited by L. Arnold, H. Crauel and J.-P. Eckmann, Lecture Notes in Mathematics, Vol.1486. Springer,Berlin, 1991, pp.263—270.

[40] R. C. L. Wolff. Local Lyapunov exponents: Looking closely at chaos. J. R. Stat. Soc. B, 54, 353—371 (1992).

[41] Z. J. Kowalik and T. Elbert. A practical method for the measurements of the chaoticity of electric and magnetic brain activity. Int. J. Bif. Chaos, 5, 475—490 (1995).

[42] H. Herzel, W. Ebeling, and Th. Schulmeister. Nonuniform chaotic dynamics and effects of noise in biochemical systems. Zeitschr. Naturf., 42a, 136—142 (1987).

[43] H. Herzel and B. Pompe. Effects of noise on a nonuniform chaotic map. Phys. Lett. A, 122, 121—125 (1987).

[44] K. Ishizaka and J. L. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell Syst. Techn. J., 51, 1233—1268 (1972).

[45] K. Ishizaka and N. Isshiki. Computer simulation of pathological vocal-cord vibrations. J. Acoust. Soc. Am. 60, 1194—1198 (1976).

[46] H. Herzel, I. Steinecke, W. Mende, and K. Wermke. Chaos and bifurcations during voiced speech. In Complexity, Chaos and Biological Evolution, edited by E. Mosekilde and L. Mosekilde, Plenum Press, New York, 1991, pp. 41—50.

[47] X. Pelorson, A. Hirschberg, R. van Hassel, A. Wijnands and Y. Auregan. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. J. Acoust. Soc. Am., 96, 3416—3431 (1994).

[48] I. R. Titze. The human vocal cords: A mathematical model. Part I. Phonetica, 28, 129—170 (1973).

[49] H. Herzel and C. Knudsen. Bifurcations in a vocal fold model. Nonlinear Dynamics, 7, 53—64 (1995).

[50] I. Steinecke and H. Herzel. Bifurcations in an asymmetric vocal fold model. J. Acoust. Soc. Am., 97, 1874—1884 (1995).

[51] I. R. Titze and D. T. Talkin. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. J. Acoust. Soc. Am., 66, 60—74 (1979).

[52] F. Alipour-Haghighi and I. R. Titze. Elastic models of vocal fold tissues. J.Acoust. Soc. Am., 90, 1326—1331 (1991).

[53] D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer, Interpretation of Biomechanical Simulations of Normal and Chaotic Vocal Fold Oscillations with Empirical Eigenfunctions. J. Acoust. Soc. Am., 95, 3595—3604 (1994).

[54] F. Alipour and I. Titze. Combined simulation of two-dimensional airflow and vocal fold vibration. In Status and Progress Report National Center for Voice and Speech, 8, 9—14 (1995).

[55] H. Herzel, K. Krischer, D. Berry, and I. Titze. Analysis of spatio-temporal pattern by means of empirical orthogonal functions. In Spatio-Temporal Patterns in Nonequilibrium Complex Systems edited by P. E. Cladis and P. Palffy-Muhoray, Addison-Wesley, Santa Fe, 1995, pp. 505—518.

[56] T. Baer. Investigation of the phonatory mechanism. ASHA Reports, 11, 38—47 (1981).

[57] D. A. Berry, H. Herzel, and I. R. Titze. Bifurcations in excised larynx experiments. J. Voice, in press.

[58] R. Kirsch. Statistische Analyse akustischer Signale. Diplomarbeit, Humbolt Universitat zu Berlin, Berlin 1995.

[59] M. M. Hess, M. Gross, and H. Herzel. Hochgeschwindigkeitsaufnahmen von Schwingungsmoden der Stimmlippen. Otorhinolaryngologia NOVA, 4, 307—312 (1994).

NCVS Status and Progress Report - 9
April 1996, 109-115

Biphonation in Voice Signals

Hanspeter Herzel, Ph.D.
Institute of Theoretical Physics, Technical University, Berlin, Germany
Robert Reuter, M.S.
Institute of Electronics, Technical University, Berlin, Germany

Abstract

Irregularities in voiced speech are often observed as a consequence of vocal fold lesions, paralyses, and other pathological conditions. Many of these instabilities are related to the intrinsic nonlinearities in the vibrations of the vocal folds. In this paper, a specific nonlinear phenomenon is discussed: The appearance of two independent fundamental frequencies termed biphonation. Several narrow-band spectrograms are presented showing biphonation in signals from voice patients, a newborn cry, a singer, and excised larynx experiments. Finally, possible physiological mechanisms of instabilities of the voice source are discussed.

Introduction

The vocal folds, together with glottal airflow, constitute a highly nonlinear self-oscillating system. According to the accepted myoelastic theory of voice production, the vocal folds are set into vibration by the combined effect of sub glottal pressure, the visco-elastic properties of the folds, and the Bernoulli effect [1,2]. The effective length, mass and tension of the vocal folds are determined by muscle action, and in this way the fundamental frequency ("pitch") and the waveform of the glottal pulses can be controlled. The vocal tract acts a s a filter transforming the primary signals into meaningful voiced speech. [3]

Normal sustained voiced sound appears to be nearly periodic, although small perturbations (in the order of a percent) are important for the naturalness of speech. Under certain circumstances, however, much larger irregularities are observed in vocalizations which are often associated with the term "roughness". In the phonetic literature phenomena such as "octave jumps", "biphonation"(two independent pitches), and "noise concentrations" have been reported for decades [4,5]. However, only within the past few years it has been pointed out that these observations might be interpreted as period-doubling bifurcations, tori, and chaos, respectively [6-10]. It has been shown using Poincare' sections and estimations of attractor dimensions and Lyapunov exponents that newborn cries [6] and vocal disorders [7-10] are intimately related to bifurcations and low-dimensional attractors.

On the one hand, subharmonics and irregularities appear occasionally in normal vocalizations (newborn cries, [6] "vocal fry" phonation [9], and speech [11]. On the other hand, vocal instabilities due to bifurcations are often symptomatic in voice pathology.[7-10] Laryngeal stroboscopy reveals that various voice disorders lead to irregular vibratory patterns of the vocal folds resulting in a rough voice quality [12]. The corresponding acoustic signals often show sudden jumps to subharmonic regimes (period-doubling or -tripling) and low-frequency modulations (tori) [7,8,10]. It is the aim of current research to understand the underlying physiological mechanisms of these bifurcations with the aid of high-speed films, excised larynx experiments [13,14], and biomechanical modeling [15,16].

Subharmonics and low-frequency modulations have been studied most extensively in previous studies. In this paper we focus on another phenomenon - the appearance of two independent audible fundamental frequencies termed "biphonation". Biphonation corresponds to a torus of the underlying dynamical system. Of course, also low-frequency modulations mentioned above represent a torus. However, modulations below, say, 70 Hz are usually not perceived as independent pitch but as tremor (1-15 Hz) or roughness (20-70 Hz). Contrarily, two independent frequencies in the range of normal pitch are perceived as a highly dissonant sound and reflect, consequently, a serious deviation from normal phonation.

A powerful tool for the analysis of independent pitches are narrow-band spectrograms. Spectrograms allow the detection of bifurcations in systems with slowly varying parameters [4,5,6,17,18]. They are based on many subsequent short-time power spectra of overlapping segments. The spectral amplitude is encoded as a grey scale and, in this way, the (dis)appearance of spectral peaks owing to bifurcations of the underlying dynamical system can be monitored. For example, period-doubling is reflected in additional peaks in the middle of the harmonics of the original pitch. Low-frequency modulations lead to side-bands of the main spectral peaks. Biphonation is characterized by nonparallel dark lines which represent independently varying spectral peaks.

In the following sections it will be shown that biphonation is a normally occurring event in a variety of voice signals including voice disorders, newborn cries, an artistic voice, and excised larynx experiments. We discuss possible physiological mechanisms leading to biphonation and review results from computer simulations.

Biphonation in Voice Disorders

The characteristic periods of vocal fold vibrations are a few milliseconds. Consequently, stationary segments in sustained vowels contain many cycles and the underlying attractor can be characterized by narrow-band spectrograms. A Hamming window of 4096 points (about 200 ms) and a shift of 512 points are used to generate the overlapping spectral segments.

Normal (almost periodic) phonation leads to a series of parallel lines (see, e.g., the segment around 1.5 s in Fig. 1). Owing to slowly changing parameters such as subglottal pressure or muscle tension, transitions to other regimes can be induced. Such sudden changes reflect bifurcations of the underlying dynamical system. Fig. 1 displays, for example, the appearance of a second series of spectral components around 2 s. This biphonation is audible as a rather dissonant sound.

Two other examples of spectrograms with independent pitches are shown in Figs. 2 and 3. Note that the dominant spectral components in these graphs are not harmonically related. A part of the time-series of the vowel displayed in Fig. 3 has been analyzed in Ref. [8]. For a relatively stationary episode around 0.5 s an attractor dimension of 2.6 has been estimated. The examples presented in

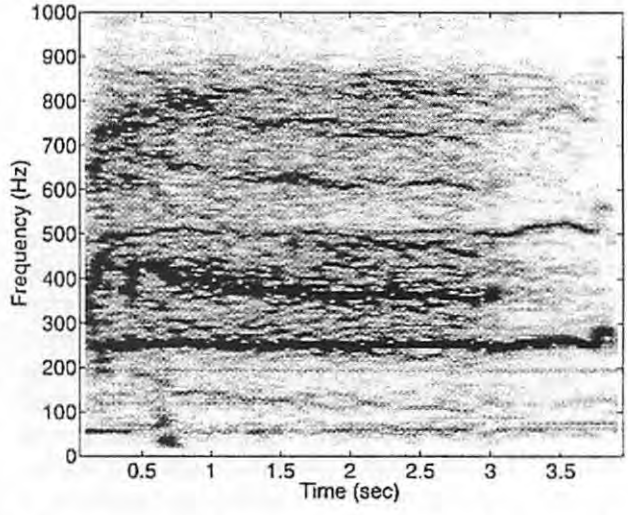


Figure 2. Biphonation in a sustained vowel "u" of a male patient with papilloma.

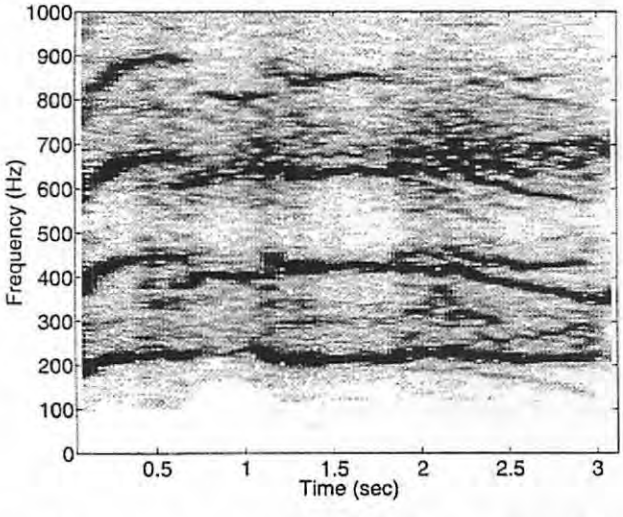


Figure 1. Spectrogram showing biphonation in a sustained vowel "u" of a male patient with edema. Note the nonparallel lines in the final second. The signals in Figures 1-3 have been recorded at the University Hospital Charite' [7,19].

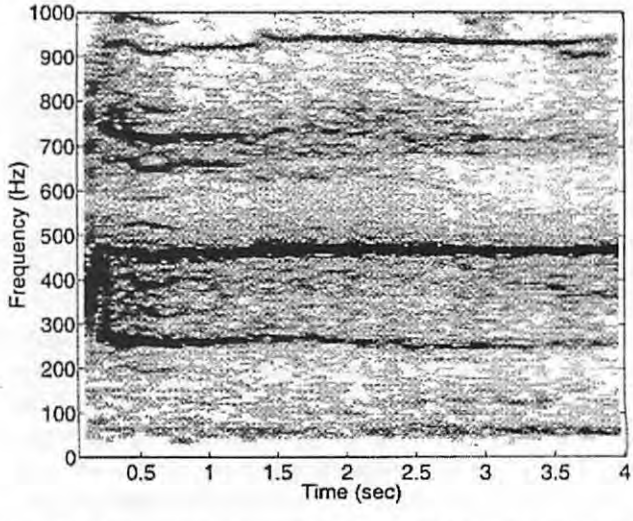


Figure 3. biphonation in a sustained vowel "u" of a female patient with papilloma.

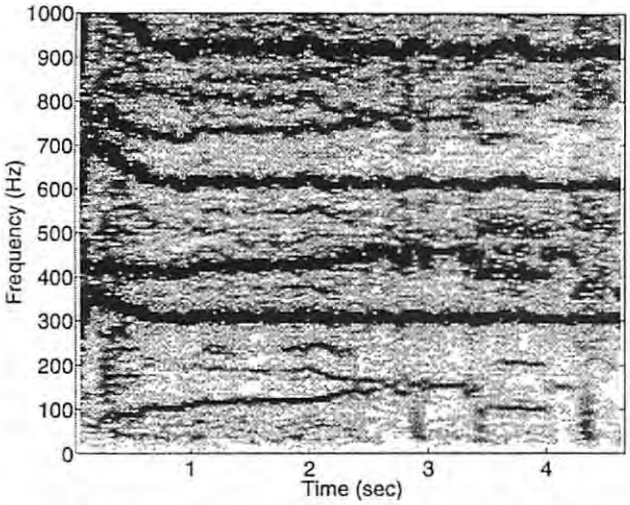


Figure 4. Biphonation and subharmonic regimes in a sustained vowel "a" due to active laryngitis.

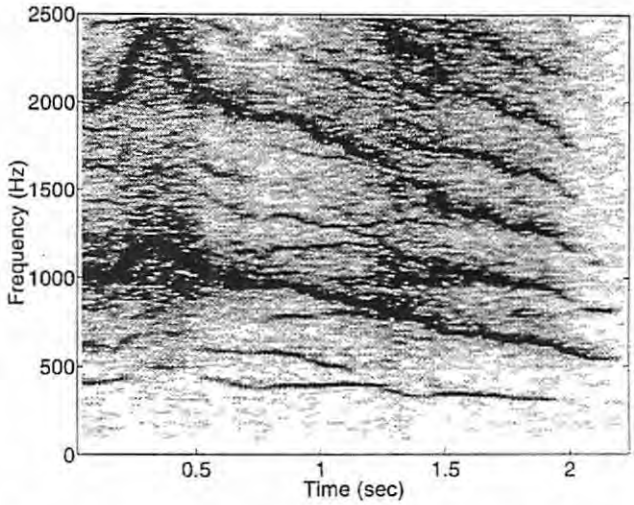


Figure 5. Spectrogram of an infant (4 years old) with papilloma. Biphonation is indicated by a series of lines at multiples of about 400 Hz and another spectral component with descending frequency. In this figure, a window of 2048 points was used.

the Figures 1-3 are chosen from a data base of 21 patients with very hoarse voices. Obviously, biphonation seems to be a rather common phenomenon in vocalizations of patients with strongly perturbed voices.

Fig. 4 shows a spectrogram recorded during acute laryngitis of one of the author s (RR). There is spectral energy at $f_1 \approx 300$ Hz and at its harmonics and another frequency $f_2 \approx 100$ Hz. Up to 2.5 s several linear combinations are discernible ($f_1 + f_2$, $2f_1 + f_2$, $3f_1 - f_2$,...). An episode of 2:1 entrainment can be seen between 3 s and 3.3 s. In addition, close to the end (3.5 s - 4 s) 3:1 entrainment can be detected. This example illustrates that different dynamic regimes can be visualized by spectrograms even in a single sustained vowel.

Biphonation in Newborn Cries

The acoustic properties of newborn cries are in a sense comparable to sustained vowels of adults. Both serve to minimize the variation in vocal fold state and, consequently, one can find relatively stationary segments of a few hundred milliseconds. Due to the smaller size of the vocal apparatus the characteristic frequencies are shifted to higher values. The pitch is usually in the range of 400-600 Hz and the main resonances of the vocal tract are between 2 kHz and 3 kHz.

The sounds of newborn infant cries are used by clinicians for the evaluation of neurological and respiratory function. Already 30 years ago nonlinear phenomena such as subharmonics, two independent pitches, frequency jumps, and "turbulence" have been described in phonations of 30 normal infants [4]. However, only recently these observations have been related to bifurcations and chaos [6].

The authors of Ref. [20] claim that markedly complicated phonation is perhaps as common as the simple basic cry. They also display two spectrograms which clearly indicate two fundamental frequencies (biphonation).

Robb and Saxman [21] analyzed 1200 non-cry vocalizations of 14 healthy infants with an age of 11-25 months. They found subharmonics, frequency shifts, and biphonation to be normally occurring phonatory events. Subharmonics appeared much more often (54 instances in all 14 children) than biphonation (5 instances in 4 children).

It is worth discussing, why these nonlinear phenomena are more frequent in infants vocalizations than in the speech of adults. A possible explanation is that due to the relatively short vocal folds and the strong airflow the role of nonlinearities is more pronounced. Moreover, the instabilities may reflect laryngeal neuromotor immaturity. In other words, older speaker learn to control those phenomena.

There is an intense research to exploit narrow-band cry spectrograms in the neonatal period for assessing the diagnosis and prognosis [22,23,24]. Chromosomal aberrations such as Down's syndrome or the cat-cry syndrome perturb newborn cries very markedly. Diseases affecting the central nervous system change cry characteristics and, therefore, cry features can indicate disturbances of brain function due to oxygen deficiency, meningitis, or hydrocephalus [22]. It was found that especially biphonation is more common in infants with cerebral lesions [23].

Hirschberg and Szende compiled in a monograph [24] more than 100 represent active sound signals from infants with various diseases. Biphonation was detected in cries of children with Soor laryngis and with Papilloma laryngis. Fig. 5 is obtained from the record which accompanies the monograph [24].

In summary, biphonation is found occasionally in vocalizations of normal healthy infants. However, it is also symptomatic for brain damages and can be exploited, therefore, as an early diagnostic tool.

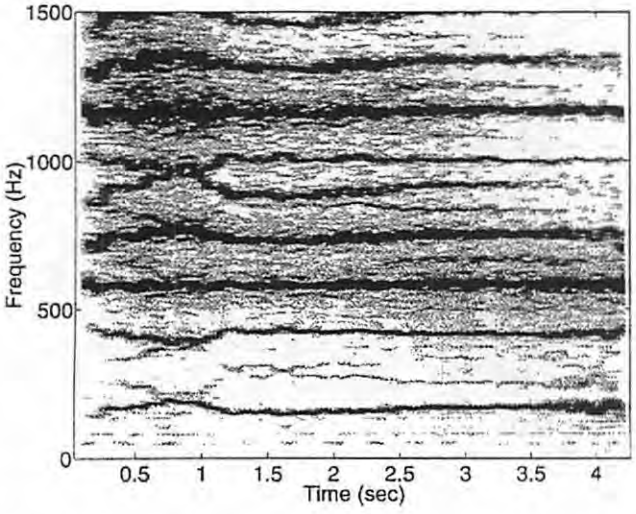


Figure 6. Voluntarily produced biphonation recorded by Pam Davis and Richard Troughear at St. Vincent's Hospital, Sydney.

Biphonation in Normal Voices

Sustained vowels of normal healthy adults are predominantly periodic. Only occasionally [11] or under extreme conditions (pressed voice, vocal fry, simulated creaky voice, inspiratory phonation) subharmonics and irregularities occur. Biphonation seems to be very rare in normal subjects. The impressive Mongolian overtone singing cannot be regarded as biphonation, since these singers use only cultivated boosting of higher harmonics in order to mimic two independent frequencies.

An example of reproducible subharmonics, amplitude modulations, and inspiratory biphonation was reported by Mazo et al. [25] for Russian lament. In these emotionally charged expressions extreme parameter regions of the human voice source are explored and, therefore, these phenomena are likely to occur.

Now we present another interesting example of biphonation in a voice without known vocal pathology. A young female Australian singer was able to produce biphonation at high pitches and intensities. Fig. 6 shows a convincing example. In addition to the intended constant pitch $f_1 \approx 600$ Hz and its harmonics another series of spectral lines is clearly visible. If we denote the frequency slightly below 200 Hz as f_2, one can easily identify various linear combinations of the two frequencies: $f_1 - f_2$, $f_1 + f_2$, $f_1 + 2f_2$, $2f_1 - f_2$, $2f_1 + f_2$,.... Unfortunately, no stroboscopic or high-speed observations were possible to analyze the generation of this remarkable sound in detail.

Excised Larynx Experiments

Experiments with human or animal larynges serve as a link between the human voice source in vivo and computer models [1,14,26]. They allow cont rolled and systematic parameter variations and easy observation of vibratory patterns.

We have examined five larynges from large (about 25 kg) mongrel dogs coming from coronary research units at The University of Iowa. The dissected larynges were mounted on an apparatus described in detail elsewhere [14]. Heated and humidified air was supplied from below as the driving force of the oscillations. The device was attached to several micrometers to control the adduction and the elongation of the vocal folds. To facilitate observation of vocal fold movement, a strobe light with adjustable frequency was placed above the glottis. The data were recorded on a color video system and afterwards digitized with 16 bit resolution and a sampling rate of 20 kHz.

In our experiments instabilities have been studied for varying subglottal pressure and for asymmetric adduction and elongation of the vocal folds. 2-parameter bifurcation diagrams can be found in Ref. [14]. Here we only summarize briefly the various dynamic regimes which have been observed for overcritical asymmetry and pressure:

- symmetric periodic phonation in head-like and chest-like registers
- periodic vibrations of the lax fold only
- subharmonics, modulations, and irregular vibrations with both folds involved
- whistle-like sound
- aphonia, i.e. vibrations ceased for very strong asymmetric tension

Typically, the parameter ranges of these regimes overlap, i.e. hysteresis is observed.

In a parameter region of medium asymmetry and pressure we have found three different phonatory modes ("registers"): A chest-like phonation revealed large amplitudes of vibration, complete closure, a pronounced phase lag between the lower and the upper edge ("mucosal wave"), and very intense harmonics. A falsetto-like phonation was characterized by small amplitude vibrations, incomplete closure, a weak mucosal wave, and less intense harmonics. A whistle-like phonation had a higher pitch, small vibration without any visible mucosal wave, and very weak harmonics. A preliminary explanation of the latter regime would be an interpretation as "vortex-induced vibrations" [14]. Sometimes spontaneous transitions between different registers appeared without external parameter changes. In particular, Fig. 7 shows the coexistence of falsetto register (note the lines at multiples of about 115 Hz) and whistle register (series of nonparallel lines).

In this case the origin of biphonation can be traced back to the coexisting registers. There are two different mechanisms at work to transfer energy from the vertical airflow to vocal fold vibrations: The phase lag between the lower and the upper edge (see [16,26,27] for details) and vortex shedding.

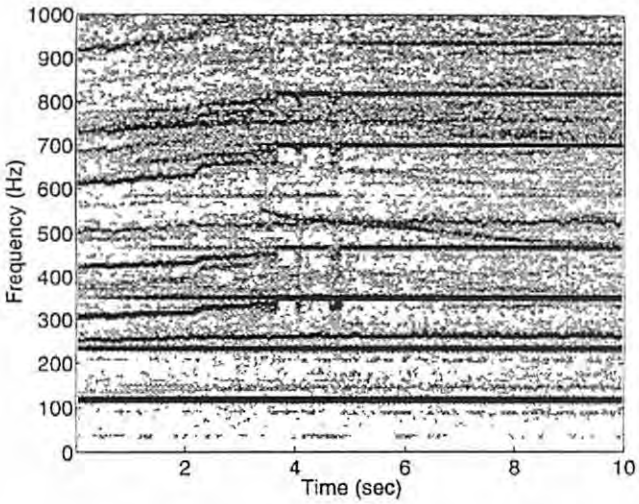


Figure 7. Biphonation in an excised larynx experiment. Details about the experimental setup and the parameter values are given in [14].

Mechanisms of Instabilities

Biphonation is a specific kind of vocal instability. Other phenomena such as subharmonics and chaos are often found in nearby parameter regions. In the following we compile possible mechanisms leading to aperiodicities.

Without making claims about their relative importance, possible candidates are listed.

Neurological Sources

Unsteadiness of muscle contraction due to variations in motor unit firing is a possible source of physiological jitter and shimmer [28]. Moreover, vibrato and vocal tremor in the frequency range of 3-12 Hz seem to be related to a neural feedback loop [29]

Modulations of higher frequencies (say, e.g., 40 Hz) and subharmonic vocalization have predominantly biomechanical origin, i.e., the muscle tension can be regarded as approximately constant. This point of view is supported by computer simulations and excised larynx experiments which exhibit these phenomena without any variations in neuromuscular control.

Aerodynamic Instabilities

Even if the flow is not predominantly turbulent (as for a breathy voice) vortex shedding and instabilities of the jet emerging from the glottis may cause perturbations [30]. The interaction of vortex shedding with acoustic resonances may generate whistle-like sound. Moreover, by means of transverse "lift forces" periodically generated vortices can induce self-sustained oscillations of vibrating structures ("vortex-induced vibrations"). Such a vibratory regime was described above for the excised larynx experiments and might be associated with the so-called whistle-register of the soprano voice [14]. Moreover, the interaction of vortices with vocal resonances might induce secondary frequencies. Biphonation as discussed above is typically associated with a large airflow generating vorticity and, therefore, it seems likely that vortex shedding is involved.

Coupling Between Vocal Folds and Sub- and Supraglottal Resonances

This coupling might play a role in extreme situations such as pressed voice where supraglottal pressure changes play an essential part in driving the vocal folds.

Mucus

It is a frequent observation in clinical practice that inappropriate management of mucus riding on vocal fold tissue leads to a perturbed voice.

Asymmetry of Left and Right Fold

For certain disorders such as unilateral vocal fold paralyses or highly asymmetric vocal fold lesions the desynchronization of both fold might be the dominant mechanism. Such asymmetry is explored extensively in computer simulations [16] and excised larynx experiments [14].

Anterior-Posterior Modes

Such horizontal motion can be seen frequently in stroboscopy and high-speed movies [31]. However, in many cases these modes are entrained by the main vibratory modes, and hence, they do not induce much irregularity.

Desynchronization of Horizontal Motion ("10-mode" related to opening and closing) and Vertical Modes ("11-mode" related to the well-known phase shift of upper and lower edge of the folds)

These two vibratory modes are dominating for normal chest voice. Their desynchronization might play a prominent role in voice disorders. For example, vocal fry, which is often symptomatic in voice pathology, seems to be related to this mechanism. Computer simulations have shown [15] that a lax cover which slows down the 11-mode gives rise to vibratory pattern known from high-speed films of vocal fry [32].

At the current state of the art, the presented list of sources of irregularities has a somewhat speculative character. It is the aim of intense research at the National Center for Voice and Speech and in other groups to substantiate the above conjectures.

Discussion

Computer models of speech production are valuable to understand the basic mechanisms of normal and pathological phonation. There are conceptionally simple

two-mass models [9,16,33,34] and sophisticated continuum models simulating the visco-elastic equations [2,8,15,35,36].

Biphonation was found in the intensively studied Ishizaka-Flanagan model [33]. In that model the bifurcations to toroidal oscillations can be induced by detuning of the two normal modes (in-phase and out-of-phase motion of both masses) [37,28]. It is worth mentioning that in a simplified version of the two-mass without vocal tract only very small instability regions were found [9,39]. This result indicates an important role of the acoustic loading of the vocal tract for the generation of biphonation.

In summary, with respect to these biomechanical models and for the excised larynx data we have some ideas about the mechanisms of biphonation. Unfortunately, the physiological mechanisms of biphonation in newborn cries and voice pathologies are still unclear.

The following tools seem to be essential for further progress in understanding the underlying mechanisms of instabilities:

Two-Parameter-Bifurcation Diagrams

This means that the instabilities are located in planes of parameters of physiological interest (subglottal pressure, muscle tension, ...). For example, voice range profiles (phonetograms) can be regarded as such diagrams where regions of sustained oscillations are separated from regions of no phonation. Bifurcation diagrams have been calculated for the asymmetric 2-mass model [16] and measured in excised larynx studies [14].

Mode Concept

The characterization of instabilities as desynchronization of vibratory modes provides a framework for the classification of biomechanically induced irregularities in models and observations.

Excised Larynx Studies

Experiments with human or animal larynges serve as a link between the human voice source in vivo and computer models [26]. They allow controlled and systematic parameter variations and easy observation of vibratory patterns.

High Speed Digital Imaging

A visualization of irregular patterns is of central importance for detecting the physiological mechanisms of instabilities.

These techniques and broad interdisciplinary cooperation will hopefully lead to a better understanding and treatment of voice disorders in the future.

Acknowledgments

Partial funding for this research was provided by the Deutsche Forschungsgemeins chaft and Grant No. P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders. We thank J. Wendler and M. Cebulla for providing data from patients with voice disorders and acknowledge gratefully P. Davis and R. Troughear for providing the voice signals of the singer. One of us (HH) thanks for the warm hospitality at the National Center for Voice and Speech in Iowa-City and the fruitful cooperation with I. R. Titze, D. Berry, and B. Story.

References

[1] van den Berg, J., J. Speech Hearing Res. 1, 227-244 (1958).

[2] Titze, I. R. and Alipour-Haghighi, F., Myoelastic Aerodynamic Theory of Phonation, (forthcoming book).

[3] Fant, G., Acoustic theory of speech production, Mouton: The Hague, 1960.

[4] Lind (Ed.), J., Newborn infant cry, Uppsala: Almquist and Wiksells Boktryck en, 1965.

[5] Kelman, A. W., Folia phoniat. 33, 73-99 (1981).

[6]. Mende, W., Herzel, H., and Wermke, K., Phys. Lett. A 145, 418-424 (1990).

[7] Herzel, H. and Wendler, J., "Evidence of chaos in phonatory samples," in EUROSPEECH, ESCA, Genova, 1991, pp.263-266.

[8] Titze, I. R., Baken, R., and Herzel, H., "Evidence of Chaos in Vocal Fold Vibration," in Vocal Fold Physiology: Frontiers in Basic Science edited by I. R. Titze, Singular Publ. Group, San Diego 1993, pp.143-188.

[9] Herzel, H., Appl. Mech. Rev. 46, 399-413 (1993).

[10] Herzel, H., Berry, D. A., Titze, I. R., and Saleh, M., J. Speech Hearing Res. 37, 1008-1019 (1994).

[11] Dolansky, L. and Tjernlund, P., IEEE Trans. AU-16, 51-56 (1968).

[12] Hirano, M., Folia phoniat. 41, 89-144 (1989).

[13] Herzel, H., Berry, D., Titze, I. R., and Steinecke I., CHAOS 5, 30-34 (1995).

[14] Berry, D. A., Herzel, H., Titze, I. R., and Story, B. H., J. of Voice, in press.

[15] Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K., J. Acoust. Soc. Am. 95, 3595-3604 (1994).

[16] Steinecke, I. and Herzel, H., J. Acoust. Soc. Am. 97, 1875-1884 (1995).

[17] Lauterborn, W. and Cramer, E., Phys. Rev. Lett. 47, 1445-1448 (1981).

[18] Liauw, M. A., Koblitz, K., Jaeger, N. I., and Plath, P., J. Phys. Chem. 97, 11724-11730 (1993).

[19] Wendler, J., Rauhut, A., and Kruger, J. Phonetics 14, 485-488 (1986).

[20] Truby, H. M. and Lind, J. Acta Paediat. Scand. Suppl. 163, 7-92 (1965).

[21] Robb, M. P. and Saxman, J., J. Acoust. Soc. Am. 83, 1876-1882 (1988).

[22] Sirvio, P., and Michelsson, K., Folia phoniat. 28, 161-173 (1976).

[23] Michelsson, K., Raes, J., and Rinne, A., Folia phoniat. 36, 219-224 (1984).

[24] Hirschberg, J., and Szende, T., Pathologische Schreistimme, Stridor und Hus tenton im Sauglingsalter, Stuttgart: Gustav Fischer Verlag, 1985.

[25] Mazo, M., Erickson, D., and Harvey, T., "Emotion and Expression: Temporal Data on Voice Quality in Russian Lament" in Vocal Fold Physiology edited by O. Fujimura and M. Hirano, Singular Publ. Group, San Diego 1995, pp.173-178.

[26] Baer, T., ASHA Reports 11, 38-47 (1981).

[27] Titze, I. R. and Strong, W., J. Acoust. Soc. Am. 57, 736-744 (1975).

[28] Titze, I. R., J. Speech Hearing Res. 34, 460-472 (1991).

[29] Smith, M. and Ramig, M.O., "Neurological disorders and the voice," National Center for Voice and Speech Progress and Status Report 7, 1995, pp. 207-227.

[30] Pelorson, X., Hirschberg, A., van Hassel, R. R., and Wijnands, A. P. J., J. Acoust. Soc. Am. 96, 3416-3431 (1994).

[31] Hess, M. M., Gross, M., and Herzel, H., Otorhinolaryngologia NOVA 4, 307-312 (1994).

[32] Moore, G. and von Leden, H., Folia phoniat. 10, 205-238 (1958).

[33] Ishizaka, K. and Flanagan, J. L., Bell Syst. Techn. J. 51, 1233-1268 (1972).

[34] Smith, M. E, Berke, G. S., Gerrat, B. R., and Kreimann, J., J. Speech Hearing Res. 35, 545-554 (1992).

[35] Titze, I. R. and Talkin, D. T., J. Acoust. Soc. Am. 66, 60-74 (1979).

[36] Alipour-Haghighi, F. and Titze, I. R., J. Acoust. Soc. Am. 90, 1326-1331 (1991).

[37] Herzel, H., Steinecke, I., Mende, W., and Wermke, K. "Chaos and bifurcations during voiced speech," in Complexity, Chaos and Biological Evolution edited by E. Mosekilde and L. Mosekilde (Plenum Press, New York, 1991), pp. 41-50.

[38] Herzel, H. "Non-Linear Dynamics of Voiced Speech" in Nonlinear Dynamics: New Theoretical and Applied Results edited by J. Awrejcewicz, Akademie Verlag, Berlin 1995, pp. 256-274.

[39] Herzel, H. and Knudsen, C., Nonlinear Dynamics 7, 53-64 (1995).

NCVS Status and Progress Report - 9
April 1996, 117-124

Regulation of Fundamental Frequency With A Physiologically-Based Model of the Larynx

Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Introduction

It is generally understood that the fundamental frequency of oscillation (F_o) of the vocal folds is controlled primarily by the cricothyroid muscles, secondarily by the thyroarytenoid muscles, and tertiarily by subglottal pressure. Increased activity in the cricothyroid (CT) muscles and increased subglottal pressure (P_s) raise F_o, whereas increased activity in the thyroarytenoid (TA) muscles may either raise or lower F_o (Larson—Kempster 1983; Titze—Luschei—Hirano 1989). The lack of unidirectionality of F_o with TA activity is a result of the complex way in which the effective length, tension, and mass change when the thyroarytenoid muscles contract. Whereas the effective tension always increases when the CT muscles elongate the folds, or when P_s increases the amplitude of vibration, the effective tension may decrease when the TA muscles shorten the folds. This is because the increased slackness of the vocal fold *cover* may dominate over the increased stiffness of the *body*.

This leads us to the *body-cover* hypothesis for F_o control first proposed by Hirano (1976) and Fujimura (1981). Simply stated, it suggests that the morphology of the vocal fold lends itself to a mechanical division between two groups of tissue layers that are set into vibration by the glottal airstream. The tissue groups are coupled in the lateral direction, but have rather different mechanical properties. The outer group, or cover, is heavily irrigated with liquid (proteoglycans). The cover can therefore propagate a surface wave that facilitates energy transfer from the glottal airstream to the vocal fold tissues (Baer 1975; Stevens 1977; Titze 1988b). Its tension is controlled by vocal fold length. The inner group, or body, is composed of the deep layer of the lamina propria (strands of ligament) and the TA muscle. It is less deformable and has active contractile properties. Its tension is therefore not only determined by length, but also by active stiffening of the muscle internally. The combined tension of the portions of the body and cover in vibration regulate the fundamental frequency F_o.

A first attempt to quantify the body-cover hypothesis was offered by Fujimura (1981), who related the elongation of the vocal folds to CT and TA contractile forces. No data on effective stiffnesses were available at the time, however, and the model was in fact more conceptual than quantitative. Since then, a number of studies have emerged on active length-tension characteristics of isolated laryngeal muscles (Alipour-Haghighi—Titze—Durham 1987; Perlman—Alipour-Haghighi 1988; Alipour-Haghighi—Titze—Perlman 1989) to justify a more quantitative look at the mechanics of F_o control. Major questions of research interest are: 1) Can a simple body-cover model predict both F_o rises and falls when the TA is activated? 2) Under what conditions does F_o rise, fall, or stay constant with TA activity? 3) What is the mechanism by which F_o changes with subglottal pressure?

This paper summarizes some recent findings by the author and his colleagues in the area of F_o control (Titze—Jiang—Druker 1988; Titze—Luschei—Hirano 1989; Titze 1989a; Titze 1991). The mechanics of cricothyroid joint rotation will first be summarized. This is followed by some data on TA stimulation and EMG recordings on several human subjects. Finally, the F_o-subglottal pressure relationship will be clarified on the basis of the ratio of vibrational amplitude to length of the vocal folds.

Mechanics of Thyroid Cartilage Rotation

Vocal fold length is regulated by the combined actions of the cricothyroid and thyroarytenoid muscles. A simplified view of this mechanism is shown in Figure 1. In parts (a) and (c), sketches of the lateral view of the larynx show how the thyroid cartilage can rotate with respect to the cricoid cartilage.

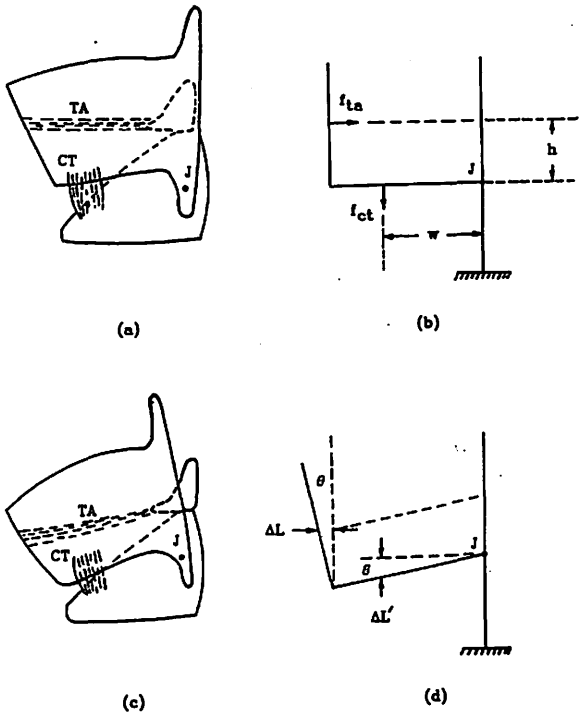


Figure 1. Lateral view of the framework of the larynx. a) and c), thyroid cartilage rotation due to cricothyroid muscle contraction. b) and d), simplified sketches for calculation of torque equilibrium.

The point of rotation is the cricothyroid joint *J*. A simplified conceptualization of this rotation is shown in parts (b) and (d) with the aid of stick figures. The cricoid cartilage and the arytenoid cartilage is represented by the anchored, vertical stick. The two cartilages are considered fused together in this simplified analysis, although slippage in the joint is possible and has been incorporated in more recent analyses. The thyroid cartilage, represented by the 90° elbow, assumes all the motion in this model. (The motion is relative and could equivalently have been assigned to the cricoid cartilage, or some combination of both, as may be the case in the human larynx.)

We see that contraction of the CT muscle rotates the elbow downward, thereby elongating the vocal fold by an amount ΔL. Contraction of the TA muscle, on the other hand, can shorten the vocal folds by creating the opposite rotation.

Assume the perpendicular distance from the point of rotation *J* to the line of action of the CT force to be *w* (Figure 1b). Similarly, assume the perpendicular distance from the point of rotation to the line of action of the TA force to be *h*. Summation of the torques around the joint yields, for static conditions,

$$wf_{ct} = hf_{ta} + k\theta \qquad (1)$$

where f_{ct} and f_{ta} are the active CT and TA forces, respectively, θ is the angle of rotation, and k is an empirically determined torsional stiffness that accounts for all of the passive restraining and restoring forces in the vocal folds around the joint. In general, none of the quantities in equation (1) is a constant. The forces f_{ct} and f_{ta} vary with muscle activity (number of recruited motor units and firing rate of each unit) and with muscle length. The moment arms w and h can vary slightly with the angle of rotation, but for small angles may be considered approximately constant. The torsional stiffness k is also likely to be a function of θ.

Notwithstanding this complexity, it is useful to begin with the simplest possible model and add higher order corrections only when theory and observation do not agree. Let us assume, therefore, that w, h, and k are constants, at least over some range of θ. Let us also recognize that θ is a small angle because the cricothyroid space is small in comparison with the dimensions w and h. We can then write

$$\Delta L = h\sin\theta \approx h\theta \qquad (2)$$

and equation (1) becomes

$$\Delta L = (h^2/k)\left(\frac{w}{h}f_{ct} - f_{ta}\right) \qquad (3)$$

This provides us with a basic beginning relationship between vocal fold elongation and the two major opposing muscle forces f_{ct} and f_{ta}. Preliminary anatomical measurements of w and h on two canine larynges (Titze—Jiang—Druker 1988) suggest that the w/h ratio is on the order of 1. This ratio also governs the relative elongations of the two muscles. If we designate $\Delta L'$ to be the shortening of the cricothyroid muscle, then by comparing the sine of equal angles θ in Figure 1(d), we see that

$$\Delta L' = (w/h)\Delta L \qquad (4)$$

This relationship will be important for relating future length-tension measurements on isolated cricothyroid muscle tissue to control of vocal fold length.

