Hunter

NCVS

National Center for Voice and Speech

University of Iowa • Denver Center for the Performing Arts
University of Wisconsin-Madison • University of Utah

Status and Progress Report

Volume 2/April 1992

NCVS Status and Progress Report
Volume 2/April 1992

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

Editorial and Distribution Information

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

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

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

Primary Sponsorship

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

Other Sponsorship

The University of Iowa
Department of Speech Pathology and Audiology
Department of Otolaryngology - Head and Neck Surgery

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

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

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

Memphis State University
Department of Audiology and Speech Pathology

The University of Illinois
Department of Speech and Hearing Science

NCVS Personnel

Administration

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

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

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

Investigators, Affiliates and Support Staff

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

Doctoral Students

Steve Austin, M.M.
Julie Barkmeier, M.S.
Todd Brennan, M.S.
David Druker, B.S.
Ann Fennell, M.S.
Mark Leddy, M.S.
Bruce Poburka, M.S.
Annie Ramos, M.S.
Alice Smith, M.A.
Brad Story, B.S.
Kenneth Tom, M.A.

Postdoctoral Students

Linnea Peterson, M.D.
David Berry, Ph.D.

Visiting Scholars

Hanspeter Herzel, Ph.D., Germany
Meijin Nakayama, M.D., Japan

Advisory Board

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

Contents

Forward

This is the second official Progress and Status Report from the National Center for Voice and Speech. We are now approximately one and a half years old and well underway in all areas of investigation. The Central Office at The University of Iowa is increasingly busy with more and more people inquiring about problems in voice and speech. Several new connections have been made with other organizations to accelerate dissemination of information and to assist with continuing education. For example, we have made formal ties with the Voice Foundation and the VoiceCare Network. We are also supporting some of the workshops given by Dr. Thomas Cleveland world-wide.

Last January, we had a very successful Center Conference. Many new ideas were presented and progress was reported in all phases of our work. Some of the papers presented at the conference appear in this volume. There was heavy emphasis on clinical investigations, which is satisfying in view of our mission to bring theory and practice into closer harmony. I was particularly impressed with the work being done by students and affiliates of the Center. Most research projects seem to be moving well, and some investigators are finding ways of generating spin-off ideas that hopefully will be supported in the future. We are most appreciative of the financial assistance granted us by the National Institute on Deafness and Other Communication Disorders.

Ingo R. Titze, Director

April 1992

Part I

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 2
April 1992, 1-15

A Quantitative Histologic Study of the Normal Human Adult Soft Palate

Sandra L. Ettema, B.S.
David P. Kuehn, Ph.D.

Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign

Abstract

This study was designed to re-examine the same soft palates studied qualitatively in a previous investigation (Kuehn and Kahane, 1990), using a true-color image analysis system. The quantified measures of the areas of specific tissue types are reported. The results indicate that: (1) tendinous tissue is prominent anteriorly and comprises about 10% of total tissue in that region, (2) the relative proportion of glandular and connective tissue is fairly uniform across the length of the soft palate averaging 22% and 36% respectively, (3) muscle tissue shows a pattern of increasing then decreasing amount from anterior to posterior with a maximum proportion of 23% in the midportion of the soft palate, (4) adipose tissue comprises 22% of total tissue area in the anterior segment and 17% of total tissue area in the posterior segment of the soft palate, (5) other tissue, primarily epithelium and vascular tissue, was found to be fairly constant anteriorly and increased in relative proportion to almost 30% of the total tissue area at the uvular base. The functional implications of these findings are discussed.

To learn about pathologic mechanisms of the velopharyngeal region, one must first have a basic understanding of the intricacies involved in the normal mechanism. Gross dissections have provided only general information about the anatomy of velopharyngeal structures (Blakeway, 1913; Bosma, 1953; Bosma and Fletcher, 1962; Browne, 1932; Fara and Dvorak, 1970; Harrington, 1944; Kriens, 1975; McMyn, 1940; Negus, 1943; Ruding, 1964; Strong, 1949; Townshend, 1940; Whillis, 1930; Wood-Jones, 1940;). Such information is important, but to establish the foundation for the construction of useful biomechanical models, more precise data involving velopharyngeal structures need to be gathered.

Although a few histologic studies of the velopharynx have been published (Azzam and Kuehn, 1977; Boorman and Sommerlad, 1985a, 1985b; Dickson, 1972; Dickson, 1975; Dickson and

Dickson, 1972; Kuehn and Azzam, 1978; Langdon and Klueber, 1978; Latham et al., 1980), Kuehn and Kahane (1990) used a more systematic, compartmentalized approach than previously employed in studying the components of the normal human adult velum. Soft palates were removed in their entirety from ten adult human cadavers, four males and six females. The extracted palates were bisected longitudinally and measured in length from the most anterior to the most posterior point. The bisected palates then were divided, according to each of their lengths, into ten blocks of equal thickness from anterior to posterior. This was done to normalize soft palate section intervals across subjects. There were no qualitative differences in gross anatomy between right and left halves or between male and female subjects. However, as expected, the male soft palates, at least in one dimension, were somewhat larger than those of the females, with thickness of the uvula being significantly larger than that of the females.

Only the face ("top" portion) of each of the ten blocks was sectioned and stained with the use of three different stains. This resulted in a sample of three hundred histologic slides that were examined, and major anatomic characteristics were noted. The typical tissue composition of an adult human soft palate was observed to consist of the following layers: 1) the oral aspect which is mostly glandular tissue with adipose tissue located somewhat laterally; 2) the two middle layers consisting of transverse muscular tissue of the levator veli palatini possibly with palatopharyngeal muscle fibers interspersed and the longitudinal musculus uvulae fibers overlying those of the levator muscle; 3) a superoanterior layer consisting of the tensor veli palatini tendon; and 4) a posteroinferior layer consisting mostly of glandular tissue.

The results of the Kuehn and Kahane (1990) study deal primarily with the structure of the velum between the lateral pharyngeal walls and show trends of structural consistency across subjects. The results provide useful qualitative information about the microscopic structure of the velum, but there are many complex aspects of the anatomy of the velum that are important in understanding its mechanization. First, as noted by Kuehn and Kahane, the soft palate is not a self-contained structure. The lateral boundaries of the velum are difficult to define although it is obvious that structures lateral to these boundaries are important in the functioning of the velum. Secondly, although the Kuehn and Kahane study does show some structural consistency across subjects, it has not been determined whether a "generic" velum will function adequately with respect to the central tendency of variability across subjects. Such between-subject variability may complicate the construction of a biomechanical model.

To provide the greatest use in construction of biomechanical models, quantified information concerning spatial distribution of tissues and characteristics of fiber density and angulation must be obtained. Recent advances in image analysis systems have made quantification of anatomic structures more practical thus enabling the collection of larger sets of data than heretofore possible. The purpose of the current investigation was to re-examine the soft palates of the same ten subjects reported in the Kuehn and Kahane (1990) study, but with the use of a true-color image analysis system thereby quantifying the areas of the specific tissue types that were reported qualitatively in the previous study.

Method

Subjects

The specimens used in this investigation were taken from ten normal human adult cadavers, four males and six females. The soft palate was removed in its entirety from each cadaver. One-

half of each palate was divided into ten blocks of equal anterior-posterior thickness. The face of each block was stained with a Mallory trichrome stain, providing a sample of one hundred slides (Kuehn and Kahane, 1990). The same slides used in the Kuehn and Kahane study were used in the present investigation.

Image Analysis System

The equipment used to collect and analyze data include the following: a PC-Based (IBM AT) true-color image analysis system (American Innovision, Videometric 150), Electrohome 38-DO51MA-YU graphics screen, single chip-Cohu RGB CCD color camera, Picon 1 :3.5 f = 100 mm macro lens, light box, and Balplan microscope. Figure 1 depicts the command screen display for the Videometric 150 (V150) image analysis system. The V150 consists of modes that allow one to organize and collect different types of data, such as area measures, and the utilities are the commands that affect these modes (American Innovision, Inc., 1988). Utility commands cause the V150 system to perform an action, such as clearing the screen, calibrating, or setting threshold levels.

The Grains mode is of particular interest in this study because it was used for measuring areas of the various tissue types. This mode is not available for monochrome image analysis systems and, therefore, represents a major technologic advancement. The Threshold utility command is used in conjunction with the Grains mode. The Threshold utility command allows interactive setting of thresholds in luma (brightness) and specific colors (hue and saturation). The Threshold command is used to "highlight" the area of interest and, thus, separate it from other areas associated with different tissue types. This feature is especially advantageous in working with a histologic stain, such as trichrome, that yields several different colors thereby facilitating differentiation among discrete tissue types. In contrast, a monochrome stain or monochrome image analysis system does not allow such a differentiated analysis.

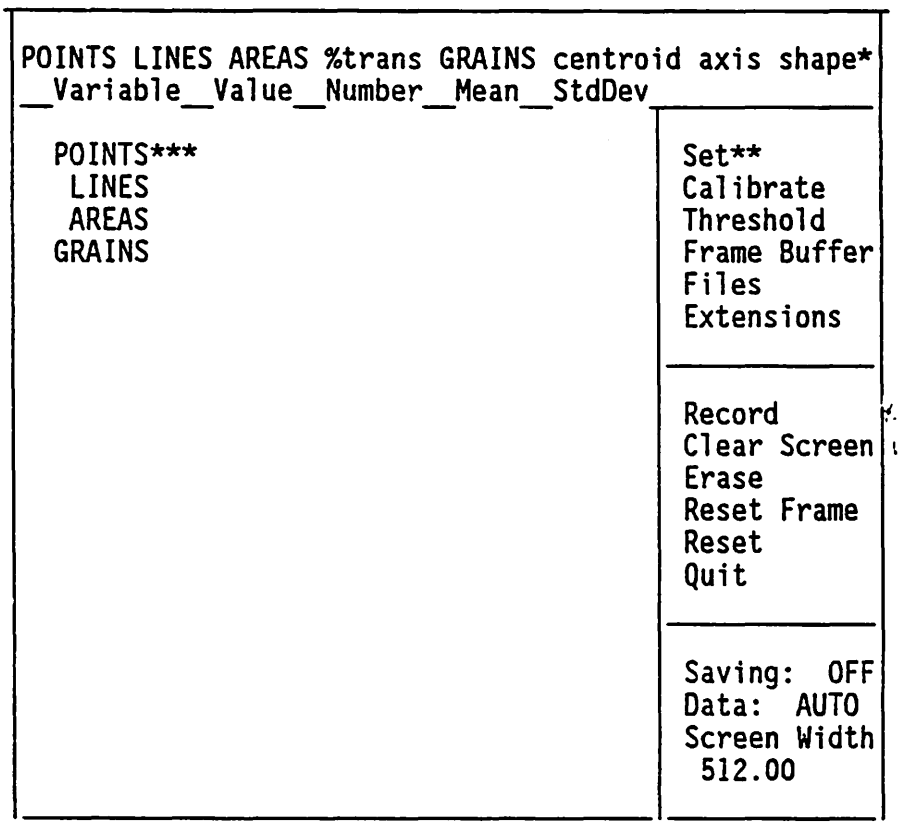

*Figure 1. PC command screen display format. * Data collection modes (top row); upper case indicates modes that are activated for a given application. ** Utility commands (right column). *** Data area; data are displayed in center of screen, which is blank in this example.*

In Grains mode, the total number of pixels highlighted on the basis of their specific color are counted within a designated area. The pixel count is converted to an area measure (mm^2 in this study) in relation to a predetermined calibration value. Nonhighlighted pixels in the field of view, such as those associated with different tissue types, artifacts, border, etc. are not included in the highlighted area measure. A major advantage of this method is that the total area of noncontiguous structures of the same composition can be determined quickly and accurately without resorting to older laborious techniques such as cutting and weighing portions of photographs.

TABLE 1. Single J-Unit values for each subject.

SUBJECTS	J-Unit standardized 1/2Lv original/3.72	single value*
M1	18.5/3.72	5.0 mm
M2	20.3/3.72	5.5 mm
M3	20.5/3.72	5.5 mm
M4	21.8/3.72	6.0 mm
F1	18.3/3.72	5.0 mm
F2	15.5/3.72	4.0 mm
F3	23.1/3.72	6.0 mm
F4	19.4/3.72	5.0 mm
F5	17.2/3.72	4.5 mm
F6	17.1/3.72	4.5 mm

*The J-unit values were doubled, and that value was used as the predetermined standardized lateral boundary. The lateral boundary thus determined corresponds approximately to the area where the lateral region of the velum blends with the lateral pharyngeal wall.

Lateral Boundary Measure

For quantified measurements to be obtained, it is obviously necessary to define the boundaries of the velum. Two of the boundaries are naturally defined, the nasal surface and the oral surface. A third boundary was predetermined in relation to the methodology used in the previous study by Kuehn and Kahane (1990), that is, the medial edge that resulted from the longitudinal bisection of the velum. The fourth boundary, specifically the lateral margin, is problematic. There is no true anatomic boundary distinguishing the soft palate from the lateral pharyngeal wall because the velar contents blend with the pharynx laterally. Therefore, an operationally defined lateral boundary was calculated for each individual velum (Jesiolowski and Kuehn, 1987).

Each of the ten specimens, used in the Kuehn and Kahane (1990) study, were standardized by the use of the length of the velum (Lv) and width of the uvular base (Wu) expressed as a ratio. The Lv/Wu ratio was calculated for each of the subjects, and the resultant values then were averaged to obtain a Lv/Wu mean value, which was equal to 3.72. The equation was adjusted to 1/2Lv / 1/2Wu average because the prepared cross sections were previously bisected sagittally. This still was equal to 3.72. The adjustment was done to allow for proportional standardization numbers which would be in closer proximity to the working dimensions. The standardization value (J-Unit) was determined for each subject by recalculating 1/2Lv / 1/2Wu using half the original individual velum length values and 3.72 (1/2 Lv original/J = 3.72). Each J-Unit value, therefore, is a variable distance in millimeters that equals half the average width of the uvular base of a soft palate with a certain length (Lv). The ratios and J-Unit values for each subject are shown in Table 1. These calculations served as the means for quantitatively establishing the lateral limits of each subject by applying two J-Units to every section of each of the ten bisected soft palates (Figure 2). For example, the lateral limit from the medial edge for subject Male-1 (M1) is 2 x 5.0 mm = 10.0 mm. Because the velum had been bisected, the two J-Units represent only one-half the velum, and four J-Units would represent the total left-to-right dimension of the velum. For example, as operationally defined in this study, the left-to-right velar dimension for Subject M1 is 4 x 5.0 mm = 20 mm.

Procedure

Each subject's data were collected individually for each section. One data file was created for each subject and subsequent buffer files were established for each section. Each slide was placed on the light box and the macro lens was adjusted to enable the investigator to view an entire velar section on a single screen. We chose this method instead of imaging through a microscope because microscopic imaging displays only a small portion of a velar section necessitating a piecemeal reconstuction of a given section.

· The first slide (i.e. section) for a particular subject was viewed and saved in a buffer file. Each slide's buffer file also contained an image of a transparent ruler in mm. The image of interest that appeared on the Electrohome screen then was calibrated, by use of the imaged ruler, to its magnification as viewed through the CCD color camera (Figure 2). Next, the values of area of tissue for T-tendinous, G-Glandular, C-Connective, M-Muscular, A-Adipose, and O-Other tissue components were calculated in that order for each section of a subject. "Other" tissue consisted primarily of epithelial and vascular tissue.

The Threshold was set and reset for each tissue type (i.e. specific color). Then the area containing that particular tissue and encompassed within two J-Units was traced and calculated using

the Grains mode. The investigator did not always include the most superficial tissue portions in the nasal or oral surface areas in the most lateral regions of the velum. Instead, lines were drawn perpendicular to the medial aspect of the velum. These lines were approximately tangent to the nasal and oral surfaces. Only structures that were contained within the perimeter bounded by those lines and the two J-Unit width were included in the analysis. For example, in reference to Figure 2, only those tissue areas encompassed in the stippled zone were measured. The investigator, when reaching the sections in or near the uvula, used the microscope as an aid to ensure correct identification of tissue types. This was done because the tissue samples were quite small in the uvular blocks.

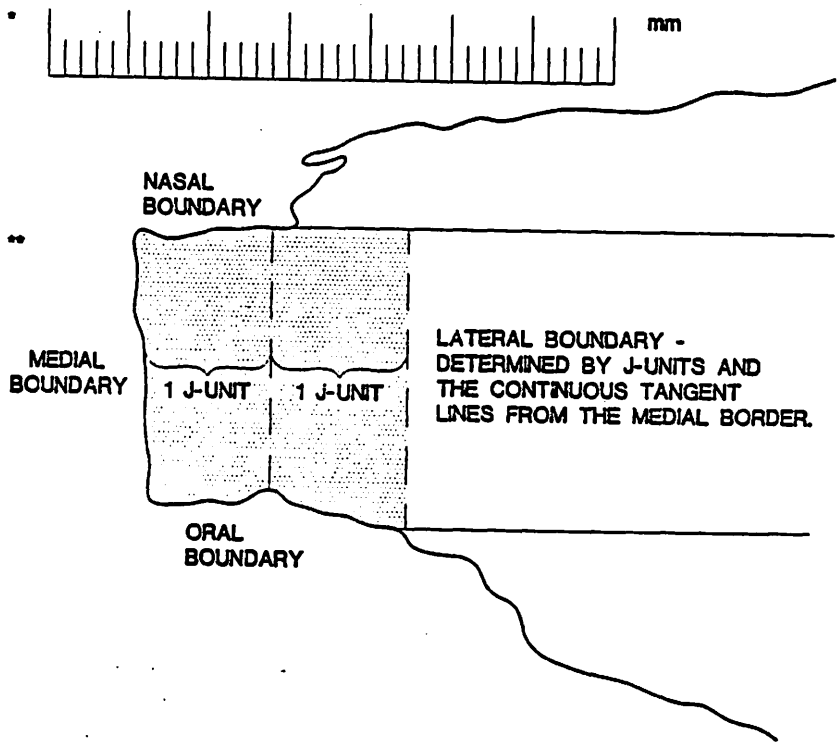

*Figure 2. Method used to determine calibration measures, J-Units, and lateral boundaries. * Imaged ruler used for calibration and determining J-Unit measures. ** Depiction of a section (i.e. slide) of the soft palate with consideration of the lateral boundary measurement; stippled area = total area measured.*

Data were recorded in a notebook and in the software package of the V150 program. The notebook records also included an estimate of total area for each specimen, which used the two J-Unit value and the dimension of the specimen from the nasal to oral surface as the area variables. These estimates were compared to the area calculations of each section (Table 2) to ensure that the calculated area was within a reasonable range thereby avoiding obvious artifact. Such a potential disparity was never encountered.

Data Analysis

Subject Female-2 (F2) was not used in the analysis of the data because this subject possessed an unusually short and thick velum (see Kuehn and Kahane, 1990, Table 1). The ten blocks of tissue, from anterior to posterior, were not accurately sectioned for Subject F2 because of the odd structure of that subject's velum. Therefore, the data for Subject F2 were not used at all, which left a total of nine subjects for the data analysis. Furthermore, averages were taken only on eight subjects for Section 10 due to the extremely small girth of the uvula in Subject F6.

Section numbers correspond to percentages along velar length (i.e. Section 1 = 0%, Section 2 = 10%, ..., Section 10 = 90%). Results were compiled for the average area, in percentages, normalized across subjects for each section and tissue type. This was obtained by first calculating the total tissue area (TA) encompassing all the various tissue types for each section of each subject (Table 2). Each specific tissue area for each section was divided by TA to get a percentage of tissue for that section and type of tissue. Calculations for average percentages across subjects for each section and tissue type then were calculated. Average area in mm^2 using absolute values and average area in mm^2 using relative values normalized by J-Unit (lateral boundary) were calculated. Relative values of area in mm^2 adjusted each subject's data in relation to the mean J-Unit value calculated, thereby standardizing the lateral boundaries of the velum. Because of the similarity in J-Unit values across subjects (Table 1), the absolute and relative (adjusted) value calculations also were very similar. For example, for glandular tissue in Section 1, absolute and relative values were 26.6% and 26.3%, respectively. Only the relative measures are reported in the Results section.

TABLE 2. Average absolute size of velar areas measured (two J-Unit segments) across subjects, with standard deviations.

	MEAN	S.D.
Section 1	90.3 mm^2	20.6
Section 2	94.4 mm^2	32.0
Section 3	110.2 mm^2	43.8
Section 4	115.9 mm^2	46.4
Section 5	116.7 mm^2	37.9
Section 6	90.2 mm^2	36.8
Section 7	68.9 mm^2	27.2
Section 8	55.0 mm^2	25.3
Section 9	30.5 mm^2	7.2
Section 10	21.0 mm^2	8.4

Reliability

Both intra-observer and inter-observer reliability data were gathered for determination of quantified tissue types in the soft palate by the two investigators. This was done by taking a random sample for each section and tissue type and recalculating the area using the same procedure as discussed above. Investigator 1 (the first author) used a random sample of n = 18 and Investigator 2 (the second author) used a random sample of n = 9 for each tissue type.

The reliability data were analyzed with the aid of a statistical software package, the System for Statistics (SYSTAT) (Wilkinson, 1990). Table 3 contains a summary of the intra- and inter-observer Pearson correlations of specific tissue types for Sections 1-9.

TABLE 3. Pearson product-moment intra-observer and inter-observer correlations of quantified tissue types for Sections 1-9. "Other" is primarily epithelial and vascular tissue.

Intra-observer correlations (Investigator 1)

Tendinous tissue	r = .94
Glandular tissue	r = .78
Connective tissue	r = .80
Muscular tissue	r = .83
Adipose tissue	r = .89
Other tissue	r = .65

Inter-observer correlations (Investigator 1 vs. Investigator 2)

Tendinous tissue	r = .87
Glandular tissue	r = .92
Connective tissue	r = .85
Muscular tissue	r = .90
Adipose tissue	r = .50
Other tissue	r = .58

Although all of these correlations are judged to be acceptable for the purposes of this investigation, those for "other" tissue within and between observers are only moderately high. One possible factor involved in the lesser reliability of measurement of the "other" tissue was the relatively small size and limited range of area of these tissues.

Although intra-observer reliabilty was high for adipose tissue, inter-observer reliability was only moderate. The lesser reliability between observers was probably related to the fact that fatty tissue is disolved during embedding of tissues and, therefore, not colored in the staining process. The fat cells appear as blank structures and are not as distinctive as the colored tissue types.

Results

Results compiled for average areas, in percentages, within the lateral boundaries operationally defined as two J-Units and normalized across subjects for each section and tissue type show definite patterns of tissue distribution and quantity across the velum. The results for the tissue types measured in this study are summarized in Figures 3-8. Detailed numerical information from which the data are summarized are available in a separate report (Ettema, 1991).

Figure 3 shows the results for the tendinous tissue. It is clear that the tendinous tissue is relatively prominent (about 10% of total tissue) anteriorly, and diminishes rather abruptly by Section 5 in all subjects. Qualitative observations from Kuehn and Kahane (1990) show that the tensor veli palatini tendon is no longer present in Section 4, but with the present image analysis technique, a minute amount of tendinous tissue was still detected in Section 5.

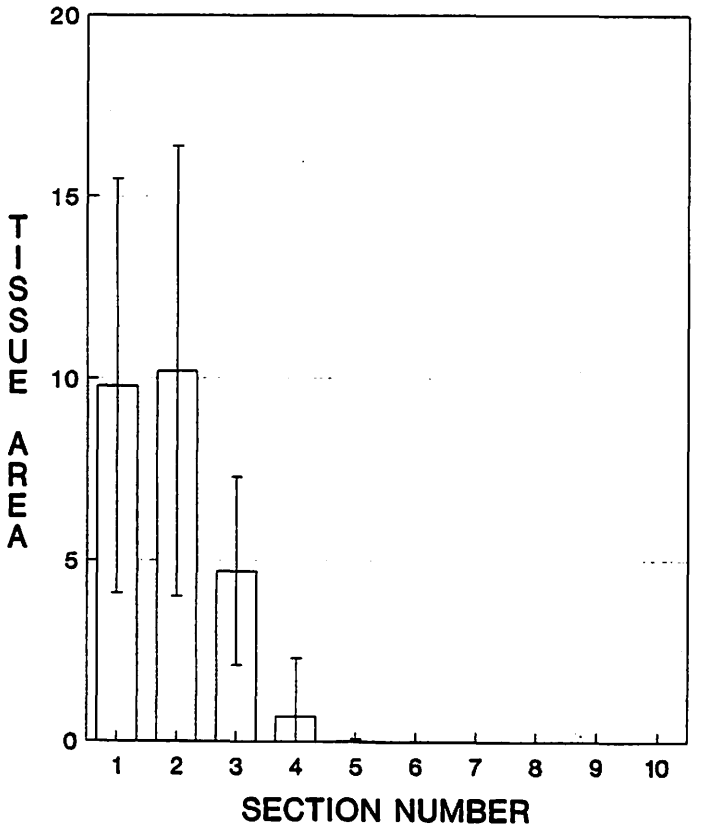

Figure 3. Tendinous tissue area averages in relative percentages with standard deviations. Section numbers correspond to percentages along velar length (i.e. Section 1 = 0%, Section 2 = 10%, ..., Section 10 = 90%).

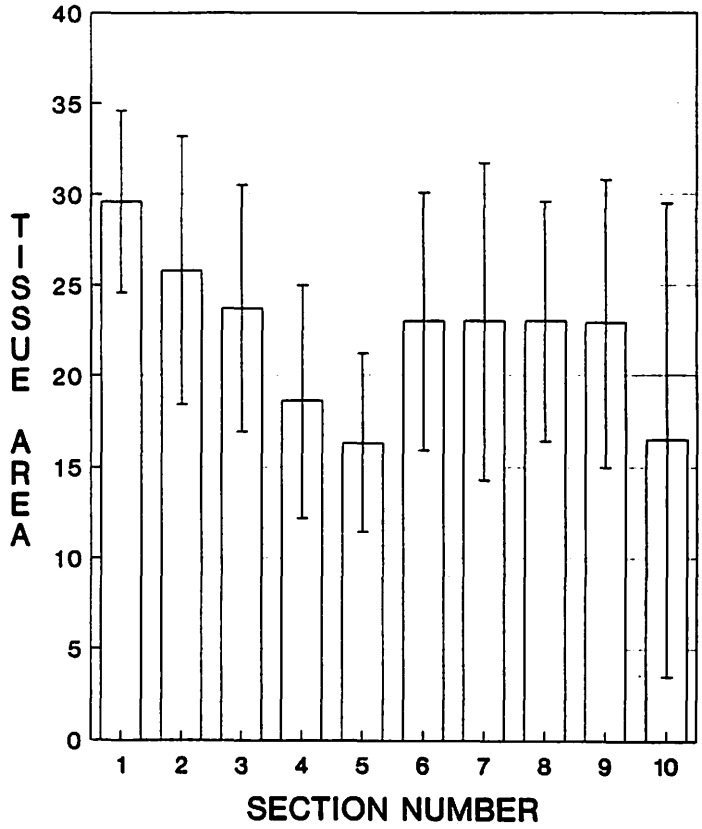

Figure 4. Glandular tissue area averages in relative percentages with standard deviations.

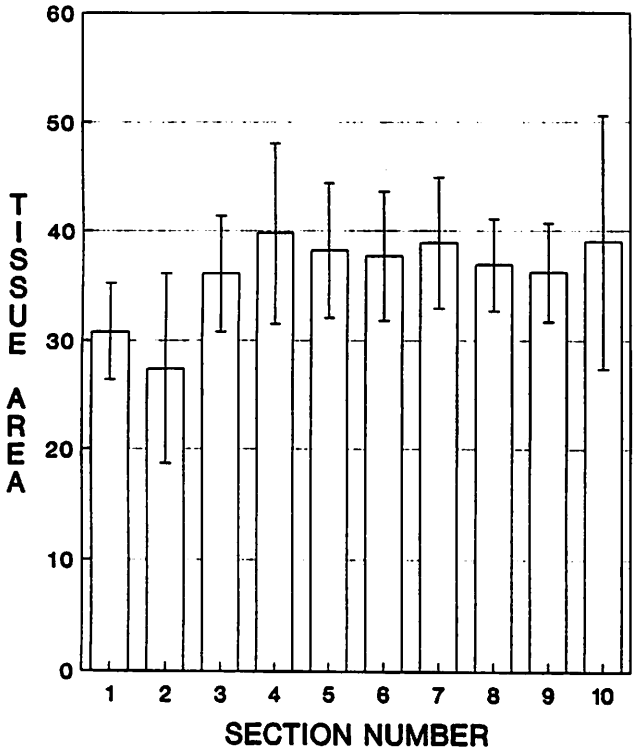

Figure 5. Connective tissue area averages in relative percentages with standard deviations.

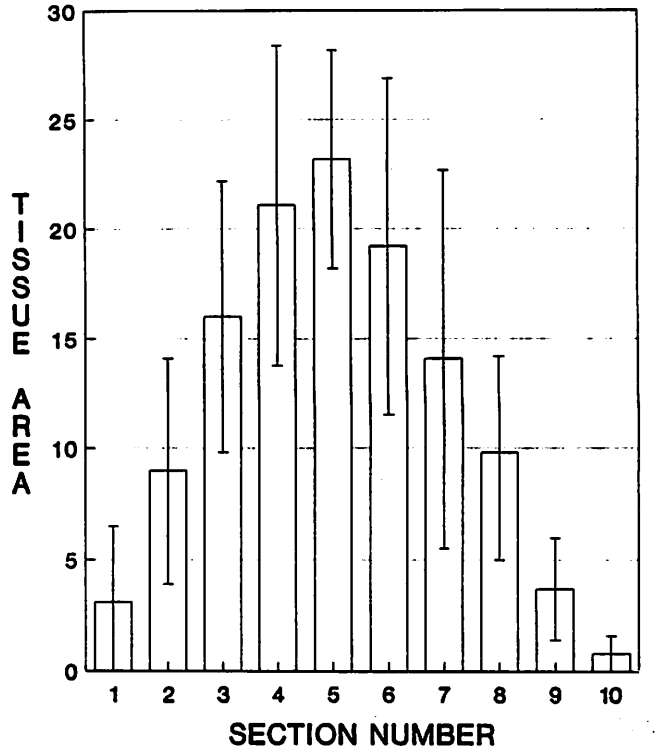

Figure 6. Muscle tissue area averages in relative percentages with standard deviations.

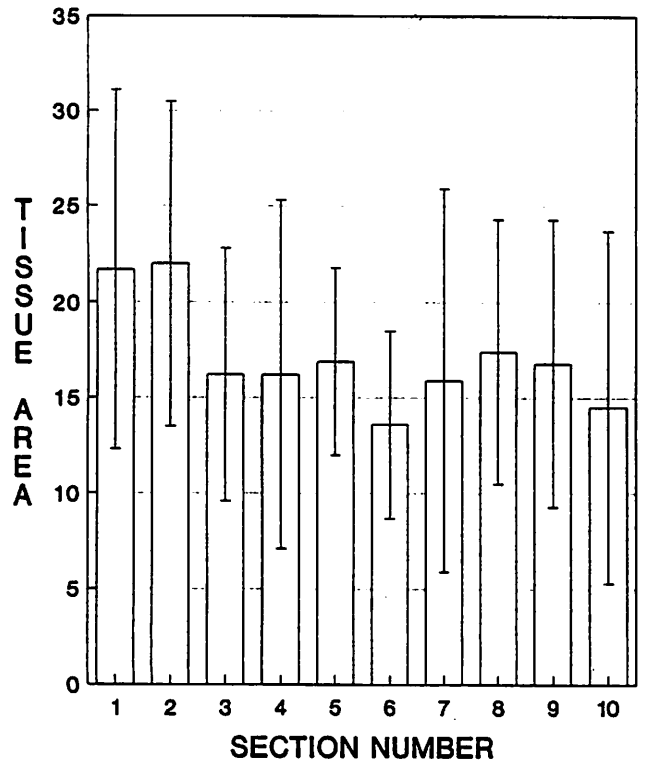

Figure 7. Adipose tissue area averages in relative percentages with standard deviations.

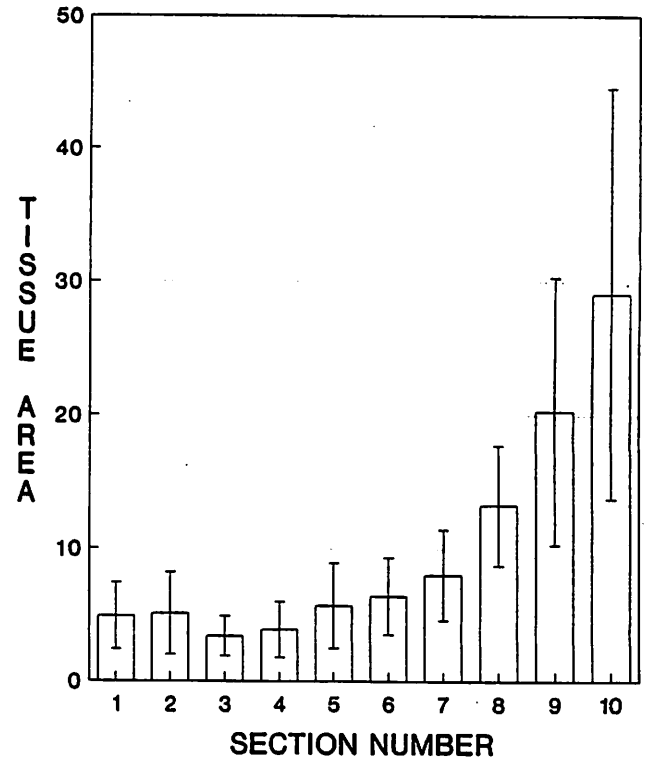

Figure 8. "Other" (primarily epithelial and vascular) tissue area averages in relative percentages with standard deviations.

Both the glandular and connective tissues (Figures 4 and 5) are quantitatively fairly uniform (mean equals 22% and 36% respectively of total tissue) in their concentration along the length of the velum. This finding is consistent with the qualitative results of Kuehn and Kahane (1990). In relation to all other tissue components in the velum, both glandular and connective tissues are present in the largest quantities. The glandular tissue is concentrated in the oral aspect of the velum anteriorly but is distributed diffusely in the more posterior regions. There are prominent areas of connective tissue beneath both the oral and nasal epithelium and less obvious areas intertwined throughout all sections in all directions. Connective tissue tended to be the least concentrated (most dispersed) of all the tissue components.

Figure 6 shows the results for muscle tissue. There is definitely a pattern of increasing then decreasing muscle quantity from anterior to posterior, with the greatest amount (about 23% of total tissue area) found in Section 5. Decreasing muscle quantity continues posteriorly until it is virtually absent in the final section of the uvular base (Section 10).

The amount of adipose tissue (Figure 7) remained at a fairly constant level both anteriorly and posteriorly but with slightly different amounts between the two regions. More adipose tissue was found anteriorly (about 22% of total tissue are for Sections 1 and 2) and then leveled off at a consistent level of about 17% of total tissue area more posteriorly (Sections 3-10).

Other tissue (Figure 8), which consists primarily of epithelial and vascular tissue, remained fairly constant in proportion throughout the first six sections and then increased to almost 30% of the total tissue at the uvular base. This marked increase in relative area was attributable mainly to greater vasculature in and near the uvula. This finding is consistent with that of the Kuehn and Kahane (1990) study.

Discussion

The main purpose of this study was to obtain quantitative information about the different tissue components in the normal human adult soft palate. This is a natural sequel to qualitative investigations of velopharyngeal anatomy. Quantitative anatomic research has been rendered more practical recently by newer technology, specifically, image analysis systems such as that used in the current study. As in any scientific endeavor, numerical data for quantifying anatomic structures are far more valuable heuristically than qualitative information.

In general, the numerical data presented in Figures 3-8 are consistent with the qualitative information of Kuehn and Kahane (1990) with regard to the proportionality and distribution of the various components in the velum. Figures 3-8 provide information not only about mean values of tissue types along the length of the velum but also, judging from the fairly large standard deviations, an indication that there is a considerable degree of variability in those tissue components across individuals.

The tensor veli palatini tendon is confined mainly to the anterior region of the velum. As the tensor tendon reduces in relative size moving form the most anterior region to the midportion, the amount of muscle tissue increases in proportion to a maximum at Section 5 (40% of the velar length). It is in the midportion of the velum, Sections 4, 5, and 6 where the musculus uvulae is in its most cohesive form (Azzam and Kuehn; Kuehn and Kahane, 1990) overlying the levator veli palatini sling. Also, fibers from the palatopharyngeus muscle add to the muscle mass in this region. These muscle masses then become greatly diminished toward and within the uvula where their importance obviously is not critical (Kuehn, Folkins, and Linville, 1988).

A marked and progressive increase in "other" tissue was observed beginning at about 60% of velar length (Section 7) reaching a maximum relative amount of nearly one-third the cross-sectional area of the uvula. Because the thickness of the nasal and oral epithelium remained fairly constant along the velar length, much of the increase in "other" tissue consisted of vascular tissue. This suggests that the uvula proper is a highly vascular structure.

Whereas clear patterns of change were measured along velar length for tendinous, muscular, and "other" tissue, glandular, connective, and adipose tissue tended to be more constant in relative concentration along velar length. However, these tissues were somewhat different in their density and location within the velum. The glandular tissue tended to be located nearer the oral surface, especially anteriorly, becoming more diffuse posteriorly. Adipose tissue tended to be less concentrated than the glandular tissue and connective tissue tended to be the most diffuse of all the tissue types.

The amount of connective tissue within the velum has received very little attention in previous investigations and deserves special discussion here. It was our initial impression in obtaining the results that the relative proportion of connective tissue, ranging from 27% to 40% (Figure 5), appeared to be somewhat high. To our knowledge, a relatively high proportion of connective tissue within the velum had not been mentioned in the literature previously. Because all of our specimens were taken from older individuals, we reasoned that the relatively high proportion of connective tissue might be due, in part, to the normal aging process. Although little has been written on the aging of the velar mechanism, Bucciante and Luria (1934) did report an increase in interstitial tissue (collagenous and elastic fibers) between muscle fibers in the musculus uvulae of individuals after the seventh decade. Therefore, it is possible that the soft palate of the older individuals studied in the current investigation may contain a greater relative proportion of connective tissue compared to younger individuals.

Another possible factor involved in the relatively high proportion of connective tissue observed in this study relates to the sensitivity of the true-color image analysis system. Using this system and the trichrome stain, connective tissue appeared as a light blue matrix interspersed among the other tissue structures. The total area of such a matrix would be extremely difficult to measure using other procedures and could easily have been underestimated in previous investigations. The fact that the intra- and inter-investigator measurement correlations for connective tissue were high (Table 3) gives us confidence that the measures are valid indicators of connective tissue amount.

The quantitative data obtained in this study provides a step forward in establishing the foundation for a functional biomechanical model of the human velopharyngeal mechanism. As an example of such a model, the velum could be considered as a "muscular-hydrostat" (Kier and Smith, 1985; Smith and Kier, 1989) with very definite muscular as well as nonmuscular layers of tissue with a significant matrix, now quantified in the present study, of noncontractile tissue interspersed. Obviously, much work remains in developing such a model but it is clear that significant progress cannot be made without further quantified information provided such as that of the current study.

In addition to its usefulness in relation to biomechanical models, a thorough understanding of the normal distribution and quantity of tissue types is likely to assume an increasingly more important role in reconstructive surgical procedures. This is so because of the increased availability of technologically more advanced imaging systems, particularly magnetic resonance imaging (MRI). Using MRI, it may be practical eventually to image palates of babies born with cleft palate before palatal surgery and emulate the normal mechanism during surgery in a more quantified fashion with regard to tissue distribution and orientation.

Acknowledgment

This study was supported in part by NIH Grant DC00976 from the National Institute on Deafness and Other Communication Disorders.

References

American Innovision, Inc. (1988). The Videometric 150: User's Manual. San Diego, CA: American Innovision, Inc.

Azzam, N.A. & Kuehn, D.P. (1977). The morphology of the musculus uvulae. Cleft Palate Journal, 14, 78-87.

Blakeway, H. (1913). Investigations of the anatomy of the palate. Journal of Anatomy and Physiology, 48, 409-416.

Boorman, J.G. & Sommerlad, B.C. (1985a). Levator palati and palatal dimples: their anatomy, relationship and clinical significance. British Journal of Plastic Surgery, 38, 326-332.

Boorman, J.G. & Sommerlad, B.C. (1985b). Musculus uvulae and levator palati: their anatomical and functional relationship in velopharyngeal closure. British Journal of Plastic Surgery, 38, 333-338.

Bosma, J.F. (1953). A correlated study of the anatomy and motor activity of the upper pharynx by cadaver dissection and cinematic study of patients after maxillofacial surgery. Annals of Otolaryngology/Rhinology/Laryngology, 62, 51-72.

Bosma, J.F. & Fletcher, S.G. (1962). The upper pharynx. Part II. Physiology. Annals of Otolaryngology/Rhinology/Laryngology, 71, 134-157.

Browne, D. (1932). The operation for cleft palate. British Journal of Surgery, 20, 7-25.

Dickson, D.R. (1972). Normal and cleft palate anatomy. Cleft Palate Journal, 9, 280-293.

Dickson, D.R. (1975). Anatomy of the normal velopharyngeal mechanism. Clinics in Plastic Surgery, 2, 235-248.

Dickson, D.R. & Dickson, W.M. (1972). Velopharyngeal anatomy. Journal of Speech and Hearing Research, 15, 372-381.

Ettema, S.L. (1991). A quantitative histologic study of the normal human adult soft palate. B.S. Thesis, University of Illinois at Urbana-Champaign.

Fara, M. & Dvorak, J. (1970). Abnormal anatomy of the muscles of the palatopharyngeal closure in cleft palates. Plastic and Reconstructive Surgery, 46, 488-497.

Harrington, R. (1944). A study of the mechanism of velopharyngeal closure. Journal of Speech and Hearing Disorders, 9, 325-245.

Jesiolowski, K.A. & Kuehn, D.P. (1987). A quantitative histological study of the nonpathologic adult human soft palate. Unpublished.

Kier, W.M. & Smith, K.K. (1985). Tongues, tentacles, and trunks: The biomechanics of movement in muscular-hydrostats. Zoological Journal of the Linnean Society, 83, 307-324.

Kriens, O. (1975). Anatomy of the velopharyngeal area in cleft palate. Clinics in Plastic Surgery, 2, 261-283.

Kuehn, D.P. & Azzam, N.A. (1978). Anatomical characteristics of palatoglossus and the anterior faucial pillar. Cleft Palate Journal, 19, 349-359.

Kuehn, D.P., Folkins, J.W., & Linville, R.N. (1988). An electromyographic study of the musculus uvulae. Cleft Palate Journal, 25, 348-355.

Kuehn, D.P. & Kahane, J.C. (1990). Histologic Study of the normal human adult soft palate. Cleft Palate Journal, 27, 26-34.

Langdon, H.L. & Klueber, K. (1978). The longitudinal fibromuscular component of the soft palate in the fifteen-week human fetus: musculus uvulae and palatine raphe. Cleft Palate Journal, 15, 337-348.

Latham, R.J., Long, R.E., & Latham, E.A. (1980). Cleft palate velopharyngeal musculature in a five-month-old infant: a three dimensional histological reconstruction. Cleft Palate Journal, 17, 1-16.

McMyn, J.K. (1940). The anatomy of the salpingopharyngeus muscle. Journal of Laryngology and Otolaryngology, 55, 1-22.

Negus, V.E. (1943). The mechanism of swallowing. Proceedings of the Royal Society of Medicine, 36, 84-97.

Ruding, R. (1964). Cleft palate: anatomical and surgical considerations. Plastic and Reconstructive Surgery, 33, 132-142.

Smith, K.K. & Kier, W.M. (1989). Trunks, tongues, and tentacles: Moving with the skeletons of muscle. American Scientist, 77, 29-35.

Strong, L.G. (1949). Muscle force components in the occlusive mechanism of the nasopharynx. Anatomical Record, 103, 510.

Townshend, R.H. (1940). The formation of Passavant's bar. Journal of Laryngology and Otolaryngology, 55, 154-165.

Whillis, J. (1930). A note on the muscles of the palate and the superior constrictor. Journal of Anatomy, 65, 92-95.

Wilkinson, J. (1990). SYSTAT: The System for Statistics. Computer Program Manual, Evanston, IL: SYSTAT, Inc.

Wood-Jones, F. (1940). The nature of the soft palate. Journal of Anatomy, 74, 147-170.

NCVS Status and Progress Report - 2
April 1992, 17-37

Vocal Intensity in Speakers and Singers

Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
and
Recording and Research Center, The Denver Center for the Performing Arts
Johan Sundberg, Ph.D.
Department of Speech Communication and Music Acoustics, The Royal Institute of Technology
Stockholm, Sweden

Abstract

Vocal intensity is studied as a function of fundamental frequency and lung pressure. A combination of analytical and empirical models is used to predict sound pressure levels from glottal waveforms of five professional tenors and twenty five normal control subjects. The glottal waveforms were obtained by inverse filtering the mouth flow. Empirical models describe features of the glottal flow waveform (peak flow, peak flow derivative, open quotient, and speed quotient) in terms of lung pressure and *phonation threshold pressure*, a key variable that incorporates the F_o dependence of many of the features of the glottal flow. The analytical model describes the contributions to SPL by the vocal tract. Results show that SPL increases with F_o at a rate of 8-9 dB/octave provided that lung pressure is raised proportional to phonation threshold pressure. SPL also increases at a rate of 8-9 dB per doubling of *excess pressure over threshold*, a new quantity that assumes considerable importance in vocal intensity calculations. For the same excess pressure over threshold, the professional tenors produced 10-12 dB greater intensity than the male nonsingers, primarily because their peak airflow was much higher for the same pressures. A simple set of rules is devised for predicting SPL from source waveforms.

Introduction

Regulation of vocal intensity in speech and singing continues to be an active research topic. Early studies by Ladefoged and McKinney (1963) and Isshiki (1964) showed that lung pressure is a primary control variable, but other mechanisms of intensity control within the larynx and above the larynx have not been well understood. In particular, the role of vocal fold adduction and formant tuning in intensity regulation are still somewhat unquantified. Studies in voice acoustics have focused primarily on spectral and temporal characteristics of the speech waveform and only secondarily on intensity level. This is understandable because many research questions in acoustic phonetics, speech coding, speech synthesis, speech transmission, and speech and voice disorders do not require consideration of signal intensity. Speech signals are routinely recorded, amplified, and filtered to obtain convenient levels of presentation to an observer (both visually and auditorially), usually making absolute power an irrelevant quantity.

Even to most speakers engaged in conversational dialogue, vocal intensity does not appear to be a major point of concern. Usually speakers are able to control the level of output to match environmental conditions. Raising and lowering one's voice seems (subjectively) to be a simple scaling problem. A little more respiratory effort, keeping all else the same, seems to accomplish the task. Animals as well as humans know instinctively how to increase sound power by increasing respiratory effort.

Closer examination of intensity control mechanisms reveals, however, that the task is neither simple nor stereotypic. There are different ways in which sound power can be changed by a speaker or a singer, some of which are more effective than others. Bouhuys *et al* (1968) and Schutte (1980), for example, showed that the glottal efficiency varied widely among singers and speakers. A greater price (in terms of vocal health and effort) may therefore be paid for some control strategies in relation to others. This may not be critical for the average conversationalist, but could have a profound impact on those who use their voices extensively in their professions, or those who have disorders that limit normal speech function. It is felt, therefore, that a somewhat detailed analysis of the acoustic power generating and controlling mechanisms is needed to better understand what constitutes optimal usage of one's voice. In addition, since speech signal power may also be correlated with voice source characteristics such as spectral balance and aperiodicity, a better understanding of the mechanisms of power regulation may have many implications for studies on speech quality.

In this paper, some of the basic theoretical underpinnings for power generation in vocal fold vibration will be explored. Specifically, the questions of interest will be (1) what is the relation between lung pressure and vocal intensity, (2) what is the relation between fundamental frequency and vocal intensity, (3) how does the vocal tract affect the power produced at the glottis, (4) how well can a classical source-filter model predict the variations in intensity on the basis of the source waveform, and (5) what do singers and nonsingers do differently in their control of intensity?

The present study begins with some simple analytical expressions for radiation of acoustic power from the mouth. This is followed by an empirical study of how power is regulated at the source. The influence of the vocal tract in regulating output power is added, and finally, a summary of the combined source-tract optimization strategies utilized by speakers and singers is given.

I. Analytic Expressions for Radiated Power and Intensity

Analytic expressions for acoustic power generation in speech have limited predictive power because they either involve too many assumptions, or they are mathematically too complex. Nevertheless, it is instructive to derive some simplified analytic expressions to set the course for later numerical or empirical work. In particular, the transfer of power through the vocal tract, which includes reflections at the mouth and glottis, can be formulated without too much complexity if fundamental frequencies remain below the first formant frequency. In addition, it can be shown explicitly (by formula) that power generation at the glottis is dependent on certain key waveshape parameters of the glottal volume velocity. This will then guide later studies involving multiple parameters in computer simulation models. We begin with analytic expressions for acoustic power radiated from the mouth.

A. Radiation from the mouth

Acoustic radiation from the mouth of a speaker or singer has been shown to be approximately equivalent to radiation from a piston in a spherical baffle (Morse, 1948; Flanagan, 1972). The simplest acoustic circuit that approximates this process (for frequencies below about 3000 Hz) is a parallel combination of resistance and inertance driven by a flow source u at the mouth (Figure 1). According to Flanagan, the values of the resistance R_r and inertance I_r are

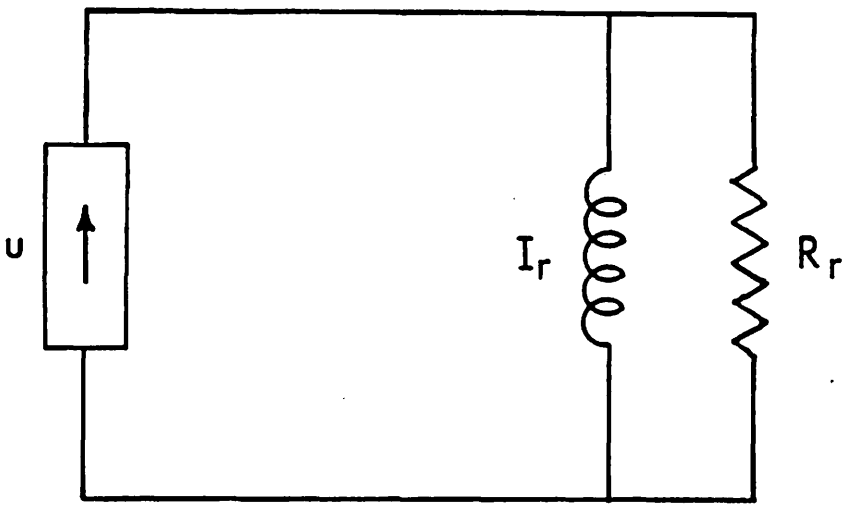

Figure 1. Parallel inertance-resistance combination for an equivalent circuit of the mouth radiation load. The mouth flow is u, the inertance is I_r, and the resistance is R_r.

$$R_r = \frac{128\ \rho c}{9\pi^3 r^2} \tag{1}$$

$$I_r = \frac{8\rho}{3\pi^2 r} \quad , \tag{2}$$

where ρ is the density of air (1.14 kg/m^3), c is the sound velocity (350 m/s for warm, moist air in the vocal tract), and r is the lip radius. The flow u in Figure 1 is in units of volume velocity (m^3/s).

At low frequencies and small lip openings $\left(\frac{\omega r}{c} \prec \frac{1}{6}\right)$ the radiation resistance is at least ten times greater than the inertive reactance. For example, at f = 1000 Hz and r = 1 cm, R_r is

1.83×10^6 kg/m^{4}s and ωI_r is 1.93×10^5 kg/m^{4}s, which gives the approximate 10:1 ratio. At the same frequency and lip radius, $\omega r/c = 0.18$, which is slightly greater than the 1/6 ratio mentioned above. Thus, considering a broad fundamental frequency range of 50 Hz - 1000 Hz for speech and singing (males and females), the radiation resistance is *guaranteed* to be much higher than the reactance at all of the fundamental frequencies, but not necessarily at the harmonics. However, since the harmonics usually decay at a rate of about 6 dB/octave, the increase in reactance for harmonics above 3-4 kHz is usually of little consequence in overall power calculations.