In a preliminary experiment (Titze—Jiang—Druker 1988), it was possible to measure length changes *in vivo* on animals (canines) as a function of peak cricothyroid and thyroarytenoid muscle contraction. In order to relate the results to the discussion presented here, let a_{ct} and a_{ta} represent the *muscle activity levels* of the CT and TA, respectively, ranging between 0 and 1 (100%) of maximum activity. Furthermore, let the elongation ΔL be divided by an appropriate vocal fold reference length L_o so that size differences between larynges can, to first order, be eliminated by normalization. Equation (3) can then be written in terms of the vocal fold strain ε, such that

$$e = \Delta L/L_o = \frac{h^2 F_{ta}}{kL_o}\left(\frac{wF_{ct}}{hF_{ta}}\,a_{ct} - a_{ta}\right) \quad (5)$$

$$= G(R\,a_{ct} - a_{ta}) \quad (6)$$

where F_{ct} and F_{ta} are the maximum forces (corresponding to maximum muscle activities) developed by the CT and TA. Equation (5) states that the effectiveness of the cricothyroid muscle in elongating the vocal folds is weighted by the ratio of the *maximum torques* (wF_{ct} and hF_{ta}) developed by the two muscles about the cricothyroid joint.

Equation (6) shows that there are two coefficients to be determined experimentally, the gain factor G and the torque ratio R. First estimates are $G = 0.1$ and $R = 3.5$ for canines (Titze—Jiang—Druker 1988), but these may be rather different for humans. In particular, the human cricothyroid muscle is much smaller than the canine CT, suggesting that R may be considerably smaller in humans.

Pitch Regulation With Vocal Fold Length

The most difficult problem that is faced in predicting fundamental frequency from basic biomechanic principles is the lack of knowledge about the effective mass in vibration. It is clear that the amplitude of vibration is maximum at the medial (glottal) surface and decreases laterally into the tissue. Vibration in the *thyrovocalis* portion of the body is less than vibration in the cover, and vibration in the *thyromuscularis* portion appears to be near zero. In the vertical dimension, it is also difficult to predict what portion of tissue is in vibration. Given this difficulty, some intermediate steps are necessary to eliminate the need for direct measurement of effective mass.

Consider the vibrating portion of the vocal fold to be a "fat string" as shown by the dashed outline in Figure 2. Let the cross-section of the fat string be A, and let it include portions of the cover and body, as indicated. Extending the simple string formula for the lowest mode of vibration to the fat string of finite cross-section, we have

$$F_o = \frac{1}{2L}(\sigma/\rho)^{1/2} \quad (7)$$

where σ is the mean longitudinal stress (mean tension per unit cross-sectional area A) and ρ is the mean tissue density, expressed in mass per unit volume.

The volumetric density can be considered constant across the body and cover. Measurements we have made on various layers of tissue (Perlman—Titze 1988) indicate values ranging between 1.02 g/cm^3 and 1.04 g/cm^3, with an average value of 1.03 g/cm^3.

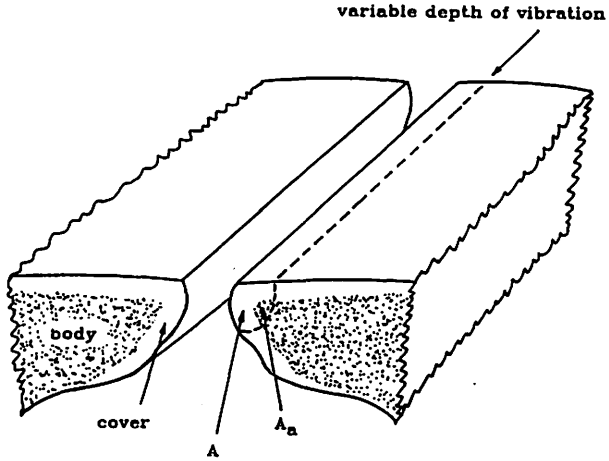


Figure 2. Schematic of vocal fold showing body and cover. The effective depth of vibration within the vocal fold is indicated by the dashed line.

With the density ρ constant in equation (7), only two variables remain that affect F_o in this simple model, L and σ. Let us assume that the mean stress σ is the sum of an active stress and a passive stress, such that

$$\sigma = \sigma_p + (A_a/A)\,\sigma_{am}a_{ta} \quad (8)$$

where σ_p is the mean passive stress over the entire vibrating cross-section A, A_a is the cross-section of the TA muscle in vibration, σ_{am} is the maximum active stress developed in the TA, and a_{ta} is the activity level of the TA, as defined previously. The equation is a direct result of summing longitudinal forces (tensions) along the vocal fold, as can be verified by multiplying both sides of the equation by A.

The passive stress can be determined indirectly from fundamental frequency measurements on human larynges (Titze 1989b),

$$\sigma_p = e^{9.2\epsilon} \quad kPa \quad , \quad (9)$$

and the maximum active stress in the thyroarytenoid muscle is estimated from data by Alipour-Haghighi—Titze—Perlman (1989)

$$\sigma_{am} = 100(1 + 0.6\epsilon) \quad kPa \quad . \quad (10)$$

Combining equations (7)-(10), we obtain

$$F_o = F_{op}\left(1 + \frac{A_a}{A}\,\frac{\sigma_{am}}{\sigma_p}\,a_{ta}\right)^{1/2} \quad , \quad (11)$$

where $F_{op} = (1/2L)(\sigma_p/\rho)^{1/2}$ is the passive fundamental frequency (when $a_{ta} = 0$).

Several observations are in order here. First, we see that the passive fundamental frequency is augmented by an active component that depends on an area ratio, a stress ratio, and the level of activity in the thyroarytenoid muscle. The area ratio can vary between 0 and 1, indicating that either none or all of the vibrating cross-section is muscle. The stress ratio can vary over larger ranges. Equation 10 gives a maximum active tetanic stress (σ_{am}) of 100 kPa at $\varepsilon = 0$ and 124 kPa at $\varepsilon = 0.4$. The passive stress (σ_p in equation 9) ranges between 1.0 and 40 kPa for the same range of ε. This yields a σ_{am}/σ_p ratio that varies from 100 at $\varepsilon = 0$ to 3.1 at $\varepsilon = 0.4$. Given these stress ranges, it is clear that the active contribution to F_o in equation (11) can range from being completely negligible (when a_{ta} is near zero) to being dominant when a_{ta} is near 1 and A_d/A is near 1. Whether or not F_o will rise or fall with increasing a_{ta} will depend on whether a drop in F_{op} due to a length decrease will be overcome by a rise in the active term in parentheses of equation (11). We will now address this question quantitatively.

Variations in F_o Based on Muscle Activities

The combined set of equations in the previous section give predictions of F_o as a function of CT and TA muscle activity levels. One parameter, A_d/A, is yet undetermined. This area ratio parameter, which quantifies the degree to which the TA muscle is in vibration, can be varied over its range of 0 to 1 to determine the effect on F_o. Results of the model are shown in Figure 3. Here a_{ct} is plotted vertically and a_{ta} horizontally. We have termed such a diagram a *muscle activation plot* (MAP). Thus, Figure 3 is a MAP for the two primary laryngeal tensor muscles CT and TA. F_o varies in 50 Hz increments from a low of 100 Hz (lower left corner) to a high of 450 Hz (upper left corner). Solid lines are for $A_d/A = 0.3$ and dashed lines are for $A_d/A = 0.6$. These ratios are estimates to represent the soft and loud conditions, respectively, in some experimental data (Titze—Luschei—Hirano 1989). In other words, less than one third of the vibrating cross-section is assumed to be TA muscle in soft phonation, whereas nearly two-thirds of the vibrating cross-section is assumed to be muscle in loud phonation.

Several observations are in order with regard to Figure 3. First, the lowest F_o is attainable with low a_{ct} and low a_{ta} (lower-left quadrant of the MAP, where speech is produced), the highest F_o is attainable with high a_{ct} and low a_{ta} (upper-left quadrant, where falsetto is produced), and intermediate values of F_o are found in all quadrants. Since the constant F_o lines are continuous, an infinite number of combinations of CT and TA activity is possible for every F_o. For example, 300 Hz can be obtained in the upper left quadrant with $a_{ta} = 0$ and $a_{ct} = 0.76$. With only a gradual increase in a_{ct} (solid line for $A_d/A = 0.3$), a_{ta} can vary all the way into the upper right quadrant. With a combined reduction in both a_{ct} and a_{ta}, the 300 Hz curve extends all the way into the lower left quadrant. Thus, the motor system is

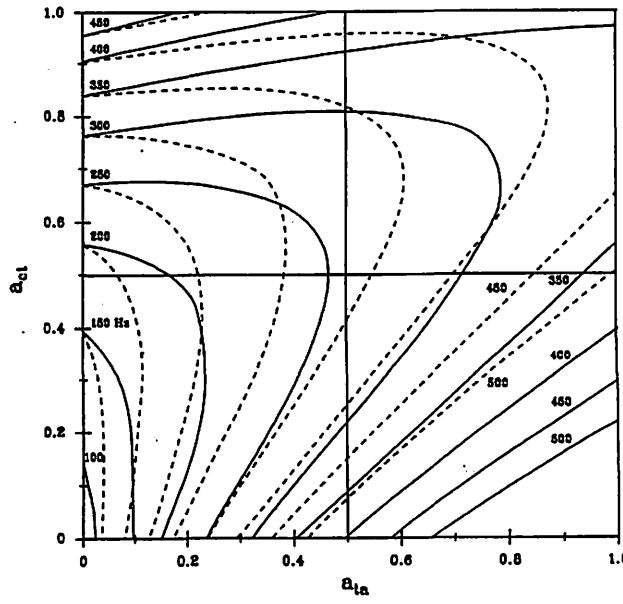


Figure 3. Theoretical muscle activation plot (MAP). Solid lines of constant F_o are for $A_d/A = 0.3$ and dashed lines for $A_d/A = 0.6$.

dealing with a peripheral mechanism that offers multiple solutions to accomplish similar tasks. These redundancies may be removed, however, when more stringent requirements are put on the type of vocalization, such as specific loudnesses, qualities, or dynamic F_o changes. For example, a 300 Hz tone may have distinctly different vocal qualities in the three quadrants. It is likely to be a falsetto sound in the upper left quadrant, a chest sound in the upper right quadrant, and a pressed sound in the lower right quadrant.

The second point to make is that changes in F_o with increased TA activity are all *positive* in the lower left quadrant. A positive increment in a_{ta} (toward the right) will approach a higher F_o curve, both for the solid lines (soft phonation) and the dashed lines (loud phonation). This is a direct result of the downward bending of the curves in this quadrant. Similar statements can be made about the lower right quadrant, but phonation is less likely to occur in this region because the vocal folds are highly adducted. Near the top of the upper quadrants, ΔF_o can be *negative* with increased a_{ta}. Displacement to the right from a given F_o curve will approach a lower F_o if the curves have a positive slope. Thus, F_o and a_{ta} are inversely related at the higher frequencies, especially for $A_d/A = 0.3$ (solid lines corresponding to soft phonation). This theoretical result agrees with experimental findings on human subjects (Titze—Luschei—Hirano 1989; Titze 1991).

The 350 Hz curve is even more interesting. Here both *positive* and *negative* changes in F_o are predicted with increasing a_{ta}. Upward sloping lines in the upper left quadrant give rise to negative ΔF_o with increasing a_{ta} for soft phonation, whereas the downward sloping lines for loud phonation give rise to positive ΔF_o. Experimental results

(Titze—Luschei—Hirano 1989) confirm this sign reversal, but data on stimulated TA muscle activity in humans are still somewhat scarce.

In EMG studies of singers performing a crescendo at constant pitch (Hirano—Vannard—Ohala 1970; Hirano 1988), the CT activity usually decreased while the TA activity increased. This indicated that lung pressure was one of the major mechanisms for increasing loudness. Given that F_0 increases at a rate of 2-6 Hz/cm H_2O with increased subglottal pressure (see below), some F_0 compensation can be achieved by reduced CT activity. Figure 3 predicts that crescendos at constant F_0 should be trajectories between the solid lines and the dashed lines if lung pressure is increased and the amplitude of vibration grows. Should TA activity stay constant in the process, the trajectories would be vertical (downward). This implies an automatic reduction in CT. If TA activity is increased together with lung pressure, the trajectories are sloping or straight lines from left to right. The slopes of the trajectories should be positive at high F_0 and zero or slightly negative at intermediate F_0s. At low fundamental frequencies, the situation is more complicated because of the highly curved nature of the constant F_0 lines in Figure 3. In the lower left quadrant, it may be necessary for TA to decrease in order to follow a trajectory from the solid to the dashed line.

Dependence of F_0 on Subglottal Pressure

It is well known that the fundamental frequency of phonation is dependent on subglottal pressure. Falsetto register included, an overall range of 1-10 Hz/cm H_2O increase in fundamental frequency with subglottic pressure has been reported under various phonatory conditions over three decades of experimentation (van den Berg 1957; Isshiki 1959; Ladefoged 1963; Fromkin—Ohala 1968; Öhman—Lodqvist 1968; Lieberman—Knudson—Mead 1969; Hixon—Klatt—Mead 1971; Baer 1979; Rothenberg—Mahshie 1986). Summarizing the changes in F_0 with P_s seem to be smallest in the high-pitched chest register (1-3 Hz/cm H_2O), greatest in the falsetto register (5-10 Hz/cm H_2O), and intermediate in the low-pitched chest register (2-6 Hz/cm H_2O). Given that subglottal pressure is the primary variable for control of vocal intensity (Fant 1983; Titze 1988a), it is evident that intensity and frequency are not controlled independently in the human vocal system. Speakers tend to raise their voice in pitch when they raise their voice in loudness, and they do it differently in different portions of their vocal range. This interaction was noted many years ago by clinicians searching for an optimal pitch for a given loudness (Fairbanks 1960).

Although this intensity-fundamental frequency dependence is not thought to be an extraordinary discovery in experimental phonetics, it has been difficult to find adequate explanations for the underlying physical nature of the phenomenon. One usually expects the amplitude and frequency of simple oscillators to be uncoupled. Musicians would find it difficult to play an instrument if a crescendo from *pp* to *ff* would raise the fundamental frequency several semitones. On a keyboard instrument, such as a piano, an intensity-frequency dependence in the vibrating string would restrict the instrument to a fixed intensity level if the frequency were to remain nominal. A tuning fork would not have much value as a calibration instrument if different pitches were obtained with different excitations. We have come to expect sound producing instruments to have two separate "knobs," one for loudness and one for pitch.

Let us push the analogy with musical instruments a little further. An inexpensive ukulele with rubber strings *does* have an intensity-frequency interaction. To get a loud sound, the rubber string has to be plucked with a large initial amplitude of vibration. As this amplitude decays gradually, the frequency of the rubber string sags. Close examination of other instruments, even those constructed with better materials, reveals that at extreme ranges of intensity, the frequency may rise or sag, albeit very little in a good instrument.

It appears that the human vocal folds are in the same category as the poorly constructed ukulele in that they are incapable of shaking this frequency-amplitude dependence, even over normal ranges of phonation. The problem is not that the biological material is inappropriate, but that the geometry of the instrument is far from ideal. Builders of string instruments learned early on that long strings under high tension are better than short strings under low tension, even though the pitches may be the same. The bodies, and particularly the necks, of stringed instruments are optimized for length (without, of course, compromising one's ability to play them). Ideally, the Adam's apple should protrude a long distance out of the neck, allowing the vocal folds to be about a meter long.

What is the relevant parameter? The ratio of the vibrational amplitude of the string to the length of the string. For a piano string this ratio is less than 0.001, for the rubber-band ukulele, it is on the order of 0.01, and for the vocal folds, it is on the order of 0.1. This accounts for the corresponding difference in amplitude-frequency dependence between these instruments. The importance of the amplitude to length ratio will now be demonstrated quantitatively.

Amplitude-Dependent Tension

In the development of the acoustic properties of ideal strings, it is assumed that the tension stays constant when the string is deformed in the transverse direction. This is a good approximation when the curved length of the string and the straight (undeformed) length of the string are very nearly the same. The tension, which is length-dependent in elastic materials, is then increased only a negligible amount

when the string assumes its maximally deformed state for a given mode of vibration. The validity of this assumption depends entirely upon the ratio A/L, where A is the transverse mode amplitude and L is the length of the string.

Although the string model is inappropriate in a number of ways for vocal fold vibration, it is surprisingly useful for the study of fundamental frequency control. This is because the major elastic restoring properties of the vocal fold tissues are longitudinal (string-like) fibers. Elastin and collagen fibers are found in large concentration in the vocal fold cover (Hirano—Kurita—Nakashima 1981), and muscle fibers make up the major portion of the body of the folds. The string model fails to capture the vertical degrees of freedom of the vocal folds, however. This can be remedied if one replaces the string model with a *ribbon model.* The finite width of the ribbon does not alter the longitudinal restoring properties much, but allows for vertical modes of vibration. In particular, it allows for a phase delay between the upper and lower portions of the vocal folds, which has been shown to be essential for efficient energy transfer between the air and the tissue (Titze 1988b; Ishizaka—Matsudaira 1972).

Because the self-oscillating mechanism of the vocal folds is not under discussion here, and because the ribbon model and the string model have similar longitudinal restoring properties, continued use will be made of the string model for the present discussion of amplitude-frequency dependence. Let us define the dynamic strain s of the string to be the fractional (percent) difference between the curved string and the straight string. In previous work (Titze 1989b), it was shown that the time-average of the dynamic strain is

$$\bar{s} = (1/8)\,\pi^2(A/L)^2 \qquad (12)$$

where A is the amplitude of vibration and L is the vocal fold length. This time-averaged value of the dynamic strain can be added to the static strain ε in any stress calculation, such as in equation (9).

Realizing that $\bar{s}$ is on the order of 0.01 (a 1 mm amplitude and a 10 mm length are typical), the exponential factor $e^{9.2\bar{s}}$ in equation (9) can be expanded as $(1 + 9.2\bar{s})$, and upon substitution of the passive stress into equation (7), the passive fundamental frequency (F_o for $a_{ta} = 0$) becomes

$$F_{op} = (8.67/L)\,(1 + 5.69\,A^2/L^2)e^{4.61L/L_o} \qquad (13)$$

Note that the A/L ratio is involved in a very explicit way. If this ratio were less than 0.001, as in a piano string, the frequency-amplitude coupling would be entirely negligible. For $A/L \approx 0.1$, however, there is an approximate 6% increase in the fundamental frequency attributed to dynamic strain. The exact value of A/L will depend on the subglottal pressure, of course, as will be shown below.

Before proceeding with the presentation of amplitude and frequency data, it is interesting to comment briefly on the theoretical dependence of F_o on the vocal fold length L. For moderate to large lengths ($.5 < L/L_o < 1.0$), the exponential factor overpowers the $1/L$ factor, causing the frequency to rise sharply with L. For small lengths, however, the exponential factor begins to asymptote toward unity, whereas the $1/L$ factor asymptotes toward infinity. More of a trade-off between these two factors is therefore possible for shorter vocal folds. Our data will show that F_o is less sensitive to L when the folds are shorter than when they are longer. In this region of lesser sensitivity of F_o to L, subglottal pressure seems to play a more dominant role in F_o control.

The relation between amplitude of vibration and subglottal pressure has not been quantified in humans. Informal observation with videostroboscopy suggests that greater pressures result in greater amplitudes, but no mathematical relationship has been proposed. Excised canine larynges have been used, however, to obtain reasonable approximations (Titze 1989b). The amplitude of vibration varies nearly as the square root of the subglottal pressure in canine larynges. An empirical model of the amplitude as a function of length and subglottal pressure is

$$A = (\alpha - \beta L/L_o)P_s^{1/2} \quad , \qquad (14)$$

where P_s is in cm H_2O, α and β are empirically determined constants, L_o is the reference length discussed previously in the context of the stress-strain curve (1.6 cm), and L is the membranous length. For $\alpha = 0.0862$ cm and $\beta = 0.0824$, the

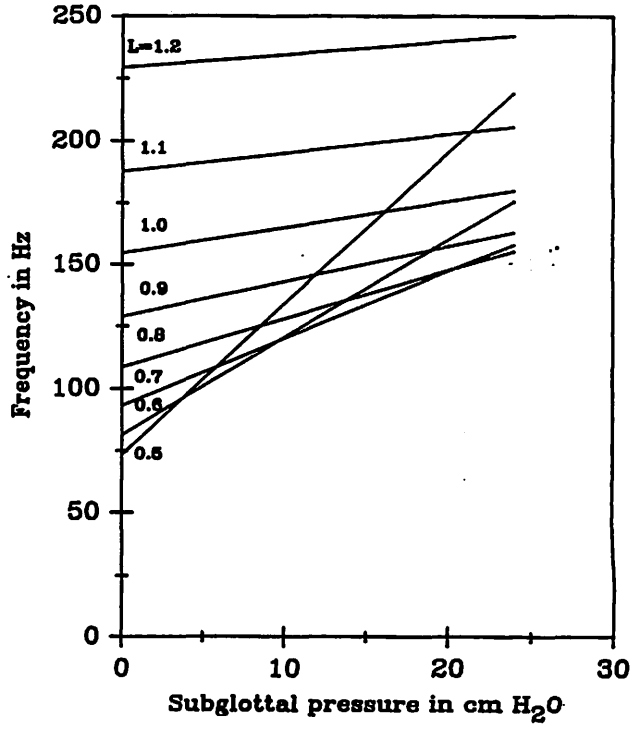


Figure 4. Predicted fundamental frequency versus subglottal pressure from equations (13) and (14). The parameter is membranous vocal fold length in cm.

best match to excised larynx data (Titze 1989b), the variation of F_o with subglottal pressure and length is shown in Figure 4. In this model, $\Delta F_o/\Delta P_s$ varies between 6.1 Hz/cm H_2O and 0.54 Hz/cm H_2O (at L = 0.5 cm and 1.2 cm, respectively).

The fact that the F_o-P_s dependence is linear is a direct consequence of the $P_s^{½}$ variation with vibrational amplitude. If we use the chain rule of differentiation

$$\partial F_o/\partial P_s = (\partial F_o/\partial A)\,(\partial A/\partial P_s) \quad , \qquad (15)$$

we see from equation (13) that $\partial F_{op}/\partial A$ varies directly with A (upon differentiation). This in turn corresponds to a $P_s^{½}$ variation according to equation (14). When $\partial A/\partial P_s$ is then evaluated from equation (14), a $P_s^{-½}$ factor appears that exactly exactly cancels the $P_{s½}$ factor. This makes the slope of the lines constant with P_s, a fortunate results that may not always occur. In explicit form, equation (15) becomes

$$\partial F_o/\partial P_s = (8.67/L^3)\,(\alpha - \beta L/L_o)^2 e^{4.61L/L_o} \quad Hz/cm\ H_2O \qquad (16)$$

which now depends only on the length L. Note the inverse cube relationship with L, which accounts for the sudden rise in the slope for small lengths (Figure 4). For larger lengths, the exponential factor is overridden by the $(\alpha\text{-}\beta L/L_o)_2$ factor, which approaches zero as L approaches L_o. Thus, the slope diminishes for higher frequencies in the chest register, as has been observed by other investigators (e.g., Baer 1975; 1979).

As mentioned earlier, the spacing between the lines for different elongations is less at shorter lengths. This is borne out clearly in Figure 4. Given that the slopes are different at these similar lengths, it would appear that F_o control by subglottal pressure would play an important role at low (speaking) pitches, but would diminish in importance at higher pitches in the chest register. In the falsetto register, on the other hand, a different effective stress-strain curve is likely to be encountered because vibration is restricted to the cover tissue of the vocal folds. The ligament tension may dominate rather than the muscle tension, which could cause a marked increase in the F_o-P_s slope. This effect could be modeled by replacing the coefficient 4.61 in the exponential factor in equation (9) by a larger number, but since insufficient data are available at present, this extension of the model will be left for future investigations.

Conclusion

An explanation has been offered for why increased thyroarytenoid muscle activity can both raise and lower F_o. When the cover is lax and the amplitude of vibration is sufficiently large to include a portion of the muscle in vibration, increased thyroarytenoid activity will raise F_o. This is because the increase in tension in the muscle outweighs the decrease in the tension in the cover that may result from a small decrease in vocal fold length. This effect increases with increasing vibrational amplitude, i.e., with increased loudness, because more of the muscle is involved in vibration.

When the cover is very tense (large cricothyroid activity with elongated vocal folds), the active tension in the TA muscle cannot match the tension in the cover and the vocal ligament. Greater contraction of the muscle will lower F_o because the small gain in muscle tension is outweighed by reduced tension in the cover that results from a small decrease in length. The length-tension curve for the cover is very steep at large elongations, making the loss of tension very dramatic with only minor reductions in length. At intermediate levels of CT and TA contraction, our results show that an increase in TA can both raise and lower F_o. This depends critically on the amount of muscle in vibration. For louder phonations, where the vibrational amplitude is larger, there is more likely to be an F_o rise.

The increase in fundamental frequency with subglottal pressure is explainable on the basis of an amplitude-dependent tension in the vocal folds. The critical parameter is A/L, the ratio of the amplitude of vibration to the length of the vocal folds. This ratio was shown to vary between approximately 0.05 and 0.5 over the range of frequencies examined, causing a variable amount of dynamic stretching of the tissue when vocal fold displacement is at its peak lateral excursion. Using a stress-strain curve derived from fundamental frequency and length measurements on human subjects, a 0.5 to 6 Hz/cm H_20 range of fundamental frequency change with subglottal pressure was predicted. This is similar to the range of values that have typically been measured in humans and excised dog larynges. A further prediction from the present analysis is that the largest changes in F_o with P_s are likely to occur when the vocal folds are very short and lax. A more precise quantification of the stress-strain curve of vocal fold tissue is needed to predict the somewhat higher F_o-P_s slopes observed in the falsetto register, but the mechanism is likely to be similar.

References

Alipour-Haghighi, F.—I.R. Titze—P.L. Durham (1987). "Twitch response in the canine vocalis muscle", *Journal of Speech and Hearing Research* 30: 290-294.

Alipour-Haghighi, F.—I.R. Titze—A. Perlman (1989). "Tetanic contraction in vocal fold muscle", *Journal of Speech and Hearing Research* 32: 226-231.

Baer, T. (1975). *Investigation of Phonation Using Excised Larynxes.* Ph.D. dissertation, Cambridge, MA: MIT.

Baer, T. (1979). "Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes", *Journal of the Acoustical Society of America* 65(5): 1271-1275.

Fairbanks, G. (ed.) (1960). *Voice and Articulation Drillbook.* New York: Harper & Row Publishers.

Fant, G. (1983). "Preliminary to analysis of the human voice source", *Quartly Progress and Status Report STL-QPSR 4.* Stockholm, Sweden: Speech Transmission Laboratory, Royal Institute of Technology, 1-27.

Fromkin, V.—J. Ohala (1968). "Laryngeal control and a model of speech production", *Working Papers in Phonetics* (UCLA) 10, 98-110.

Fujimura, O. (1981). "Body-cover theory of the vocal fold and its phonetic implications", in: K.N. Stevens—M. Hirano (eds.), *Vocal fold physiology.* Tokyo: University of Tokyo Press, 271-288.

Hirano, M. (1976). "Morphological structure of the vocal cord as a vibrator and its variations", *Folia Phoniatrica* 26: 89-94.

Hirano, M. (1988)."Vocal mechanisms in singing: Laryngological and phoniatric aspects", *Journal of Voice* 2(1): 51-69.

Hirano, M.—W. Vannard—J. Ohala (1970). "Regulation of register, pitch, and intensity of voice", *Folia Phoniatrica* 27: 1-20.

Hirano, M.—S. Kurita—T. Nakashima (1981). "The structure of the vocal folds", in: K. Stevens— M. Hirano (eds.), *Vocal fold physiology.* Tokyo: University Tokyo Press, 33-43.

Hixon, T.J.—D.H. Klatt—J. Mead (1971). "Influence of forced transglottal pressure on fundamental frequency", *Journal of the Acoustical Society of America* 49: 105(A).

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

Isshiki, N. (1959)."Regulatory mechanism of the pitch and volume of voice", *Otolaryngology-Rhinology-Laryngology* (Kyoto) 52: 1065-1094.

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

Larson, C.R.—G.B. Kempster (1983). "Voice fundamental frequency changes following discharge of laryngeal motor units", in: I.R. Titze—R.C. Scherer (eds.), *Vocal fold physiology: Biology, acoustics, and phonatory control.* Denver, CO: The Denver Center for the Performing Arts: 91-104.

Lieberman, P.—R. Knudson—J. Mead (1969). "Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation, *Journal of the Acoustical Society of America* 45: 1537-1543.

Öhman, S.—J. Lindqvist (1968). "Analysis and synthesis of prosodic pitch contours", *Zeitschrift für Phonetik* 21: 164-170.

Perlman, A.L.—I.R. Titze (1988). "Development of an *in vivo* technique for measuring elastic properties of vocal fold tissue, *Journal of Speech and Hearing Research* 31: 288-298.

Perlman, A.L.—F. Alipour-Haghighi (1988). "Comparative study of the physiological properties of the vocalis and cricothyroid muscles, *Acta Otolarygology* 105: 372-378.

Rothenberg, M.—J. Mahshie (1986). "Induced transglottal pressure variations during voicing", *Journal of Phonetics* 14: 365-371.

Stevens, K.N. (1977). "Physics of laryngeal behavior and larynx modes", *Phonetica* 34: 264-279.

Titze, I.R. (1988a). "Regulation of vocal power and efficiency by subglottal Pressure and glottal width", in: O. Fujimura (ed.), *Vocal physiology.* New York NY: Raven Press: 227-238.

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

Titze, I.R.—J. Jiang—D. Druker (1988). "Preliminaries to the body-cover theory of pitch control", *Journal of Voice* 1(4): 314-319.

Titze, I.R. (1989a). "Physiologic and acoustic differences between male and female voices", *Journal of the Acoustical Society of America* 85(4): 1699-1707.

Titze, I.R. (1989b). "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.—E.S. Luschei—M. Hirano (1989). "The role of the thyroarytenoid muscle in regulation of fundamental frequency", *Journal of Voice* 3(3): 213-224.

Titze, I.R. (1991). "Mechanisms underlying the control of fundamental frequency", in: J. Gauffin—B. Hammarberg (eds.), *Vocal fold physiology: Acoustic perceptual and physiological aspects of voice mechanisms.* San Diego: Singular Publishing Group, Inc.: 129-138.

van den Berg, J.W. (1957). "Sub-glottal pressure and vibrations of the vocal folds", *Folia Phoniatrica* 9: 65-71.

NCVS Status and Progress Report - 9
April 1996, 125-132

Vocal Tract Imaging: A Comparison of MRI and EBCT

Brad Story, Ph.D.
Department of Speech Pathology and Audiology, University of Iowa
Eric Hoffman, Ph.D.
Department of Radiology, University of Iowa College of Medicine
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, University of Iowa

Abstract

Vocal tract imaging for the vowels /i/ and /a/ using both EBCT and MRI was carried out for one subject (29 yr. old male, native of midwestern United States) using an Imatron C-150 electron beam CT scanner and a GE Signa 1.5 Tesla scanner, respectively. Each image set was analyzed using a general display and quantitation package called VIDATM (Volumetric Image Display and Analysis). The image analysis consisted of segmenting the airspace from the surrounding tissue, obtaining a 3-D vocal tract shape via shape based interpolation, and finally using an iterative bisection algorithm to determine the vocal tract area function. The results show that the 3-D representations of the vocal tract shapes derived from EBCT show subtle deformations of the airway by articulatory structures and teeth that are not observed in the MRI based representations. Shaded surface renderings of each vocal tract shape and for each imaging technique are shown and the apparent trade-offs between the two imaging methods are discussed.

Introduction

In the field of speech research there has long been a desire to obtain a body of morphological information about the shape of the vocal tract as well as other airways such as the nasal tract and trachea to aid in further understanding voice and speech production. Modern imaging techniques are beginning to be used to acquire three-dimensional shape information about the vocal tract and associated airways. Volumetric imaging of the airways relies on the ability of an imaging technique to acquire a series of image slices, in one or more anatomical planes, through a desired volume. This process can be performed with either magnetic resonance imaging (MRI) or electron beam computed tomography (EBCT). Post-processing of the images typically includes segmentation of the air space from the surrounding tissue and a 3-D reconstruction of the airway shape. Once the 3-D structure of the airway has been identified, it can be measured and analyzed in many ways. It should be noted that during speech production a speaker will traverse many different vocal tract shapes in a short period of time, often overlapping one shape into another (i.e. coarticulation). Unfortunately, neither MRI nor EBCT can acquire a volume image set fast enough to capture the dynamically changing vocal tract shape. Thus, the use of present, commercially available, imaging techniques can be used only to study static vocal tract shapes.

MRI is an attractive imaging technique primarily because it poses no known danger to the human subject being imaged. Since human speech comprises a large number of vowel and consonant shapes it is often desirable to acquire a large number of image sets. Thus, the human subject may need to be exposed to hours of imaging, but fortunately, with no apparent risk. However, MRI has a number of distinct disadvantages. The scanning time required for acquisition of a full image set is on the order of several minutes. When this is coupled with the pauses required for subject respiration, an image set may require 10 to 15 minutes to complete. Thus, the image set and subsequent reconstruction of the airway is based on a large number of repetitions of a particular shape. Also, because of various imaging artifacts associated with air/water interfaces, the air-tissue interface can be poorly defined when imaged via MRI techniques, and quite often the faster MRI

techniques yield the worst air/water interface artifacts. Additionally, teeth and bone are poorly imaged by MRI because of the low hydrogen (water) concentration, making these structures appear to be the same gray scale value as air.

Baer et al.[1] demonstrated the use of MRI to directly measure the vocal tract shape for four vowels (2 subjects). This study was the first demonstration of 3-D reconstructions of the vocal tract shape using imaging techniques. Moore[2] also performed an MRI study in which sagittal and coronal image sets of the vocal tract were obtained for three vowels and two consonants. The study investigated correlations between cavity volumes and resonance frequencies of the vocal tract, as well as other variables such as lip length and lip area. However, it did not include three-dimensional reconstructions of the vocal tract shapes. Another MRI study of the vocal tract was performed by Sulter et al.[3]. Their primary interest was in correlating the vocal tract length measured for an /e/ vowel to resonance frequencies (formants) predicted by theory. They also imaged the vowels /i/ and /a/ from which they measured cavity volumes. Dang et al.[4] used MRI to produce 3-D reconstructions of the nasal tract passages and sinus cavities. The nasal tract morphology was subsequently used to model the acoustic characteristics of the nasal system. MR volume imaging of the fricative consonants has been recently reported by Narayanan et al.[5] and Narayanan[6]. These studies provide the most accurate information to date of the constrictions and air channels that produce the turbulence generated sound characteristic of fricative consonants. Beautemps et al.[7], Greenwood et al.[8], and Lakshminarayanan et al.[9] have all concentrated on using MRI to acquire midsagittal sections of the vocal tract from which they measure the tract widths and convert them to cross-sectional areas using an empirically derived formula. Since only a midsagittal section is acquired, a 3-D reconstruction of the vocal tract shape is not possible. However, acquisition of a single slice provides a fast and efficient method of obtaining useful shape information of the vocal tract albeit not as complete as a full volume image set.

If image quality were the only relevant consideration, EBCT would be the preferred method because; 1) it allows for a sharp definition of the air-tissue interface at the walls of the vocal tract; 2) images teeth; and 3) a volume scan of the vocal tract with a much finer pixel resolution than MRI can be completed within tens of seconds. However, since EBCT uses ionizing radiation, only a limited number of volume scans of the vocal tract region are considered to be safe. For this reason EBCT has been used sparingly in speech production studies. However, Kiritani et al.[10] reported a feasibility study using EBCT and Perrier et al.[11] have used EBCT to acquire a sparse (noncontiguous) set of slices through various vocal tract regions. They combined this with a sectioned vocal tract cast of a cadaver in an attempt to formulate an accurate midsagittal width-to-area transformation.

It is apparent that imaging the vocal tract with MRI is a trade-off of image quality and accuracy for safety of the subject. EBCT imaging brings with it a radiation risk while providing potentially much better vocal tract images. It is the intent of this paper to demonstrate the quality of 3-D vocal tract reconstructions obtained with image sets from both MRI and EBCT to evaluate the information loss in the MRI based data sets. The results will be presented primarily from a qualitative standpoint in that the discussion will center around the similarities and differences of 3-D vocal tract reconstructions of the vowels /i/ and /a/ based on the two imaging techniques. Particular attention will be given to the areas of the vocal tract that contain teeth and also the epilaryngeal and lower pharyngeal regions where the airway can be quite complicated but also very small.

Image Acquisition and Analysis

Scanning Parameters and Protocol

Volumetric imaging of the vocal tract using MRI for a single subject has been completed for 22 different phoneme configurations and also for the nasal tract and trachea[12,13]. However, only the vowels, /i/ and /a/, will be considered in this paper. The subject (BS) was a 29 year-old male with no history of speech or voice disorders and is native to the midwestern United States. The subject is also an author of this study. Additionally, the vowels /i/ and /a/ were imaged for subject BS using EBCT.

The MR images were acquired using a General Electric Signa 1.5 Tesla scanner. As discussed above, MRI produces a blurred boundary between tissue and air[1]. However, the acquisition mode can be chosen and the pulse sequence parameters adjusted to provide reasonable air-tissue interfaces. The parameters shown in Table 1 were empirically found to provide acceptable images of the airway. Using these parameters, a 26 slice series of 5 mm thick contiguous, parallel, axial sections was gathered in an interleaved acquisition. This image set, which extended from just cephalad of the hard palate down to the first tracheal ring, could be acquired with 4 minutes and 16 seconds of scan time. However, because the subject was required to phonate during image acquisition, the actual amount of time required to image one shape was approximately 10 minutes, allowing for respiration during pauses in scanning.

Table 1.
MR Parameters for Vocal Tract Image Acquisition

Mode = Fast Spin-Echo, using an anterior neck coil	
TE = 13 msec	(echo delay time)
TR = 4000 msec	(repetition time)
ETL = 16 msec	(echo train length)
FOV = 24 cm	(field of view)
NEX = 2	(number of excitations)
image matrix =	256 x 256 pixels
resolution =	0.938 mm/pixel
slice thickness =	5 mm

All EBCT images were obtained using an Imatron C-15014 scanner (Imatron Corp. South San Francisco). For a single volume scan of the vocal tract, a series of 3 mm thick contiguous, parallel, axial images were acquired which extended from just cephalad of the hard palate down to the first tracheal ring. The axial images captured the full A-P (anterior-posterior) and lateral extent of the airway within each slice. The field of view (FOV) for each slice was 21 cm while the pixel matrix was 512 x 512. This provided a pixel dimension in the axial plane of 0.410 mm, close to the theoretical limit of the scanner resolution. The time required to scan the desired 38 slice volume using the EBCT was approximately 20 seconds, which was too long to comfortably phonate without strain. Thus, the subject phonated for 10 seconds at the end of which scanning was halted to allow for respiration and then restarted.

The protocol for collecting the images with each of the two techniques was similar. In both cases the subject was positioned in a comfortable supine position on the patient table and was required to phonate at a comfortable pitch and loudness. The requirement of phonation is necessary in order for the subject to maintain the desired tract shape. More details concerning the scanning protocols can be found in Story et al.[14] ,Story[12], and Story et al.[13]

Image Analysis

All image analysis operations were performed with a general image display and quantitation package called VIDA (Volumetric Image Display and Analysis) which has been developed (and continues to be enhanced) by researchers in the Division of Physiologic Imaging at the University of Iowa[16].

The image analysis process included three main steps: 1) segmentation of the airway from the surrounding tissue, 2) three-dimensional reconstruction of the airway by shape-based interpolation, and 3) determination of an airway centerline and subsequent extraction of cross- sectional areas assessed from oblique sections calculated to be locally perpendicular to the airway centerline to create an "area function". However, the area function analysis will not be discussed in this paper. More detailed explanations of this analysis process can be found in Story et al.[14], Story[12] and Story et al.[13] and also at the Internet web site - http://everest.radiology.uiowa.edu/tutor/app/vocal/vocal.html. A brief explanation of the image analysis is given here.

Airway Segmentation: The image data sets were transferred from the GE Signa scanner or the EBCT scannner to a Unix- based workstation (via magnetic tape) and translated into a file format recognized by VIDA. When files are in VIDA format they can be read into VIDA s shared memory structure with a conversion from 16 to 8 bit per pixel gray scale resolution (simply to reduce unnecessary data size). The upper gray level cutoff was determined by computing a pixel histogram on several slices sampled throughout the data set and choosing the cutoff point so that all pixel values present in the image set were below the cutoff.

The airway was segmented from the surrounding tissue by setting all voxels considered to be in the airway to a unique gray scale value. The first step in this process was to attenuate all gray scale values in an image slice by 5 percent. This ensured that no voxel in the image volume had a value of 255 (the brightest value for an 8-bit gray-scale range). Next, a gray scale value that represented the threshold dividing air and tissue was determined by choosing the voxel intensity that was halfway between the darkest part of the airway and the brightest part of the tissue surrounding the airway as determined by region of interest analysis. Validation of this approach using the image data sets for a phantom study is discussed in Story[12] and Story et al.