The low frequency radiation circuit of Figure 1 allows the radiated power to be calculated in a rather simple way. Since nearly all the flow $\dot{u}$ goes through I_r, the instantaneous pressure at the mouth is, appropriately,

$$p_r = I_r \dot{u} \tag{3}$$

where $\dot{u}$ is the time derivative of the airflow through the lips. With this pressure across the radiation resistance R_r, the instantaneous acoustic power radiated is

$$\mathcal{P} = p_r^2/R_r \tag{4}$$

$$= (I_r \dot{u})^2/R_r \tag{5}$$

$$= \frac{\rho}{2\pi c} \dot{u}^2 \quad , \tag{6}$$

with substitutions of equations (1) and (2) for R_r and I_r. Equation (6) is an important relationship between the radiated acoustic power and the flow derivative at the lips. Immediately one can establish that the radiated power increases 6 dB whenever the flow derivative is doubled. It is interesting to point out that, aside from the physical constants ρ and c, this power is completely specified by the flow derivative. This relationship will become a basic building block from which power expressions involving source and vocal tract will be deduced in subsequent discussions.

To examine radiated power in the frequency domain, consider that the mouth pressure has frequency components of the form

$$p_r = P_o \sin(\omega t + \theta) \quad , \tag{7}$$

where P_o is the amplitude and θ is the phase at a given frequency ω. The average radiated power over one cycle at this component frequency is then

$$\bar{\mathcal{P}}(\omega) = \frac{1}{T} \int_o^T p_r^2/R_r \, dt \tag{8}$$

$$= \frac{1}{2} P_o^2/R_r \quad . \tag{9}$$

The same result can be obtained by performing the complex conjugate operation

$$\bar{\mathcal{P}}(\omega) = \frac{1}{2} P_r P_r^* / R_r \tag{10}$$

where P_r is the Laplace or Fourier transform of the radiated pressure. (Recall that P_r is generally a complex quantity and that radiated power must be a real quantity; hence the complex conjugate operation.)

The Laplace transform P_r is obtained from Figure 1 by solving for the pressure across the resistance in terms of the total flow

$$P_r = \frac{sR_rU}{s+R_r/I_r} \quad . \tag{11}$$

Thus, the power of any frequency component can be computed from equation (10) provided that U, the Laplace (or Fourier) transform of the mouth flow u, is known. This can be derived from a model of the glottal flow and the vocal tract transfer function, which will now be discussed.

B. Vocal tract transfer function

Fant (1960) has shown that the vocal tract transfer function for non-nasalized vowels can be modeled as an all-pole system of the form

$$\frac{U(s)}{U_g(s)} = \frac{1}{\prod_{i=1}^{\infty} (1-s/s_i)(1-s/s_i^*)} \quad , \tag{12}$$

where $U_g(s)$ is the Laplace transform of the glottal flow and (s_i, s_i^*) are pairs of complex conjugate poles representing the formants of the vocal tract. More specifically, s_i can be written as

$$s_i = -\frac{\omega_i}{2Q_i} + j\omega_i \quad , \tag{13}$$

where ω_i is the radian formant frequency and Q_i is the quality of the resonance. Transmission of sound energy through the vocal tract wall is neglected in this all-pole series expansion of the transfer function.[1] Combining the vocal tract and the radiation transfer functions (equations 11 and 12), the ratio of radiated pressure to glottal flow becomes

$$\frac{P_r(s)}{U_g(s)} = \frac{sR_r}{(s+R_r/I_r)\prod_{i=1}^{\infty}(1-s/s_i)(1-s/s_i^*)} \tag{14}$$

$$= H(s) \tag{15}$$

This is the classical source-filter transfer function. From H(s) the power can be calculated with equation (10) if the source spectrum is known.

C. Glottal source waveform and spectrum

Consider first a low frequency power calculation. Equation (12) shows that for $s \to o$, the vocal tract offers a unity transfer function, with the mouth flow being identical to the glottal flow. Acoustically, the vocal tract has negligible length and the glottis is effectively connected to the lips.

Assume now that the flow pulse has temporal characteristics as shown in the top waveform of Figure 2. The flow derivative $\dot{u}$, shown below the u waveform, has been approximated by straight line segments in the open phase. Such an approximation is reasonable for evaluating total power, but would not be valid if the spectral components of the power were to be represented accurately. These spectral components depend critically on waveshape (Ananthapadmanabha, 1984).

Let T_p be the duration of positive $\dot{u}$ and T_n the duration of negative $\dot{u}$. Assuming there is no drift in the mean flow from period to period, the area under $\dot{u}$ during T_p must equal the area under $\dot{u}$ during T_n. This implies that if $\dot{u}_m$ is the maximum negative derivative as shown, then $\dot{u}_m T_n/T_p$ is the maximum positive derivative. According to equation (6), the approximate time-average glottal power is then

$$\bar{\mathcal{P}}_g = \frac{\rho}{2\pi c}\,\frac{1}{T}\int_0^T \dot{u}^2 dt \tag{16}$$

$$= \frac{\rho}{2\pi cT}\left[\tfrac{1}{3}\left(\frac{T_n}{T_p}\right)^2 T_p\dot{u}_m^2 + \tfrac{1}{3}\,T_n\dot{u}_m^2\right] \tag{17}$$

$$= \frac{\rho}{6\pi c}\,\frac{T_n}{T}\left(\frac{T_n}{T_p}+1\right)\dot{u}_m^2 \tag{18}$$

$$= \frac{\rho}{6\pi c}\,\frac{Q_o}{Q_s}\,\dot{u}_m^2 \tag{19}$$

[1]The wall compliance and associated losses have been discussed and evaluated (Fujimura & Lindqvist-Gauffin, 1971) and can be approximated by a modification of equation (13).

where Q_o is the open quotient, defined as $(T_p + T_n)/T$, and Q_s is the skewing ratio (speed quotient), defined as T_p/T_n. Equation (19) would represent the glottal power, and approximately the speech power, if the glottal flow (as observed by inverse filtering under realistic vocal tract loading) were to radiate directly into free-space (no vocal tract). $\bar{\mathcal{P}}_g$ will henceforth be referred to simply as the *approximate glottal power*. It will prove to be of value in estimating the source power from inverse-filtered glottal waveforms. Quantities like Q_o, Q_s, and $\dot{u}_m$ can all be obtained from glottal waveform shapes and can thus be utilized to assess power regulation at the source.

There are several problems with attempting to generalize this approximate source power to high fundamental frequencies. First of all, an impedance mismatch (high source impedance, low radiation impedance) was assumed in decoupling the vocal tract for the source and the radiator. This assumption of impedance mismatch is not valid near the formant frequencies, where the input impedance to the vocal tract is high. Furthermore, the vocal tract load changes from being inertive to resistive to compliant as F_o moves through F_1, making the power calculations in equations (6) and (19) invalid. For this reason, fundamental frequencies of subjects (to be discussed later) were kept below about half of the first formant frequency.

It is instructive to compare the approximate glottal power to the power derived from a sinusoidal source of the same fundamental frequency and flow amplitude. For a sinusoidal source, $\dot{u}$ can be written as

$$\dot{u} = \dot{u}_m \cos 2\pi F_o t \quad . \tag{20}$$

The time-averaged value of the square of the cosine is 1/2, making the time-averaged power of the sinusoidal source

$$\bar{\mathcal{P}}_o = \frac{\rho}{4\pi c} \dot{u}_m^2 \tag{21}$$

according to equation (6). Comparing this result with equation (19) for $Q_o = 1.0$ and $Q_s = 1.0$ (the asymptotic condition for a sinusoidal waveform shape), the low frequency power from the sinusoid appears to be 50% larger than the low frequency power estimated from the complex waveform. This difference results from the straight-line approximations made in the integration of $\dot{u}^2$. For a sinusoidal $\dot{u}$ waveform, this approximation is obviously not correct. Fortunately, typical $\dot{u}$ waveforms are more triangular than sinusoidal (Figure 2 is a typical inverse-filtered waveform). This makes the 50% error in the power calculation a worst case situation. In subsequent discussions, it should be kept in mind, however, that equation (19) may underestimate the approximate glottal power by 1-2 dB if the flow waveform approaches a sinusoid.

The sinusoidal source power in equation (21) is in agreement with conventional formulae for radiation from simple sources. If u_m is the peak value of the flow, then $\dot{u}_m = 2\pi F_o u_m$ and

$$\bar{\mathcal{P}}_o = \frac{\rho\pi}{c} F_o^2 u_m^2 \quad , \tag{22}$$

which is the classic low-frequency formula for power radiated from a hemisphere or from a small piston in an infinite baffle (Kinsler & Frey, 1950; Morse, 1948). It is also twice the power radiated from a pulsating sphere (an isotropic radiator). Equation (22) shows clearly that radiated power from a pulsating source is proportional to the square of the frequency and the square of the source strength (u_m). Each of these variables contribute an increase of 6 dB of power when they are doubled.

Consider now a simple spectral model for the flow pulse. A logical approach would be to approximate the rising and falling portions of the flow or flow derivative in Figure 2 with mathematical functions (pieces of sinusoids, polynomials, or exponentials; e.g. Fant, Liljencrants & Lin, 1985), and to Fourier analyze this stylized waveform. Unfortunately, the resulting closed-form expressions for the Fourier coefficients are usually too complex to be of much help. (For an example that uses a truncated sinusoid model, see Titze, 1988b.) Spectra for more realistic glottal flow models are usually computed numerically (e.g., Ananthapadmanabha, 1984). A highly simplified one-parameter source spectrum will be used here by making the classical assumption that the amplitudes of spectral components of the flow pulse decrease exponentially with harmonic number n, such that

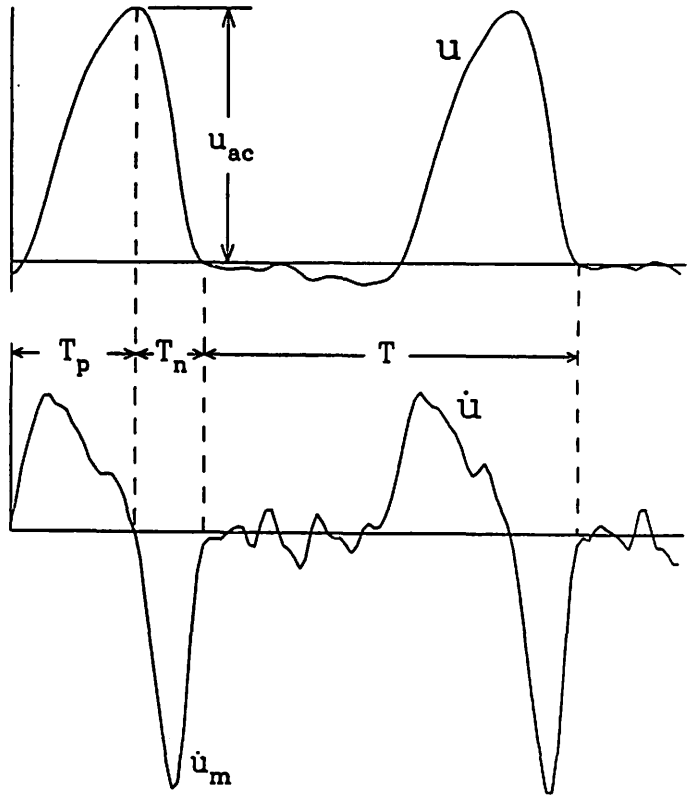

Figure 2. Two successive glottal flow pulses (top) and the corresponding flow derivative (bottom). The flow derivative is approximated by straight line segments (dashed) in the opening and closing phases.

$$U_g(nF_o) = k(nF_o)^{-\alpha} . \qquad (23)$$

In this equation, α is the parameter to be varied and k is a constant to be evaluated later. This source spectrum neglects several of the subtleties discussed recently by Ananthapadmanabha (1984) and by Fant and Lin (1987). Primarily, a broad, low frequency prominence and a series of spectral zeros are omitted, but it is expected that these subtleties do not substantially affect the first-order power calculation attempted here.

The constant α is customarily expressed in dB/octave, i.e.,

$$\alpha = \frac{\text{spectral slope in dB/oct.}}{20 \log_{10} 2} . \qquad (24)$$

A spectral slope of 12 dB/octave yields an α of 2.0, which is typical for normal male phonation in speech. Some estimates of the spectral slopes for singers and nonsingers will be given later.

D. Analytic expression for vocal power

The source spectrum $U_g(nF_o)$ in equation (23) is a real quantity and remains separate from the transfer function H(s) (equations 14 and 15) in a power calculation. The radiated power of the n-th harmonic is then

$$\bar{\mathcal{P}}_n = \left(\frac{k^2}{2R_r}\right)(nF_o)^{-2\alpha} \left|H(nF_o)\right|^2 \tag{25}$$

according to equations (10), (11), (12), and (14). Note that the magnitude-squared operation in (26) is the effect of the complex conjugate multiplication in equation (10).

All quantities in equation (25) are known except the constant k. This can be evaluated by considering the low-frequency asymptotic condition previously discussed. Equating the low frequency power in (19) to the summation of all of the harmonic components $\sum\bar{\mathcal{P}}_n$, with $\bar{\mathcal{P}}_n$ as given in equation (25), the constant k^2 becomes

$$k^2 = \bar{\mathcal{P}}_g\ (2R_r)(2\pi F_o I_r)^{-2}\ F_o^{2\alpha}/\sum_{n=1}^{\infty} n^{2-2\alpha} \quad . \tag{26}$$

Strictly speaking, the summation of harmonic terms cannot go to infinity because the low-frequency approximation would not hold. However, only a few terms in the summation are significant for $\alpha \geq 2$ (i.e., for source spectra that fall 12 dB/octave or more). The harmonic amplitudes decrease by at least $1/n^2$, making only a few low harmonics significant in the sum. This is obviously valid only at asymptotic F_os, where the harmonics do not reach the formant regions.

Substituting k^2 back into equation (26) and again summing over all n, the total radiated power at normal F_os becomes

$$\bar{\mathcal{P}} = \bar{\mathcal{P}}_g\ G(F_o) \quad , \tag{27}$$

where $G(F_o)$ is defined as the *power gain* over the approximate glottal power,

$$G(F_o) = \frac{\sum_{n=1}^{\infty} n^{-2\alpha}\ \left|H(nF_o)\right|^2}{(2\pi F_o I_r)^2 \sum_{n=1}^{\infty} n^{2-2\alpha}} \tag{28}$$

The use of the term *power gain* here is not to suggest that the vocal tract can actually increase the source power, like an amplifier. Since the tract is a passive system, it can only facilitate a better power transfer (i.e., a better impedance match) between the glottis and free space. The term *gain*, therefore, refers to a power increase only *vis a vis* direct radiation by the larynx into free space with the same glottal pulse.

E. Conversion from radiated power to intensity

In order to express radiated power in the more common sound pressure level (SPL) units, a conversion from power to intensity is needed. If isotropic, far-field radiation is assumed, the sound pressure level in dB is

$$SPL = 10 \log_{10} \frac{I}{I_o} \tag{29}$$

$$= \log_{10} \frac{10^{12}\ \bar{\varphi}}{4\pi R^2} \tag{30}$$

$$= 120 - 10 \log_{10} (4\pi R^2) + 10 \log_{10} \bar{\varphi} \quad , \tag{31}$$

where R is the mouth-to-microphone distance in meters and $\bar{\varphi}$ is the power in watts (equation 27). The reference intensity I_o is the standard 10^{-12} watt/m^2. Some correction to this far field, isotropic radiation expression is necessary if the mouth-to-microphone distance is small or if the microphone is highly directive. For 0.5 m and a typical cardioid or omnidirectional pattern, the expression is expected to be accurate to within ± 2 dB, and equation (31) simplifies to

$$SPL = 115 + 10 \log_{10} \bar{\varphi} \quad . \tag{32}$$

This expression is rather simple to remember: *One watt of radiated power is 115 dB SPL at 0.5 meter*. Some speakers, and especially singers, can produce 90-100 dB at that distance, suggesting that they radiate in excess of 0.01 watt of acoustic power.

F. Summary of analytic results

This concludes the analytic derivations for power and intensity. Critical source parameters in the expressions are $\dot{u}_m$, Q_o, Q_s, F_o, and α. Critical vocal tract parameters are r, the mouth orifice radius, together with the formant frequencies ω_i and the resonant qualities Q_i. These parameters cannot all be varied independently by a speaker. Some further empirical relationships are needed to explore the combined control strategies used by individuals in intensity regulation. The following section is devoted to an investigation of how the parameters affect each other and the vocal intensity.

II. Experimental Studies

Nothing has been said so far about lung pressure[2], the primary control variable for vocal intensity. All of the glottal parameters $\dot{u}_m$, Q_o, Q_s, F_o, and α are expected to vary with lung pressure.

[2]*The term lung pressure is used here for brevity instead of mean subglottal pressure. The difference is negligible for any consideration here, and lung pressure is conceptually the simplest to deal with.*

Some data from previous studies on human subjects and excised canine larynges will now be used to develop empirical relations and to combine them with the analytical model presented above.

A. Measurements of intensity, waveform parameters, and spectral slope

In a recent study by Sundberg, Scherer, and Titze (in press), glottal flow waveforms were obtained by inverse filtering the oral airflow (Rothenberg, 1973). The subjects were professional and amateur singers. Data from a subset of these subjects, five professional tenors, will be reported here. One singer was an operatic tenor, one a recording artist and recitalist of classical song, two were musical theatre singers, and one was a counter tenor. The tenors were chosen because they best matched the low F_o phonations of a group of control subjects, namely the male nonsingers studied by Holmberg, Hillman, & Perkell (1988).

All subjects in the Sundberg *et al* study repeated the syllable [pæ] five times on each of five pitches, spanning about 1½ to 2 octaves. There were four loudness conditions on each pitch (soft, medium soft, medium loud, and loud). In the Holmberg *et al* study, there was one pitch level (normal speaking pitch) and three loudness levels (soft, normal, and loud). Subjects were 25 males and 20 females, a larger group than the singers.

Sound pressure level

The sound pressure level (SPL) was measured for all subjects, using the linear scale of a B&K sound level meter (Type 2230). In the Sundberg *et al* study, the mouth to microphone distance was 0.5 m, whereas in the Holmberg *et al* study, it was 0.15 m. This required a correction of -10.4 dB (a reduction) for the Holmberg *et al* SPL levels to match the Sundberg *et al* data (see equation 31 for renormalization with a different R).

Figure 3 shows a collection of SPLs measured for the five tenors as a function of F_o. There are 60 data points, representing three frequencies and four loudnesses for each of the five tenors (some of the data points are double valued and the highest point, 101 dB, falls just outside the plotting range). These measurements were then averaged over the five subjects and over three F_o clusters, the low F_os (105, 110, 131 Hz; mean 115

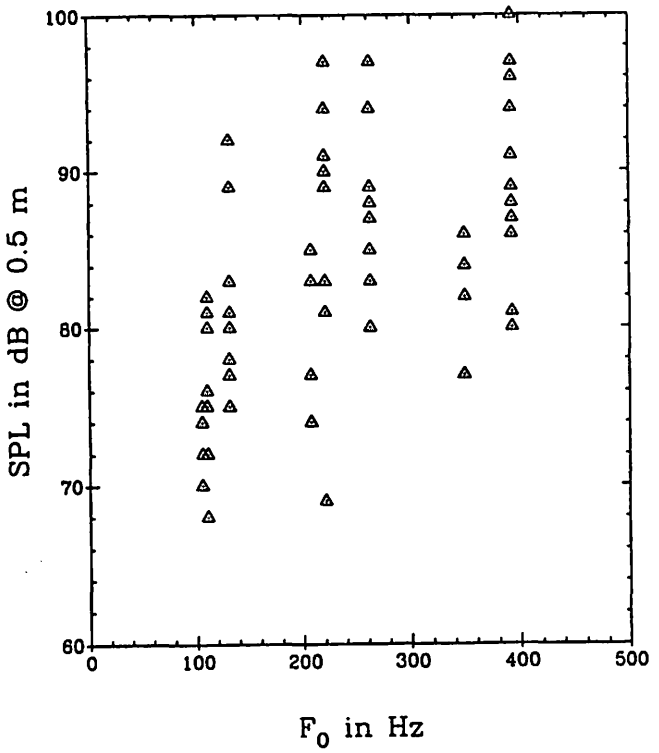

Figure 3. Sound pressure levels (SPLs) measured on five professional tenors at a variety of fundamental frequencies and loudness conditions.

Hz), the medium F_os (207, 220, 262 Hz; mean 234 Hz) and the high F_os (349, 392 Hz; mean 383 Hz). Thus, the 60 data points were collapsed into 12 points (three frequencies x four loudnesses) for subsequent presentations. The Holmberg *et al* data had already been averaged in their reported tables.

Lung pressure

Lung pressure was estimated from oral pressure during mouth occlusion (Rothenberg, 1973; Smitheran & Hixon, 1981). Commercially available pressure and flow transducers (Glottal Enterprises) were used, after calibration, as described by Sundberg *et al* (in press). Figure 4 shows the averaged lung pressures in kPa produced by male subjects as a function of F_o. (Recall 1 kPa ≈ 10 cm H_2O.) Open triangles, circles, and squares represent the low, medium, and high F_os for the tenors, respectively, and the four loudness conditions are represented by progressively higher pressures. As a control, the averaged pressures from the Holmberg *et al* male nonsingers are shown by filled triangles. These pressures represent only three loudnesses at one F_o (116 Hz). Female pressures for nonsingers were similar, but they do not serve as a control because there were no female singers in this study.

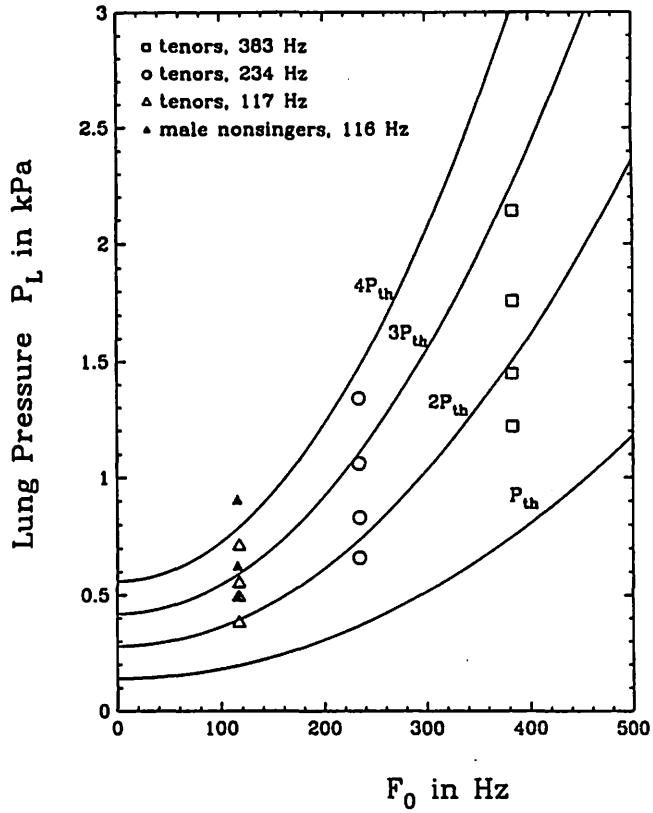

Figure 4. Lung pressure P_L for five professional tenors at three averaged frequencies and for male nonsingers at one frequency. The curves represent the phonation threshold pressure and multiples thereof.

Two observations stand out. First, the lung pressures rise systematically with F_o. The smallest P_L at 383 Hz is greater than the largest P_L at 117 Hz. The range of P_L between softest phonation and loudest phonation also increases with F_o, from about 0.5 kPa at 117 Hz to more than 2.0 kPa at 383 Hz. Second, the professional tenors used smaller pressures, on average, than the male nonsingers at a comparable F_o (compare filled with unfilled triangles). These results, and several others to be described below, can be explained on the basis of a phonation threshold pressure P_{th}. This quantity, along with its multiples $2P_{th}$, $3P_{th}$, and $4P_{th}$, is plotted by the curved lines in Figure 4. Phonation threshold pressure, the minimum lung pressure required to initiate vocal fold vibration, has become a key parameter in the development of empirical relations for glottal aerodynamics, and will be used later to derive relations for glottal source power.

Glottal airflow

After inverse filtering and digitization of the oral flow (20 kHz, 16 bits), each glottal flow waveform and its time derivative were displayed on a computer terminal. Figure 2 is an example of two cycle of a flow waveform u and its derivative $\dot{u}$. The following measures were extracted from these waveforms: the alternating flow u_{ac}, the maximum negative flow derivative $\dot{u}_m$, the duration of increasing flow T_p (positive flow derivative), the duration of decreasing flow T_n (negative flow derivative), and the period T.

Spectral slope

Some estimates of the spectral slope of the inverse filtered glottal waveforms were made from the data of the five tenors. Four consecutive cycles of the u waveforms were windowed and Fourier analyzed (discrete Fourier transforms, with the number of points being determined by the 20 kHz sampling frequency and the four cycle window). By viewing all the spectra, it was determined that harmonic energy above about 1500 Hz (above the second formant) was essentially negligible. For high F_o, therefore (350-400 Hz), the spectral slope was measured simply as the difference (in dB) between the levels of the first and second harmonics. For lower F_os (100-250 Hz), two spectral slopes were computed and averaged. The first slope was again the difference between the levels of the first and second harmonics, whereas the second slope was based on all harmonics above the second (up to about 1200 Hz). This averaging of two spectral slopes was necessary to approximate both the low frequency roll-off and the high frequency roll-off of the spectrum with a single parameter, to the extent that this is possible. (For a discussion of multiple parameter spectral measures, see Ananthapadmanabha, 1984). Results of these spectral calculations are shown in Figure 5 (data points). The data are quite scattered because spectral slope measurements are not very precise, even when averaged. Nevertheless, the trend with lung pressure is seen rather well. The curves represent an empirical model to be described below.

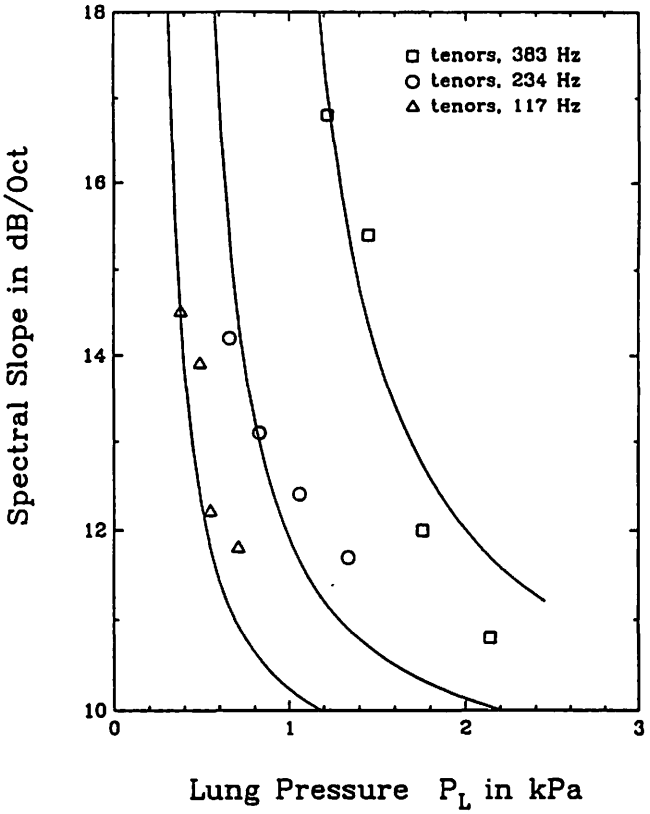

Figure 5. Spectral slope versus lung pressure for five professional tenors at three frequencies. The curves represent equation (38).

B. Empirical relations

In order to explain the dependence of SPL on F_o and P_L, some empirical relations are first restated from another study (Titze, in review):

$$u_{ac} = k_1 \frac{P_L - P_{th}}{P_{th}} (1-\cos\pi Q_o) \tag{33}$$

$$Q_o = k_2 + (1-k_2) \frac{P_{th}}{P_L} \tag{34}$$

$$Q_s = 1.0 + k_3(2P_m - P_L - P_{th})(P_L - P_{th}) \tag{35}$$

$$\dot{u}_m = k_4 u_{ac} F_o (Q_s+1)/Q_o \tag{36}$$

All quantities on the left side of these equations have been defined in forgoing sections. New quantities on the right side, k_1, k_2, k_3, k_4, and P_m are empirically determined constants, summarized in Table 1. Two of the constants are dimensionless (k_2 and k_4), while all others have dimensions that involve flow in m^3/s and pressure in kPa.

Table 1. Empirical Constants in Equations (34)-(37)

Subject Group	k_1	k_2	k_3	k_4	P_m
tenors	0.0003	0.40	0.50	1.5	2.0
male nonsingers	0.00009	0.45	0.55	1.9	2.0
female nonsingers	0.000045	0.62	3.00	1.9	0.65

Note the role that the pressure difference P_L-P_{th} plays in equations (33) and (35). It is the quantity that drives the time-varying flow. As P_L approaches P_{th}, the time-varying flow goes to zero, and the open quotient and speed quotients approach 1.0.

All of the F_o variation of u_{ac}, Q_o, and Q_s is contained in P_{th}. This threshold pressure is approximated by the quadratic form (Titze, in review)

$$P_{th} = 0.14 + 0.06(F_o/\bar{F}_o)^2 \quad \text{kPa} \quad , \tag{37}$$

where $\bar{F}_o$ is the mean speaking fundamental frequency (120 Hz for males and 190 Hz for females). The curves in Figure 4 (for male subjects) were generated with equation (37). Note that typical pressures for speakers and singers range between $2P_{th}$ and $4P_{th}$, with a few exceptions at extreme loudnesses and fundamental frequencies.

Based on this threshold pressure, an empirical model for spectral slope is

$$\text{Spectral slope} = k_5 + 5\,\frac{P_{th}}{P_L - P_{th}} \quad \text{dB/octave} \quad , \tag{38}$$

with k_5 being 9.0 for singers and 11.0 for nonsingers (males and females). This yields a range of 19 dB/octave to 10.7 dB/octave for the singers as lung pressure is varied from $1.5P_{th}$ to $4P_{th}$ (Figure 5). For the nonsingers, the spectral slope is 2.5 dB/octave higher overall, ranging from 21 dB/octave at $1.5P_{th}$ to 12.7 dB/octave at $4P_{th}$.

With the use of equation (33) and (36), the approximate glottal power $\bar{\mathcal{P}}_g$ in equation (19) can be computed as a function of P_L and F_o,

$$\bar{\mathcal{P}}_g = \frac{\rho}{6\pi c}\,(k_1 k_4)^2\,\left(\frac{P_L - P_{th}}{P_{th}}\right)^2\,F_o^2\,\frac{(Q_s+1)^2}{Q_s}\,\frac{(1-\cos\pi Q_o)^2}{Q_o} \quad . \tag{39}$$

This power would increase with the square of (P_L-P_{th}) if F_o, Q_o, and Q_s were all held constant. It would then yield a 6 dB increase per doubling of the pressure difference (P_L-P_{th}). Note also that the F_o^2 factor would yield a 6 dB increase in power per octave rise in fundamental frequency if other variables were held constant. These hypothetical cases are nearly realized because the two factors involving Q_s and Q_o in equation (39) are relatively weak functions of P_L and do not contribute much to the power variation with either P_L and F_o. A more detailed analysis of this will now be given by including the vocal tract model in the power calculations.

III. Variation of SPL with Lung Pressure and Fundamental Frequency

The combination analytical-empirical model for vocal intensity (equation 27, with $\bar{\varphi}_g$ being empirical and $G(F_o)$ being analytical) can now be tested against data obtained from singers and nonsingers. Figure 6 shows average values of SPL for the tenors of three frequencies and for male and female nonsingers at one frequency each. In this and all subsequent graphs, the following symbols are used to identify measurement and theory: open triangles, circles, and squares represent the tenors at 117 Hz, 234 Hz, and 383 Hz, respectively; filled triangles represent the male nonsingers; asterisks represent the female nonsingers; solid lines represent the model for the tenors; dashed lines represent the model for nonsingers.

As far as the vocal tract parameters are concerned, the mouth radius was set to 1.0 cm, formant frequencies were 700 Hz, 1080 Hz, and 2600 Hz for the first three formants of a simulated /a/ vowel, and all resonance Q_s were set to 10. These parameters were not varied because the questions of interest dealt with lung pressure and fundamental frequency.

Several observations are in order with regard to Figure 6. First, there is the expected nonlinear relation between SPL and P_L. Sound pressure level is a logarithmic expression for vocal power (equation 32), which has no direct proportion to lung pressure. Second, the tenors produced intensities 10-15 dB higher than the nonsingers at the same lung pressures. This is a direct result of u_{ac} being three to four times greater for the tenors than the nonsingers (k_1 in equation 33, with the numerical value given in Table 1). Each doubling of this alternating flow increases the SPL by 6 dB. Thus, for a 4:1 ratio in peak flows, the difference in SPL would be 12 dB, which is the approximate difference seen in Figure 5 in the 0.4-0.6 kPa region. Third, SPL values for male and female nonsingers do not differ substantially. An approximate 2:1 ratio in the alternating flows between males and females is nearly offset by a 1:1.7 ratio in F_o (radiated power being proportional to the square of both). Thus, SPL values are only 1-2 dB lower for females than for males.

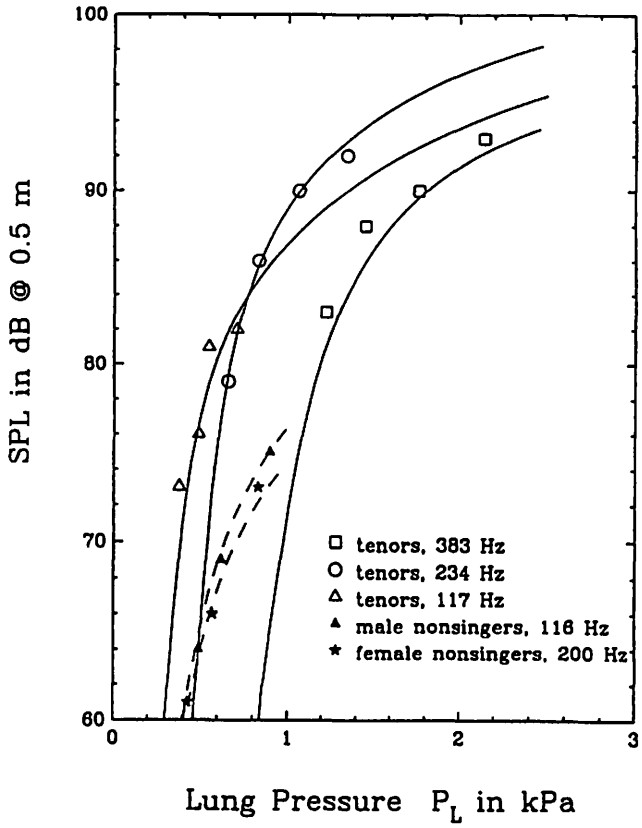

Figure 6. Sound pressure level versus lung pressure for professional tenors and nonsingers. Lung pressure is plotted on a linear scale. The curves represent the model described in the text.

Attempts have been made in the past to find simple mathematical relations between SPL and P_L by plotting the two variables on a log-log scale. This serves no useful purpose unless the phonation threshold pressure is taken into account. Figure 7 is an example of such a log-log display. The curves appear straighter only because the upper range of P_L is increased from 3 kPa to 10 kPa. There is no single slope (in terms of dB per doubling of P_L) that describes the relation. At low P_L, for example, SPL can rise as much as 10-15 dB per doubling of P_L, whereas at larger P_L the increase is more like 6-9 dB.

An improved display is obtained by plotting the *excess pressure over threshold*, P_L-P_{th}, logarithmically on the x-axis (Figure 8). This not only straightens out the curves, but tends to collapse the data of the tenors across F_o. If a single straight line were to be fit to all the data, the slope would be about 8-9 dB per doubling of P_L-P_{th}. The same slope applies to the male and female nonsingers. It appears, therefore, that the excess pressure over threshold may be an appropriate quantity to consider for a linear relation between SPL and logarithmic driving pressure.

For a clear separation of both the P_L and F_o effect on SPL, a further modification of the log-log plot is proposed. In Figure 9, the ratio $(P_L\text{-}P_{th})P_{th}$ is used as the variable along the abscissa. The quantity is the *fractional excess pressure over threshold.* For example, a value of 1.0 implies a 100% increase in P_L over P_{th}, or an absolute value of $2P_{th}$. A value of 4.0 implies a 400% increase in

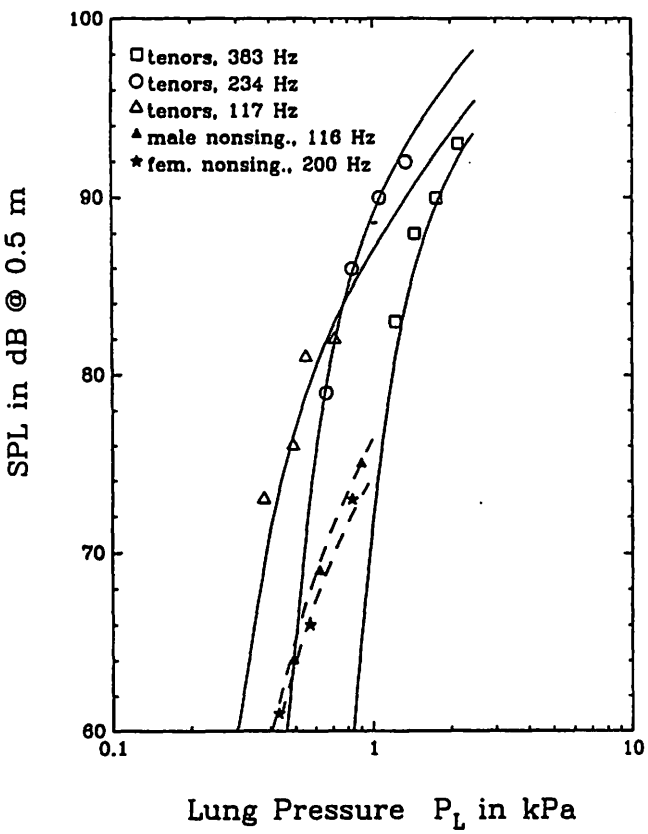

Figure 7. Sound pressure level versus lung pressure for professional tenors and nonsingers. Lung pressure is plotted on a logarithmic scale. The curves represent the model described in the text.

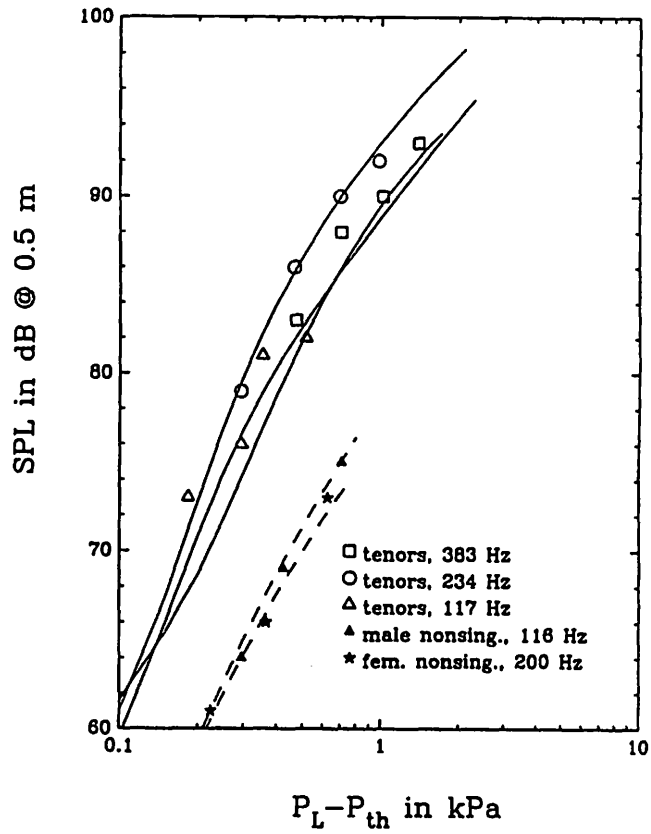

Figure 8. Sound pressure level versus excess lung pressure over threshold (on a log scale) for professional tenors and nonsingers. The curves represent the model described in the text.

P_L over P_{th} and a value of 0.5 implies a 50% increase in P_L over P_{th}. Most of the fractional excess pressures used by the subjects in Figure 9 fall between 0.5 and 4.0. The nonsingers (males and females) have no phonations for which P_L falls below 100% excess over threshold. The tenors, on the other hand, used a few pressures between 50% and 100% above P_{th} for the lowest intensities (recall also Figure 4 for the same information).

The attractive feature of Figure 9 is the vertical stacking of the data (in terms of a family of nearly parallel lines). The approximate 8-9 dB rise per doubling of excess pressure over threshold is preserved, but a clearer picture of the F_o effect emerges. Increases in SPL with F_o appear to be more systematic. They do not display the crossing of constant F_o lines seen in Figure 6-8.

To elaborate further, Figure 10 is another version of Figure 9, but with the constant F_o lines adjusted to simple multiples of 100 Hz and the data points removed for clarity. The model represents the singers only so that the approximate glottal power ($\bar{\mathcal{P}}_g$) could be shown for comparison (dot-dashed lines). The glottal power was converted to equivalent SPL values at 0.5 m mouth to microphone distance. For any of the three F_os, differences between the solid lines and dot-dashed lines measure the vocal tract "gain" (equations 27-28). Note that this gain is on the order of 10 dB, but varies with pressure and F_o.

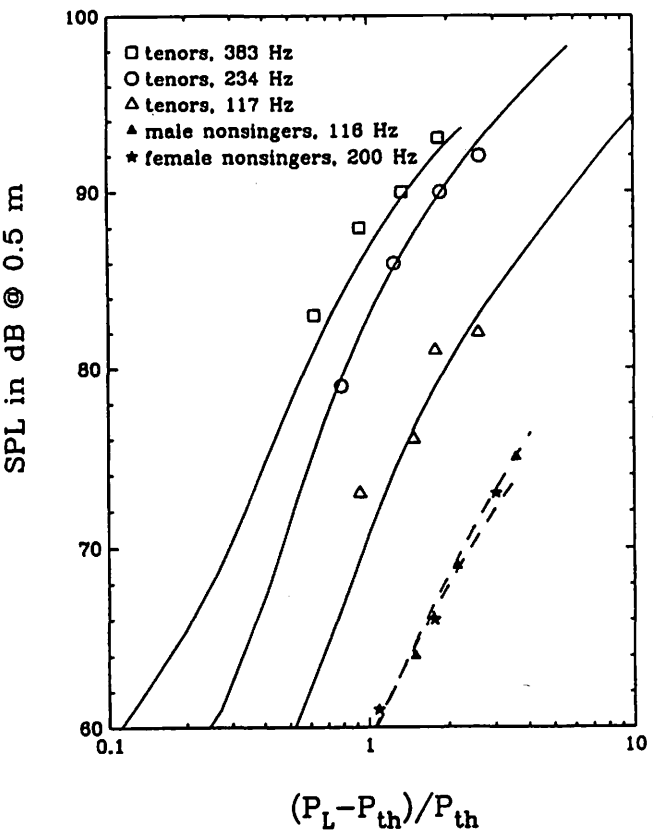

Figure 9. Sound pressure level versus fractional excess lung pressure over threshold *(on a log scale) for tenors and nonsingers. The curves represent the model as described in the text.*

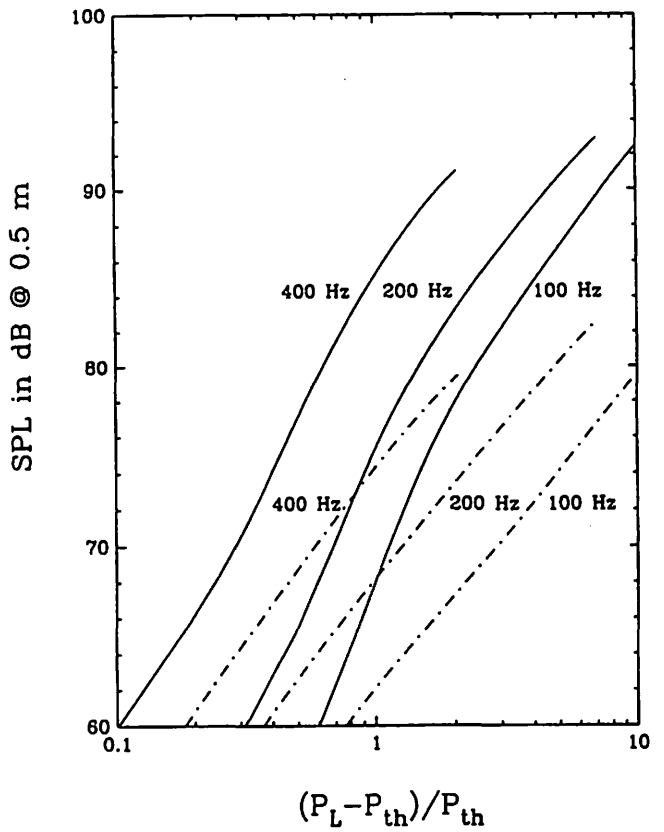

Figure 10. Computed sound pressure level versus fractional excess lung pressure over threshold *(on a log scale) at three values of F_o. Dot-dashed lines represent the approximate glottal power (the power radiated at very low F_o or in the absence of a vocal tract).*

Of particular interest is the difference in slopes between the vocal tract and no vocal tract conditions. The glottal power increases at a constant rate of 6 dB per doubling of excess pressure over threshold (recall equation 39). This verifies the previous assertion that the factors involving Q_s and Q_o explicitly in equation (39) are weak functions of P_L. With the vocal tract added, the steeper slope (8-9 dB per doubling of excess pressure over threshold) is a direct consequence of the *changing spectral slope* with P_L (equation 38). As P_L increases, relatively more harmonic power is generated (due to waveform skewing), which is boosted collectively by the formants to increase the SPL-P_L slope by another 2-3 dB. This additional slope is no longer present, however, when P_L becomes large, as seen by the bending of the solid curves in Figure 10 at the top of the graph. This asymptotic condition is evident also from equation (38), where a constant spectral slope k_s is approached as P_L becomes large.

Constant F_o lines in Figure 10 were deliberately chosen an octave apart to show more clearly the change in SPL with F_o. The glottal power increases at a rate of 6 dB/octave at any constant pressure value, as seen by the equal vertical spacing of the dot-dashed lines. This is a direct result of the F_o^2 variation of $\bar{\mathcal{P}}_g$ (equation 39). With the vocal tract included, the spacing between octaves is less uniform because harmonics may be affected differently by the formants at different F_os. A ripple effect is experienced in SPL with varying F_o.

Figure 11 shows this variation of SPL with F_o more directly. Data points are shown only for the tenors, since F_o variations for the nonsingers were not available. Two sets of model curves are shown. Solid lines represent SPL with the vocal tract included and dot-dashed lines again represent the approximate glottal power $\bar{\mathcal{P}}_g$ (converted to equivalent SPL at 0.5 m). The two intensities are each shown for two values of P_L, one for $P_L=2P_{th}$ and one for $P_L=3P_{th}$. Note that there is an exact 6 dB/octave rise in $\bar{\mathcal{P}}_g$, as previously observed.

When the vocal tract is included, the SPL-F_o slope is about 8-9 dB per octave. The additional 2-3 dB per octave rise comes from the fact that more and more of the harmonic power gets boosted by the formants as F_o increases. In addition, large fluctuations in SPL (on the order of ±3-5 dB) are observed for higher F_os when a single harmonic is turned or detuned by the first formant. For example, the largest peak in Figure 11 occurs at 350 Hz, where the second harmonic (700 Hz) lines up with the first formant. Progressively smaller peaks for lower F_os occur when the third, fourth, and fifth harmonics etc. line up with the first formant.

In the region of F_o between 100 and 400 Hz, the effective gain of the vocal tract is again seen to be on the order of 10-15 dB, but

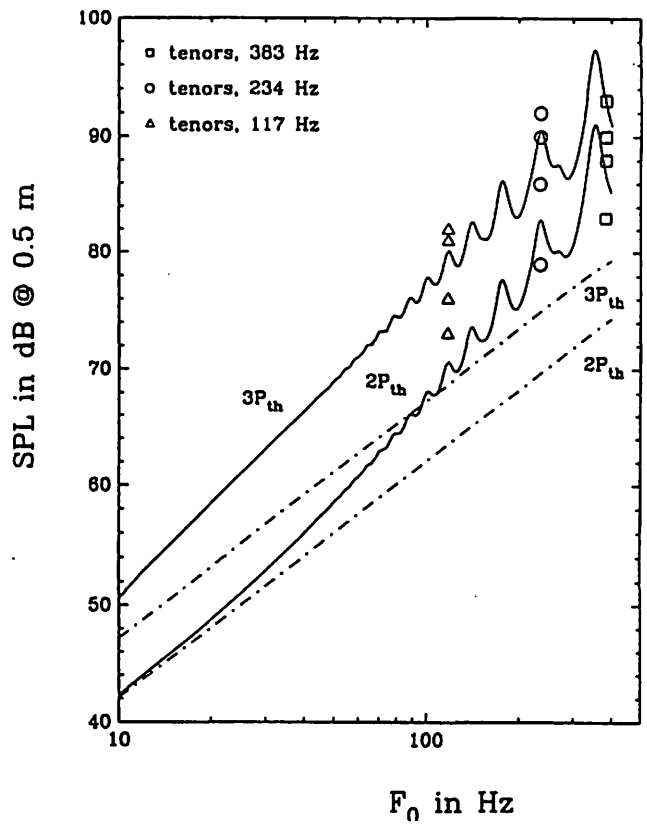

Figure 11. Sound pressure level versus fundamental frequency (on a log scale) computed at two values of P_L ($2P_{th}$ and $3P_{th}$). Dot-dashed lines represent the approximate glottal power (the power radiated either at very low F_o or in the absence of a vocal tract).

can reach nearly 20 dB when the second harmonic is tuned by F_1. Singers may be tempted to ride the peaks of the ripples shown in Figure 11 if they have the freedom to alter the vowels while maximizing intensity. Interpretations for impedance matching (or formant tuning) when F_o approaches (or goes beyond) F_1 are not appropriate on the basis of this model, as stated earlier.

A comment about the absolute value of the approximate glottal power is in order. Consider the SPL range between the dot-dashed lines at around 100 Hz in Figure 11. This 60-65 dB range for $\bar{\mathcal{P}}_g$ compares favorably with the power measured on excised larynges radiating into free space (Titze, 1988b). According to equation (32), 65 dB corresponds to a power of 10μwatts, which was the power radiated by excised larynges (without a vocal tract) at low to moderate subglottal pressures. This is an added confirmation that the models derived here have validity.

Conclusions

The questions posed at the beginning of this paper can now be addressed with a few concise answers. The first question, "What is the relation between lung pressure and vocal intensity?" has a simple answer only if the effective source pressure is taken to be the *excess pressure over threshold.* The answer is then: an 8-9 dB increase in SPL per doubling of this excess pressure over threshold. The first 6 dB of this SPL slope comes from an increase in the glottal flow amplitude, whereas the remaining 2-3 dB comes from a decreasing spectral slope (or equivalently, a less symmetric glottal waveshape). These conclusions were reached earlier by Fant (1982), but threshold pressure was not accounted for in his formulation.

The second question, "What is the relation between fundamental frequency and vocal intensity?" has a very similar answer: an 8-9 dB increase in SPL per octave of F_o rise, assuming that the lung pressure remains proportional to the phonation threshold pressure. This qualification requires the absolute lung pressure to rise quadratically with F_o.

The third question, "How does the vocal tract affect the power produced at the glottis?" does not have a simple answer. On average, the glottis would radiate 10-15 dB less power if connected directly to the lips. As an impedance matcher between the glottis and free space, then, the vocal tract provides an effective 10-15 dB "gain." Additionally, however, and more importantly for active control of intensity, vocal tract formants can be tuned to harmonics of the source to get another 3-5 dB increase in SPL at higher F_os.

The fourth question, "How well can a classical source-filter model predict the variations in intensity with F_o and lung pressure?" can be answered only by saying that the source-filter concept remains viable. The predictions made by the combination analytical and empirical models are sufficiently close to the measured SPL data (± 2 dB in all cases) that errors in the theory cannot be easily separated from measurement error.

The last question, "What do singers and nonsingers do differently in their control of intensity?" seems to hinge on one major observation: Singers apparently can achieve 3-4 times greater time-varying flows for the same lung pressures. This results in a predictable 10-12 dB difference in SPLs between singers and nonsingers. Presumably, singers learn how to lower their effective glottal impedance to transfer more power from the source to the vocal tract for a given lung pressure. One possibility is a different mode of vibration that produces greater amplitudes of vibration for the same driving pressures (less energy loss in the tissue). Another possibility is an adjustment in the epilaryngeal region (near the false folds) that produces a more favorable input impedance of the vocal tract.

As a practical application, SPL can be estimated from inverse-filtered glottal source waveforms by first estimating the approximate glottal power (equation 19 in the text), then adding a vocal tract gain (about 10 dB at 100 Hz rising to about 15 dB at 400 Hz), and finally adding (or subtracting) a formant ripple (±1 dB at 100 Hz rising to ±5 dB at 400 Hz) that depends on whether a harmonic of the source is tuned or detuned by the first formant.

Acknowledgements

This work was supported by grant No. DC 00387-03 from the National Institutes of Health, a grant from The Voice Foundation, The Fulbright Commission, and Svärdska Stiftelsen. The authors thank Linnie Southard and Julie Lemke for manuscript preparation.

References

Ananthapadmanabha, T. (1984). Acoustic analysis of voice source dynamics. Quarterly progress and status report STL-QPSR2-3. Royal Institute of Technology (KTH), Stockholm, Sweden, 1-52.

Bouhuys, A., Mead, J., Proctor, D., & Stevens, K. (1968). Pressure flow events during singing. Ann. NY Acad. Sci., 155, 165-176.

Fant, G. (1960). The acoustic theory of speech production. Moutan & Co., The Hague, The Nether lands.

Fant, G., Liljencrants, J., & Lin, Q. (1985). A four parameter model of glottal flow. Quarterly Progress and Status Report STL-QPSR 4, Speech Transmission Laboratory, Royal Institute of Technology (KTH) Stockholm, Sweden, 1-22.

Fant, G. (1982). Preliminaries to analysis of the human voice source. Quarterly Progress and Status Report STL-QPSR 4. Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden.

Fant, G., & Lin, Q. (1987). Glottal voice source-vocal tract acoustic interaction. Quarterly progress and status report STL-QPSR 1. Royal Institute of Technology (KTH), Stockholm, Sweden, 13-27.

Flanagan, J. (1972). Speech analysis, synthesis, and perception. New York: Springer Verlag.

Fujimura, O., & Lindqvist-Gauffin, J. (1971). Sweep-tone measurements of vocal tract characteris tics. J. Acoust. Soc. Amer., 49, 541-558.

Gauffin, J., & Sundberg, J. (1989). Spectral correlates of glottal voice source waveform characteris tics. J. Speech Hear. Res., 32, 556-565.

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. J. Acoust. Soc. Am., 84(2), 511-529.

Isshiki, N. (1964). Regulatory mechanisms of voice intensity variation. J. Speech Hear. Res., 7, 17-29.

Kinsler, L., & Fry, A. (1950). Fundamentals of acoustics. New York: Wiley.

Ladefoged, P., & McKinney, N. (1963). Loudness, sound pressures, and subglottal pressure in speech. J. Acoust. Soc. Am., 35, 454-460.

Morse, P. (1948). Vibration and sound. Reprinted for the Acoustical Society of America by the American Institute of Physics, New York (1976).

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. J. Acoust. Soc. Am., 53(6), 1632-1645.

Schutte, H. (1980). The efficiency of voice production. Groningen: Kemper.

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

Sundberg, J., Scherer, R., & Titze, I. (in press). Loudness regulation in male singers. Journal of Voice.

Titze, I.R. (1988b). Regulation of vocal power and efficiency by subglottal pressure and glottal width. In O. Fujimura (Ed.), Vocal fold physiology: Voice production, mechanisms, and functions (pp. 227-238). New York: Raven Press.

Titze, I.R. (in review). Phonation threshold pressure: A missing link in glottal aerodynamics. J. Acoust. Soc. Amer.

NCVS Status and Progress Report - 2
April 1992, 39-58

Voice Source Model for Continuous Control of Pitch Period

Paul H. Milenkovic, Ph.D.
Department of Electrical and Computer Engineering, The University of Wisconsin-Madison

Abstract

The voiced speech waveform may be synthesized by exciting an LPC vocal tract filter with a pulse waveform patterned after naturally occuring glottal airflow pulses. Such a pulse waveform may be generated by computing samples of a piecewise polynomial curve at equally spaced time intervals. In this type of synthesis, the pitch period is commonly restricted to an integer multiple of the sample interval. We present a method for removing this restriction, permitting both pulse duration and pitch period to be varied over continuous time. Aliasing distortion is prevented by computing the sample values of pulses that have been low pass filtered in continuous time prior to sampling. Applications of this technique include modelling glottal pulses by least squares fit to inverse filter waveforms, the synthesis of calibration waveforms for evaluating measures of speech waveform jitter, the perceptual evaluation of low levels of waveform jitter, and the synthesis of the singing voice.

Introduction

Voiced speech may be synthesized by applying excitation pulses to a digital filter according to the source-filter model of speech production. The natural glottal pulse contributes zeroes to the speech spectrum (Mathews *et al.*, 1961) which are understood to have perceptual relevance. Patterning the shape of the excitation pulses after the glottal pulses observed by inverse filtering was reported by Rosenberg (1971) and by Holmes (1973) to improve the naturalness of synthetic speech.

We have reported (Milenkovic, 1986) on the use of pulses described by piecewise polynomial curves to model the voice source. The voice source signal in this instance refers to the first derivative of glottal airflow as obtained by inverse filtering the speech waveform measured with a pressure microphone; the model pulses were patterned after the shape of the first derivative of the glottal airflow pulse. The parameters controlling the pulse shape were adjusted to fit the model to the inverse filtered speech waveform in a least squares sense.

Fully natural sounding speech synthesis may need to account for the effects of source-tract interaction. The vocal tract driving point impedance has a transform zero at DC related to the inertia of the air column, and this impedance zero accounts for the skewing of the glottal airflow pulse relative to a more symmetric glottal opening area pulse (Rothenberg, 1983). The vocal tract impedance also has peaks at the formant frequencies; these peaks result in formant ripple being added to the glottal airflow pulse (Ananthapadmanabha and Fant, 1982).

Recent examples of source interactive synthesizers are reported by Sondhi and Schroeter (1988) and by Pinto *et al.* (1989). The Sondhi-Schroeter synthesizer is based on an articulatory model that produces a vocal tract area function from which both the vocal tract transfer function as well as the driving point impedance may be computed. The synthesizer described by Pinto *et al* is based on a formant model where the formant frequencies are used to specify an equivalent circuit model of the driving point impedance.

Source interactive speech synthesis is restricted in its application by the need for articulatory information. The formant frequencies do not specify a unique vocal tract area function (Atal *et al*, 1978), and the different area functions with the same formant frequencies produce nearly similar transfer functions but widely different driving point impedances (Milenkovic, 1984). While the impedance peak frequencies are coincident with the formant frequencies, the impedance zero frequencies are not set by the formant frequencies, and the location of these zeroes has a profound effect on the amplitudes of the impedance peaks affecting the formant ripple.

The effect of source-tract interaction on the acoustic waveform can, to a degree, be reproduced by a source-filter synthesizer. The skewing of the glottal airflow pulse is easily incorporated into a mathematical function describing a pulse shape. Formant ripple, however, is related to the dissipation of the formant oscillations into losses in the glottal constriction and subglottal acoustic system (Flanagan, 1972). Klatt and Klatt (1990) suggest using a pitch synchronous adjustment of bandwidths in a formant synthesizer to approximate this effect. On the other hand, the LPC multi-pulse synthesis method (Atal and Remde, 1982) is able to match the acoustic speech waveform in a least-squares sense using an LPC vocal tract filter that is fixed over one or more pitch period cycles. A model combining an LPC derived filter with a small number of source parameters may have sufficient degrees of freedom to approximate acoustic waveforms influenced by source-tract interaction.

Efforts to improve the quality of source-filter synthesized speech remain a worthwhile endeavor. In light of work in source-interactive synthesizers, the source-filter model is still employed to study the influence of the voice source on synthesis quality (Childers and Lee, 1991). While the source-filter synthesizer may never be as parsimonious as a source interactive synthesizer, the parameters controlling a source-filter synthesizer may be more readily determined by analysis of the acoustic speech waveform in the absence of x-ray measurements of vocal tract geometry. While the source-filter synthesizer may never give a perfect reproduction of natural speech waveforms, we seek to remedy sources of artifact where possible.

The process of fitting pulses to the voice source in discrete time is a source of artifact we seek to remedy. Fitting synthesizer pulses to the inverse filter waveform takes place in discrete time because that is the form in which we acquire speech waveforms for analysis. Because of discontinuous

derivatives at the points of glottal opening and glottal closure, the synthesizer pulses will undergo aliasing distortion upon computing sample values. This aliasing distortion becomes evident when the pulse endpoints are not aligned with the sample points. The endpoints of the naturally occuring glottal pulses, however, are not necessarily aligned with the sample points.

We have modelled a glottal pulse with an endpoint lying between two sample points as a weighted average of pulses with endpoints placed at sample points on either side of the desired endpoint (Milenkovic, 1986). The pulses in the model need to be low pass filtered well below the Nyquist frequency to make this averaging of pulses bracketing the desired endpoint a reasonable approximation to a pulse with the desired endpoint. This scheme gave good least squares fits between synthetic and natural glottal pulses. The pitch period estimates obtained by using the pulse weights to interpolate between sample points gave values for jitter well below the sample interval, jitter being the cycle-to-cycle flucuation in the pitch period. While the resynthesized speech waveform matched the large scale oscillations of the dominant formants in the natural speech waveform, the high frequencies were missing on account of the low pass filter used to interpolate pulse endpoints.

We focus attention on devising a glottal pulse model of controllable bandwidth and spectrum slope that can be aligned between sample points. One solution would be to oversample the synthesizer pulse shape, digitally low pass filter, and then resample at the rate of the digital speech recording as reported by Hartwell and Prezas (1981). The pulse endpoints would still be aligned on sample intervals, but the sample intervals would be much shorter. Another solution would be to put fillets or smooth corners on the pulse endpoints. This method was demonstrated by Titze (1986) to improve the naturalness of a synthesized singing voice. Of course the smooth corners significantly modify the voice source spectrum as well as reduce the aliasing artifact.

We have developed a technique for analytically applying a low pass anti-alias filter to pulses described by piecewise polynomial curves. The filter is effectively applied in continuous time before the sample values are obtained. We are able to compute sample values of polynomial pulse that is low pass filtered close to the Nyquist frequency, allowing continuous control of pulse duration and pitch period with minimal levels of aliasing distortion. We also have control over the spectral shape of the anti-alias filtering.

While the derivation of the technique may seem complicated, the actual computation of these pulse shapes is quite straightforward. Sample values are obtained by multiplying vectors containing powers of time *t*, sine, and cosine values by matrices of constant coefficients. The computations compare quite favorably with computing pulse shapes by oversampling and digital low pass filtering. There is no design compromise between time resolution and complexity because the method gives continuous control over pulse shape and pulse duration.

While the new technique for computing excitation pulses is motivated by fitting excitation pulses to the inverse filter derived voice source, other applications include synthesizing low jitter speech waveforms for evaluating measures of voice aperiodicity (Milenkovic, 1987; Yumoto *et al.*, 1982; Muta *et al.*, 1988; Titze *et al.* 1987; Hillenbrand, 1987), the perceptual evaluation of low levels of jitter on voice quality (Hillenbrand, 1988), and the synthesis of the singing voice (Titze, 1986).

Methods

We begin by deriving a formula for computing samples of a pulse waveform that has been low-pass filtered. We represent a pulse waveform which starts at $t=0$ and ends at $t=T$ as the piecewise polynomial function $p(t)u(t) - p(t)u(t-T)$ where $u(t)$ is the unit step function and where the polynomial $p(t)$ has components

$$p(t) = w_1 p_1(\tau) + \ldots + w_m p_m(\tau), \tag{1}$$

for $\tau = t/T$ and

$$p_i(\tau) = p_{i0} + p_{i1}\tau + \ldots + p_{in}\tau^n. \tag{2}$$

Synthesizer excitation waveform samples $p_{t_d}[n]$ for a pitch period beginning at t_d will be obtained according to $p_{t_d}[n] = p(n\,t_s - t_d)$ where t_s is the time between samples. The parameter T controls pulse length while the parameters w_1 through w_m control pulse shape.

The pulse waveform may be expressed as

$$p(t)u(t) - p(t)u(t-T) = wPT_n u_t^T - wPBT_n u_{t-T}^T, \tag{3}$$

where w is the row vector of pulse shape parameters w_i, P is the matrix of polynomial weights p_{ij}, T_n is a diagonal matrix of elements $1, T^{-1}, \ldots, T^{-n}$, u_t is the vector $[u(t), t\,u(t), \ldots, t^n u(t)]$, u_{t-T} is the vector $[u(t-T), (t-T)\,u(t-T), \ldots, (t-T)^n u(t-T)]$, and B is a lower triangular matrix of binomial coefficients b_{ij} coming from the expansion

$$\begin{aligned} \tau^i &= (1 + (\tau - 1))^i \\ &= b_{i0} + b_{i1}(\tau - 1) + \ldots + b_{ii}(\tau - 1)^n. \end{aligned} \tag{4}$$

Anti-alias filtering

We compute samples of the pulse at discrete times $t = n\,t_s$. In order to suppress aliasing, we need to low pass filter to a frequency at or below $1/(2t_s)$. This low pass filtering needs to be performed in continuous time prior to computing sample values of the filtered pulse. To show how this is done, we note that the filtered pulse

$$\begin{aligned} p_h(t) &= h(t) * \{p(t)u(t) - p(t)u(t-T)\} \\ &= h(t) * \{p(t)u(t)\} - h(t) * \{p(t)u(t-T)\} \\ &= wPT_n h_t^T - wPBT_n h_{t-T}^T, \end{aligned} \tag{5}$$

where * denotes convolution, h_t is the vector $[h(t)*u(t), h(t)*t\,u(t), \ldots, h(t)*t^n u(t)]$ and h_{t-T} is the vector $[h(t)*u(t-T), h(t)*(t-T)u(t-T), \ldots, h(t)*(t-T)^n u(t-T)]$. Hence, filtering a piecewise polynomial pulse reduces to evaluating $h(t)*t^i u(t)$ where $h(t)$ is the filter impulse response. To evaluate $h(t)*t^i u(t)$, we define the linear operator

$$\mathcal{L}\{g(t)\} = \int_{-\infty}^{t} g(u)du, \tag{6}$$

we note that

$$\begin{aligned} u(t) &= \mathcal{L}\{\delta(t)\} \\ t^i u(t) &= i!\mathcal{L}^{i+1}\{\delta(t)\} \end{aligned} \tag{7}$$

and that

$$h(t) * t^i u(t) = i!\mathcal{L}^{i+1}\{h(t)\}. \tag{8}$$

This last expression may be evaluated in closed form for any filter impulse response *h(t)* that is *i* times integrable.

The frequency sampling design (Gold and Jordan, 1969, Rabiner *et al.*, 1970) gives a finite duration impulse response that is closed form integrable. The impulse response has the formula

$$h(t) = h_0 + h_1 \cos(\pi t/t_h) + \ldots + h_k \cos(k\pi t/t_h), \quad -t_h < t \leq t_h, \tag{9}$$

where impulse response *h(t)* has duration $2\ t_h$.

We evaluate $i!\mathcal{L}^{i+1}\{h(t)\}$, essentially the normalized *i* fold integration of *h(t)*, in the regions $t < -t_h$, $-t_h < t \leq t_h$, and $t_h < t$. In first region, the expression evaluates to zero. In the second region, we will obtain a linear combination of sines and cosines as well as powers of *t* through t^{i+1}. In the third region, *h(t)* is zero valued, $\mathcal{L}\{h(t)\}$ will be a constant, and successive integrations will generate powers of *t* through t^i. It follows that

$$h_t = \begin{cases} 0 & t < -t_h \\ Cc_t^T + Ss_t^T + Ut_{n+1}^T & -t_h < t \leq t_h \\ Vt_n^T & t_h < t \end{cases}, \tag{10}$$

$$h_{t-T} = Cc_{t-T}^T + Ss_{t-T}^T + Ud_{n+1}^T, \quad T - t_h < t \leq T + t_h, \tag{11}$$

where we define the vectors

$$\begin{aligned} c_t &= [\cos(\pi t/t_h), \ldots, \cos(k\pi t/t_h)] \\ s_t &= [\sin(\pi t/t_h), \ldots, \sin(k\pi t/t_h)] \\ t_{n+1} &= [1, t, \ldots, t^{n+1}] \\ t_n &= [1, t, \ldots, t^n] \\ c_{t-T} &= [\cos(\pi(t-T)/t_h), \ldots, \cos(k\pi(t-T)/t_h)] \\ s_{t-T} &= [\sin(\pi(t-T)/t_h), \ldots, \sin(k\pi(t-T)/t_h)] \\ d_{n+1} &= [1, (t-T), \ldots, (t-T)^{n+1}] \end{aligned}$$

and where *C*, *S*, *U*, and *V* are matrices of constant coefficients.

The constant coefficient matrices may be computed by noting that rows i of matrices C, S, U, and V contain the coefficient weights of cosines, sines, and powers of t making up $i!\mathcal{L}^{i+1}\{h(t)\}$. That $i!\mathcal{L}^{i+1}\{h(t)\} = i\,\mathcal{L}\{(i\text{-}1)!\,\mathcal{L}^{i}\{h(t)\}\}$ means row i may be computed from row i-1 by following the rules for integrating sines, cosines, and powers of t

$$\int \sin(j\pi t/t_h)dt = -(t_h/j\pi)\cos(j\pi t/t_h) + K \tag{12}$$

$$\int \cos(j\pi t/t_h)dt = (t_h/j\pi)\sin(j\pi t/t_h) + K \tag{13}$$

$$\int t^{j-1}dt = (1/j)t^j + K, \tag{14}$$

where K is the constant of integration in each expression. Matrix row $i= 0$ is thus given by

$$\begin{array}{lcll} C_{0j} & = & 0 & 1 \le j \le k \\ S_{0j} & = & (t_h/j\pi)h_j & 1 \le j \le k \\ U_{01} & = & h_0 & \\ U_{0j} & = & 0 & 1 < j \le n+1 \\ V_{0j} & = & 0 & 1 \le j \le n \end{array} \tag{15}$$

while rows $i=1$ through $i=n$ are given by

$$\begin{array}{lcll} C_{ij} & = & -(it_h/j\pi)S_{i-1\,j} & 1 \le j \le k \\ S_{ij} & = & (it_h/j\pi)C_{i-1\,j} & 1 \le j \le k \\ U_{ij} & = & (i/j)U_{i-1\,j-1} & 1 \le j \le n+1 \\ V_{ij} & = & (i/j)V_{i-1\,j-1} & 1 \le j \le n. \end{array} \tag{16}$$

The values of the integration constants K are obtained by forcing h_t to be continuous at $t=-t_h$ and $t=t_h$. From Equation 10, this condition is satisfied if $C\,c_t^T + S\,s_t^T + U\,t^T_{n+1} = 0$ at $t=-t_h$ and if $C\,c_t^T + S\,s_t^T + U\,t^T_{n+1} = V\,t_n^T$ at $t=-t_h$. It follows that for rows $i=0$ through $i=n$

$$-U_{i0} = \sum_{j=1}^{n+1}(-t_h)^j U_{ij} + \sum_{j=1}^{k}(-1)^j C_{ij} \tag{17}$$

$$V_{i0} = \sum_{j=0}^{n+1}(t_h)^j + \sum_{j=1}^{k}(-1)^j C_{ij} - \sum_{j=1}^{n+1}(t_h)^j V_{ij}. \tag{18}$$

Combining Equations 5, 10, and 11, the filtered polynomial pulse reduces to

$$p_h(t) = \begin{cases} c_0 c_t^T + s_0 s_t^T + u_0 t_{n+1}^T & -t_h < t \le t_h \\ v t_n^T & t_h < t \le T - t_h \\ v t_n^T - (c_1 c_{t-T}^T + s_1 s_{t-T}^T + u_1 d_{n+1}^T) & T - t_h < t \le T + t_h \end{cases}, \quad (19)$$

with coefficient vectors

$$\begin{aligned} c_0 &= w P T_n C & s_0 &= w P T_n S & u_0 &= w P T_n U \\ v &= w V \\ c_1 &= w P B T_n C & s_1 &= w P B T_n S & u_1 &= w P B T_n U. \end{aligned}$$

These coefficient vectors need to be computed for each pulse of new duration T and shape parameters w.

Once T and w are fixed, the filtered polynomial pulse consists of two "fillets", the regions $-t_h < t \le t_h$ and $T-t_h < t \le T+t_h$ where the pulse is constructed from cosines, sines, and powers of t, together with the region $t_h < t \le T-t_h$ where the pulse is simply a polynomial, of the same form as the original $p(t)$ but where the polynomial weights have been adjusted. The computational complexity compares favorable with, for example, simply "tacking on" fillet regions to the original polynomial pulse $p(t)$. The new technique, however, offers greater control over the spectral shape of the resulting pulse.

Basis functions

We control the shape of the piecewise polynomial polynomial pulse by adjusting the w vector. The elements of the w vector are weights for combining basis function polynomials $p_i(\tau)$ to obtain a representation of the first derivative of the glottal pulse. As we indicated earlier, we will be modelling an inverse filter signal that is the first derivative of the estimated glottal airflow.

The form of the basis functions is derived from a series of assumptions about the glottal pulse shape. We assume that the glottal flow starts at zero at the leading edge of the glottal pulse and ends at zero at the trailing edge. In the case where the glottis does not close completely on account of a glottal chink (Bless *et al.*, 1986), we assume that glottal flow starts and ends at the same constant value. We will regard the glottal waveform to be a smooth function except at the points of glottal opening and closing, where the derivatives may be discontinous.

The dominant aspects of the excitation signal influencing the speech spectrum that we incorporate into our model are the pulse duration T together with the discontinuities in the first two derivatives of the glottal wave at the glottal pulse endpoints. In the case of a glottal pulse which has a continuous first derivative between its endpoints, the asymptotic frequency spacing of the Laplace transform zeroes is controlled by the pulse duration T while the asymptotic damping of these zeroes is controlled by the ratio of the closing edge to the opening edge slopes (Mathews, *et al.*, 1961; Flanagan, 1972). Our model will control the slope as well as the local curvature of glottal airflow at each pulse endpoint in order to better fit observed inverse filter waveforms.

The slope and curvature of the discontinuities of the glottal pulse are equivalent to the step and slope discontinuities of the glottal pulse derivative at its endpoints. We seek four basis functions of minimum polynomial order to exercise independent control over the step and slope discontinuities at

the two endpoints. That the glottal pulse starts and ends at the same level imposes the contraint that the integrated value of each of the four basis functions be zero. The minimum polynomial order meeting these conditions is *n=4:* each polynomial has five coefficients.

The four basis functions are polynomials $p_i(\tau)$ through $p_4(\tau)$ for $\tau = t/T$ where we specify boundary values

$$\begin{aligned} p_1(0) &= 1 \qquad p'_2(0) = 1 \\ p_3(1) &= 1 \qquad p'_4(1) = -1, \end{aligned} \tag{20}$$

where the remaining boundary values are set to zero. We add the contraint

$$\int_0^1 p_i(\tau)d\tau = 0. \tag{21}$$

For each $p_i(\tau)$, evaluating Equation 2 at the boundaries gives

$$\begin{aligned} p_i(1) &= p_{i0} + p_{i1} + p_{i2} + p_{i3} + p_{i4} \\ p'_i(1) &= p_{i1} + 2p_{i2} + 3p_{i3} + 4p_{i4} \\ p_i(0) &= p_{i0} \\ p'_i(0) &= p_{i1} \\ 0 &= p_{i0} + \tfrac{1}{2}p_{i1} + \tfrac{1}{3}p_{i2} + \tfrac{1}{4}p_{i3} + \tfrac{1}{5}p_{i4}. \end{aligned} \tag{22}$$

Solving this linear system of equations for i=1 to i=4 results in

$$P = \begin{bmatrix} 1 & 0 & -18 & 32 & -15 \\ 0 & 1 & -4.5 & 6 & -2.5 \\ 0 & 0 & -12 & 28 & -15 \\ 0 & 0 & -1.5 & 4 & -2.5 \end{bmatrix}, \tag{23}$$

where P is the matrix of elements p_{ij}, and

$$PB = \begin{bmatrix} 0 & 0 & -12 & -28 & -15 \\ 0 & 0 & -1.5 & -4 & -2.5 \\ 1 & 0 & -18 & -32 & -15 \\ 0 & -1 & -4.5 & -6 & -2.5 \end{bmatrix}, \tag{24}$$

where B is the lower triangular matrix of binomial coefficients introduced in Equation 4. Remembering that P is the matrix of coefficients of the polynomials p_i expressed as functions of τ and PB is the matrix of coeffcients of the same polynomials expressed as functions of τ - 1, the symmetry properties $p_i(-\tau) = p_3(\tau - 1)$ and $p_2(-\tau) = p_4(\tau - 1)$ account for the observed similarities between P and PB.

Representative synthesizer pulses are depicted in Figure 1; the Fourier transform of these pulses are seen in Figure 2. The pulse weights are w_1=50, w_2=0, w_3=-300, and w_4=1000, and the sampling

rate is 10 kHz. The pulses were produced under three conditions. The first condition is with no filter. The second condition is with frequency sampling filter t_h=0.6ms, h_0=0.5, h_1=1, h_2=1, h_3=1, h_4=0.3904, which is a single transition sample low pass design. The third condition is with the filter t_h=0.6 ms, h_0=0.5, h_1=0.5, which is a filter with a raised cosine impulse response. This third condition makes the smoothing effect of the filter more visually apparent.

Close examination of the glottal flow pulses in Figure 1, obtained by integrating the sampled glottal flow derivative, reveals that the integrated pulse synthesized without the filter does not start and finish at the same level of closed glottis flow. The integrated pulses synthesized with an anti-alias filter do indeed start and finish at the same level. The anti-alias filter helps the sampled pulse of the glottal flow derivative to maintain the same zero-DC constraint imposed on the continuous time version of the pulse.

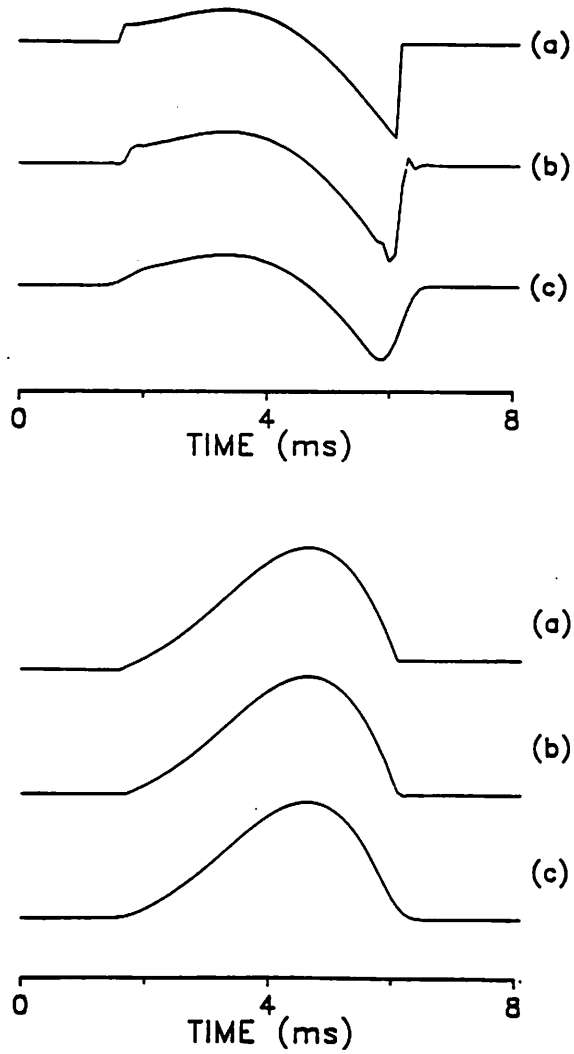

Figure 1: Synthesized glottal pulses: a) unfiltered, b) frequency sampling design anti-alias filter, and c) raised cosine filter. The top panel is glottal flow derivative, and the bottom panel is glottal flow.

Results

We demonstrate three applications of the proposed voice source pulse model. The first is modelling the glottal wave by matching the pulse model to the inverse filtered acoustic waveform. The second is the use of the pulse model to synthesize waveforms used in the calibration of measures of voice perturbation. The third is the perceptual evaluation of low levels of voice perturbation in speech synthesis.

Modelling the glottal wave

An estimate of the glottal excitation waveform is obtained by LPC inverse filter analysis of the acoustic speech waveform. We recorded the acoustic wave with a Panasonic model WM-063T electret microphone cartridge (Digi-Key part P9932) AC coupled to a preamplifier with a high pass cutoff frequency of 2 Hz. The microphone is held in close proximity to the subject's lips (5 cm) by a boom attached to a pair of safety glasses worn by the subject. The microphone is held off-axis to avoid air-blast artifact. Placing the microphone in the near field maximizes the ratio of direct sound energy to the energy of sound reflected from the room, minimizing artifact resulting from reverberation.

The output of the microphone preamplifier was low pass filtered at 4 kHz with a 4-pole Butterworth filter and sampled at 8.33 kHz using an A/D with 12 bits resolution. The digital speech recording is high pass filtered by two applications of a single pole high pass digital filter with a 20 Hz cutoff. The high pass filter is first applied to the data in the forward direction and then is applied

to the data in the backward direction, resulting in a 2-pole filter with zero phase (Milenkovic, 1987). Use of a zero phase high pass filter minimizes the waveform droop, making the distinction between the open and closed phases of the glottal pulse more apparent.

The inverse filter is determined by applying 10 coefficient covariance method LPC analysis to the speech wave, preemphasized by taking the first difference (Markel and Gray, 1976). This analysis is applied to a 20 ms interval of the waveform. Applying the resulting LPC inverse filter to the unpreemphasized speech waveform recorded with a pressure sensing microphone gives an estimate of the first derivative of the glottal airflow (Milenkovic, 1987). Integrating the inverse filter signal gives an estimate of the AC component of glottal flow.

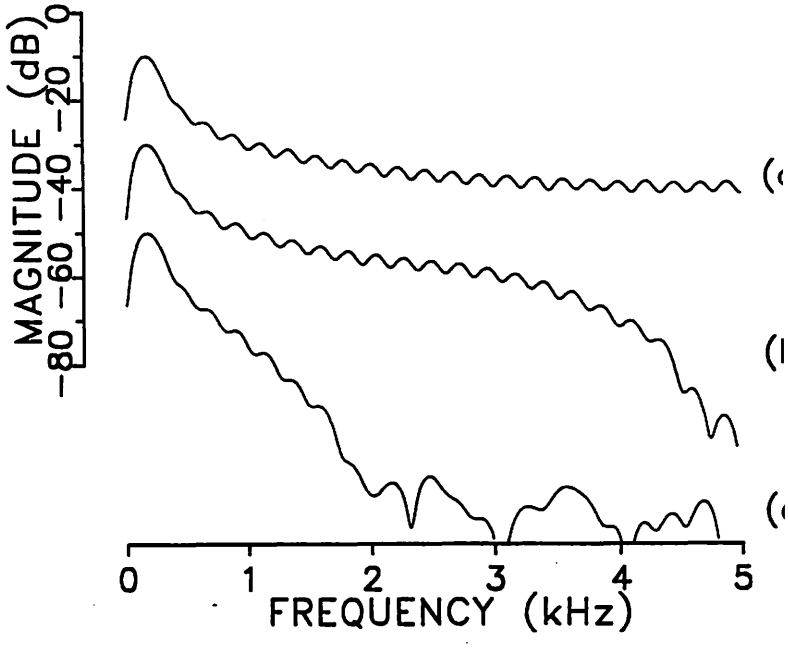

Figure 2: Fourier magnitude spectrum of the glottal flow derivative for synthesized pulses. Synthesis conditions: a) unfiltered, b) frequency sampling design anti-alias filter, and c) raised cosine filter.

We collected data from four male and four female subjects in order to obtain representative waveforms. We instructed our subjects to sustain the vowel /a/ for 1 second. We employed the vowel /a/ because the high first formant frequency of this vowel reduces the artifact caused by misalignment of the inverse filter zeroes with the vocal tract poles. Figure 3 shows both Fourier as well as LPC spectra of the preemphasized speech signals for one male and two female subjects. We show one female subject having a spectrogram similar to the male subject. We show another female subject with a spectrogram that is most dissimilar to the male subject, exhibiting the elevated first harmonic together with broadened formant bandwidths described as typifying female vowels (Klatt, 1987; Klatt and Klatt, 1990).

A computer program identified successive pitch periods by comparing the large downward stroke of the inverse filter signal against a threshold. The downward stroke identifies the approximate location of glottal closure. The computer program then fit a 5 parameter pulse model to the inverse filter waveform in each pitch period. The 5 parameters are t_1 and t_2, the times of glottal opening and glottal closing, as well as w_1, w_3, and w_4, the weights of basis polynomials p_1, p_3, and p_4. We excluded p_2 by fixing w_2 at zero as we were able to obtain good fits without this basis function.

In fitting the pulse model to the inverse filter waveform, we employed a raised cosine filter with parameters $t_h = .24$ *ms*, $h_0 = 0.5$, $h_1 = 0.5$. This filter deemphasizes the higher frequencies and has been empirically selected for its effectiveness in the pulse fitting process. One may readily change the h coefficients to substitute a broader band filter for later speech synthesis using the fitted glottal pulses.

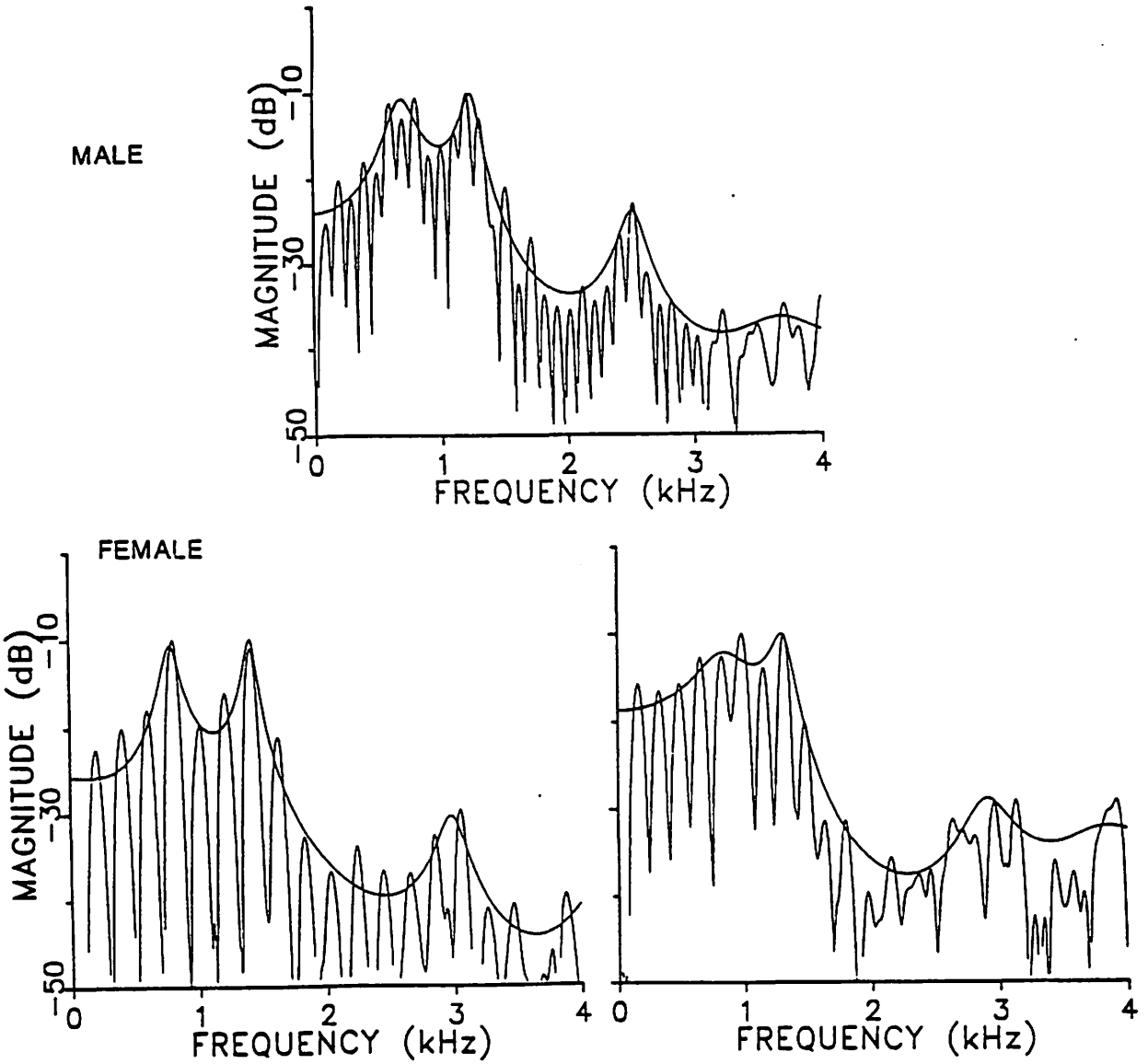

Figure 3: Fourier and LPC spectra of the vowel /a/.

The parameters t_1 and t_2 may be adjusted over continous time because of the anti-alias property of the method for computing pulse samples. For each trial value of t_1 and of t_2, we obtain values of w_1, w_3, and w_4 by linear regression of the inverse filter waveform against the scaled, filtered, and time shifted basis polynomials p_1, p_3, and p_4. We first step through values of t_2 with t_1 held fixed, then step through values of t_1 with t_2 held fixed in order to find values of t_1 and t_2 that give the best least squares regression on the linear weights w_1, w_3, and w_4. We repeat this process, each time reducing the step size used to adjust t_1 and t_2, until we obtain optimal values for all five parameters.

The results of fitting the pulse model are shown in Figure 4. With the second female subject, we observed cycle-to-cycle fluctuation in the value of t_1, the glottal opening time. Is is not unique to females; we show results for a male subject exhibiting this condition in Figure 5. This figure includes a contour plot, indicating contour lines of equal least squares regression on the w's for different values of the endpoints t_1 and t_2. The contour plot was obtained by sampling t_1 and t_2 on a grid

ten times denser than the .12 ms sample spacing and by employing the Surfer software package (Golden Software, Golden, Colorado). This contour plot indicates that there are multiple local minima in t_1, and these minima are of nearly the same magnitude. This suggests that the precise moment of glottal opening is ambiguous given the degrees of freedom in our model and the degree to which the model fits inverse filter waveforms. This ambiguity is inherent in the fit between the model and the data and does not appear to be an artifact of the search procedure used to find optimal values of t_1 and t_2.

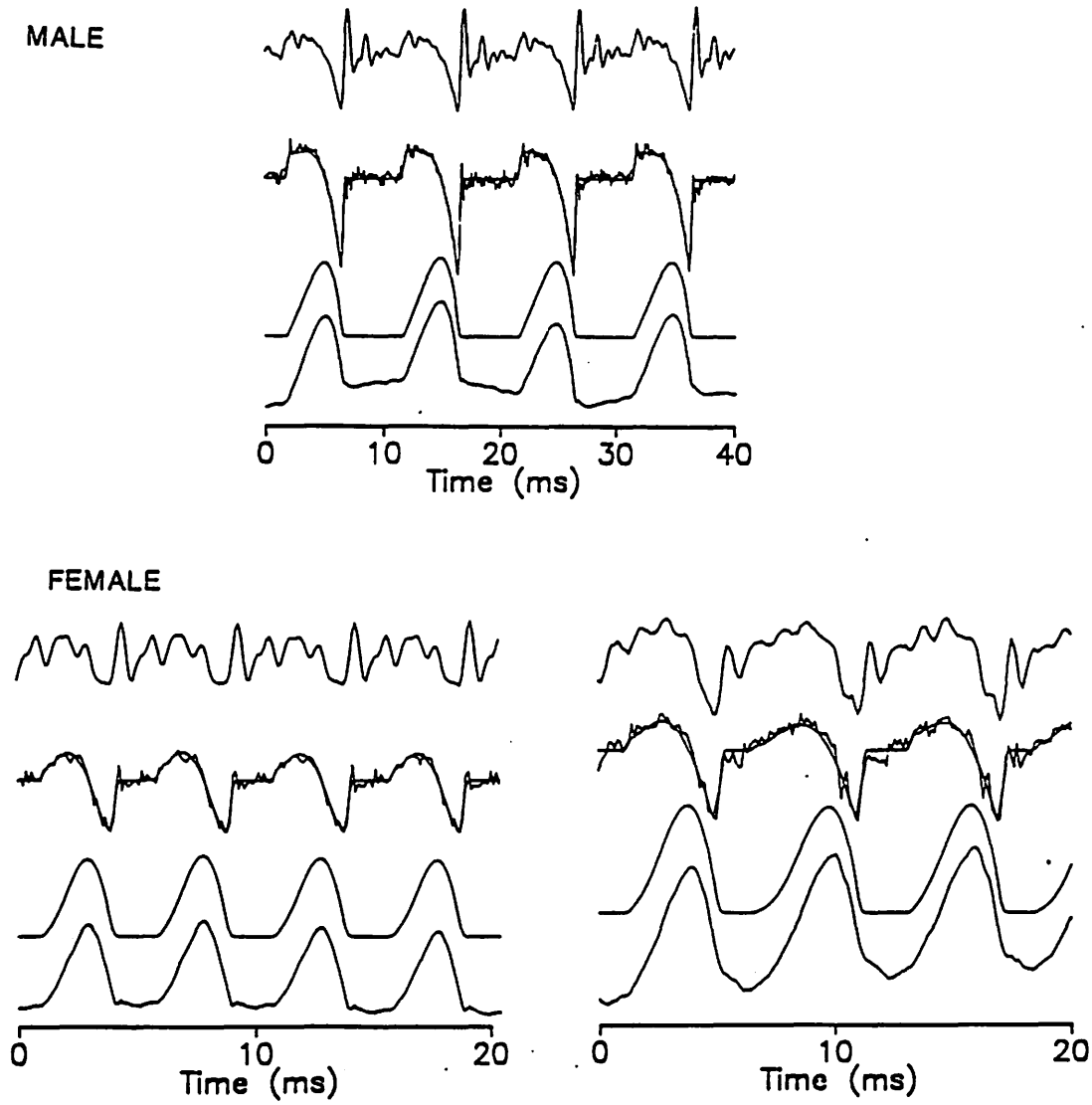

Figure 4: Voice source waveforms: upper) free field acoustic pressure, upper middle) superimposed pulse model and inverse filter estimate of the first derivative of glottal airflow, lower middle) integrated pulse, and lower) inverse filter estimate of glottal flow.

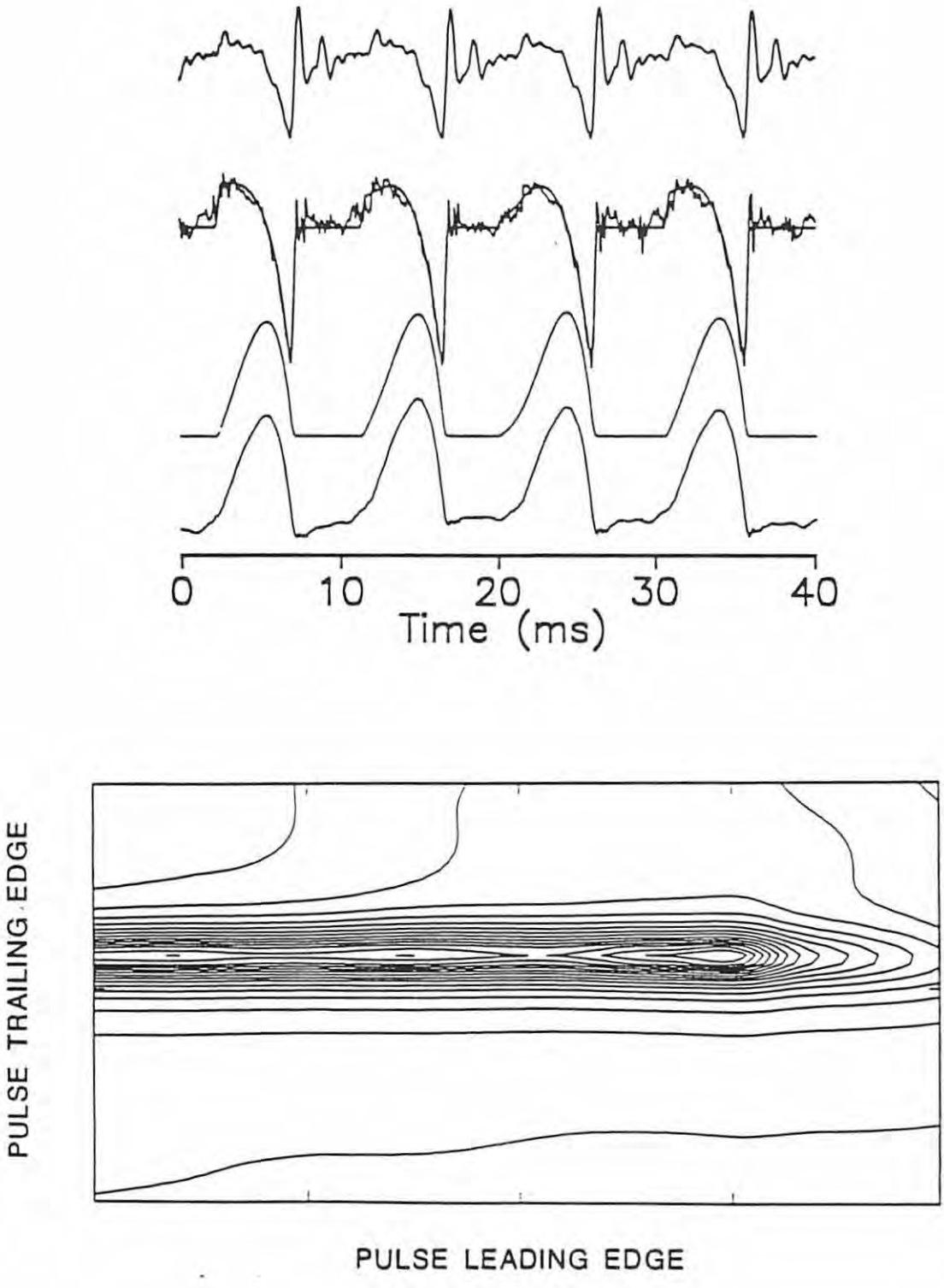

Figure 5: Male subject exhibiting ambiguity in the time of glottal opening. The bottom panel is a contour plot of the least squares fit between the pulse model and the glottal flow derivative as a function of the pulse endpoints. The tick marks are placed one sample point (.12 ms) apart.

Calibration of voice perturbation

We used the voice source pulse model to synthesize constant pitch synthetic vowels for purposes of calibrating measures of jitter and shimmer. The naturally occuring jitter may be as low as 10 μs while shimmer may be under 1 percent (Milenkovic, 1987). Ideally, one would like a synthetic voice source where jitter and shimmer may be controlled to well below these levels in order to determine the noise floor on the algorithms used to measure jitter and shimmer. This synthetic voice source should be capable of producing pitch periods that are not restricted to an integer number of sample points because the natural voice source does not have this restriction.

We employed the same pulse shape and h(t) filter depicted in Figure 1. The pulse weights were set at w_1=50, w_2=0, w_3=-300, and w_4=1000, we fixed the open portion of the glottal pulse at .55 the pitch period, and we selected a sampling rate of 10 kHz. The anti-alias filter was given the specifications t_h=0.6 ms, h_0=0.5, h_1=1, h_2=1, h_3=1, h_4=0.3904, which is a single transition sample low pass design. The voice source waveform produced in this manner was fed into a 12 coefficient LPC synthesizer. The LPC coefficients were obtained by covariance method analysis applied to a 20 ms duration of the preemphasized speech waveform, applied to the vowel /a/ produced by a male subject.

Three synthesis conditions were used. The first condition is synthesis without the use of the h(t) filter where the pulse endpoints were rounded to the nearest sample point; this is the pitch synchronous condition. The pitch periods were then adjusted to maintain the average pitch period value. The second condition is synthesis without the h(t) filter where the voice source pulse is computed asynchronously for a pitch period that is not an integer number of sample points. The voice source pulse is treated as a continuous time waveform which is sampled without regard for the resulting aliasing. In the third condition, the voice source pulse is computed asynchronously, but this time with the benefit of the anti-alias filter h(t).

Synthesis pitch periods were selected from two ranges. The range around 8 ms is representative of a male voice while the range around 4 ms is representative of a female voice. For each range, pitch period values were selected that were non-integer numbers of sample points. The values 8.05 ms and 4.05 ms are midway between integer multiples of the sampling interval, and we expect synthesis jitter to be greatest at these values.

Table 1: Measured jitter and shimmer for the synthesized vowel /a/ sampled at 10 kHz.

	Pitch synchronous		**Pitch asynchronous**			
	No filter		No filter		Low pass filter	
Pitch period (ms)	Jitter (μs)	Shimmer (percent)	Jitter (μs)	Shimmer (percent)	Jitter (μs)	Shimmer (percent)
8.02	35	.31	34	2.27	0	.13
8.05	70	.60	69	3.23	0	.09
4.02	42	3.28	40	3.61	1	.15
4.05	87	8.23	86	3.46	1	.18

The values of jitter and shimmer are obtained from a 250 ms interval of a synthesized vowel of constant pitch; the analysis was performed according to a method described by Milenkovic (1987). Results are summarized in Table 1. In this analysis procedure, jitter and shimmer are obtained by computing the short term autocorrelation function and using parabolic interpolation to find the peak. The interpolated peak of the short term autocorrelation function provides an accurate measure of pitch period from which to compute jitter, and the amplitude of this peak is used to compute shimmer.

In the synthesis conditions without an h(t) filter we observe considerable shimmer, the value of which varies with the pitch range as well as the synthesis method. In the pitch synchronous synthesis without the *h(t)* filter, shimmer is attributable to the jittered overlap of the vocal tract impulse response between successive pitch periods. Considerable reduction in both jitter and shimmer is obtained by using the h(t) filter to compute the voice source pulses.

Perceptual effect of aliasing artifact

The question arises whether the reduction in jitter and shimmer afforded by the new source model is perceptually significant. We present results of listening tests on the ability of subjects to detect differences between vowel tokens synthesized with and without the anti-aliasing filter. These tests constitute a demonstration that the proposed voice source model has perceptual relevance.

We synthesized 500 ms tokens of the vowel /a/ sampled at 10 kHz according to the procedure of the previous section. We applied an exponential taper to the amplitude envelope of both the onset and offset of each vowel token. The pulse amplitudes at onset were given by $A_0(1 - .5^n)$ where A_0 is the steady state pulse amplitude and n is the pitch period number after onset. The pulse amplitudes at offset were given by $A_0 (.8)^n$ where n is the pitch period number after offset.

We compared two synthesis conditions. The first condition was pitch asynchronous with the anti-alias filter. The second condition was pitch synchronous without the anti-alias filter, pitch synchronous being the case where we round the pulse endpoint values to the nearest sample point. In this second synthesis condition, we followed the pulse synthesis with the application of a digital filter to match the spectral shaping of the first synthesis condition.

The two synthesis conditions have closely matched spectra as seen in Figure 6, eliminating spectrum envelope difference as a perceptual clue. The two spectra, however, differ in the amount the harmonic peaks are separated from the apparent noise floor; this difference is inherent in the two synthesis methods as this spectrographic effect is not readily compensated by post filtering. We seek to determine if such spectrum differences are perceptually apparent.

We synthesized three different vowel tokens. The first was a constant low pitch with a pitch period of 8.05 ms. The second was a falling high pitch with an initial pitch period of 4 ms and a final pitch period of 4.4 ms. The third was a falling low pitch with an initial pitch period of 8 ms and a final pitch period of 8.8 ms.

We first conducted informal listening evaluation of these tokens. The informal evaluation indicated that the first condition, with a true constant pitch period, gave a less rough sound quality than the second condition, with a constant average pitch period maintained by aligning pulse endpoints at the nearest sample points. The three tokens, as we have ordered them, turn out to be ranked in order of most apparent to least apparent for this effect.

The differences between synthesis conditions were apparent when the tokens were output over a loudspeaker. The differences vanished when the tokens were output over headphones. Hillenbrand

(1988) reported the same effect in listening evaluation of low levels of voice perturbation. Wilde *et al.* (1986) indicated that the roughness of jittered sounds is most apparent in a reverberant sound field or in sound presented over headphones with artificial reverberation. On the basis of these results, we decided to conduct our formal listening evaluation free field over a loudspeaker.

We employed 6 male subjects, ages 18 to 30 years, for our formal listening trials. We presented pairs of vowel tokens over a loudspeaker placed 6 feet from the subject. The listening was performed in a room typical of an office environment. The pairs of tokens were random selections from the set of ordered pairs *{AA,AB,BA,BB}* where *A* denotes the first synthesis condition and *B* denotes the second. The two 500 ms tokens were separated by 200 ms. After each presentation, the subject indicated Same, Different, or Unable to Decide by the appropriate key press on a computer keyboard. The next token pair was presented as soon as the subject indicated a response for the previous pair.