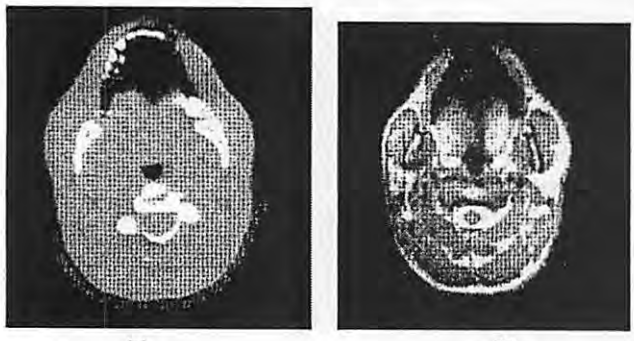
Once the threshold value had been determined the actual segmentation process could be performed. A seeded region growing algorithm[16] was used that changed the gray scale values of all voxels below the threshold value to the brightest possible gray level[17,18]. This effectively set the airway to a single color that, because of the previous attenuation process, was unique to the rest of the image volume. The seeded region growing process required the airway region to be non-continuous with air outside the body. For the axial slices in the pharyngeal section this was not a problem but in the mouth region where some slices have an open mouth condition the algorithm would "leak" the brightest pixel value out of the airway and into the "blackness" (air) outside the image. To correct the problem an artificial boundary that defined the mouth termination of the vocal tract was constructed in the midsagittal plane according to Mermelstein[19].

Volume Reconstruction: To reconstruct the three-dimensional shape of the vocal tract airway, the segmented image set was interpolated by shape based interpolation[20]. Only the gray scale value of the airway was interpolated while all other colors were ignored. This had the effect of "stripping" all tissue away from the airway and creating the same voxel resolution in the axial direction as the other directions. The edges of the interpolated airway image were extracted and used to perform surface rendering of the airway, creating a digital "cast" of the vocal tract. Surface rendering produces a high quality three-dimensional representation of the airway that can be rotated and magnified to show many different perspectives. It is only a qualitative tool, but nevertheless an important step in the image analysis because it shows the quality of the segmentation process and provides three-dimensional views of each vocal tract shape.

Results

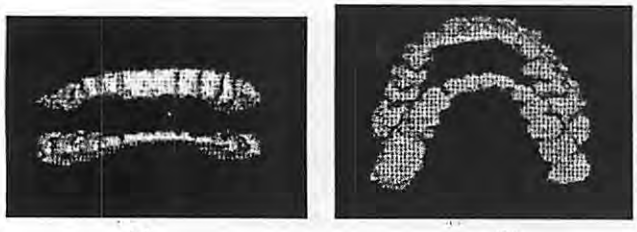
Segmentation and Exclusion of the Teeth

An axial slice from each of the EBCT and MR image sets for the /a/ vowel are shown in Figure 1. The slices are at similar locations within the vocal tract. The EBCT image is clearly of higher resolution and definition than that from MRI. In particular, the boundary between the tissue and air is very sharp and the teeth are easily separated from the air space. An advantage of collecting an EBCT image set is that it provides a convenient means of measuring the dimensions of the teeth, in terms of their extent into the airway, which can then be used as an aid in segmenting the air from tissue and teeth in the MR image sets. Specifically, using the EBCT image set, a "digital cast" of the teeth can be obtained by segmenting them from all other tissue and air and then applying shape-based interpolation to extract only the teeth from the image set. A shaded surface rendering of the teeth from the /a/ vowel is shown in Figure 2. The dimensions of this "cast" of the teeth can now be measured with a general region of interest analysis program[16] and then be used to estimate the exclusion of the teeth in the MR image sets. Figures 3 and 4 show two perspective views of the 3-D reconstruction of the /a/ vowel using the MRI data. The most inferior portion of each figure is the superior part of the trachea. The airway then traverses upward through the glottis (the narrowest portion of the airway), into the pharynx, around the 90 degree bend into the oral cavity, and finally terminates at the lips as represented by the flat plane that is visible only in Figures 3a and 4a. Figure 3 was reconstructed assuming that the teeth did not exist (i.e. all of the space occupied by the teeth was segmented as air) while Figure 4 represents the same vowel shape but after using the dimensions of the teeth obtained from the EBCT images to estimate their extent and consequently eliminate their contribution to the MR derived air space image. Note that since the dimensions of the teeth cannot change, their dimensions measured from the EBCT data can also be used in segmenting any other vowel shape acquired with MR. Thus, one EBCT image set provides the necessary information for estimating the location and size of the teeth in any MR image set when relating the teeth dimensions to their soft tissue surroundings.



(a) (b)

Figure 1. Axial slices from the oral region of the vocal tract for the vowel /a/: a) EBCT, b) MRI.

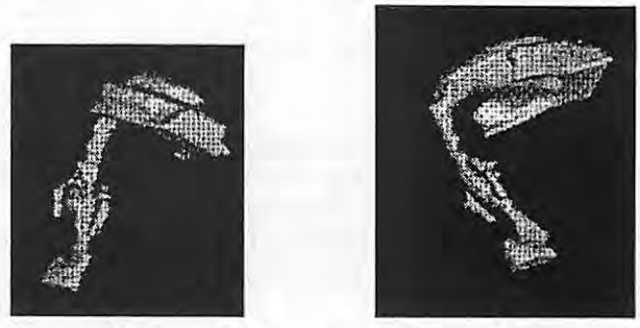


(a) (b)

Figure 2. A "digital cast" of the teeth using the EBCT image set for the vowel /a/: a) frontal view, b) slightly inferior view.

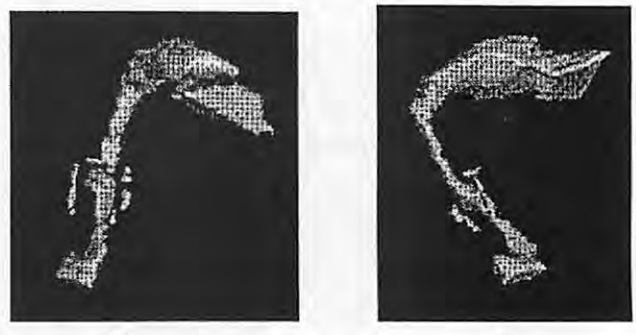
Comparison of Vowel Shape Reconstruction from EBCT and MRI

Utilizing the processes of segmentation, shape-based interpolation, and edge detection, the surface rendered airways of the EBCT and MRI versions of the /a/



(a) (b)

Figure 3. Shaded surface display (using MRI) of the vowel /a/ including the teeth: a) left tilted frontal view - flat surface in the upper right portion of the picture is the mouth termination plane, b) left tilted, inferior view - the "fins" that are visible on the left and right sides of the oral region are due to the included presence of the teeth.



(a) (b)

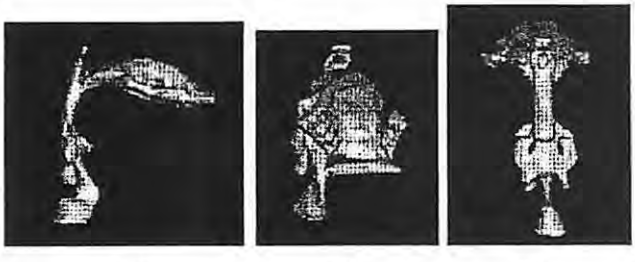
Figure 4. Shaded surface display (using MRI) of the vowel /a/ excluding the teeth by using the "digital cast" in Figure 2 to estimate their dimensions: a) left tilted frontal view - flat surface in the upper right portion of the picture is the mouth termination plane, b) left tilted, inferior view.

vowel were produced and are each shown from three perspectives in Figures 5 and 6, respectively. In general, the MRI shapes are similar to the EBCT versions but the higher resolution of the EBCT is apparent. Note also that the the cross-sectional area of the mouth termination appears to be larger in the MRI version than for the EBCT. This is due to the fact that the image sets for each of the two methods were acquired during entirely separate productions of the vowel (in fact the scanning for EBCT occured approximately 3 months prior to the MRI scans) and it is generally known that many variations of one particular vowel shape can produce an acceptable vowel sound. It is often the case that reduction of cross-sectional area in the front part of the tract (oral region) can be countered with an expansion of the lower or back part of the vocal tract (lower pharyngeal region). A close examination of Figures 5a and 6a suggests that, relative to the MRI, the EBCT version of the vowel does in fact have a slightly expanded pharyngeal section possibly to counter the smaller mouth termination.

The deformations of the airway due to the presence of the teeth are very apparent in the EBCT version of /a/ (Figure 5). The surface rendering of the /a/ vowel using MR (Figure 6) does not show the subtle indentations of the teeth but the volume consumed by the teeth appears to be largely removed. Figure 5 also shows that this production of the /a/ vowel was slightly nasalized as indicated by the presence of the short airway that branches off the main airway just as the vocal tract bends into the oral region. The small "hole" in the velar airway is likely caused by contact of the velum with the posterior pharyngeal wall.

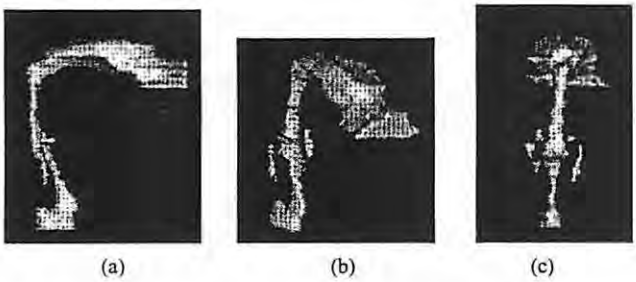
Other features of interest are the detail in which the epi-laryngeal tube, piriform sinuses, and valleculae are represented by the EBCT version and the lack of detail in which the MRI version represents these same structures. Figure 7 shows expanded views of this region of the vocal tract for both methods. It is obvious that the EBCT version provides an airway structure that is defined well enough to allow measurements of the cross-sectional areas and lengths of both the piriform sinuses and valleculae as well as the epi-laryngeal tube. Additionally, the location where the piriform sinus space joins the main vocal tract space is quite apparent. In contrast, the MRI version of this region of the tract is not as well represented. The piriform sinuses are visible but are shown disconnected from the main vocal tract. It is possible that during some types of vowel productions they could be "squeezed" off but a more plausible explanation is that the cross-sectional areas of the various airways in this region small enough that the blurring of the images due to the poor air-tissue interface obscures much of the important information. A hint of the valleculae is visible but certainly not with enough detail to make a meaningful measurement.

The surface renderings of the vowel /i/ are shown in Figure 8 for EBCT data and in Figure 9 for the MRI. Like the /a/ vowel, the general shapes of the /i/ for both of the methods are similar showing a widened pharynx and tightly constricted oral cavity. The presence of the teeth are not as apparent in Figure 8 as they were in Figure 5 because for an /i/ vowel the tongue blade is typically in contact with the molars and the tongue tip may contact the lower incisors, and hence there would be no air around them to be segmented. The upper incisors are the main portion of the teeth that are contact with the air and their presence is depicted by the slight indentation near the termination of the vocal tract which can be seen in Figure 8b. EBCT also brings out the contours of the pharyngeal walls and shows how the valleculae and piriform sinuses are reconfigured for the case of a



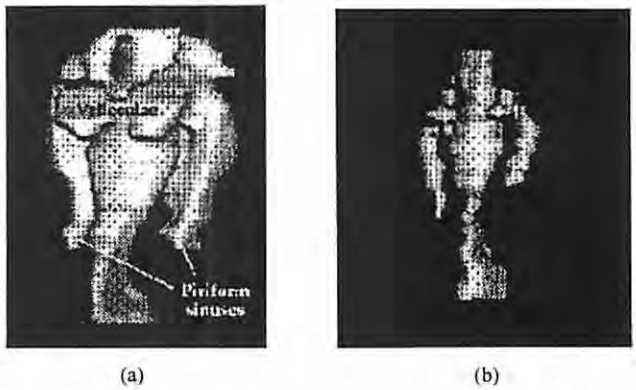
(a) (b) (c)

Figure 5. Shaded surface displays of the vowel /a/ using the EBCT image set: a,b, and c) various projections of the airway.



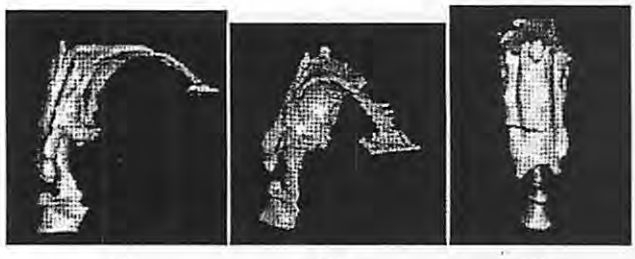
(a) (b) (c)

Figure 6. Shaded surface displays of the vowel /a/ using the MR image set: a,b, and c) various projections of the airway.



(a) (b)

Figure 7. Shaded surface displays of the vowel /a/ expanded in the epi-laryngeal and lower pharyngeal region: a) EBCT, b) MRI.



(a) (b) (c)

Figure 8. Shaded surface displays of the vowel /i/ using the EBCT image set: a,b, and c) various projections of the airway.

(a) (b) (c)

Figure 9. Shaded surface displays of the vowel /i/ using the MR image set: a,b, and c) various projections of the airway.

widened pharynx. The MRI version of /i/ shown in Figure 9 appears to provide a better representation of the piriform sinuses than for the /a/ vowel. This is due to the wider, more open, pharynx; i.e. the cross-sectional areas are larger so that the blurred air-tissue interface is less detrimental than in the previous case of the /a/. It also appears that the method of excluding the teeth by measuring their dimensions from the EBCT images worked well in the sense that the final tract shape in the oral region is quite similar to that reconstructed from the EBCT image set.

A Note on Area Functions and Vowel Simulations

The /a/ and /i/ vowel shapes acquired with both imaging methods were further analyzed to determine the "area function" using an iterative bisection tube geometry analysis technique[12,13,15,16]. An area function is defined as the cross-sectional area of the vocal tract as a function of the distance from the glottis. The area functions for each case are given in Figure 10. They begin at a point just above the glottis (at 0.0 cm) and progress out to the open mouth. Area functions derived from the MRI image sets are denoted by the solid line and the EBCT versions by the dashed line.

The /a/ vowel from the EBCT exhibits a piriform and valleculae region (located approx. 2-5 cm above the glottis) that is nearly three times larger than the MRI. The EBCT /a/ also shows larger area than the MRI throughout the pharyngeal and oral cavities. However, the cross- sectional

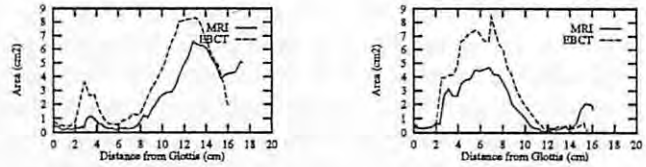


Figure 10. Area functions for /a/ and /i/ from MRI (solid) and EBCT (dashed) images: a) /a/, b) /i/.

area at the mouth termination drops to 1.8 cm2 for the EBCT and rises to 5 cm2 for the MRI. The two /i/ area functions are similar in the regions just above the glottis and also within the highly constricted region of oral cavity. However, the pharyngeal section is about 3 cm2 greater and the lip area about four times smaller for the EBCT than for the MRI. If the MRI area functions were uniformly smaller across the full length of the vocal tract, it would be suspected that the MRI derived areas had been underestimated (or possibly that the EBCT derived areas had been overestimated). A uniform difference is largely the case, except that the MRI area functions are larger at the lips than the corresponding EBCT area function; this is true for both vowels. The larger open mouth area and smaller pharyngeal region for the MRI area functions suggests the possibility that the subject s production of each vowel was slightly different during image acquisitions for each of the methods.

Subsequent comparisons of simulated vowel sounds, using the area functions in a wave-reflection type of vocal tract model[21,22] , and recorded natural speech sounds produced by the subject BS showed that the MRI based area functions actually produced a better spectral match (LPC based spectra) to the natural speech than did the EBCT versions[12]. It should be noted, however, that the natural speech was not recorded simultaneously with either of the imaging sessions. Thus, it is possible that, if the sound produced during the EBCT image acquisition had been recorded, a spectral analysis may have shown formant locations more in line with the EBCT based simulation. Unfortunately, data does not exist to test this hypothesis.

The idea that separate productions of the same vowel may utilize different area functions fits with the classic phenomena of the the "many-to-one" mapping for an acoustic tube; i.e. a particular acoustic waveform can be produced by an infinite number of acoustic tube configurations. Mermelstein[23] and Mrayati et al.[24] have both shown in detail how an acoustic tube can produce the same spectral pattern of acoustic resonances with different tract configurations. Mrayati et al. call this phenomena "region synergy" and they suggest that the vocal tract (or any acoustic tube) can be divided into regions that can cooperate with one another to maintain a particular formant pattern. One type of cooperation or "synergy" is to counter a decrease in area of a back section with an increase in area of a front section. From the area functions for the EBCT and MRI, it appears that such a manipulation of the vocal tract may have occured.

Conclusions

Both electron beam computed tomography (EBCT) and magnetic resonance imaging (MRI) have been used to image the vocal tract airway for the vowels /a/ and /i/ of one male subject. From the EBCT image sets both the airway and teeth were separately segmented. Measurements of the segmented teeth were used to aid in excluding them during the airway segmentation of the MR image sets. Shaded surface renderings of each EBCT based vocal tract shape showed subtle deformations of the space occupied by the teeth, contours of the oral and pharyngeal walls, and the detail of the region containing the epilarynx, piriform sinuses, and valleculae. In contrast, the surface renderings based on MR image sets show much less detail in all regions of the vocal tract but the similarity in gross vocal tract shape is certainly close enough to produce acceptable information for studies where minute detail is not important. The disadvantages of MRI are primarily a long scan time, somewhat blurred air-tissue interface, and no imaging of teeth. However, it poses no known risks to a human subject and is thus an attractive method for imaging the human vocal tract. The image quality that can be obtained from EBCT is superior to MRI because it produces a sharp definition of the air-tissue interface and images the teeth.

Scanning time is also greatly reduced. However, since EBCT uses ionizing radiation, only a limited number of full vocal tract volume scans are considered safe for one subject. Thus, EBCT is more suited to studies where only a few vocal tract image sets are collected but possibly for a larger pool of subjects. Because of the potential health risks posed by EBCT, future vocal tract imaging studies will probably continue to utilize MRI most heavily. However, using EBCT in a limited fashion may complement a more extensive data set acquired with MRI.

Acknowledgements

The authors would like to thank Carol Gracco for providing the initial MRI scanning parameters from which we developed those shown in Table 1. This study was supported in part by grant No. P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders.

References

1. Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W., "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," J. Acoust. Soc. Am., 90, 799-828, 1991.

2. Moore, C. A., "The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images," JSHR, 35, 1009-1023, 1992.

3. Sulter, A., Miller, D., Wolf, R., Schutte, H., Wit, H., and Mooyaart, E., "On the relation between the dimensions and resonance characteristics of the vocal tract: A study with MRI," Mag. Res. Imag., 10, 365-373, 1992.

4. Dang, J., Honda, K., and Suzuki, H., "Morphological and acoustical analysis of the nasal and the paranasal cavities," J. Acoust. Soc. Am., 96(4), 2088-2100, 1994.

5. Narayanan, S. S., Alwan, A. A., and Haker, K., "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am., 98(3), 1325-1347, 1995.

6. Narayanan, S. S., "Fricative consonants:An articulatory, acoustic, and systems study," Ph. D. thesis, UCLA, Dept. Elect. Engineer., Los Angeles, CA.

7. Beautemps, D., Badin, P., and Laboissiere, R., "Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data," Speech Communication, 16, 27-47, 1995.

8. Greenwood, A. R., Goodyear, C. C., and Martin, P. A., "Measurements of vocal tract shapes using magnetic resonance imaging," IEEE Proceedings-I, 139(6), 553-560, 1992.

9. Lakshminarayanan, A., Lee, S., and McCutcheon, M., "MR imaging of the vocal tract during vowel production," J. Mag. Res. Imag., 1(1), 71-76, 1991.

10. Kiritani, S., Tateno, Y., Iinuma, T., and Sawashima, M., "Computer tomography of the vocal tract," in Dynamic Aspects of Speech Production, M. Sawashima and F. Cooper (Eds), Univ. of Tokyo Press, Tokyo, 203-206, 1977.

11. Perrier, P., Boe, L-J, and Sock, R., "Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: Modeling the transition with two sets of coefficients," J. Speech and Hearing Research, 35, 53-67, 1992.

12. Story, B. H., "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph. D. Dissertation, University of Iowa, Dept. of Speech Pathology and Audiology, 1995.

13. Story, B., Titze, I., and Hoffman, E., "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am., (in review), 1996.

14. Boyd, D. P., and Lipton, M. J., "Cardiac computed tomography," Proceedings of the IEEE, 71, 298-307, 1983.

15. Story, B. H., Hoffman, E. A., and Titze, I. R., "Speech simulation based on MR images of the vocal tract," SPIE Proc. Physiology and Function from Multidimensional Images, 2433, San Diego, CA, 26 Feb. - 2 Mar., 1995.

16. Hoffman, E. A., Gnanaprakasam, D., Gupta, K. B., Hoford, J. D., Kugelmass, S. D., and Kulawiec, R. S., "VIDA: An enviroment for multidimensional image display and analysis," SPIE Proc. Biomed. Image Proc. and 3-D Microscopy, 1660, San Jose, CA, 10-13 Feb., 1992.

17. Hoffman, E. A., Sinak, L. J., Robb, R. A., and Ritman, E. L., "Noninvasive quantitative imaging of shape and volume of lungs," American Physiological Society, 1414-1421, 1983.

18. Udupa, J. K., "Computer aspects of 3D imaging in medicine: A tutorial," in 3D Imaging in Medicine, J. K. Udupa and G. T. Herman (Eds.), CRC Press, 1-64, 1991.

19. Mermelstein, P., "Articulatory model for the study of speech production," J. Acoust. Soc. Am., 53(4), 1070-1082, 1973.

20. Raya, S. P., and Udupa, J. K., "Shape-based interpolation of multidimensional objects," IEEE Trans. Med. Imag., 9, 32-42, 1990.

21. Kelly, J. L., and Lochbaum, C. C., "Speech Synthesis," Proc. 4th Int. Congr. Acoust., Paper G 42, 1-4, 1962.

22. Liljencrants, J., "Speech Synthesis with a Reflection-Type Line Analog," DS Dissertation, Dept. of Speech Comm. and Music Acous., Royal Inst. of Tech., Stockholm, Sweden, 1985.

23. Mermelstein, P., "Determination of vocal tract shape from measured formant frequencies," J. Acoust. Soc. Am., 41(5), 1283-1294, 1966.

24. Mrayati, M., Carre, R., and Guerin, B., "Distinctive regions and modes: A new theory of speech production," Speech Comm., 7, 257-286, 1988.

NCVS Status and Progress Report - 9
April 1996, 133-141

Volumetric EBCT Imaging of the Vocal Tract Applied to Male Falsetto Singing

Kenneth Tom, Ph.D.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Eric Hoffman, Ph.D.
Department of Radiology, The University of Iowa Hospitals and Clinics
Brad Story, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

As part of an analysis by synthesis approach to studying vocal intensity control in falsetto register, volumetric imaging of the vocal tract (the upper airway from the glottis to the lips) using electron beam computed tomography was performed on a classically trained singer, a countertenor, who uses a falsetto singing technique. Eight pitch and loudness conditions were imaged, a subset of which will be presented here. Each set of scans consisted of contiguous 3 mm axial "slices" encompassing the arch of the hard palate superiorly and the first tracheal ring inferiorly. Images were analyzed in three stages: image segmentation, three-dimensional airway reconstruction and airway measurement. The vocal tract airway was segmented from surrounding tissue by assigning airway voxels a unique gray scale value. Reconstruction of the vocal tract in three dimensions was accomplished using shape based interpolation on the segmented images. Cross-sectional areas and vocal tract length were acquired from shape based interpolated data. Vocal tract area functions derived from these measurements were used to simulate the subject's phonations, which in turn allowed estimation of glottal and supraglottal contributions to vocal intensity.

Vocal Tract Imaging and Voice Production in Falsetto Singing

Speech and song produced in falsetto register has long been a part of many of the world's cultures. Functionally, falsetto phonation represents the upper third to two-thirds of the pitch range in adult males, a significant portion of the speech mechanism's physiologic range. Nonetheless, our understanding of the physiology of vocal intensity control in falsetto register is limited. In the area of speech physiology, there are few comprehensive studies regarding vocal intensity control in falsetto speech or singing. Most studies examining physiologic control of vocal intensity *per se* have avoided high pitched or falsetto phonation, focusing on relatively low pitched speech-type phonation in the chest register[1-3]. This is to a large degree due to practical and theoretical limitations of currently available measurement and analysis techniques used to obtain estimates of glottal behavior and vocal tract resonances. As a result, detailed information for higher pitched phonation in falsetto register has been virtually unavailable. The goal of the current study is to examine the intensity control mechanism in male falsetto singing using volumetric imaging of the vocal tract in the context of an analysis by synthesis approach.

Vocal intensity control can be understood in terms of the specific contributions of the three physiological subsystems involved in producing voice: the lungs, the vocal folds, and the supraglottal articulatory vocal tract. To quantify their relative contributions to changes in vocal intensity, changes in (1) lung (subglottal) pressure, (2) vocal fold adduction and (3) the effect of structural changes in the articulatory vocal tract need to be measured at different pitch and loudness conditions. Techniques to measure or estimate the first variable, subglottal pressure, are readily available. Direct measurement of transglottal airflow, from which glottal adductory measures can be derived, is not feasible. For lower pitched (below approximately 300 Hz) chest register voice, the latter two measures (vocal fold adduction

and the effect of articulatory vocal tract changes) have been estimated with some success using acoustic filtering techniques to identify vowel resonances (formant frequencies) of the vocal tract.

The effectiveness of such techniques, e.g. digital linear predictive coding (LPC) methods, in estimating the location and bandwidth of vocal tract resonances (formants) from the spectrum of the wide bandwidth oral airflow waveform depends on having a relatively high density of harmonics with which to "sample" the resonances, a situation which occurs in low pitched phonation. Once these formants are specified, the time-varying glottal airflow can be estimated by inverse filtering the formants; i.e., removing the contributions of vocal tract resonances from the wide bandwidth oral airflow. The duty cycle (open quotient) of this time-varying glottal airflow is a measure related to the degree of laryngeal valving or vocal fold adduction. Changes in open quotient from one phonatory condition to another can indicate changes which occur at the level of laryngeal valving. With higher pitched falsetto register voice, however, LPC techniques as described above generally fail. There are far fewer harmonics with which to "sample" formants; at very high pitches, there may be no harmonics whatsoever within the bandwidth of a formant.

In order to avoid the limitations of current linear inverse filtering approaches to estimating time-varying glottal airflow, an analysis by synthesis paradigm using a physiologic model of voice synthesis was implemented. In this model[4] (SPEAK, v. 2.0), glottal dynamics interact with acoustic pressures from the vocal tract. Use of the model required "fitting" the vocal tract in the model to that of the subject. This entailed the acquisition of detailed physical measurements of the subject's vocal tract (length, cross-sectional areas) for each phonatory condition; hence the need for volumetric imaging. Actual phonatory measures from the subject for subglottal pressure and fundamental frequency were used for those parameter values in the synthesis. Glottal parameters were adjusted such that electroglottographic (EGG) waveforms (implied vocal fold contact patterns) and glottal configuration of the model would be as close as possible to the subject's EGG and videostroboscopic data. Simulated glottal airflow waveforms produced by the model would "parallel" those of the subject and represent a form of "non-linear inverse filtered" flow. Relative glottal adduction, which is one of the three salient variables in intensity variation, could then be derived from the duty cycle of the simulated time-varying glottal airflow.

A boost in acoustic intensity can also occur as a result of structural manipulation of the articulatory tract[5]. The shape and length of the vocal tract can be altered by positioning the lips, tongue and jaw such that a resonant frequency of the vocal tract approaches or overlaps an harmonic of the vibratory source formed by the vocal folds. In order to determine the effects of such "formant tuning," formant frequencies and bandwidths need to be specified. Using measurements based on volumetric images of the vocal tract during a particular phonatory condition, a vocal tract analogous to the subject's vocal tract during that phonation can be modeled. Vowel formant characteristics can be observed in the frequency response (to a pulse excitation) of the modeled vocal tract.

In this manner, volumetric imaging of the vocal tract makes it possible to estimate two of the three production variables related to vocal intensity control. Detailed measurements of the vocal tract, specifically the length and cross-sectional areas along that length, make it possible to use an analysis by synthesis approach to estimate (1) glottal adduction as well as (2) the degree to which articulatory changes contribute to overall intensity.

Image Acquisition

Three-dimensional imaging of the vocal tract can be accomplished with either electron beam computed tomography (EBCT) or magnetic resonance imaging (MRI). For highest resolution imaging of the vocal tract EBCT is preferable. Compared with MRI imaging, the air-tissue boundary is captured with greater resolution, and bony tissue without hydrogen content, such as teeth, are more effectively imaged with EBCT. A sixty-slice volumetric study of contiguous 3 mm slices can be acquired in about 45 seconds, which also reduces the chance of movement artifact blurring images. The disadvantage associated with EBCT is its use of ionizing radiation, which therefore limits the number of scans considered safe. In terms of reducing the risk of any adverse side effects, MRI has the clear advantage. No long-lasting, hazardous biologic effects have been observed in humans from short term exposures to the magnetic fields used in commercially available MRI scanners.

Images of static vowel shapes have been acquired using MRI and reconstructed in three dimensions[6-7]. Both of these studies noted the relatively long image acquisition times required. Story [8] recently used MRI imaging parameters that were an improvement over those used in past studies. Nonetheless, these image sets still required 5-6 minutes of actual scanning time. Given pauses for inhalations required for reiterations of a target phoneme, image capture involved well over 7-8 minutes of repeated execution of a phonation task, all while trying to avoid movement of vocal tract structures. Since only tissue with hydrogen content is imaged in MRI, gray values associated with teeth are difficult to distinguish from airspace. Movement artifact also becomes a great concern when a phonation must be reiterated 30 times or more.

Pilot studies performed with MRI by the authors produced severely blurred images due to the nature of the singing tasks required (7-8 minutes of repeated high effort phonation). It was concluded that MRI was not a viable alternative for our imaging needs.

Subject Selection

For the purposes of this study, a subject who was able to readily and consistently produce falsetto phonations throughout his fundamental frequency and intensity ranges was required. Since it was not technically feasible to record acoustic, aerodynamic, videostroboscopic and physiologic data simultaneous to volumetric imaging, the participation of a subject who is accustomed to producing repeated falsetto phonations in a consistent manner across different acoustic environments greatly increased the reliability of data recordings. Singers trained in classical vocal technique acquire vocal production habits which allow them to produce tones of similar frequency, intensity and timbre reliably across acoustic environments. Therefore, the subject chosen for this study was a 45 year old adult male who has had extensive singing training in the Western classical tradition, including an earned Doctor of Musical Arts degree, 12 years of vocal study as a baritone, and 6 years of study as a countertenor. He is an active performer and voice teacher. He has sung as a countertenor for the past 11 years performing early music using a falsetto singing technique to vocalize in the alto range. Medical and recent health history was unremarkable. The subject has no history of speech or hearing disorders, nor cardiac abnormalities. His speaking range is within the normal baritone range. The subject's fundamental frequencies range from $C\#_2$ (69 Hz) to G_4 (392 Hz) in chest register and from D_3 (147 Hz) to D_5 (587 Hz) in falsetto register. He is a native speaker of General North American English.

EBCT Image Acquisition

In order to minimize risks associated with radiation exposure, the total number of phonatory conditions to be imaged (pitch/loudness conditions) was limited to eight volume sets. The vowel /a/ sung in falsetto register at three pitch levels (low/C_4- 262 Hz, medium/F_4-349 Hz, and high/ B-$flat_4$-466 Hz), each at moderately low intensity (*mezzopiano*) and very loud intensity (*fortissimo*) and spoken in chest register at comfortable and loud intensity levels, were scanned using an Imatron C-150 scanner[9]. The scout scan preceded the eight image sets. The human equivalent radiation exposure for the entire protocol (eight volume studies and a scout scan) was calculated to be 0.17 rem/rad. A subset of these will be presented here.

At each phonatory condition to be scanned, the subject was positioned comfortably in a supine position on the imaging table. His neck was supported and stabilized with a rolled towel. The head position was realigned for each scan such that the Frankfort plane was perpendicular to the imaging table, and a light beam locator was centered between the pupils of the subject's eyes. Each phonatory condition required approximately 45 seconds of actual scanning time. This required numerous reiterations of the target vowel with pauses for inhalation before each one. Before each phonatory condition was scanned, the subject performed a few trial utterances to gauge how many seconds he could comfortably prolong the utterance for that pitch/ loudness condition. The radiology technician timed pauses during scanning such that he stopped before the end of a vowel reiteration and reinitiated imaging as soon as the subject began the next reiteration, as monitored over the intercom. This avoided the use of any foot or hand signaling which could have caused movement artifact.

Imaging Parameters: Sixty 3 mm contiguous slices encompassing at least the hard palate superiorly, the first tracheal ring inferiorly, the lips anteriorly, the posterior pharyngeal wall posteriorly, and the buccal walls of the vocal tract left and right were acquired in each volume set. The pixel matrix was 512x512 and the field of view (FOV) for each slice was 21 cm. The resolution in the plane of imaging was 0.41 mm, which is near the theoretical limit of the scanner's resolution.

The accuracy of the image acquisition and analysis procedures using the Imatron C-150 scanner has been assessed by imaging and analyzing a tubular phantom of known dimensions[8]. The phantom consisted of three connected sections of air-filled tubing set in a closed water filled plastic enclosure. Known and measured cross-sectional areas of the phantom differed by 1.8% - 2.0%.

Image Analysis

Image analysis was accomplished in three stages: image segmentation, three-dimensional airway reconstruction, and airway measurement. These procedures were performed using an image display and measurement package called VIDA™ (Volumetric Image Display and Analysis), which was developed by Hoffman and colleagues[10]. The analysis techniques have been described in detail by Story[8]. A general manual, which explains the use of all VIDA modules can be accessed via the Internet at the following URL address: http://everest.radiology.uiowa.edu:8080.

Airway Segmentation

The axial image data were transferred via magnetic tape from the Imatron C-150 scanner to a computer workstation for analysis. Figure 1 is an example of an axial (in the imaging plane) EBCT image of the lower pharyngeal airway above the glottis. The vocal tract airway was segmented or differentiated from surrounding tissue by assigning all airway voxels to a unique gray scale value. In this case, the brightest color value was chosen. Once a 5% reduction had been made on the entire image set, this brightest value could be reassigned. A threshold value for the border between tissue and air was determined by evaluating brightness densities in an area of tissue to air transition and selecting the 50% point between the brightest area of tissue and the darkest area of the airway. Once this threshold was determined, actual segmentation could begin.

Figure 1. EBCT Image - Axial Slice.

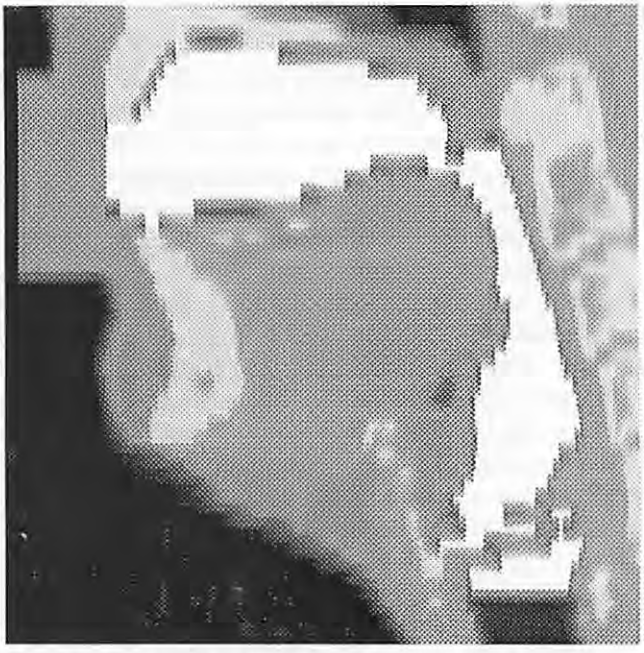


Figure 3. Rectangle painted to define mouth termination brightest pixels = vocal tract.

Figure 2. Segmentation of axial slice light scatter from dental work.

The process of segmentation consists of the following sequence of steps: (1) a mouth termination plane was determined using the procedure defined by Mermelstein[11] and "painted" on the midsagittal plane, as illustrated in figure 3. When this painted bar was extended to all slices, it served as a delimiter for the anterior portion of the mouth, fencing off the oral airspace. (2) Because of lateral rounding of the lips, there can still be areas of "leakage" from the oral to the outside airspace, which confuses the automatic algorithm for re-coloring (2-D Fill) the vocal tract airspace to its unique color value. To address this, such airway "leaks" were contained by drawing a line using the same color value as the "fence" rectangle to close the gap on axial slices in which this occurs. This procedure was somewhat subjective and consequently introduced some error in the mouth region. (3) In addition to closing such airway "leaks" into the outer airspace, manual editing was also required to address light scatter created by imaging of metallic dental fillings. Figure 2 illustrates the large amount of such scatter that can occur on axial slices in the mouth region. Knowledge of the subject's dentition and familiarity with patterns of scatter typical of such images aided in the admittedly subjective procedure of outlining the air-tissue border manually, where light scatter obscured the threshold value for automatic detection. This may have contributed additional error in the mouth region. Once these editing steps were completed, automatic detection of vocal tract air space on each axial slice served to define the entire vocal tract as one unique brightness level.

Volume Reconstruction

Reconstruction of the vocal tract in three dimensions was accomplished by using a process called shape based interpolation[12] on the segmented image set. Rather than interpolating gray scale values as in cubic voxel interpolation, shape based interpolation defines the shape of a structure by its edges. The gray scale values associated with the edges of the airway's shape are selected and all other

Figure 4. Volume reconstruction sagittal view with piriform sinuses.

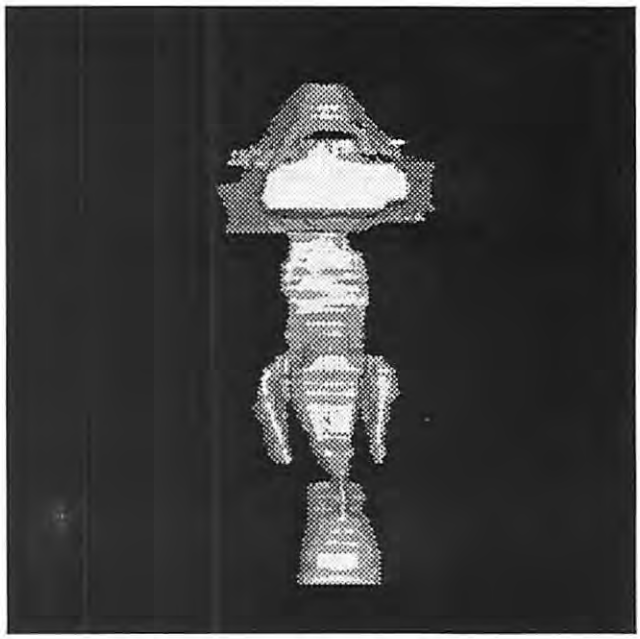


Figure 5. Volume reconstruction anterior view with piriform sinuses.

values ignored. This has the effect of "sculpting out" the vocal tract shape from the surrounding tissue and providing the same voxel resolution (0.41 mm) in the axial dimension as in the coronal and sagittal dimensions.

The edges of the interpolated airway can be used to perform surface renderings of the vocal tract, creating a "digitally molded sculpture" of the airway. Graphically represented as a three-dimensional object using shading, surface renderings can be displayed at any number of angles or magnification levels. The display cannot be measured directly but allows the user to assess the quality of the segmentation procedure and to observe three-dimensional views of the vocal tract's outer shape. Figures 4 through 6 show surface renderings of the vocal tract (with piriform sinuses intact) associated with the B-flat$_4$ (466Hz) moderately soft condition from sagittal, anterior and posterior orientations.

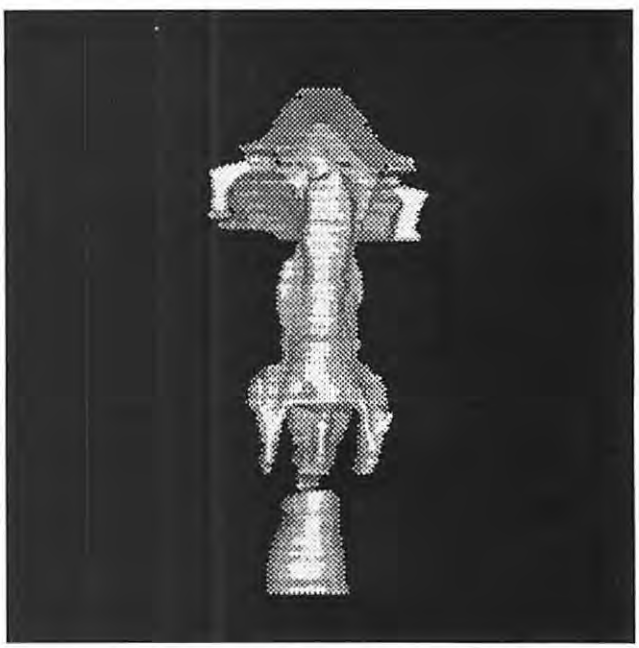


Figure 6. Volume reconstruction posterior view with piriform sinuses.