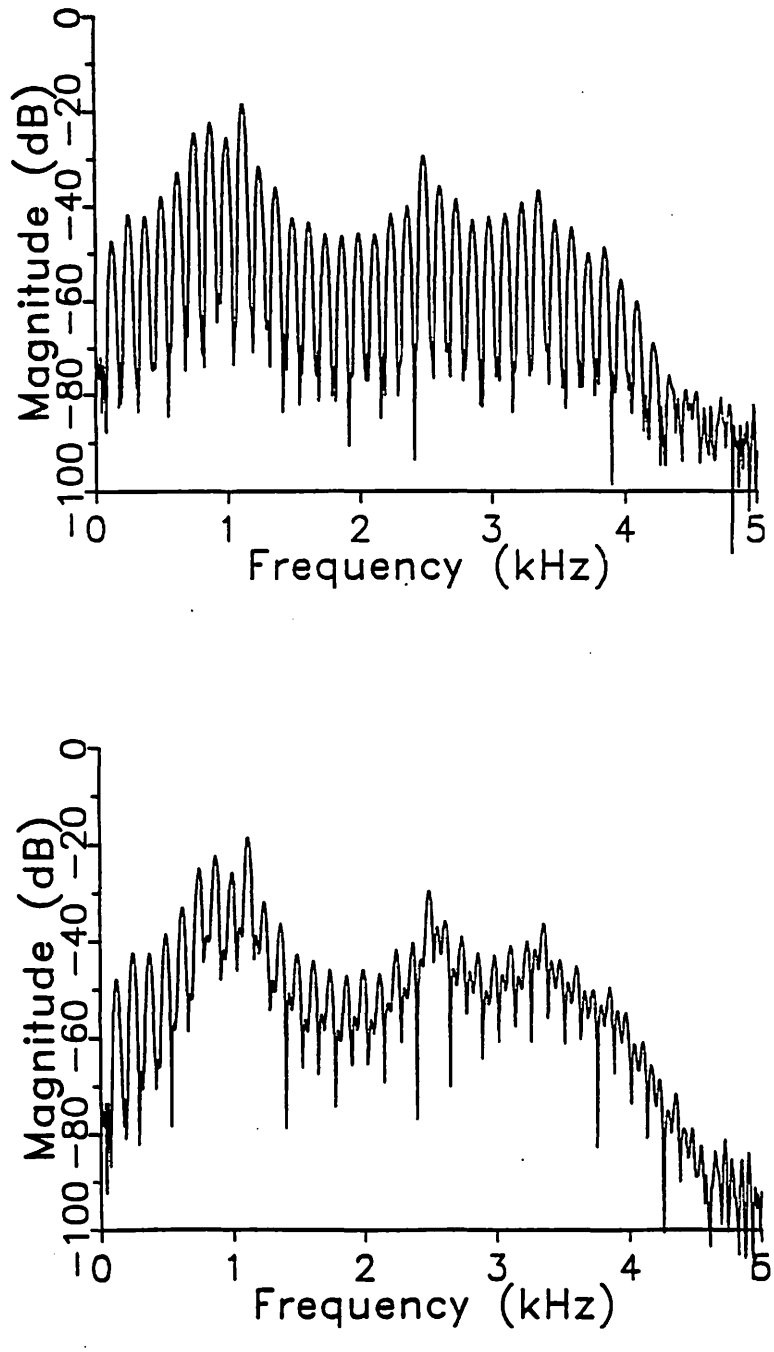

Figure 6: Fourier analysis of the synthetic vowel /a/. Use of the anti-alias filter to synthesize a vowel with a pitch period that is a non-integer multiple of the sample interval is shown at the top. Rounding the glottal pulse endpoints to the nearest sample position and then post filtering to match the spectrum envelope is shown at the bottom.

The subjects were instructed to listen for a smooth/rough distinction in the vowel tokens. They were not given any prior training in terms of practice trials. Each subject was presented with 100 trials of each of the three vowel tokens, starting with constant low pitch and ending with falling low pitch. The results are summarized in Table 2. A result of 66 or more correct answers is significant at $p < .001$ for the binomial distribution. The first token was correctly discriminated by all 6 subjects, the second by 5 subjects, and the third by 4 subjects. The

subject who failed to discriminate the second token also failed to discriminate the third token, consistent with our ordering of the difficulty of the tasks.

Table 2. Discriminations between aliased voice source and anti-alias filtered voice source made by 6 subjects.

	Constant low pitch			Falling high pitch			Falling low pitch		
	C	I	A	C	I	A	C	I	A
Subject									
1	96	2	2	97	1	2	75	18	7
2	94	4	2	95	2	3	80	14	6
3	93	6	1	96	4	0	74	26	0
4	82	13	5	98	1	1	58	40	2
5	98	1	1	99	0	1	86	8	6
6	98	2	0	64	36	0	43	50	7

C: Correct; I: Incorrect; A: Abstain

Discussion

The proposed scheme for synthesizing voice source pulses with endpoints controllable between sample times offers definite advantages when fitting pulses to the inverse filter waveform. It permits us to treat the glottal opening and closing times as real instead of integer parameters in optimization algorithms for determining the glottal opening and closing times that produce an optimal fit. The more precise specification of the pulse endpoints can improve the fit of the source model to the inverse filter waveform. It also aids in mapping the least square error surface as a function of the pulse endpoint locations, giving insight into the location of local minima and the effect of these minima on an optimization algorithm.

The map of the least square error as a function of pulse endpoints indicates an inherent ambiguity in the glottal opening transient. The glottal closing transient is marked by a strong global optimum of the error. The glottal closing transient, however, can be marked by multiple local minima of nearly the same least square error. This ambiguity of the glottal opening transient suggests that the pulse model has more degrees of freedom than can be resolved from one pitch period cycle rather than the fault being with the optimization algorithm. Imposing a smooth variation in the glottal opening time between pitch period cycles might be a way to resolve this ambiguity.

We were able to fit inverse filter waveforms of both male and female subjects with our pulse model, even though we have not encorporated the exponential tail following the glottal closing transition, a feature of the LF model (Fant and Lin, 1988). The effect of this exponential tail is to change the spectral tilt of the speech signal. Klatt and Klatt (1990) propose to add the spectral tilt by filtering a source pulse without the exponential tail; the filtered pulse will exhibit such a tail. In our case, where covariance LPC analysis is applied to an analysis frame containing at least one pitch period, this spectral tilt is incorporated into the vocal tract filter instead of the source pulse.

The acoustic speech waveform is readily resynthesized by exciting an LPC vocal tract filter with the sequence of pulses extracted from the inverse filter waveform. The voice source pulses can have

a bandwidth close to the Nyquist bandwidth for the sample rate used. The perceptually significant jitter and shimmer that results from aligning the pulse endpoints at discrete sample positions is eliminated.

In synthesis applications where one is not required to align each pitch period with the pitch period in a natural utterance, the pitch period could be rounded to the nearest integer number of sample intervals. While perceptual tests with constant pitch sounds suggest a pitch discrimination threshold of .3 Hz (Flanagan and Saslow, 1958), experiments on sounds with varying pitch indicate a threshold of 2 Hz (Klatt, 1973), and experiments with sentences indicate a threshold of 4 Hz (Mack and Gold, 1986).

These studies of pitch discrimination, however, address the issue of pitch shifts between pairs of syllable or sentence length utterances. Pitch fluctuations between successive pitch periods, known as jitter, is readily perceptable as roughness at the level of 1 percent (Hillenbrand, 1988) (1 Hz for a pitch of 100 Hz). The jitter levels of natural speech have average values on the order of .5 percent (Horii, 1979) and have been measured as low as .1 percent (Milenkovic, 1987). Fine control of the pitch period is required to produce jitter at or below these levels.

While Klatt and Klatt (1990) argue against the low levels of jitter in natural speech being perceptually significant in sentence productions, they incorporate a pitch "flutter" into their synthesizer, a sinusoidal variation in pitch over a long time interval. The accurate synthesis of flutter would benefit from our proposed technique. It can be readily demonstrated that slow pitch changes in the Klatt synthesizer (Klatt, 1980) can be heard as a series of steps, and our technique can eliminate this effect. The fine control of pitch is also important in the synthesis of the singing voice to eliminate beats when combining synthesized singing with other musical sounds.

The ability to exercise continuous control of pitch is an important tool in both the perceptual as well as algorithmic measurement of voice perturbation. One no longer needs to make a judgement as to what level of oversampling is adequate in perceptual experiments of pitch and of jitter and shimmer. The proposed synthesis method can also provide an accurate standard for comparing different methods of measuring jitter and shimmer.

References

Ananthapadmanabha, T. V., and Fant, G. (1982). "Calculation of true glottal flow and its components," Speech Communication 1, 167-184, North Holland Publishing Company.

Atal, B. S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique," J. Acoust. Soc. Am. 63, 1535-1555.

Atal, B. S., and Remde, J. R. (1982). "A new model of LPC excitation for producing natural-sounding speech at low bit rates," IEEE ICASSP Proceedings, 614-617.

Bless, D. M., Biever, D., and Shaikh, A. (1986). "Comparisons of vibratory characteristics of young adult males and females," Proceedings of the International Conference on Voice, Kurume, Japan, Vol. 2, 46-54.

Childers, D. G., and Lee, C. K. (1991). "Vocal quality factors: Analysis, synthesis, and perception," J. Acoust. Soc. Am. 90, 2394-2410.

Fant, G., and Lin, Q.-G. (1988). "Frequency domain interpretation and derivation of glottal flow parameters," STL-QPSR 2-3, 1-21.

Flanagan, J.L. (1972). *Speech Analysis, Synthesis and Perception* (Springer-Verlag, Berlin), 2nd ed..

Flanagan, J. L., and Saslow, M. G. (1958). "Pitch discrimination in synthetic vowels," J. Acoust. Soc. Am. 30, 435-442.

Gold, B., and Jordan, K. L. (1969). "A direct search procedure for designing finite duration impulse response filters," IEEE Trans. Audio Electroacoust. AU-17, 33-36.

Hartwell, W. T., and Prezas, D. P. (1981). "A pulse driving function generator for LPC synthesis of voiced segments of speech," IEEE Trans. Acoust. Speech Signal Proc. ASSP-29, 1113-1116.

Hillenbrand, J. (1987). "A methodological study of perturbation and additive noise in synthetically generated voice signals," J. Speech Hear. Res. 30, 448-461.

Hillenbrand, J. (1988). "Perception of aperiodicities in synthetically generated voices," J. Acoust. Soc. Am. 83, 2361-2371.

Holmes, J. N. (1973). "The influence of glottal waveform on the naturalness of speech from a parallel formant synthesizer," IEEE Trans. Audio Electroacoust. AU-21, 298-305.

Horii, Y. (1979). "Fundamental frequency perturbation observed in sustained phonation," J. Speech Hear. Res. 22, 5-19.

Klatt, D. H. (1973). "Discrimination of fundamental frequency contours in synthetic speech — implications for models of pitch perception," J. Acoust. Soc. Am. 53, 8-16.

Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. 67, 971-995.

Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality varia tions among female and male talkers," J. Acoust Soc. Am. 87, 820-857.

Mack, M., and Gold, B. (1986). "The effect of linguistic content upon the discrimination of pitch in monotone stimuli," J. Phonetics 14, 333-337.

Markel, J. D., and Gray, A. H., Jr. (1976). *Linear Prediction of Speech,* (Springer-Verlag, Berlin).

Mathews, M. V., Miller, J. E., and David, E. E., Jr. (1961). "Pitch synchronous analysis of voiced sounds," J. Acoust Soc. Am. 33, 179-186.

Milenkovic, P. (1984). "Vocal tract area functions from two-point acoustic measurements with formant frequency constraints," IEEE Trans. Acoust. Speech Signal Proc. ASSP-32, 1122-1135.

Milenkovic, P. (1986). "Glottal inverse filtering by joint estimation of an AR system with a linear input model," IEEE Trans. Acoust. Speech Signal Proc. ASSP-23, 28-42.

Milenkovic, P. (1987). "Least mean square measures of voice perturbation," J. Speech Hear. Res. 30, 529-538.

Muta, H., Baer, T., Wagatsuma, K., Muraoka, T., and Fukuda, H. (1988). "A pitch—synchronous analysis of hoarseness in running speech," J. Acoust. Soc. Am. 84, 1292-1301.

Pinto, N. B., Childers, D. G., and Lalwani, A. L. (1989). "Formant speech synthesis: improving production quality," IEEE Trans. Acoust. Speech Signal Process. ASSP-37, 1870-1887.

Rabiner, L. R., Gold, B., and McGonegal, C. A. (1970). "An approach to the approximation prob lem for nonrecursive digital filters," IEEE Trans. Audio Electroacoust. AU-18, 83-106.

Rosenberg, A. E. (1971). "Effect of glottal pulse shape of the quality of natural vowels," J. Acoust. Soc. Am. 49, 583-590.

Rothenberg, M. (1983). "An interactive model for the voice source," in *Vocal Fold Physiology,* D. M. Bless and J. H. Abbs, eds. (College Hill Press, San Diego).

Sondhi, M. M., and Schroeter, J. (1988). "A hybrid time-frequency domain articulatory synthe sizer," IEEE Trans. Acoust. Speech Signal Proc. ASSP-35, 955-967.
Titze, I. R. (1986). "Three models of phonation," J. Acoust Soc Am. 79(S1), S81-S82.
Titze, I. R., Horii, Y., and Scherer, R. C. (1987). "Some technical considerations in voice perturbation measurements," J. Speech Hear. Res. 30, 252-260.
Wilde, M. D., Martens, W. L., Hillenbrand, J., and Jones, D. R. (1986). "Externalization mediates changes in the perceived roughness of sound signals with jittered fundamental frequencies," Proceedings of the 1986 International Computer Music Conference, the Hague.
Yumoto, E., Gould, W. J., and Baer, T. (1982). "Harmonics-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. 71, 1544-1550.

Acknowledgement

This work was supported by grant P60 DC00976 from the National Institute on Deafness and Other Communication Disorders.

NCVS Status and Progress Report - 2
April 1992, 59-81

Coordination of the Respiratory and Laryngeal Systems in Breathing and Vocalization

Pamela J. Davis, Ph.D.
Cumberland College of Health Sciences, The University of Sydney, Australia
Donald Bartlett, Jr., M.D.
Department of Physiology, Dartmouth Medical School
Erich S. Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Introduction

In organisms as complex as vertebrates, structures serving a particular physiological system often play roles in other major systems as well. Obvious examples are the shared equipment of the digestive and respiratory tracts at the head end of the organism and of the genital and urinary tracts at the other end. Shared structures of this sort require a sort of biological time-sharing with regard to function. One can breathe or swallow, but not at the same time without embarrassing consequences.

A more subtle co-ordination of functions must exist within the system that we use for both respiration and phonation. Breathing and vocalization are not mutually exclusive, but are interdependent. The co-ordination of movements for these two purposes and the integration of reflexes that serve them constitute an important problem in regulatory physiology. Since the larynx is not only a respiratory valve, but also the organ of speech and song, we hope the discussion will be of interest to students of both systems.

Because vocal and respiratory science have evolved nearly independently, it is not surprising that certain responses that actually influence both systems are historically entrenched as the property of one or the other. A good example is the coordinated set of reflex responses mediated by pulmonary stretch receptors. These are the well-known Hering-Breuer reflexes, which have been firmly associated with the regulation of respiratory tidal volume for more than a century (Breuer, 1970). Despite this association, these reflexes appear to be at least as effective in maintaining upper airway

patency as in regulating lung volume (Sica, et al., 1984; Kuna, 1986; Bartlett and St. John, 1988), and as noted later in the chapter, they may well play a role in the co-ordination of vocalization as well.

Our purpose is to examine some factors that influence both breathing and phonation. We begin by presenting a brief review of the respiratory physiology of the larynx. We then report some recent ideas and findings about the influence of "respiratory" reflexes during evoked vocalizations in animals and the apparent gating of protective reflexes during phonation. The closing section consists of comments and discussion about the importance of reflex control of motor activity during speech and other complex movements.

Respiratory Physiology of the Larynx

In is early evolution, the larynx arose as a respiratory valve that contributed critically to the emergence of air-breathing in amphibians and reptiles (Bartlett, 1989). In humans and other mammals it has undergone further evolutionary development in relation to vocalization, but it retains important respiratory functions even in these species.

Laryngeal Breathing Movements

Laryngeal movements in concert with breathing were noted by Galen (Galen, 1968), but serious interest in the larynx as a respiratory organ has only developed fairly recently. Animals that hold their breath for long periods for diving or for postural stabilization use the larynx as a valve for complete closure of the airway (Bartlett and Birchard, 1983; Boelaert, 1941; Olsen, et al., 1969; Scholander, 1940). It is a very good valve when used in this way; pressures of over 250 mmHg may be sustained briefly during coughing or weightlifting in humans (Compton, et al., 1973; Sharpey-Schafer, 1953).

The larynx is not just an open or closed valve, however, even if we continue to view it as a respiratory rather than a vocal organ. The larynx acts as a variable resistor, which regulates airflow in and out of the lungs. Most variations in laryngeal resistance occur at the level of the vocal folds: contraction of the posterior cricoarytenoid (PCA) muscles abducts the folds and lowers resistance, whereas contraction of the thyroarytenoids (TA) and other vocal fold adductors narrows the glottic slit and raises resistance (Proctor, 1964, 1980; Sasaki, 1984). The cricothyroid (CT) muscle has a complex action, tilting the thyroid cartilage ventrally and caudally with respect to the cricoid, thus lengthening and slightly adducting the vocal folds (Arnold, 1961). Simultaneous CT and PCA activation renders the glottic airway slightly larger than during PCA activation alone (Horiuchi and Sasaki, 1978; Konrad and Rattenborg, 1969).

Extrinsic mechanisms may alter laryngeal resistance in some circumstances (Fink, 1974, 1975; Fink, et al., 1956), but most evidence indicates that intrinsic muscle activity is much more important (Brancatisano, et al., 1983, 1984). The balance between abductor and adductor activities varies greatly. Considerable relative adduction of the vocal folds may result from relaxation of the PCA muscles even in the absence of adductor muscle activity. In some studies, the vocal folds have been found to be more widely separated during all phases of the respiratory cycle than they are after muscle paralysis or death, suggesting a net abductor influence, even during expiration (Bartlett, 1979; Bartlett, et al., 1973; Semon, 1891). Adductor activity is clearly present in some circumstances, but the most important and consistent mechanism underlying the respiratory movements is the phasic activity of the intrinsic abductor muscles, the PCAs (Brancatisano, et al., 1984).

During quiet breathing in humans, inspiration and expiration are fundamentally different in mechanical and energetic terms. Inspiration is powered by the diaphragm and other muscles of the ventilatory bellows, and the rate of inspiratory airflow is determined by the force generated by these muscles and by the resistance and elastance of the system. The larynx may influence inspiration to the extent that it influences resistance (Bartlett, et al., 1973; England, et al., 1982). Normally the vocal folds are widely separated during inspiration, and dynamic passive collapse of the laryngeal airway is prevented by the cricoid ring, so laryngeal resistance is low and has little influence on inspiratory flow (Brancatisano, et al., 1983; England, et al., 1982).

In contrast, the energy that drives expiratory flow is recovered from potential energy stored in the system during the previous inspiration. Expiratory flow is determined by the recoil pressure of the respiratory system and its resistance rather than by ongoing muscle activity. In most circumstances, there is a remarkable matching between the duration of expiratory flow and the time before the next inspiration, so there is no end-expiratory pause. The inherent time constant for passive emptying of the tidal volume is short (Brody, 1954; Gautier, et al., 1973). Thus the matching of expiratory flow to expiratory time is achieved by a variable retardation of expiratory flow by post-inspiratory pliometric contraction of inspiratory muscles during early expiration (Brody, 1954; Gautier, et al., 1973; Gesell and White, 1938; Murphy, et al., 1959; Petit, et al., 1960), and by expiratory narrowing of the glottis (Bartlett, et al., 1973; England, et al., 1982). The laryngeal braking of expiratory airflow is most prominent at rest and is much reduced in conditions with increased airflow, such as exercise, hypercapnia and voluntary hyperventilation (Bartlett and Knuth, 1984; Bartlett, et al., 1973; England and Bartlett, 1982; Murakami and Kirchner, 1972). Infants use the laryngeal braking mechanism extensively (Fisher, et al., 1982; Harding, 1980; Mortola, et al., 1982). Their expiratory duration is brief so their end-expiratory lung volume is advantageously held above the passively determined functional residual capacity (Kosch and Stark, 1984; Mortola, 1987).

As might be expected, stimuli that influence breathing have effects on the respiratory movements of the larynx as well. As already noted, increases in ventilation resulting from hypercapnia or exercise are associated with a widely patent glottic airway, owing to increased expiratory PCA activity (Bartlett, 1979; Bartlett, et al., 1973; Campbell, et al., 1963; England and Bartlett, 1982; McCaffrey and Kern, 1980; Murakami and Kirchner, 1972). By contrast, stimulation of the carotid chemoreceptors by hypoxia or cyanide, which also increases ventilation, causes expiratory narrowing of the glottis under some conditions (Bartlett, 1980; Dixon, et al., 1974; England, et al., 1982). The physiological significance of this response is not clear, but it may contribute to the increase in end-expiratory volume that has been reported to occur in hypoxic states (Bouverot and Fitzgerald, 1969).

Pulmonary receptors with afferents in the vagus nerves have profound effects on breathing movements, including those of the larynx. Lung inflation inhibits or delays inspiratory bursts of respiratory muscle activity, including those of the PCA. Inflation also elicits a sustained, tonic PCA activation, however, in concert with the activity of the abdominal expiratory muscles (Bartlett, et al., 1973). In this setting, the PCA shows reflex responses that are characteristic of both inspiratory and expiratory muscles. This is appropriate as both inspiratory and expiratory airflows traverse the glottic airway.

Another approach to determining the reflex effects of lung inflation is to prevent inflation during a single respiratory cycle, usually by temporary occlusion of the airway. This maneuver, which was first used by Breuer (Breuer, 1970), results in vigorous inspiratory PCA activity, reflecting inflation-induced inhibition of this activity during unhindered cycles (Cohen, 1975; Kuna, 1986; Sica, et al., 1984; Van Lunteren, et al., 1984). These responses appear to depend on slowly adapting pulmonary stretch receptors.

Pulmonary irritant receptors and C-fibers also have reflex connections with laryngeal muscles. Stimulation of these receptors enhances expiratory glottic narrowing or may induce sustained closure of the laryngeal airway.

The laryngeal responses to stimulation of intercostal afferents are complex and depend on the anatomic level at which the stimuli are applied (Remmers, 1973; Sherrey, and Megirian, 1974). These responses may have more to do with postural than with ventilatory control.

Laryngeal Afferents and Their Reflex Actions

The larynx has an unusually rich endowment of sensory nerve endings. The recurrent laryngeal nerves (RLNs) contain a few sensory fibers (Sampson and Eyzaguirre, 1964; Suzuki and Kirchner, 1969), but the overwhelming majority run in the internal branches of the superior laryngeal nerves (SLNs) (Mathew, et al., 1984; Widdicombe, 1986; Wyke and Kirchner, 1976). Sant'Ambrogio and his colleagues (1983) have noted that in the cat, this branch contains about 2200 myelinated sensory fibers (Dubois and Foley, 1934). For comparison, the entire cervical vagus contains about 3000 such fibers (Agostoni, et al., 1957).

A reasonable speculation would be that this extensive innervation has evolved in accordance with the requirements of sensory feedback for vocalization. This interpretation is called into question, however, by the fact that animals with little vocal ability, such as rabbits and snakes, also have large a number of sensory nerve fibers in the larynx (Bartlett, unpublished; Davis and Nail, 1987). These nerves may be important for breathing and airway protection as well as for vocalization.

Although conscious perception of sensations arising in the larynx is a common experience, remarkably little formal study of these sensations has been done (Widdicombe, 1986). Misplaced food particles and inhaled insects elicit characteristic sensations, presumably by stimulation of receptors in the laryngeal epithelium. Laryngeal sensations of uncertain origin accompany emotional experiences: Robert Frost (1963) wrote that a poem "begins as a lump in the throat," but he did not instruct us about the specific stimuli or pathways that are involved.

In contrast to the relative neglect of conscious perceptions arising in the larynx, many investigators have studied the responses of laryngeal receptors to a wide range of stimuli, chiefly by recording the activity of single afferent fibers in the SLN (Andrew, 1956; Boggs and Bartlett, 1982; Boushey, et al., 1974; Davis, 1986; Davis and Nail, 1987; Fisher, et al., 1985; Kirchner and Wyke, 1965; Mårtensson, 1964; Mathew, et al., 1984; Sampson and Eyzaguirre, 1964; Sant'Ambrogio, et al., 1986, 1983; 1985a, b; Shingai, 1977, 1979, 1980; Suzuki and Kirchner, 1968). The results of these studies are confusing owing both to the inherent complexity of the receptors and their responses, and to the use of different classification schemes by different investigators.

Mechanoreceptors

The most thoroughly described general category of laryngeal receptors are those responding to mechanical displacement or deformation. In their early study of SLN afferents, Sampson and Eyzaguirre (1964) identified "touch" receptors, which responded to light mechanical stimulation of the epithelium, and "deep" mechanoreceptors, which they suggested were in the laryngeal muscles or joints. Boushey and associates (1974) also identified two groups of receptors, but used different criteria. Their "group 1" fibers had little or no spontaneous activity and adapted quickly to sustained mechanical stimuli. "Group 2" units were spontaneously active and adapted slowly and incompletely to sustained stimuli.

More recently, Davis and Nail (1987) used a similar binary classification scheme, identifying "silent" and "tonic" receptors from among those SLN afferent fibers which could be stimulated by light touch of the laryngeal mucosa in the cat and rabbit. These fibers had small receptive fields, often less than 1 mm in diameter, and were most densely located on the apex of the arytenoid cartilage, the ary-epiglottic fold, and the base of the laryngeal surface of epiglottis.

These authors also provided a detailed description of the responses of touch-sensitive SLN afferents to static and dynamic mechanical stimulation with a servo-controlled deforming probe. These afferents displayed a remarkable dynamic sensitivity and could be entrained to discharge in a 1:1 manner with the probe at frequencies as high as 400 Hz for periods of 1 second when stimulated at probe amplitudes less than 100 m. They were generally very broadly tuned, in that they would respond over a wide range of stimulation frequencies within a sensitive range of stimulation amplitudes (<100 m). The dynamic sensitivity of these receptors from the various locations would ensure a high probability of the movement of a foreign particle into the larynx being signalled at most velocities and as described above, on most mucosal surfaces but in particular those around the laryngeal inlet. These observations led Davis and Nail (1987) to suggest that the dynamic sensitivity of touch-sensitive SLN afferents was appropriate for the detection of the ingress of foreign particles, and the provocation of defensive reflexes.

The most extensive analysis of SLN afferents has been carried out by Sant'Ambrogio and his colleagues (Mathew, et al., 1984; Sant'Ambrogio, et al., 1986, 1983, 1985a, b) using a spontaneously breathing animal preparation with a T-tube tracheotomy. This preparation enabled them to direct the respiratory airflow through the animal's upper airway or to divert it through the tracheotomy during different respiratory cycles, and to determine the responses of individual laryngeal receptors to changes in transmural pressure, airflow and laryngeal movements. "Pressure" receptors, similar to the "group 2" receptors of Boushey et al. and the "tonic" receptors of Davis and Nail, were found to increase their discharge in response to collapsing or distending transmural pressures. "Flow" receptors, so called because they responded to changes in laryngeal airflow, have subsequently been found to be "cold" receptors, responsive to epithelial cooling during inspiration.

Chemoreceptors

Some laryngeal receptors respond dramatically when water or other fluids are placed in the laryngeal lumen. The characteristic response is immediate, intense activity, which adapts only to a small extent, and is sustained until the offending liquid is washed out of the larynx with saline or some other non-stimulating fluid. The chemical basis for the activation of these "water receptors" varies with species, but seems to depend on the removal of chloride ions from the epithelial surface in dogs and rabbits (Boggs and Bartlett, 1982; Boushey, et al., 1974; Shingai, 1977, 1979).

The classification of receptors by stimulus modality is somewhat arbitrary as some receptors respond to both mechanical and chemical stimuli. As reported by Boushey et al. (1974) and Davis (1986), the activity of some "group 2" or "tonic" mechanoreceptors is inhibited by CO_2. A recent more extensive analysis of this phenomenon (Bartlett and Knuth, unpublished) indicates that many receptors in cats are inhibited by CO_2 concentrations as low as 3%, thus suggesting that receptor discharge is modulated by the CO_2 in expired air with every breath.

Reflex responses

The most prominent and characteristic reflex responses to laryngeal mucosal stimulation is immediate closure of the glottis (Bartlett, et al., 1981; Szereda-Przestaszewska and Widdicombe, 1973). This response protects the lower respiratory tract from contamination with undesirable materials. Sustained glottic closure is incompatible with breathing, however, and it is not surprising that protective closure of the glottis is often brief or incomplete and is coordinated with the breathing cycle.

Weak mechanical or chemical stimulation of the laryngeal epithelium exaggerates the expiratory adduction of the vocal folds, but may not result in complete airway occlusion (Suzuki and Sasaki, 1976, 1977a, b; Szereda-Przestaszewska and Widdicombe, 1973). More intense stimulation causes apnea and glottic closure, often interrupted by coughing (Jiménez-Vargas, et al., 1962; Szereda-Przestaszewska and Widdicombe, 1973). The constellation of responses may include bradycardia (Tomori and Widdicombe, 1969), bronchoconstriction (Boushey, et al., 1972; Nadel and Widdicombe, 1962; Tomori and Widdicombe, 1969) and increased respiratory tract mucus secretion (Phipps and Richardson, 1976). Apnea is particularly striking in neonatal animals, whereas coughing becomes more prominent with maturation.

Two other responses attributable to laryngeal receptors should be mentioned. One of these is the enhancement of inspiratory activities of several pharyngeal dilator muscles by collapsing transmural pressures in the upper airway (Mathew, 1984; Mathew and Farber, 1983; Mathew, et al., 1982, 1984a, 1984b; Van Lunteren, et al., 1984). This response, which is greatly reduced by SLN section, is well designed to relieve inspiratory pharyngeal obstruction, such as that occurring in obstructive sleep apnea. The specific receptors responsible for this response have not been defined, but the best candidates are the "group 2", "tonic" or "pressure" receptors.

Finally, intralaryngeal CO_2 has been shown to inhibit breathing (Bartlett, et al., 1990; Boushey and Richardson, 1973). Recent analysis of this response suggests that the inhibition of tonically active mechanoreceptors is probably responsible (Bartlett and Knuth, unpublished). This finding is unexpected since inhibition of breathing is the characteristic response to stimulation rather than inhibition of laryngeal receptors. The system is clearly more complicated than many have supposed, and much remains to be learned about it.

Reflex Effects on the Larynx During Vocalization

Successful vocalization, whether produced by a kitten signalling distress to its mother or an operatic tenor communicating the pain of lost love, depends upon the proper coordination of the laryngeal muscles and the respiratory system. Reflex effects produced by afferents from the respiratory system appear to be important to such coordination. Other laryngeal reflexes, however, such as those that protect the airway against the entry of foreign material, would be highly disruptive if they occurred during vocalization. They have to be controlled in some manner.

Effects Mediated by Receptors in the Lungs

Sensory input from the respiratory system has been shown to be capable of modifying the pattern of activity in laryngeal muscles during vocalization evoked by electrical stimulation of the midbrain tegmentum. Testerman (1970) was the first to demonstrate a role for pulmonary vagal afferents in the control of CT muscle activity during vocalization evoked by electrical stimulation of the midbrain tegmentum in cats. Garrett and Luschei (1987) subsequently demonstrated that the duration of TA and CT muscle activity during evoked vocalization in anesthetized cats was strongly influenced by either opening the trachea to the atmosphere or by occluding the proximal trachea. In contrast, the application of large subglottic pressure modulations during vocalization had no effect on the activity of these muscles.

More recently, it has been shown that electrical stimulation of a more circumscribed region of the midbrain, comprising the periaqueductal gray (PAG) and adjacent tegmentum (stereotaxic coordinates A 4.1 - P 0.9) elicits vocalization as a component of coordinated defense reactions (Bandler, 1988). A more precise localization of the PAG neurons mediating vocalization has been achieved using the technique of intracerebral microinjection of excitatory amino acids (Bandler, 1982; for a review see Bandler et al., 1991; Jürgens and Richter, 1986). These amino acids include the endogenous neurotransmitters L-glutamate and L-asparate, as well as other compounds such as D,L-homocysteic acid (DLH) and kainic acid. These chemicals excite central nervous system neurons and their dendritic processes directly, unlike electrical stimulation which has the disadvantage that is also activates axons of passage (Goodchild et al., 1982; Lipski et al., 1988).

A single microinjection of DLH at a dose of 40 nmol (in 200 nL of phosphate buffer, 0.02M, pH 7.4) into the lateral PAG of a freely moving or unanesthetized, precollicular decerebrate cat results in a sequence of episodic vocalization for 1-2 minutes. Injections of kainate acid at concentrations as low as 940 pmol will produce effects for 20 or 30 minutes (Bandler and Carrive, 1988; Carrive et al., 1987; Zhang et al., 1991). Vocalizations that can be elicited from the caudal two-thirds of the lateral PAG have a natural acoustic structure and are indistinguishable from spontaneous vocalizations. A typical EMG pattern, resulting from a microinjection within the lateral PAG (A1.6) which evoked a howling sequence is shown in Figure 1. Howling was accompanied by an increase in the activity of the thyroarytenoid (TA), cricothyroid (CT), and internal intercostal (IIC) muscles, the latter exposed by dissection of the overlying external layer in the 9-10 rib space. There was also an increase in inspiratory activity of the diaphragm (D) and posterior cricoarytenoid (PCA) muscles (Davis and Zhang, 1991).

Between each vocal episode there is one, or sometimes two, inspirations. The sequence of inspiration and vocalization is repeated throughout the duration of the effect. The peak tracheal air pressures range from approximately 4 cm H_2O in hissing to 40 cm H_2O in howling and the expiratory duration from 0.3 seconds in hissing to 2.0 seconds in howling. It is not known what determines the length of these vocalizations, but following on the previous observations (Garrett and Luschei, 1987; Testerman, 1970), it seems likely that the respiratory pattern associated with the vocalization evoked by midbrain stimulation is influenced by the sensory discharge from the respiratory system.

A recent study of vocalization evoked by DLH stimulation of the lateral PAG in the precollicular decerebrate, unanesthetized cat (Davis, Zhang, Carrive and Bandler, unpublished) has confirmed and extended these previous reports of the effects of pulmonary reflexes during evoked vocalization. In this study the air pressure in the lungs and airways was varied during episodes of evoked vocalization.

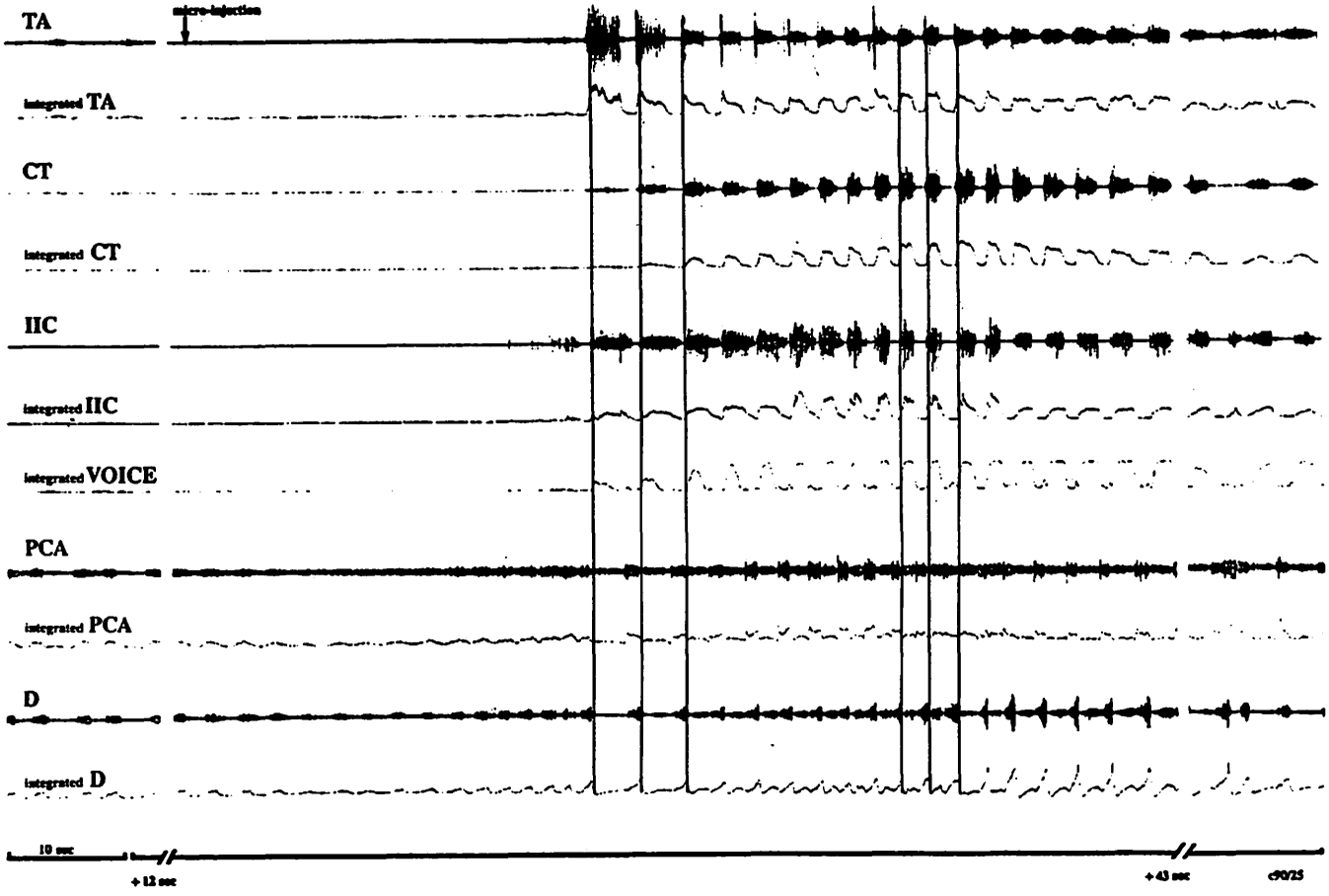

Figure 1. Laryngeal and respiratory EMG during a sequence of 38 howls evoked by DLH microinjection into the lateral PAG of the unanesthetized decerebrate cat. To facilitate comparison of the records, vertical lines marking several voice onsets have been added to the raw and "integrated" (0.1 s time constant) data from the posterior cricoarytenoid (PCA), thyroarytenoid (TA) cricoarytenoid (CT), internal intercostal (IIC), and diaphragm (D) muscles. The acoustic microphone signal of the voice has been smoothed, and 12 and 43 seconds of the record have been omitted as indicated on the figure. Time calibration: 10 seconds.

The application of static inflation pressures to the lungs during an episode of vocalization was associated with prolonged expiratory durations, an effect which was abolished by bilateral vagotomy. A range of inflation pressures was tested, encompassing most of the physiological range usually observed during the various types of vocalization. Frequently, the prolonged expiratory duration was associated with sustained discharges in the TA, CT, IIC and abdominal muscles, responses which at least for the two laryngeal muscles were dissimilar to control inflations delivered in a similar manner prior to DLH injection.

These findings indicate that the control of the expiratory duration during vocalization is linked to the pressure of the air in the lungs and airways. The discharge of vagal afferent fibers would be stimulated by the elevated transmural pressures associated with vocalization or stimulated during the inflation tests. Although vagal pulmonary stretch receptor afferent fibers transmit signals related to lung volume in humans and laboratory animals (Guz and Trenchard, 1971), human respiratory control appears to rely little upon the Hering-Breuer reflexes, at least for volumes below about 2 times normal eupneic tidal volume (Clark and von Euler, 1971). However, it is possible that these respiratory reflexes may be used to regulate the duration of vocalization according to the availability of air in the lungs.

Functionally, it is important to be able to vocalize for a duration determined by linguistic demands and by the availability of air in the lungs. There is evidence that during phonation, the inspiratory depth is related to the length of the intended utterance (Gelfer et al., 1983). Indeed, the effectiveness of some types of vocalization, such as singing or various species-specific calls, would be significantly impaired by inappropriately-timed inspiratory refills. In this context it is significant that vocalization has been found to be associated with an inhibition of the ventilatory response to CO_2 (Bunn and Mead, 1971; Phillipson et al., 1978).

In future studies it will be important to clarify the processes involved in ventilatory inhibition during speech. An animal model of vocalization may be well suited to such a study. A CNS-determined suppression of the ventilatory drive during vocalization would be functionally useful, as discussed above, and may also serve to enhance the pulmonary volume-related reflex inspiratory inhibition.

Stimulation of Respiratory Defensive Reflexes During Vocalization

Recent observations by Zhang, Bandler and Davis (unpublished) have indicated that an interaction occurs when mechanical probing of the larynx and trachea is carried out during vocalization evoked by DLH stimulation of the lateral PAG in the decerebrate cat. In this preparation, the tracheal cannula enabled the cat to breathe through its larynx while a length of tubing could be inserted through an otherwise occluded port, either rostrally to the larynx, or caudally into the trachea and lungs. Mechanical stimulation of the trachea to approximately the depth of the carina, a stimulus associated with coughing as previously reported (Widdicombe, 1954), was ineffective in provoking coughing during vocalization. Note on the tracheal pressure record in Figure 2 (part B, left panel) that coughing was not evoked during the period of vocalization following the DLH injection. In contrast, coughing was evoked by the same mechanical stimulation 120 seconds later (part B, right panel), when by comparison with the control test, it appeared likely that the vocalization would have subsided. Similar results were obtained by mechanical stimulation of the larynx in two other cats, however the adequacy of mechanical probing of the laryngeal mucosa was always uncertain during vocalization.

Additionally, as illustrated in Figure 2, during mechanical stimulation of the trachea there was a marked disruption to the sequence of evoked vocalization and to the pattern of the intercostal, diaphragm, and laryngeal muscle activity. In particular, the tracheal stimulation was associated with a recruitment of inspiratory-related CT muscle activity, activity which was not observed during DLH injections without mechanical stimulation, although CT exhibited expiratory bursts for the few vocalizations that were produced (see four large tracheal pressure peaks on the record). Although a regular sequence of howling and hissing was observed in the control sequence, there was a smaller number of irregularly-spaced vocalizations produced during the tracheal stimulation. The latency to vocalize was prolonged during tracheal stimulation from 20 seconds after the completion of the DLH injection in the control to 35 seconds during stimulation. In addition, the respiratory rate was faster during tracheal stimulation.

Different laryngeal reflexes, manifest as swallows, coughs, apnea or expiratory efforts were observed in response to laryngeal stimulation of sleeping and awake dogs (Sullivan et al., 1978) The type of reflex elicited was dependent on the dog's sleep-wakefulness state. It is possible that the interaction effect described above between vocalization and defensive reflexes represents a similar type of modulation of the sensitivity of defensive respiratory reflexes. Alternatively, the neural processes involved in the reception of irritant stimuli from the trachea and the motor pattern genera-

tion for coughing may be mutually inhibitory with that for vocalization. In this context it is important to note that the vocalization was evoked in decerebrate, unanesthetized cats by chemical stimulation of the PAG, and thus without the possible confounding effects of electrical stimulation. Obviously, it will be important in future research to ascertain the level of this apparent inhibition of defensive reflexes in higher order sensory neurons.

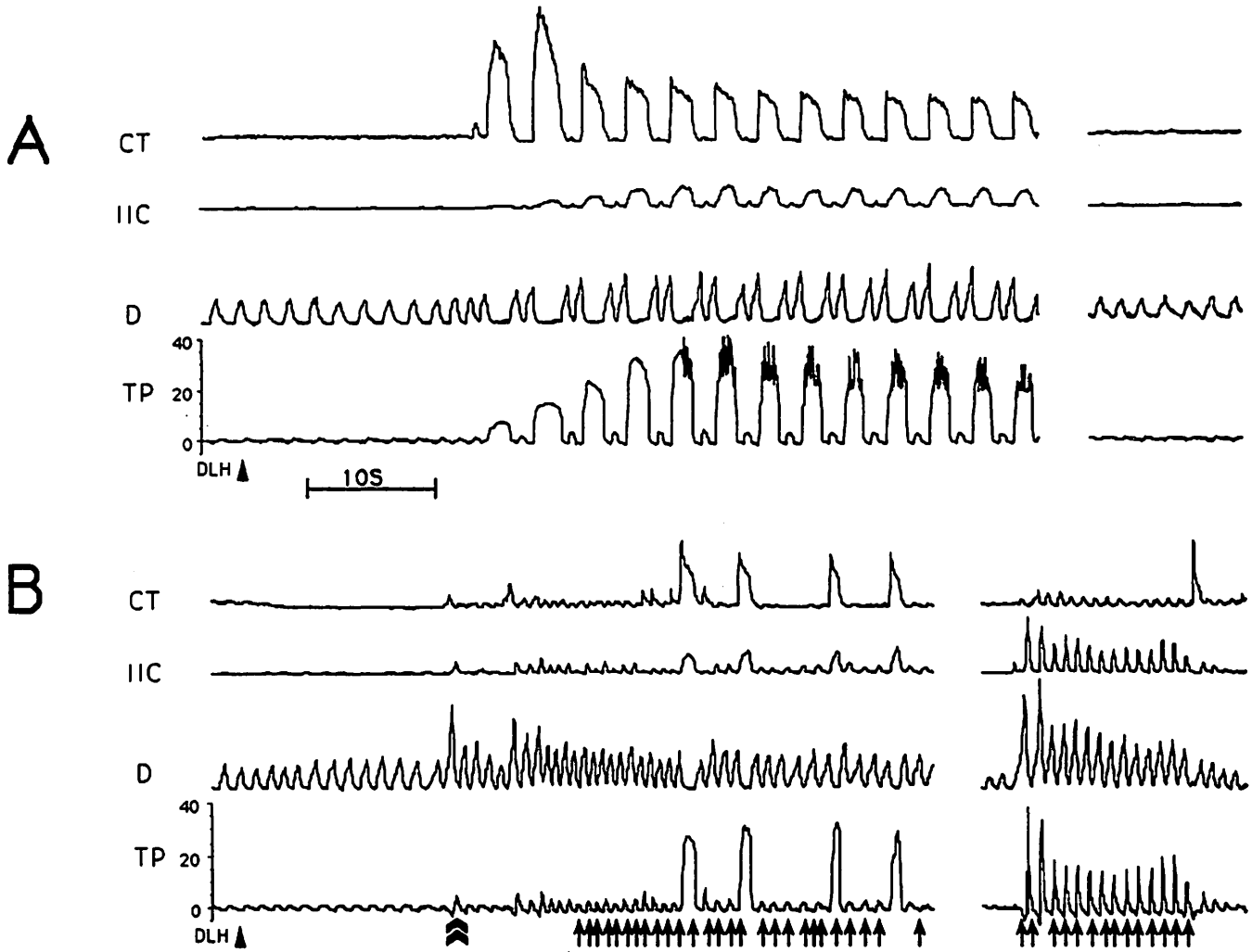

Figure 2. Tracheal pressure (TP cm H_2O) and "integrated" EMG of the cricothyroid (CT), internal intercostal (IIC) and diaphragm (D) muscles during part of a sequence of howl-like vocalizations evoked by two separate DLH microinjections (arrowheads) into the lateral PAG of a decerebrate cat. The upper panel (A) represents the muscle and pressure changes associated with the first half of a howling sequence as well as a record obtained several minutes after the DLH effect was completed. The lower panel (B) shows the result obtained 20 seconds after the DLH injection was completed when mechanical stimulation of the trachea was carried out (see text). This stimulation was repeated for most of the sequence shown in part B, although 63 seconds of the record has been omitted between the two panels. Each stimulation of the trachea was marked by a signal on the voice record (not shown) except for the initial 10 seconds. These recorded signals are represented by the sequence of single arrows and the commencement of the mechanical stimulation by a double arrow. Note that in the right panels of B, which was associated with coughing, that each mechanical stimulus is associated with prominent bursts of activity in the intercostals and diaphragm muscles as well as large increases in tracheal pressure. The time scale (10 seconds) is common to all panels.

Mechanisms of Reflex Suppression During Vocalization

The ability of SLN mechanoreceptors to signal at vocalization frequencies suggests that they may be stimulated by the rapid oscillations in local pressure gradients at the vocal fold vibration frequency. Such signals may be particularly important in the control of the vocal folds at high pitch or intensity, occasions when human vocal control is modified by topical anaesthesia of the larynx (Gould and Okamura, 1974).

Garrett (1986) recorded the discharge patterns of SLN nerve fibers during vocalization elicited by electrical stimulation of the midbrain tegmentum in the anesthetized cat. Fifteen of 35 fibers were stimulated during vocalization and 11 were inhibited. From the former group, the discharge of 4 fibers was entrained to the same frequency of the vocal fold vibration. This study provided data relating to the mechanical stimulation of the field of one fiber that was entrained by vocalization. It demonstrated slowly adapting discharges when probed, but it discharged at over 400 Hz for over one second during vocalization. The high dynamic sensitivity observed by Davis and Nail (1987) for the "tonic" fiber group as well as those "silent" fibers which displayed more slowly adapting responses to steady indentation, would predict such an unusual response. However these mechanoreceptive fibers also produce sustained discharges when exposed to chemical agents such as distilled water and irritant vapors (Davis and Nail, 1988). If these fibers are indeed from the same class as those studied during vocalization by Garrett (1986), then the potential for ambiguity exists. They would signal both mechanical events related to vocalization as well as the presence of certain chemical agents via a period of sustained discharge.

As noted in the previous sections, there are many receptors in the lungs, respiratory muscles, and vocal tract that could affect laryngeal activity associated with speech and singing. The production of a cough would clearly interfere with this process. The data illustrated in Figure 2 suggest that the cough reflex elicited by mechanical stimulation is suppressed before and during the vocalization elicited by stimulation of the PAG. The suppression occurring before the first vocalization would imply the existence of reflex "gating" within the nervous system. To understand this process completely, however, it is useful to consider additional ways in which laryngeal reflexes may be modified by vocalization.

One possibility is that punctate stimulation, such as that produced by a small inhaled particle (a cookie crumb, for example), leads to far more local tissue deformation than does general vibration of the whole laryngeal region. It is possible that these sensitive mechanoreceptors don't actually respond very well to the vibration associated with phonation. Although Garrett showed that some laryngeal mechanoreceptors were highly activated during vocalization, he also observed a sizeable group of fibers whose activity decreased during phonation. In future studies of this type, it will be very useful to characterize the receptive fields of fibers that are studied during vocalizations. Admittedly, this will be a very difficult experiment because the conditions necessary to evoke vocalization in an animal preparation (an intact larynx) make it very difficult to apply quantitative mechanical stimulation to the receptive field of a mechanoreceptor.

The technical difficulty of using single-fiber recording to evaluate the responsiveness of laryngeal receptors during vocalizations may warrant, at least for the time being, other experimental strategies. An alternative approach, which could provide information on the general responsiveness of laryngeal receptors to the mechanical stimuli associated with vocalization, would be to record from the whole SLN during evoked vocalization in an experimental animal. If vocalization stimulates many mechanoreceptors, the nerve "traffic," as reflected in the integrated signal or as potentials

phased locked to the fundamental frequency, ought to increase significantly during phonation. If such changes were observed, one would not know, by this observation alone, whether the activated receptors were primarily mucosal receptors or receptors located in deeper tissues. However, if application of a topical anesthetic to the larynx produced a large decrease in the SLN activity provoked by phonation, then this activity could be ascribed to mucosal receptors. If not, then the activity would be more reasonably ascribed to receptors in deep tissues.

If the mechanical stimuli associated with vocalization do in fact stimulate mechanoreceptors that ordinarily lead to protective reflex responses, they could fail to evoke this reflex during vocalization because these receptors are synaptically connected to produce a mutually inhibitory "center-surround" type of receptive field. In this hypothetical mechanism, protective reflexes would involve an interneuron excited by "center" stimulation, but inhibited by simultaneous stimulation of the surrounding tissue. During phonation, the widespread mechanical stimuli would excite all the receptors, but would only weakly stimulate, if at all, the interneuron. Other sensory systems, such as retinal ganglion cells, are known to be organized in this manner.

If sensitive laryngeal mechanoreceptors respond to the mechanical events associated with vocalization, and are not synaptically organized so as to make interneurons more sensitive to punctate than diffuse stimuli, then one has to conclude that defensive reflexes are suppressed by descending control signals. Such control could take place presynaptically. Sessle (1972, 1973) has demonstrated presynaptic afferent depolarization (PAD) of SLN afferent synaptic endings in nucleus solitarius. This effect was produced by mechanical stimulation of teeth, and was thus interpreted in the context of coordination between the masticatory and laryngeal systems. However, descending systems giving rise to vocalization could also use the PAD effect to reduce the sensitivity of the cough reflex.

Postsynaptic inhibition or disfacilitation of reflex interneurons is also a mechanism for suppressing a reflex. Evaluation of these possibilities is made very difficult by the lack of knowledge on the synaptology of the cough reflex, in particular the location and defining response properties of its interneurons. Another mechanism for suppressing a reflex is direct postsynaptic inhibition of the motoneurons that are involved. If it is assumed that the same laryngeal motoneurons are activated for vigorous defensive responses as are activated for vocalization, then one can dismiss this idea as the mechanism for suppression of the cough response during vocalization. Although it is as yet unproved, it seems very likely that the same motoneurons are involved in both activities. These motoneurons are very active during vocalization, so they are obviously excitable.

Although the synaptic arrangements necessary for hypothetical "gating" of reflexes have been demonstrated in many motor systems, experiments that have provided irrefutable demonstration that such "gating" actually takes place in intact conscious animals and humans are few in number. It is a difficult experiment. One has to demonstrate that the level of motoneuron pool excitability is as high or higher during the "suppressed" condition as it is in the control condition. Further, it has to be shown (or believed) that the eliciting stimulus is the same in both conditions. Lund and Rossignol (1981) have, however, provided convincing evidence that the jaw-opening reflex in rabbits is suppressed during mastication. There is thus solid evidence in intact animals that at least some reflexes involving brainstem neural systems are suppressed in circumstances where they would be inappropriate. It seems likely that further research will confirm the suggestion, illustrated in Figure 2, that the cough reflex is suppressed or at least greatly attenuated during vocalization.

Contributions of Vocal Tract Propioceptors to Vocalization

As noted in proceeding sections, the laryngeal mucosa and tissues of the lung are well-supplied with receptors that have reflex connections that play an important role in protecting the airway from foreign material and in regulating breathing. The laryngeal muscles and articulations of animals and humans also contain receptors (Baken, 1971; Kirchner and Wyke, 1965; Larson et al., 1974; Nagai, 1989; Okamura and Katto, 1988; and Rossi and Cortesina, 1965). Some receptors found in laryngeal muscles have been described as "typical" spindles, but other were spiral endings around muscle fibers. Besides this contingent of proprioceptors, the walls of the trachea contain stretch receptors whose axons exhibit graded rates of discharge to positive pressures in the range of the pressures used in speaking and singing (Bartlett et al., 1976; Mathew et al., 1984; and Ogilvie et al., 1989). Although little is actually known of the reflex effects, if any, of these receptors, they have figured prominently in the hypothesis of Wyke (1985) that the activity of laryngeal and respiratory muscles during speech and singing is reflexly regulated by inputs from these receptors. Warren (1986) has also hypothesized that the speech of cleft palate speakers is degraded by reflexes driven by pressure receptors that regulate the oral pressures during speaking to values that are too high for these individuals.

These are legitimate hypotheses, and follow naturally from the ideas of noted neurophysiologists such as Sherrington and Granit, who long ago canonized the principle that proprioceptors, particularly those from muscles, are very important for regulation of movements. If one believes in the basic correctness of this principle, and additionally believes that all motor systems in the body are controlled using the same principles, then the proof of the existence of proprioceptors in the vocal tract might be sufficient to accept the views of Wyke and Warren. It is nevertheless reasonable to ask what empirical evidence there is to support these ideas.

One major problem with acceptance of Wyke and Warren's views is that the "servo hypothesis," (Merton, 1953), presently seems incorrect. Hulliger, in a comprehensive review of muscle spindles and their connections (1984) has stated, with reference to the Merton hypothesis, "This possibility can now safely be dismissed..." One line of investigation of the Merton servo hypothesis has been to assess the effects of interrupting sensory feedback upon motor control. Such an experiment, which relates directly to the question of voluntary laryngeal control, was performed by Tanabe, Kitajima, and Gould (1975). These scientists percutaneously bilaterally anesthetized the internal branch of the SLN in human subjects, and assessed the effect upon the ability of the subjects to control their voice before and after the anesthesia. There were no statistically reliable quantitative changes in the ability of these three subjects to produce the initial part of a phonation "on pitch" and at the expected intensity. Following the anesthesia, however, there was an increase in the amplitude of small slow fluctuations of the fundamental frequency during the phonation following the anesthesia. The authors of the study concluded that "it is reasonable to assume that the reflex through the internal branch of the SLN plays the important role for the precise control of the fundamental frequencies without, however, affecting the overall or gross motor control of phonation."

Deafferentation experiments with animals (Finocchio and Luschei, 1988; Goodwin and Luschei, 1974; Polit and Bizzi, 1979; Wetzel et al., 1976) have reached conclusions which are similar to the one noted above. Very different motor systems have been involved in these studies: arm control in monkeys, hindlimb movements during walking in cats, and mandibular control in monkeys. In particular, studies of mandibular movement in monkeys after destroying afferents in the tract of the mesencephalic nucleus of the fifth nerve, which selectively removed jaw muscle

spindle afferents, revealed only two obvious detrimental effects. These are an increase in the amplitude of low frequency "tremor" of either mandibular biting force or mandibular position, and a small but statistically reliable increase of the mean mandibular tracking error. On the other hand, these same animals chewed quite normally, performed complex mandibular tracking tasks with great skill right from the beginning of post-operative testing, and appeared to be quite normal in all respects other than the subtle affects noted above.

These results are mentioned here because they call into question the basic correctness of the idea that muscle spindle afferents in general play a major role in controlling the activity of motoneurons. If it cannot reasonably be assumed that laryngeal muscle spindles have strong reflex connections to laryngeal motoneurons, then such connections would have to demonstrated for Wyke's proposition to be tenable. There is no positive evidence for a strong stretch reflex of laryngeal muscles. There is, however, negative evidence (Testerman, 1970). The laryngeal deafferentation experiment of Tanabe, Kitajima and Gould is also very significant in this regard because the conclusions were based upon the study of human subjects, and the examination of a skilled vocal response.

As noted earlier, there are pressure receptors in the trachea. Could they be the source of reflex control of the subglottic pressure for speech and singing? Most of these receptors have exhibited dynamic sensitivity. During phonation, such receptors are exposed to large pressure fluctuations at the fundamental frequency, so their ability to signal mean pressures in the vocal tract could be compromised. Nevertheless, it is reasonable to suppose that a composite pressure-related afferent signal could be extracted by the nervous system. For Warren's hypothesis to be tenable, however, it needs to be demonstrated that pressure fluctuations lead to reflex changes in respiratory and laryngeal muscles that lead to stability of the pressure in the vocal tract. Several studies of vocal tract pressure perturbations during phonation (Baken and Orlikoff, 1987; Mead and Reid, 1988; Perkell, 1976; and Sears and Newson Davis, 1968) have noted a variety of short latency changes in muscle activity and/or phonation frequency. None of the responses reported, however, was of a magnitude, duration, or of the correct "sign," to have been effective in restabilizing the vocal tract pressure. Additionally, Moon and Luschei (unpublished observations) have been unable to detect any EMG response in expiratory muscles of human subjects during voluntary control of oral pressure or oral airflow in response to very large sudden oral pressure perturbations.

Generally speaking, the circumstantial evidence for reflex control of the vocal tract during phonation and speech is quite meager. Direct experimental tests of such reflex regulation that could have revealed strong reflex control of laryngeal behavior, had it been present, have failed to do so. Such studies do not rule out, however, more subtle reflex effects that might be involved in fine control.

Some General Thoughts about Receptors and Reflexes

Although powerful reflexes have historically been regarded as "good," for motor control, the fact is that powerful reflexes in combination with fast muscles and significant conduction delays would tend to promote oscillation and general instability rather than stability. Any time a receptor senses an event which it must control, and there is a significant time which elapses before it can produce a restoring effect, there exists a frequency of perturbation at which the response arrives at just the time necessary to reinforce rather than oppose the perturbation. If a perturbation occurs at this frequency (and where there is "noise" in the system this "bad" frequency will exist to some

degree) the whole system will go into high amplitude oscillation at this frequency if the reflex gain is high. One can argue this point from a theoretical, quantitative, approach. A much more "picturesque" confirmation of this fact comes from attempts to build a mechanical servo system intended to produce high frequency force or positional displacements when a significant delay in the feedback path is present: the result is a very expensive vibrator. It is likely that, if our voluntary motor behavior depended upon continuously active high gain "servo-like" reflexes, we would be a quivering and shaking pile of bones and muscle most of the time. Any attempt to speak or phonate would break into uncontrollable oscillation of intensity and pitch at a frequency primarily determined by the "lag term" introduced by nerve conduction delays and the period of time before an increase in muscle activity produces its maximum mechanical response. However, a slowly responding effector system may use high gain reflexes and still retain stability because the effector cannot vary its output significantly at the "bad" frequency. That is, the response of the entire system is "rolled off" at a frequency well below the resonant frequency.

It seems possible, and perhaps likely, that we execute practiced movements, particularly those that require complex, rapid, sequences of muscle activity in an "open loop" manner. This is not a new view (Lashley, 1917). It is a disturbing view to some scientists, however, because it seems to leave as "worthless" the rich sensory innervation of muscles, joints and tendons. This view does not, however, force one to suppose that receptors are unimportant in motor control. It is likely that they respond during these movements, but it is also likely that their discharge rate has little to do with movement "parameters" as defined by our engineering-oriented way of looking at things. It would be extremely difficult for the nervous system to make sense of this information by attention to one receptor, or even one class of receptor. The analogous situation would be trying to understand speech by studying the output of a single narrow-band filter applied to the speech signal. Rather, the total sensory consequence of movements may be regarded as a "collage," or, for the musically inclined, a "symphony".

The synaptic consequences of many receptors having weak reflex connections may be viewed as creating a complex "background" of synaptic excitation and inhibition of motoneuron pools. In the period of motor learning, as neonates, this composite effect must be as bewildering and meaningless as the pattern of speech sounds that falls on the ear of a baby. Yet the nervous system can learn to "understand" this complex composite neural signal just as it learns to understand speech. It is with respect to this signal, and the composite synaptic background to motoneuron pools produced by reflexes, that the nervous system evolves the open-loop motor "programs" responsible for skilled movements. Once learned, however, these "programs" may be executed with reasonable accuracy without sensory feeback. A good example of this phenomenon is the retention of intelligible speech by individuals who become completely deaf after normal speech has been acquired. This view does not "throw away" receptors and reflexes. Rather, it takes them out of the extremely narrow context in which they have been historically regarded.

Not all receptors and their reflex connections need operate in a mere "background" role, however. In some systems, in some situations, reflexes may assume the dominant mode of control of motoneuron pools. The key word is "some." The laryngeal protective reflexes, respiratory regulation reflexes, the eyeblink reflex, vestibular reflexes, and muscle spindle mediated reflexes in some muscle-limb systems, are examples. Holding this view makes it quite possible to suppose, for example, that the jaw stretch reflex has a very important functional role in keeping the mandible from "bouncing" due to inertial loading during locomotion (Goodwin and Luschei, 1978) and, at the same time, to suppose that the few muscle spindles found in some muscles have little special impor-

tance for motor control. As noted previously, however, they may have significance in terms of subtle background excitation of motoneuron pools, and perhaps as part of the sensory "collage" that is produced when using the vocal tract.

A corollary to the view expressed here is that there is not "a function" to be ascribed to a receptor like a muscle spindle or a mechanoreceptor. A receptor might, in engineering terms, be like a transducer that can be used for many different tasks. In an automobile, a pressure transducer can be used as part of a feedback system to control the engine speed. It might also be used as a monitoring device to warn the driver if the oil pressure has dropped to a low level. Or it might sit unused most of the time in the glove compartment of the car, waiting to be used to check the tire pressure. In the last case, it cannot be cannot concluded that the transducer is useless even though it might be removed without any serious consequences for the operation of the car. For most people, its use would be a matter of convenience. Some drivers might choose not to use it at all, merely waiting for a friendly passer-by to tell them their tire looks flat. It is possible that quite a few receptors in our body are of this last "rarely used" category. Note, however, that if the car is a racing automobile, and the income of the driver is dependent upon winning races, then frequent precise checking of the tire pressure might be the difference between surviving or becoming "extinct" in this profession.

References

Agostoni, E., Chinnock, J.E., De Burgh Daly, M., and Murray, J.G. (1957). Functional and histological studies of the vagus nerve and its branches to the heart, lungs and abdominal viscera in the cat. *Journal of Physiology (London), 135,* 182-205.

Andrew, B.L. (1956). A functional analysis of the myelinated fibers of the superior laryngeal nerve of the rat. *Journal of Physiology (London), 133,* 420-432.

Arnold, G.E. (1961). Physiology and pathology of the cricothyroid muscle. *Laryngoscope, 71,* 687-752.

Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. *Journal of the Acoustic Society of America, 65,* 1271-1275.

Baken, R.J. (1971). Neuromuscular spindles in the intrinsic muscles of a human larynx. *Folia Phoniatrica, 23,* 204-210.

Baken, R.J., and Orlikoff, R.F. (1987). Phonatory response to step-function changes in supraglottal pressure. In T. Baer, C. Sasaki and K. Harris (eds.), *Laryngeal Function in Phonation and Respiration* (pp 273-290). Boston: College-Hill Press.

Bandler, R. (1988). Brain mechanisms of aggression as revealed by electrical and chemical stimulation: Suggestion of a central role for the midbrain periaqueductal grey region. In A. Epstein and A. Morrison (eds.), *Progress in Psychobiology and Physiological Psychology: Vol. 13* (pp 67-153). New York: Acad Press.

Bandler, R., Carrive, P., and Zhang, S.P. (1991). Integration of somatic and autonomic reactions within the midbrain periaqueductal grey: Viscerotopic, somatotopic and functional organization. In G. Holstege (ed.), *Role of the forebrain in sensation and behavior. Progress in Brain Research, Vol. 87* (pp 269-305). Amsterdam: Elsevier Science Publishers.

Bartlett, D., Jr. (1979). Effects of hypercapnia and hypoxia on laryngeal resistance to airflow. *Respiratory Physiology, 37,* 293-302.

Bartlett, D., Jr. (1980). Effects of vagal afferents on laryngeal responses to hypercapnia and hypoxia. *Respiratory Physiology, 42,* 189-198.

Bartlett, D., Jr. (1989). Respiratory functions of the larynx. *Physiological Reviews, 69,* 33-57.

Bartlett, D., Jr., and Birchard, G.F. (1983). Effects of hypoxia on lung volume in the garter snake. *Respiratory Physiology, 53,* 63-70.

Bartlett, D., Jr., and Knuth, S.L. (1984). Human vocal cord movements during voluntary hyperventilation. *Respiratory Physiology, 58,* 289-294.

Bartlett, D., Jr., Knuth, S.L., and Knuth, K.V. (1981). Effects of pulmonary stretch receptor blockade on laryngeal responses to hypercapnia and hypoxia. *Respiratory Physiology, 45,* 67-77.

Bartlett, D., Jr., Leiter, J.C., and Knuth, S.L. (1990). Control and actions of the genioglossus muscle. In F.G. Issa, P.M. Suratt and J.E. Remmers (eds.), *Sleep and Respiration, Section III: Upper Airway Muscles: Vol. 345, Progress in Clinical and Biological Research* (pp 99-108). New York: Wiley-Liss.

Bartlett, D., Jr., Remmers, J.E., and Gautier, H. (1973). Laryngeal regulation of respiratory airflow. *Respiratory Physiology, 18,* 194-204.

Bartlett, D., Jr., Sant'Ambrogio, G., and Wise, J.C.M. (1976). Transduction properties of tracheal stretch receptors. *Journal of Physiology (London), 258,* 421-432.

Bartlett, D., Jr., and St. John, W.M. (1988). Influence of lung volume on phrenic hypoglossal and mylohyoid nerve activities. *Respiratory Physiology, 73,* 97-110.

Boelaert, R. (1941). Sur la physiologie de la respiration des lacertiens. *Archives Internationales de Physiologie et de Biochimie, 51,* 379-437.

Boggs, D.F., and Bartlett, D., Jr. (1982). Chemical specificity of a laryngeal apneic reflex in puppies. *Journal of Applied Physiology, 53,* 455-462.

Boushey, H.A., and Richardson, P.S. (1973). The reflex effects of intralaryngeal carbon dioxide on the pattern of breathing. *Journal of Physiology (London), 228,* 181-191.

Boushey, H.A., Richardson, P.S., and Widdicombe, J.G. (1972). Reflex effects of laryngeal irritation on the pattern of breathing and total lung resistance. *Journal of Physiology (London), 224,* 501-513.

Boushey, H.A., Richardson, P.S., Widdicombe, J.G., and Wise, J.C.M. (1974). The response of laryngeal afferent fibers to mechanical and chemical stimuli. *Journal of Physiology (London), 240,* 153-175.

Bouverot, P., and Fitzgerald, R.S. (1969). Role of the arterial chemoreceptors in controlling lung volume in the dog. *Respiratory Physiology, 7,* 203-215.

Brancatisano, T.P., Collett, P.W., and Engel, L.A. (1983). Respiratory movements of the vocal cords. *Journal of Applied Physiology, 54,* 1269-1276.

Brancatisano, T.P., Dodd, D.S., and Engel, L.A. (1984). Respiratory activity of posterior cricoarytenoid muscle and vocal cords in humans. *Journal of Applied Physiology, 57,* 1143-1149.

Breuer, J. (1970). Self-steering of respiration through the nervus vagus (English translation by Elisabeth Ullmann). In R. Porter (ed.), *Breathing: Hering-Breuer Centenary Symposium* (pp 365-394). London: Churchill.

Brody, A.W. (1954). Mechanical compliance and resistance of the lung-thorax calculated from the flow recorded during passive expiration. *American Journal of Physiology, 178,* 189-196.

Bunn, J.C., and Mead, J. (1971). Control of ventilation during speech. *Journal of Applied Physiology, 31,* 870-872.

Campbell, C.J., Murtagh, J.A., and Raber, C.F. (1963). Chemical agents and reflex control of laryngeal glottis. *Annals of Otology, Rhinology and Laryngology, 72,* 589-604.

Carrive, P., Dampney, R.A.L., and Bandler, R. (1987). Excitation of neurones in a restricted portion of the midbrain periaqueductal grey elicits both behavioral and cardiovascular components of the defence reaction in the unanesthetized decerebrate cat. *Neuroscience Letters, 81,* 273-278.

Clark, F.J., and von Euler, C. (1972). On the regulation of depth and rate of breathing. *Journal of Physiology (London), 222,* 267-295.

Cohen, M. I. (1975). Phrenic and recurrent laryngeal discharge patterns and the Hering-Breuer reflex. *American Journal of Physiology, 228,* 1489-1496.

Compton, D., Hill, P. M., and Sinclair, J. D. (1973). Weight-lifters' blackout. *Lancet, 2,* 1234-1237.

Davis, P. J. (1986). *Observations on the innervation and control of the larynx.* Unpublished doctoral dissertation, University of New South Wales, Kensington, Australia.

Davis, P. J. and Nail, B. S. (1987). Quantitative analysis of laryngeal mechanosensitivity in the cat and rabbit. *Journal of Physiology (London), 388,* 467-485.

Davis, P. J. and Nail, B. S. (1988). The sensitivity of laryngeal epithelial receptors to static and dynamic forms of mechanical stimulation. In O. Fujimura (ed.), *Vocal Physiology: Voice Production, Mechanisms and Functions* (pp 1-18). New York: Raven Press.

Davis, P. J., and Zhang, S. P. (1991). What is the role of the midbrain periaqueductal gray in respiration and vocalization? In A. Depaulis and R. Bandler (eds.), *The Midbrain Periaqueductal Gray Matter: Functional, Anatomical and Immunohistochemical Organization.* (pp 57-66). New York: Plenum Publishing Corp.

Dixon, M., Szereda-Przestaszewska, M., Widdicombe, J. G., and Wise, J. C M. (1974). Studies on laryngeal calibre during stimulation of peripheral and central chemoreceptors, pneumothorax and increased respiratory loads. *Journal of Physiology (London), 239,* 347-363.

Dubois, F. S., and Foley, J. O. (1934). Experimental studies on the vagus and spinal accessory nerves in the cat. *Anatomic Record, 64,* 285- 307.

England, S. J., and Bartlett, D., Jr. (1982). Changes in respiratory movements of the human vocal cords during hyperpnea. *Journal of Applied Physiology, 52,* 780-785.

England, S. J., Bartlett, D., Jr., and Daubenspeck, J. A. (1982). Influence of human vocal cord movements on airflow and resistance in eupnea. *Journal of Applied Physiology, 52,* 773-779.

England, S. J., Bartlett, D., Jr., and Knuth, S. L. (1982). Comparison of human vocal cord movements during isocapnic hypoxia and hypercapnia. *Journal of Applied Physiology, 53,* 81-86.

Fink, B. R. (1974). Spring mechanisms in the human larynx. *Acta Oto-Laryngologica (Stockholm), 77,* 295-304.

Fink, B. R. (1975). *The Human Larynx: A Functional Study.* New York: Raven.

Fink, B. R., Basek, M., and Epanchin, V. (1956). The mechanism of opening of the human larynx. *Laryngoscope, 66,* 410-425.

Finocchio, D.V. and Luschei, E.S. (1988). Characteristics of complex voluntary mandibular movements in the monkey before and after destruction of most jaw muscle spindle afferents. *Journal of Voice, 2,* 279-290.

Fisher, J. T., Mathew, O. P., Sant'Ambrogio, F. B., and Sant'Ambrogio, G. (1985). Reflex effects and receptor responses to upper airway pressure and flow stimuli in developing puppies. *Journal of Applied Physiology, 58,* 258-264.

Fisher, J. T., Mortola, J. P., Smith, J. B., Fox, G. S., and Weeks, S. (1982). Respiration in newborns. Development of the control of breathing. *American Review of Respiratory Diseases, 125,* 650-657.

Frost, R. (1963). Letter to Louis Untermeyer dated January 1, 1916. In, *The Letters of Robert Frost to Louis Untermeyer,* (p 22). New York: Holt, Rinehart and Winston.

Galen. (1968). *On the Usefulness of the Parts of the Body* (M. T. May, Trans) (p 354). Ithica, NY: Cornell University Press.

Garrett, J. D. and Luschei, E. S. (1987). Subglottic pressure modulation during evoked phonation in the anesthetized cat. In T. Baer, C. Sasaki and K. Harris (eds.), *Laryngeal Function in Phonation and Respiration* (pp 139-153). Boston: College-Hill Press.

Garrett, J.D. (1986). *The response of laryngeal mechanoreceptors innervated by the superior laryngeal nerve during evoked phonation in the anesthetized cat.* Unpublished doctoral dissertation, The University of Iowa, Iowa City, IA.

Gautier, H., Remmers, J. E., and Bartlett, D., Jr. (1973). Control of the duration of expiration. *Respiratory Physiology,* 18, 205-221.

Gelfer, C., Harris, K., Collier, R., and Baer, T. (1983). Is declination actively controlled? In I.R. Titze and R.C. Scherer, (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control* (pp 113-126). Denevr: The Denver Center for the Performing Arts.

Gesell, R., and White, F. (1938). Recruitment of muscular activity and the central neurone afterdischarge of hyperpnea. *American Journal of Physiology, 122,* 48-56.

Goodchild, A.K., Dampney, R.A.L., and Bandler, R. (1982). A method of evoking physiological responses by stimulation of cell bodies, but not axons of passage, within localized regions of the nervous system. *Journal of Neuroscience Methods, 6,* 351-363.

Goodwin, G.M. and Luschei, E.S. (1974). Effects of destroying spindle afferents from the jaw muscles on mastication in monkeys. *Journal of Neurophysiology, 37,* 967-81.

Goodwin, G.M., Hoffman, D.S., and Luschei, E.S. (1978). The strength of the reflex response to sinusoidal stretch of monkey jaw closing muscles during voluntary contraction. *Journal of Physiology (London), 279,* 81-111.

Gould, W.J., and Okamura, H. (1974). Interrelationships between voice and mucosal reflexes. In B. Wyke (ed.), *Ventilatory and Phonatory Control Systems* (pp 347-369). Oxford: Oxford University Press.

Guz, A., and Trenchard, D.W. (1971). Pulmonary stretch receptor activity in man: A comparison with dog and cat. *Journal of Physiology (London), 213,* 329-343.

Harding, R. (1980). State-related and developmental changes in laryngeal function. *Sleep, 3,* 307-322.

Horiuchi, M., and Sasaki, C. T. (1978). Cricothyroid muscle in respiration. *Annals of Otology, Rhinology and Laryngology, 87,* 386-391.

Hulliger, M. (1984). The mammalian muscle spindle and its central control. *Reviews of Physiology Biochemistry and Pharmacology (Berlin), 101,* 1-110.

Jiménez-Vargas, J., Miranda, J., and Mouriz, A. (1962). Physiology of the cough. Differentiation of constrictor and dilatory reflexes of the laryngeal sphincter. *Revista espanola de Fisiologia, 18,* 7-21.

Jürgens, U., and Richter, K. (1986). Glutamate-induced vocalization in the squirrel monkey. *Brain Research, 373,* 349-358.

Kirchner, J. A., and Wyke, B. (1965). Afferent discharges from laryngeal articular mechanoreceptor. *Nature (London), 205,* 86-87.

Konrad, H. R., and Rattenborg, C. C. (1969). Combined action of laryngeal muscles. *Acta Oto-Laryngologica (Stockholm), 67,* 646-649.

Kosch, P. C., and Stark, A. R. (1984). Dynamic maintenance of end-expiratory lung volume in full-term infants. *Journal of Applied Physiology, 57,* 1126-1133.

Kuna, S. T. (1986). Inhibition of inspiratory upper airway motoneuron activity by phasic volume feedback. *Journal of Applied Physiology, 60,* 1373-1379.

Larson, C.R., Sutton, D., and Lindeman, R.C. (1974). Muscle spindles in nonhuman primate laryngeal muscles. *Folia Primatologica, 22,* 315-323.

Lashley, K.S. (1917). The accuracy of movement in the absence of excitation from the moving organ. *American Journal of Physiology, 43,* 169-194.

Lipski, J., Bellingham, M.C., West, M.J., and Pilowski, P. (1988). Limitations of the technique of pressure microinjection of excitatory amino acids for evoking responses from localized regions of the CNS. *Journal of Neuroscience Methods, 26,* 169-179.

Lund, J.P., and Rossignol, S. (1981). Modulation of the amplitude of the digastric jaw opening reflex during the masticatory cycle. *Neuroscience, 6,* 95-98.

Mårtensson, A. (1964). Proprioceptive impulse patterns during contraction of intrinsic laryngeal muscle. *Acta Physiologica Scandinavica, 62,* 176-194.

Mathew, P. (1984). Upper airway negative-pressure effects on respiratory activity of upper airway muscles. *Journal of Applied Physiology, 56,* 500-505.

Mathew, O. P., Abu-Osba, Y. K., and Thach, B. T. (1982). Genioglossus muscle responses to upper airway pressure changes: afferent pathways. *Journal of Applied Physiology, 52,* 445-450.

Mathew, O. P., and Farber, J. P. (1983). Effect of upper airway negative pressure on respiratory timing. *Respiratory. Physiology, 54,* 259-268.

Mathew, O. P., Sant'Ambrogio, G., Fisher, J. T., and Sant'Ambrogio, F. B. (1984). Laryngeal pressure receptors. *Respiratory Physiology, 57,* 113-122.

Mathew, O. P., Sant'Ambrogio, G., Fisher, J. T., and Sant'Ambrogio, F. B. (1984). Respiratory afferent activity in the superior laryngeal nerves. *Respiratory Physiology, 58,* 41-50.

McCaffrey, T. V., and Kern, E. B. (1980). Laryngeal regulation of airway resistance. I. Chemoreceptor reflexes. *Annals of Otology, Rhynology and Laryngology, 89,* 209-214.

Mead, J., and Reid, M.B. (1988). Respiratory muscle activity during repeated airflow interruption. *Journal of Applied Physiology, 64,* 2314-2317.

Merton, P.A. (1953). Speculations on the servo-control of movement. In J.L. Malcolm, J.A.B. Gray, and G.E.W. Wolstenholme (eds.), *The Spinal Cord* (pp 247-255). Boston: Little, Brown, and Co.

Mortola, J. P. (1987). Dynamics of breathing in newborn mammals. *Physiological Review, 67,* 187-243.

Mortola, J. P., Fisher, J. T., Smith, B., Fox, G., and Weeks, S. (1982). Dynamics of breathing in infants. *Journal of Applied Physiology, 52,* 1209-1215.

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

Murphy, A. J., Koepke, G. H., Smith, E. M., and Dickinson, D. G. (1959). Sequence of action of the diaphragm and intercostal muscles during respiration. II. Expiration. *Archives of Physi cal Medicine and Rehabilitation, 40,* 337-342.

Nadel, J. A., and Widdicombe, J. G. (1962). Reflex effects of upper airway irritation on total lung resistance and blood pressure. *Journal of Applied Physiology, 17,* 861-865.

Nagai, T. (1989). Confirmation of encapsulated nerve structures in the human vocal cord. *Acta Oto-Laryngologica (Stockholm), 107,* 278-282.

Ogilvie, M.D., Bogen, D.K., Galante, R.J., and Pack, A.I. (1989). Response of stretch receptors to static inflations and deflations in an isolated tracheal segment. *Respiratory Physiology, 75,* 289-308.

Okamura, H. and Katto, Y. (1988). Fine structure of muscle spindle in interarytenoid muscle of human larynx. In O. Fujimura (ed.), *Vocal Physiology: Voice Production, Mechanisms and Functions.* New York: Raven Press.

Olsen, C. R., Hale, F. C., and Elsner, R. (1969). Mechanics of ventilation in the pilot whale. *Respiratory Physiology, 7,* 137-149.

Perkell, J.S. (1976). Responses to an unexpected suddenly induced change in the state of the vocal tract. *Research Laboratory of Electronics, MIT Report, 117,* 273-281.

Petit, J. M., Milic-Emili, G., and Delhez, L. (1960). Role of the diaphragm in breathing in conscious normal man: an electromyographic study. *Journal of Applied Physiology, 15,* 1101-1106.

Phipps, R. J., and Richardson, P. S. (1976). The effects of irritation at various levels of the airway upon tracheal mucus secretion in the cat. *Journal of Physiology (London), 261,* 563-581.

Phillipson, E.A., McClean, P.A., Sullivan, C.E., and Zamel, N. (1978). Interaction of metabolic and behavioral respiratory control during hypercapnia and speech. *American Review of Respiratory Diseases, 117,* 903-909.

Polit, A., and Bizzi, E. (1979). Characteristics of motor programs underlying arm movements in monkeys. *Journal of Neurophysiology, 42,* 183-194.

Proctor, D.F. (1964). Physiology of the upper airway. In, *Handbook of Physiology. Respiration. Section 3. Volume 1. Chapter 8* (pp 309-345). Washington, D.C.: American Physiological Society.

Proctor, D.F. (1980). *Breathing. Speech and Song.* Vienna: Springer-Verlag.

Remmers, J.E. (1973). Extra-segmental reflexes derived from intercostal afferents: phrenic and laryngeal responses. *Journal of Physiology (London), 233,* 45-62.

Rossi, G. and Cortesina, G. (1965). Morphological study of the laryngeal muscles of man. *Acta Oto-Laryngologica (Stockholm), 59,* 575-592.

Sampson, S., and Eyzaguirre, C. (1964). Some functional characteristics of mechano-receptors in the larynx of the cat. *Journal of Neurophysiology, 27,* 464-480.

Sant'Ambrogio, G., Brambilla-Sant'Ambrogio, F., and Mathew, O.P. (1986). Effect of cold air on laryngeal mechanoreceptors in the dog. *Respiratory Physiology, 64,* 45-56.

Sant'Ambrogio, G., Mathew, O.P., Fisher, J.T., and Sant'Ambrogio, F.B. (1983). Laryngeal receptors responding to transmural pressure, airflow and local muscle activity. *Respiratory Physiology, 54,* 317-330.

Sant'Ambrogio, G., Mathew, O.P., and Sant'Ambrogio, F.B. (1985a). Role of intrinsic muscles and tracheal motion in modulating laryngeal receptors. *Respiratory Physiology, 61,* 289-300.

Sant'Ambrogio, G., Mathew, O.P., Sant'Ambrogio, F.B., and Fisher, J. T. (1985b). Laryngeal cold receptors. *Respiratory Physiology, 59,* 35-44.

Sasaki, C., and English T. (1984). Physiology of the larynx. In G.M. English (ed.),*Otolaryngology. Volume 3* (pp 1-26). Philadelphia: Harper and Row.

Scholander, P.F. (1940). Experimental investigations on the respiratory function in diving mammals and birds. *Hvalrådets Skrifter det Norske Videnskaps- Akademi Oslo, 22,* 1-131.

Sears, T. and Newsom Davis, J. (1968). The control of respiratory muscles during voluntary breath ing. *Annals of the New York Academy of Science, 155,* 183-190.

Semon, F. (1891). On the position of the vocal cords in quiet respiration of man, and on the reflex-tonus of their abductor muscles. *Procedings of the Royal Society of London Series B-Biological Sciences, 48,* 156-159.

Sessle, B.J., and Storey, A.T. (1972). Periodontal and facial influences on the laryngeal input to the brain stem of the cat. *Archives of Oral Biology, 17,* 1583-1595.

Sessle, B.J. (1973). Presynaptic excitability changes in single laryngeal primary afferent fibers. *Brain Research, 53,* 333-342.

Sharpey-Schafer, E.P. (1953). The mechanism of syncope after coughing. *British Medical Journal, 2,* 860-863.

Sherrey, J.H., and Megirian, D. (1974). Spontaneous and reflexly evoked laryngeal abductor and adductor muscle activity of cat. *Experimental Neurology, 43,* 487-498.

Shingai, T. (1977). Ionic mechanism of water receptors in the laryngeal mucosa of the rabbit. *Japanese Journal of Physiology, 27,* 27-42.

Shingai, T. (1979). Physiochemical study of receptive mechanism of laryngeal water fibers in the rabbit. *Japanese Journal of Physiology, 29,* 459 470.Shingai, T. (1980). Water fibers in the superior laryngeal nerve of the rat. *Japanese Journal of Physiology, 30,* 305-307.

Sica, A.L., Cohen, M.I., Donnelly, D.F., and Zhang, H. (1984). Hypoglossal motoneuron responses to pulmonary and superior laryngeal afferent inputs. *Respiratory Physiology, 56,* 339-357.

Sullivan, C.E., Murphy, E., Kozar, L.F., and Phillipson, E.A. (1978). Waking and ventilatory responses to laryngeal stimulation in sleeping dogs. *Journal of Applied Physiology, 45,* 681-689.

Suzuki, M., and Kirchner, J.A. (1968). Afferent nerve fibers in the external branch of the superior laryngeal nerve in the cat. *Annals of Otology, Rhinology and Laryngology, 77,* 1059-1070.

Suzuki, M., and Kirchner, J.A. (1969). Sensory fibers in the recurrent laryngeal nerve. *Annals of Otology, Rhinology and Laryngology, 78,* 1-30.

Suzuki, M., and Sasaki, C.T. (1976). Initiation of reflex glottic closure. *Annals of Otology, Rhinology and Laryngology, 85,* 382-386.

Suzuki, M., and Sasaki, C.T. (1977a). Effect of various sensory stimuli on reflex laryngeal adduction. *Annals of Otology, Rhinology and Laryngology, 86,* 30-36.

Suzuki, M., and Sasaki, C.T. (1977b). Laryngeal spasm: a neuropysiologic redefinition. *Annals of Otology, Rhinology and Laryngology, 86,* 150-158.

Szereda-Przestaszewska, M., and Widdicombe, J.G. (1973). Reflex effects of chemical irritation of the upper airways on the laryngeal lumen in cats. *Respiratory Physiology, 18,* 107-115.

Tanabe, M., Kitajima, K., and Gould, W.J. (1975). Laryngeal phonatory reflex: The effect of anesthetization of the internal branch of the superior laryngeal nerve: Acoustic aspects. *Annals of Otology, Rhinology and Laryngology, 84,* 206-213.

Testerman, R.L. (1970). Modulation of laryngeal activity by pulmonary changes during vocalization in cats. *Experimental Neurology, 29,* 281-297.

Tomori, Z., and Widdicombe, J.G. (1969). Muscular, bronchomotor and cardiovascular reflexes elicited by mechanical stimulation of the respiratory tract. *Journal of Physiology (London), 200,* 25-50.

Van Lunteren, E., Strohl, K.P., Parker, D.M., Bruce, E.N., Van de Graaff, W.B., and Cherniak, N.S. (1984). Phasic volume-related feedback on upper airway muscle activity. *Journal of Applied Physiology, 56,* 730-736.

Warren, D.W. (1986). Compensatory speech behaviors in individuals with cleft palate: A regulation/control phenomenon? *Cleft Palate Journal, 23,* 251-260.

Widdicombe, J.G. (1954). Respiratory reflexes elicited by inflation of the lungs. *Journal of Physiology (London), 123,* 105-115.

Widdicombe, J.G. (1986). Reflexes from the upper respiratory tract. In N.S. Cherniack and J.G. Widdicombe (eds.) *Handbook of Physiology. The Respiratory System. Control of Breathing. Volume II* (pp 363-394). Bethesda,MD: American Physiological Society.

Wetzel, M.C., Atwater, A.E., Wait, J.V., and Stuart, D.G. (1976). Kinematics of locomotion by cats with a single limb deafferented. *Journal of Neurophysiology, 39,* 667-678.

Wyke, B.D. (1983). Reflexogenic contributions to vocal fold control systems. In I.R. Titze and R.C. Scherer (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control* (pp 138-141). Denver: The Denver Center for the Performing Arts.

Wyke, B.D., and Kirchner, J.A. (1976). Neurology of the larynx. In R. Hinchcliffe and D. Harrison (eds.), *Scientific Foundations of Otolaryngology* (pp 546-574). Chicago: Heinemann.

Zhang, S.P., Bandler, R., and Carrive, P. (1990). Flight and immobility evoked by excitatory amino acid microinjection within distinct parts of the subtentorial midbrain periaqueductal grey in the cat. *Brain Research, 520,* 73-82.

NCVS Status and Progress Report - 2
April 1992, 83-91

Measurements of Mucosal Wave Propagation and Vertical Phase Difference in Vocal Fold Vibration

Ingo R. Titze, Ph.D.
Jack J. Jiang, M.D., Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Tzu-Yu Hsiao, M.D.
Department of Otolaryngology, National Taiwan University Hospital

Abstract

Examination of surface wave properties in the vocal fold mucosa is becoming to be an important part of assessment of vocal function. A key wave property is propagation velocity, which determines the phase delay between upper and lower margins of the vocal folds. Excised canine larynges were used in this study to measure this phase delay, and therewith propagation velocity. The motion of two fleshpoints was tracked stroboscopically. Differential displacements between the fleshpoints were matched to displacements of a model. A least-squared fit of the data to the model provided the numerical values of propagation velocity, which varied from 0.5 m/s to about 2.0 m/s, depending on fundamental frequency. The corresponding phase delay along the medial surface of the vocal folds varied from about 60°/mm to 30°/mm.

Introduction

It is now well established that self-oscillation of the vocal folds is facilitated by a phase delay in movement of the upper margin of folds with respect to the lower margin. Ishizaka & Matsudaira (1972) demonstrated this by deriving analytical expressions for conditions of oscillation with a two-mass model of the vocal folds. Subsequently, Titze (1988) confirmed the derivation by modeling a surface wave that propagates vertically along the medial surface of the vocal fold mucosa. A finite propagation velocity of the mucosal wave supplies the phase delay, and thus the two approaches

have merged conceptually into one theory of vocal fold vibration. The theory is based on small-amplitude assumptions, with flow-induced vibration being sustained around a small glottal opening.

More recently, the clinical importance of the mucosal wave has been described by Bless *et al.* (1987). Patients with vocal fold lesions and scarred tissue in the mucosa show a reduced amplitude of the mucosal wave. Furthermore, the *phonation threshold pressure* (the minimum lung pressure required to establish phonation at a given pitch) has been shown to be related to the mucosal wave velocity (Titze, 1988; Verdolini-Marston *et al.* 1990), suggesting that "ease of phonation" is facilitated by a greater phase delay.

Given these theoretical and clinical observations, it seems logical to conduct an investigation to measure the phase delay and the associated mucosal wave velocity. In particular, it is important to know how these quantities vary with fundamental frequency (F_o). This would then allow better interpretation of phonation threshold pressure, which is known to vary nonlinearly with F_o (Titze, in press). One criterion for successful phonosurgery might be the lowering of phonation threshold pressure (Wexler *et al.* 1989), which expands the range of intensity that can be achieved in phonation.

Relations Between Mucosal Wave Velocity and Vertical Phase Delay

Hiroto (1966) was one of the first to point out that the maximum glottal width, as seen from above, depends on time-delayed movements between the upper and lower lips of the vocal folds. The glottal width is generally a time and space dependent function that varies in both the anterior-posterior direction and the inferior-superior direction. It can be identified at two vertically separated fleshpoints on the medial surface of the folds, as shown in Figure 1(a). At these points, the time variation of the glottis is modeled as

$$g_1(t) = Max[0,\ g_{01}+2a_1\sin(2\pi F_o t)] \tag{1}$$

$$g_2(t) = Max[0,\ g_{02}+2a_2\sin(2\pi F_o t-\phi)] \quad , \tag{2}$$

where the subscript 1 refers to the lower fleshpoint and subscript 2 to the upper fleshpoint. The quantities g_{o1} and g_{o2} are pre-phonatory glottal widths, a_1 and a_2 are the amplitudes of (assumed) sinusoidal fleshpoint motions in the medial-lateral direction, and ϕ is the vertical phase delay. An assumption of symmetry between left and right vocal fold movement is implied in the equations, as indicated by the factor of 2 in front of the amplitudes of vibration. Time variations $g_1(t)$ and $g_2(t)$ are shown in Figure 1(b) for two cycles of vibration. (Some features of this figure will be explained later).

It should be kept in mind that the specific glottal widths $g_1(t)$ and $g_2(t)$ are defined only for one common anterior-posterior position. A more complete model of static and dynamic glottal shapes, involving variations along the length of the vocal folds, has been proposed (Titze, 1989), but this additional complexity is not needed for the present study.

Assuming a surface wave to be propagating vertically upward in the mucosa with a velocity c, the time delay between the two fleshpoints separated by a vertical distance z is

$$\tau = \frac{z}{c} \tag{3}$$

and the corresponding phase delay is this time delay normalized to the fundamental period of vibration T,

$$\phi = 2\pi\frac{\tau}{T} = 2\pi F_o\tau = 2\pi F_o z/c \qquad (4)$$

By knowing the fleshpoint separation z (Figure 1a) and the fundamental frequency F_o, a measurement of the phase angle ϕ will reveal the propagation velocity c of the mucosal wave. The experimental procedure for obtaining these measures will now be described.

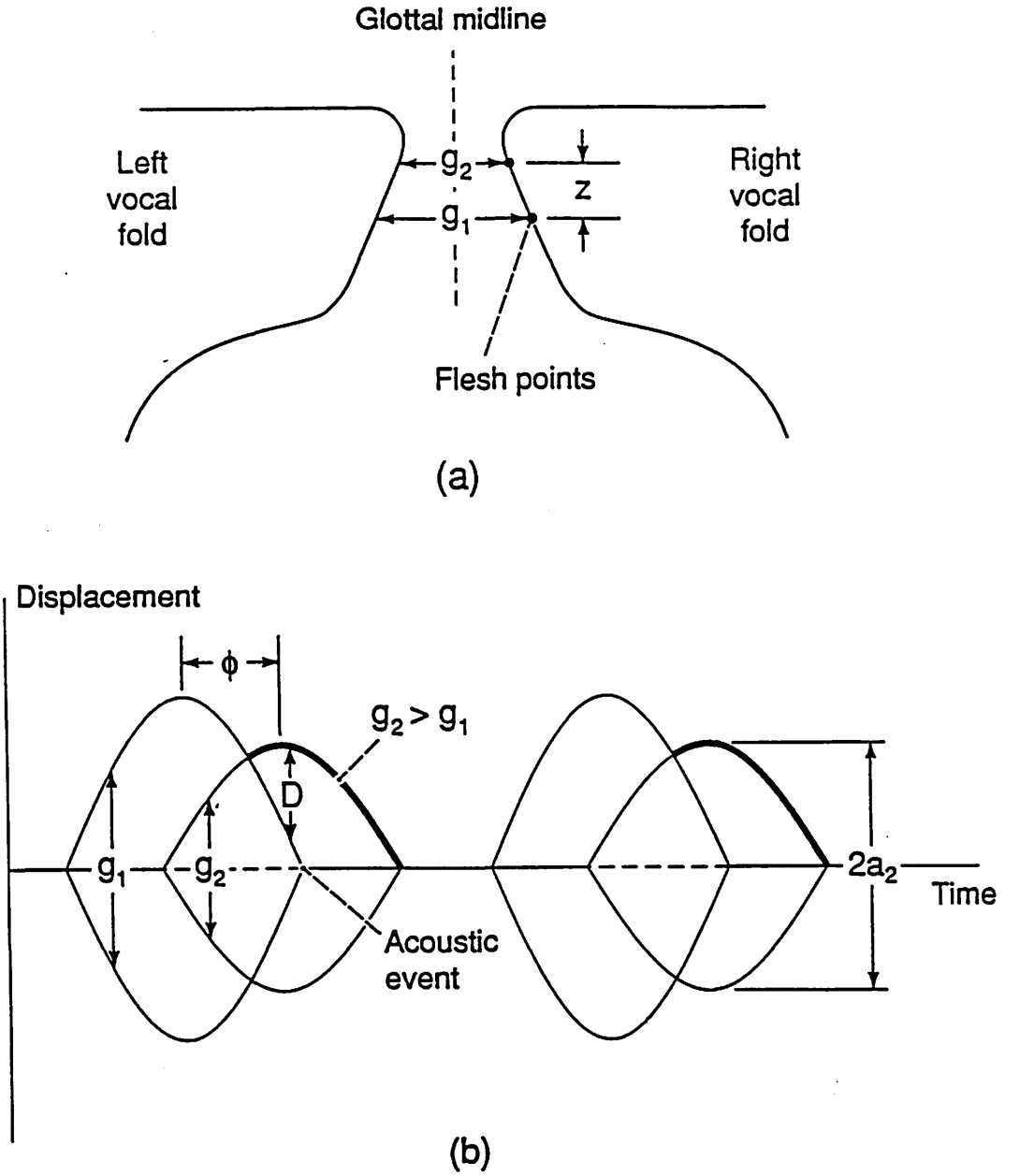

Figure 1. Illustration of the movement of fleshpoints on the upper and low lips of the vocal folds. (a) frontal section through vocal folds, (b) displacements versus time for glottal widths. The heavy lines illustrate portions of the cycle over which measurements can be made.

Materials and Methods

Excised canine larynges were harvested from mongrel dogs (body weights 20 to 23 Kg). The dogs had been sacrificed in a cardiovascular research laboratory. Only grossly normal larynges were used in this study. From approximately 10 larynges that were harvested, five were judged normal and kept for experimentation. Rejection was based on insufficient size, evidence of pathology, or obvious asymmetries across the midline.

Techniques for dissecting, mounting, and controlling the glottal configuration of the larynx are described elsewhere (Durham *et al.* 1987). Pressurized air with controlled temperature and humidity was used to induce self-oscillation of the vocal folds, as described by Baer (1975), Durham *et al.* (1987), and Cooper (1986).

A custom-designed triggering circuit was used to phase-lock the flashes of a stroboscopic light source (Pioneer DS 330-ST) to specific events in the period of vibration. The stroboscopic image of the top view of the folds, which filled the entire screen of a 19 inch video monitor, could thus be frozen to make measurements (in screen coordinates) of glottal dimensions at any given phase in the cycle. Magnification was 10:1, based on an object-lens distance of 45 cm (Sony DXC-102 video camera with a 90 mm microlens).

Prior to mounting the larynx, two small suture marks had been placed at the middle of the membranous vocal folds using 9-0 nylon stitches. One mark was at the upper lip and the other approximately 2 mm lower. The vertical distance between these two marks was measured with a micrometer (0.1 mm accuracy) and constituted the distance z in Figure 1(a) and in equations (3) and (4).

Different subglottal pressures and different elongations of the vocal folds were used to obtain different fundamental frequencies. The two audio channels of a high fidelity video cassette recorder were used to record two time-locked acoustic signals alongside the stroboscopic video image. One acoustic signal was the audio signal from a microphone (Realistic 33-1056A) placed 5 cm from the vocal folds, and the second signal was a timing signal obtained from a light sensor that converted the stroboscopic flashes into electric signals. In this manner, video events could be correlated with a constant acoustic event in the vibratory cycle. Figure 2 shows an example of the two time-locked audio signals. The interval t defines the time between the strobe flash (top trace) and the initial burst of acoustic pressure and glottal closure (bottom trace). The burst of acoustic pressure was taken to be the constant reference event, also shown in Figure 1(b). Phase differences between fleshpoint movements could then be determined relative to this reference event, and relative to each other.

Data Reduction

During video playback, the lateral displacement D between the two marks was measured for each of a number of phases. According to equations (1) and (2), this displacement is

$$D = g_2 - g_1 = (g_{02} - g_{01}) + 2a_2\sin(2\pi F_0 t - \phi) - 2a_1\sin(2\pi F_0 t) \quad . \tag{5}$$

The quantity D and the phase angle ϕ are shown in Figure 1(b). Only those phases where $g_2 > g_1$ could be used for measurement (thick portion of g_2) because in all other cases the upper mark shadowed the lower mark. Given that the glottis is closed over about half of the period, and given that

over about one quarter of the period the bottom mark is shadowed, the average measurement period was only about one quarter of a cycle (duration of thick line).

The acoustic signals from audio playback were displayed on a digital oscilloscope (DATA 6000). The stroboscopic reference time t was measured along with each value of D. More than 10 sets of D versus t data could be obtained for each pitch of a given larynx. To determine additional unknown factors in equation (5), the maximal glottal width at the upper fleshpoint ($2a_2$, shown in Figure 1b) was also measured separately on the TV screen. Finally, the quantity g_{02}-g_{01} was determined to be negligible because the marks were only about 1.5-2.0 mm apart and the pre-phonatory glottis was nearly rectangular. This left only a_1 and ϕ as unknowns.

Using nonlinear regression curve fitting (Marquardt-Levenberg algorithm in Sigmaplot 4.0 program), the data were fitted to equation (5), where a_1 and ϕ were parameters to be optimized under program control. The criterion for optimization was the combined least squared difference between the measured and calculated D's for the 10 or more data points.

Examples of best-case and worst-case matches between measurement and model of D as a function of t are given in Figures 3 and 4, respectively. In the best case (Figure 3), a sinusoidal difference function for D can readily be seen, both for the data and the model (the difference between two sinusoids of different phase and different amplitude is also a sinusoid). In the worst case (Figure 4), the sinusoidal nature of the data was obscured, but the model was nevertheless deemed appropriate. Note that the curves follow the pattern one would predict from Figure 1(b). The difference function D is maximum near t = 0, the acoustic reference event, and diminishes in both directions on the thick line.

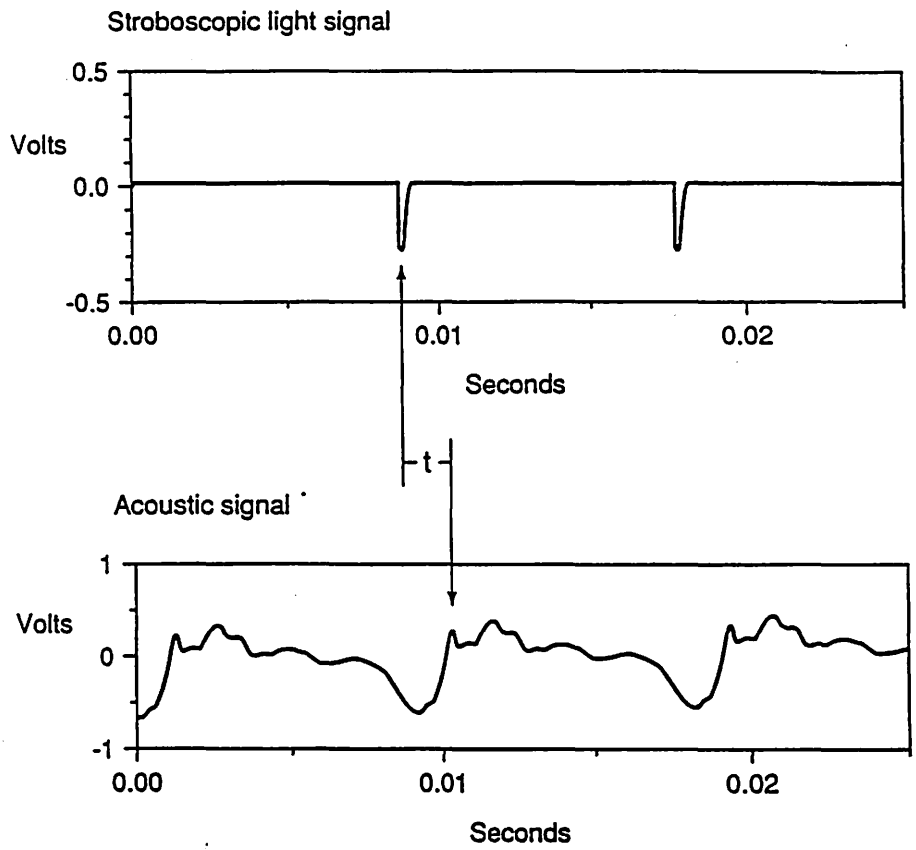

Figure 2. Display of simultaneously recorded strobe flash signal (upper trace) and microphone signal (lower trace). The time interval t is the relative time between two identifiable events in the cycle.