Vocal Tract Measurement

To acquire area functions (cross-sectional areas) from the shape based interpolated data set, an algorithm originally designed to study the upper airway during sleep[13] was used. This algorithm computes cross-sectional areas from oblique sections perpendicular to the local airway long axis. The user defines the end points of a tube-shaped structure, in this case the mouth and the glottis, and the iterative bisection algorithm computes a series of centroids for oblique cross sections of the tube. The three-dimensional spatial coordinates of these centroids can be triangulated using a three-dimensional Pythagorean theorem, and once "connected" become the measured centerline of the tube being analyzed. Vocal tract lengths were recalculated averaging lateral coordinates (x-coordinates) to reduce sharp short-term directional changes in the centerline that add unrealistic acoustic length. Acoustically propagating waves in the vocal tract (of the wavelengths related to frequencies found in the spectrum of speech) do not travel following small directional perturbations in the lateral dimension.

The series of calculations used to model wave propagation in the computer simulation requires the use of an even number of equally long sections. The sampling

frequency used in the computer simulation (44.1 kHz) also requires that each segment be 0.396825 cm thick. Consequently the length of the final discretized vocal tract must be an even integer multiple of 0.396825 cm. A vocal tract of 42 sections, for example, would be 16.66 cm in length; 44 sections 17.46 cm in length; and 36 sections 18.25 cm in length.

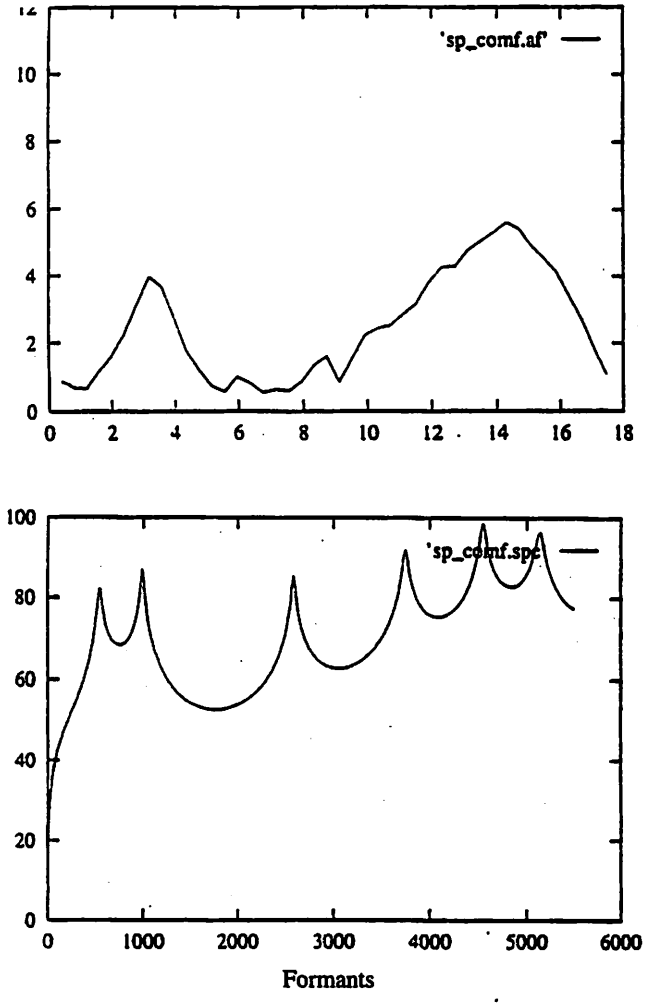
The process of discretization consisted of (1) making a choice from the discretized vocal tract lengths available to best fit the measured data, (2) normalizing measured data to this length, (3) fitting the data to a cubic spline curve and (4) sampling the cubic spline curve at equally spaced intervals of 0.396825 cm.

Piriform sinuses: The piriform sinuses could not be effectively measured using the iterative bisection method. However, since the axis of wave propagation in the piriform sinuses is roughly perpendicular to the axial plane, a pixel counting method could be used to measure cross-sectional areas. Area functions were obtained by sampling every five slices (every 2.05 mm) from the interpolated image data.

Results

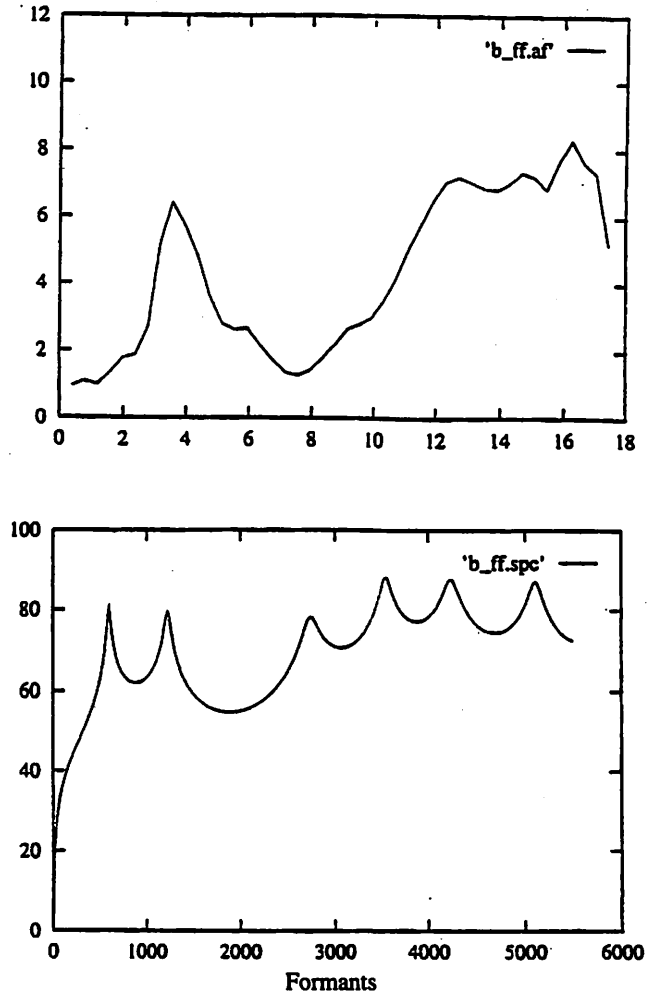
Surface renderings

Surface renderings of volume reconstructions for the B-flat $_4$ (466 Hz), moderately soft condition are presented in figures 4-6; which are sagittal, anterior and posterior views, respectively. The airspace of the vocal tract is represented graphically as an apparently solid gray-scale object, with the articulatory speech organs "stripped away". In figure 4 the rising and falling curve at the topmost portion of the object is the shape of the airspace beneath the arch of the hard palate and velum. In the same figure, the rounded



F1	543.00 Hz
F2	993.42 Hz
F3	2585.22 Hz
F4	3747.33 Hz

Figure 7. Comfortable speech, chest register area functions (above); frequency response (below).



F1	600.64 Hz
F2	1229.63 Hz
F3	2751.09 Hz
F4	3552.55 Hz

Figure 8. B-flat$_4$ (466 Hz) very loud, falsetto register area functions (above); frequency response (below).

fist-like shape in the upper portion of the figure is the air within the buccal cavity. Its inferior edge, which curves downward into a plateau and then rises again into an arc, is the shape of the air superior to the tongue. The blunt termination plane of the oral airway, per Mermelstein's procedure[11] can be seen in figures 4 and 5. In figures 5 and 6, the two cone shaped objects in the lower half of the image lateral to the main airway are the piriform sinuses. In figure 6, the two short projections on the left and right sides of the buccal cavity are the air pockets between the laterally positioned teeth/gum ridges and the mucosa of the buccal walls. Figure 6 also offers a detailed view of the transition of the lower posterior pharynx into the piriform sinuses.

Vocal Tract Length

Table 1 lists the vocal tract lengths as measured and smoothed in the x-direction, and the differences between measured and the normalized (17.46 cm) length used in simulation.

Table 1.
Vocal Tract Lengths

Condition	Measured Lengths	Difference from Normalized Length
Bb_4ff	18.16 cm	+0.70 cm
Bb_4mp	17.02 cm	-0.44 cm
F_4ff	17.48 cm	+0.02 cm
F_4mp	16.99 cm	-0.47 cm
C_4ff	16.81 cm	-0.65 cm
C_4mp	17.29 cm	-0.17 cm
Speech loud	16.55 cm	-0.91 cm
Speech comf	17.39 cm	-0.07 cm

*ff = very loud, mp = moderately soft

Reliability Measures

Intra-examiner reliability was assessed by having the first author segment, reconstruct and measure the vocal tract associated with the F_4/349 Hz, very loud condition at three different times. On each occasion, he began with the original EBCT images. Measured vocal tract lengths were 17.58 cm, 17.43 cm, and 17.39 cm on these three analyses. Maximum normalized error in this measure was 1.09%.

Cross-Sectional Areas

Figures 7 and 8 illustrate the area functions (upper diagrams) and the frequency responses corresponding to the area functions (lower diagrams) for two of the eight phonatory conditions. The upper diagrams indicate changes in cross-sectional area along the length of the vocal tract. Beginning on the left, measured units on the x-axis represent distance in centimeters above the glottis. On the y-axis, measured units represent cross-sectional area in square centimeters. The lower diagrams show the frequency response of each vocal tract with frequency (in Hz) plotted on the x-axis and intensity (in dB power) on the y-axis.

Figure 7 represents speech phonation of the vowel /a/ in chest register at a comfortable intensity level. The fundamental frequency was approximately 117 Hz. The cross-sectional area (upper diagram) just above the glottis is approximately 1 cm^2, then drops slightly before climbing to a peak of 4 cm^2 at a point 3.5 cm above the glottis. There is an almost symmetric decline in area until just below 5 cm above the glottis. The cross-sectional area from just below 5 cm above the glottis to 9 cm above the glottis remains small, less than 2 cm^2. It then increases gradually from 1 cm^2 to 5.8 cm^2 from 9 cm above the glottis to just above 14 cm above the glottis. The area then gradually reduces to about 1 cm^2 at the mouth termination.

Figure 8 represents a high-pitched sung phonation of the vowel /a/ in falsetto register at a very loud intensity level. The fundamental frequency was 466 Hz (B-flat$_4$). The cross-sectional area (upper diagram) just above the glottis is approximately 1 cm^2, rises to 2 cm^2 at 2 cm above the glottis, then very sharply increases to 6.2 cm^2 just below 4 cm above the glottis. There is an almost symmetric sharp decline in area to less than 2 cm^2 from 4 cm above the glottis to 7.5 cm above the glottis. The area increases from 2 cm^2 to 7 cm^2 from the point 8 cm above the glottis to 12.5 cm above the glottis, then varies within a smaller range until it closes down to an area of approximately 5 cm^2 at the mouth termination.

For the high pitched, very loud sung falsetto tone, the increase in cross-sectional area in the laryngopharynx 3-4 cm above the glottis is noticeably larger than in speech phonation at a comfortable intensity. The cross-sectional area of the buccal cavity overall and the mouth opening are also noticeably larger.

Formant Structure

The first four vowel resonances or formants of each phonatory condition are listed below each frequency response (lower diagram) in figures 7 and 8. The first and second formants are required for vowel perception and the third and forth contribute mostly to overall timbre.

For the comfortable speech condition (figure 7), the first formant (F1) was 543 Hz, the second formant (F2) 993 Hz and the third formant (F3) 2585 Hz. Normative average values for adult male F1, F2 and F3 for the vowel /a/ are 730 Hz, 1090 Hz, and 2440 Hz, respectively[14]. The subject's values for both F1 and F2 are noticeably lower; almost 200 Hz lower for F1 and 100 Hz lower for F2. F3 for the subject was more than 100 Hz above the norm. The subject's F1 and F2 values are indicate a more rounded /a/ allophone.

For the high pitched very loud sung falsetto tone (figure 8), the first formant (F1) was 601 Hz, the second formant (F2) 1230 Hz and the third formant (F3) 2751 Hz. Normative average values for adult male F1, F2 and F3 for the vowel /a/ are 730 Hz, 1090 Hz, and 2440 Hz, respectively[14]. The subject's F1 was more than 100 Hz below the

norm. The F2 and F3 were both higher than the norm. The F1 and F2 indicate a phonetically neutralized vowel quality, probably due to the more open buccal cavity.

There is more frequency separation between F1 and F2 and less frequency separation between F3 and F4 in the high pitched very loud sung falsetto tone. The higher formants in this condition (figure 8 lower diagram) F3, 4 and 5, appear to be more closely clustered than in the comfortable speech condition. The countertenor subject, who was trained previously for many years as a baritone, may be using an expanded supraglottic posture associated with the so-called "singer's formant" in the 3000 Hz region when singing tones at high intensity, even in falsetto register.

Discussion

The use of volumetric imaging has been essential in our attempt to circumvent the limitations of current techniques for estimating transglottal flow and formant identification. In our analysis by synthesis approach, vocal tract measurements acquired from volumetric imaging provided input for a physiologic computer model, which then simulated the subject's phonatory behaviors. This procedure enabled us to accomplish two goals: (1) to simulate time-varying glottal flow analogous to that of a specific subject, (first order estimates of glottal adduction are derived from simulated glottal flow) and (2) to specify formant characteristics, which makes it possible to map out the proximity of vocal tract resonances and voice source harmonics with regard any formant tuning effects.

Volumetric imaging of the vocal tract during actual phonation results in a 3-D volume which is effectively a composite of several iterations of a particular phonatory target. By keeping imaging time to a minimum and using a subject whose vocal training lends itself to producing consistent repetitions, we were able to acquire images of excellent quality. The high resolution of the EBCT images provided vocal tract measurements that were more than adequate for our purposes. Although overall results varied, our efforts at simulating the subject's phonations were moderately successful. We were able to "match" subject utterances with good to very good results, using acoustic spectra as the primary "matching" criterion and subsequently derive estimates of glottal and supraglottal behaviors.

Although the results of the entire study are based on data from one subject, the data may not be atypical of male speakers with a similar, but untrained falsetto phonation pattern, trained countertenors or other falsetto singers. A comprehensive description of falsetto intensity control from a subject who has developed an optimal technique for producing falsetto phonations across a range of frequencies and intensity levels may give us insight into a falsetto type that is not only reliable but perhaps vocally healthier than other falsetto phonation types (that is less injurious to the tissues of the vocal folds). The results of this study may therefore have implications for the use of falsetto phonation in voice therapy and in the training of singing voice.

It is probable that a number of different strategies exist for controlling vocal intensity in the falsetto register, each with unique advantages and disadvantages. Further studies are needed to examine such intra-individual and inter-individual differences in vocal intensity control strategies, not only to acquire a better understanding of the relevant peripheral mechanisms, but to guide clinical practitioners and vocal trainers in how they use "falsetto voice" in their work.

Acknowledgements

This research was supported in part by research grant number 5 R01 DC 00387-08 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

We would like to express our gratitude to Dr. William Stanford, Dr. Brad Thompson, and the members of the Division of Physiologic Imaging, Department of Radiology at the University of Iowa College of Medicine, especially Gopal Sundarmoorthy, Jehangir Tajik and Neil D'Souza for their contributions to this study.

References

1. E. Holmberg, R. Hillman and J. Perkell, "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," *Journal of the Acoustical Society of America*, Vol. 84 (2), pp. 511-529, 1988.

2. E. Holmberg, R. Hillman, J. Perkell, P.C. Guiod and S.L. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustical spectral measures of female voice," *Journal of Speech and Hearing Research*, Vol. 38, pp. 1212-1223, 1995.

3. E.T. Stathopoulos and C.M. Sapienza, "Respiratory and laryngeal function of women and men during vocal intensity variation," *Journal of Speech and Hearing Research*, Vol. 36, pp. 64-75, 1993.

4. I.R. Titze, SPEAK, version 2.0 (software), 1995.

5. J. Sundberg, "Formant technique in a professional female singer," *Acustica*, Vol. 32, pp. 89-96, 1975.

6. T. Baer, J.C. Gore, L.C. Gracco and P.W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *Journal of the Acoustical Society of America*, Vol. 90 (2), pp. 799-828, 1991.

7. C.A. Moore, "The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images," *Journal of Speech and Hearing Research*, Vol. 35, pp. 1009-1035, 1992.

8. B.H. Story, *Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract* (unpublished doctoral dissertation), University of Iowa, 1995.

9. D.P. Boyd and M.J. Lipton, "Cardiac computed tomography," *Proceedings of the IEEE*, Vol. 71, pp. 298-307, 1983.

10. E.A. Hoffman, D. Gnanaprakasam, K.B. Gupta, J.D. Hoford, S.D. Kugelmass and R.S. Kulawiec, "VIDA: an environment for multidimensional image display and analysis," *SPIE Proceedings, Biomedical Image Processing and 3-D Microscopy*, Vol. 1660, pp. 1-19, 1992.

11. P. Mermelstein, "Articulatory model for the study of speech production," *Journal of the Acoustical Society of America*, Vol. 53, pp. 1070-1082, 1973.

12. J.K. Udupa, "Computer aspects of 3-D imaging in medicine: A tutorial," in J.K. Udupa and G.T. Herman (Eds.), *3D Imaging in Medicine*, Boca Raton, CRC Press, 1991.

13. E.A. Hoffman and W.B. Gefter, "Multimodality imaging of the upper airway: MRI, MR spectroscopy, and ultrafast x-ray CT," in F.G. Issa, P.M. Suratt, and J.E. Remmers (Eds.), *Sleep and Respiration*, New York, Wiley-Liss, Inc., 1990.

14. G.E. Peterson and H.L. Barney, "Control methods used in a study of vowels," *Journal of the Acoustical Society of America*, Vol. 24, pp. 175-184, 1952.

NCVS Status and Progress Report - 9
April 1996, 143-158

Anatomy and Physiology of Normal and Disordered Velopharyngeal Function for Speech

Jerald Moon, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
David Kuehn, Ph.D.
Department of Speech and Hearing Sciences, The University of Illinois at Urbana-Champaign

The production of speech involves the coordinated activity of a number of structures within the vocal tract. Those moving structures lying above the larynx that are used to modify the airstream are referred to as articulators. Movements of the lips, tongue, and jaw modify vocal tract geometry, resulting in time varying alterations in vocal tract resonance properties. Of no less importance as an articulator is the velopharyngeal mechanism. The velopharyngeal mechanism is a muscular valve that modulates the degree of coupling between the oral and nasal cavities. Described simply, it is open for nasal sounds and closed for oral sounds. However, as will be outlined in this chapter, its control is not quite so simple. Its functional status becomes an issue clinically in those cases where the valve does not, or cannot, adequately separate the oral and nasal cavities. The consequences of such dysfunction will be addressed in other chapters.

For those interested in helping patients with dysfunctional velopharyngeal mechanisms either surgically, prosthetically, or behaviorally, a good working knowledge of normal velopharyngeal anatomy and physiology is essential. One needs also to be knowledgeable about the coordination of velopharyngeal function with other systems that are involved in speech production (i.e. respiratory system, larynx, other articulators). Not unlike the auto mechanic setting out to diagnose and then repair a defective engine, the speech-language pathologist needs to know what the parts are, how they should work together, what is likely to happen within the engine should a part or parts fail to perform properly, and what the consequences are on overall performance.

The goal of this chapter is to present a) the anatomy of the normal velopharyngeal mechanism, b) physiologic aspects of normal velopharyngeal function, and c) an introduction to the consequences of palatal clefting and its repair on the form and function of the velopharyngeal valve.

Normal Velopharyngeal Anatomy

The anatomy of the velopharynx has been described in detail elsewhere (Cassell, Moon, and Elkadi, 1990; Cassell and Elkadi, 1995; Dickson, 1975; Dickson and Dickson, 1972; Dickson, Grant, Sicher, DuBrul, and Paltan, 1974; Kuehn, 1979; Kuehn and Dalston, 1988). Only the most salient aspects will be discussed here.

Surface Anatomy

The velopharyngeal mechanism is a muscular valve which extends from the posterior surface of the hard palate to the posterior pharyngeal wall. The mechanism is comprised of the soft palate, or velum, the lateral pharyngeal walls, and the posterior pharyngeal wall. It is situated in that portion of the vocal tract referred to as the velopharynx (Figure 1). The anterior boundary of the velopharynx, referred to as the oropharyngeal isthmus, is bounded laterally by the anterior faucial pillars, superiorly by the undersurface of the soft palate, and inferiorly by the tongue. The superior boundary of the velopharynx is defined as a line extending from the hard palate, through the elevated velum, to the posterior pharyngeal wall. When at rest, the soft palate "hangs" down from its attachment to the hard palate into the pharynx.

Musculature

Levator Veli Palatini. The levator veli palatini (Figure 2) arises from the base of the skull; specifically, from the apex of the petrous portion of the temporal bone in an area anterior and medial to the carotid canal. It courses along

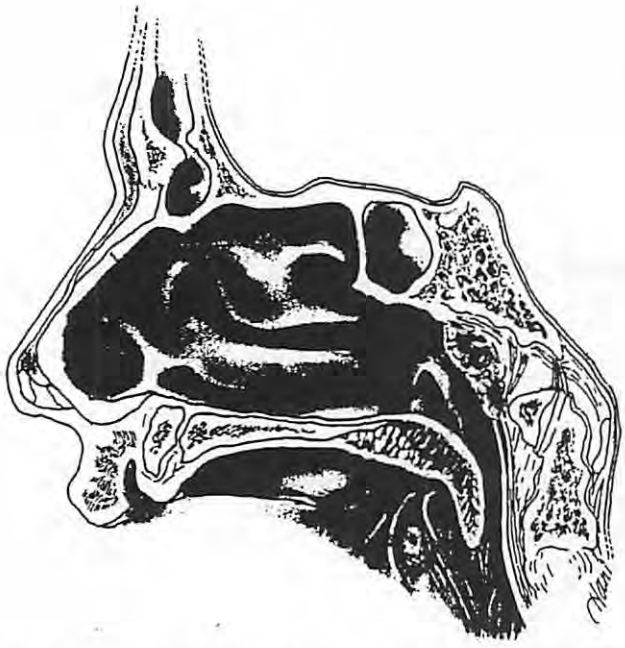


Figure 1. Lateral view of the nasal cavity and velopharynx (1 - soft palate, 2 - anterior faucial pillar, 3 - posterior faucial pillar, 4 - palatine tonsil).

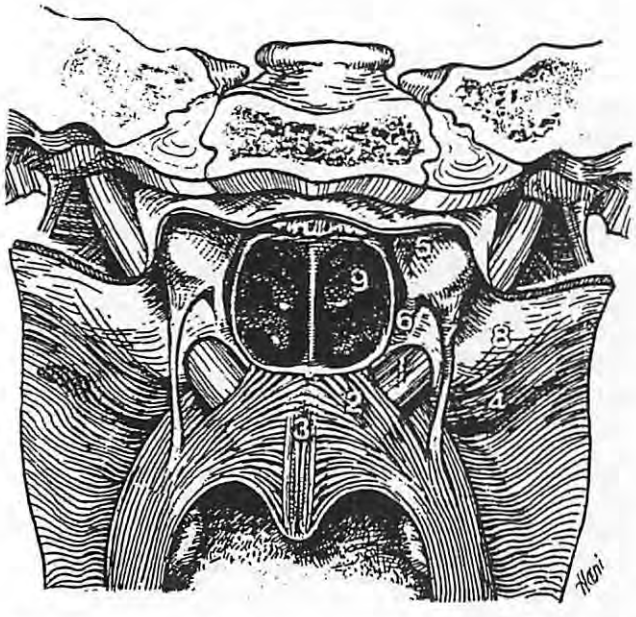


Figure 2. Muscles of the pharynx, viewed from behind. (1 - levator veli palatini, 2 - palatopharyngeus, 3 - musculus uvulae, 4 - palatopharyngeus (horizontal), 5 - torus tubarius, 6 - salpingopalatine fold, 7 - salpingopharyngeus, 8 - pharyngobasilar fascia, 9 - nasal conchae).

the under surface of the auditory (Eustachian) tube in an anterior, medial, inferior direction. Along with its fellow from the opposite side, it forms a sling and inserts into the middle approximately 40% of the velum (Boorman and Sommerlad, 1985).

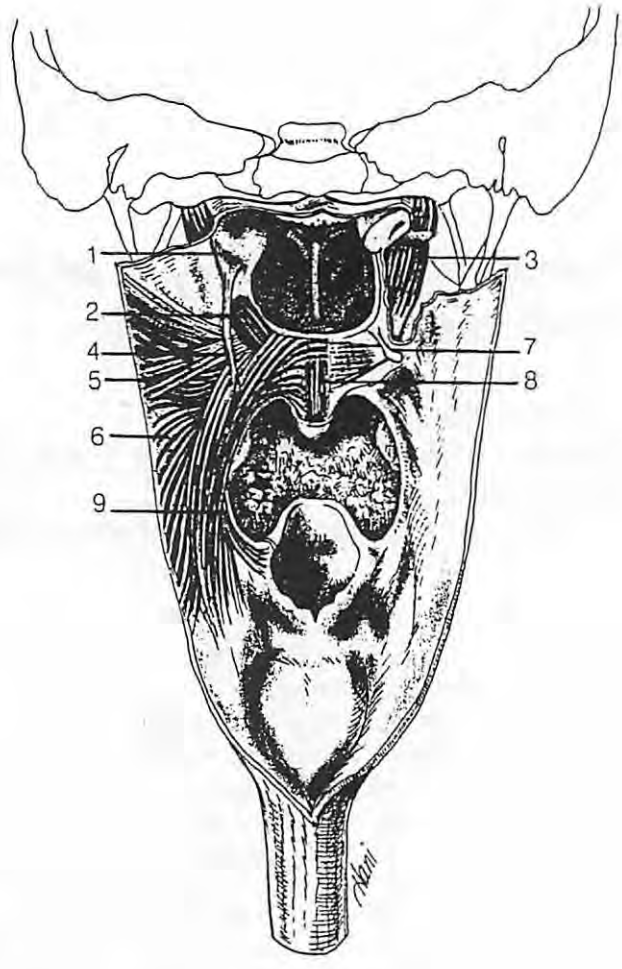


Figure 3. Posterior view of the velopharynx with mucosa of left side removed. (1 - tubal cartilage, 2 - levator veli palatini, 3 - tensor veli palatini, 4 - salpingopharyngeus, 5 - palatopharyngeus (horizontal), 6 - superior constrictor, 7 - pterygoid hamulus, 8 - musculus uvulae, 9 - palatopharyngeus).

Of all the velopharyngeal muscles, the levator veli palatini is perhaps the most consistent with regard to size, shape, and location. In cross-section, it is shaped as a slightly flattened cylinder with its larger diameter measuring almost 1 cm in the adult (Azzam and Kuehn, 1977; Kuehn and Azzam, 1978). In both lateral and frontal perspective, it enters the velum at an angle of approximately 45% relative to a vertical line. It is ideally suited to accomplish its namesake, that is to elevate the velum. In so doing, the velum is pulled along a superior and posterior path that may be linear or somewhat curved (Kent, Carney, and Severeid, 1974; Kuehn, 1976).

Tensor Veli Palatini: The tensor muscle (Figure 3,4) is located within the pterygoid fossa between the medial and lateral pterygoid plates. The muscle arises from a small depression, the scaphoid fossa, and adjacent area at the superior margin of the pterygoid fossa. Fibers also arise

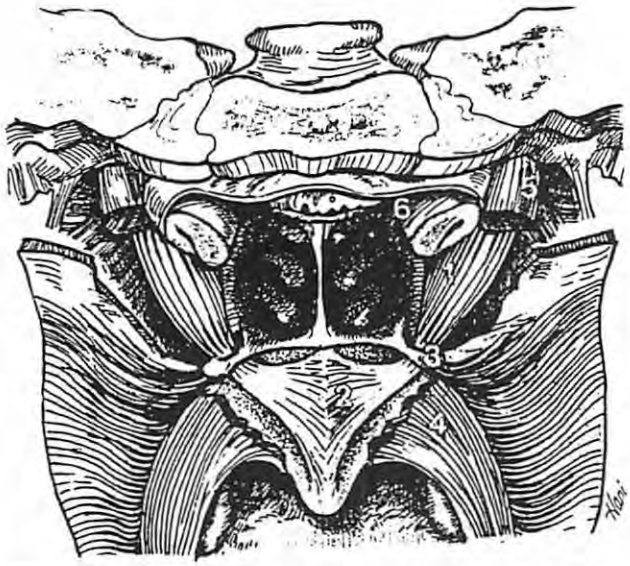


Figure 4. Posterior view of velopharyngeal musculature. Levator veli palatini and musculus uvulae have been removed. (1 - tensor veli palatini, 2 - tensor aponeurosis, 3 - pterygoid hamulus, 4 - palatopharyngeus , 5 - levator veli palatini (cut), 6 - tubal cartilage and pharyngobasilar fascia).

from portions of the lateral pterygoid plate and the spine of the sphenoid bone. Rood and Doyle (1978) described two separate groups of fibers associated with the tensor muscle, a medial group that they termed the "dilator tubae" and a lateral group that they designated as the tensor veli palatini proper. A few fibers of the lateral group are continuous with the tensor tympani muscle superiorly. The medial fibers attach directly to the auditory tube but the lateral more vertically inclined fibers do not. These fibers are continuous with a tendon that winds around the hamulus of the medial pterygoid plate and inserts into the superior and anterior region of the velum.

There is good agreement in the literature that the tensor muscle is the primary muscle of auditory tube dilation and probably interacts with the tensor tympani muscle owing to the anatomic connection. There is less agreement, however, with regard to the possible functional role of the tensor in relation to velopharyngeal activity. Any force distributed by the tensor muscle to the velum must be done so in relation to the hamulus as a focal point. Thus, it can be argued on purely anatomic grounds that tensor muscle force is poorly positioned, at least in the adult human, to either raise or lower the velum because the hamulus is close to the pivotal line of the velum and the leverage for either raising or lowering the velum from that locus would be unfavorable (Kuehn, 1990).

In addition to limitations imposed on the tensor muscle in relation to its locus of force delivery to the velum, the degree of free movement of the tensor tendon relative to the hamulus remains uncertain. Thus, the amount of "tensing" of the anterior portion of the velum that would be provided by the tensor tendon remains unclear. The presence of a bursa that would lend itself to a pulley action around the hamulus was not found in a study by Ross (1971) involving 32 gross dissections and 8 microscopic examinations. Moreover, Ettema (1993) did not find a membranous sac, indicative of a bursa, and did find that fibers of the tensor tendon inserted directly and firmly onto the bony tissue of the hamulus in a microscopic study of four human adult specimens.

Musculus Uvulae. The musculus uvulae (Figure 2,3) is the only intrinsic muscle of the velum. As such, it is completely contained within the velum and does not extend beyond its boundaries. Contrary to descriptions in many anatomy textbooks, the musculus uvulae, also called the uvular muscle, does not take its origin from the posterior border of the hard palate. Instead, it arises rather abruptly posterior to the border of the hard palate, in the region of the palatine aponeurosis (Azzam and Kuehn, 1977). Its structure can be likened to a flexible double-barrel shotgun at the midlength of the velum with its two bundles in close apposition near the nasal surface of the velum. It is in its most cohesive form in an area overlying the levator sling. Posteroinferiorly, the muscle bundles divide and subdivide becoming quite diffuse as a few fibers enter the uvula. The uvula proper, the free structure that hangs from the soft palate, contains very few muscle fibers (Ettema and Kuehn, 1994; Kuehn and Kahane, 1990). Therefore, the designation of this muscle as a "uvular" muscle is somewhat of a misnomer and rather misleading. A possible role for musculus uvulae contraction in velopharyngeal closure will be presented later. In addition to muscle activity, the bulk of the musculus uvulae would tend to fill in the region of contact between the velum and posterior pharyngeal wall for which its absence would leave a void. Indeed, it is often observed that in some individuals with repaired cleft palate, a central gap remains, which would be indicative of an absent or underdeveloped musculus uvulae.

Palatoglossus. The palatoglossus muscle (Figure 5,6) is located within the anterior faucial pillar and is positioned to either assist in velar lowering or tongue elevation. The designation of origin versus insertion is somewhat arbitrary in that the muscle attaches to two movable structures. Thus, depending on which structure is more stable at the moment, the velum or tongue, the palatoglossus muscle could exert a force in either direction. However, Kuehn and Azzam (1978) reported that the velar attachment varies across individuals. In some individuals the attachment is more anterior, near the posterior border of the hard palate, and in other individuals it is more posterior, near the uvula. Thus, depending on the location of velar attachment, the palatoglossus may provide a somewhat different level of force to the tongue versus soft palate.

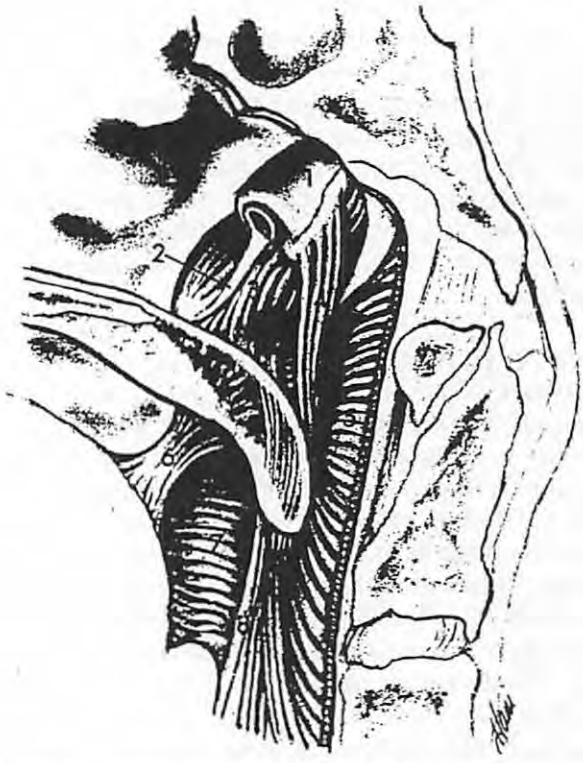


Figure 5. Lateral view of velopharyngeal musculature. (1 - tubal cartilage, 2 - salpingopharyngeal fascia, 3 - levator veli palatini, 4 - salpingopharyngeus, 5 - palatopharyngeus (horizontal), 6 - palatoglossus, 7 - superior constrictor, 8 - palatopharyngeus).

The palatoglossus muscle is somewhat diffuse within the anterior faucial pillar and thus appears to have limited force potential. Muscle spindles have been found within palatoglossus (Kuehn, Templeton, and Maynard, 1990) and stretching of the muscle could, perhaps, be involved in initiating the swallow reflex.

Palatopharyngeus. The potential functional significance of the palatopharyngeus muscle (Figure 3 - 6) is probably the least understood and most perplexing of all the velopharyngeal muscles. There is not even agreement on whether the palatopharyngeus is primarily a muscle of velopharyngeal port opening or closing. Classical anatomy textbook descriptions indicate that the muscle is contained within and helps to form the posterior faucial pillar. Its fibers course from the velum downward through the pillar and gradually terminate along the length of the pharynx with a few of its fibers reaching the thyroid cartilage of the larynx. That portion of the muscle that attaches to the thyroid cartilage has been designated as the "palatothyroideus" (Cassell and Elkadi, 1995; Cassell, Moon, and Elkadi, 1990). Such a description lends itself to the likelihood that the attachment to the velum is the more mobile, and thus the insertion, and that action of the muscle would be primarily in lowering the velum owing to the vertically descending orientation of the fibers. Moreover, the palatopharyngeus could assist in elevating the lower portion of the pharynx or the larynx via its connections to the thyroid cartilage.

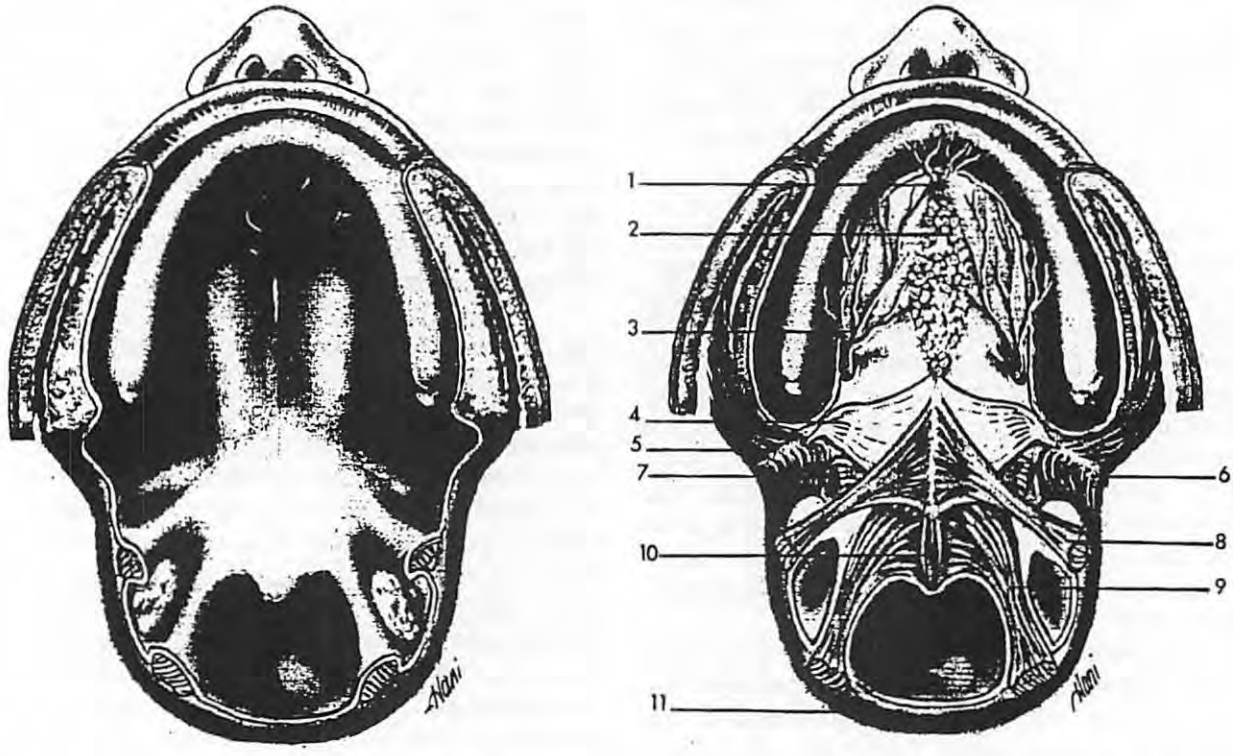


Figure 6. Ventral view of the velopharynx in horizontal section. (1 - incisive fossa, 2 - palatine mucous glands, 3 - greater palatine nerves and vessels, 4 - pterygoid hamulus and tensor veli palatini, 5 - buccinator, 6 - pterygomandibular raphe and superior constrictor, 7 - levator veli palatini, 8 - palatoglossus, 9 - palatopharyngeus , 10 - musculus uvulae, 11 - superior constrictor).

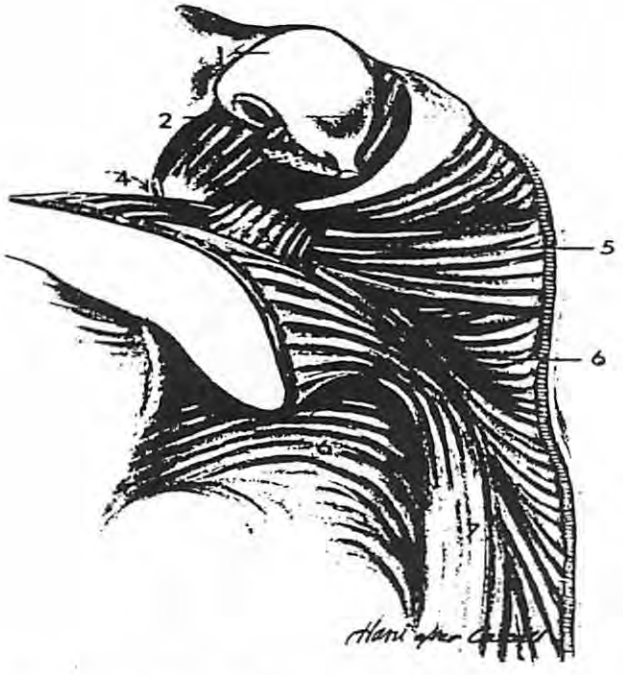


Figure 7. Oblique view of the velopharyngeal musculature (1 - Eustachian tube, 2 - tensor veli palatini, 3 - levator veli palatini, 4 - pterygoid hamulus, 5 - palatopharyngeus (horizontal), 6 - superior constrictor, 7 - palatopharyngeus).