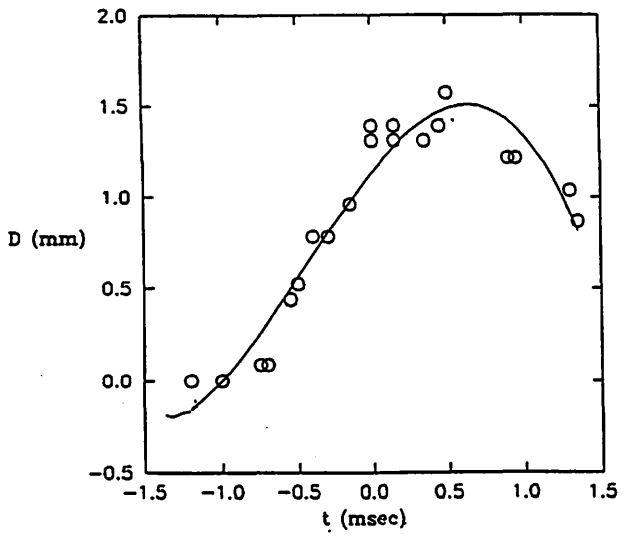

Figure 3

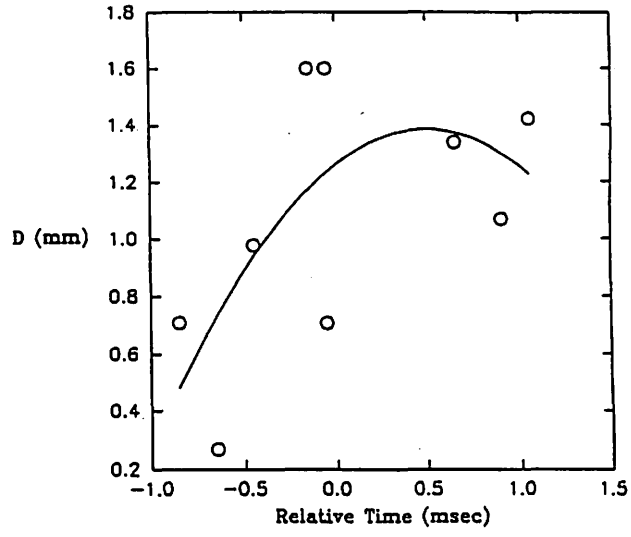

Figure 4

Figure 3. Distance D between two marks on the vocal folds as a function of relative time t for the best case. The solid curve is the model given by equation (5). Figure 4. Distance D between two marks on the vocal folds as a function of relative time t for the worst case. The solid curve is the model given by equation (5).

Results

Table 1 shows data of mucosal wave velocity c and phase delay per mm (ϕ /z) from 5 excised canine larynges. At least four different fundamental frequencies were obtained on each larynx. The columns of data are related by the equation

$$\frac{\phi}{z} = 360\frac{F_o}{c} \quad , \tag{6}$$

which is equation (4) expressed in degrees rather than radians. Note that the velocities range from about 0.6 m/s to about 2.2 m/s, with higher values consistently corresponding to higher fundamental frequencies. The phase delay per mm decreases with F_o.

Figures 5 and 6 show the data of Table 1 graphically. Except for larynx L5, which shows an unusually steep rise of c with F_o (and a correspondingly steep fall of ϕ/z), the results suggest that wave velocity in the mucosa may increase linearly with F_o. A three-fold increase in c, from 0.5 m/s to 1.5 m/s, is observed over the 100-180 Hz range. If L5 were excluded, the increase would still be more than two-fold. Phase delay may decrease linearly, but at a relatively moderate rate of decline (less than 30% over the 100-180 Hz range, and less if L5 were excluded).

The rate of decline of ϕ/z with F_o may actually be over-estimated. As the vocal folds are elongated, they usually shrink in thickness to preserve overall tissue volume. A 30% elongation, for example, would decrease the separation z of the marks by 30%. This would then keep ϕ/z nearly constant with F_o and c would increase proportional to F_o. Further investigation of this effect is needed.

Table 1. Mucosal Wave Velocity and Corresponding Phase Delay

# of larynx	F_o (Hz)	Velocity c (m/sec)	Phase delay ϕ/z (Degree/mm)
1	95	0.587	58.3
	107	0.877	43.9
	114	0.673	61.0
	119	0.704	60.9
	123	0.858	51.6
	131	0.778	60.6
	134	0.817	59.1
	140	0.847	59.5
	157	0.946	59.8
	173	1.466	42.5
2	129	0.866	53.6
	136	0.934	52.4
	154	1.047	53.0
	163	1.124	52.2
	178	1.139	56.3
3	126	0.859	52.8
	141	0.940	54.0
	158	1.063	53.5
	163	1.086	54.1
	181	2.081	31.3
4	127	1.012	45.2
	134	1.147	42.1
	139	1.254	39.9
	148	2.177	24.5
5	137	0.951	51.8
	148	1.469	36.3
	152	1.764	31.0
	155	2.001	27.0

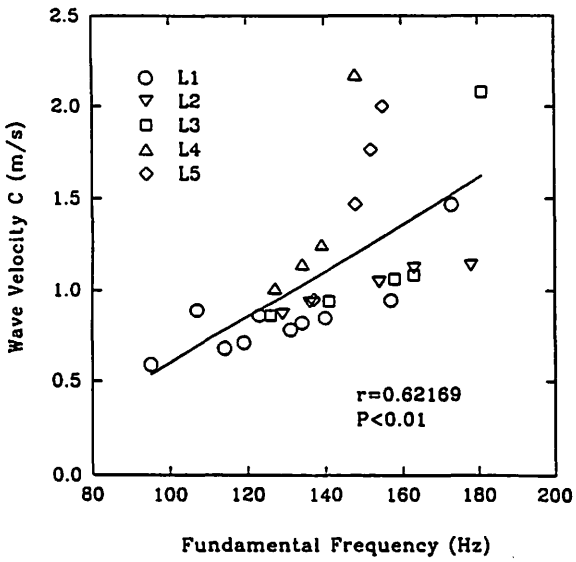

Figure 5

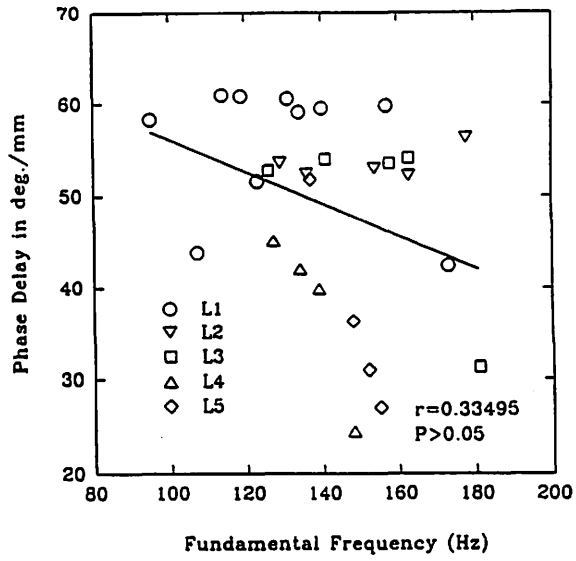

Figure 6

Figure 5. Wave velocity c versus fundamental frequency F_o for all five larynges. Figure 6. Phase delay in degrees per mm as a function of fundamental frequency F_o for all five larynges.

Discussion and Conclusions

In experiments conducted on excised canine larynges, Baer (1975) summarized his findings with respect to mucosal wave velocity as follows:

"The derived propagation velocities are on the order of 1.0 m/s. This value may be compared with the value of 1.6 m/s reported by van Gierke *et al.* (1952) on human skin and in such places as the thigh or forearm. As discussed earlier, however, wave velocity on the more lateral parts of the superior surface drops to 0.3 to 0.5 ms, perhaps because the membranes are more slack or because the effective thickness is greater."

It is clear from Figure 5 that Baer's value 1.0 m/s is in the center of the range measured here. Our findings, therefore, agree with his. Slower wave velocities (0.1 to 0.5 m/s) were measured by Hirano *et al.* (1981) on the superior surface. These findings do not conflict with our results in light of Baer's comments that mucosal waves slow down considerably as they propagate laterally. It is important, therefore, to specify where the wave velocity is measured. Hirano *et al.* did this carefully in their study.

The importance of quantifying the mucosal wave velocity on the medial surface stems from the fact that phonation threshold pressure, the minimum pressure required to establish phonation, depends critically on this propagation velocity (Titze, 1988). Thus, by knowing how c varies with F_o on this surface, a better understanding of phonation threshold pressure may emerge in the future. Direct or indirect estimates of this pressure may then lead to better tools for assessment of normal and abnormal vocal function.

Acknowledgement

This work was supported by the National Institutes of Health, grant No. P60- DC00976. The authors thank Julie Lemke for manuscript preparation.

References

Baer, T. (1975). Investigation of Phonation Using Excised Larynges. [Dissertation]. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Bless, D., Hirano, M., & Feder, R. (1987). Videostroboscopic evaluation of the larynx. Ear, Nose and Throat Journal, 66(7).

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

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

Hirano, M., Kakita, Y., Kawasaki, H., Gould, W., & Lambiase, A. (1981). Data from High-Speed Motion Picture Studies. In: K. Stevens & M. Hirano (Eds.), Vocal Fold Physiology (pp. 85-91). University of Tokyo Press.

Hiroto, I. (1966). The mechanism of phonation - Its pathophysiological aspects. Pract. Otolaryngol. (Kyoto) 39, 229-291. (In Japanese).

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

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

Titze, I. (1989). A four-parameter model of the glottis and vocal fold contact area. Speech Communication, 8(3), 191-201.

Titze, I. (in press). Phonation threshold pressure. Journal of the Acoustical Society of America.

Verdolini-Marston, K., Titze, I., & Druker, D. (1990). Changes in phonation threshold pressure with induced conditions of hydration. Journal of Voice, 4(2), 142-151.

von Gierke, H., Oestreicher, H., Franke, E., Parrack, H., & von Wittern, W. (1952). Physics of vibrations in living tissues. Journal of Applied Physiology, 4, 886-900.

Wexler, D., Jiang, J., Gray, S., & Titze, I. (1989). Phonosurgical studies: Fat-graft reconstruction of injured canine vocal folds. Annals of Otology, Rhinology, and Laryngology, 98, 668-673.

NCVS Status and Progress Report - 2
April 1992, 93-106

Levator Veli Palatini Muscle Activity in Relation to Intranasal Air Pressure Variation

David P. Kuehn, Ph.D.
Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign
Jerald B. Moon, Ph.D.
John W. Folkins, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Continuous positive airway pressure (CPAP) therapy can be used to reduce hypernasality by elevating the air pressure in the nasal cavities during speech. The purpose of this study was to determine whether increased intranasal air pressure loads the major muscle of velopharyngeal closure, the levator veli palatini. Nine subjects, four with cleft palate and five without cleft palate, were studied. Electromyographic activity was measured from the levator veli palatini muscle with several levels of air pressure delivered to the nasal cavities using a commercially available CPAP instrument. It was found that levator veli palatini activity was significantly greater for the positive air pressure conditions compared to the atmospheric pressure conditions for both subject groups. This indicates that the levator veli palatini muscle acts against the resistive load produced by the increased intranasal air pressure. The results support the use of CPAP therapy as a method of resistance exercise for strengthening velopharyngeal closure muscles.

Kuehn (1991) has described a new therapy procedure for reducing hypernasality using a device that delivers positive air pressure to the nasal cavities during speech. The device is used routinely in treating patients with obstructive sleep apnea syndrome (Schmidt-Nowara, 1984). Such treatment is referred to as continuous positive airway pressure or, more commonly, CPAP. The increased air pressure, relative to the prevailing atmospheric pressure, helps to maintain a patent upper airway during sleep in apnea patients. The elevated air pressure presumably provides a

resistance against the velopharynx and other anatomical sites thereby opposing closure of the airway (Kuna and Sant'Ambrogio, 1991).

Theoretically, when intranasal air pressure is increased, greater effort would be required during speech to close the velopharyngeal orifice. Such a condition could be used as a resistance exercise that might strengthen the velopharyngeal closure muscles. It is well-known that resistance exercises (e.g., weight-lifting) increase muscle strength (Atha, 1981; DiNubile, 1991). Komi (1986) pointed out that muscle strength can be increased by almost any method if the frequency or amount of activity exceed sufficiently those existing during the normal activation of the muscles. Resistance exercises are used commonly to strengthen normal muscles and such exercises also comprise an important aspect of muscle rehabilitation used by physical therapists (Kisner and Colby, 1990). The CPAP therapy procedure described by Kuehn (1991) could be considered as a type of resistance exercise even though a weight is not physically attached to the velopharyngeal structures.

There are several assumptions underlying CPAP therapy for reducing hypernasality that are consistent with those of any physical therapy exercise program for accomplishing other goals, specifically: 1) the velopharygeal closure muscles are subjected to a resistive load, in this case, increased intranasal air pressure, 2) such loading can be used in a systematic program of progressive resistance exercise, 3) resistance exercise eventually will strengthen the velopharyngeal closure muscles, and 4) stronger velopharyngeal closure muscles will be functionally useful in reducing hypernasality. We were interested primarily in addressing the first of these assumptions in the current study.

We chose to study electromyographic (EMG) activity of one muscle, the levator veli palatini, as a representative of the muscles of velopharyngeal closure. The levator muscle is considered to be the primary elevator of the velum and the major muscle of velopharyngeal closure (Fritzell, 1969; Bell-Berti, 1976; Kuehn et al., 1982). Therefore, in a resistance exercise involving velopharyngeal closure, the levator muscle is likely to be affected.

In keeping with the assumptions expressed above, any substantial increase in the activity produced by the levator muscle beyond its habitual level should increase its strength. Positive correlations have been reported between EMG levels and muscle force for those conditions in which muscle length and shortening velocity are constant (Bigland and Lippold, 1954; Bouisset, 1973; Milner-Brown and Stein, 1975). Because the aerodynamic loading employed in this study is not likely to change the length or shortening velocity of the muscles appreciably, it appears justified to compare normalized EMG levels across tasks. If greater EMG levels in levator are associated with increased levels of intranasal air pressure, it would indicate that the levator muscle contracts more forcefully to overcome the increased air pressure resisting velopharyngeal closure. We need to know this to test the first assumption listed above. Moreover, in reference to the second assumption, we were interested in determining whether subjects with cleft palate typically contract the levator muscle at or near its maximum level during speech or whether the muscle operates with a margin of reserve capacity that could be exploited in using progressive resistance exercises. Therefore, the purpose of this study was to measure the effect of increasing intranasal air pressure on the electromyographic activity of the levator veli palatini muscle during velopharyngeal closing movements.

Method

Subjects

In part, we were interested in determining whether <u>any</u> speaker's velopharyngeal mechanism would increase its activity in relation to increased loading thereby suggesting that the

velopharyngeal mechanism can be controlled, at least in some individuals, in relation to environmental factors. More specifically, however, the application of the CPAP resistance exercise program is aimed at individuals who do not have structurally or functionally normal mechanisms. So, it was important to determine whether structurally abnormal velopharyngeal mechanisms respond to changes in intranasal air pressure in a fashion similar to speakers with normal mechanisms. Two groups of subjects were studied, those with cleft palate (C) and those without cleft palate (N).

The subjects are described in Table 1. The subjects without cleft palate consisted of five normally speaking adults, three females and two males. These subjects did not report a history of speech, language, hearing, or craniofacial disorders. The investigators noted no abnormalities in the subjects' speech at the time of data collection.

The subjects with cleft palate consisted of four adults, three females and one male. All of these subjects had a history of borderline velopharyngeal inadequacy, but none were sufficiently severe to warrant a pharyngeal flap. Each of these subjects was evaluated subjectively by the first author at the time of data collection. Hypernasality was judged on a 7-point scale, with 1 being normal (i.e., no hypernasality) and 7 being extremely severe hypernasality. Subject C1 was assessed as 2, Subject C2 as 3, Subject C3 as 3, and Subject C4 as 3. Thus, some hypernasality was detected for all four subjects, with Subject C1 being the mildest and the other three being about equal in severity. It was observed by oral examination that the soft palate did move during phonation for all four subjects with cleft palate.

Table 1. Subject summary

Subjects	Sex	Age	Cleft Type
Noncleft			
N1	F	38	
N2	F	33	
N3	F	25	
N4	M	37	
N5	M	37	
Cleft			
C1	F	23	Unilateral lip and palate
C2	F	27	Hard and soft palate
C3	M	22	Unilateral lip and palate
C4	F	22	Submucous

Instrumentation

The device used to produce continuous positive airway pressure (CPAP) is available commercially (SleepEasy III, Respironics, Inc). The main components consist of an air flow generator, a valve system, a flexible hose, and a nasal mask. Further details are provided by Kuehn (1991). During the experiment, the output pressure was adjusted manually with the nasal mask on the subject. Pressure levels were measured with a U-tube water manometer.

Five air pressure conditions were employed. In one condition, the subjects did not wear the mask (mask off, 0 pressure condition). In the other four conditions with the mask on, intranasal air pressures delivered were 0 (machine turned off), 7.5, 10, and 12.5 cmH_2O. The sequence of these five conditions was randomly determined and different for each subject.

Stainless steel wire electrodes, 110 m in diameter, were used for recording EMG activity. The wires were inserted perorally using 1/2 inch 30 gauge hypodermic needles. The needles were inserted at an angle following the course of the levator muscle, that is, in a superior, lateral, and posterior direction. The two wires for bipolar recording were placed approximately 4 mm apart and 10 mm deep into the right levator veli palatini muscle. Placement criteria included EMG activity that was observed in association with sustained [s] production. The EMG signals were amplified using Biocommunications Electronics preamplifiers (model 301) and amplifiers (model 205).

Mask pressure was sensed with a Honeywell Microswitch pressure transducer (model 162PC01D) and amplified with a Biocommunications Electronics amplifier (model 205). The audio signal from a dynamic microphone was amplified using a Nakamichi preamplifier and Tascam tape recorder (model 22-4). EMG activity, pressure, and audio signals were monitored on a Tektronix storage oscilloscope (model 5111A) and recorded on a Sony digital instrumentation recorder (model PC108M).

Table 2. Task summary.

Task	Description	Pressure Values
[ansi]	Five pressure conditions each with 10 repetitions of [ansi]; marked pause etween [ansi] words	1) 0, mask off 2) 0, mask on 3) 7.5 cmH_2O 4) 10.0 cmH_2O 5) 12.5 cmH_2O
[sasasa...]	Ascending-descending condition with [sa] repeated continuously; no marked pause between syllables	1) From 5.0 cmH_2O to 12.5 cmH_2O back to 5.0 cmH_2O
Swallowing	Drinking water from a cup	1) 0, mask off

Tasks

The tasks are summarized in Table 2. The speech sample included the utterance [ansi] spoken in a natural manner with the second syllable stressed. The utterance was produced ten times during each pressure condition. The juxtaposition of [a] and [n] produced an opening of the velopharyngeal orifice. The sequence of [n] to [s] enhanced the probability of a forceful velopharyngeal closure gesture with a strong levator veli palatini contraction for [s]. Each production of [ansi] was followed by a pause of about two seconds; therefore, the utterances were produced as ten isolated words for each pressure condition.

In a second speech task, [sa] (isolated [s] for one subject) was repeated without a marked pause between syllables. The subjects were instructed to repeat the syllable continuously in a comfortable manner. This speech sample was chosen because [s] typically is associated with

velopharyngeal closure and the vowel facilitated subsequent segmentation during data analysis. During the production of [sa] utterances, pressure delivered to the nasal cavities was increased gradually from 5 to 12.5 cmH_2O then decreased gradually back to 5 cmH_2O. This task will be referred to as the ascending-descending condition. At least two such sequences were recorded for each subject.

A third task was swallowing. We were interested in determining whether systematic differences in levator activity occur between speech and swallowing and, if so, whether the differences are similar between individuals with and without cleft palate. Each subject was instructed to swallow saliva or water at various points in time during the experimental session.

Experimental Procedure

The oral cavity was sprayed lightly with 4% lidocaine topical anesthetic and the hooked-wire electrodes were inserted into the levator veli palatini muscle. The subject then was familiarized with the CPAP machine and allowed to practice speaking with the mask on and with different air pressures delivered to the nasal cavities.

The subjects were prompted to produce [ansi] by one of the experimenters using a hand signal to keep a fairly constant pace. The subject was instructed to hold the mask firmly against the face for an entire series of ten [ansi] productions, but to remove the mask between each pressure condition. Each subject produced all five pressure conditions with swallowing gestures interspersed between conditions. The ascending-descending conditions were recorded at the end of each experimental session.

Data Analysis

EMG signals were initially full-wave rectified and smoothed with a 40 ms time constant. Mask pressure signals also were smoothed with a 40 ms time constant. Rectified and smoothed EMG signals, smoothed pressure signals, and the audio signal were digitized with a rate of 1000 samples per second using a laboratory computer and commercially available analog-to-digital conversion software. Data then were displayed and analyzed using custom graphics and analysis routines.

The level of levator EMG activity associated with velar elevation for [s] in each [ansi] utterance was measured as follows. The EMG activity associated with the [s] burst of the syllable [si] was identified by one of the investigators. Starting and ending time points were marked using the cursor (Figure 1) and

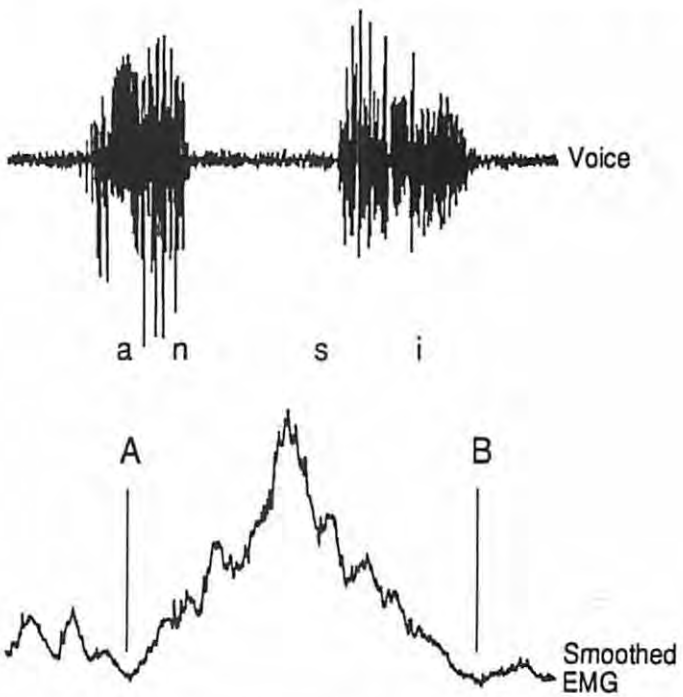

Figure 1. EMG area measure. Levator veli palatini trace in the lower portion and voice channel in upper portion for the utterance [ansi]. Vertical lines at A and B show cursor positions for beginning and ending points of measurement. The area under the trace between A and B was used as the dependent variable.

supplied to a custom computer algorithm that computed area under the EMG waveform for that time segment. EMG area values subsequently were normalized as a percentage of the largest area value produced for each subject on each production within a given pressure condition.

Interobserver reliability of EMG area values was determined by two observers who each measured four different sets of ten [ansi] values. A Pearson correlation coefficient of 0.97 was obtained between data from the two observers.

We were interested in determining whether levator activity between utterances might be influenced by the prevailing air pressure applied to the nasal cavities. If so, this would suggest that the levator muscle adjusts its tonic level of activity in response to changing environmental conditions. To assess the effects of mask pressure on tonic levels of levater EMG activity between [ansi] productions, the average level of levator muscle activity was computed over a 200 ms window in the interval between the offset of one [ansi] production and the onset of the next [ansi] production. One of the investigators marked the center point of the window that was associated with stable levator activity. The average level of EMG activity was calculated from this point.

Peak levels of activity during [ansi] productions were compared to those observed during swallowing to assess the relative level of levator activity during speech compared to that during a nonspeech task. The same starting and ending points used in the calculation of EMG area were supplied to a custom computer algorithm that searched for, and recorded, the peak level of EMG activity.

The entire ascending-descending pressure sequences involving [sa] repetitions were displayed on the computer screen. Obvious changes in levator EMG amplitude associated with change in pressure were noted, but no specific measurements were made on data from the ascending-descending task.

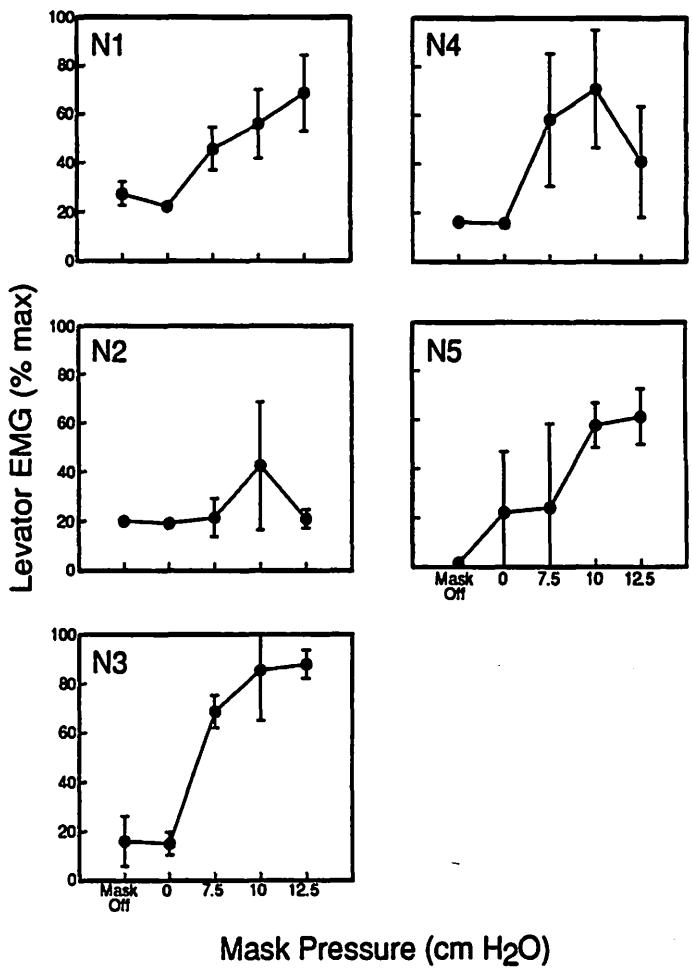

Figure 2. Noncleft palate subjects, average levator veli palatini EMG area levels in relation to mask (intranasal) air pressure values for [ansi] productions. Levator EMG measures are expressed as a percentage of the maximum area value across [ansi] productions for each subject.

Results

Word Production

The results for the productions of [ansi] are summarized in Tables 3a, 3b, and Figures 2 and 3. A split-plot analysis of variance was conducted with group (cleft, noncleft) and subject's given group as the among subject sources of variation and with mask pressure condition (mask off, 0, 7.5, 10.0, and 12.5 cmH_2O) as the within subject

source. The interaction of group-by-mask pressure also was examined to test whether the pattern of EMG variation across mask pressure conditions was similar in the two subject groups. Three designed contrasts were obtained partitioning the interaction source of variation. The first tested whether the mask off versus mask on 0 cmH_2O condition differed within and across groups. The second and third contrasts tested linear and nonlinear (quadratic and cubic) polynomial variations in levator EMG among the mask pressure conditions respectively. A final statistical analysis used a split-plot analysis of covariance design and utilized the group and subject given group sources to examine mask pressure as a quadratic regressor variable.

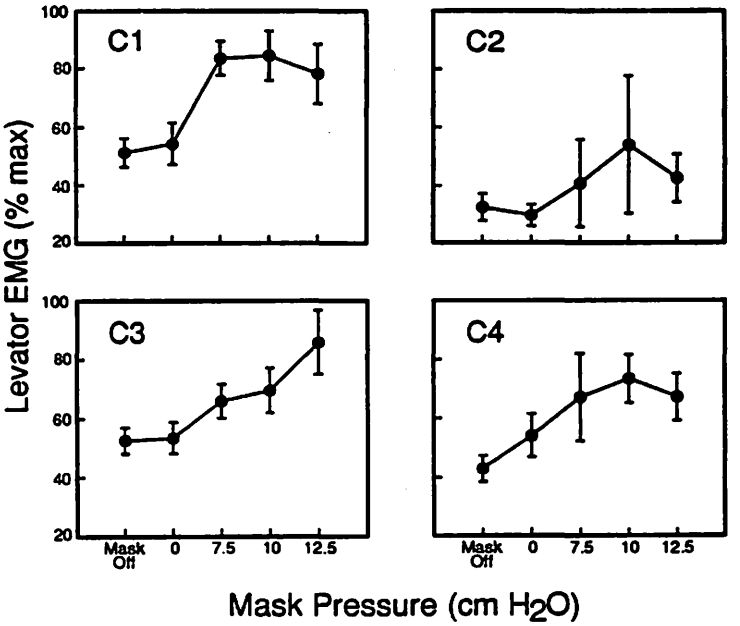

Figure 3. Cleft palate subjects, average levator veli palatini EMG area levels in relation to mask (intranasal) air pressure values for [ansi] productions. Levator EMG measures are expressed as a percentage of the maximum area value across [ansi] productions for each subject.

It was found that the zero pressure conditions, mask off and mask on 0 cmH_2O, were not significantly different from each other. Within both subject groups, each zero pressure condition was significantly different ($p<.05$) from all of the greater-than-zero pressure conditions. That is, compared to the zero pressure conditions, introduction of positive air pressure to the nasal cavities was associated with significantly greater levels of levator veli palatini EMG activity regardless of the magnitude of positive pressure introduced. In addition, for the subjects without cleft palate, levator activity was significantly greater at 10 than at 7.5 cmH_2O, and at 10 compared to 12.5 cmH_2O. For the subjects with cleft palate, there were no significant differences among the greater-than-zero pressure comparisons.

Inspection of data from individual subjects in Figures 2 and 3 reveals a similar pattern in the function relating levator EMG activity to mask pressure for six of the nine subjects. For these six subjects, there is a general increase in EMG activity with increases in pressure but then a decrease in EMG activity at the highest pressure, 12.5 cmH_2O, compared to the 10 cmH_2O condition. Across subjects within each group, the data conformed to a parabolic curve. However, the curve was significantly steeper for the subjects without cleft palate compared to the group with cleft palate. For example, levator activity increased relatively more between 0 pressure and 7.5 cmH_2O for the subjects without cleft palate than for the subjects with cleft palate.

Table 3. Results of analysis of variance comparing levator veli palatini EMG levels across pressure conditions. * $p<0.05$. Off = 0 pressure, mask off; 0 = 0 pressure, mask on; 7.5, 10.0, 12.5 = the corresponding mask pressure in cm H_2O.

A. [ansi] analysis; Subjects without cleft palate

	0	7.5	10.0	12.5
Off		*	*	*
0	.	*	*	*
7.5		.	*	
10.0			.	*

B. [ansi] analysis; Subjects with cleft palate

	0	7.5	10.0	12.5
Off		*	*	*
0	.	*	*	*
7.5		.		
10.0			.	

C. Interword analysis; Subjects without cleft palate

	0	7.5	10.0	12.5
Off		*	*	*
0	.	*	*	*
7.5		.	*	*
10.0			.	*

D. Interword analysis; Subjects with cleft palate

	0	7.5	10.0	12.5
Off		*	*	*
0	.	*	*	*
7.5		.		
10.0			.	

Interword Levels

The results for midpoint values for the intervals between [ansi] utterances are shown in Tables 3c, 3d, and Figures 4 and 5. These data represent the tonic levels of levator veli palatini muscle activity between isolated words. The statistical analyses were the same as those used for the intraword comparisons described above.

The interword results are similar to the intraword results. No significant difference was found between the zero pressure conditions, mask off versus mask on, for either group. However, each zero pressure condition was significantly different from all greater-than-zero pressure conditions for both groups. That is, for the intervals between words, the tonic muscle activity for levator was greater with elevated intranasal pressure levels compared to atmospheric intranasal pressure conditions. For the subjects without cleft palate, levator activity was significantly greater for 10.0 cmH_2O compared to 7.5 cmH_2O, greater for 12.5 cmH_2O compared to 7.5 cmH_2O, and greater for 10.0 cmH_2O compared to 12.5 cmH_2O. There were no significant differences among greater-than-zero pressure comparisons for the subjects with cleft palate.

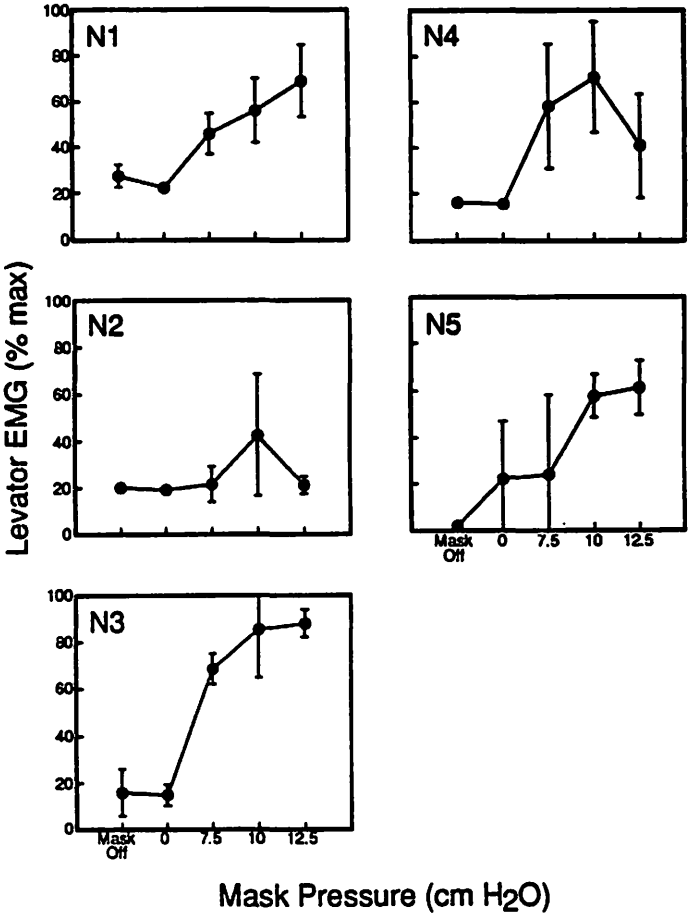

Figure 4. Noncleft palate subjects, average levator veli palatini EMG area levels in relation to mask (intranasal) air pressure values for intervals between [ansi] productions. Levator EMG measures are expressed as a percentage of the maximum area value across all interword intervals for each subject.

Ascending-Descending Condition

For the ascending-descending condition, mask pressure was varied continuously and gradually from 5 up to 12.5 and back to 5 cmH_2O while the subject repeated the syllable [sa] with only a brief pause between syllables. Figure 6 shows examples for SubjectsN5 and C3. Both subjects demonstrated discrete pulses of activity for mask pressure and levator muscle activity associated with repeated [sa] syllables. Subject N5 shows greater mask pressure changes than Subject C3 for each syllable. Although velar position was not monitored in this study, it is likely that greater velar movement occurred for Subject N5 than for Subject C3 during the syllable repetition task. Each dip in the pressure curve suggests a velar lowering movement resulting in a lowered mask pressure.

Only three of the nine subjects, N1, N3, and N4, exhibited some increase in levator EMG activity levels across the gradually changing pressure. Figure 7 shows two ascending-descending sequences for Subject N3. Subject N3 repeated isolated [s] sounds whereas the other subjects repeated [sa] syllables. A slight elevation in the overall EMG interference pattern can be observed in association with the elevated portion of the mask pressure curve. Subject N3 does not show the discrete EMG and pressure pulses associated with individual syllables that are evident for Subjects

N5 and C3 (Figure 6). Presumably, the strategy of N3 was to elevate the velum and keep the levator active without appreciable lowering movements between utterances. It is interesting to note, however, that for one of her ascending-descending sequences (Figure 7) a marked reduction in levator activity, indicated by the dip in the EMG trace, is followed closely in time by a pronounced dip in the pressure trace suggesting that the velopharyngeal port opened momentarily.

The mask pressure at the beginning of the ascending-descending condition was about 5 cmH_2O. Most of the subjects began the sequences at a levator EMG activity level that was higher than that for the zero pressure, mask off condition. Table 4 shows the average levator EMG peak activity levels across the second, third, and fourth syllables compared to the average peak levels across [s] in [ansi] for the zero pressure, mask off condition. Subjects N1 and C1 began the ascending-descending series with a levator EMG level lower than the average value for [ansi] zero pressure, mask off condition.

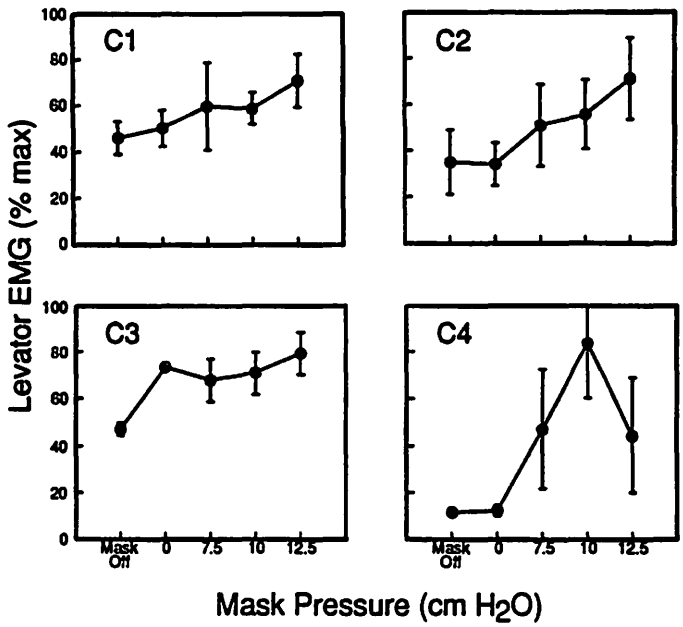

Figure 5. Cleft palate subjects, average Levator veli palatini EMG area levels in relation to mask (intranasal) air pressure values for intervals between [ansi] productions. Levator EMG measures are expressed as a percentage of the maximum area value across all interword intervals for each subject.

Table 4. Average peak levator veli palatini EMG values in arbitrary units above baseline. Off = [ansi] utterances, 0 pressure, mask off condition. Asc-Des = ascending-descending series, [sa] syllables, pressure approximately 5.0 cmH_2O. * = those subjects exhibiting a tendency to increase levator activity with an increase in mask pressure during the ascending-descending condition.

Subject	Off	Asc-Des
N1	567	253 *
N2	480	573
N3	534	708 *
N4	506	612*
N5	654	691
C1	375	306
C2	334	350
C3	636	704
C4	786	896

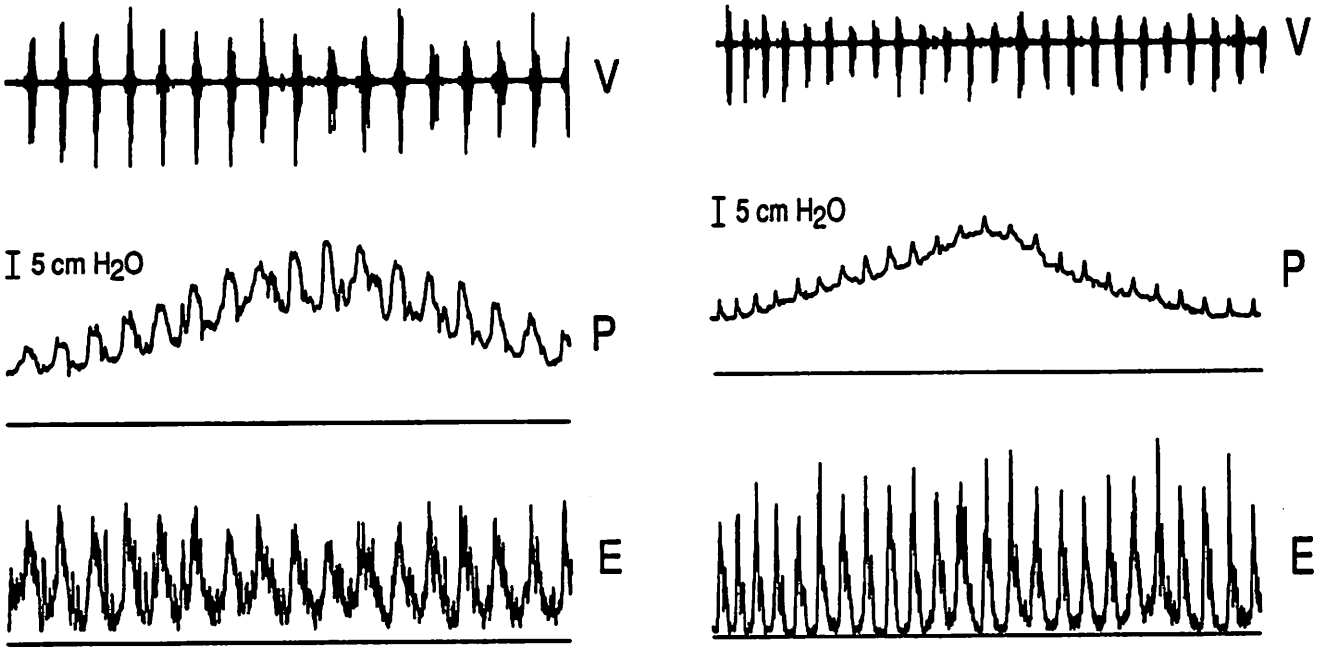

Figure 6. Ascending-descending series. V = voice trace, P = pressure trace, E = EMG trace. A.., at left, Subject N5 (without cleft palate) B., at right, Subject C3 (with cleft palate).

Swallowing Compared to Speech

Table 5 shows levator EMG peak levels for swallowing compared to the [ansi] zero pressure, mask off condition and to the single maximum peak level across the entire [ansi] data set. Levator EMG activity was greater during swallowing than during [s] in [ansi] for Subjects N1, N3, N4, N5, C1, and C2. For two of these subjects, N1 and C1, levator activity for swallowing exceeded the highest peak value across the [ansi] utterances. Levator EMG activity during swallowing was less than during speech for Subjects N2, C3, and C4. For all subjects in both groups, the maximum peak value across all [ansi] utterances, which typically occurred during the 10 cmH_2O pressure condition, was well above the average value across the zero pressure, mask off condition. As shown in Table 5, no clear difference emerged in the swallowing and speech data for the two groups. Across subjects, both groups behaved in a similar fashion.

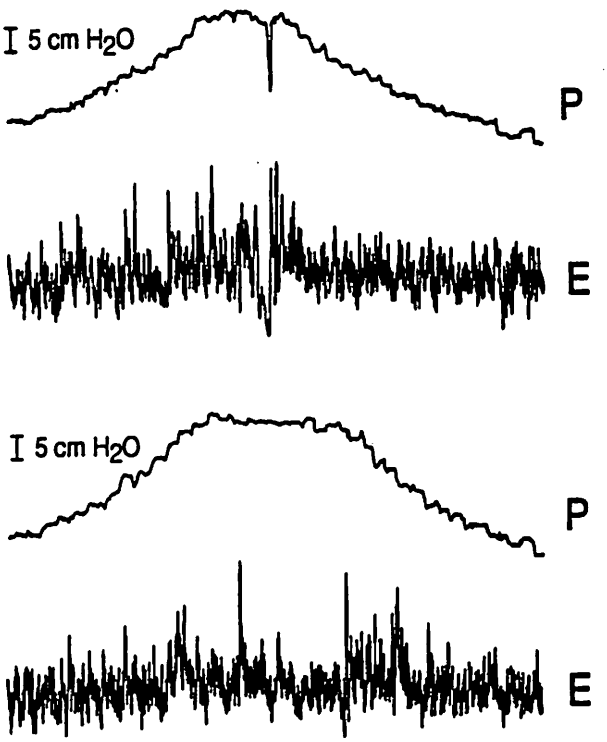

Figure 7. Two ascending-descending series. P = pressure trace, E = EMG trace. Subject N3 (without cleft palate). Upper pair shows dip in both P trace and E trace indicating a momentary opening of the velopharyngeal orifice.

Table 5. Levator veli palatini peak EMG levels during swallowing compared to speech. Swallow = average peak value across all swallowing gestures. Off = average peak value over ten [ansi] productions, 0 pressure, mask off. Maximum = the maximum peak value in the subject's entire [ansi] data set.

A. Subjects without cleft palate

	N1	N2	N3	N4	N5
Swallow	712	441	558	1142	686
Off	567	480	534	506	654
Maximum	696	762	1338	1665	1041

B. Subjects with cleft palate

	C1	C2	C3	C4
Swallow	715	656	420	652
Off	375	334	636	786
Maximum	675	1185	1248	1338

Discussion

The major finding in this study is that levator veli palatini muscle activity is greater with a positive air pressure delivered to the nasal cavities than without such positive pressure. Thus, with the introduction of a positive pressure via the CPAP device, the levator muscle contracts more forcefully and, when subjected to such a resistance exercise situation, it may increase its strength in closing the velopharyngeal orifice. This lends support to the therapy procedure proposed by Kuehn (1991).

The optimum CPAP pressures to be used in resistance exercises cannot be determined from this study and await further investigation. However, this study does provide information about the limits of pressure that might be useful. Several subjects produced no increase, or actually a decrease, in levator EMG activity at the highest pressure (12.5 cmH_2O) compared to the intermediate pressures. It may be that the velopharyngeal mechanism decreased its activity in relation to a resistance that was generally too great to overcome. This suggests that such higher levels of pressure (e.g., above 10 cmH_2O) should be avoided at least during initial therapy sessions.

Results for the intervals between words showed that the tonic levator muscle activity was significantly higher for the greater-than-zero conditions than for the zero pressure conditions. This suggests that the velopharyngeal mechanism may be primed in relation to the prevailing intranasal pressure. The levator muscle may be postured in a ready position to overcome increased intranasal air pressure that is introduced. This may be an added benefit in CPAP resistance therapy in that levator motor units would tend to be active continuously in the presence of increased intranasal pressure. Folkins (1985) has discussed changes in speech motor behavior that might accrue as a result of environmental factors and referred to such changes as "forced-variation flexibility."

The results pertaining to the ascending-descending series were somewhat surprising. Instead of increasing levator muscle activity in concert with gradually increasing intranasal pressure, levator

activity did not change appreciably during the ascending-descending condition for most subjects. That is, most subjects adopted a strategy that seemed to result in rather constant levator activity regardless of the changing circumstance. Whereas the levator EMG levels for the ascending-descending series typically were above those for the zero pressure condition (Table 4), the former were typically well below the maximum levator EMG levels as judged by comparing ascending-descending values in Table 4 to maximum values in Table 5. Therefore, if the motive is to elicit high levels of levator EMG activity, gradually increasing intranasal pressure during syllable repetition appears to be less effective than utilizing a constant positive pressure with production of isolated words.

With regard to the general operating level of the levator muscle for speech and swallowing, it was thought that in the subjects without cleft palate, the levator muscle might exhibit less activity for speech than for swallowing. In contrast, the levator muscle in individuals with cleft palate might function closer to its maximum activity level for speech in an attempt to achieve velopharyngeal closure and, therefore, the difference between levator activity for speech versus swallowing might be less than in normal speakers. It was found that levels of levator EMG activity were not consistently larger or smaller for swallowing compared to speech across subjects (Table 5). Therefore, swallowing does not necessarily appear to elicit the greatest amount of activity from levator in either subject group. In addition, the results do not lend support to the notion that the levator muscle in individuals with cleft palate generally operates near its maximum capacity during speech.

All subjects demonstrated a large difference between their typical levator activity levels for speech with normal intranasal air pressure compared to their peak activity levels in the presence of elevated intranasal pressure. This suggests that the levator muscle possesses some reserve in that the operating range observed was below the maximum capability. This appeared to be the case for the subjects with cleft palate as well as without cleft palate, even though all of the former exhibiting some degree of hypernasality. This finding suggests that the levator muscle can be induced to contract more forcefully allowing the possibility of long-term increases in strength through progressive resistance exercises.

Consistent differences between the subject groups with and without cleft palate generally were not found. In fact, the only significant difference between groups was that the function relating changes in levator activity to changes in intranasal pressure was steeper in the subjects without cleft palate. It is possible that the operating range for levator is a bit greater in normal speakers than in speakers with cleft palate.

The results of this study help to interpret the effects of CPAP therapy in attempts to reduce hypernasality. Further studies are warranted to determine whether the levator veli palatini muscle increases its strength over time with long-term utilization of increased intranasal air pressure. If levator muscle strength is increased, does this lead to tighter velopharyngeal closure perhaps over relatively longer speech segment intervals and, more importantly, is such modification in velopharyngeal closure related to a decrease in hypernasality?

References

Atha J. Strengthening muscle. Exer Sport Sci Rev 1981; 9: 1-73.

Bell-Berti F. An electromyographic study of velopharyngeal function in speech. J Speech Hear Res 1976; 19: 225-240.

Bigland B, Lippold OCJ. Motor unit activity in the voluntary contraction of human muscle. J Physiol 1954; 125: 322-335.

Bouisset S. EMG and muscle force in normal motor activities. In: Desmedt JE, ed. New developments in electromyography and clinical neurophysiology. Basel: Karger, 1973: 547-585.

DiNubile NA. Strength training. Clin Sports Med 1991; 10: 33-62.

Folkins JW. Issues in speech motor control and their relation to the speech of individuals with cleft palate. Cleft Palate J 1985; 22: 106-122.

Fritzell B. The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. Acta Otolaryngol 1969; 250 (Suppl): 1-81.

Kisner C, Colby LA. Therapeutic exercise (2nd ed). Philadelphia: FA Davis, 1990: 62.

Komi PV. Training of muscle strength and power: Interaction of neuromotoric, hypertrophic, and mechanical factors. Int J Sports Med 1986; 7 (Suppl): 10-15.

Kuehn DP. New therapy for treating hypernasal speech using continuous positive airway pressure (CPAP). Plast Reconstr Surg 1991; 88: 959-966.

Kuehn DP, Folkins JW, Cutting CB. Relationships between muscle activity and velar position. Cleft Palate J 1982; 19: 25-35.

Kuna ST, Sant'Ambrogio G. Pathophysiology of upper airway closure during sleep. JAMA 1991; 266: 1384-1389.

Milner-Brown HS, Stein RB. The relation between the surface electromyogram and muscular force. J Physiol 1975; 246: 549-569.

Schmidt-Nowara WW. Continuous positive airway pressure for long-term treatment of sleep apnea. Amer J Diseases Child 1984; 138: 82-92.

Acknowledgment

This study was supported in part by PHS Research Grants DE-05837, the National Institute of Dental Research, DC-00976, and DC-00085 from the National Institutes on Deafness and Other Communication Disorders.

NCVS Status and Progress Report - 2
April 1992, 107-117

Rapid Voice Tremor, or "Flutter," in Amyotrophic Lateral Sclerosis

Arnold E. Aronson, Ph.D.
Speech Pathology Section, Mayo Clinic and Mayo Foundation
Lorraine Olson Ramig, Ph.D.
Department of Communication Disorders and Speech Science, University of Colorado
and
Recording and Research Center, Denver Center for the Performing Arts
William S. Winholtz, A.A.S.
Recording and Research Center, Denver Center for the Performing Arts
Sandra R. Silber, B.Sc.
Speech Pathology Section, Mayo Clinic and Mayo Foundation

Abstract

In an attempt to clarify the origin and frequency characteristics of a rapid voice tremor, or "flutter," in patients with amyotrophic lateral sclerosis (ALS), eight patients (four male and four female; ages 42 to 70 years) who had ALS and rapid voice tremor and an age- and sex-matched control group of eight subjects were asked to sustain the vowel /a/ and their voice was recorded for later analysis. Each segment of phonation was demodulated into amplitude and frequency components. From each subject's 8-second amplitude and frequency signals, an FFT analysis was done on a 1-second segment previously identified perceptually as having the most apparent tremor, or "flutter."

The results showed that patients with ALS had multiple combinations of levels and frequencies for amplitude and frequency modulations in comparison with control subjects, who had consistently low levels of modulations.