Compared to the vertically oriented fibers, a very different structural and functional picture emerges in relation to the transversely oriented (horizontal) fibers described by Cassell, Moon, and Elkadi, 1990, and Dickson and Maue-Dickson, 1982. Figure 7 shows horizontal palatopharyngeus fibers that are quite high in the pharynx which could provide a sphincteric action thereby assisting in velopharyngeal closure. Cassell, Moon, and Elkadi suggest that confusion about the function of the "palatopharyngeus" muscle could be alleviated if the two muscle masses simply were given different names, that is, palatopharyngeus for the horizontal mass and palatothyroideus for the vertical mass. This argument would be more compelling if it can be demonstrated that the primary flow of the vertical fibers is into the thyroid cartilage with negligible attachment of fibers into the lateral walls of the pharynx, thus obviating the need for "pharyngeal" in the name of that muscle mass. With the exception of recent work authored by Cassell (Cassell et al., 1990; Cassell and Elkadi, 1995), vertical and horizontal fibers of the palatopharyngeus are not typically differentiated. To avoid confusion when reading other chapters in this text, we will differentiate the two fiber groups on the basis of orientation, and not with different names. Therefore, "palatopharyngeus" will be used to refer to the vertically oriented muscle fibers. Horizontally oriented palatopharyngeus muscle fibers will be referred to as "palatopharyngeus (horizontal)".

A different perspective on the potential functional significance of the palatopharyngeus muscle has been offered by Ettema and Kuehn (1994). They suggested that palatopharyngeus might function as a muscular hydrostat (Kier and Smith, 1985; Smith and Kier, 1989) by squeezing the contents of the posterior aspect of the velum so that it conforms to the configuration of the posterior pharyngeal wall thus providing a better seal during velopharyngeal closure. This will elaborated on later in the chapter.

Superior Pharyngeal Constrictor. The superior pharyngeal constrictor muscle (Figure 8) embraces the upper pharynx. It arises from several locations anteriorly and inserts into the pharyngeal raphe in the midline of the posterior pharyngeal wall. The raphe inserts onto the pharyngeal tubercle on the basilar part of the occipital bone. The various anterior origin sites, from the pterygoid hamulus to the posterior border of the tongue, have been described by Cassell and Elkadi (1995). These authors identified four sections of the muscle based on their respective origin from superior to inferior as follows: musculus pterygopharyngeus, musculus buccopharyngeus, musculus mylopharyngeus (mylohyoid line of the mandible), and musculus glossopharyngeus. There has been some confusion concerning the most superior fibers of the superior constrictor presumably due to the mingling of these fibers with the transverse fibers of the palatopharngeus muscle. Thus, some authors have described fibers of the superior constrictor muscle as entering the velum (Dickson, 1975).

Salpingopharyngeus. The salpingopharyngeus muscle (Figure 5) arises from the inferior border of the torus tubarius at the upper level of the pharynx and courses vertically along the lateral pharyngeal wall where its fibers gradually terminate. Some of the fibers may reach the larynx. The muscle is located within the salpingopharyngeal fold on each side of the pharynx along the lateral aspect of the pharyngeal walls. Dickson (1975) pointed out that the salpingopharyngeal fold may actually be void of muscle fibers and that the fold consists largely of glandular and connective tissue. Owing to its small size, if present, its lateral location, and vertical orientation, it is difficult to ascribe an important role to this muscle. Several authors have suggested that the salpingopharyngeal "muscle" simply consists of a few wayward fibers from other muscles in the region such as the palatopharyngeus.

The Velum

The velum, or soft palate, is clearly delineated superiorly and inferiorly, less so anteriorly, and indistinctly laterally. Thus, it is difficult to specify the lateral margins of the velum proper. This led Ettema and Kuehn (1994) to operationally define the lateral margins of the velum. Using that operational definition, they determined an average left-to-right dimension of 22.0 mm for the adult male cadaver

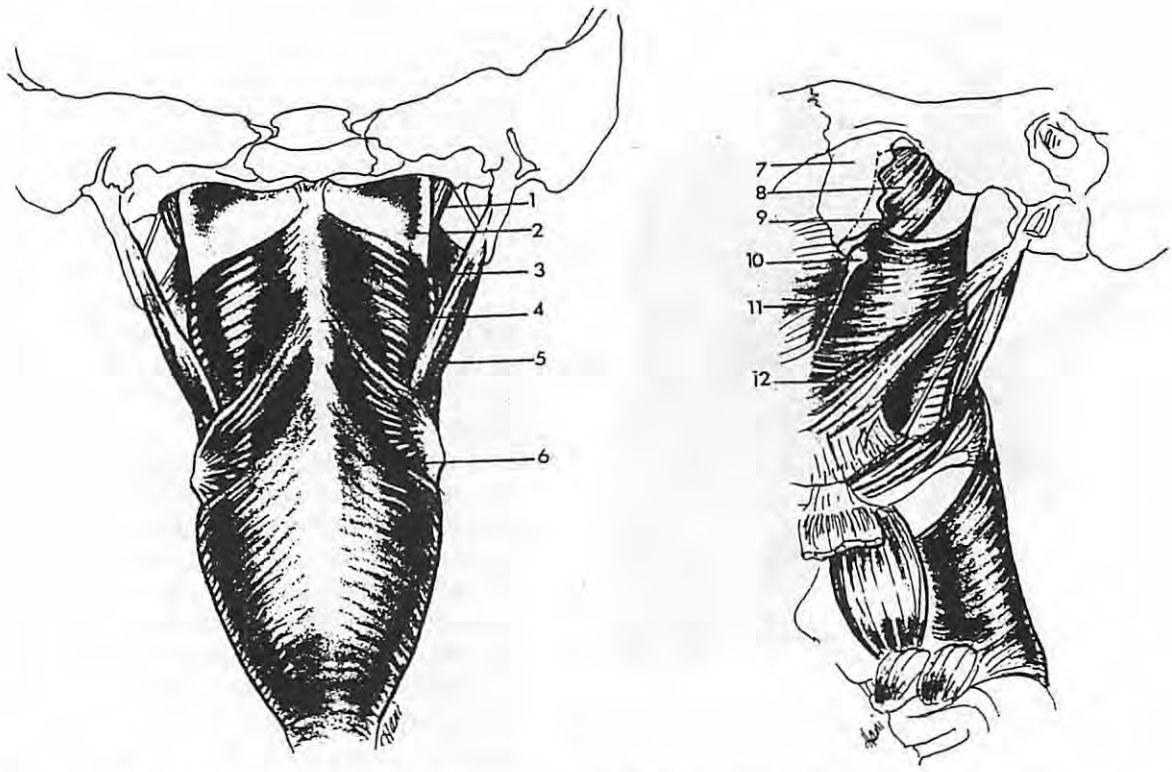


Figure 8. Posterior (left) and lateral (right) views of the pharyngeal constrictors (1 - levator veli palatini, 2 - pharyngobasilar fascia, 3 - superior constrictor, 4 - insertion of middle constrictor on pharyngeal raphe, 5 - styolpharyngeus, 6 - inferior constrictor, 7 - lateral pterygoid plate, 8 - tensor veli palatini, 9 - levator veli palatini, 10 - pterygoid hamulus, 11 - superior constrictor, 12 - styloglossus).

specimens and 19.3 mm for the adult female cadaver specimens. In a previous study (Kuehn and Kahane, 1990) using the same subjects, it was determined that the velar length of the males averaged 40.5 mm and the average length for the females was 36.8 mm. At its thickest location, the velum averaged 12.8 mm for the males and 12.5 mm for the females.

The major components of the velum are arranged in layers (Ettema and Kuehn, 1994; Kuehn and Kahane, 1990). The oral surface consists of stratified squamous epithelium. The nasal surface consists of pseudostratified ciliated columnar epithelium anteriorly and stratified squamous epithelium posteriorly where the velum contacts the posterior pharyngeal wall.

Anteriorly, the velum consists of the tensor tendon superiorly, very few (if any) muscle fibers just below the tensor tendon, and a large investment of glandular tissue inferiorly with some adipose tissue especially laterally. The palatine aponeurosis is located just below the nasal surface and extends about 1 cm posteriorly from its attachment to the posterior border of the hard palate (Cassell and Elkadi, 1995; Ettema and Kuehn, 1994). The aponeurosis is a mixture of fibrous connective tissue and the medial extension of the tensor veli palatini tendon.

The midportion of the velum is invested with muscle fibers that are both longitudinally oriented, those of the musculus uvulae, and also those oriented primarily side-to-side, those of the levator veli palatini with perhaps a blend of other fibers such as those from the palatopharyngeus. However, Kuehn and Kahane (1990) remarked that the palatopharyngeus muscle appeared to be quite prominent in the more lateral aspects of the velum. The longitudinal fibers are rather superficial and the side-to-side fibers are deep, within the central core of the velum. Ettema and Kuehn (1994) measured the quantity of muscle tissue within the velum and found the proportion of muscle tissue to be maximum at 40% of the length of the velum from anterior to posterior. This coincides with the region where the musculus uvulae is most cohesive and overlies the levator sling.

The quantity of muscle tissue tapers both anteriorly and posteriorly from the midportion of the velum so that there are very few muscle fibers either in the most anterior or the most posterior segments. Therefore, the uvula is largely void of muscle tissue. The uvula, as the most anterior portion of the velum, consists of a relatively large proportion of connective, glandular, and adipose tissues. However, unlike the anterior velum, the uvula consists of a much larger proportion of vascular tissue which could provide a warming function for the uvula thus avoiding a situation of excessive cooling of the pendulous structure.

Motor Innervation and Sensory Feedback Mechanisms

Motor Innervation. The classic description of innervation of the velopharyngeal musculature is that, with the exception of the tensor veli palatini which is innervated by the mandibular branch of the trigeminal nerve, all other muscles receive their innervation via the pharyngeal plexus (Kennedy and Kuehn, 1989). The exact contribution to the pharyngeal plexus by specific cranial nerves to the velopharnx has been debated in the literature (Dickson, 1975). Cassell and Elkadi (1995) concluded, on the basis of their own dissections and a review of the literature, that motor supply to the velopharynx is derived mainly from the pharyngeal branches of the glossopharyngeal and vagus nerves via the brainstem nuclei ambiguous and retrofacialis. The authors acknowledged that interconnecting branches of the facial and hypoglossal nerves might contribute to the motor innervation. The facial nerve has been implicated in innervating the velopharyngeal muscles in several studies that involved a clinical series (Ibuki, Tamaki, Matsuya, and Miyazaki, 1981), human fetal and adult dissection (Kacirkova, 1979), and animal nerve stimulation (Ibuki, Matsuya, Nishio, Hamamura, and Miyazaki, 1978; Nishio, Matsuya, Ibuki, and Miyazaki, 1976; Nishio, Matsuya, Machida, and Miyazaki, 1976). The hypoglossal nerve has been implicated in innervating the palatoglossus muscle in studies involving the uptake of horseradish peroxidase (Cassell and Elkadi, 1995).

Sensory Feedback Mechanisms. It has been reported that cutaneous sensory nerve endings diminish in quantity from the anterior to the posterior regions of the oropharyngeal cavity (Grossman and Hattis, 1964; Kanagasuntheram, Wong, and Chan, 1969). Sensory supply of the hard palate and velopharyngeal region has been described in detail in recent reports (Cassell and Elkadi, 1995; Liss, Kuehn, and Hinkle, 1994). Sensory innervation of the hard and soft palate is largely provided by the lesser and greater palatine nerves. These nerves feed into the maxillary division of the trigeminal nerve. The glossopharyngeal nerve supplies the faucial and pharyngeal regions. The facial and vagus nerves might also contribute to sensory innervation. Cassell and Elkadi (1995) point out, however, that even though the peripheral distribution of sensory fibers may travel along different cranial nerve routes, they all appear to terminate in the spinal nucleus of the trigeminal nerve. Presumably, this applies to sensations involving pain, temperature, and touch in that the mesencephalic nucleus of the trigeminal nerve has been implicated in receiving proprioceptive information from muscle spindles in the head.

Muscle spindles have been found in the levator veli palatini (Liss, 1990), the tensor veli palatini (Korner, 1941; Kuehn, Templeton, and Maynard, 1990; Winkler, 1964), the palatoglossus (Bossy and Vidic, 1967; Kuehn, Templeton, and Maynard, 1990; Liss,1990), but not in the palatopharyngeus (Bossy and Vidic, 1967; Kuehn, Templeton, and Maynard, 1990), the musculus uvulae (Kuehn, Templeton, and Maynard, 1990), the salpingopharyngeus (Kuehn, Templeton, and Maynard, 1990), or in the superior pharyngeal constrictor (Bossy and Vidic, 1967; Kuehn, Templeton, and Maynard, 1990). It should be noted that spindles found in the tensor and palatoglossus have been described as being similar to those located in other skeletal muscles of the body. However, typical spindles were not found in levator veli palatini in attempts by Kuehn, Templeton, and Maynard (1990) and Winkler (1964). A subsequent search by Liss (1990) using somewhat different procedures than those of previous investigators enabled identification of smaller spindles within the levator muscle. It is possible that other muscles of the velopharyngeal region for which spindles have not been found thus far might also contain smaller, atypical spindles compared to those found in other skeletal muscles.

The functional role of spindles within the velopharyngeal musculature is unknown and it is possible that spindles may have very different roles depending on the muscle within which they are located. Thus, because the tensor veli palatini is the primary muscle of auditory tube dilation, it is possible that spindles within that muscle may function in relation to air pressure balance and aeration of the middle ear cavity. As mentioned previously, it is possible that spindles within the palatoglossus may be involved in initiating the swallowing reflex. Spindles in the levator veli palatini may be involved in timing, signaling the onset and offset of velar raising and lowering gestures. Much work needs to be done in determining the exact role of these sensory mechanisms. Furthermore, virtually no information exists concerning other possible feedback mechanisms such as that provided by Golgi tendon organs.

Cleft Palate Anatomy

Most of the information involving cleft palate anatomy has been derived from fetal material, stillborns, or patients on the operating table. A listing of many of these articles appears in Kuehn and Dalston (1988) and Kriens (1990). Succinct descriptions of cleft palate anatomy appear in Dickson (1972) and Latham, Long, and Latham (1980). Dickson (1972) pointed out that there are two aspects that characterize anatomy of the cleft mechanism, lack of a palatine aponeurosis anteriorly and a general longitudinal flow of muscles, often called the cleft muscle of Veau, inserting onto the margin of the cleft hard palate. With regard to the latter, Latham, Long, and Latham (1980) identified these longitudinal fibers as belonging to the levator veli palatini, the palatopharyngeus, and the hemiuvular muscles. With regard to the levator muscle in particular, fibers course from one rigid attachment to another thereby leading potentially to a situation of isometric contraction.

Therefore, it is logical that the fibers should be freed from their attachments to the bony hard palate and moved into a position where they could more effectively raise the velum. Indeed, many surgeons attempt to accomplish this during primary palatoplasty.

Physiology of the Velopharyngeal Mechanism

In order for perceptually appropriate speech to be produced a number of structures in the vocal tract must move within the vocal tract space to "shape" the acoustic signal from the larynx. These movements result in spectral patterns that we perceive as different speech sounds. The vocal tract structures are referred to as articulators. These supraglottal articulators include the tongue, upper lip, lower lip, mandible, and the velopharyngeal mechanism. Closing and opening gestures of the velopharyngeal valve act as a variable resistor, coupling and decoupling the nasal and oropharyngeal cavities during speech in order to systematically alter the spectral characteristics of the acoustic signal. This discussion of velar physiology will be divided into two general sections: velopharyngeal movement patterns, and notions regarding velopharyngeal control.

Velopharyngeal Movement Patterns

Primary velopharyngeal movements may be divided into two basic groups: (1) movements of the velum towards (elevation) and away from (lowering) the posterior pharyngeal wall, and (2) mesial movements of the lateral pharyngeal walls. Anterior movements of the posterior pharyngeal wall, and bulging of the nasal surface of the velum may also play a role in velopharyngeal function for speech.

Velar Elevation. The velum is maintained in a lowered position when breathing at rest. While some velar elevation may be observed during nasal sound production (Moll and Shriner, 1967), especially during connected speech, the velopharyngeal port is still held open to allow for the movement of air between the oral and nasal cavities. The velum is moved posteriorly and superiorly towards the posterior pharyngeal wall during the production of oral speech sounds. The mid portion of the velum is typically elevated the highest, while contact with the posterior pharyngeal wall is typically accomplished by the third quadrant (Bzoch, Graber, and Aoba, 1959). Velar height and displacement vary during connected speech as a function of phonetic context. For example, high vowels tend to be associated with a more elevated velum than low vowels (Moll, 1962).

The muscle primarily responsible for velar elevation is the levator veli palatini. Numerous investigations have provided electromyographic data in support of this role for the levator muscle. Fritzell (1969, 1979) observed that contraction of the levator veli palatini was associated with movement of the mid portion of the velum upward and backward. Similar observations have been reported by Dickson and Dickson (1972), Lubker, Fritzell, and Lindqvist (1970), Bell-Berti (1976), and Kuehn (1979).

There is some evidence that variations in velar height observed during different speech sounds may be associated with variations in the magnitude of levator veli palatini muscle activity (Fritzell, 1969, 1979; Bell-Berti, 1976). However, correlations reported by Fritzell (1969) ranged from 0.46 to 0.94. Fritzell offered competing influences of other muscles as one possible explanation for the low correlations. Specifically, he suggested that superior constrictor, palatoglossus, and palatopharyngeus muscles may influence velar position in some subjects. To assess these possible influences, Kuehn, Folkins, and Cutting (1982) compared electromyographic activity from the levator veli palatini, palatoglossus, palatopharyngeus and superior constrictor to velar position as observed on lateral x-rays. Levator muscle activity alone did not appear directly related to velar position. Instead, systematic interactions were observed among levator veli palatini, palatoglossus, and palatopharyngeus muscle activity. Kuehn et al. suggested that a trading relationship might exist among these three muscles in positioning the velum. This notion will be expanded upon later in the chapter. Suffice it to say at this point that velopharyngeal closure may be more complicated than once thought.

Velar Lowering. Levator muscle activation results in movement of the velum towards the closed position. A logical assumption would be that a reduction or cessation of levator muscle activity would be associated with velar lowering. Numerous investigators have, in fact, observed this relationship (Fritzell, 1969, 1979; Lubker et al., 1970; Bell-Berti, 1976). The precise mechanism involved in velopharyngeal opening remains open to some question. Bell-Berti (1976) suggested that velar lowering was accomplished by the suppression of levator veli palatini activity and the natural tendency of tissue to return to its rest position. In support of this notion, Kuehn and Azzam (1978) did report the presence of elastic fibers in the anterior faucial pillars.

Kuehn and Azzam (1978) posited that restoration of the soft palate to its open configuration could be accomplished on the basis of three forces: muscle contraction, gravity, and tissue elasticity. There is evidence that palatoglossus and perhaps palatopharyngeus may be involved in velar lowering gestures (Lubker et al., 1970; Iglesias et al., 1980; Kuehn and Azzam, 1978). Based on anatomical data collected from dissection of 25 cadaver heads, Kuehn and Azzam (1978) suggested that the palatoglossus has a favorable mechanical ability to lower the velum. Fritzell (1979) and Lubker et al. (1970) observed palatoglossus activity during palate lowering. Fritzell suggested that, in addition to the possible influence of gravitational forces, Palatoglossus pulls the palate down for nasal

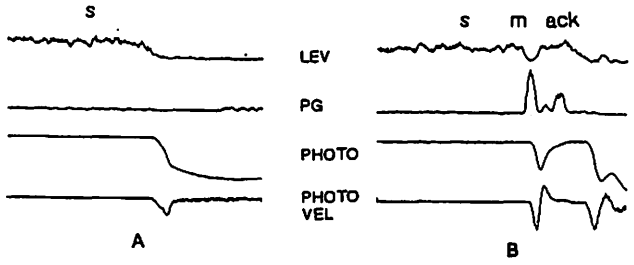


Figure 9. Rectified and smoothed levator veli palatini (LEV) and palatoglossus (PG) muscle activity, phototransducer output of velopharyngeal cross sectional area (PHOTO) and rate of change of velopharyngeal cross sectional area (PHOTO VEL) during production of [s] (left) and "smack" (right). Note the reciprocal activity of LEV and PG preceding velopharyngeal opening (PHOTO) for /m/.

sound production and at the end of phonation. Fritzell also noted a small amount of palatoglossus activity at the onset of phonation, perhaps related to pre-phonatory posturing of the velopharyngeal mechanism. Preliminary data collected from our laboratory appears to support the role of palatoglossus in velopharyngeal opening gestures, at least in some instances. Figure 9 shows rectified and smoothed electromyographic data from the levator veli palatini and palatoglossus muscles, along with a measure of velopharyngeal port area and the velocity of opening area changes during two speech tasks; sustained /s/ and the word "smack". In both instances, the onset of velopharyngeal opening is associated with a cessation of levator veli palatini activity. The velopharyngeal opening gesture associated with the cessation of /s/ production does not involve activation of palatoglossus in this example. However, the onset of velar lowering prior to the nasal consonant in "smack" is preceded by a large burst of activity in palatoglossus that occurs simultaneously with the cessation of levator activity. Of additional interest is the observation that increasing velopharyngeal port area associated with the nasal in "smack" and the burst of palatoglossus activity occurs with a higher velocity than opening at the cessation of the sustained /s/ where no palatoglossus activity was observed. While these data are preliminary in nature, the patterns observed point to the complexity of velar lowering. It may be the case that palatoglossus is activated in some instances (i.e. when quick velar lowering is required) and not in others.

Velar Bulging. Most descriptions of the physiology of velopharyngeal opening and closing gestures focus on the role of levator veli palatini, palatoglossus, and palatopharyngeus. Historically, contraction of the musculus uvulae was thought to shorten the uvula (Wells, 1971), an action that would appear to oppose velopharyngeal closure. More recently, Kuehn, Folkins, and Linville (1988) proposed two possible roles for the musculus uvulae during speech. As a stiffness modifying mechanism, the MU might control the velar-distorting forces of the levator veli palatini. Levator veli palatini contraction with a compliant (relaxed) MU might result in distortion of the velum by stretching the top layer upward instead of the entire velum. Contraction of the MU would provide internal stiffness to avoid such distortion. As a velar extensor, the MU might act as a two layered flexible beam. This notion is based on the fact that MU lies in the top half the curved velum and are attached anteriorly to the palatal aponeurosis (Figure 3). When contracted, a compressional force would be exerted along the nasal or top side of the velum. Since the oral or bottom side of the velum is relatively compliant, this compressional force would straighten the curved velum and extend it posteriorly.

Lateral Pharyngeal Walls. Mesial movements of the lateral pharyngeal walls contribute to velopharyngeal closure. However, it is not clear how this is accomplished. Current explanations for lateral pharyngeal wall movement focus on activity of the levator veli palatini and superior constrictor. Dickson and Dickson (1972) suggested that levator veli palatini muscle activity was responsible for both velar elevation and lateral pharyngeal wall movement. The later notion was based on their observation of inward deflection of the torus tubarius and knowledge of the relative position of levator veli palatini and the torus tubarius(see Figure 5). Others (Honjo et al., 1976; Niimi et al., 1982; Isshiki et al., 1985) have provided support for this theory. Proponents of this model argue against the influence of the superior constrictor based on observations that lateral pharyngeal wall movement occurs at a level in the pharynx above the region where superior constrictor fibers lie (Dickson and Dickson, 1972).

Alternatively, Skolnick (1969, 1970), Skolnick et al. (1973), and Shprintzen et al. (1975) have suggested that superior constrictor is responsible for lateral pharyngeal wall movement. Their argument was based on the observation that maximal mesial movements of the lateral pharyngeal walls occurred below the velar eminence associated with levator veli palatini muscle insertion into the velum. In agreement, Iglesias, Kuehn, and Morris (1980) found that the greatest mesial displacements along the lateral pharyngeal walls for all of their subjects and all of the speech sounds studied occurred at the level of and just below the plane of the hard palate. In addition, positive but low correlations were found between velar and lateral pharyngeal wall displacements for most of their normal subjects. These results do not support the hypothesis that the levator veli palatini muscle is solely involved in both velar and lateral pharyngeal wall movements. Such activity could be accounted for by contraction of the superior constrictor in conjunction, perhaps, with the horizontal fibers of the palatopharyngeus assisting in velopharyngeal closure gestures.

Posterior Pharyngeal Wall. Iglesias et al. (1980) found that the extent of posterior pharyngeal wall displacement in their group of normal subjects was very small. The largest mean displacement across subjects occurred for /z/,

1.3 mm, at the highest level of the pharynx. All other mean velaues across speech sounds and subjects were less than 1.0 mm. However, in some individual, a pronounced bulging forward of the posterior pharyngeal wall may be observed. In the 1800's, Passavant described such a bulging forward of the posterior pharyngeal wall in a cleft palate subject. This bulge, known as Passavant's ridge has also been observed in normal speakers (Calnan, 1957), although it occurs less frequently in normals. It has been suggested (Dickson and Dickson, 1972; McWilliams et al., 1990) that superior constrictor may contribute to the ridge. Cassell et al. (1991) suggested that the horizontal fibers of palatopharyngeus may contribute to the ridge. McWilliams et al. (1990) concluded on the basis of their extensive literature review that Passavants ridge does contribute to velopharyngeal closure in some individuals who demonstrate the ridge. They state further that the ridge does not develop in all individuals with velopharyngeal dysfunction. Casey and Emrich (1988) summarized the prevalence of Passavants ridge in their review of eight studies reporting on such prevalence. The average reported prevalence was 23% for subjects with cleft palate and 15% for subjects with intact palate.

Velopharyngeal Control

In 1985, McWilliams addressed the "state of the art" in velopharyngeal valving from the point of view of those interested in the diagnosis and treatment of individuals with velopharyngeal valving difficulties. She stated that despite advances "in knowledge about VP valving, relatively simplistic views of the mechanism are still commonly held. We often think of closure as being either achieved or not achieved" (p. 29). She stated further that motor aspects of velopharyngeal valving, including the strength of the velopharyngeal seal, present additional unanswered questions. Finally, McWilliams proposed that "we must focus our efforts on learning more about the VP valve as part of an integrated vocal system, and we must determine the extent to which structural and functional attributes of the whole system work in concert with VP valving to create speech variations." (p. 32).

The velopharyngeal valve is an articulator that must function according to principles of neuromotor programming and whose activity must be coordinated with the movements of other articulators in order to achieve acceptable speech. A number of studies have been conducted since McWilliams' 1985 comments that have addressed motor control of the velopharyngeal mechanism. These will be reviewed in this section.

Coordination of Muscle Activity within the Velopharyngeal Mechanism. Based on their anatomic position, the levator, palatoglossus and palatopharyngeus might act in somewhat of an agonist - antagonist role (Figure 10), although not in the same sense as agonist and antagonist

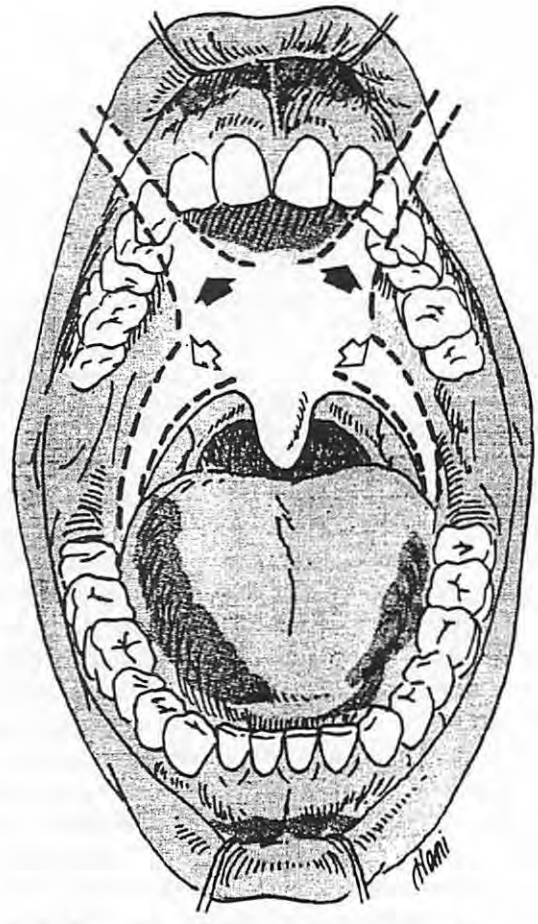


Figure 10. Representation of the relative orientation of the levator veli palatini and palatoglossus muscles, viewed from the front.

limb muscle working around a joint. Within this framework, upward movement of the velum via contraction of the levator veli palatini could be opposed by a downward pull via contraction of palatoglossus. One might argue that the interactions between levator veli palatini and palatoglossus depicted in Figure 9 support this conceptualization.

One might also consider the muscles of the velopharyngeal mechanism to constitute the elements of a coordinative structure. That is, position control of the velopharyngeal mechanism is flexible. Within this framework, varying combinations of muscle activation among the constituent muscles may lead to the same velopharyngeal position. As stated earlier, Kuehn et al. (1982) observed systematic interactions among levator veli palatini, palatoglossus, and palatopharyngeus muscle activity in relation to velar height. Kuehn et al. suggested that a trading relationship might exist among these three muscles in positioning the velum. Following up on this suggestion, Moon et al. (1994) assessed the relative contributions of the levator veli palatini, palatoglossus and palatopharyngeus relative to a range of positions of the velopharynx during vowel production. Four normal speakers were asked to phonate the

vowels /a/ and /i/ at either 50% or 75% velopharyngeal closure, using visual feedback of velopharyngeal aperture size as provided by a phototransducer (Dalston, 1982). Their results indicated that velopharyngeal positioning could best be predicted by considering activation levels of all three muscles as compared to any one of the muscles considered separately. This is not to suggest however, that earlier reports of the importance of the levator muscle to velar elevation are invalid. Based on anatomic position alone, velopharyngeal closure must be greatly influenced by levator veli palatini activity.

Considering the muscular anatomy of the velopharyngeal mechanism, in particular the apparent interrelated activity of some velopharyngeal muscles during velar positioning tasks, the velopharyngeal mechanism might be considered as a muscular hydrostat (Ettema and Kuehn, 1994; Moon and Jones, 1991). A muscular hydrostat is a structure of constant volume that does not undergo a volume change as its muscles contract (Kier and Smith, 1985). However, distribution of the surfaces of the hydrostat may be changed via muscle contraction (Ettema and Kuehn, 1994). An elephant's trunk is one example of a muscular hydrostat. Muscle contractions can, however, alter internal parameters such as stiffness. Kuehn et al.'s (1988) theory of musculus uvulae contraction acting to stiffen the velum in order to allow the levator veli palatini muscle to elevate the velum more efficiently is consistent with a conceptualization of the velopharyngeal mechanism as a muscular hydrostat. The muscular hydrostat notion is not necessarily compatible with the conceptualization of the levator and palatoglossus and palatopharyngeus acting as an agonist-antagonist pair. It is compatible with the notion that relative activation levels of different muscles can effect alterations in velopharyngeal positioning. In addition, different relative activation levels of the velopharyngeal muscles may be associated with the same velopharyngeal position and shape. Further, previous notions regarding the role of individual muscles based, in part, on anatomic positioning may be modified when considering hydrostatic principles. For example, Ettema and Kuehn (1994) suggest that contraction of the palatopharyngeus muscle could actually aid velopharyngeal closure "by squeezing the contents of the more posterior aspects of the velum and forcing the posterior nasal surface of the velum to conform to the concavity of the surrounding pharyngeal walls, much like forcing a water balloon to conform to the walls of a cylindrical container" (p. 311).

The degree and nature of interaction between the velopharyngeal muscles during velopharyngeal adjustments would be expected to vary from individual to individual for a number of reasons. Considering the normal population, there exists some variability on muscle orientation and insertion. For example, in those individuals for whom the palatoglossus attaches close to the hard palate, velar lowering might rely more heavily on other forces such as gravity and other muscle activity. For those individuals in whom the palatoglossus muscle is located closer to the uvula, the muscle may be more actively involved in velar lowering and less important as a synergist in tongue elevation. When considering the population with repaired palatal clefts, one must assume even more variability in muscle orientation, insertion point, mass, etc. If coordinated activity amongst the velopharyngeal muscles is a critical component of the control of velopharyngeal movement, it is possible that abnormalities in the muscle architecture and geometry within the repaired cleft soft palate may render its control less than optimal. However, there are no reports of systematic study of these notions.

Closure Force. One aspect of velopharyngeal function that has received little attention is the force of closure, or contact between the nasal surface of the velum and the posterior pharyngeal wall. Nusbaum et al. (1935) studied the firmness of velar-pharyngeal occlusion during production of various vowels. This was accomplished by applying air pressure to the subject's nasal cavity during vowel production until the seal between the velum and the posterior pharyngeal wall was broken, allowing the impounded air to be vented to the oral cavity. Closure pressure ranged from about 7 cm H_2O for low vowels to 23 cm H_2O for high vowels. Goto (1977) used a tube-like pressure sensing bulb placed in the velopharyngeal port to record closure force. Again, high vowels were associated with greater closure forces. Stop consonants were produced with greater closure force than vowels. Moon, Kuehn, and Huisman (1994) reported on the development and initial use of a new velopharyngeal closure force sensing bulb. In their initial report, bulb design characteristics were tested and reported. In addition, closure forces associated with vowel production were presented. As observed by previous investigators, high vowels were produced with greater closure forces than low vowels.

Yet to be tested is the relationship between velopharyngeal closure force and relative activation levels of the velopharyngeal muscles. The results of these studies are important to our understanding of the physiologic basis for observed variations in closure force. For example, are greater closure forces observed on high vowels a result of increased levator veli palatini muscle activity, or simply a decrease in inferiorly directed traction on the velum produced by the lowered tongue via the anterior faucial pillar.

Fatigue. Some muscles used for speech production appear to be able to produce forces that far exceed those necessary or typically used for speech. For example, lip muscle forces observed during speech represent only about 10-20 % of maximum attainable forces (Barlow and Abbs, 1983). Cook, Mead, and Orzalesi (1964) observed maximal pulmonary pressures between 146 and 237 cm H_2O in their study group, compared to the 7 - 10 cm H_2O typically associated with comfortable effort vowel prolongation. The

range of forces or pressures above that typically observed during speech may be viewed as one's reserve capacity. Kent et al. (1987) noted that a reduced reserve capacity may impair a speaker's "flexibility" and render speech a taxing activity. There are no reports of typical versus maximal range of velopharyngeal closure forces produced by either normal speakers or those with repaired palatal clefts in the literature. However, recent work by Kuehn and Moon (1994, 1995) has shed some light on ranges of muscle activation level observed in these two populations. For the 10 normal speakers studied (Kuehn and Moon, 1994), levator veli palatini muscle activity for speech tended to occur in the lower region of its operating range as determined in relation to a blowing task (see Figure 11a). That is, levator activity levels for speech in relation to the total range for blowing suggests a relatively low effort on the part of the levator muscle during speech. This is consistent with the notion that speech does not require a great deal of effort. This also suggests that normal speakers have a significant reserve capacity, and the flexibility to utilize that reserve should the circumstances dictate it.

A different levator activation pattern was observed in speakers with repaired palatal clefts (Kuehn and Moon, 1995). In general, the 5 speakers studied used levator muscle activation levels in the mid to upper region of its operating range as determined in relation to the blowing task (see Figure 11b). In contrast to the normal speakers, velar elevation might be considered to be a more effortful task for these speakers. In addition, these speakers would appear to have a greatly reduced reserve capacity.

These findings have important implications for our understanding of both normal and disordered velopharyngeal physiology. The speakers with palatal clefts studied by Kuehn and Moon (1995) were working at or near the physiologic limits of levator muscle activity when producing speech. They also demonstrated more variability in activation levels across repeated productions, perhaps as a result of having to function closer to the physiologic limits of the muscle. Functioning at or near the maximal activation level during a repetitive activity such as speech might be expected to increase the likelihood of muscle fatigue. As stated by Kuehn and Moon (1994), the velum may be less likely to reach its intended target more consistently and in a more timely fashion when the levator veli palatini muscle is overtaxed to the level approaching fatigue. Although not yet addressed in the laboratory, this may offer an explanation for the observation of inconsistent velopharyngeal closure in some speakers. That is, the increased variability in muscle activation level and potential for muscle fatigue associated with near maximal muscle activation levels during speech may result in closure being achieved on one attempt, but not on another.

For some patients with velopharyngeal dysfunction, attempts to strengthen the levator veli palatini might extend the operating range and increase the difference between the functional range for speech and the threshold of fatigue. Resistance exercises such as CPAP treatment (Kuehn, 1991) discussed elsewhere in this text represent attempts to accomplish this goal.

Interarticulator Coordination. Many previous studies of interarticulator coordination have focussed on a search for underlying rules of coordination that lead to an appropriate end result regardless of prevailing environmental conditions. For speech, the nervous system is thought to establish consistent relations among movements and the muscle activity of different structures to reduce the complexity of control while still producing appropriate speech output. A large body of data exists concerning interrelations among the lips and jaw during bilabial articulation (e.g. Folkins et al., 1988; Folkins and Abbs, 1975; Folkins and Linville, 1983; Folkins and Zimmermann 1981, 1982). Much less

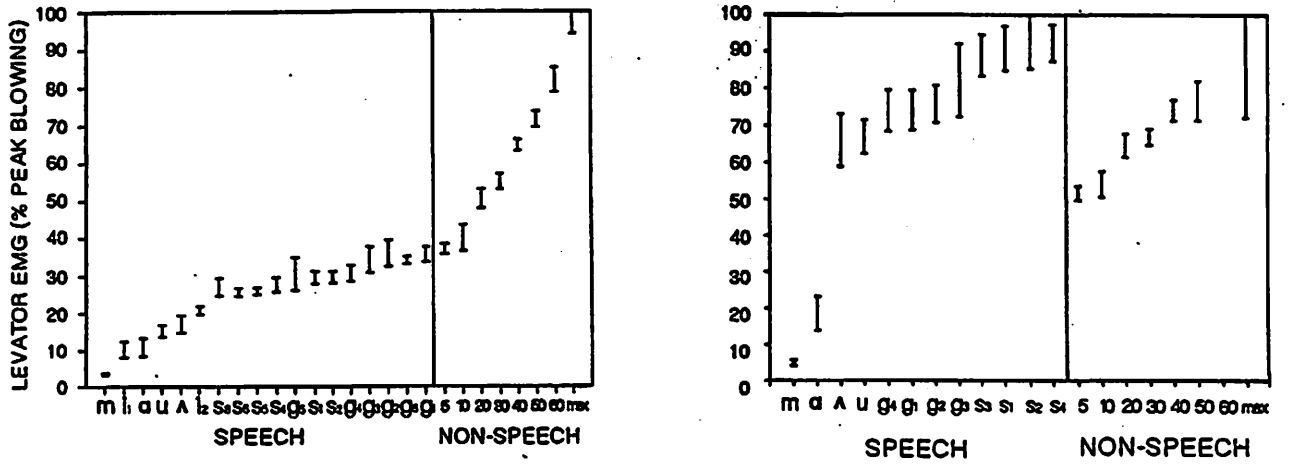


Figure 11. Levator veli palatini muscle activity (95% confidence intervals) for speech and nonspeech tasks. Subscripts under phonetic symbols represent different occurences on various phonetic contexts. Non speech tasks involved blowing through a tube to generate different levels of intraoral air pressure (left - representative normal speaker, right - speaker with repaired palatal cleft).

work has been conducted on coordination between the velum and other articulators.

Velar/Lingual Coordination. Kuehn (1976) observed an interaction between the velum and tongue in one normal subject such that velar velocity depended on the timing of velar elevation relative to tongue elevation. The later the elevation of the velum relative to tongue tip elevation for /s/, the greater the velocity of velar elevation.

Velar/Laryngeal Coordination. Based on observation of normal subjects, Kent and Moll (1969) reported that during VNCV syllables, velar elevating gestures tended to begin earlier, or were executed more rapidly when the stop consonant was voiceless. These timing relations have obvious implications for the adequacy of perceived voice quality (i.e., resonance). Zimmermann et al. (1984) stated that "if resistance to airflow into the nasal cavity is critical to perception of nasality then the time at which decoupling of the oral and nasal cavities is achieved in relation to the activation of the sound source may be critical to the perception of nasality" (p. 298). They found that their cleft palate speaker with abnormal nasal resonance reached velopharyngeal closure later relative to voice onset time than did the speaker with normal nasal resonance. Warren, Dalston, and Mayo (1994) observed that the most important factor in perceived nasality for subjects with small velopharyngeal openings appeared to be the length of time the velopahryngeal port was open. The results of EMG studies suggest further that the velopharyngeal and laryngeal mechanisms might demonstrate functional coordination. Two of Seaver and Kuehn's (1980) subjects exhibited higher levator veli palatini muscle activities during syllables with voiceless consonants. Palatopharyngeus muscle activity parallelled levator veli palatini activity, although the onsets, offsets, and major peaks in activity often occurred later in time than did the corresponding levator veli palatini activity. Seaver and Kuehn (1980) speculated that this might represent some active adjustment of the larynx for voiced sounds in the utterance. Ushijimi and Hirose (1974), however, saw no differences in levator veli palatini activity as a function of voicing. Bell-Berti (1976) observed some phonetic context dependent relationships between levator veli palatini activity and voicing. Intersubject variation in palatopharyngeus activity during the production of voiced versus voiceless stops was also observed. Bell-Berti related these differences to a need to control pharyngeal cavity size in order to maintain transglottic pressure differentials necessary for voiced stop production.

The relationships between changes in vocal fundamental frequency or intensity and articulatory characteristics have been investigated in a number of studies, (e.g. Tucker, 1963). In addition, observations of articulatory changes accompanying variations in linguistic stress (e.g., Kuehn and Moll, 1976; Ostry et al., 1983) may partially reflect changes in these prosodic variables as well as be related to the durational changes which accompany stress variations. Previous results indicate that the vocal tract becomes more "open" (lower jaw and tongue positions). Although some observations suggest that the vocal tract also may become more open as fundamental frequency is increased, the findings have varied with the types of subjects used (trained or untrained voices) and the identity of the vowel sounds produced. In addition, changes in fundamental frequency and intensity often have been experimentally confounded with each other and with other dimensions with which they tend to co-vary. The only specific study of the effects of pitch and intensity on velar function was that of Tucker (1963) who found increased velar elevation on sustained vowels produced at higher intensity (effort) levels but no systematic changes with fundamental frequency. It has been suggested that articulatory changes with intensity (or stress) may simply reflect the laryngeal adjustments needed to change vocal intensity. Because of the structural relationships between the larynx and the articulators, these adjustments may represent separate actions that help to increase intensity, or may reflect a general increase in muscular activity level in the entire speech production system. It might be hypothesized that increasing either intensity or fundamental frequency of vowels would have no effect on velar activity, since vowels are produced with an open vocal tract. That is, alterations in frequency or intensity would not be expected to modulate the aerodynamic demands placed on the velopharyngeal mechanism to maintain an adequate seal. However, alterations in jaw position with changes in fundamental frequency and/or intensity might be expected to apply forces to the velum. If the velopharyngeal mechanism operates within a coordinative structure framework with the other articulators, the forces might induce alterations in velar muscle activity.

Compensatory Interactions. Impaired function of a given articulator may have a significant effect on the performance of one or more other articulators. Altered performance characteristics may represent an attempt by the speaker to compensate for impaired function somewhere else in the vocal tract. For example, Warren et al. (1988) stated that speakers with palatal dysfunction exert greater respiratory effort during speech. They proposed that cleft palate speakers may compensate for palatal dysfunction by increasing respiratory effort in order to elevate intraoral air pressures. Active compensation at the laryngeal level has also been reported. For example, glottal stop articulation has been hypothesized as a compensatory valving mechanism associated with velopharyngeal dysfunction. Zajac and Linville (1989) suggested that altered laryngeal valving may be reflected in increased cycle-to-cycle variations in voice fundamental frequency (jitter). Hamlet (1973) suggested that "glottal tightness" secondary to hypernasality may be associated with lower values of observed vocal fold open quotient. Altered performance characteristics might also reflect purely physical consequences of an interaction between the glottal source and the vocal tract filter. Following

the nonlinear source-filter theory of Flanagan (1958), coupling of the nasal cavity might acoustically load the glottal source and influence its output. While coupling of the oral and nasal cavities may have an adverse effect on vocal fold behavior (e.g. in cases of velopharyngeal dysfunction), it may also be the case that oral-nasal coupling facilitates laryngeal function.

Compensatory actions of the tongue in response to velopharyngeal dysfunction have also been reported. A comprehensive review of these studies is provided by McWilliams et al. (1990). Relative to normal lingual articulatory patterns, speakers with palatal clefts have been observed to elevate and retract the tongue. These patterns were thought to represent attempts to obturate the velopharyngeal defect (Shohara, 1942; Powers, 1962). Disrupted timing relationships between the velum and tongue have also been observed in patients with velopharyngeal dysfunction (Wada et al., 1970).

Summary

The velopharyngeal mechanism is a complex articulator. Its normal anatomy has been well documented. However, there remain some questions about the arrangement of muscle bundles that may be of some importance when considering the characteristics of velopharyngeal closure (e.g. palatopharyngeus versus palatothyroideus). Anatomy of the cleft palate has also been reasonably well documented.

Unfortunately, our understanding of the physiology of both normal and disordered velopharyngeal function lags behind our knowledge of its anatomy. While roles have been assigned to each of the constituent muscles, recent studies have demonstrated that the control of velopharyngeal gestures is more complicated than can be explained by observation of one muscle or another. Further, the velopharyngeal mechanism does not function in isolation. Its control is undoubtedly affected by actions of other articulators, especially the tongue. Although complex in nature, velar control during speech would appear to be a relatively low effort task. Recent studies investigating velopharyngeal function within a speech motor control framework have added to our understanding of the physiology of this valve. There remains much to be learned. This is especially true about the disordered velopharyngeal mechanism. Palatal clefting must be expected to result in significant alterations in both velopharyngeal anatomy and physiology. Yet, velopharyngeal biomechanics and kinematics have not been systematically studied in this population. Research efforts directed towards a comprehensive understanding of normal and disordered velopharyngeal physiology are a necessity if we are to improve in our attempts to optimize attempts to produce speech by those with repaired palatal clefts.

Acknowledgement

Figures 1,2,3,4,5,6,7,8,10 are taken from Cassell et al. (1991) with permission. Our thanks to Hani Elkadi for the original drawings, and to the authors for permssion to reproduce those figures here.

References

Azzam, N. A., and Kuehn, D. P. (1977). The morphology of musculus uvulae. **Cleft Palate Jour, 14**, 78-87.

Barlow,S., and Abbs, J. (1983). Force transducers for the evaluation of labial, lingual, and mandibular motor impairments. **Jour Speech Hear Res, 26**, 616-621.

Bell-Berti, F. (1976). An electromyographic study of velopharyngeal function in speech. **Jour Speech Hear Res, 19**, 225 - 240.

Boorman, J.G. and Sommerlad, B.C. (1985). Levator palati and palatal dimples: their anatomy, relationship and clinical significance. **British Jour Plastic Surgery, 38**, 326-332.

Bossy, J. and Vidic, B. (1967). Les muscles du pharynx chez l'homme. **Archives d'Anatomie, d'Histologie, d'Embryologie, (Strasbourg), 50**, 273-284.

Bzoch, K., Graber, T., and Aoba, T. (1959). A study of normal velopharyngeal valving for speech. **Cleft Palate Bull, 9**, 3 - 4.

Calnan, J. (1957). Modern views on Passavant's ridge. **British Jour Plastic Surgery, 10**, 89.

Casey, D., and Emrich, L. (1988). Passavant's ridge in patients with soft palatectomy. **Cleft Palate Jour, 25**, 72 - 77.

Cassell, M.D., Moon, J.B., and Elkadi, H. (1990). Anatomy and physiology of the velopharynx. In Bardach, J. and Morris, H.L. (eds.) **Multidisciplinary Management of Cleft Lip and Palate**, Saunders: Philadelphia.

Cassell, M.D. and Elkadi, H. (1995). Anatomy and physiology of the palate and velopharyngeal structures. In Shprintzen, R.J. and Bardach, J. (eds.). **Cleft Palate Speech Management: a Multidisciplinary Approach**, Mosby: St. Louis.

Cook, C., Mead, J., and Orzalesi, M. (1964). Static volume-pressure characteristics of the respiratory system during maximal efforts. **Jour Applied Physiol, 19**, 1016 - 1022.

Dalston, R. (1982). Photodetector assessment of velopharyngeal activity. **Cleft Palate Jour, 19**, 1-8.

Dickson, D.R. (1972). Normal and cleft palate anatomy. **Cleft Palate Jour, 9**, 280-293.

Dickson, D.R. (1975). Anatomy of the normal velopharyngeal mechanism. **Clinics in Plastic Surgery, 2**, 235-247.

Dickson, D.R. and Dickson, W.M. (1972). Velopharyngeal anatomy. **Jour Speech Hear Res, 15**, 372-381.

Dickson, D.R., Grant, J.C.B., Sicher, H., DuBrul, E.L., and Paltan, J. (1974). Status of research in cleft palate anatomy and physiology. **Cleft Palate Jour, 11**, 471-492.

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

Ettema, S.L. (1993). A quantitative histologic study of the pterygoid hamulus and its surrounding region. Master of Arts Thesis, University of Illinois at Urbana-Champaign.

Ettema, S.L. and Kuehn, D.P. (1994). A quantitative histologic study of the normal human adult soft palate. **Jour Speech Hear Res, 37,** 303-313.

Flanagan, J. (1958). Some properties of the glottal sound source. **Jour Speech Hear Res, 1,** 99 - 116.

Folkins, J.,Linville, R.,Garrett, J., and Brown, K. (1988). Interactions in the labial musculature during speech. **Jour Speech Hear Res, 31,** 253-264.

Folkins, J., & Abbs, J. (1975). Lip and jaw motor control during speech: responses to resistive loading of the jaw. **Jour Speech Hear Res, 18,** 207-220.

Folkins, J., & Linville, R. (1983). The effects of varying lower-lip displacement on upper lip movements: implications for the coordination of speech movements. **Jour Speech Hear Res, 26,** 209-217.

Folkins, J., & Zimmermann, G. (1981). Jaw-muscle activity during speech with the mandible fixed. **Jour Acoust Soc Amer, 69,** 1441-1445.

Folkins, J., & Zimmermann, G. (1982). Lip and jaw interactions during speech: responses to perturbation of lower-lip movement prior to bilabial closure. **Jour Acoust Soc Amer, 71,** 1225-1233.

Fritzell, B. (1969). The velopharyngeal muscles in speech. **Acta Otolaryngologica, Suppl. 250.**

Fritzell, B. (1979). Electromyography in the study of the velopharyngeal function - a review. **Folia Phoniatrica, 31,** 93 - 102.

Goto, T. (1977). Tightness in velopharyngeal closure and its regulatory mechanism. **Jour Osaka Univ Dental Soc, 22,**1-19.

Grossman, R.C. and Hattis, B.F. (1964). Oral mucosal sensory information and sensory experience. In Bosma, J.F. (Ed.), **First Symposium on Oral Sensation and Perception.** Springfield, IL:Thomas.

Hamlet, S. (1973). Vocal compensation: an ultrasound study of vocal vibration in normal and nasal vowels. **Cleft Palate Jour, 10,** 267 - 285.

Honjo, I. Harada, H., and Kumazasa, T. (1976). Role of the levator veli palatini muscle in movement of the lateral pharyngeal wall. **Arch Oto Rhinol Laryngol , 212,** 93 - 98.

Ibuki, K., Matsuya, T., Nishio, J., Hamamura, Y., and Miyazaki, T. (1978). The course of facial nerve innervation for the levator veli palatini muscle. **Cleft Palate Jour, 15,** 209-214.

Ibuki, K., Tamaki, H., Matsuya, T., and Miyazaki, T. (1981). Velopharyngeal closure in patients with facial paralysis: thefiberscopic examination of the velopharyngeal movements. **Cleft Palate Jour, 18,** 100-109.

Iglesias, A., Kuehn, D. P., & Morris, H. L. (1980). Simultaneous assessment of pharyngeal wall and velar displacement for selected speech sounds. **Jour Speech Hear Res, 23,** 429-46.

Isshiki, N., Harita, Y, and Kawano, M. (1985). What muscle is responsible for lateral pharyngeal wall movement? **Annals Plast Surg, 14,** 224 - 227.

Kacirkova, J. (1979). The muscles of the soft palate: innervation, morphology, syntopia. **Folia Morphologica, 27,** 162-166.

Kanagasuntheram, R., Wong, W.C., and Chan, H.L. (1969). Some observations on the innervation of the human nasopharynx. **Jour Anatomy, 104,** 361-376.

Kennedy, J. G., and Kuehn, D. P. (1989). Neuroanatomy of Speech. In D. P. Kuehn, M. L. Lemme, & J. M. Baumgartner (Eds.), **Neural Bases of Speech, Hearing, and Language.** Boston: College-Hill (Little, Brown and Company).

Kent, R., & Moll, K. (1969). Vocal-tract characteristics of the stop cognates. **Jour Acoust Soc Amer, 46,** 1549-1555.

Kent, R.D., Carney, P.J., and Severeid, L.R. (1974). Velar movement and timing: evaluation of a model for binary control. **Jour Speech Hear Res, 17,** 470-488.

Kent, R.D., Kent, J.F., & Rosenbek, J.C.(1987). Maximum performance tests of speech production. **Jour Speech Hear Dis, 52,** 367-387.

Kier, W.M. and Smith, K.K. (1985). Tongues, tentacles, and trunks: the biomechanics of movement in muscular-hydrostats. **Zoological Jour Linnean Society, 83,** 307-324.

Komer, F. (1941). Die musculi tensor und levator veli palatini. **Zeitschrift fur Anatomie und Entwicklungsgeschichte, 111,** 508-532.

Kriens, O. (1990). Anatomy of the cleft palate. In Bardach, J. and Morris, H.L. (eds.) **Multidisciplinary Management of Cleft Lip and Palate,** Saunders: Philadelphia.

Kuehn, D. P. (1976). A cineradiographic investigation of velar movement variables in two normals. **Cleft Palate Jour, 13,** 88-103.

Kuehn, D. P. (1979). Velopharyngeal anatomy and physiology. **Ear, Nose, & Throat Jour, 58,** 316-21.

Kuehn, D.P. (1990). Commentary on "Observations on a Role for the Tensor Veli Palatini Muscle in Intrinsic Palatal Function." **Cleft Palate Jour, 27,** 318-319.

Kuehn, D. (1991). New therapy technique for treating hypernasal speech using continuous positive airway pressure (CPAP). **Plast Reconst Surg, 88,** 959-966.

Kuehn, D., & Moll, K. (1976). A cineradiographic study of VC and CV articulatory velocities. **Jour Phonetics, 4,** 303-320.

Kuehn, D. P., and Azzam, N. A. (1978). Anatomical characteristics of palatoglossus and the anterior faucial pillar. **Cleft Palate Jour, 15,** 349-59.

Kuehn, D., Folkins, J., and Cutting, C. (1982). Relationships between muscle activity and velar position. **Cleft Palate Jour, 19,** 25 - 35.

Kuehn, D. P., and Dalston, R. M. (1988). Cleft palate and studies related to velopharyngeal function. In H. Winitz (Ed.), **Human Communication and its Disorders: A review 1988,** Vol. 2. Norwood, NJ: Ablex Publishing Corporation.

Kuehn, D. P., Folkins, J. W., and Linville, R. N. (1988). An electromyographic study of the musculus uvulae. **Cleft Palate Jour, 25,** 348-55.

Kuehn, D. P. and Kahane, J. C. (1990). Histologic study of the normal human adult soft palate. **Cleft Palate Jour, 27,** 26-34.

Kuehn, D.P., Templeton, P.J., and Maynard, J.A. (1990). Muscle spindles in the velopharyngeal musculature of humans. **Jour Speech Hear Res, 33**, 488-493.

Kuehn, D., & Moon, J. (1994) Levator veli palatini muscle activity in relation to intraoral air pressure variation. **Jour Speech Hear Res, 37.**, 1260-1270.

Kuehn, D., & Moon, J. (1995). Levator veli palatini muscle activity in relation to intraoral air pressure variation in cleft palate subjects. **Cleft Palate-Cranio Jour**, in press.

Latham, R.A., Long, R.E., and Latham, E.A. (1980). Cleft palate velopharyngeal musculature in a five-month-old infant: a three-dimensional histological reconstruction. **Cleft Palate Jour, 17**, 1-16.

Liss, J.M. (1990). Muscle spindles in the human levator veli palatini and palatoglossus muscles. **Jour Speech Hear Res, 33**, 736-746.

Liss, J.M., Kuehn, D.K., & Hinkle, K.P. (1994). Direct training of velopharyngeal musculature. **Jour Medical Speech-Language Path, 2**, 243-251.

Lubker, J., Fritzell, B., and Lindqvist, J. (1970). Velopharyngeal function: an electromyographic study. **Quart Prog Stat Rep - Speech Trans Lab (Stockholm), 4**, 9 - 20.

McWilliams, B. (1985). Unresolved issues in velopharyngeal valving. **Cleft Palate Jour, 22**, 29-33.

McWilliams, B., Morris, H., and Shelton, R. (1990). **Cleft Palate Speech**, Toronto: Decker.

Moon, J., and Jones, D. (1991). Motor control of velopharyngeal structures during vowel production. **Cleft Palate-Craniofacial Jour, 28**, 267-273.

Moon, J., Smith, A., Folkins, J., Lemke, J., & Gartlan, M. (1994) Coordination of velopharyngeal muscle activity during positioning of the soft palate. **Cleft Palate-Craniofacial Jour, 31**, 45-55.

Moon, J., Kuehn, D., and Huisman, J. (1994). Measurement of velopharyngeal closure force during vowel production. **Cleft Palate-Cranio Jour, 31**, 356-363.

Moll, K. (1962). Velopharyngeal closure on vowels. **Jour Speech Hear Res, 5**, 30-37.

Moll, K., and Shriner, T. (1967). Preliminary investigation of a new concept of velar activity during speech. **Cleft Palate Jour, 2**, 58-69.

Niimi, S., Bell-Berti, F., and Harris, K. (1982). Dynamic aspects of velopharyngeal closure. **Folia Phoniatrica, 34**, 246 - 257.

Nishio, J., Matsuya, T., Ibuki, K., and Miyazaki, T. (1976). Roles of the facial, glossopharyngeal and vagus nerves in velopharyngeal movement. **Cleft Palate Jour, 13**, 201-214.

Nishio, J., Matsuya, T., Machida, J., and Miyazaki, T. (1976). The motor nerve supply of the velopharyngeal muscles. **Cleft Palate Jour, 13**, 20-30.

Nusbaum, E., Foly, L., & Wells, C. (1935). Experimental studies of the firmness of the velar-pharyngeal occlusion during the production of the English vowels. **Speech Monographs, 2**, 71-80.

Ostry, D., Keller, E., & Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: kinematics of tongue dorsum movements in speech. **Jour Exp Psych, Human Percep Perf, 9**, 622-636.

Powers, G. (1962). Cinefluorographic investigation of articulatory movements of selected individuals with cleft palates. **Jour Speech Hear Res., 5**, 59 - ??

Rood, S.R. and Doyle, W.J. (1978). Morphology of tensor veli palatini, tensor tympani, and dilator tubae muscles. **Annals Oto Rhinol Laryngol, 87**, 202-210.

Ross, M.A. (1971). Functional anatomy of the tensor palati. **Archives Otol, 93**, 1-8.

Seaver, E., & Kuehn, D. (1980). A cineradiographic and electromyographic investigatopn of velar positioning in non-nasal speech. **Cleft Palate Jour, 17**, 216-226.

Shohara, H. (1942). Speech rehabilitation in a case of post-operated cleft palate and malocclusion. **Jour Speech Hear Dis, 7**, 381-388.

Shprintzen, R., McCall, G., Skolnick, L., and Lemicone, R. (1975). Selective movement of the lateral aspects of the pharyngeal walls during velopharyngeal closure for speech, blowing, and whistling in normals. **Cleft Palate Jour, 12**, 51-58.

Skolnick, L. (1969). Video velopharyngography in patients with nasal speech, with emphasis on lateral pharyngeal wall motion in velopharyngeal closure. **Radiology, 93**, 747 - 755.

Skolnick, L. (1970). Videofluoroscopic examination of the velopharyngeal portal during phonation in lateral and base projections - a new technique for studying the mechanics of closure. **Cleft Palate Jour, 7**, 803 - 816.

Skolnick, L. McCall, G., and Barnes, M. (1973). The sphincteric mechanism of velopharyngeal closure. **Cleft Palate Jour, 10**, 286 - 305.

Smith, K.K. and Kier, W.M. (1989). Trunks, tongues, and tentacles: moving with the skeletons of muscle. **American Scientist, 77**, 29-35.

Tucker, L. (1963). Articulatory variations in normal speakers with changes in vocal pitch and effort. M.A. thesis, University of Iowa.

Ushijima, T., and Hirose, H. (1974). Electromyographic study of the velum during speech. **Jour Phonetics, 2**, 315 - 326.

Wada, T., Yasumoto, M., Ikeoka, N., Fujika, Y, & Yoshinaga, R. (1970). An approach for the cinefluorographic study of articulatory movements. **Cleft Palate Jour, 7**, 506 - 522.

Warren, D., Dalston, R., Morr, K., and Smith, L. (1988). Respiratory and temporal responses to velopharyngeal inadequacy. Paper presented at annual meeting of the American Speech-Lang-Hear Assoc, Boston.

Warren, D., Dalston, R., and Mayo, R. (1994). Hypernasality and velopharyngeal impairment. **Cleft Palate-Cranio Jour, 31**, 257 - 262.

Wells, C. (1971). **Cleft Palate and its Associated Speech Disorders.** New York: McGraw-Hill.

Winkler, G. (1964). L'equipment nerveux du muscle tenseur du voile du palais.**Archives d'Anatomie, d'Histologie, d'Embryologie,(Strasbourg), 47**, 311-316.

Zajac, D., and Linville, R. (1989). Voice perturbations of children with preceived nasality and hoarseness. **Cleft Palate Jour, 26**, 226 - 231.

Zimmermann, G., Karnell, M., and Rettaliata, P. (1984). Coordination of the velum with jaw, tongue and voicing in two cleft palate speakers with different nasality ratings. **Jour Phonetics, 12**, 297-306.

NCVS Status and Progress Report - 9
April 1996, 159-180

Behavioral Therapy for Speakers with Velopharyngeal Impairment

Lucrezia Tomes, Ph.D.
David Kuehn, Ph.D.
Department of Speech and Hearing Sciences, University of Illinois at Urbana-Champaign
Sally Peterson-Falzone, Ph.D.
Center for Craniofacial Anomalies, University of California--San Francisco

For more than 50 years, speech-language clinicians and clinical researchers have attempted to reduce hypernasality and audible nasal emission of air by using behavioral treatments [e.g., 5, 9, 46, 199]. Although the effectiveness of these treatments is still largely in doubt [100, 123, 133, 157, 158], many clinicians and researchers continue to use, develop, and advocate them.

To provide background for the current use and further study of behavioral treatments of hypernasality and nasal emission of air, this chapter will (a) describe the goals of these treatments, (b) review their controversial history, (c) summarize current opinions and evidence concerning these treatments, (d) explore why it is reasonable to continue searching for effective behavioral approaches to change velopharyngeal function even though years of work have produced little data to support these approaches, and (e) discuss considerations for future clinical research in this area.

Goals Of Behavioral Treatments

Behavioral treatments of hypernasality and nasal emission of air traditionally have had several goals [138, 158]:

(1) to change velopharyngeal muscles by increasing their strength, endurance, or mass,

(2) to change control of velopharyngeal activity by improving muscle coordination, rate of velar movement, or consistency of velopharyngeal closure, and

(3) to change respiratory, laryngeal, or oral articulatory behaviors to reduce speech nasalization without necessarily improving velopharyngeal function.

This chapter will focus primarily on treatments having the first two goals, that is, treatments intended to improve velopharyngeal function and thereby improve speech. However, treatments to reduce perceived nasalization by changing respiratory, laryngeal, or oral articulatory behaviors will also be mentioned briefly because they are often used in conjunction with treatments to improve velopharyngeal function.

Identifying the specific goals of proposed treatments to improve velopharyngeal function can be difficult. As Starr [157] has noted, most writers describing these treatments have not identified which aspect of function they desire to change. Even when they do identify the aspect of function to be changed, they usually do not measure this variable directly. Rather, they appear to assume that changes either in velopharyngeal muscles or in muscle control will be reflected in reductions in velopharyngeal orifice size, perceived nasalization, or both. This assumption is related to another complication in identifying the goals of various behavioral treatments: treatment goals may overlap. For example, techniques to strengthen velopharyngeal muscles may also enable a speaker to achieve velopharyngeal closure more consistently.

It is often presumed but not stated that reaching these goals should result in improvements in conversational speech. Other typically unstated assumptions are that the newly taught behaviors should be accomplished by the speaker without becoming fatigued and that these new behaviors should not have deleterious effects on laryngeal structure or function [22, 100, 161].

Despite these complexities in stating treatment goals, identifying the specific goals targeted by a treatment may be necessary to ensure the selection of appropriate treatments. Various specific treatment goals rest on different underlying assumptions regarding the nature of a speaker's velopharyngeal mechanism prior to therapy and its potential to change with therapy. For example, some

treatments require that the speaker achieve velopharyngeal closure during some speaking tasks prior to therapy while others do not. If treatment goals are not considered, the treatments provided may be ineffective because they do not match the needs and abilities of the people being treated.

In attempts to meet goals such as those listed above, many different behavioral treatments have been proposed. If the history of clinical experience with these treatments and research studying their effects is examined, the evolution of current behavioral treatments of hypernasality and nasal emission may be appreciated.

History Of Behavioral Treatments

Examination of textbooks and journal articles published over the last half century reveals three generally successive phases of opinions about the likely effectiveness of behavioral treatments to reduce hypernasality and audible nasal emission of air [123]. These phases can be summarized roughly as follows: Phase 1--1940s to 1960s: "Yes; we can improve velopharyngeal closure through physical exercises and speech therapy and we can alter other speaker behaviors so that the perception of nasalization is reduced," Phase 2--1960s to 1970s: "No; we have been naive in thinking that we can improve velopharyngeal closure by behavioral means, but perhaps therapy can develop speech that is perceived as less nasal," and Phase 3--1970s to 1990s: "Although definitive data are lacking, perhaps for some individuals, velopharyngeal closure may be improved without sending them back to the operating room."

Phase 1: Speech Therapy Works

From the 1940s through the 1950s and early 1960s, many clinicians believed that nonspeech exercises would increase the strength or voluntary control of the velopharyngeal mechanism for speech. During this period, clinicians commonly recommended activities such as blowing, whistling, sucking, swallowing, cheek puffing, and playing wind instruments [6, 18, 70, 90, 94, 106, 110, 172, 173, 185, 186, 188]. Most of these recommendations were made with the expectation that velopharyngeal closure obtained during these activities would transfer automatically to speech and that these methods would benefit all speakers with velopharyngeal impairment. Other writers claimed that velopharyngeal closure for speech sometimes improved as a side benefit of articulation therapy [e.g., 117, 190] or that feedback on the presence or absence of nasal airflow--shown by condensation on mirrors or by movement of feathers or other objects--could lead to improved velopharyngeal closure [e.g., 69, 106, 186]. To elicit improved velopharyngeal closure during speech, some [46, 47] advocated using greater effort by pushing the fists forcefully downward simultaneous with producing vowels or other short utterances.

As it became apparent that velopharyngeal function could not always be improved, clinicians began to suggest that using a greater oral opening would result in speech perceived to be less hypernasal [6, 96, 175]. McDonald and Koepp-Baker [96] stated that increasing oral opening would decrease oral impedance to sound and airflow. They reasoned that if oral impedance was reduced relative to nasal impedance, speech might be perceived as less nasal even though velopharyngeal closure movements remained relatively unchanged.

All of these recommendations during Phase 1 were based predominately on clinical impressions. Some of these impressions were questioned in Phase 2 when researchers began to study these treatments.

Phase 2: Speech Therapy May Not Improve Velopharyngeal Closure but May Reduce Perceived Nasalization Treatments to Improve Velopharyngeal Function

Skepticism about the effectiveness of these treatments. During the 1960s and 1970s, behavioral treatments to improve velopharyngeal closure fell out of favor primarily because research during that time did not support their use. In an often cited study using sagittal cinefluorography, Moll [102] showed that normal speaking individuals did not typically close the velopharyngeal port during sucking and cheek puffing. He also found that for most of his subjects velar movements used during blowing and gagging differed from those used during speech. Others [45, 62, 99, 119] also found differences between speech and nonspeech velopharyngeal movements in speakers with velopharyngeal impairments. These results led clinicians to question the validity of using nonspeech activities as the bases for improving velopharyngeal closure movements for speech.

Other research examined subjects' responses to treatment and failed to support the idea that nonspeech exercises or speech therapy procedures were associated with improvements in velopharyngeal function. Massengill, Quinn, Pickerell, and Levinson [94] investigated the effects of blowing, sucking, and swallowing exercises in conjunction with intensive articulation therapy on velopharyngeal gap during sustained vowels. Subjects with repaired clefts failed to show improvement with blowing and sucking exercises. However, Massengill and his colleagues reported positive results for subjects who performed swallowing exercises requiring them to practice prolonged swallows and to monitor each swallow by placing a finger on the larynx. Unfortunately, these researchers neglected to document homogeneity across subject groups performing the various types of exercises. Furthermore, in a study of the same types of exercises, Powers and Starr [126] reported poor results for subjects with repaired clefts who reportedly achieved closure during a blowing task prior to receiving the treatment. These subjects showed no significant change in

velopharyngeal gap or perceived nasality associated with the exercises. Change in velopharyngeal function associated with articulation therapy was studied by Shelton and his co-workers [142]. Their subjects, who also had repaired palatal clefts, failed to show smaller velopharyngeal gap sizes or increased posterior pharyngeal wall movements. All of these studies were limited in that only a lateral view of velopharyngeal movements was used to measure change.

Because clinical impressions that therapy improved velopharyngeal function were not supported by research, many concluded that velopharyngeal impairment could not be improved by speech therapy methods [e.g., 126, 132, 143, 156]. Emphasis was placed on treatments involving surgery or long-term use of prostheses because of the realization that many speakers had received prolonged speech therapy that failed to change velopharyngeal function or reduce hypernasality or audible nasal emission of air.

Continued interest in treatments to improve velopharyngeal function. Notwithstanding the negative findings for treatments involving nonspeech exercises and articulation therapy and the resulting skepticism, an undercurrent of interest in behavioral treatments remained and new procedures were proposed and studied during the 1960s and 1970s. Clinicians and researchers devoted considerable study to the effects of approaches intended to stimulate increased velopharyngeal movement.

The most commonly studied method of this type was the use of a temporary speech prosthesis. This prosthesis consisted of a segment constructed to follow the contour of the velum, possibly elevating it slightly, and an obturator bulb added to the posterior, pharyngeal portion of the prosthesis to prevent nasal airflow during speech. Blakeley [9] explained that a primary purpose of the prosthesis was to require that speech be produced with appropriate oral air pressures and orally directed airflow, thereby facilitating acquisition of plosive and fricative consonants. He and others reported that use of obturators was associated with not only the expected changes in speech as a result of the obturation of the velopharyngeal port, but also longitudinal changes in how the velopharyneal muscles behaved. Many authors reported increased pharyngeal wall motion in speakers with velopharyngeal impairment who had been fitted with obturators, in some cases allowing a reduction in obturator size or complete elimination of the appliance [7, 8, 9, 11, 21, 24, 43, 57, 58, 81, 95, 101, 130, 176, 183, 184, 197]. Cole [22] noted that some authors [e.g., 43, 130] believed that increased muscle activity resulted from resistance of the obturators to the movement of the palatal and pharyngeal muscles, while others [7, 21] felt that the presence of the obturator in the nasopharynx somehow served to "stimulate" the musculature.

Although the reports of successful obturator reduction and elimination of prostheses were widespread, few systematically gathered data were reported. Weiss [183] reported that 23 of a series of 125 patients experienced reduction and complete elimination of the prosthesis with little or no remaining nasality [184] but he did not present specific data describing his patients' articulation, resonance, or velopharyngeal closure behaviors. Blakeley [9] reported his clinical observation that usually the lateral aspects of the appliance were reduced more than the posterior aspect but he also presented no supporting data. In a study using sagittal cinefluorography, Shelton and his co-workers found greater anterior movement of the posterior pharyngeal wall as bulb size was reduced for 2 of 19 subjects [146]. In another study, they found that 1 of 3 subjects developed greater posterior pharyngeal wall movement in association with a program of therapy that involved interchange of obturators of various sizes, but the change in movement was not judged to be clinically significant [147].

Reports that the presence of obturators was associated with improved velopharyngeal function prompted other clinicians to explore a variety of methods for stimulating increased velopharyngeal movement. Lubit and Larsen [88, 89] devised an inflatable "palatal exerciser" which elevated the velum superiorly and posteriorly and pressed against the posterior pharyngeal wall. Although they claimed that patients benefitted from the exerciser, if the device had actually "exercised" the palate, it would have provided resistance against muscles lowering the velum and thus would have exercised these muscles rather than the muscles elevating the velum. Resistance to muscles elevating the velum would have been reduced by the device, although the opposite would be needed for resistance exercise of those muscles. Massengill, Quinn, and Pickerell [93] advocated use of "palatal stimulators" which resembled palatal lifts and reported data from five patients showing that velopharyngeal gaps were either reduced or eliminated after wearing the "stimulator" for one year.

Yules and Chase [200] reported that reduced hypernasality and improved velopharyngeal closure for speech were associated with a treatment program incorporating electrical stimulation of the velum, feedback of nasal acoustic energy, and a home program of tactile stimulation of the velopharyngeal area. However, a follow-up study in their laboratory failed to replicate their earlier positive results [182]. In addition, Peterson [120] later questioned their methodology based on her finding that electrical stimulation did not consistently elicit palatal elevation in either normal subjects or those with velopharyngeal impairment.

Shelton [138] was among the first to recognize that lack of a conscious sense of palatal position limited the success of behavioral treatments to change velopharyngeal movements. In a series of studies, he and his co-workers attempted to increase awareness of palatal position and to train voluntary movement of the velum and pharyngeal walls. Normal speaking children and adults learned to

voluntarily elevate the velum [145]. However, subjects' reports of whether or not they elevated their palates did not consistently agree with observations of their velar elevation made by judges who monitored the subjects' velar activity. In addition, while some subjects reported sensing their palatal movements, others said they were unable to do so. Furthermore, when two normal adult subjects were tested for their ability to elevate their velums to various degrees, they were only able to elevate to two different height categories [144]. Shelton and his colleagues interpreted these findings as evidence that sensation of palate position and movement is imprecise and that the velar elevation training provided in these studies would not likely facilitate improved velopharyngeal closure during speech. In a related study, Tash et al. [161] found that normal speaking children and children with velopharyngeal impairment but no history of palatal clefting could be trained to produce voluntary pharyngeal wall movements during production of /a/, but the movements did not generalize to spontaneous speech.

In 1979, as Phase 2 shifted into Phase 3, Cole [22] reviewed the literature on behavioral treatments of velopharyngeal function. He categorized these treatments into three groups of "muscle training" procedures: (a) indirect muscle training which included articulation therapy and yawning, swallowing, and gagging exercises, (b) semidirect muscle training which included blowing and sucking exercises, and (c) direct muscle training which used devices to "touch, stroke, manipulate, stimulate, or apply resistance to the palatal and pharyngeal muscles" (p. 335) to bring involuntary velopharyngeal movements under voluntary control. According to Cole, indirect methods were unlikely to effect improvement in velopharyngeal function. However, he maintained that semidirect or direct muscle training procedures should be tried if velopharyngeal closure was inconsistent or the residual gap was "relatively small" (p. 330). He indicated that semidirect training activities might be more effective if they were "performed against considerable resistance" (p. 334)--for example, tasks might involve blowing a baseball rather than cotton. In addition, he described direct muscle training procedures for teaching individuals to gradually bring under their voluntary control velopharyngeal movements present during the gag reflex. To make patients aware of sensations associated with contraction of the velopharyngeal muscles, Cole used long-handled spoons or cotton applicators to touch and stroke the velopharyngeal area to elicit small muscle movements, but stopping just short of eliciting the gag reflex. As patients became more tolerant of the stimulating devices, he also used the devices to provide resistance against pharyngeal wall movements. Cole recommended incorporating performance feedback into both semidirect and direct muscle training tasks, and he emphasized his belief that this was the most important component of muscle training programs. Cole provided no empirical evidence supporting the effectiveness of his recommended treatments. He also did not consider earlier work indicating that the velopharyngeal mechanism is not likely to be under the same neural control for gag as for speech [62] and does not show the same pattern of behavior in these two activities [153].

Three years after Cole's review, Ruscello [132] reviewed "palatal training procedures" under the same categories proposed by Cole. Ruscello reached the general conclusion that "present clinical treatments directed to the palatal mechanism do not have empirical support and consequently cannot be successful on a routine basis" (p. 191).

Treatments to Reduce Perceived Nasalization by Changing Respiratory, Laryngeal, or Oral Articulatory Behaviors

In agreement with McDonald and Koepp-Baker [96] and others during Phase 1, clinicians during Phase 2 also wrote that hypernasality could be reduced by increasing the range of oral motion [e.g., 104, 107, 143, 187]. An acoustic model described by Lindblom and Sundberg [83] revealed that speakers who were least hypernasal tended to use the greatest oral openings. Fletcher [41] obtained comparable results for speakers with repaired palatal clefts. Warren [178] indicated that increasing the size of the oral opening resulted in less nasal airflow even if the size of the velopharyngeal orifice remained unchanged.

Other changes in articulatory behavior recommended to reduce nasalization included use of light articulatory contacts [e.g., 111, 143, 187, 191] and slow speaking rate [e.g., 52]. Clinicians and researchers also noted that speakers with good oral articulation skills tended to be perceived as less nasal than those with poor oral articulation and, therefore, suggested that improving oral articulation skills alone might lead to reduced hypernasality [e.g., 36, 149, 171].

Clinicians during Phase 2 suggested changing loudness and modal pitch to reduce perceived hypernasality [e.g., 39, 201]. However, there was disagreement regarding whether increased or decreased loudness and pitch would contribute to reducing hypernasality [e.g., 26, 39, 61].

During Phase 2, Fletcher [40] introduced TONAR, a device designed to measure an acoustic correlate of hypernasality. TONAR consisted of a nasal microphone and an oral microphone isolated from each other by a lead enclosure to limit sound at one microphone from reaching the other microphone. The ratio of the acoustic energy at the nares to the sum of the acoustic energy at the nares and the oral opening was termed "nasalance" [42]. Fletcher [40] described his use of TONAR to provide feedback during therapy to reduce hypernasality. Fletcher instructed his subjects to watch the nasalance trace as they spoke and to keep it under certain threshold values. He presented case studies for two individuals, one with velopharyngeal im-

pairment secondary to a repaired cleft palate and the other with hypernasality following tonsillectomy and adenoidectomy. Although nasalance decreased during their treatment, Fletcher provided no evidence that velopharyngeal function improved during the course of their treatment.

Phase 3: Certain Kinds of Speech Therapy May Work for Some Individuals, but Supporting Data are Lacking

In the mid-1970s and early 1980s, the number of reports advocating use of behavioral treatments of velopharyngeal impairment began to increase. The treatments in this third phase are a mixture of old and new. Some of the treatments described and advocated during Phase 2 continue to be used in Phase 3 and ongoing interest in them was probably motivated by undocumented claims of clinical success during Phase 2. Most of the other treatments in Phase 3 are new approaches that were made possible by two developments in instrumentation for viewing the velopharyngeal system: (a) replacement of cinefluorography, which had been typically carried out only in the lateral view, with multiview videofluoroscopy, and (b) use of videoendoscopy--first oral, then nasal--for viewing movements of the velopharyngeal system without the dangers of radiation. These and other new instruments allowed clinicians and researchers to observe patients' pretreatment velopharyngeal function in greater detail so that treatments could be more closely matched to patients' needs and abilities, to provide feedback on velopharyngeal function during treatment, and to quantify the results of treatment.

Phase 3 opinions about the probable effectiveness of the treatments represent a refinement of older points of view. The number and variety of behavioral treatments described during Phase 3 reflect hope that these approaches may be useful in treating velopharyngeal impairment. However, recent reviewers of the literature in this area are hesitant to endorse any specific behavioral treatment because the supporting data are lacking. While they acknowledge that the available reports of treatment effectiveness indicate that some individuals with velopharyngeal impairment are helped with the procedures, they emphasize that the procedures are not likely to help all those with velopharyngeal impairment and clinicians and researchers continue to be unable to identify those who can be helped [e.g., 54, 100, 133, 158].

Treatments to Improve Velopharyngeal Function

Use of temporary prostheses. During Phase 3, temporary prostheses continue to be used and advocated to improve velopharyngeal function [66, 98, 150]. Israel, Cook, and Blakeley [66] reported fitting temporary prostheses with pharyngeal obturator bulbs in children as young as 2 1/2 years of age and indicated that their patients wear the obturators for 3 to 5 years. During this time, the child receives articulation therapy. After articulation skills are age appropriate, or nearly so, obturator reduction is attempted. In addition to reducing the obturator, they may also recommend complete removal of the appliance for a trial period of time, such as a week or month, if a child's parents report that the child's speech is perceptually the same with or without the appliance. Israel and his colleagues stated that in their experience, for 25 to 45% of the patients, the obturator could eventually be completely removed without the need for subsequent surgery. McGrath and Anderson [98] reported that various centers estimate that 3 to 60% of their patients need no further physical management after removal of the temporary prosthesis. Israel et al. expressed their opinion that in the presence of the obturator "the sphincter muscles may learn to constrict further during speaking than they normally would have without the prosthesis" (p. 207). They emphasized that they do not believe that the muscles hypertrophy, but that the muscles "'learn' to function more voluntarily" (p. 207).

According to Israel and his associates, even when surgery is ultimately needed to replace an obturator, the experience with the temporary obturator is of benefit to the patient. While wearing the obturator, the patient learns normal articulation skills. As a result, the patient is not likely to direct speech sounds through the nose after surgery and, therefore, the surgeons use a surgical procedure that is "less obstructing to the nasal airway" (p. 208).

Wolfaardt and his co-workers [196] studied the effects of a procedure that involved systematic reduction of the amount of time a palatal lift appliance was worn by 32 patients with velopharyngeal impairment. In addition to wearing a palatal lift, their patients received: (a) listening training to develop their awareness of normal and deviant nasalization, (b) nasalance feedback, (c) nasal airflow feedback, (d) articulation therapy to correct consonant errors, including compensatory errors, and (e) therapy to increase jaw range of motion as a means of maximizing oral-nasal resonance balance. Of the 32 patients studied, 14 were able to discontinue using the palatal lift while maintaining speech with appropriate oral-nasal resonance balance. Unfortunately, Wolfaardt et al. did not report the ages of these patients or the etiologies of their velopharyngeal impairment.