In an attempt to quantify the tremor, or "flutter," in ALS, amplitude and frequency modulations were not clearly or predominantly represented at one point along the spectrum. Nevertheless,

these frequency and amplitude modulations are more prominent in patients with ALS than in normal subjects.

The origins of these aberrant frequency and amplitude modulations in ALS patients remain obscure, although speculation is that they are of peripheral rather than central nervous system origin.

Tremor is pervasive in neurologic disease. It also occurs in a normal nervous system. There is no uniform definition or classification of tremor, but students of the phenomenon agree that it is an involuntary, periodic oscillation of a body member, continuous and rhythmic in appearance and distinguishable from other involuntary movements. The frequency, waveform, and amplitude of tremor are known to be determined by the underlying pathophysiologic mechanism causing it.[1]

Normal individuals produce a physiologic tremor within a frequency range of 6 to 12 Hz.[2,3] Abnormal tremors in neurologic disease can be classified by frequency band. In the band from 1.5 to 3 Hz are found the coarse tremors in cerebellar ataxia involving the head, trunk, and upper extremities. In the band from 4 to 6 Hz are tremors that occur in parkinsonism. Patients who have Parkinson's disease produce voice tremor calculated to occur within the range of 4 to 7 Hz.[4]

Essential tremor of the limbs ranges from 4 to 9 Hz, and that of the head is about 4 Hz.[1] Organic voice tremor, a laryngeal form of essential tremor, can become so severe that during the adductor phase of the vocal fold tremor complete voice arrests can occur, producing a form of adductor spastic (spasmodic) dysphonia.[5]

Tremors in the band from 8 to 12 Hz, the highest found in humans, are interpreted as exaggerations of physiologic tremor and are most commonly associated with anxiety. However, high-frequency tremors also occur in central nervous system degenerative diseases and multifocal infarcts, in peripheral neuropathy in which there are individual violent muscle twitches, and in many states of drug intoxication.[1]

Typically, the bulbar form of amyotrophic lateral sclerosis (ALS) produces a flaccid dysarthria, spastic dysarthria, or both, depending on the stage of the disease. In fully developed flaccid-spastic dysarthria, the patients's phonation has a strained, wet hoarseness, resonation is hypernasal, and articulation is slow and imprecise.[6] The physical signs that underlie the dysarthria consist of unilateral or bilateral flaccid weakness of one or both true vocal folds; spastic hyperadduction of the false vocal folds; unilateral or bilateral weakness of the soft palate; weakness, atrophy, and fasciculation of the tongue; and weakness of the lips.

Subsequently, Aronson[7] added a further clinical observation to these already established dysarthric signs in ALS by noting that on vowel prolongation many ALS patients had a rapid voice tremor, or "flutter," that was almost undetectable during conversational speech. This tremor and its frequency and amplitude had not been reported previously, and a physiologic explanation for its presence is yet to be established. Acoustic spectrographic recordings of patients with ALS who have rapid tremor, or "flutter," produce sonograms with an obvious disturbance in the vowel /a/ prolongation (Fig. 1A) when compared with recordings of a normal subject (Fig. 1B). To the ear, the "flutter" is rhythmic, but its complexity and lack of demarcation among fluctuations on the sonogram prevent visual quantification.

This study was done to quantify the frequency of perceptual "flutter" in patients with ALS by means of a fast Fourier transform (FFT) analysis. Knowledge of the frequency band of flutter might provide further understanding of the site or physiology of the dysfunction producing it.

Method and Procedures

From 32 patients with ALS identified as having rapid voice tremor, or "flutter," on vowel prolongation, 8 (4 male and 4 female; ages 42 to 70 years) were chosen for analysis (Table 1). They were selected from the larger group of patients with ALS because of the unusual prominence of their tremors and their ability to sustain vowel prolongations for at least 10 seconds. An age- and sex-matched control group of eight subjects was also selected for this study. All control subjects were free of neurologic disease or speech disorders of any kind.

From each subject's prolongation of the vowel /a/, an 8-second segment was chosen from reel-to-reel recordings and digitized at 20,000 samples per second into a 16-bit analog-to-digital converter to facilitate repeatability during analysis (Fig. 2). These digitized segments were reconverted to analog at 20,000 samples per second through a 16-bit digital-to-analog converter and fed into the Vocal Demodulator, an instrument for analysis of low-frequency amplitude and frequency modulations.[8] The amplitude and frequency demodulated signals were digitized into separate channels of a DATA 6000 Universal Waveform Analyzer. The 14-bit converter was set to 555 samples per second.

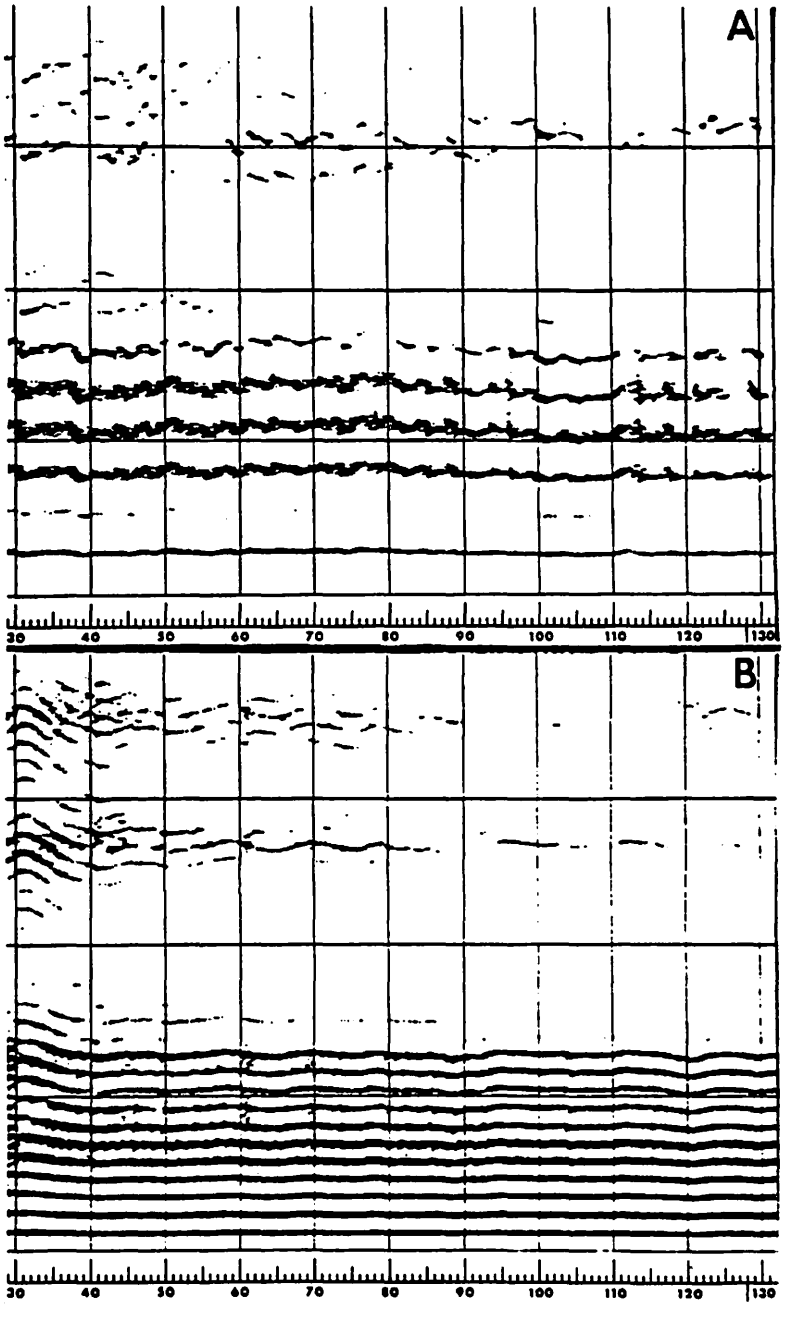

Fig. 1. A, Sonogram of vowel prolongation in a patient with ALS. B, Sonogram of vowel prolongation in a normal subject.

Table 1. Clinical data for eight patients with ALS who had rapid voice tremor

Case	Sex	Age, year-month: At exam	At onset of gen sx	At onset of sp sx	Bulbar signs only	Bulbar & spinal signs	Tongue fascic	Dysarthria: Flaccid only	Spastic only	Mixed sp/fl	Phonatory Defects: Strained-hoarse	Breathy	Wet-hoarse	Hyper-nasal	Articulation defect
1	F	68-6	64-11	68	-	+	+	-	-	-	+	-	-	+	+
2	F	67-4	66-2	65-11*	-	+	+	-	-	+	+	+	-	+	+
3	M	54	54-1	54-6*	+	-	+	-	-	+	+	-	-	-	-
4	M	63-9	62-3	62-5	-	+	+	-	-	+	-	-	+	+	+
5	M	51-10	49-10	51-4	-	+	+	-	-	+	-	+	-	+	-
6	F	69-8	69-4	68-10*	-	+	+	-	-	+	+	-	+	+	+
7	F	55-10	55-2	54-11*	-	+	+	-	-	+	+	-	-	-	+
8	M	41-9	61-3	41-1*	+	-	+	-	-	+	+	-	-	+	+

*Speech changes preceded other neurologic signs.

Abbreviations: exam, examination; fascic, fasciculations; gen sx, general symptoms; sp/fl, spastic/flaccid; sp sx, speech symptoms; +, present; - absent.

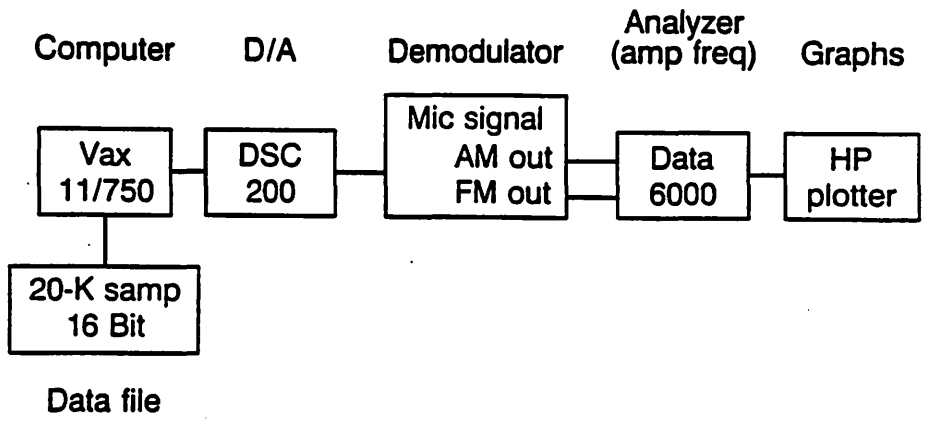

Fig. 2. A block diagram of instrumentation for analysis of acoustic data. D/A, digital to analogue; DSC, Digital Sound Corporation; HP, Hewlett-Packard; Mic, microphone; samp, sampling rate; Vax, computer.

From each subject's 8-second amplitude and frequency signals, an FFT analysis was done on a 1-second segment identified perceptually as having apparent tremor, or "flutter." Because the amplitude and frequency of "flutter" varied within the 8-second sample, a 1-second analysis duration was selected to allow a precise sampling of the "flutter" and to avoid averaging the sample with tremor components. Spectral frequency resolution was 0.54 Hz. Dominant spectral peaks were identified in the range of 1 to 25 Hz by two experimenters. The frequency (Hz) and amplitude (volts) for each peak were recorded, and spectral data were plotted on an HP 7475A plotter. For the patients with ALS, peaks below 25% of the most prominent peak were excluded. All identifiable peaks were recorded for the control group. Seventy-five percent of the ALS and control data were reanalyzed for assessment of measurement reliability. Pearson product moment correlation coefficients for repeated measures of these data are summarized in Table 2.

Table 2. Pearson Product Moment Correlation Coefficients for Repeated Measurements of FFTs From Demodulated Amplitude and Frequency Data

	Patient data	Control data
Amplitude spectral peaks, Hz	0.87	0.97
Amplitude spectral peaks, mV	0.99	0.99
Frequency spectral peaks, Hz	1.00	0.80
Frequency spectral peaks, mV	0.99	0.91

Results

Waveform and spectral data for each patient with ALS and corresponding control subject are given in Figures 3 through 18. Amplitude and frequency demodulation and corresponding spectra are represented in four traces. Panel A1 is the demodulated signal for amplitude, and panel A2 is the FFT of that signal. Panel B1 is the demodulated signal for frequency, and panel B2 is the FFT of that signal.

In patients with ALS, it is apparent that there were multiple combinations of levels and frequencies for amplitude and frequency modulations. For example, case 1 (Fig. 3) had regular amplitude and frequency modulations at a single frequency. This finding was apparent in both the demodulated data according to waveform similarities (Fig. 3, panels A1 and B1) and the corresponding spectral data (Fig. 3, panels A2 and B2). The prominent spectral peaks for both amplitude and frequency for this subject were 9.8 Hz. In contrast, case 2 (Fig. 5) had prominent amplitude and frequency modulations, these occurring at various multiple frequencies. This observation was apparent both in the demodulated data, as demonstrated by the waveform dissimilarities (Fig. 5, panels A1 and B1), and in the corresponding spectral data (Fig. 5, panels A2 and B2). Multiple prominent spectral peaks for this subject's phonation ranged from 1.1 to 25 Hz for amplitude and from 1.1 to 17 Hz for frequency. Case 3 (Fig. 7) had a higher level of modulation in amplitude than in frequency. This finding was apparent both in the magnitude dissimilarities of the demodulated waveforms (Fig. 7, panels A1 and B1) and in the corresponding spectral peaks (Fig. 7, panels A2 and B2).

In contrast to the patients with ALS, the control subjects had consistently low levels of modulations. Four of the eight control subjects (Fig. 4, 6, 14, 16) had a regular low-level frequency modulation, which is apparent in both the demodulated data and the spectral data. The prominent frequencies of these modulations ranged from 2.2 to 4.5 Hz.

Individual data for patients and controls are summarized in Table 3. These data include spectral peaks and levels measured for each subject. For patients with ALS, amplitude spectral peaks ranged from the cutoff frequency of 1.1 to 25 Hz, and frequency peaks ranged from 1.1 to 24.9 Hz. For control subjects, amplitude spectral peaks ranged from 1.1 to 24.9 Hz, and frequency peaks ranged from 1.1 to 25 Hz. The amplitude of the spectral peaks for patients with ALS ranged from 32.5 to 400.0 mV for amplitude and from 17.7 to 637.0 mV for frequency. For controls, the amplitude of the spectral peaks ranged from 9.7 to 91.8 mV for amplitude and from 7.3 to 134.8 mV for frequency. Patients had an average of four to five peaks in amplitude and frequency per analysis, and controls had an average of three to four peaks.

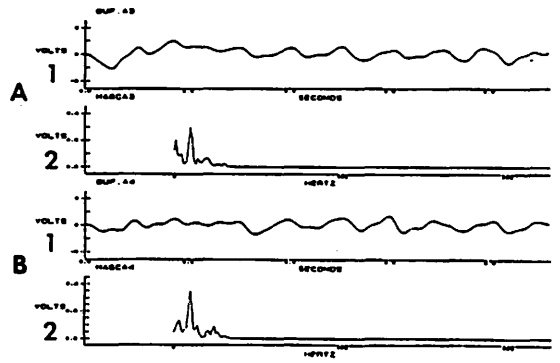

Figure 3

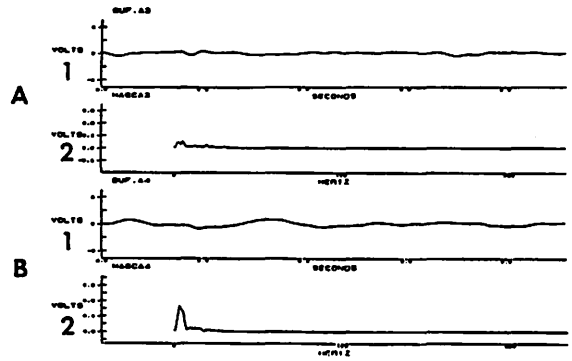

Figure 4

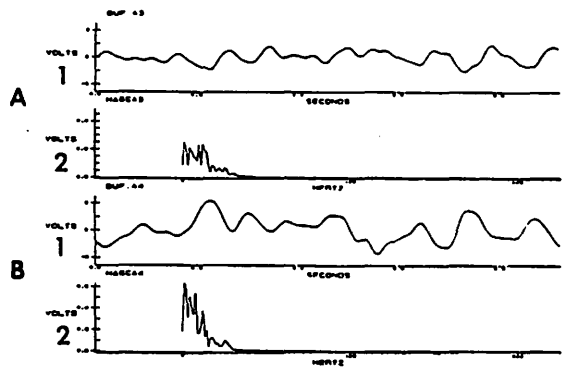

Figure 5

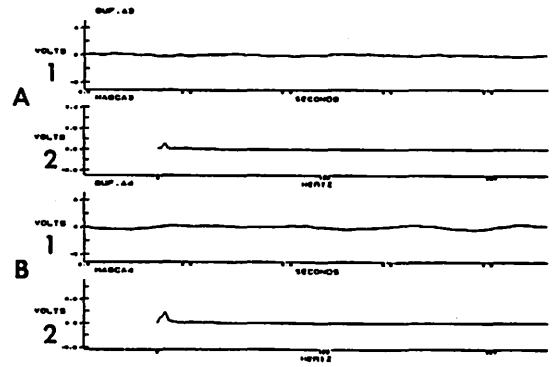

Figure 6

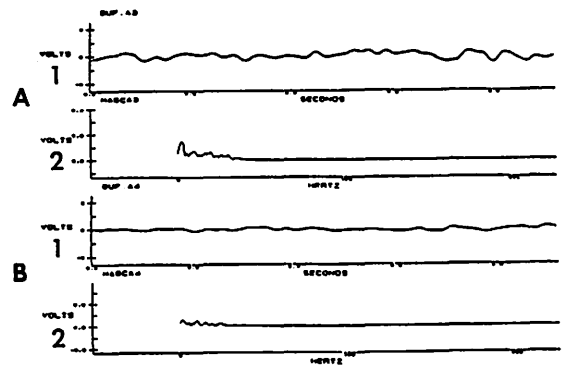

Figure 7

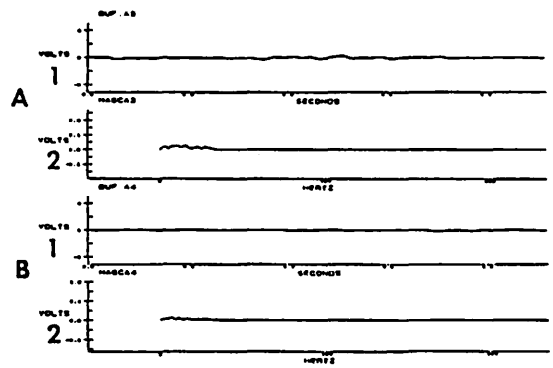

Figure 8

Figs. 3-8. Demodulated amplitude (A: panel 1, contour; panel 2, spectra) and frequency (B: panel 1, countour; panel 2, spectra). Figures 3, 5 and 7, patients with ALS. Figures 4, 6 and 8, control subjects.

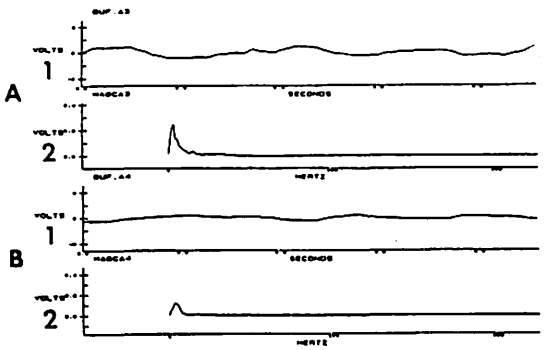

Figure 9

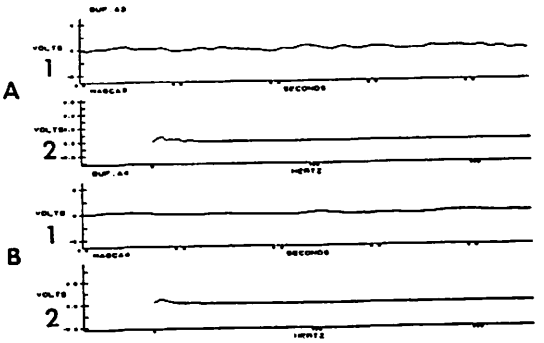

Figure 10

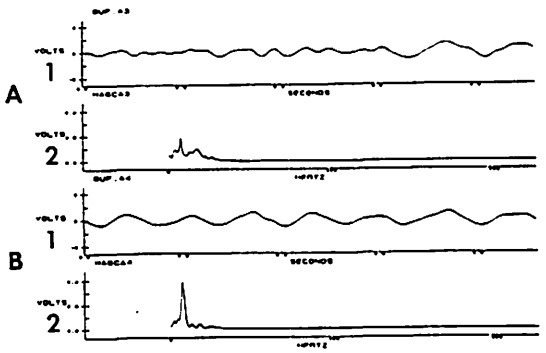

Figure 11

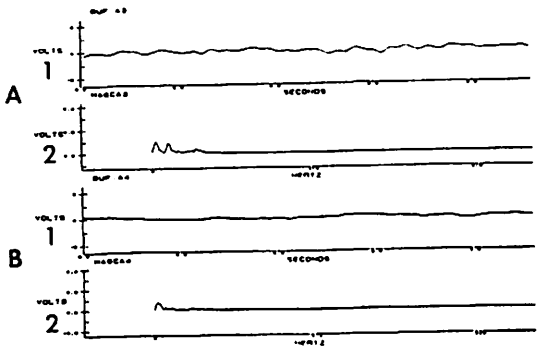

Figure 12

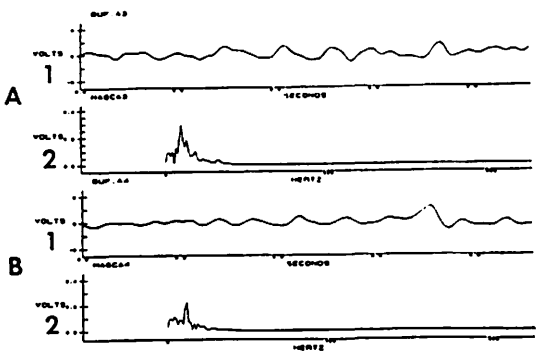

Figure 13

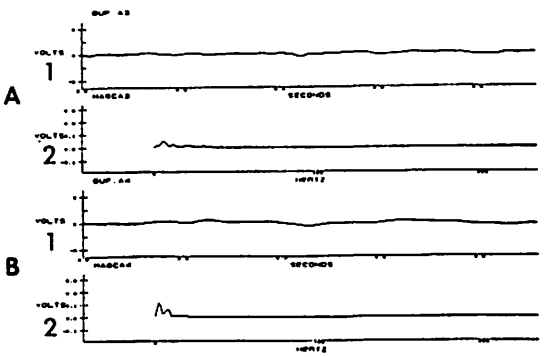

Figure 14

Figs. 9-14. Demodulated amplitude (A: panel 1, contour; panel 2, spectra) and frequency (B: panel 1, countour; panel 2, spectra). Figures 9, 11 and 13, patients with ALS. Figures 10, 12 and 14, control subjects.

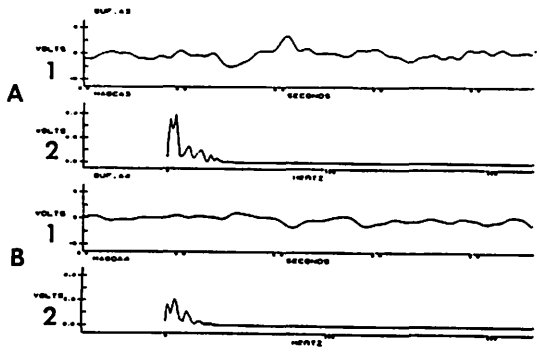

Figure 15

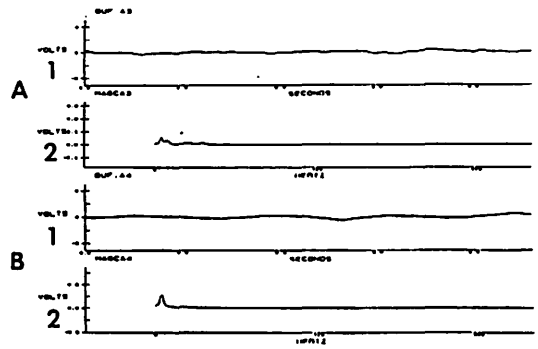

Figure 16

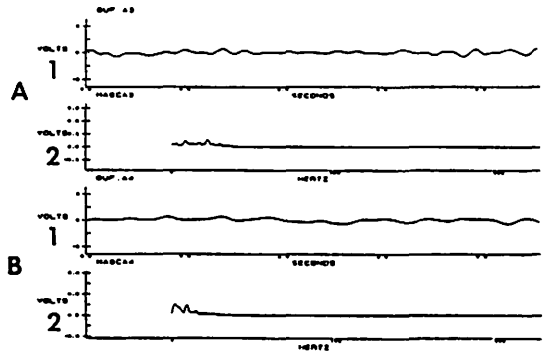

Figure 17

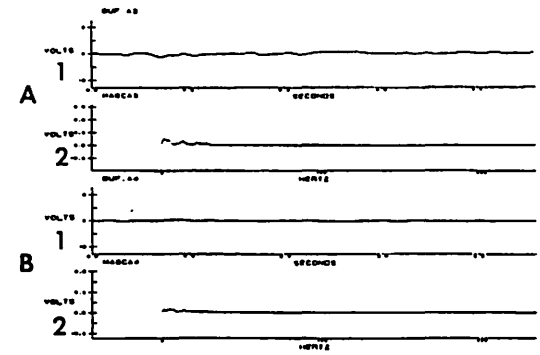

Figure 18

Figs. 15-18. Demodulated amplitude (A: panel 1, contour; panel 2, spectra) and frequency (B: panel 1, countour; panel 2, spectra). Figures 15 and 17, patients with ALS. Figures 16 and 18, control subjects.

Table 3. Frequency (Hz) and Amplitude (mV) for Amplitude and Frequency Spectral Peaks and Mean Fundamental Frequency (Hz) for Eight ALS Subjects with Vocal Tremor ("Flutter") and Eight Age- and Sex-matched Control Subjects[1]

	Amplitude spectral peaks		Frequency spectral peaks		
Case	Hz	mV	Hz	mV	Fo[2]
1 (ALS, Fig. 3)	1.1	204.5	3.3	134.4	286
	4.3	99.2	9.8	352.7	
	9.8	298.5	14.1	95.0	
	19.5	69.1	23.9	87.9	
2 (control, Fig. 4)	2.2	91.8	1.1	106.4	245
	9.8	33.7	3.3	126.8	
			4.3	131.0	
			9.8	51.0	
3 (ALS, Fig. 5)	1.1	249.8	1.1	637.0	326
	4.3	209.2	4.3	507.0	
	9.8	231.7	7.6	540.0	
	11.9	234.3	11.9	370.7	
	17.4	85.7	14.1	190.3	
	21.7	68.0	17.4	126.4	
	25.0	69.5			

Table 3, continued

Case	Amplitude spectral peaks		Frequency spectral peaks		Fo[2]
	Hz	mV	Hz	mV	
4 (control, Fig. 6)	4.3	42.7	3.3	134.8	268
	13.0	25.3	13.0	24.7	
	19.5	23.1			
	24.9	24.3			
5 (ALS, Fig. 7)	2.2	148.0	1.1	58.9	111
	5.4	60.6	2.2	51.4	
	9.7	76.1	5.4	34.4	
	16.3	55.5	7.6	30.2	
	19.5	62.4	9.7	52.5	
	23.9	32.5	10.9	43.5	
			15.2	34.7	
			18.4	17.7	
			23.9	25.9	
6 (control, Fig. 8)	3.3	33.2	3.3	13.3	119
	8.7	29.7	7.6	15.8	
	14.1	18.3	18.5	7.3	
	17.4	20.3	25.0	6.9	
	22.7	11.0			
7 (ALS, Fig. 9)	4.3	261.3	4.3	126.1	116
8 (control, Fig. 10)	6.5	83.8	1.1	26.3	146
	10.8	39.7	3.3	43.0	
	13.0	34.4	6.5	35.7	
	16.2	44.6	10.8	15.9	
	23.8	27.1	24.9	9.9	
9 (ALS, Fig. 11)	4.3	92.6	7.5	389.7	122
	7.7	186.9			
	13.0	82.3			
	17.4	101.8			
10 (control, Fig 12)	2.2	76.7	3.3	52.1	117
	4.3	58.2	9.8	37.1	
	6.5	55.7			
	11.9	38.8			
11 (ALS, Fig. 13)	2.2	94.3	2.2	97.7	175
	4.3	96.9	4.3	109.4	
	6.5	124.7	7.6	86.6	
	9.8	293.0	11.9	222.2	
	13.0	181.7	16.3	56.6	
	18.4	100.3	18.4	57.3	
12 (control, Fig. 14)	4.3	9.7	2.2	100.0	365
	9.7	38.6	6.5	37.2	
13 (ALS, Fig. 15)	2.2	363.2	1.1	166.7	166
	5.4	400.0	5.4	209.2	
	13.0	133.6	13.0	112.6	
	20.6	101.7			
14 (control, Fig. 16)	1.1	35.3	3.3	112.4	198
	5.4	48.6			
	10.9	24.4			
	16.3	22.4			
	23.8	17.8			
15 (ALS, Fig. 17)	1.1	30.3	2.2	106.0	166
	3.3	30.6	4.3	73.2	
	8.7	52.0	9.8	95.8	
	13.0	26.3	15.2	36.4	
	21.7	60.5			
	25.0	23.4			
16 (control, Fig. 18)	3.3	61.1	4.3	31.4	105
	7.6	27.5	7.6	28.1	
	10.9	22.0	10.9	18.6	
	14.1	35.4			
	20.6	15.9			
	23.8	18.3			

[1]Data were obtained from fast Fourier transform analysis of amplitude and frequency demodulated signals.
[2] Fo, fundamental frequency

Discussion

To the experienced listener, the tremor, or "flutter," during vowel prolongation in patients with ALS sounds rapid, somewhere between 7 and 10 Hz. The intent of this study was to attempt to quantify these oscillations.

Results of demodulation and spectral analysis revealed that ALS patients with perceptible tremor, or "flutter," produced both frequency and amplitude modulations of varied frequencies and magnitudes. In certain patients, amplitude modulations were most apparent, in other patients frequency modulations were most apparent, and in still other patients frequency and amplitude modulations were equal and in phase. The prominent frequencies spanned the entire analysis range of 0 to25 Hz; most patients had amplitude or frequency peaks within the range of 6 to 12 Hz. In contrast to the control subjects, the spectral peaks of the ALS patients were much more prominent.

The origins of these aberrant frequency and amplitude changes in the patients with ALS remain obscure. Although the ear hears these changes as some kind of tremor, or "flutter," quantification does not clearly establish that these fluctuations belong within the classification of "tremor," although whether they may be a matter of semantics. Of interest, regardless of what these perturbations are called, is their origins. We are fairly certain that in patients with purely spastic (that is, pseudobulbar) bilateral upper motor neuron dysarthria, the dysphonia does not contain audible tremor, and in light of the fact that spasticity is the only other known supranuclear motor impairment in ALS, it seems unlikely that this "flutter" is of central origin. Alternatively, at the lower motor neuron level there is evidence that patients who have peripheral neuropathy can have tremorwithin the range of 8 to 12 Hz.[9] One explanation for its presence is that in subclinical peripheral neuropathy in which the stretch reflex is preserved, physiologic tremor is enhanced. In a study by Said et al.[10] of 14 patients with tremor associated with acquired peripheral neuropathies of different origin, all patients showed electrophysiologic evidence of denervation, minimal weakness, a nd tremor frequency in different limbs ranging from 4 to 12 Hz. (In none of these studies was voice tremor described.) One of the 14 patients with tremor in the study by Said et al. had motor neuron disease (ALS) involving both the central and the peripheral nervous systems.

In patients with polyneuropathy, a type of action tremor has been described involving distal muscles in which accelerometric tracings are synchronous with electromyographic activity at a rate of 6 to 8 Hz, and with an improvement in the neuropathy there is a reduction in the amplitude of the tremor.[1]

An alternative to the hypothesis that the laryngeal and extralaryngeal tremor in ALS and peripheral neuropathy is due to an enhancement of physiologic tremor is that the tremor is a sign of loss of motor units resulting in intermittent absence of motor unit firing that, when affecting the intrinsic laryngeal muscles, is perceived as a tremor, or "flutter." In our fiberoptiscopic observations of the pharynx and larynx during "flutter" in ALS patients prolonging the vowel /a/, a rapid tremorous movement of the entire vocal tract could be clearly seen. The true vocal folds oscillated rapidly, as did the supraglottic musculature. Whether these tremorous movements were synchronous with the voice tremor, or "flutter," was not determined, and this issue needs to be explored, electromyographically and acoustically, in future studies of this unusual phenomenon.

References

1. Findley LJ, Capildeo R. Movement disorders: tremor. New York: Oxford University Press, 1984.

2. Lippold OCJ. Oscillation in the stretch reflex arc and the origin of the rhythmical, 8-12 c/s component of physiological tremor. J Physiol (Lond) 1970; 206:359-82.

3. Marsden CD. The mechanisms of physiological tremor and their significance for pathological tremors. prog Clin Neurophysiol 1978;5:1-16.

4. Ramig LA, Scherer RC, Titze IR, Ringel SP. Acoustic analysis of voices of patients with neurologic disease: rationale and preliminary data. Ann Otol Rhinol Laryngol 1988;97:146-72.

5. Aronson AE, Hartman DE. Adductor spastic dysphonia as a sign of essential (voice) tremor. J Speech Hear Disord 1981;46:52-58.

6. Darley FL, Aronson AE, Brown JR. Motor speech disorders. Philadelphia: WB Saunders Company, 1975.

7. Aronson AE. Clinical voice disorders: an interdisciplinary approach. New York: Brian C. Decker, 1980.

8. Winholtz WS, Ramig LO. Vocal tremor analysis with the vocal demodulator J. Speech Hear Res (in press).

9. Shahani BT. Electromyography in CNS disorders: central EMG. Boston: Butterworth Publishers, 1984.

10. Said G, Bathien N, Cesaro P. Peripheral neuropathies and tremor. Neurology 1982; 32:480-5.

NCVS Status and Progress Report - 2
April 1992, 119-137

Vocal Tremor Analysis with the Vocal Demodulator

William S. Winholtz, A.A.S.
Recording and Research Center, Denver Center for the Performing Arts
Lorraine Olson Ramig, Ph.D., CCC-SP
Dept. of Communication Disorders and Speech Science, University of Colorado-Boulder
Recording and Research Center, Denver Center for the Performing Arts

Abstract

Acoustic analysis of vocal tremor has the potential to make significant quantitative and diagnostic contributions to the study of vocal disorders. This paper presents a new device for analysis of vocal tremor. The Vocal Demodulator produces amplitude and frequency demodulated outputs and measures the frequency and level (percent) of low frequency tremor components in sustained phonation. A standard microphone is used to transduce the voice signal for input to the Demodulator. The input fundamental frequency (Fo) range is 70Hz to 1200Hz and frequency response of the amplitude and frequency demodulation is 2.5Hz to 25Hz. Five parameters are displayed in real-time: Fo, amplitude modulation frequency, amplitude modulation level, frequency modulation frequency, and frequency modulation level. Validation, calibration, and reliability data from synthesized test signals with modulation, as well as phonation from subjects with vocal tremor, individuals producing vibrato, and normal voice are presented. Research and clinical applications of this device are suggested.

Some pathological voices are characterized by low frequency tremor, a modulation of the acoustic voice signal. It is useful for the researcher and clinician to know more about these modulations because the frequency, level and signal waveshape may relate to the type, severity and etiology of a disorder. Vocal tremor has been associated with neurological disorders such as essential tremor, Parkinson's disease, and cerebellar ataxia (Aronson, 1985). Vocal tremor may be the first symptom

or the only symptom of a neurological disease (Brown & Simonson, 1963; Findley & Gresty, 1988; Meeuwis & Baarsma, 1985). Furthermore, the frequency, amplitude, and regularity of vocal tremor may differ among diseases of different neural subsystems (Aronson, 1985; Lebrun, Devreux, Rousseau, & Darimont, 1982; Ludlow, Bassich, Connor, & Coulter, 1986). Therefore, analysis of vocal tremor may make important contributions to early and differential diagnosis of neurological diseases and, consequently, to treatment decisions.

The primary non-invasive quantification of vocal tremor has been through acoustic analysis. Most of the acoustic data on vocal tremor have been obtained from visual inspection of oscillographic displays of waveform data (Brown & Simonson, 1963; Hachinski, Thomson, & Buch, 1975) or graphic level recorder displays of amplitude contours of sustained vowel phonation (Hartman, Overholt, & Vishwanat, 1983; Massey & Paulson, 1982). Consequently, the bulk of acoustic data on vocal tremor involves only visually quantifiable amplitude modulations which are difficult to measure. Without means to reduce harmonic and formant interaction, the previous analyses may include the effects of these components on the modulation signal. Furthermore, these types of analyses have considered only the amplitude modulation component and lacked analysis of the equally important frequency modulation component.

Recently, researchers have begun to measure the amplitude and frequency modulation components of the acoustic signal (Ludlow, Bassich, Connor, & Coulter, 1986; Ramig & Shipp, 1987) and apply spectral analysis to the modulating signals (Hartman, Abbs, & Vishwanat, 1988; Philippbar, Robin, & Luschei, 1989). These more recent data suggest that vocal tremor may involve a number of components (Freund & Dietz, 1978; Hunker & Abbs, 1984; Philippbar, Robin, & Luschei, 1989) rather than the single frequency usually reported from earlier oscillographic studies. These more precise analyses may facilitate distinctions among tremors accompanying different neurological diseases and contribute to our understanding about underlying physiologic bases of vocal tremor within the speech mechanism (Brin, Fahn, Blitzer, Ramig, & Stewart, in press).

In one approach to the demodulation of sustained phonation, Philippbar, Robin, and Luschei (1989) used the Visi-Pitch, a hardware device for extracting frequency and intensity of the Fo signal. However, problems with this approach include non-linear conversion of the intensity and frequency output signals as well as inconsistencies from one Visi-Pitch to another (Horii, 1983). Furthermore, the outputs of the Visi-Pitch have a logarithmic function and will produce different results from standard demodulation (e.g., Schrader, 1985) which is a linear process. In another demodulation approach, Ludlow, Bassich, Connor, and Coulter (1986) used a combination of hardware and software. A low pass filter for the input signal was set just below the first formant and optimized for each subject. Modulation levels were calculated by dividing the amplitude (or frequency) average linear trend by shimmer (or jitter), multiplied by 100. Modulation frequency was measured by a "slew rate" algorithm which determined the period between the maximum positive slopes for each cycle. Neither of these approaches to demodulation included verification and calibration data using standard modulation test signals as input.

In order to maximize effectiveness of demodulation for vocal tremor analysis, the following characteristics were considered necessary: a) demodulation should be based on standard theory that is verifiable with existing test equipment; b) measurement of demodulation should be calibrated in standard units; c) there should be automatic adjustment of critical parameters used for conditioning the input signal (e.g., gain, filtering) to improve consistency of operation across highly variable subject input signals; d) data should be processed in real-time to facilitate data collection and provide immediate feedback to enhance clinical utility.

The current study evaluates a hardware demodulation process based on standard modulation theory and further quantifies the demodulation process characteristics in terms of input range, frequency response, signal-to-noise ratio, level accuracy, frequency accuracy, and sensitivity.

Standard Signal Modulation

In the field of electronic communication, modulation is used to broadcast various types of signals. A radio frequency signal, the carrier, is modulated by another signal which contains the information to be transmitted. After transmission, the composite signal is demodulated, which separates the information from the carrier. In standard modulation theory, the modulating frequency is more than a factor of 10 below the carrier frequency. This separation between the modulating and carrier frequencies provides more cycles of carrier signal per cycle of modulating signal, or more definition to the modulating signal. If the separation factor is reduced to 6, an error of approximately -2% is introduced into the demodulated level and increases as the separation factor is reduced further.

Figure 1a is an example of amplitude modulation. The modulation waveform has been superimposed on the Fo signal. The amplitude of the Fo signal is varied by the modulating signal. Figure 1b is an example of frequency modulation with the modulation waveform drawn below the Fo signal. Period of the Fo signal is varied by the modulation. The following formulae are used for calculating levels of standard amplitude and frequency modulation:

$$\text{Amplitude Modulation Level (\%)} = \frac{V_{max} - V_{min}}{V_{max} + V_{min}} \times 100 \qquad (1)$$

$$\text{Frequency Modulation Level (\%)} = \frac{F_{o\ deviation} - F_{o\ average}}{F_{o\ average}} \times 100 \qquad (2)$$

In Figure 1a the modulation waveform has been marked with the maximum and minimum levels. The differences between the Fo peaks and the modulation waveform may be due to the low sampling frequency used to display the low frequency modulation signal riding on the higher frequency Fo signal. Amplitude modulation level for Figure 1a is 30.9%.

$$\frac{930mV - 491mV}{930mV + 491mV} \times 100 = 30.9\%$$

In Figure 1b, arrows indicating the regions for average Fo (125Hz) and maximum deviation (167Hz) are displayed. Frequency modulation level for Figure 1b is 33.6%.

$$\frac{167\text{Hz} - 125\text{Hz}}{125\text{Hz}} \times 100 = 33.6\%$$

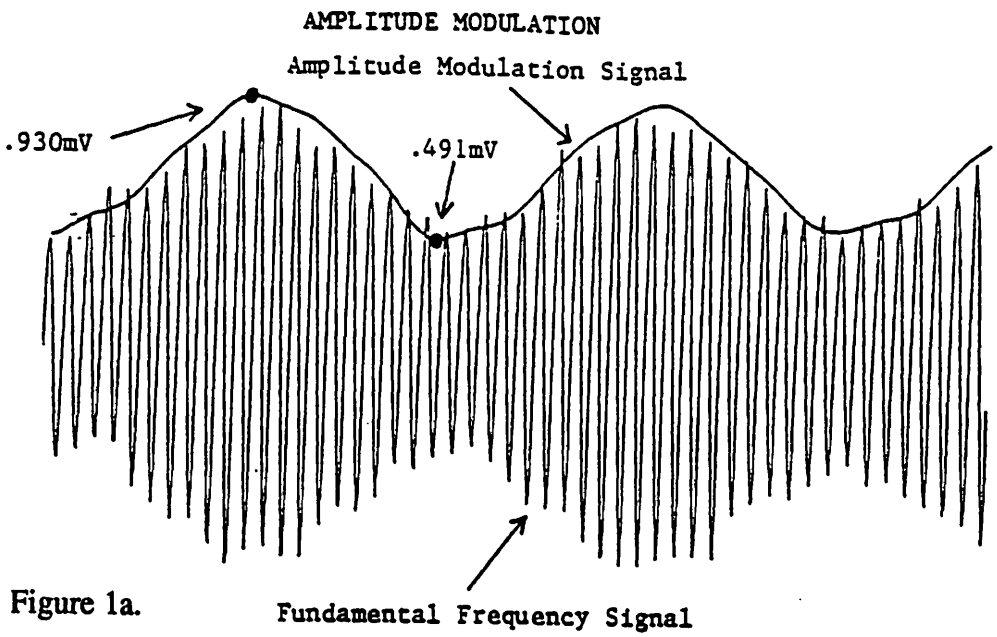

Figure 1a.

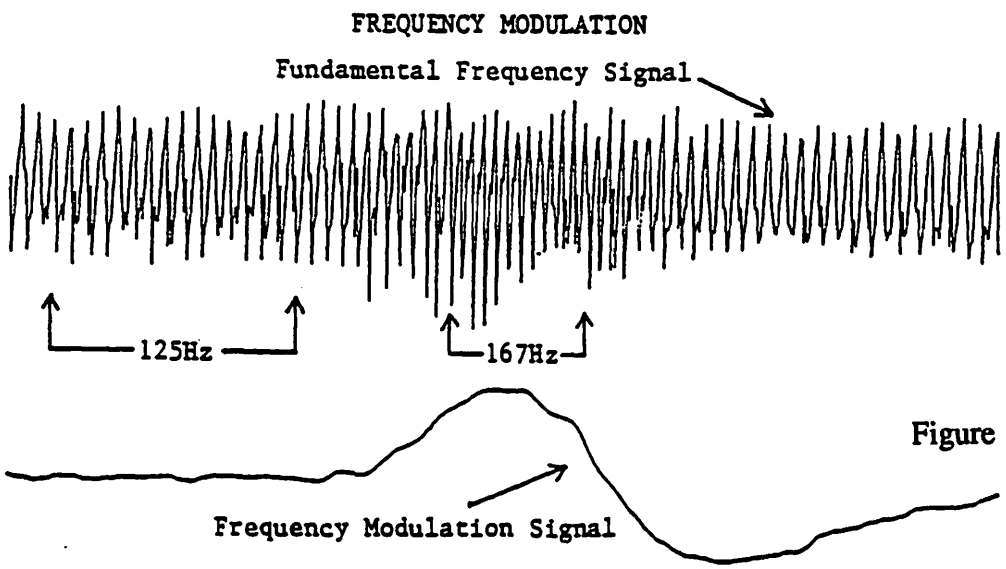

Figure 1b.

Figure 1a. Example of amplitude modulation with the modulating signal superimposed on the fundamental frequency signal. The maximum and minimum modulation voltages are indicated for calculating modulation level. Figure 1b. Example of frequency modulation with the modulating signal drawn below the fundamental frequency signal. Regions of average frequency and peak deviation are indicated for calculating frequency modulation level.

Vocal Signal Modulation

In sustained phonation, the Fo is the carrier which is modulated by different sources in the vocal tract. Probable sources for vocal modulation are the mechanism controlling intensity of the Fo and the mechanism for controlling period of the Fo.

When the frequency of modulation is close to the Fo, it is in the range of cycle-to-cycle variations in amplitude and frequency (shimmer and jitter) and it is not possible to separate and reconstruct the signals causing standard modulation. When the frequency of modulation is much lower, in the tremor range for example, modulation components can be extracted and separated for analysis. When the tremor frequency is more than a factor of 10 below the Fo, formulae for percent of standard modulation can be used to measure level. Application of these formulae to analysis of vocal modulations are a step towards measurement standardization of vocal tremor level.

Description

The Vocal Demodulator (patent pending) was designed as a means of analyzing vocal tremor components of the Fo in sustained phonation. Figure 2 is a functional diagram of the Vocal Demodulator. Input to the Demodulator is typically a microphone signal but it may be driven by other sources that produce standard modulation within the ranges of the Demodulator. Line level signals produced by various recording devices (e.g., DAT, reel-to-reel) may also be used for input. The output from the Demodulator is in two forms. First, separate analog outputs are provided for amplitude and frequency modulation components. Second, the Demodulator analyzes and digitally displays in real-time the values of five parameters: Fo, amplitude modulation frequency, amplitude modulation level, frequency modulation frequency, and frequency modulation level. There are no user adjustments in operation of the Demodulator.

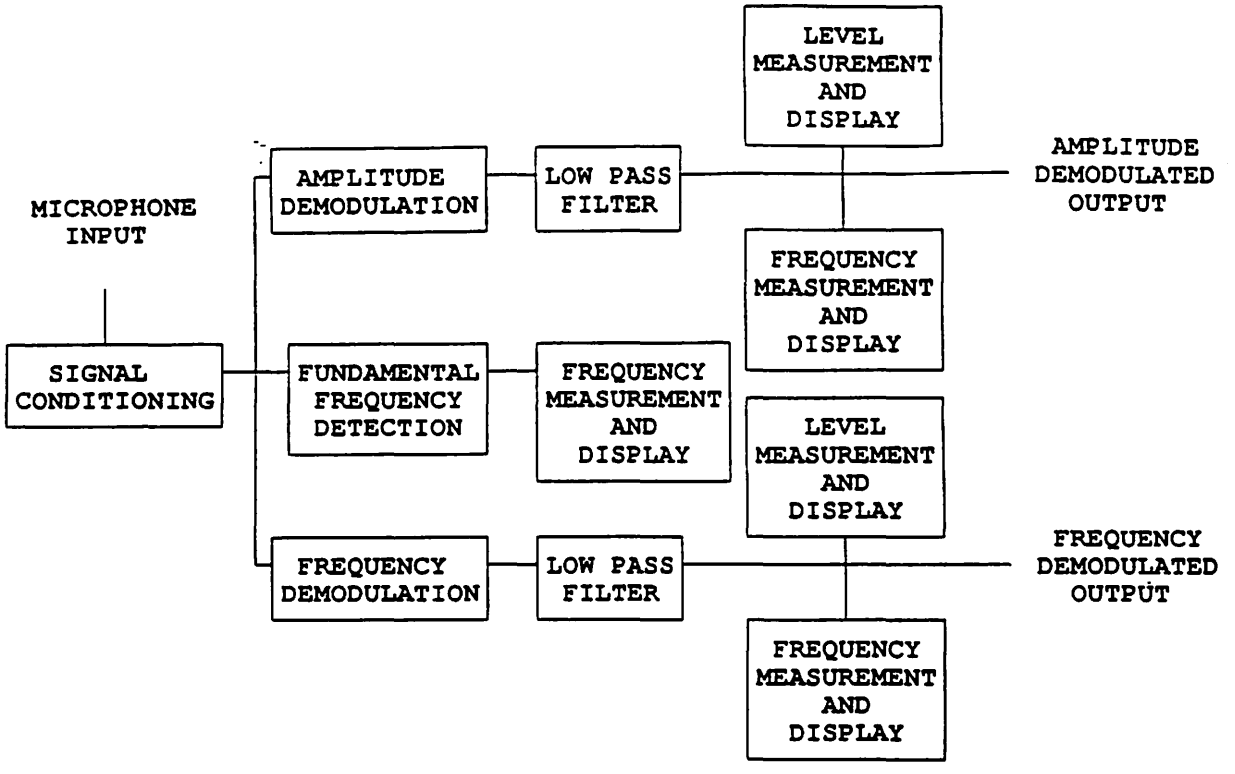

Figure 2. Functional diagram of the Vocal Demodulator

A microphone signal of sustained phonation contains a mix of harmonics and formants produced by the physical properties of the speech and voice mechanism. Since Fo is a reflection of vocal fold oscillation it is of primary interest for demodulation. A low pass filter automatically set above each subject's Fo by the Demodulator reduces the effects of subject dependent upper harmonic and formant energy that would contaminate the demodulation process. Next, the signal is amplitude demodulated and frequency demodulated, then low pass filtered at 25Hz to remove residual Fo components.

Fo is detected with a zero crossing technique and averaged over 1s intervals. The amplitude demodulation detects variations in peak intensity of the Fo signal. A peak voltage detection circuit is used to follow the envelope of the Fo signal and is then filtered to reduce the effects of short-term amplitude perturbation near in frequency to the Fo. Similarly, the frequency demodulation detects peak variations of the Fo period. The frequency demodulation circuit produces a voltage proportional to deviation from the Fo and is also filtered to reduce effects of short-term frequency perturbation near Fo.

In the demodulated sustained phonation of many pathological subjects there may be multiple prominent frequencies that exist at different intensities (Brin et al., in press). For analysis of individual frequencies and levels of signals with multiple frequencies, the demodulated outputs can be analyzed with a spectrum analyzer or Fast Fourier Transform (FFT). When this type of analysis is not available, the Demodulator offers a compromise method generating a single value of frequency. This frequency is a weighted average of individual spectral peaks in the range of 2.5 to 25 Hz, sampled over 1s intervals. A comparison between individual frequencies according to their intensities is used to determine their significance. The greater the intensity of an individual frequency the more significance it has in the average value given for the combined frequencies, formulae 3.

$$F_{\text{weighted average}} = \frac{(f_1 xi_1)+(f_2 xi_2)+\ldots(f_n xi_n)}{i_1+i_2+\ldots i_n} \qquad (3)$$

Level measurement is based on full wave rectification of the demodulated signals, averaged over 0.5s intervals. This allows a short-term average for stability while also providing some measurement of short irregular jumps in level. The level displays for amplitude and frequency demodulation are calibrated in percent of the Fo according to formulae 1 & 2 presented earlier for modulation level. Amplitude modulation level is related to Fo intensity and frequency modulation level is related to Fo period. This method provides uniform measures that are not dependent on a subjects Fo intensity and period.

When the amplitude demodulated output is displayed on an oscilloscope or in the Figures, a positive going trace indicates an increase in intensity of the Fo signal, a negative going trace indicates a decrease in intensity. For the frequency demodulated output, a positive going trace indicates an increase in frequency of the Fo signal, a negative going trace indicates a decrease in frequency.

Methods

Validity of the Demodulator was tested using two procedures. In the first procedure, demodulation was compared to a standard modulation source of known level and frequency accuracy.