Palatal training appliance. Another oral prosthesis said to improve velopharyngeal function is the "palatal training appliance" (PTA), which is used in the United Kingdom [34, 159, 160, 164]. First described by Tudor and Selley [169], it consisted of an acrylic plate covering the hard palate and a U-shaped wire loop extending posteriorly under the soft palate. To provide feedback on palatal movement, a "visual speech aid" (VSA) was devised by replacing the U-shaped loop with two electrodes connected to a light bulb. When a patient lifted the soft palate off the electrodes, the light went out. Tudor and Selley believed that the PTA increased "sensory perception of the soft palate

. . . the patients became aware of its movements and could control them voluntarily" (p. 119). Five of their 11 dysarthric speakers reportedly established "intelligible" speech within three weeks. However, no supporting data were given.

According to recent descriptions of the PTA [137, 159, 160], the U-shaped loop is adjusted to touch the soft palate at rest without elevating it and the end of the loop is located at the point where maximal palatal elevation would be expected to occur. The PTA is generally worn at all times. Practice sessions with the VSA occur several times a day. Ongoing speech therapy teaches patients to use visual, auditory, tactile, and nasal airflow feedback; improve articulation accuracy and speed; discriminate between accurate and distorted productions; and alternate between old and new productions.

Stuffins [160] summarized treatment results for 26 patients whose hypernasality and audible nasal emission of air was related to palatal clefting, neurological disorders, or an unknown etiology. These patients wore the PTA for periods ranging from 3 months to 3 1/2 years. Gradual withdrawal of the appliance was accomplished by increasing the amount of time per day that the patients did not wear their appliances. Forty-two percent of the patients reportedly developed "normal speech and resonance" and an additional 30% made "major improvements" in speech.

Recently, Selley et al. [137] speculated that the PTA may "reduc[e] tongue humping" (p. 382), thus allowing oral airflow to increase and nasal airflow to decrease. Stuffins [160] conjectured that if the "posterior bunching movement of the palatoglossus muscle" (p. 113) is inhibited, the levator palatini muscle would be better able to elevate the velum.

Feedback treatments. Many forms of feedback have been used in behavioral treatments of velopharyngeal impairment during Phase 3. Feedback is often provided by instruments for assessing velopharyngeal function that have been introduced in Phase 3.

By the mid-1970s, clinicians were beginning to investigate how the videoendoscopic image could be used as a form of feedback in the effort to change velopharyngeal function. In 1966, Taub [162] first described the "oral panendoscope" for viewing the velopharyngeal system and in 1975 Shelton and his co-workers [148] successfully taught normal speakers to voluntarily produce closure on nonspeech tasks by having them view the velum through a video playback system linked to the panendoscope. However, these authors were cautious about generalizing their results on normal speakers to speakers with velopharyngeal closure problems. In 1978, Shelton and his colleagues [141] used video feedback provided by the panendoscope to teach two individuals with velopharyngeal impairment to "more frequently approximat[e] complete closure" (p. 6) as they produced vowels and syllables. Shelton did not speculate on whether this type of training would be effective in improving velopharyngeal closure in spontaneous speech. The oral endoscope could be used to provide feedback during production of only certain speech sounds because it interfered with speech articulation and because its view was limited to the oral surface of the velum.

With the advent of rigid nasopharyngoscopy and then flexible fiberoptic nasopharyngoscopy, clinicians and researchers could view the velopharyngeal mechanism without interfering with speech. Since the mid-1970s, clinicians in Japan, Canada, Germany, and the United States have reported using feedback of videonasendoscopic images in treatments to improve velopharyngeal function [17, 63, 115, 194, 195, 198]. Many of these clinicians report success using this type of feedback with small numbers of patients but provide no quantitative data indicative of velopharyngeal function before, during, or after the course of treatment.

The report that used the most subjects was that of Yamaoka, Matsuya, Miyazaki, Nishio, and Ibuki [198] who treated 59 patients with repaired palatal clefts. These subjects attended hour-long sessions provided biweekly for one year. During each session, subjects received nasopharyngoscopic feedback of their velopharyngeal movements as they repeated the activities for which they had not reached complete velopharyngeal closure during the previous session. Some subjects also received "ordinary speech therapy" to correct articulation problems. Following treatment, 59 percent of the subjects demonstrated better velopharyngeal closure during vowels and consonants than was demonstrated before treatment.

Some treatments to improve velopharyngeal function during Phase 3 incorporate feedback of nasal airflow as was also done before Phase 3. These treatments during Phase 3 generally use the See Scape (manufactured by PRO-ED) [49] or an oscilloscope to provide visual feedback of nasal airflow. Usually, this approach is used in combination with other treatment methods [e.g., 25, 112, 196]. The amount of nasal airflow is a function not only of velopharyngeal orifice size but also of nasal pathway resistance and respiratory effort [177, 181]; changes in nasal airflow may be controlled by changes in respiratory effort without changing velopharyngeal orifice size. Therefore, the presence or absence of nasal airflow can be used to indicate that the velopharynx is opened or closed but variations in amount of nasal airflow cannot be used to indicate varying degrees of velopharyngeal closure. For this reason, feedback of nasal airflow must be interpreted carefully or accompanied by feedback of oral air pressure [e.g., 38].

Moller and his co-workers [103] experimented with use of a strain gauge and associated instruments to track velar movement. They provided feedback on this movement to a pre-adolescent boy with a repaired palatal cleft, hypernasality, and evidence of a short velum. He reportedly learned to increase his velar movement during /u/, however, his hypernasality and velopharyngeal gap size did not change,

possibly because adenoid involution occurred during the course of his 15-session, five-week treatment.

Two types of feedback treatment incorporating photoelectric detection of light associated with velopharyngeal function have been suggested. Künzel [80] provided feedback of information obtained from a "velograph." His velograph consisted of a cold light source and photocell inserted transnasally and positioned above the velum; the photocell registered the amount of light reflected off the elevated velum, thus indexing the extent of velar elevation. Künzel used feedback of velograph data in his treatment of four individuals with repaired cleft palates who had shown the ability to attain velopharyngeal closure "at least sometimes during speech" (p. 99) prior to treatment. He reported that these four subjects learned to "actively control their velopharyngeal mechanism under feedback conditions, two of them also under nonfeedback conditions" (p. 98).

Dalston and Keefe [28] used a "photodetector" to provide feedback related to size of the velopharyngeal opening. The photodetector [27] measures the amount of light transmitted throught the velopharynx by using a light source positioned below the velopharyngeal port and a light detector positioned above the port. Dalston and Keefe stated that treatment incorporating photodetector feedback provided by a microcomputer enabled a small group of normal speakers and speakers with repaired cleft palates to alter velopharyngeal movements. They also indicated that hypernasality and audible nasal emission of air was reduced or eliminated for the speakers with repaired clefts. However, they provided no data to support these statements.

Whistling-blowing technique. Shprintzen and his co-workers described a behavioral treatment that used nasal airflow feedback. In addition, they also incorporated use of multiview videofluoroscopy to provide a rationale for their procedure and to show that velopharyngeal closure had improved with treatment. Using multiview videofluoroscopy, Shprintzen, Lencione, McCall, and Skolnick [153] observed that velopharyngeal closure patterns were similar for blowing, whistling, and speech. Subsequently, Shprintzen, McCall, and Skolnick [154] treated four individuals who initially achieved closure during blowing and whistling but not speech. Three of the subjects had a history of palatal clefting and the fourth had a history of hearing loss and velopharyngeal impairment secondary to adenoidectomy. All subjects were said to have frequent audible nasal emission of air or nasal snorts and two were perceived to be severely hypernasal. To elicit velopharyngeal closure during vowel phonation, the subjects were instructed to blow or whistle and, as the blowing or whistling continued, to add simultaneous phonation of a vowel. The "scape-scope" was used to inform the subjects whether nasal airflow was present during vowel production. Later treatment steps required productions of consonants, phrases, sentences, and spontaneous speech as the use of blowing or whistling was extinguished.

Shprintzen and his associates presented what they deemed to be representative traces of multiview videofluoroscopic images before and after treatment to show that all subjects demonstrated greater palatal elevation and pharyngeal wall movement after treatment. In addition, they stated that clinical evaluations revealed that all subjects' speech improved during the treatment period. Shprintzen and his colleagues speculated that these speakers were able to improve velopharyngeal function with behavioral treatment because they had experienced velopharyngeal impairment as a result of an "error in learning."

Shprintzen later [151] observed that speakers who achieved velopharyngeal closure during blowing or whistling also achieved closure during prolonged /s/ and /f/ even though they did not achieve closure during /s/ and /f/ in running speech. Therefore, he advocated using prolonged /s/ or /f/ as a starting point from which to shape use of appropriate velopharyngeal closure during production of these and other speech sounds. He claimed that this procedure was as effective as the use of blowing to shape improved velopharyngeal closure.

Shaping improved velopharyngeal closure from closure achieved on other pressure consonants. This shaping approach has been recommended [e.g., 1, 53, 86, 87, 112, 124, 129, 170] to elicit improved velopharyngeal function for individuals who achieve closure during some, but not all, pressure consonants prior to therapy. The individual begins by producing one of the speech sounds associated with relatively good velopharyngeal closure. Then, during the production of this sound, he or she makes tongue movements as directed by the clinician or researcher, and thereby changes this speech sound into a sound that had been produced with relatively poor velopharyngeal closure prior to treatment. The intent is to maintain the velopharyngeal posture reached for the first sound throughout the production of the second sound.

Hall and Tomblin [53] used this approach to treat a first-grade student who consistently nasalized /s/ and /z/, but who orally produced all other nonnasal sounds including /f/, /θ/, and /t/. They first instructed the child to produce /f/ and "slide" to the /s/ while continuing the oral airstream. This failed to stimulate nonnasal production of /s/, as did attempts to "slide" from /θ/ to /s/. Next, they told the boy to produce a /t/ and then release the tongue-palate contact of the /t/ into the fricative /s/. This elicited an /s/ produced without audible nasal emission of air, and presumably with velopharyngeal closure. They used a variety of articulation therapy techniques to help the child successfully establish nonnasal production of /s/ and generalize it to conversational speech.

Others [112, 194, 195] have presented case reports in which this approach was used to treat individuals with

velopharyngeal impairment secondary to a variety of conditions including palatal clefting. Feedback of nasal airflow information [53, 112] or videonasopharyngoscopic images [194, 195] was provided as part of these treatments.

Articulation therapy. During Phase 3, another type of articulation therapy has been advocated to improve velopharyngeal closure. Hoch et al. [63] claimed that improved velopharyngeal function is sometimes associated with therapy to eliminate compensatory articulation errors such as glottal and pharyngeal substitutions. To support this claim, they summarized Shprintzen's unpublished comparison of pretreatment and posttreatment nasopharyngoscopic and multiview videofluoroscopic examinations of patients who received therapy to correct compensatory errors. Of his 72 patients with "near or total absence of velar and lateral pharyngeal wall motion" before treatment, 34 showed "improve[d] . . . degree of velopharyngeal movement" after treatment.

Ysunza et al. [199] published results that were compatible with the clinical experience of Hoch et al. and Shprintzen. Ysunza and his co-workers studied 31 children with repaired palatal clefts who used compensatory substitutions and who were hypernasal. They gave these children 6 to 15 months of speech therapy to replace glottal and pharyngeal substitutions with oral articulatory placements. They then examined the effects of this therapy on velopharyngeal function by comparing results of pretreatment and posttreatment videonasopharyngoscopic and multiview videofluoroscopic examinations. Ratios of velar movement and ratios of lateral pharyngeal wall movement[1] were significantly greater after treatment than before and velopharyngeal gap size was significantly smaller after treatment than before. Thirteen of the children achieved complete velopharyngeal closure during the posttreatment examination.

These reports suggest that the initial step in management for speakers who use glottal stops and pharyngeal articulations should be speech therapy to replace these articulations with oral placements rather than surgery to correct a presumably incompetent velopharyngeal port [1, 63, 167, 199]. Some of those who use glottal and pharyngeal places of articulation may not be closing the velopharyngeal port during these sound productions not because they have an incompetent velopharyngeal mechanism but because velopharyngeal closure is not needed to produce these glottal and pharyngeal articulations. When the individual learns oral articulatory placements, velopharyngeal closure becomes necessary to achieve adequate intraoral air pressure for the sounds. If velopharyngeal function improves as the individual learns oral articulatory placements--that is, if the individual leanrs to use correctly a velopharyngeal mechanism that is anatomically and physiologically adequate--then neither surgery nor obturation is needed. Broen et al. [15] described how they used this type of articulation therapy to explore whether improved velopharyngeal function was possible without surgery or obturation.

Resistance training using Continuous Positive Airway Pressure (CPAP). An implicit assumption involved in the procedures described above is that the individual being treated possesses the inherent anatomic and physiologic capability to improve velopharyngeal functioning for speech. This view assumes that the individual can "learn" how to use the mechanism in a more appropriate manner thereby improving functioning for speech. However, it is possible that there are physical limitations that must be overcome before additional modification in behavior can be achieved. An example of such a limitation might be insufficient muscle strength or endurance coupled with a threshold of fatigue that is relatively low. An obvious approach to alleviate such a problem is to increase muscle strength or endurance using a resistance exercise regimen focused on the muscles of the soft palate. Although this concept is not new during Phase 3, until recently, resistance exercise of the muscles of the velum was not applied in a practical or theoretically sound manner.

In 1991, Kuehn [73] proposed a resistance type of treatment for strengthening the muscles of velopharyngeal closure that utilizes well-known principles of exercise physiology [13, 16, 97]. Resistance to the muscles is provided by delivering heightened air pressure to the nasal cavities using a commercially available device referred to generically as continuous positive airway pressure (CPAP). CPAP is used routinely to treat individuals with obstructive sleep apnea [136].

Major principles of the treatment proposed by Kuehn [73] include overloading, progressive resistance, and specificity of the exercise regimen, all of which are known to increase muscle strength. Muscles are subjected to a load greater than that they typically work against. This is called "overloading." As the muscles practice working against this load, they adapt to the load and strength is increased. The greater the overload that the muscles adapt to, the greater the strength gain achieved for the muscle or muscle group being exercised. As the muscles adapt to a given load, the "intensity" of the exercise program is increased by increasing the size of the load, such as increasing the weight of an object lifted. Thus, the muscles are again overloaded. This process is repeated in gradually increasing increments; this gradual increase in exercise intensity is called "progressive resistance training." Such an increase is necessary as muscles adapt to each load bearing capacity. If progressive resistance does not occur, further increases in strength will not be achieved.

[1] Larger ratios are consistent with greater movement of the velum or lateral pharyngeal walls. See Golding-Kushner et al. [50] for further explanation of these ratio measures.

Specificity involves exercising the particular muscles and, presumably, activating the corresponding motor nerve pathways that will be used during a given task. This is an important consideration with regard to the muscles of velopharyngeal closure because nonspeech activities, such as swallowing and blowing, have been used frequently in attempts to strengthen the muscles. However, it is known that nonspeech tasks, such as swallowing, involve very different velopharyngeal movement patterns compared to that for speech [102, 153]. Thus, it is advantageous to provide resistance to the muscles of velopharyngeal closure during speech rather than nonspeech activities if increases in strength during speech are to be expected.

The muscles of velopharyngeal closure appear to be vulnerable to fatigue because velar elevation during speech is a repetitive activity that may occur frequently within a relatively short amount of time. For individuals with a normal mechanism, levator veli palatini muscle activity for speech occurs at a fairly low level within its total operating range [76] and thus is not likely to be affected by fatigue. However, for individuals with surgically repaired cleft palate, levator activity for speech occurs at a relatively high level within its total operating range [77] and, presumably, is much more susceptible to the effects of fatigue.

It is possible that individuals with marginal velopharyngeal impairment, in particular, may tend to elevate the velum to a position involving muscle activity that is just below the muscles' threshold of fatigue. This would tend to avoid a situation in which fatigue occurs and velar elevation becomes very difficult or impossible resulting in severe hypernasality. Unfortunately, the velar position achieved may not be sufficient for tight velopharyngeal closure and some degree of hypernasality might result. The degree of velopharyngeal impairment may vary for a given individual in relation to a muscle's momentary proximity to the threshold of fatigue which, in turn, could be influenced by such factors as the individual's general energy level. If fatigue effects are at issue, it appears that individuals with velopharyngeal impairment for speech would benefit from exercises that increase muscle endurance as well as strength. Endurance refers to a muscle's ability to maintain a constant tension over prolonged or repeated contraction [2].

The CPAP procedure [73] could potentially increase both muscle strength as well as endurance because it overloads the levator veli palatini muscle [78] and involves a regimen with a large number of repetitions of velar elevation. The CPAP instrumentation consists of an air pressure generating source and a hose and mask assembly. The mask is placed over the subject's nose, and the increased air pressure is delivered directly to the nasal cavities. Specially constructed speech utterances of the form VNCV, such as /insi/, are used for drill work. These nonsense words are designed to lower the velum during the nasal consonant (N) followed by a rigorous elevation of the velum for the following pressure consonant (C) much like the power lift in weight training. The heightened air pressure in the nasal cavities provide the "weight" against which the muscles of velopharyngeal closure must work.

Therapy sessions are conducted in the subject's home and involve one session per day, six sessions per week, with a total duration of eight weeks. The time of each session progresses from 10 minutes during the first week to 24 minutes during the eighth week. The air pressure delivered to the nasal cavities also increases in accordance with the principle of progressive resistance exercise such that the lowest pressure is 4.0 cm H_2O during the first week and the highest pressure of 8.5 cm H_2O is reached during the eighth week.

The efficacy of this therapy procedure currently is being tested.

Treatments to Reduce Perceived Nasalization by Changing Respiratory, Laryngeal, or Oral Articulatory Behaviors

During Phase 3, clinicians and researchers have continued to suggest methods for reducing perceived hypernasality and audible nasal emission of air by altering respiratory, laryngeal, or oral articulatory behaviors without necessarily improving velopharyngeal function. Methods suggested during earlier phases are still recommended. These include using a greater range of oral motion [12, 25, 91, 92, 100, 111, 112, 124, 158, 170, 174, 178, 179, 192], using light contacts for consonants [25, 91, 100, 127, 170, 174], speaking at a slower rate [100, 111, 192], acquiring good oral articulation skills [54, 124, 158, 170, 192], and altering loudness and pitch [12, 25, 39, 91, 100, 112, 125, 158, 192]. However, authors during this phase have warned that increasing loudness or altering pitch might lead to phonation disorders [e.g., 100].

Although these methods have been suggested often in the literature, Starr [158] recently pointed out that clinical researchers have yet to show that they effectively reduce perceived nasalization. In fact, research related to some of these procedures suggests that they may not yield the expected reductions in hypernasality for all hypernasal speakers. For example, studies pertaining to speaking rate and hypernasality indicate that speakers with repaired palatal clefts are not necessarily more nasal when using speaking rates faster than their typical rate [67, 68]. Other investigators found that normal speakers were judged to be more nasal when reading slowly, in word-by-word fashion, than when reading at their typical rate [23] and noted that some normal speakers occasionally opened the velopharyngeal port during nonnasal utterances produced at slow rates [20, 163]. If this latter behavior is also present in those with velopharyngeal impairment, speaking at a slow rate might increase hypernasality. Even increasing the range of oral motions may not decrease hypernasality for all who have velopharyngeal impairment. Some authors [100, 122] have

warned that when speakers move the mandible and tongue downward to speak with a greater-than-typical oral opening, mechanical connections between those structures and the velum may cause the velopharyngeal port to open more, thus increasing hypernasality.

During Phase 3, various types of feedback have been suggested for use in treatments to reduce hypernasality. Such feedback includes vibrational data from nasal accelerometers or nasal/voice accelerometric ratios [64, 84, 114, 128] and "feedback filtering" [19, 48]. However, Redenbaugh & Reich [128] cautioned that clinical use of vibrational data awaits further study of accelerometric ratios in normal and hypernasal speakers.

Probably the most popular device for providing feedback related to hypernasality is the updated version of TONAR. During Phase 3, TONAR has been replaced by the Nasometer™ (Kay Elemetrics Corporation). The Nasometer™ uses a nasal microphone and an oral microphone separated from each other by a plate positioned above the upper lip. Like TONAR, the Nasometer™ measures nasalance and has been recommended to provide feedback in behavioral treatment to reduce hypernasality. Current research has addressed use of the Nasometer™'s nasalance as an index to speech nasality [29, 55]. Also, recently, several authors have emphasized that nasalance probably reflects not only extent of oral-nasal coupling, but also degree of oral opening, height of tongue carriage, respiratory effort, patency of the nasal airway, or some combination of these [41, 42, 55, 100]. Presumably, changes in one or more of these variables during treatment might induce reduced nasalance. Although there are no published studies of the effects of treatments using the Nasometer™ to provide nasalance feedback, Starr [158] described an unpublished study by K. Burrell who used this type of feedback in the treatment of two individuals with inconsistent mild hypernasality. She found that nasalance measured during sentence lists was less after treatment than before it and listener judgments indicated that hypernasality decreased. For one subject with a repaired cleft, neither lateral nor frontal videofluoroscopic images revealed change in velopharyngeal movements. However, for the other subject, who had no history of clefting, lateral videofluoroscopic evaluation indicated no velar-pharyngeal wall contact pretherapy, but consistent contact after therapy.

Overview of Current Opinions and Evidence Concerning Behavioral Treatments to Improve Velopharyngeal Function

The above historical review indicates that behavioral treatments to improve velopharyngeal function are being studied and used today. Although much remains to be learned about which treatments are likely to be effective and who these treatments might help, this section will summarize current opinion and evidence on these questions.

Which treatments may be effective?

Starr [157] concluded from his review that treatments consisting of speech activities are more likely to be effective than treatments consisting of nonspeech activities. These treatments would have the most face validity [73]. Virtually all currently advocated treatments focus on improving velopharyngeal function during speech activities. Some treatments that do include a nonspeech activity either combine nonspeech and speech behaviors [e.g., 154] or use the nonspeech activities for only a short time early in treatment to teach young clients the desired oral direction of airflow [e.g., 170].

Currently, the most often advocated treatment component is provision of instrumental feedback related to velopharyngeal function or movement [157]. By providing a speaker with such feedback, the clinician is supplementing the sensory information the speaker can receive via audition and somesthesis [32]. Because awareness of velopharyngeal position and kinesthesis appears to be poor [e.g., 138, 140], use of instrumental feedback may be a logical treatment approach. Results of studies incorporating feedback have been described as promising [100, 133, 157]; however, reviewers of this literature agree that additional research is needed to document the effectiveness of these procedures. In particular, future studies need to assess the relative effectiveness of various components of treatments incorporating feedback. As Starr [158] pointed out, many individuals receiving feedback as one treatment component have simultaneously received other types of therapy such as articulation therapy [e.g., 196, 198], shaping from blowing [e.g., 154], or shaping from other consonants produced with closure [e.g., 194]. Using two or more treatments simultaneously complicates the interpretation of which treatment component contributed to change in velopharyngeal function. Using feedback alone would appear to be useful only if the patient reaches velopharyngeal closure in at least some contexts prior to treatment or quickly acquires the ability to reach closure in some contexts in the early stages of treatment. Furthermore, to successfully resolve speech problems associated with velopharyngeal impairments, feedback procedures must be combined with other procedures that develop skilled velopharyngeal movements during automatic conversational speech [100].

McWilliams et al. [100] reviewed the literature describing muscle training, articulation therapy, obturator reduction, and biofeedback as methods of improving velopharyngeal function. They stated that obturator reduction "appears to be a more promising technique of developing velopharyngeal motion and incorporating it into speech" (p. 409) than the other three techniques they reviewed. However, they emphasized that more data are needed to support the effectiveness of this treatment.

The therapy procedure involving muscle training using CPAP [73] had not been published yet at the time of the McWilliams et al. [100] review. Compared to other recently advocated procedures, the primary theoretical advantage of the proposed resistance training using CPAP is that it may alleviate physical limitations of the velopharyngeal mechanism by increasing muscle strength, endurance, or both. An additional advantage is that the procedure can be conducted in the subject's home thereby permitting frequent and convenient treatment sessions. However, as for other currently advocated behavioral treatments to improve velopharyngeal function for speech, efficacy of the CPAP procedure must still be demonstrated.

Who may profit from behavioral treatments?

Clinicians and researchers interested in behavioral treatments to improve velopharyngeal function acknowledge that only some individuals with velopharyngeal impairment are likely to have the potential to profit from these treatments. Among those judged most likely to benefit are individuals with inconsistent or borderline velopharyngeal "competency" [22, 25, 63, 80, 100] or borderline velopharyngeal "adequacy" [179], individuals with mild-to-moderate consistent velopharyngeal "dysfunction" [87], individuals who have recently received secondary palatal surgery [25, 170, 195], individuals with "phoneme-specific" nasal emission of air [1, 121], and individuals who achieve velopharyngeal closure during oral blowing prior to treatment [100, 154, 158, 166, 198]. However, being a member of any of these categories does not guarantee improvement of velopharyngeal function with behavioral treatment [74, 100, 170]. Anatomic or physiologic impairment of the mechanism may limit speakers' ability to profit from these treatments even for those who belong to one of the above categories. For example, if a palate is too short, velopharyngeal closure will not be achieved no matter how well the palate or pharyngeal walls move.

Most clinicians and researchers using behavioral methods to improve velopharyngeal function have administered these treatments to speakers with repaired cleft palates [100, 132, 133, 157, 158]. Others have used these methods to treat velopharyngeal impairment and hypernasality accompanying neurologic problems [e.g., 11, 79, 113, 169] or hearing loss [44], but reports of applications of the methods to velopharyngeal problems associated with these disorders are relatively infrequent. Often, the etiology of velopharyngeal impairments treated behaviorally is unknown [e.g., 85, 134].

Several authors claim that phoneme-specific nasal emission of air is usually relatively easy to resolve [e.g., 25, 63, 166, 170]. However, others have written that this is "sometimes a difficult habit to break" [1] and may not change even after extensive traditional articulation therapy [121, 194]. Some children with phoneme-specific nasal emission of air learn to correct this problem when producing single words or sentences but continue to have difficulty reaching velopharyngeal closure for sibilant and fricative consonants in conversational speech [170].

Although both children and adults have been treated with behavioral procedures, most of the subjects of reported treatments have been children. In addition, based on experience with a series of 125 patients, Weiss [183] stated that younger patients are more likely than older patients to experience successful obturator reduction and elimination.

Why Continue to Search for Effective Behavioral Treatments of Velopharyngeal Impairment?

As is obvious from this review, much of what constitutes current thoughts on these treatments is still opinion, supported by little or no conclusive research. Considering the number of years devoted to use and study of the procedures, there seems to be little hope of showing that these procedures are effective. Why, then, is it reasonable to continue to search for ways of behaviorally treating velopharyngeal impairment? There are three answers to this question: (a) there is a clinical need for noninvasive treatments, (b) the velopharyngeal musculature should be amenable to voluntary change, and (c) several studies describing the nature of velopharyngeal impairments suggest that certain speakers may have the potential for improved closure without surgery.

There is a Clinical Need for Noninvasive Treatments

Apparently, many clinicians and researchers have continued to use and develop behavioral treatments because they have believed that some speakers with velopharyngeal impairment have speech disabilities needing management alternatives other than surgery or long-term use of palatal prostheses [14, 122]. These speakers include those whose velopharyngeal-related speech disabilities are inconsistent or relatively mild, those who have pharyngeal flaps but still demonstrate nasalized speech, and those who have phoneme-specific nasal emission of air.

Physicians and speech-language clinicians may be uncertain whether some of these speakers have velopharyngeal mechanisms capable of supporting normal speech [56] and, therefore, they may be reluctant to recommend surgery because it is a risky, costly, permanent alternative that may create other problems such as impaired nasal respiration. Even if surgery is recommended, the individual or his or her careproviders may be unwilling to agree to surgery as an appropriate treatment alternative for a problem they consider to be relatively minor [100]. Therefore, for these speakers, surgery may be excessive treatment.

In other cases, clinicians may judge that surgical or prosthetic treatment will not resolve the speech disability because it is related not to problems of velopharyngeal anatomy or physiology but to inappropriate learning [e.g.,

100] or the persistence of behaviors learned earlier in life when problems of anatomy or physiology did exist [e.g., 170]. For these speakers, surgery or long-term use of prostheses would be ineffective treatment of the velopharyngeal-related speech disability.

Rather than recommend possibly excessive or ineffective physical management for these individuals, clinicians often recommend trial behavioral treatment to improve the function of the velopharyngeal mechanism during speech [25, 87, 100, 133, 136].

Velopharyngeal Movements Should Be Amenable to Voluntary Change

The musculature of the velopharyngeal mechanism is of the skeletal type. Therefore, speakers should be able to voluntarily alter its movements and learn new movement patterns [100]. It is known that at least some individuals without velopharyngeal impairment can elevate the velum upon command apart from speech or vegetative activities [76, 144, 145]. Normal speakers have also been observed to modify speech-related velopharyngeal movements in certain experimental contexts [e.g., 105, 155].

Structural and physiologic characteristics of the velopharyngeal mechanism, as well as other variables such as intelligence and hearing sensitivity [158], may affect a person's ability to alter movements of the velopharyngeal mechanism. Speakers with impairments of velopharyngeal anatomy or physiology may not have as much ability to modify velopharyngeal movements as speakers with normal velopharyngeal mechanisms but they should be able to make modifications within the limits of their mechanisms [78, 100]. If an individual has not yet tried movements that the mechanism is capable of or has not yet learned to use specific movements in appropriate contexts, behavioral treatments may be a reasonable approach to improving velopharyngeal function [100].

There is Evidence Suggesting that Velopharyngeal Closure May Improve Without Surgery

While the studies to be considered here did not involve behavioral treatments to improve velopharyngeal function, they provided information compatible with the belief that some speakers with velopharyngeal impairments may have the potential to improve closure. They showed that individuals with speech disorders secondary to repaired palatal clefts may demonstrate substantially better velopharyngeal performance under certain conditions. In addition, they revealed that some speakers with velopharyngeal impairments who have perceptually good or normal speech use atypical velopharyngeal behaviors that may have been acquired as compensatory behaviors.

Better Velopharyngeal Function Under Certain Conditions

Research has shown that some individuals with speech disorders secondary to repaired palatal clefts demonstrate substantially greater range of velopharyngeal movements and greater muscle activity during certain speech tasks than others. The circumstances under which these types of intrasubject variability occurred suggest that these individuals might be capable of acquiring improved velopharyngeal closure for speech without surgery.

Henningsson and Isberg showed that some speakers with repaired clefts demonstrate large intraspeaker variability in velopharyngeal closure movements as viewed with cineradiography. In one study, they [60] found that speakers who used some glottal stops and some oral stops failed to close for the glottal stops but achieved closure, or more nearly approximated closure, for the oral stops. In another study, they found that speakers with fistulas failed to close completely for obstruents produced anterior to the fistulas but did achieve velopharyngeal closure for obstruents produced posterior to the fistulas [65]. Moreover, temporary closure of the fistulas was associated with complete, or nearly complete, velopharyngeal movements for obstruents anterior to the fistula [65]. The better velopharyngeal closure observed during oral stops, obstruents posterior to fistulas, or obstruents produced with the fistula covered was accomplished primarily by increased medial movement of the lateral pharyngeal walls and, to a lesser extent, by increased elevation of the velum. These studies demonstrated that absence of appropriate velopharyngeal movements during a glottal stop or during an obstruent produced anterior to an open fistula does not necessarily indicate that the speaker is incapable of velopharyngeal closure in other contexts.

Another study that showed intraspeaker variability of velopharyngeal behavior was that of Kuehn, Moon, and Folkins [78]. They recorded electromyographic (EMG) activity in the levator veli palatini muscle while continuous positive airway pressure (CPAP) was directed intranasally. They wanted to know whether the positive airway pressure was associated with increased activity in the levator muscle. Subjects in this study were four hypernasal speakers with repaired clefts and five normal speakers without clefts. Kuehn and his co-workers measured EMG activity in the levator muscle during speech produced when nasal pressure was equal to atmospheric pressure and when several levels of positive air pressure were introduced intranasally. They found that levator activity for both groups of subjects was significantly greater while the subjects were receiving the positive intranasal air pressure compared to the atmospheric pressure condition. This finding indicated that the levator muscles of even the hypernasal speakers had some reserve potential that they did not use under the normal speaking condition.

If speakers such as those described in these studies can be helped to learn use of their optimal velopharyngeal behaviors, their speech might be improved. Anecdotal reports and research evidence described earlier in this chapter suggest that this might be possible. For some speakers, articulation therapy to eliminate glottal stops reportedly leads to improved velopharyngeal movements [e.g., 25, 51, 63, 199]. Similarly, preliminary tests of Kuehn's [73] use of CPAP to provide resistance exercise of velopharyngeal closure muscles have suggested that velopharyngeal function was improved for some speakers. Perhaps other behavioral treatment methods would also facilitate speakers' use of any untapped potential for improved velopharyngeal function.

Compensatory Behaviors

Some speakers with repaired clefts and perceptually good or normal speech use velopharyngeal behaviors different from those typically used by speakers with repaired clefts and nasalized speech or by normal speakers without clefts. These findings led investigators to speculate that some speakers acquire atypical velopharyngeal behaviors to compensate for velopharyngeal impairments, thereby avoiding nasalized speech.

Karnell, Folkins, and Morris [71] observed velopharyngeal movement patterns of four adults who were judged to have "touch closure in the midsagittal plane" during production of /s/ in a sentence context. Two of the subjects were judged to be more nasal than the other two subjects who were perceived to have mild hypernasality. For these four subjects velopharyngeal function during consonant-vowel-consonant (CVC) utterances composed of oral consonants was observed on cinefluorographic films. During the CVC utterances, the two subjects judged to be most nasal showed a lower velar height during the vowel than during the consonants, which is the movement pattern typical of normal speakers. In contrast, the two subjects judged to be least nasal showed no decrease in velar height during the vowel; that is, they maintained a relatively high velar position throughout the syllable.

One explanation offered by Karnell et al. was that "velopharyngeal movements normally expected in CVC utterances may be avoided by some speakers with cleft palate in order to minimize perceptible nasalization" (p. 63); that is, the two subjects who maintained a relatively high velar position may have developed compensations for anatomic or physiologic limitations imposed by their repaired clefts. This study offered no information about how these speakers came to acquire the apparently compensatory behavior of keeping the velum elevated during the entire syllable. Nevertheless, Karnell et al. speculated that when these variations in movement patterns are better understood, "behavioral adjustments in movement timing" perhaps could be helpful to speakers demonstrating marginal velopharyngeal competence.

Dalston, Warren, and Smith [30] compared the air pressure and nasal airflow patterns of two groups of adults as they spoke the word "hamper." One group had repaired cleft palates and the other group were normal speakers without clefts. None of the subjects in either group had perceptible hypernasality or nasal emission of air. All subjects' velopharyngeal closure met the respiratory requirements for speech; that is, their estimated velopharyngeal orifice areas were less than 5 mm^2 [e.g., 180]. During "hamper" the group of speakers with repaired clefts had larger mean velopharyngeal orifice areas but less mean nasal airflow than did the group of speakers without clefts. Dalston et al. further analyzed some of their data to show that the presence of less nasal airflow in the group of speakers with repaired clefts was not related to greater nasal pathway resistance in this group. Dalston and his co-workers reasoned that these speakers may have used less nasal airflow during /m/ to avoid producing /p/ with audible nasal emission. They also speculated that these speakers reduced nasal airflow either by using less respiratory effort or by opening the velopharyngeal port less for /m/ and closing it sooner for /p/. Thus, these authors, like Karnell and his colleagues, suggested that their subjects with repaired palatal clefts may have acquired some type of compensatory velopharyngeal behaviors that improved their speech. If these or other atypical velopharyngeal behaviors are acquired compensatory behaviors, perhaps behavioral treatments could be devised to facilitate acquisition of such behaviors in other speakers with speech disability secondary to velopharyngeal impairment.

Clinical and Research Considerations for Testing the Effectiveness of Behavioral Treatments

Clearly, additional information is required to guide clinical use of behavioral treatments to improve velopharyngeal function. Clinicians and researchers need to systematically observe and describe the results of behavioral procedures for modifying velopharyngeal closure movements and to identify the characteristics that distinguish individuals for whom the procedures are successful and unsuccessful. Testing the effectiveness of these treatments is complicated by a variety of clinical and research considerations.

Clinical Considerations Protecting Individuals from Long-Term Ineffective Treatments

Because of the lack of conclusive research demonstrating that behavioral treatments of velopharyngeal impairment are effective for more than a few individuals, clinical use of these treatments should continue to be experimental at this time. Clinicians and researchers should insure that clients and subjects understand the nature of the experimental treatments, particularly that it is not known whether

the treatments are effective. In addition, they should explain other treatment alternatives that may be implemented instead of, or following, the experimental behavioral treatment [157].

Individuals agreeing to undergo behavioral treatments should be told the approximate duration of the treatment if no progress is made. Many authors [14, 22, 54, 87, 100, 118] have recommended that these treatments should not be continued longer than three to six months if no progress occurs. However, some treatments may elicit progress in much less time--one or a few sessions [e.g., 12, 63, 100, 108]--if they elicit progress at all. Morris [108] stated that if appropriate oral-nasal resonance does not begin to generalize to conversational speech in about 12 hours of therapy, then a physiologic deficit may be preventing progress; however, he indicated that more than 12 hours of therapy may be needed to bring about generalization in very young clients. Boone and McFarlane [12] expressed their belief that treatment for velopharyngeal impairments should be relatively intensive--that is, there should be a minimum of three sessions per week.

Developing Treatment Goals and Procedures to Be Studied

From a clinical perspective, successful treatments of velopharyngeal impairment must resolve, or at least alleviate, any velopharyngeal-related speech disability in spontaneous speech. Thus, neither improved velopharyngeal movements nor reduced nasalization during prolonged vowels or sentences read aloud is a satisfactory clinical outcome by itself. Even improved velopharyngeal function during spontaneous speech is not a satisfactory outcome unless it is associated with reduced nasalization in spontaneous speech. Therefore, experimental behavioral treatments should either aim to change nasalization in conversational speech or subsequent studies will need to show how treatment outcomes other than this may eventually contribute to reduced nasalization in spontaneous speech. To ascertain whether treatments have made clinically significant improvements in speech, measures of perceived nasalization should be included as dependent variables.

For many clients, successful therapy to reduce nasalization is likely to require combinations of treatments. This is expected because some currently advocated treatments are designed to facilitate improved velopharyngeal function only at certain stages of the therapy process and because clients differ in their ability to respond positively to the various types of treatments. For example, even if various kinds of feedback are shown to be useful in eliciting new velopharyngeal behaviors, other procedures such as negative practice, listening training, or training in self-monitoring probably will be necessary to insure that the new behaviors become automatic in daily conversational speech [e.g., 12, 39, 112, 170, 192]. And, if after receiving behavioral and physical management to improve velopharyngeal function, an individual's speech continues to be nasalized, treatments to change other speech mechanism behaviors--such as those controlling range of oral movements--may be tried in an attempt to further reduce listeners' perceptions of nasality. Such clinical use of treatment combinations complicates efforts to study the effectiveness of various treatment components. However, at this time studies of individual treatment components (such as feedback procedures to elicit improved velopharyngeal movements during speech) and studies of treatment combinations to resolve the communication problem in spontaneous speech would be useful additions to the literature.

Research Considerations

Treatment Effectiveness Studies

Previous reviewers of this literature have pointed out that experimental, rather than descriptive, research designs are needed to show whether or not these behavioral treatments improve velopharyngeal function. McWilliams et al. [100] called for studies controlling variables potentially affecting treatment outcome and Witzel [193] noted that effectiveness studies should be "strictly controlled, randomized clinical trials" (p. 152). To our knowledge, no studies of these treatments have compared performance of experimental and control groups [see also 193].

Group studies of these treatments are difficult to conduct primarily because of limited subject availability. The number of potential candidates for behavioral treatments seen by any one clinician or researcher is typically small. Moreover, the population of those with velopharyngeal impairment is heterogeneous with respect to many variables that may influence treatment outcome (e.g., oral and velopharyngeal anatomy and physiology, hearing, and intelligence). These problems significantly hinder researchers' efforts to form appropriate experimental and control groups. According to Van Demark and Hardin [170], treatment research "controlling all variables is essentially impossible to conduct" (p. 805).

To address the problem of insufficient numbers of subjects, clinical researchers may study a series of subjects over time, continuing until an adequate sample size has been studied [152]. Adequate sample sizes may be even more easily obtained if researchers from several institutions conduct treatment studies collaboratively [152].[2] Although these two strategies may increase sample size, neither eliminates the problem of subject heterogeneity or the difficulty

[2] Kuehn and his co-workers from six cleft palate centers are currently using these strategies to obtain subjects for an investigation of the efficacy of the CPAP therapy described previously in this chapter.

of controlling for a multitude of relevant subject variables. Nevertheless, strategies such as these may allow researchers to compare experimental and control groups with sufficient statistical power to enable certain conclusions to be drawn concerning treatment outcomes.

Because the need for data-based research regarding these treatments is so great, clinicians and researchers who have access to only small numbers of potential subjects with velopharyngeal impairment should consider developing studies of single subjects or small groups of subjects [116, 139, 158]. Experimental single-subject studies may use withdrawal and reinstatement of treatment, alternating treatments, changing performance criteria, or multiple baselines. These design features allow investigators to exert experimental control and make inferences about whether the treatment actually caused improvement for the subject or subjects being studied.

There have been at least two unpublished attempts to use experimental single-subject studies to examine behavioral treatments of velopharyngeal impairment. Leeper [82] mentioned using single-subject designs to demonstrate progress in treatment using the Nasometer to provide nasalance feedback. However, she did not provide subject data. Kuehn and Hinkle [75] currently are analyzing data within a single-subject design study in which five subjects with hypernasality were treated with CPAP. Each subject was enrolled in the study for eight months. Audiotape recordings and measures of nasalance were obtained at the beginning of each month. Baseline measures were obtained followed by a two-month therapy period which, in turn, was followed by post-treatment measures. The time period within the eight-month interval during which treatment was provided was staggered across subjects, making this study's design a variation of the multiple baselines across subjects' design.

No experimental single-subject studies of behavioral treatments of velopharyngeal impairment have appeared in the literature. This is probably because applying single-subject designs to test these treatments involves several practical difficulties. These difficulties include the need to continuously measure velopharyngeal behavior during an experimental single-subject study, clinicians' hesitance to withdraw treatment if it appears to be successful, inability to reliably observe differences in slowly changing behavior with withdrawal or changing criterion designs, and crossover effects between adjacent treatment phases. Despite these difficulties, development of innovative single-subject studies should be possible under some circumstances.

Technically, a cause-and-effect relation between treatment and improved velopharyngeal function cannot be inferred unless an experimental design has been used and these designs should be employed whenever feasible. However, because of the difficulties faced in both group and single-subject experimental studies, many investigations are likely to be nonexperimental case studies. Even some studies with relatively large sample sizes may not include experimental controls. For these nonexperimental studies, investigators should use procedures that would logically increase the internal validity of their conclusions even though experimental controls have not been used. Such procedures include: (a) describing changes in velopharyngeal behavior and speech quantitatively rather than only anecdotally; (b) assessing velopharyngeal behaviors and speech at several points in time to describe the baseline and to observe the course of changes associated with treatment; and (c) reporting treatment-associated changes observed in more than one subject because the more individuals showing changes with treatment, the more unlikely an extraneous event is responsible for the changes [72, 157].

Prediction Studies

Because behavioral treatments are not likely to improve velopharyngeal function for all those with velopharyngeal impairment, clinical researchers should explore how to predict who may profit from these treatments. When large numbers of subjects are available, multivariate statistical methods may be used to learn which subject variables distinguish individuals for whom the treatments are successful and unsuccessful. When few subjects are available, investigators may address the prediction problem by describing in detail characteristics of subjects and then tentatively predicting that other individuals with similar characteristics would respond similarly to the same treatment [59].

The subject characteristic most often mentioned as possibly predictive of improved velopharyngeal closure with treatment is inconsistent velopharyngeal closure prior to treatment [e.g., 22, 25, 63, 100, 118, 133, 154, 189]. To examine whether inconsistent closure is predictive of progress with treatment, pretreatment patterns of inconsistent closure during spontaneous speech should be precisely described. Subjects' ability to reach closure inconsistently also may be probed in contexts other than spontaneous speech by observing whether velopharyngeal function improves under various stimulability conditions [e.g., 109, 127, 165]. The predictive value of various patterns of inconsistent velopharyngeal closure during both spontaneous speech and stimulability conditions are worthy of study.

Although inconsistency of closure during pretreatment spontaneous speech and stimulability tasks may be related to potential to improve with treatment, it is unlikely that inconsistency alone would predict progress with high accuracy. Several authors [e.g., 100, 170] have warned that, in their clinical experience, occasional velopharyngeal closure does not necessarily indicate that an individual is capable of normal velopharyngeal closure. They believe that, for some individuals, limitations of anatomy or physiology preclude consistent velopharyngeal closure. For this

reason, the predictive value of subject variables such as palatal length, pharyngeal depth, and pretreatment velar and pharyngeal wall movement also deserves study.

Velopharyngeal performance early in treatment should also be considered as a possible predictor of response to continued treatment [cf. 3, 33]. That response early in treatment might contribute to predicting progress later in treatment is compatible with often-made clinical recommendations for trial therapy prior to making the final decision for secondary surgery.

A variety of other subject variables may also strengthen predictions about who may profit from these behavioral treatments. These variables include: etiology of velopharyngeal impairment, age at intervention, chronological age, concomitant surgical or dental treatments, oral and velopharyngeal anatomy and physiology, hearing, and intelligence.

Summary

The literature supporting the effectiveness of behavioral treatments of velopharyngeal impairment is a diverse blend of clinical opinion and data-based studies. Some clinicians and investigators who develop and advocate these behavioral treatments give no quantitative results of their proposed treatments; most offer minimal pretreatment and posttreatment data. This literature constitutes, at best, very weak support for using behavioral methods to improve velopharyngeal function.

Nevertheless, we believe that barring prohibitive anatomic or physiologic constraints, it should be possible, using appropriate behavioral procedures, to strengthen muscles of velopharyngeal closure, to increase range of velar and pharyngeal wall movement, to increase force of velopharyngeal closure, or perhaps to modify the timing or pattern of velopharyngeal closure. However, much more information is needed to develop and support such procedures and to match treatment approaches with individuals' deterministic anatomic and physiologic limitations. We need more basic information regarding the detailed anatomy of the velopharyngeal mechanism and physiologic capabilities and limitations given those structural characteristics. And, we need more quantitative data regarding the effectiveness of various behavioral treatments.

Because of the current lack of research supporting these procedures, they should continue to be considered experimental. We encourage both clinicians and researchers to pursue the best studies possible with their available resources. Experimental group designs should be used whenever feasible. When these designs are not practical, information pertaining to the effectiveness of these treatments may be obtained from other research approaches such as single-subject experimental designs and well-documented clinical case studies [cf. 4, 59]. Results of this research should help clinicians select and provide appropriate treatments for those with speech problems associated with velopharyngeal impairment.

References

1. Albery, L. Approaches to the Treatment of Speech Problems. In J. Stengelhofen (Ed.), Cleft Palate: The Nature and Remediation of Communication Disorders. New York: Churchill Livingstone, 1989.

2. Anderson, T., and Kearney, J.T. Effects of three resistance training programs on muscular strength and absolute and relative endurance. Research Quarterly Exercise Sport. 53:1, 1982.

3. Arndt, W. B., Elbert, M., and Shelton, R. L. Prediction of articulation improvement with therapy from early lesson sound production task scores. Journal of Speech and Hearing Research. 14:149, 1971.

4. Bardach, J., and Kelly, K. M. Reflections on research. Cleft Palate-Craniofacial Journal. 28:130, 1991.

5. Berry, M. F. Correction of cleft-palate speech by phonetic instruction. Quarterly Journal of Speech. 14:523, 1928.

6. Berry, M. F., and Eisenson, J. Speech Disorders: Principles and Practices of Therapy. New York: Appleton-Century-Crofts, 1956.

7. Blakeley, R. W. Temporary speech prosthesis as an aid in speech training. Cleft Palate Bulletin. 10:63, 1960.

8. Blakeley, R. W. The complementary use of speech prostheses and pharyngeal flaps in palatal insufficiency. Cleft Palate Journal. 1:194, 1964.

9. Blakeley, R. W. The rationale for a temporary speech prosthesis in palatal insufficiency. British Journal of Disorders of Communication. 4:134, 1969.

0. Blakeley, R. W. The Practice of Speech Pathology: A Clinical Diary. Springfield, Illinois: C. C. Thomas, 1972.

11. Blakeley, R. W., and Porter, D. R. Unexpected reduction and removal of an obturator in a patient with palate paralysis. British Journal of Disorders of Communication. 6:33, 1971.

12. Boone, D. R., and McFarlane, S. C. The Voice and Voice Therapy. (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 1988.

13. Bowers, R.W., and Fox, E.L. Sports Physiology (3rd ed.). Dubuque, Iowa: Wm. C. Brown, 1992.

14. Bradley, D. P. Congenital and Acquired Velopharyngeal Inadequacy. In K. R. Bzoch (Ed.), Communicative Disorders Related to Cleft Lip and Palate (3rd ed.). Boston, Massachusetts: College-Hill, 1989.

15. Broen, P. A., Doyle, S. S., and Bacon, C. K. The velopharyngeally inadequate child: phonologic change with intervention. Cleft Palate-Craniofacial Journal. 30:500, 1993.

16. Brooks, G.A., and Fahey, T.D. Fundamentals of Human Performance. New York: Macmillan, 1987.

17. Brunner, M., Stellzig, A., Decker, W., Komposch, G., and Wirth, G. Cognitive, evaluative processes and resultant psychosocial effects in video feedback therapy with nasopharyngoscope. Presented at the 52nd annual meeting of the American Cleft Palate-Craniofacial Association, Tampa, Florida, 1995.

18. Buck, M., and Harrington, R. Organized speech therapy for cleft palate rehabilitation. Journal of Speech and Hearing Research. 14:43, 1949.

19. Burzynski, C., and Starr, C. D. Effects of feedback filtering on nasalization and self-perception of nasality. Journal of Speech and Hearing Research. 28:585, 1985.

20. Bzoch, K. R. Variations in velopharyngeal valving: The factor of vowel changes. Cleft Palate Journal. 5:211, 1968.

21. Cole, R. Cleft lip and palate: Problems and progress. Chicago Medical School Quarterly. Fall:129, 1966.

22. Cole, R. Direct Muscle Training for the Improvement of Velopharyngeal Activity. In K. R. Bzoch (Ed.), Communicative Disorders Related to Cleft Lip and Palate. Boston: Little, Brown, 1979.

23. Colton, R. H., and Cooker, H. S. Perceived nasality in the speech of the deaf. Journal of Speech and Hearing Research. 11:553, 1968.

24. Cooper, H. K., Cooper, J. A., Mazaheri, M., and Millard, R. T. Psychological, Orthodontic, and Prosthetic Approaches in Rehabilitation of the Cleft Palate Patient. In Dental Clinics of North America. Philadelphia: W. B. Saunders, 1960.

25. Coston, G. N. Therapeutic management of speech disorders associated with velopharyngeal incompetence. Journal of Childhood Communication Disorders. 10:75, 1986.

26. Counihan, D. T., and Cullinan, W. L. Some relationships between vocal intensity and rated nasality. Cleft Palate Journal. 9:101, 1972.

27. Dalston, R. M. Photodetector assessment of velopharyngeal activity. Cleft Palate Journal. 19:1, 1982.

28. Dalston, R. M., and Keefe, M. J. The use of a microcomputer in monitoring and modifying velopharyngeal movements. Journal for Computer Users in Speech and Hearing. 3:159, 1987.

29. Dalston, R. M., Warren, D. W., and Dalston, E. T. Use of nasometry as a diagnostic tool for identifying patients with velopharyngeal impairment. Cleft Palate-Craniofacial Journal. 28:184, 1991.

30. Dalston, R. M., Warren, D. W., and Smith, L. A. The aerodynamic characteristics of speech produced by normal speakers and cleft palate speakers with adequate velopharyngeal function. Cleft Palate Journal. 27:393, 1990.

31. Daly, D. A., and Johnson, H. P. Instrumental modification of hypernasal voice quality in retarded children. Journal of Speech and Hearing Disorders. 39:500, 1974.

32. Davis, S. M., and Drichta, C. E. Biofeedback: Theory and Application to Speech Pathology. In N. J. Lass (Ed.), Speech and Language: Advances in Basic Research and Practice. New York: Academic Press, 1980. Vol. 3.

33. Diedrich, W. M., and Bangert, J. Articulation Learning. Houston, Texas: College-Hill, 1980.

34. Duxbury, J. T., and Graham, S. M. Palatal training aids for velopharyngeal insufficiency: An interdisciplinary approach. Dental Update. 12:609, 1985.

35. Dworkin, J. P. Motor Speech Disorders. A Treatment Guide. Chicago: Mosby, 1991.

36. Eisenson, J., and Ogilvie, M. Speech Correction in the Schools (4th ed.). New York: Macmillan, 1977.

37. Ellis, R. E., Flack, F. C., Curle, H. J., and Selley, W. G. A system for the assessment of nasal airflow during speech. British Journal of Disorders of Communication. 13:31, 1978.

38. Ericsson, G. Personal communication, 1994.

39. Fisher, H. B. Improving Voice and Articulation. Boston: Houghton Mifflin, 1975.

40. Fletcher, S. G. Contingencies for bioelectronic modification of nasality. Journal of Speech and Hearing Disorders. 37:329, 1972.

41. Fletcher, S. G. Diagnosing Speech Disorders from Cleft Palate. New York: Grune & Stratton, 1978.

42. Fletcher, S. G., Adams, L. E., and McCutcheon, M. J. Cleft Palate Speech Assessment Through Oral-Nasal Acoustic Measures. In K. R. Bzoch (Ed.), Communicative Disorders Related to Cleft Lip and Palate (3rd ed.). Boston: College-Hill, 1989.

43. Fletcher, S. G., Haskins, R., and Bosma, J. A movable bulb appliance to assist in palatopharyngeal closure. Journal of Speech and Hearing Disorders. 25:249, 1960.

44. Fletcher, S. G., and Higgins, J. M. Performance of children with severe to profound auditory impairment in instrumentally guided reduction of nasal resonance. Journal of Speech and Hearing Disorders. 45:181, 1980.

45. Flowers, C. R., and Morris, H. L. Oral-Pharyngeal Movements During Swallowing and Speech. Cleft Palate Journal. 10:181, 1973.

46. Froeschels, E. A contribution to pathology and therapy of dysarthria due to certain cerebral lesions. Journal of Speech and Hearing Disorders. 8:301, 1943.

47. Froeschels, E., Kastein, S., and Weiss, D. A. A method of therapy for paralytic conditions of the mechanisms of phonation, respiration, and glutination. Journal of Speech and Hearing Disorders. 20:365, 1955.

48. Garber, S., and Moller, K. T. The effects of feedback filtering on nasalization in normal and hypernasal speakers. Journal of Speech and Hearing Research. 14:271, 1979.

49. Glaser, E. R., and Shprintzen, R. J. Review of See-Scape: Instrument and Manual by C. C. Publications, Tigard, Oregon. Cleft Palate Journal. 16:213, 1979.

50. Golding-Kushner, K. J. (Working Group Coordinator). Standardization for the reporting of nasopharyngoscopy and multiview videofluoroscopy: A report from an international working group. Cleft Palate Journal. 27:337, 1990.

51. Golding-Kushner, K. J. Commentary on "A cineradiographic study of velopharyngeal movements for deviant versus nondeviant articulation" by G. Henningsson and A. Isberg. Cleft Palate-Craniofacial Journal. 28:117, 1991.

52. Greene, M. C. L. The Voice and Its Disorders (3rd ed.). New York: Pitman Publishing Corporation, 1972.

53. Hall, P. K., and Tomblin, J. B. Case study: Therapy procedures for remediation of a nasal lisp. Language, Speech, and Hearing Services in Schools. 6:29, 1975.

54. Hardin, M. A. Intervention. Clinics in Communication Disorders. 1:12, 1991.

55. Hardin, M. A., Van Demark, D. R., Morris, H. L., and Payne, M. M. Correspondence between nasalance scores and listener judgments of hypernasality and hyponasality ratings. Cleft Palate-Craniofacial Journal. 29:346, 1992.

56. Hardin, M. A., Morris, H. L., and Van Demark, D. R. A study of cleft palate speakers with marginal velopharyngeal competence. Journal of Communication Disorders. 19:461, 1986.

57. Harkins, C., Harkins, W., and Harkins, J. Principles of Cleft Palate Prosthesis. New York: Columbia University Press, 1960.

58. Harkins, C. S., and Koepp-Baker, H. Twenty-five years of cleft palate prosthesis. Journal of Speech and Hearing Disorders. 13:23, 1948.

59. Hegde, M. N. Clinical Research in Communicative Disorders. Boston: College-Hill, 1987.

60. Henningsson, G. E., and Isberg, A. M. Velopharyngeal movement patterns in patients alternating between oral and glottal articulation: A clinical and cineradiographical study. Cleft Palate Journal. 23:1, 1986.

61. Hess, D. A. Pitch, intensity, and cleft palate voice quality. Journal of Speech and Hearing Research. 2:113, 1959.

62. Hixon, T. J., and Hardy, J. C. Restricted motility of the speech articulators in cerebral palsy. Journal of Speech and Hearing Disorders. 29:294, 1964.

63. Hoch, L., Golding-Kushner, K., Siegel-Sadewitz, V. L., and Shprintzen, R. J. Speech therapy. Seminars in Speech and Language. 7:313, 1986.

64. Horii, Y., and Monroe, N. Auditory and visual feedback of nasalization using a modified accelerometric method. Journal of Speech and Hearing Research. 26:472, 1983.

65. Isberg, A., and Henningsson, G. Influence of palatal fistulas on velopharyngeal movements: A cineradiographic study. Plastic and Reconstructive Surgery. 79:525, 1987.

66. Israel, J. M., Cook, T. A., and Blakeley, R. W. The use of a temporary oral prosthesis to treat speech in velopharyngeal incompetence. Facial Plastic Surgery. 9:206, 1993.

67. Jones, D. L., and Folkins, J. W. Effect of speaking rate on judgments of disordered speech in children with cleft palate. Cleft Palate Journal. 22:246, 1985.

68. Jones, D. L., Folkins, J. W., and Morris, H. L. Speech production time and judgments of disordered nasalization in speakers with cleft palate. Journal of Speech and Hearing Research. 33:458, 1990.

69. Kanter, C. E. Four devices in the treatment of rhinolalia aperta. Journal of Speech and Hearing Disorders. 2:73, 1937.

70. Kanter, C. E. The rationale of blowing exercises for patients with repaired cleft palates. Journal of Speech Disorders. 12:281, 1947.

71. Karnell, M. P., Folkins, J. W., and Morris, H. L. Relationships between the perception of nasalization and speech movements in speakers with cleft palate. Journal of Speech and Hearing Research. 28:63, 1985.

72. Kazdin, A. E. Drawing valid inferences from case studies. In A. E. Kazdin (Ed.), Methodological issues and strategies in clinical research. Washington, D.C.: American Psychological Association, 1992.

73. Kuehn, D. P. New therapy for treating hypernasal speech using continuous positive airway pressure (CPAP). Plastic and Reconstructive Surgery. 88:959, 1991.

74. Kuehn, D. P., and Dalston, R. M. Cleft palate and studies related to velopharyngeal function. In H. Winitz (Ed.), Human communication and its disorders, a review 1988. Norwood, New Jersey: Ablex Publishing Corporation, 1988.

75. Kuehn, D. P., and Hinkle, K. Manuscript in preparation, 1995.

76. Kuehn, D.P., and Moon, J.B. Levator veli palatini muscle activity in relation to intraoral air pressure variation. Journal of Speech and Hearing Research. 37:1260, 1994.

77. Kuehn, D.K., and Moon, J.B. Levator veli palatini muscle activity in relation to intraoral air pressure variation in cleft palate subjects. Cleft Palate-Craniofacial Journal. In press.

78. Kuehn, D. P., Moon, J. B., and Folkins, J. W. Levator veli palatini muscle activity in relation to intranasal air pressure variation. Cleft Palate-Craniofacial Journal. 30:361, 1993.

79. Kuehn, D. P., and Wachtel, J. M. CPAP therapy for treating hypernasality following closed head injury. In J. A. Till, K. M. Yorkston, and D. R. Beukelman (Eds.), Motor Speech Disorders: Advances in Assessment and Treatment. Baltimore: Paul H. Brookes Publishing Co, 1994.

80. Künzel, H. J. First application of a biofeedback device for the therapy of velopharyngeal incompetence. Folia Phoniatrica. 34:92, 1982.

81. Lange, B. R., and Kipfmueller, L. J. Treating velopharyngeal inadequacy with the palatal lift concept. Plastic and Reconstructive Surgery. 43:467, 1969.

82. Leeper, L. H. Nasometer applications in a clinical training program. Nasometer Notes. Pine Brook, New Jersey: Kay Elemetrics Corp. Undated.

83. Lindblom, B. E. F., and Sundberg, J. E. F. Acoustical consequences of lip, tongue, jaw, and larynx movement. Journal of the Acoustical Society of America. 50:1166, 1971.

84. Lippmann, R. Detecting nasalization using low cost accelerometer. Journal of Speech and Hearing Research. 14:314, 1981.

85. Lotz, W. K. Personal communication, 1990.

86. Lotz, W. K., and Netsell, R. Behavioral treatment of velopharyngeal function. Paper presented at the meeting of the Nebraska Speech-Language-Hearing Association, Omaha, Nebraska, 1986.

87. Lotz, W. K., and Netsell, R. Treatment of velopharyngeal dysfunction. Paper presented at the American Speech-Language-Hearing Convention, New Orleans, Louisiana, 1987.

88. Lubit, E. C., and Larsen, R. E. The Lubit palatal exerciser: A preliminary report. Cleft Palate Journal. 6:120, 1969.

89. Lubit, E. C., and Larsen, R. E. A speech aid for velopharyngeal incompetency. Journal of Speech and Hearing Disorders. 36:61, 1971.

90. Luchsinger, R., and Arnold, G. Voice. Speech and Language: Clinical Communicology. Belmont, California: Wadsworth Publishing Co., 1965.

91. Mason, R. M., and Helmick, J. W. Residual hypernasality in repaired cleft palate. Journal of Communication Disorders. 12:431, 1979.

92. Massengill, R., and Phillips, P. P. Cleft Palate and Associated Speech Characteristics. Lincoln, Nebraska: Cliffs Notes, Inc., 1975.

93. Massengill, R., Jr., Quinn, G. W., and Pickerell, K. L. The use of a palatal stimulator to decrease velopharyngeal gap. Annals of Otology, Rhinology, and Laryngology. 80:135, 1971.

94. Massengill, R., Jr., Quinn, G. W., Pickerell, K. L., and Levinson, C. Therapeutic exercise and velopharyngeal gap. Cleft Palate Journal. 5:44, 1968.

95. Mazaheri, M., and Mazaheri, E. H. Prosthodontic aspects of palatal elevation and palatopharyngeal stimulation. Journal of Prosthetic Dentistry. 35:319, 1976.

96. McDonald, E. T., and Koepp-Baker, H. Cleft palate speech: An integration of research and clinical observation. Journal of Speech and Hearing Disorders. 16:9, 1951.

97. McGlynn, G. Dynamics of Fitness (2nd ed.). Dubuque, Iowa: Wm. C. Brown, 1990.

98. McGrath, C. O., and Anderson, M. W. Prosthetic treatment of velopharyngeal incompetence. In J. Bardach and H. L. Morris (Eds.), Multidisciplinary Management of Cleft Lip and Palate. Philadelphia: W. B. Saunders, 1990.

99. McWilliams, B. J., and Bradley, D. Ratings of velopharyngeal closure during blowing and speech. Cleft Palate Journal. 2:46, 1965.

100. McWilliams, B. J., Morris, H. L., and Shelton, R. L. Cleft Palate Speech (2nd ed.). Toronto: B. C. Decker, 1990.

101. Millard, R. T. Training for Optimal Use of the Prosthetic Speech Appliance. In W. C. Grabb, S. W. Rosenstein, and K. R. Bzoch (Eds.), Cleft Lip and Palate: Surgical, Dental, and Speech Aspects. Boston: Little, Brown, 1971.

102. Moll, K. L. A cinefluorographic study of velopharyngeal function in normals during various activities. Cleft Palate Journal. 2:112, 1965.

103. Moller, K. T., Path, M., Werth, L. J., and Christiansen, R. L. The modification of velar movement. Journal of Speech and Hearing Disorders. 38:323, 1973.

104. Moncur, J. P., and Brackett, I. P. Modifying Vocal Behavior. New York: Harper & Row, 1974.

105. Moon, J. B., and Jones, D. L. Motor control of velopharyngeal structures during vowel production. Cleft Palate-Craniofacial Journal. 28:267, 1990.

106. Morley, M. E. Cleft Palate and Speech (1st ed.). Baltimore: Williams and Wilkins, 1945.

107. Morley, M. E. Cleft Palate and Speech (7th ed.). Baltimore: Williams and Wilkins, 1970.

108. Morris, H. L. Some questions and answers about velopharyngeal dysfunction during speech. American Journal of Speech-Language Pathology. 1:26, 1992.

109. Morris, H. L., and Smith, J. K. A multiple approach for evaluating velopharyngeal competency. Journal of Speech and Hearing Disorders. 27:218, 1962.

110. Moser, H. Diagnostic and clinical procedures in rhinolalia. Journal of Speech Disorders. 7:1, 1942.

111. Mysak, E. D. Pathologies of Speech Systems. Baltimore: Williams and Wilkins, 1976.

112. Netsell, R. W., and Lotz, W. K. Evaluation and treatment of children with speech and voice disorders. Workshop presented at the University of Arizona, Tucson, Arizona, 1991.

113. Netsell, R., and Rosenbek, J. Treating the Dysarthrias. In R. Netsell, A Neurobiologic View of Speech Production and the Dysarthrias. San Diego, California: College-Hill Press, 1986.

114. Nickerson, R., Kalikow, D., and Stevens, K. Computer-aided speech training for the deaf. Journal of Speech and Hearing Disorders. 41:120, 1976.

115. Nishio, J., Yamaoka, M., Matsuya, T., and Miyazaki, T. How to exercise the velopharyngeal movement by the velopharyngeal fiberscope. Japanese Journal of Oral Surgery. 20:450, 1974. (abstract from Cleft Palate Journal. 13:310, 1976)

116. Noll, J. D. Remediation of Impaired Resonance Among Persons with Neuropathologies of Speech. In N. Lass, L. McReynolds, J. Northern, and D. Yoder (Eds.), Speech, Language, and Hearing: Pathologies of Speech and Language. Philadelphia: W. B. Saunders, 1982. Vol. III.

117. Nylen, B. O. Cleft palate and speech: A surgical study including observation on velopharyngeal closure during connected speech, using synchronized cineradiograph and sound spectrography. Acta Radiol. Suppl 203, 1961.

118. Pannbacker, M. Speech therapy for cleft palate speakers. Language, Speech, and Hearing Services in the Schools. 4:157, 1973.

119. Peterson, S. J. Velopharyngeal closure: Some important differences. Journal of Speech and Hearing Disorders. 38:89, 1973.

120. Peterson, S. J. Electrical stimulation of the soft palate. Cleft Palate Journal. 11:72, 1974.

121. Peterson, S. J. Nasal emission as a component of the misarticulation of sibilants and affricates. Journal of Speech and Hearing Disorders. 40:106, 1975.

122. Peterson-Falzone, S. J. Resonance Disorders in Structural Defects. In N. J. Lass, L. V. McReynolds, J. L. Northern, and D. E. Yoder (Eds.) Speech, Language, and Hearing. Philadelphia: W. B. Saunders, 1982. Vol. 2.

123. Peterson-Falzone, S. J. Speech Disorders Related to Craniofacial Structural Defects: Part 1. In N. J. Lass, L. V. McReynolds, J. L. Northern, and D. E. Yoder (Eds.) Handbook of Speech-Language Pathology and Audiology. Philadelphia: B. C. Decker, 1988.

124. Peterson-Falzone, S. J. Speech Disorders Related to Craniofacial Structural Defects: Part 2. In N. J. Lass, L. V. McReynolds, J. L. Northern, and D. E. Yoder (Eds.) Handbook of Speech-Language Pathology and Audiology. Philadelphia: B. C. Decker, 1988.

125. Powers, G. L. Cleft Palate. Austin, Texas: PRO-ED, 1986.

126. Powers, G. L., and Starr, C. D. The effects of muscle exercises on velopharyngeal gap and nasality. Cleft Palate Journal. 11:28, 1974.

127. Rampp, D. L., Pannbacker, M., and Kinnebrew, M. C. VPI Velopharyngeal Incompetency. A Practical Guide for Evaluation and Management. Tulsa: Modern Education Corp., 1984.

128. Redenbaugh, M. A., and Reich, A. R. Correspondence between an accelerometric nasal/voice amplitude ratio and listeners' direct magnitude estimations of hypernasality. Journal of Speech and Hearing Research. 28:273, 1985.

129. Riski, J. E. Functional Velopharyngeal Incompetence: Diagnosis and Management. In H. Winitz (Ed.), Treating Articulation Disorders: For Clinicians By Clinicians. Austin, Texas: PRO-ED, 1984.

130. Rosen, M. S., and Bzoch, K. R. The prosthetic speech appliance in rehabilitation of patients with cleft palate. Journal of the American Dental Association. 57:103, 1958.

131. Rosenbeck, J. C., and LaPointe, L. L. The Dysarthrias: Description, Diagnosis, and Treatment. In D. F. Johns (Ed.), Clinical Management of Neurogenic Communicative Disorders. Boston: Little, Brown and Co., 1978.

132. Ruscello, D. M. A selected review of palatal training procedures. Cleft Palate Journal. 19:181, 1982.

133. Ruscello, D. M. Modifying Velopharyngeal Closure Through Training Procedures. In K. R. Bzoch (Ed.), Communicative Disorders Related to Cleft Lip and Palate (3rd ed.). Boston: College-Hill, 1989.

134. Ruscello, D. M., Shuster, L. I., and Sandwisch, A. Modification of context-specific nasal emission. Journal of Speech and Hearing Research. 34:27, 1991.

135. Schmidt-Nowara, W.W. Continuous positive airway pressure for long-term treatment of sleep apnea. American Journal of Diseases of Children. 138:82, 1984.

136. Schneider, E., and Shprintzen, R. J. A survey of speech pathologists: Current trends in the management of velopharyngeal insufficiency. Cleft Palate Journal. 17:249, 1980.

137. Selley, W. G., Zananiri, M.-C., Ellis, R. E., and Flack, F. C. The effect of tongue position on division of airflow in the presence of velopharyngeal defects. British Journal of Plastic Surgery. 40:377, 1987.

138. Shelton, R. L. Therapeutic exercise and speech pathology. ASHA. 5:855, 1963.

139. Shelton, R. L. Modification of Oral-Facial Function During Speech. In ASHA Report No. 5, Speech and the Dentofacial Complex: The State of theh Art. Washington, D. C.: American Speech and Hearing Association, 1970.

140. Shelton, R. L. Oral Sensory Function in Speech Production and Remediation. In K. R. Bzoch (Ed.). Communication Disorders Related to Cleft Lip and Palate (3rd ed.). Boston, Massachusetts: College-Hill, 1989.

141. Shelton, R. L., Beaumont, K., Trier, W. C., and Furr, M. L. Videoendoscopic feedback in training velopharyngeal closure. Cleft Palate Journal. 1:6, 1978.

142. Shelton, R. L., Chisum, L., Youngstrom, K. A., Arndt, W. B., and Elbert, M. Effect of articulation therapy on palatopharyngeal closure, movement of the pharyngeal wall, and tongue posture. Cleft Palate Journal. 6:440, 1969.

143. Shelton, R. L., Hahn, E., and Morris, H. L. Diagnosis and Therapy. In D. R. Spriestersbach and D. Sherman (Eds.), Cleft Palate and Communication. New York: Academic Press, 1968.

144. Shelton, R. L., Harris, K., Sholes, G., and Dooley, P. A Study of Nonspeech Voluntary Palate Movements by Scaling and Electromyographic Techniques. In J. Bosma (Ed.), Second Symposium on Oral Sensation and Perception. Springfield, Illinois: C. C. Thomas, 1970.

145. Shelton, R. L., Knox, A. W., Elbert, M., and Johnson, T. S. Palate Awareness and Nonspeech Voluntary Palate Movement. In J. Bosma (Ed.), Second Symposium on Oral Sensation and Perception. Springfield, Illinois: C. C. Thomas, 1970.

146. Shelton, R. L., Lindquist, A. F., Arndt, W. B., Elbert, M., and Youngstrom, K. A. Effect of speech bulb reduction on movement of the posterior wall of the pharynx and posture of the tongue. Cleft Palate Journal. 8:10, 1971.

147. Shelton, R. L., Lindquist, A. F., Knox, A. W., Wright, V. L., Arndt, W. B., Elbert, M., and Youngstrom, K. A. The relationship between pharyngeal wall movements and exchangeable speech appliance sections. Cleft Palate Journal. 8:145, 1971.

148. Shelton, R. L., Paesani, A., McClelland, and Bradfield, S. S. Panendoscopic feedback in the study of voluntary velopharyngeal movements. Journal of Speech and Hearing Disorders. 40:232, 1975.

149. Sherman, D., and Hall, P. K. Nasality and precision of articulation. Perceptual and Motor Skills. 46:115, 1978.

150. Shprintzen, R. J. Nasopharyngoscopy. In K. R. Bzoch (Ed.). Communicative Disorders Related to Cleft Lip and Palate (3rd ed.). Boston, Massachusetts: College-Hill, 1989.

151. Shprintzen, R. J. Research revisited. Cleft Palate Journal. 26:148, 1989.

152. Shprintzen, R. J. Fallibility of clinical research. Cleft Palate-Craniofacial Journal. 28:136, 1991.

153. Shprintzen, R. J., Lencione, R. M., McCall, G. N., and Skolnick, M. L. A three dimensional cinefluoroscopic analysis of velopharyngeal closure during speech and nonspeech activities in normals. Cleft Palate Journal. 11:412, 1974.

154. Shprintzen, R., McCall, G. M., and Skolnick, L. A new therapeutic technique for the treatment of velopharyngeal incompetence. Journal of Speech and Hearing Disorders. 40:69, 1975.

155. Siegel-Sadewitz, V. L., and Shprintzen, R. J. Nasopharyngoscopy of the normal velopharyngeal sphincter: An experiment of biofeedback. Cleft Palate Journal. 19:194, 1982.

156. Spriestersbach, D. C., Dickson, D. R., Fraser, F. C., Horowitz, S. L., McWilliams, B. J., Paradise, J. L., and Randall, P. Clinical research in cleft lip and palate: The state of the art. Cleft Palate Journal. 10:113, 1973.

157. Starr, C. D. Treatment by Therapeutic Exercises. In J. Bardach and H. L. Morris (Eds.), Multidisciplinary Management of Cleft Lip and Palate. Philadelphia: W. B. Saunders, 1990.

158. Starr, C. D. Behavioral Approaches to Treating Velopharyngeal Closure and Nasality. In K. T. Moller and C. D. Starr (Eds.), Cleft Palate: Interdisciplinary Issues and Treatment--For Clinicians By Clinicians. Austin, Texas: PRO-ED, 1993.

159. Stengelhofen, J. Working with Cleft Palate. Great Britain: Winslow Press, 1990.

160. Stuffins, G. M. The Use of Appliances in the Treatment of Speech Problems in Cleft Palate. In J. Stengelhofen (Ed.), Cleft Palate: The Nature and Remediation of Communication Problems. New York: Churchill Livingstone, 1989.

161. Tash, E. L., Shelton, R. L., Knox, A. W., and Michel, J. F. Training voluntary pharyngeal wall movements in children with normal and inadequate velopharyngeal closure. Cleft Palate Journal. 8:277, 1971.

162. Taub, S. The Taub oral panendoscope: A new technique. Cleft Palate Journal. 3:328, 1966.

163. Thompson, A. E., and Hixon, T. J. Nasal air flow during normal speech production. Cleft Palate Journal. 16:412, 1979.

164. Thompson, R. P. J., Ferguson, J. W., and Barton, M. The role of removable orthodontic appliances in the investigation and management of patients with hypernasal speech. British Journal of Orthodontics. 12:70, 1985.

165. Tomes, L. A. Velopharyngeal Orifice Areas During Tasks Used Clinically to Stimulate Improved Velopharyngeal Closure. Dissertation, University of Arizona, Tucson, 1994.

166. Trost-Cardamone, J. E. Effects of velopharyngeal incompetence on speech. Journal of Childhood Communication Disorders. 10:31, 1986.

167. Trost-Cardamone, J. E., and Bernthal, J. E. Articulation Assessment Procedures and Treatment Decisions. In K. T. Moller and C. D. Starr (Eds.), Cleft Palate: Interdisciplinary Issues and Treatment--For Clinicians By Clinicians. Austin, Texas: PRO-ED, 1993.

168. Tucker, L. J. Articulatory Variations in Normal Speakers with Changes in Vocal Pitch and Effort. Master's Thesis, University of Iowa, Iowa City, 1963.

169. Tudor, C., and Selley, W. G. A palatal training appliance and visual aid for use in the treatment of hypernasal speech. British Journal of Disorders of Communication. 9:117, 1974.

170. Van Demark, D. R., and Hardin, M. A. Speech Therapy for the Child with Cleft Lip and Palate. In J. Bardach and H. L. Morris (Eds.), Multidisciplinary Management of Cleft Lip and Palate. Philadelphia: W. B. Saunders, 1990.

171. Van Hattum, R. J. Communication Therapy for Problems Associated with Cleft Palate. In S. Dickson (Ed.), Communication Disorders. Remedial Principles and Practices. Glenview, Illinois: Scott, Foresman & Co., 1974.

172. Van Riper, C. Speech Correction: Principles and Methods. New York: Prentice-Hall, 1947.

173. Van Riper, C. Speech Correction: Principles and Methods (4th ed.). New York: Prentice-Hall, 1963.

174. Van Riper, C. Speech Correction: Principles and Methods (6th ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 1978.

175. Van Riper, C., and Irwin, J. V. Voice and Articulation. Englewood Cliffs, New Jersey: Prentice-Hall, 1958.

176. Warren, D. W. A physiologic approach to cleft palate prosthesis. Journal of Prosthetic Dentistry. 15:770, 1965.

177. Warren, D. W. Nasal emission of air and velopharyngeal function. Cleft Palate Journal. 4:148, 1967.

178. Warren, D. W. The determination of velopharyngeal incompetency by aerodynamic and acoustical techniques. Clinics in Plastic Surgery. 2:299, 1975.

179. Warren, D. W. Perci: A method for rating palatal efficiency. Cleft Palate Journal. 16:279, 1979.

180. Warren, D. W. Assessment of aerodynamic velopharyngeal performance. In K. R. Bzoch (Ed.), Communicative Disorders Related to Cleft Lip and Palate (3rd ed.). Boston: College-Hill, 1989.

181. Warren, D. W., and Ryon, W. E. Oral port constriction, nasal resistance, and respiratory aspects of cleft palate speech. Cleft Palate Journal. 4:38, 1967.

182. Weber, J., Jr., Jobe, R. P., and Chase, R. A. Evaluation of muscle stimulation in the rehabilitation of patients with hypernasal speech. Plastic and Reconstructive Surgery. 46:173 1970.

183. Weiss, C. E. Success of an obturator reduction program. Cleft Palate Journal. 8:291, 1971.

184. Weiss, C. E. The significance of Passavant's pad in post-obturator patients. Folia Phoniatrica. 24:51, 1972.

185. Wells, C. Improving the speech of the cleft palate child. Journal of Speech Disorders. 10:162, 1945.

186. Wells, C. Practical techniques for speech training for cleft palate cases. Journal of Speech and Hearing Disorders. 13:71, 1948.

187. Wells, C. C. Cleft Palate and Its Associated Speech Disorders. St. Louis: McGraw-Hill, 1971.

188. West, R., Ansberry, M., and Carr, A. The Rehabilitation of Speech. New York: Harper, 1957.

189. Westlake, H. Speech Learning in Cleft Palate Children. In R. Lencione (Ed.) Cleft Palate Habilitation. Syracuse: Syracuse University Press, 1968.

190. Westlake, H., and Rutherford, D. Cleft palate. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.

191. Wilson, D. K. Voice Problems of Children. Baltimore, Maryland: Williams & Wilkins, 1972.

192. Wilson, D. K. Voice Problems of Children (3rd ed.). Baltimore, Maryland: Williams & Wilkins, 1987.

193. Witzel, M. A. Craniofacial anomalies. Seminars in Speech and Language. 11:145, 1990.

194. Witzel, M. A., Tobe, J., and Salyer, K. The use of nasopharyngoscopy biofeedback therapy in the correction of inconsistent velopharyngeal closure. International Journal of Pediatric Otorhinolaryngology. 15:137, 1988.

195. Witzel, M. A., Tobe, J., and Salyer, K. E. The use of videonasopharyngoscopy for biofeedback therapy in adults after pharyngeal flap surgery. Cleft Palate Journal. 26:129, 1989.

196. Wolfaardt, J. F., Wilson, F. B., Rochet, A., and McPhee, L. An appliance based approach to the management of palatopharyngeal incompetency: A clinical pilot project. Journal of Prosthetic Dentistry. 69:186, 1993.

197. Wong, L. P., and Weiss, C. E. A clinical assessment of obturator-wearing cleft palate patients. Journal of Prosthetic Dentistry. 27:632, 1972.

198. Yamaoka, M., Matsuya, T., Miyazaki, T., Nishio, J., and Ibuki, K. Visual training for velopharyngeal closure in cleft palate patients; a fiberscopic procedure (preliminary report). Journal of Maxillofacial Surgery. 11:191, 1983.

199. Ysunza, A., Pamplona, C., and Toledo, E. Change in velopharyngeal valving after speech therapy in cleft palate patients. A videonasopharyngoscopic and multi-view videofluoroscopic study. International Journal of Pediatric Otorhinolaryngology. 24:45, 1992.

200. Yules, R. B., and Chase, R. A. A training method for reduction of hypernasality in speech. Plastic and Reconstructive Surgery. 43:180, 1969.

201. Zimmerman, J. D., and Canfield, W. H. Language, Speech, and Hearing Therapy. In R. B. Stark (Ed.), Cleft Palate. A Multidiscipline Approach. New York: Harper & Row, 1968.