In the second procedure, human phonation signals were input to the Demodulator and a correlation of level and frequency measurements were used between a Data 6000 universal waveform analyzer and the Demodulator. Reliability was tested by evaluating the consistency of data produced by the demodulated outputs for repeated inputs of the same section of phonation.

Validity procedure with test signals

Validity of the Vocal Demodulator was evaluated using an HP-8904A Multifunction Synthesizer with a 0.1% modulation level accuracy. The synthesizer was programmed for amplitude or frequency modulation with a carrier sinewave of 200 Hz and a modulation sinewave of 10Hz and then input to the Demodulator. The synthesizer produced different levels of modulation and corresponding levels from the Demodulator were noted. Two ranges of levels were used to provide higher accuracy in ranges of normal use. For amplitude modulation, ranges were 0.5% to 30% in 0.5% steps, and 31% to 50% in 1% steps. Frequency modulation ranges were 0.5% to 10% in 0.1% steps, and 11% to 50% in 1% steps. Accuracy of the frequency measurement circuits was determined by an HP-5316A Universal counter. An output from the Demodulator was input to the counter and simultaneous readings from the Demodulator and Universal counter were noted for different modulation frequencies. Frequency response was determined as the upper and lower frequencies where the demodulation level dropped 3dB from a reference level at 10Hz. Input frequency range, set by design limitations, was the upper and lower Fo where demodulation would be distorted. Signal-to-noise was measured by a ST1500 Audio Analyzer as the difference in the demodulated output levels between maximum input modulation and no input modulation. Cross modulation effects were measured as the residual output level of demodulation caused by a high input modulation level of the opposite type. Squarewave modulation of the Fo was used to measure the delay of the outputs with respect to instantaneous changes of input modulation.

Validity, comparison, and reliability with human phonation signals

Subjects.

Thirty-six individuals were selected to be subjects in this study. Individuals were selected who presented a wide range of vocal modulations from a variety of sources. Samples of vocal tremor were obtained from twelve individuals with neurological diseases including Parkinson's disease, essential tremor, spasmodic dysphonia, and spinal muscular atrophy. The ages of these subjects ranged from 53 to 74 years. Samples of vocal vibrato were obtained from twelve professional singers. Their ages ranged from 25 to 45 years. Twelve individuals without history of neurological or phonatory disorders served as a control group. Their ages ranged from 25 to 48 years. There were six males and females in each of these groups.

Validity procedure.

Each subject produced sustained phonation of the vowel /a/. Approximately eight seconds of the phonation were digitized at 20K samples per second into a 16-bit A/D and D/A converter and VAX computer. A play-back from the converter was input to the Demodulator with the amplitude and frequency demodulated signals input to the waveform analyzer. The waveform analyzer was programmed to sample at 1.8ms, which produced a 0.54 Hz spectral resolution, and one second sampling window similar to the Demodulator. It was not possible to synchronize the one second sampling windows of the waveform analyzer and Vocal Demodulator yielding a random time window error that could have been a maximum of one second. The variability of the signal being

measured and the inability to synchronize the sampling windows made a direct comparison of data from each device impossible. Therefore, a correlation of average data over many samples from approximately the same section of phonation was used to effectively reduce the sampling window error between the Demodulator and waveform analyzer.

A single frequency comparison of six target frequencies, 3Hz, 6Hz, 9Hz, 12Hz, 15Hz, and 18Hz, was used to match similar demodulated frequencies from the subjects. A section of phonation that produced a demodulated frequency near each of the target frequencies was repeated 10 times. Spectral measurements from the waveform analyzer were noted with frequency measurements from the Demodulator and the average value calculated for each set of 10 samples. Then a correlation of the average values for each of the target frequencies was computed. Next, the same type of test was conducted for sections of phonation with multiple prominent frequencies. The spectral peaks from the waveform analyzer were noted as to their relative intensity with respect to the frequency with the highest intensity. Four approximation categories of level were given to the peaks (100%, 75%, 50%, and 25%) and a weighted average frequency was calculated and noted with the frequency value from the Demodulator.

A similar procedure was used for level correlation. Target levels of 5%, 10%, 15%, 20%, 25% for amplitude demodulation and 1%, 2.5%, 5%, 7.5%, 10% for frequency demodulation were used. To compare similar units, the frequency demodulated levels in percent were converted to Hertz of deviation which were correlated with the peak voltage readings from the waveform analyzer.

To verify accuracy of the Demodulator Fo detection, a correlation was computed between the subject's Fo data from a random 1s section of phonation analyzed by the Demodulator and from a FFT by the waveform analyzer.

Comparison of tremor, vibrato, and control subjects.

A 1s section of phonation from each subject was analyzed by the Demodulator for frequency and level data of subjects with vocal tremor, producing vibrato, and with normal voice to determine the medians and distributions for each group.

Reliability.

Reliability of the Demodulator for human phonation was assessed by repeating an input signal and correlating the FFT data from the amplitude and frequency demodulated output signals. A 1s section of phonation was input to the Demodulator and individual frequency and amplitude data from FFTs of the demodulated outputs were noted. The procedure was repeated using the same section of phonation for input and a correlation was performed on the data from the two FFTs. This procedure was carried out on data from five subjects with vocal tremor randomly selected from the subject pool.

Results

Validity procedure with test signals

For the validity procedure using test signals (Table 1) the mean error between the synthesized modulation levels and Demodulator levels for amplitude modulation in the first range (0.5% to 30%) was 0.112%, and 0.560% for the second range (31% to 50%). For frequency modulation level

the mean error was 0.074% in the first range (0.5% to 10%), and 0.308% in the second range (11% to 50%). Frequency response for amplitude and frequency demodulation was 2.5Hz to 25Hz (-3dB). Signal-to-noise ratio for amplitude and frequency demodulation was greater than 57dB. Input Fo range was 70Hz to 1200Hz. Cross modulation of 90% amplitude modulation with no frequency modulation produced frequency demodulation of less than 0.5%. An input signal with 20% frequency modulation and no amplitude modulation produced amplitude demodulation of 2.5%. Squarewave modulation indicated a 20ms delay. Resolution of the modulation level circuits was 0.1% for amplitude and frequency with a minimum sensitivity of 0.5%. The demodulated frequency measurement circuits had an accuracy of +,- 0.15Hz, with a minimum level sensitivity of 0.5%. Fo resolution was 1Hz.

Table 1. Results of validity procedure with standard test signals.

Modulation Test	Vocal Demodulator Parameter		
	Amplitude	Frequency	Fo
Level mean error	0.112%	0.074%	
Frequency response	2.5-25Hz	2.5-25Hz	70-1200Hz
Signal to noise ratio	>57dB	>57dB	
Cross modulation	2.5%(20% FM)	<0.5%(90% AM)	
Output delay	20mS	20mS	
Level resolution	0.1%	0.1%	
Frequency resolution	+,-0.15Hz	+,-0.15Hz	1Hz
Sensitivity	0.5%	0.5%	

Validity, comparison, and reliability with human phonation signals

Figures 3 thru 9 are examples of demodulation for the control, tremor, and vibrato groups. In the Figures the first (top) trace is the amplitude modulation signal. The second trace is a FFT of the amplitude modulation signal. The third trace is the frequency modulation signal. The fourth trace is a FFT of the frequency modulation signal. The FFTs are displayed on a linear scale in order to show more detail of signals typically having a small dynamic range. In the FFT traces, a comparison is shown between the weighted average frequency of the Demodulator, vertical arrow, and spectral display of the waveform analyzer. For example, in Figure 4 trace 2, the weighted average frequency is 8.3Hz.

$$\frac{(5.4\text{Hz} \times 520\text{mV}) + (9.7\text{Hz} \times 250\text{mV}) + (14.1\text{Hz} \times 200\text{mV})}{520\text{mV} + 250\text{mV} + 200\text{mV}} = 8.3\text{Hz}$$

In Figure 4 trace 4, the weighted average frequency is 5.4Hz.

$$\frac{5.4\text{Hz} \times 400\text{mV}}{400\text{mV}} = 5.4\text{Hz}$$

Discrepancies between the vertical arrow and spectral display of the waveform analyzer may be due to sampling window error.

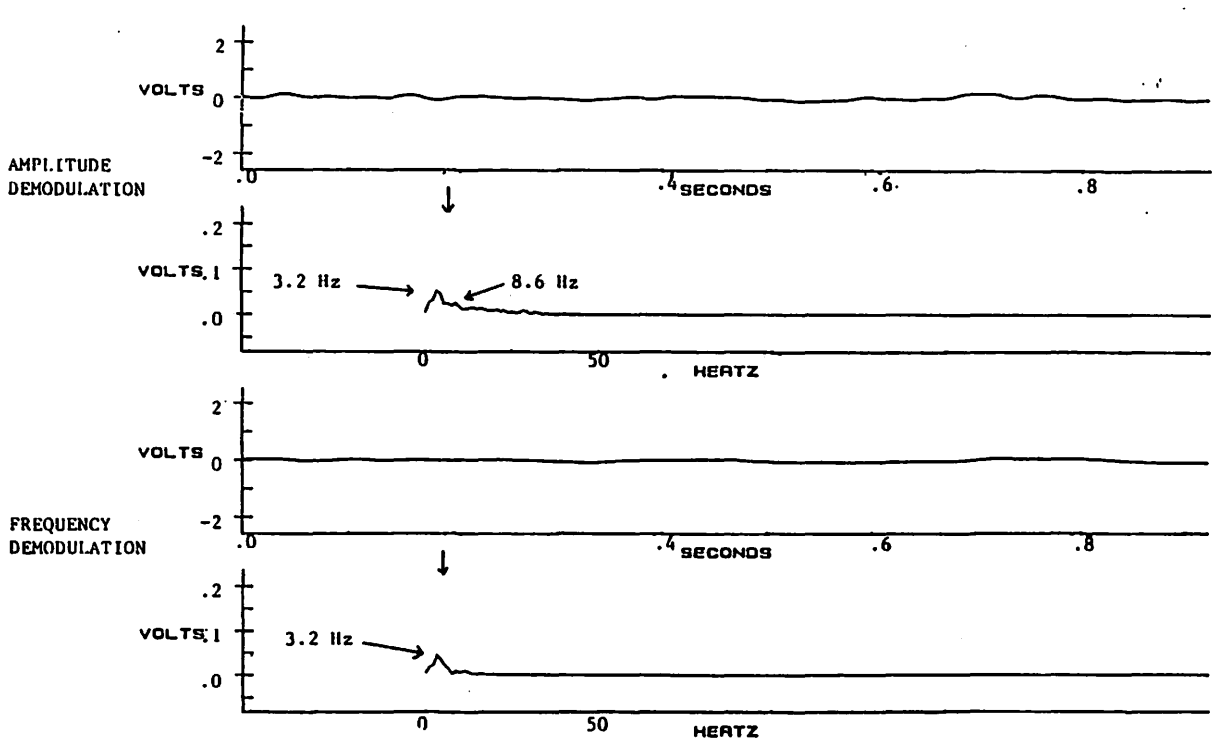

Figure 3. Demodulation of a subject with normal voice. In figures 4 thru 10 the first (top) trace is the amplitude modulation signal, the second trace is a FFT of the amplitude modulation signal, the third trace is the frequency modulation signal, and the fourth trace is a FFT of the frequency modulation signal. The vertical arrow in the FFT traces represents the weighted average frequency from the Demodulator. Vocal Demodulator display: Fo = 191Hz; Amplitude Modulation: freq (Hz) = 6.0; level (%) = 2.4: Frequency Modulation: freq (Hz) = 4.6; level (%) = 0.8

Figure 3 is an example of demodulation for a normal subject. Figures 4 thru 7 are examples of various types, levels, and frequencies of demodulation for subjects with vocal tremor. Figures 4 and 5 are examples of demodulation for subjects producing vibrato. In Figure 6, a subject's phonation consists primarily of amplitude modulation with low level frequency modulation at about the same frequency. In Figure 7, a subject's phonation is primarily frequency modulation with low levels of amplitude modulation at multiple frequencies. Figure 8 illustrates a subject's phonation with high levels of modulation at the same frequency. Figure 9 illustrates a subject's phonation with high levels of modulation at apparently unrelated frequencies.

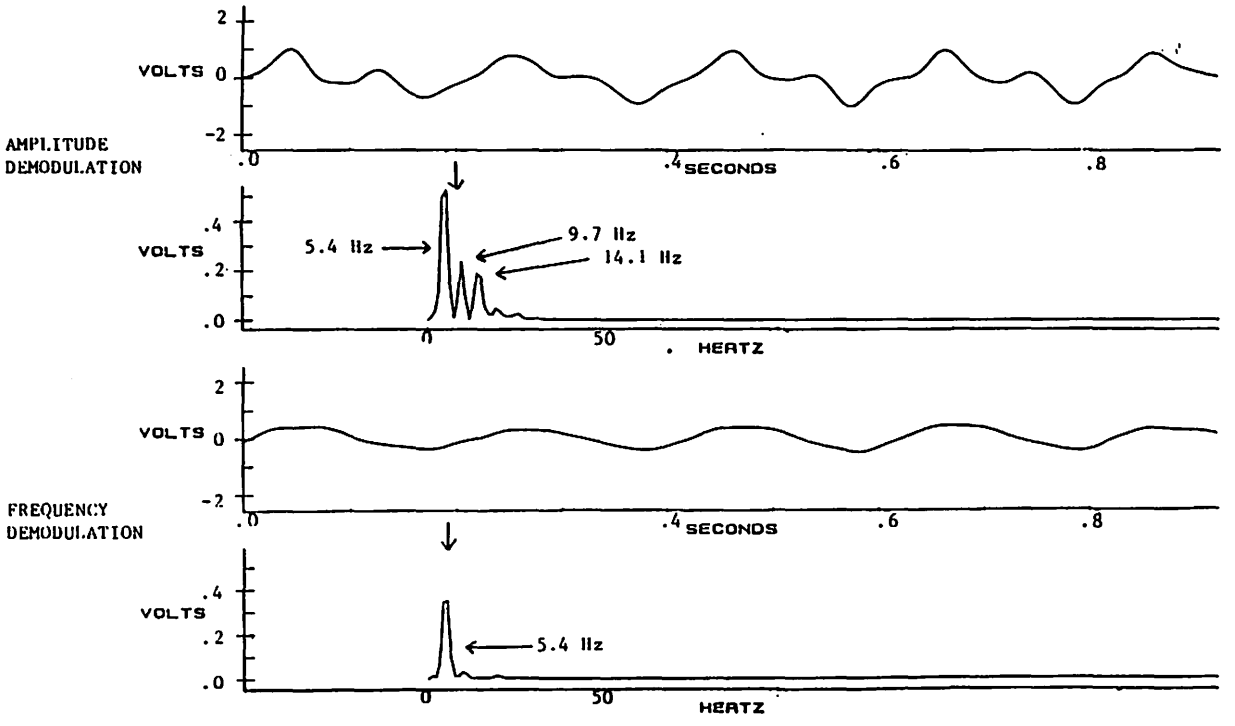

Figure 4. Demodulation of a female subject producing vibrato. Vocal Demodulator display: Fo = 374Hz; Amplitude Modulation: freq (Hz) = 7.5; level (%) = 28.3: Frequency Modulation: freq (Hz) = 5.1; level (%) = 3.4

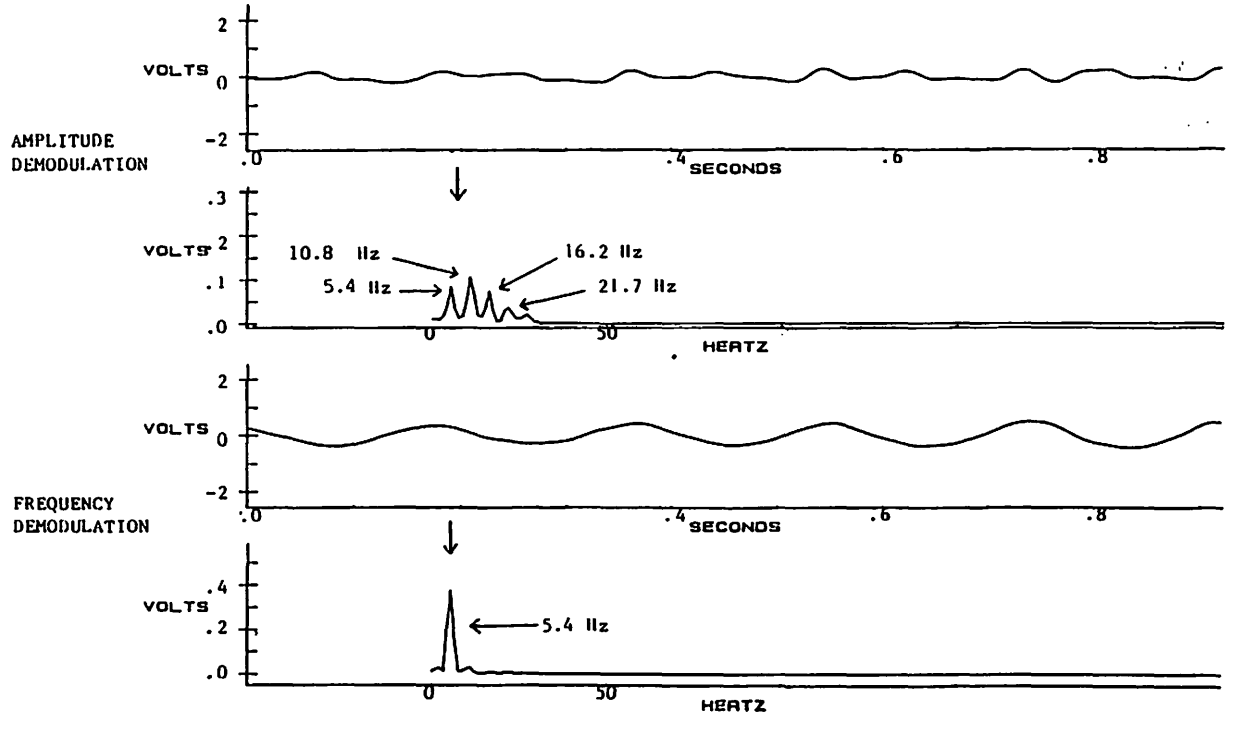

Figure 5. Demodulation of a male subject producing vibrato. Vocal Demodulator display: Fo = 249Hz; Amplitude Modulation: freq (Hz) = 6.0; level (%) = 6.7: Frequency Modulation: freq (Hz) = 4.5; level (%) = 4.2

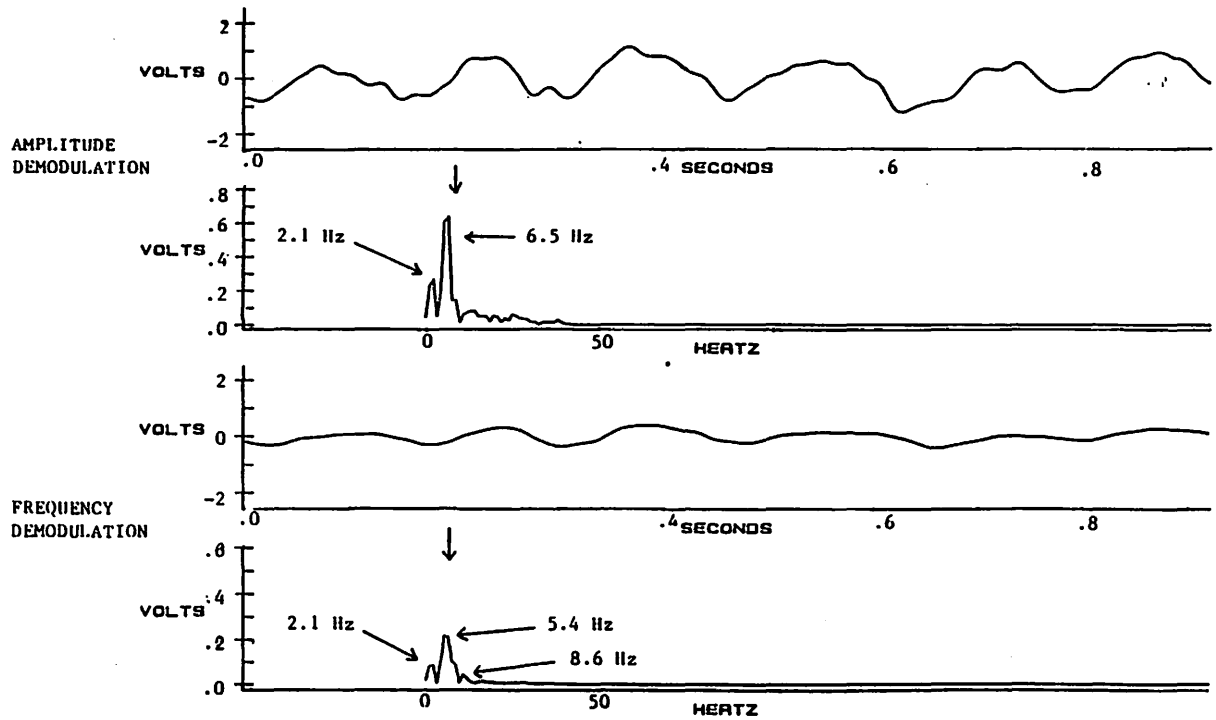

Figure 6. Demodulation of a subject with Parkinson's disease producing primarily amplitude vocal tremor. Vocal Demodulator display: Fo = 176Hz; Amplitude Modulation: freq (Hz) = 7.0; level (%) = 19.4: Frequency Modulation: freq (Hz) = 6.0; level (%) = 3.1

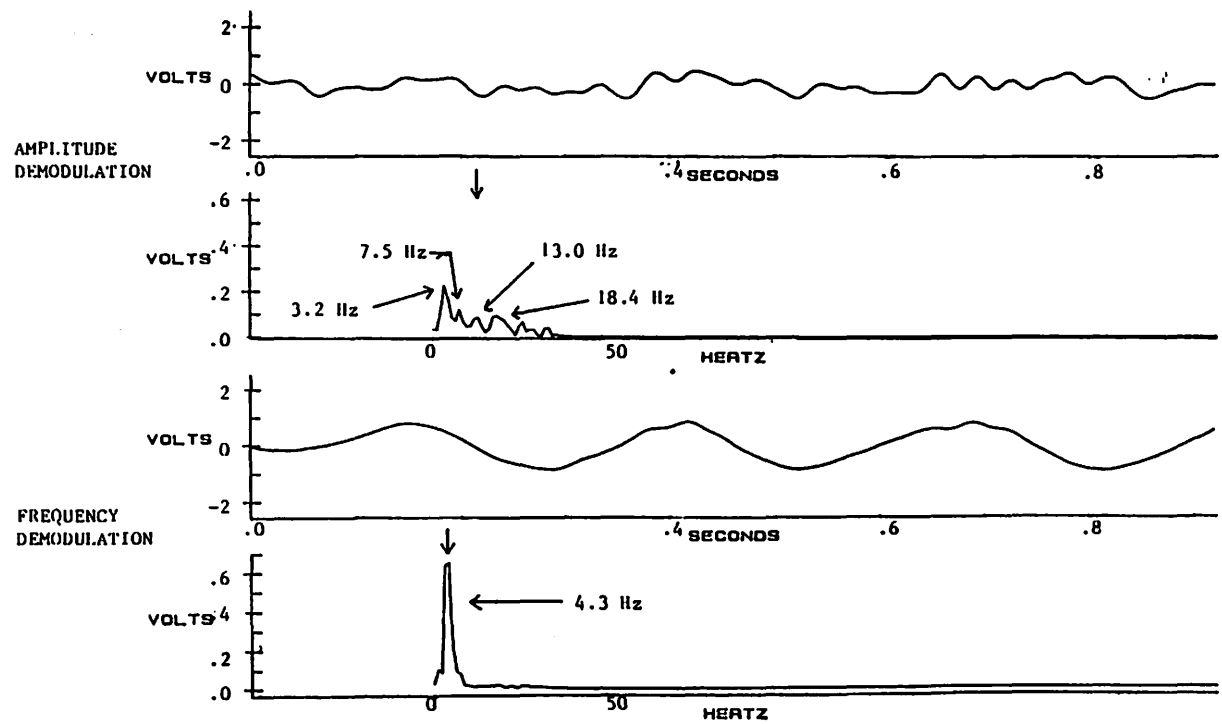

Figure 7. Demodulation of a subject with Parkinson's disease producing primarily frequency vocal tremor. Vocal Demodulator display: Fo = 167Hz; Amplitude Modulation: freq (Hz) = 12.6; level (%) = 8.8: Frequency Modulation: freq (Hz) = 6.0; level (%) = 6.0

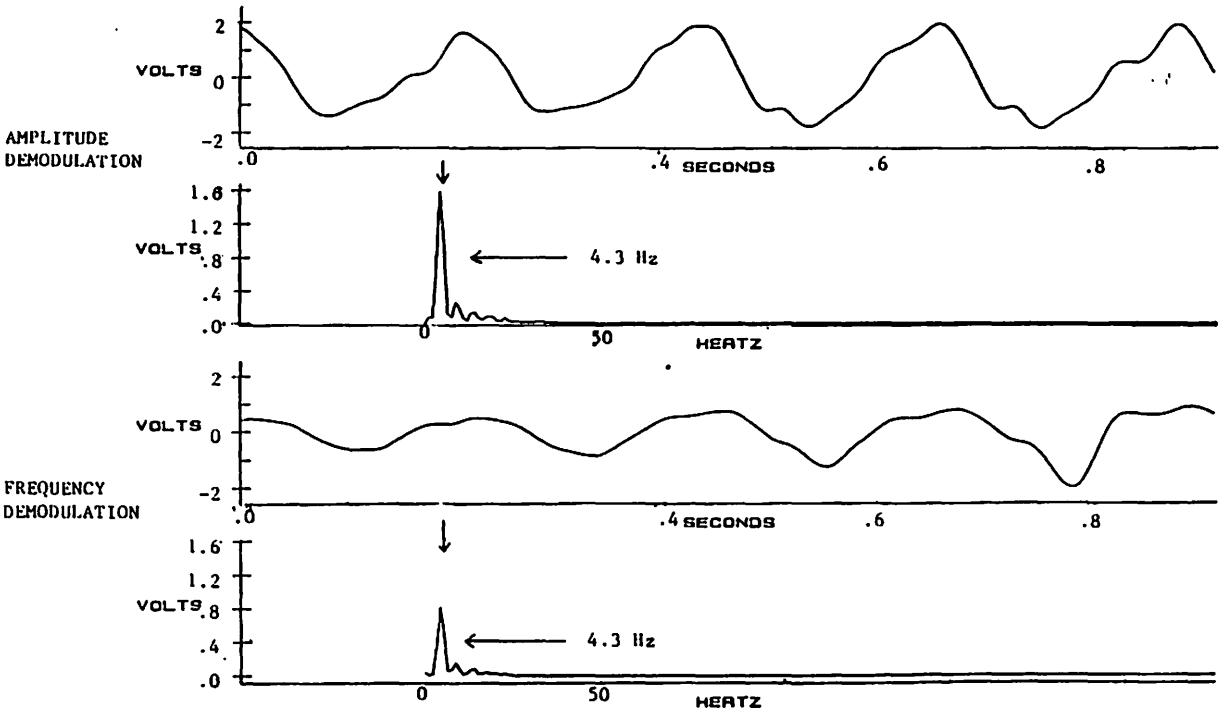

Figure 8. Demodulation of a subject with essential tremor producing the same frequency for amplitude and frequency vocal tremor. Vocal Demodulator display: Fo = 232Hz; Amplitude Modulation: freq (Hz) = 4.9; level (%) = 38.9: Frequency Modulation: freq (Hz) = 4.5; level (%) = 9.3

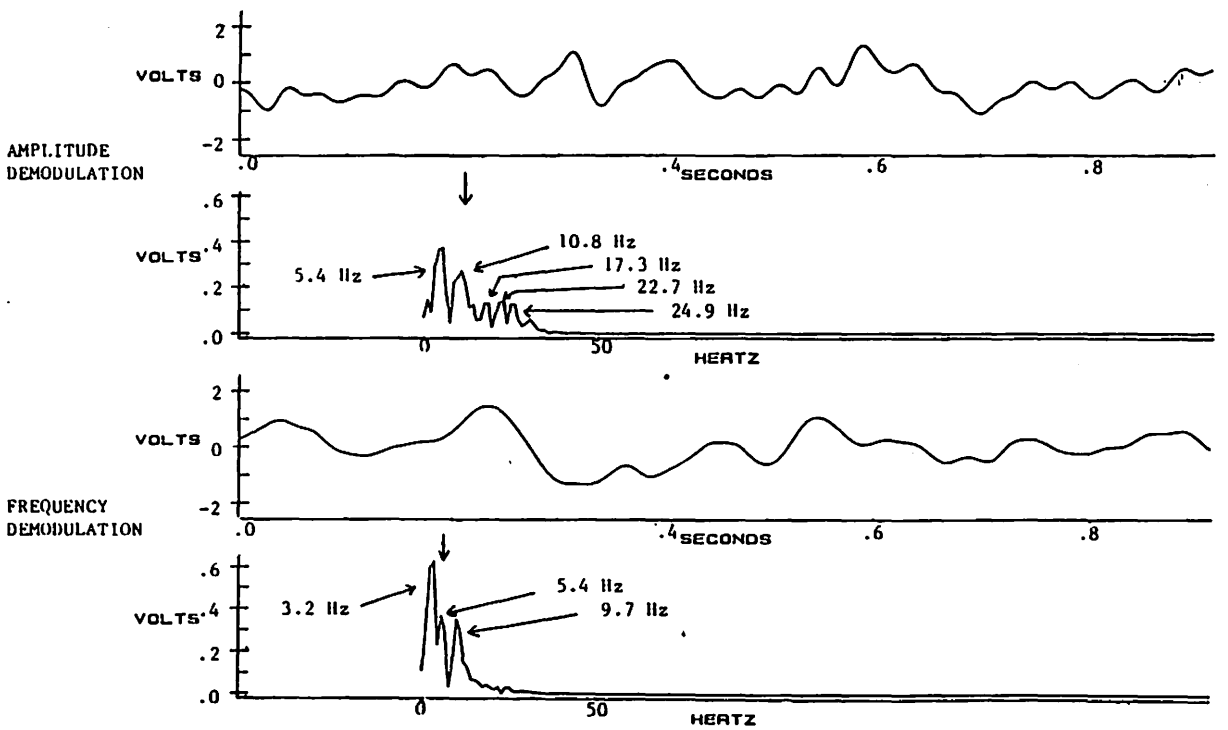

Figure 9. Demodulation of a subject with Parkinson's disease producing multiple frequencies for amplitude and frequency vocal tremor. Vocal Demodulator display: Fo = 224Hz; Amplitude Modulation: freq (Hz) = 9.2; level (%) = 20.4: Frequency Modulation: freq (Hz) = 5.5; level (%) = 10.7

Validity

Table 2 summarizes the correlation of measurements from the waveform analyzer and Demodulator for the validity procedure with human phonation signals. The correlation coefficient (Pearson product moment) of mean values for a single demodulated frequency was .998 for 120 samples. The correlation coefficient of mean values for multiple demodulated frequencies was .996 for 120 samples. The amplitude tremor level correlation coefficient was .995 for 50 samples. A function comparison of voltage and Hertz of deviation along with level in percent is given for frequency tremor level. The frequency level correlation coefficient was .997 for 50 samples.

In the Fo correlation, 24 of the subjects had an Fo within the 250Hz spectral window of the waveform analyzer. The sample period of the waveform analyzer was then changed to increase the spectral window and include subjects with an Fo above 250Hz. The correlation coefficient for Fo detection was .999 for 36 samples. The Demodulator could not detect Fo for one male subject using vibrato. Further analysis of the subjects acoustic signal did not reveal any obvious explanation. Another male subject was substituted for analysis.

Table 2. Correlation data of level and frequency measures from the Waveform analyzer and Demodulator for human phonation signals.

	MEAN	SD	MIN	MAX	N	r
			Single Tremor Frequency			
Wavform analyzer	10.60	4.26	4.8	17.5	120	.998
Demodulator	10.35	4.33	4.3	17.9	120	
			Multiple Tremor Frequencies			
Waveform analyzer	9.40	4.34	4.1	16.8	120	.996
Demodulator	9.31	4.50	3.7	17.1	120	
			Amplitude Tremor Level			
Waveform analyzer (mV)	569.20	267.56	251	921	50	.995
Demodulator (%)	13.86	6.66	5.6	22.5	50	
			Frequency Tremor Level			
Waveform analyzer (mV)	847.20	533.39	154	1520	50	.997
Demodulator (Hz)	11.12	6.62	2.0	19.1	50	
Demodulator (%)	5.06	3.19	1.4	9.3	50	
			Fundamental Frequency (Hz)			
Waveform analyzer	238.75	125.13	103	574	36	.999
Demodulator	239.17	125.58	102	573	36	

Table 3. Group comparison of Vocal Demodulator data for amplitude and frequency modulation components for twelve tremor, twelve vibrato, and twelve control subjects.

	MEDIAN	MIN	LOWER HINGE	UPPER HINGE	MAX
			Amplitude Modulation Frequency (Hz, weighted average)		
Male					
Tremor	9.5	6.0	6.5	10.0	12.6
Vibrato	6.3	5.0	5.0	7.0	10.5
Control	10.2	6.3	7.0	13.0	13.0
Female					
Tremor	7.0	4.5	5.4	8.3	8.5
Vibrato	5.8	5.0	5.5	7.5	11.8
Control	8.1	6.5	7.0	9.5	13.2
			Amplitude Modulation level (%)		
Male					
Tremor	12.4	8.2	8.8	20.2	22.9
Vibrato	17.9	6.7	12.0	19.6	34.1
Control	5.8	2.1	2.1	7.4	12.6
Female					
Tremor	23.7	13.7	16.4	37.4	45.0
Vibrato	16.3	12.0	12.6	28.3	38.1
Control	4.8	2.1	3.3	6.9	13.3
			Frequency Modulation Frequency (Hz, weighted average)		
Male					
Tremor	5.5	4.0	4.0	6.5	6.6
Vibrato	5.0	4.5	5.0	5.3	5.7
Control	6.6	4.8	5.5	8.0	10.6
Female					
Tremor	5.8	5.0	5.0	6.5	6.9
Vibrato	5.1	5.0	5.0	5.4	6.0
Control	5.0	4.7	4.9	6.1	6.6
			Frequency Modulation level (%)		
Male					
Tremor	4.8	3.3	3.4	6.0	6.1
Vibrato	3.4	1.8	2.5	4.2	4.2
Control	1.5	0.5	0.9	1.8	2.8
Female					
Tremor	7.4	2.5	4.3	10.7	11.8
Vibrato	3.3	2.3	3.0	3.4	4.5
Control	1.1	0.5	0.8	1.3	1.4
			Fundamental Frequency (Hz)		
Male					
Tremor	135	106	131	151	167
Vibrato	231	145	201	267	390
Control	119	106	117	147	147
Female					
Tremor	212	175	177	233	349
Vibrato	452	374	404	488	573
Control	252	197	242	270	367

Note. See text for discussion of tested significant differences.

Comparison of tremor, vibrato, and control groups.

Table 3 summarizes the data from the tremor, vibrato, and control subjects. In order to assess differences in modulation characteristics among groups, Kruskal Wallis nonparametric analysis of variance tests were carried out on the variables Fo, amplitude modulation frequency, amplitude modulation level, frequency modulation frequency and frequency modulation level for the factors group and sex. Nonparametric analyses were chosen because of non-normal distributions of our means.

For all group comparisons, the factor of sex was statistically significant only for the variable Fo. Fo for females (median=265Hz) was statistically significantly higher than for males (median=145Hz) across all groups ($p<.05$).

Between group differences were statistically significant for the variable amplitude modulation frequency. Vocal tremor subjects and control subjects had significantly higher modulation frequencies (median=8.15Hz,range=4.5Hz-12.6Hz; and median=9.25Hz,range=6.3Hz-13.2Hz, respectively) than did the vocal vibrato subjects (median=6.0,range=5.0Hz-11.8Hz)($p<.05$). There were no significant group differences on the variable frequency modulation frequency (tremor median=5.75Hz,range=4.0Hz-6.9Hz; vibrato median=5.0Hz,range=4.5Hz-6.0Hz; control median=5.5,range=4.7Hz-10.6Hz)($p>.05$).

The levels of amplitude and frequency modulations were statistically significantly different among the groups. For level of amplitude modulation, vocal tremor subjects and vibrato subjects had statistically significantly higher levels (median=18.0%,range=8.2%-45.0%; and median=16.9%,range=6.7%-38.1%, respectively) than did the control subjects (median=5.0%, range=2.1%-13.3%)($p<.05$). For level of frequency modulation, vocal tremor subjects and vibrato subjects had statistically significantly higher levels (median=5.45%,range=2.5%-11.8%; and median=3.25%,range=1.8%-4.5%, respectively) than did the control subjects (median=1.25%, range=0.5%-2.8%) ($p<.05$).

Reliability

The correlation coefficient for repeated testing (reliability) ranged from .993 to 1.0 for 140 amplitude and frequency FFT samples from the five subjects (amplitude .997, .993, .996, .999, .995; frequency 1.0, 1.0, 1.0, 1.0, 1.0).

Discussion

A new device for analysis of vocal modulation components of the Fo has been described. The validation indicates that the device accurately quantifies modulation components based on comparisons with a high performance multifunction synthesizer and waveform analyzer, as well as in application to human phonation from subjects with vocal tremor, vibrato, and normal voice.

When comparing the groups of subjects, the frequencies of the amplitude and frequency modulations were higher in the tremor and control groups, than in singers producing vocal vibrato. Since the frequency of vocal vibrato has consistantly been reported in the range of 5Hz to 7Hz, this is not surprising. The control subjects' data fell within the range of normal physiologic tremor. Because of the wide range of etiologies for the vocal tremor patients, any more specific interpretation of these tremor data would not be meaningful.

The median levels of amplitude modulations as well as their ranges were much greater in the tremor and vibrato subjects when compared to the control group. Variation among singers in level of amplitude modulation was similar to that of patients with vocal tremor.

In median levels of frequency modulation, the tremor and vibrato subjects were higher than a control group, with the range for the tremor subjects being the greatest. The tremor and vibrato subjects had smaller ranges in their levels of frequency modulation than in their levels of amplitude modulation.

For these data, greater between group differences existed for levels of modulation than for frequencies of modulation. Since earlier analyses of vocal modulation consisted primarily of the average amplitude modulation frequencies, the levels of modulation as reported here may add new significance to vocal modulation analysis.

Presentation of human subject data in this paper was designed to demonstrate the applicability of the Vocal Demodulator. Future research will study larger groups of subjects to investigate hypotheses underlying physiologic sources for vocal modulations.

Application

The Vocal Demodulator has many potential applications. In clinical use, the Demodulator as a stand alone unit would give a real-time indication of patients' vocal stability in terms of frequencies and levels of modulation. This may be particularly useful in allowing a more quantitative, on-line diagnosis of vocal tremor. For patients with voice disorders which include phonatory instabilities, the Vocal Demodulator may provide useful on-line biofeedback as various techniques are implemented to improve voice production.

For research purposes, the analog outputs from the Demodulator can be used as input to a two channel digital oscilloscope, spectrum analyzer, or waveform analyzer providing simultaneous analysis of modulation components. This type of analysis may be useful in relating events in the waveshape to primary and secondary mechanisms generating vocal tract modulations. Another research application may be the study of frequency distributions of vocal modulation components similar to work by Aronson, Ramig, Winholtz, and Silber (in submission) carried out to quantify higher frequency vocal modulation components ('flutter') in patients with amyotrophic lateral sclerosis.

Application of the Demodulator to vibrato, may provide additional insight into the relationship between modulation characteristics and the perceptual quality of vibrato. Furthermore, the Demodulator could serve as an instructional aid by providing feedback of vibrato characteristics to singing students, allowing real-time adjustment of their technique.

Some acoustic measures, such as shimmer and jitter, used for non-invasive voice analysis can be unreliable because of interactions between types of modulation and the harmonic and formant structure of the vocal tract (Hillenbrand, 1987). The signal processing of the Demodulator appears to reduce interaction between types of low frequency modulation as demonstrated by the various examples of demodulation presented earlier. The significant instability of the dysphonic voice often precludes many types of acoustic analysis, leaving the clinical researcher at a loss as to quantification of vocal instability in certain populations. In this study, 97% of the subjects were analyzable with the Vocal Demodulator. Because the Vocal Demodulator analyzes low frequency (long-term) phonatory instability it may offer a feasible way to quantify dysphonic voice.

Previous studies have shown a relationship between amplitude of the Fo and subglottal pressure (Hachinski, Thomsen, & Buch, 1975), and a relationship between period of the Fo and neuromuscular control of the cricothyroid muscle (Larson & Kempster, 1985). It is possible that different combinations of amplitude tremor level, amplitude tremor frequency, frequency tremor level, and frequency tremor frequency can be related to different tremor sources in the vocal tract.

The tremor level could be related to the extent or progression of neurologic disorder and the frequency of tremor could suggest involvement of a particular neural subsystem.

The Vocal Demodulator could assist by tracking changes in pre-and post- behavioral, neuropharmacological or surgical treatment by quantifying levels and frequencies of tremor components. If research can link the demodulation characteristics to sources in the vocal tract, the Demodulator could function as a diagnostic tool for vocal disorders.

The Demodulator provides calibrated real-time measurement of five parameters in standard units at 0.5% base accuracy, and separate analog outputs for external analysis. It appears that the Vocal Demodulator can function as a robust tool for measuring and analyzing vocal tremor and may therefore make significant contributions to quantification and diagnosis of vocal tremor disorders.

Acknowledgements

The authors thank Ingo Titze and Ron Scherer for their helpful review of an earlier manuscript of this paper. Manuscript preparation was supported in part by NIH grant RO1 DC01150-01. Application of the Demodulator was supported in part by NIH grant RO1 DC00387-04. The data presented here were supported in part by OE-NIDRR grant H133G00079.

References

Aronson, A. (1985). Clinical Voice Disorders (2 ed.). New York: Thieme.

Aronson, A., Ramig, L.O., Winholtz, W.S., & Silber, S.R., (in press). Acoustic characteristics of vocal flutter in patients with amyotrophic lateral sclerosis. Annals of Otology, Rhinology and Laryngology.

Brin, M., Fahn, S., Blitzer, A., Ramig, L.O., & Stewart, C., (in press). Movement Disorders. In A. Blitzer, C. Sasaki, S. Fahn, M. Brin, & K. Harris (Eds.), Neurolaryngology. New York: Thieme

Brown, J.R., & Simonson, J. (1963). Organic voice tremor: a tremor of phonation. Neurology,13,520-525.

Findley, L.J., & Gresty, M.A. (1988). Head, facial and voice tremor. In J. Jankovic & E. Tolosa (Eds.), Facial Dyskinesias: Advances in Neurology (pp. 239-253). New York: Raven Press.

Freund, H.J., & Dietz, V. (1978). The relationship between physiological and pathological tremor. In J. Desmedt (Ed.), Physiological Tremor, Pathological Tremor and Clonus (pp. 66-89). Basel, Switzerland: s Karger AG.

Hachinski, V.C., Thomsen, I.V., & Buch, N.H. (1975). The nature of primary vocal tremor. Journal of Neurological Science,2,195-197.

Hartman, D.E., Abbs, J., & Vishwanat, B (1988). Clinical investigations of adductor spastic dysphonia. Annals of Otorhinolaryngology,97,(3),247-252.

Hartman, D.E., Overholt, S., & Vishwanat, B. (1983). A case of vocal cord nodules masking essential (voice) tremor. Archives of Otolaryngology,108,52-53.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. Journal of Speech and Hearing Research,30,448-461.

Horii, Y. (1983). Automatic analysis of voice fundamental frequency and intensity using the visipitch. Journal of Speech and Hearing Research,26,467-471.

Hunker, C., & Abbs, J. (1984). Physiological analysis of Parkinsonian tremors in the oral facial system. The Dysarthrias, 69-100.

Larson C., & Kempster, G. (1985). Voice fundamental frequency changes following discharge of laryngeal motor units. In I.R. Titze & R.C. Scherer (Eds.), Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control (pp. 91-104). Denver: The Denver Center for the Performing Arts.

Lebrun, Y., Devreux, F., Rousseau, J.J., & Darimont, P. (1982). Tremulous speech. Folia Phoniatrica,34,134-142.

Ludlow, C.L., Bassich, C.J., Connor, N.P., & Coulter, D.C. (1986). Phonatory characteristics of vocal fold tremor. Journal of Phonetics,14,509-515.

Massey, E.W., & Paulson, G. (1982). Essential vocal tremor: response to therapy. Neurology,32,A113.

Meeuwis, C.A., & Baarsma, E.A. (1985). Essential (voice) tremor. Clinical Otolaryngology,10,54.

Philippbar, S.A., Robin, D.A., & Luschei, E.S. (1989). Limb, jaw, and vocal tremor in Parkinson's patients. In K.M. Yorkston and D.R. Beukelman. (Eds.) Recent advances in clinical dysarthria (pp. 65-197). Boston: College Hill Publication.

Ramig, L.A., & Shipp, T. (1987). Comparative measures of vocal tremor and vocal vibrato. Journal of Voice,2,162-167.

Shrader, R.L. (1985). Electronic Communication. New York: McGraw- Hill.

NCVS Status and Progress Report - 2
April 1992, 139-155

Contrasts in Voice Characteristics with Adductor and Abductor Spasmodic Dysphonia

Kiyoshi Makiyama, MD.
Otolaryngology-Head and Neck Surgery, University of Wisconsin-Madison
Waisman Center, University of Wisconsin-Madison
Diane M. Bless, Ph.D.
Otolaryngology-Head and Neck Surgery, University of Wisconsin-Madison
Waisman Center, University of Wisconsin-Madison
Department of Communicative Disorders, University of Wisconsin-Madison
Charles N. Ford, M.D.
Otolaryngology-Head and Neck Surgery, University of Wisconsin-Madison

Abstract

This retrospective study reports subjective and objective measurements of vocal function on twenty-two subjects with spasmodic dysphonia (SD). Five types of data were collected: 1)perceptual voice judgments, 2)indirect laryngoscopic examinations, 3)videostroboscopic examinations, 4)vocal function examinations, and 5)acoustic recording. Data were compared for fourteen adductor spasmodic dysphonia (AD-SD) subjects and eight abductor spasmodic dysphonia (AB-SD) subjects, and sixty-nine subjects without voice complaints. Results demonstrate differences between the AD¬SD and AB-SD, and SD groups and normal group: perceptual characteristics differed between the AD-SD and AB-SD subjects; mean airflow rate was higher in the AB-SD group; and voice source was poorer for the AB-SD male subjects and the AD-SD female subjects. This suggests that a combination of subjective and objective observations expand description of this disorder and should facilitate documentation of change resulting from treatment.

Introduction

The clinical entity known today as "spasmodic dysphonia" has a long and complex history. It was first described by Traube in 1871 as "spastic form of psychogenic hoarseness."[1] Traube's article set the framework for treatment and description of the problem for the next one hundred years. During this period, numerous authors described spasmodic dysphonia (SD) and its treatment. In general authors seemed to concur that a set of perceptual vocal symptoms helped identify a psychogenic form of hoarseness that was resistant to treatment: the perceptual symptoms included a squeezed, strained, choked, staccato, stutteringlike, jerky, grunting, groaning, effortful, pinched and grating voice that resulted from hyperadduction of the true and false folds[2]; the treatment included psychotherapy, voice-speech therapy, acupuncture, hypnosis, biofeedback, respiratory therapy, electroshock therapy, meditation, tranquilizers, muscle relaxants, megavitamins, and chiropractic[3]. Clinicians were both puzzled and troubled by the failure of individuals with SD to respond to treatment: they raised questions about the treatment methods; they raised questions about the etiology; and they raised questions about the diagnosis.

In 1973 Aronson suggested that SD was not one but three or four disorders: Adductor spasmodic dysphonia (AD-SD), Abductor spasmodic dysphonia (AB-SD), Mixed type and Essential voice tremor[4]. He characterized the AB-SD voice as one with sudden interruptions of brief moments of breathy or whispered segments. Since the mechanism of hoarseness was different it was also reasonable to assume that the response to therapy would also differ. Searches for treatment were largely unsuccessful until Dedo severed the recurrent laryngeal nerve of one of his patients[5]. The result was remarkably fluent speech; absent were the staccato, stutterlike and jerky speech characteristic of the SD patient. The results had significant impact on the course of treatment of SD: it stimulated investigations to look for possible neurogenic bases for the disorders, and it stimulated considerable discussion about treatment. The focal points of treatment discussions centered around who to treat and the longterm results of treatment. The treatment continues to be controversial because different centers report different results. These contradictory results may be related to differences in technique, subject selections, or criteria for success. Thus far both subject selection and criteria for success have been based on subjective judgements, making it difficult to make cross institutional comparisons. An obvious solution to this problem is the application of vocal assessment tools to describe the voice in a manner that can be easily replicated. To date objective assessment of the voice of SD speakers has been limited to acoustic and aerodynamic measure[6-16]. Acoustic analysis has demonstrated clear changes between the AD-SD and AB-SD groups. Wolfe and Bacon[6] showed spectrographic evidence of a breakdown in format structure, or the addition of fricative fill superimposed upon resonance bars for the AB-SD subjects, compared to widely and irregularly spaced vertical striations for the AD-SD subjects. Zwitman[7] found aspiration time between initial consonant and subsequent vowel seemed to be an additional distinctive feature between the two groups. Aronson[14] used sonograms to describe the characteristics of his AD-SD patients. He observed irregular breaks in the fundamental frequency, and noise in the harmonic structure on prolonged /a/ vowels.

Hartman and Aronson[9] reported acoustic studies for 17 subjects with AB-SD. They revealed four patterns of amplitude variation in oscillographic tracings during sustained phonation: the vowel /a/ was prolonged steadily with random changes in amplitude; the vowel was sustained with intensity changes that seemed to approach some kind of rhythmicity; the vowel was prolonged with mean rhythmic changes; and there were moments of intermittent breathy dysphonia and adductor voice

arrests during vowel prolongation as well as rhythmic variations in intensity. They also reported on 22 subjects with AD-SD on whom they observed both voice arrests and tremor cycles on the oscillographic tracings, in contrast to the 30 essential voice tremor subjects who only exhibited the tremor cycles[10]. An additional oscillographic analysis of AD-SD patients was reported by Hartman, Abbs and Vishwanat[15].These investigators noted irregular changes in intensity with variable amplitude for their AD-SD subjects. All of these reported study results appear to imply that acoustic analyses helps describe SD vocal characteristics but still has limitations: the observations made included both qualitative and quantitative data and were restricted to a few measurements; and the acoustic stadies did not provide any simultaneous data or airflow or image of the larynx.

Merson and Ginsberg[8] measured fundamental frequency and airflow rate on two AB-SD patients and on one normal subject. The mean fundamental frequency for the sustained vowel /a/ was higher in the SD subjects than in the normal subject. The mean airflow rate on vowel and sentence utterance was higher in the SD subjects than in the normal subject, and excessive variation in airflow throughout the sentence utterance was observed in the SD subjects. Ludlow, Naunton and Bassich[13] measured mean frequency perturbation (jitter) and mean amplitude perturbation (shimmer) on four AD-SD patients. Jitter value were from 0.0678 to 0.2481 msec and shimmer value were from 8.02 to 20.62 %. These data were higher than in age- and gender-matched normal control subjects. Their study included laryngeal video recordings using a laryngo-fiberscope to judge the abnormalities of larynx. Davis, Boone, Carroll, Darveniza, and Harrison[16] reported on phonatory airflow during sustained phonation, maximum phonation time, and speaking fundamental frequency during the reading of the whole Rainbow passage by 23 AD-SD subjects. The mean airflow was 90.5 ml/sec for the 14 male subjects, and 77.2 ml/sec for the 6 female subjects. Moreover three patterns of phonatory airflow seemed to emerge: oscillating, irregular, and steady. The maximum phonation times were from 4.4 to 24.2 sec for the males, from 17.2 to 26.1 sec for the females. The mean fundamental frequency was 162 Hz for the 16 male subjects, 134 Hz for the 7 female subjects, and there were no significant differences when subjects were compared to age- and gender-matched controls.

Thus, studies to date have 1) identified at least three forms of SD 2) demonstrated that the groups can be differentiated with acoustic and aerodynamic analysis and 3) proven to be useful to judge the degree of symptoms, to infer the need for treatment, and to evaluate the effectiveness of treatment. In spite of the demonstrated merits the application of voice analysis is limited. Two problems common to all of these studies is the small number of subjects, and/or that only one or two parameters were measured. Limited parameter testing means that clinicians in other facilities, wishing to make direct comparisons, must have similar equipment for making measurements of Fo or formant structure. Many clinics do not have equipment for acoustic analysis making clinical comparisons and generalization of results difficult. Consequently, the purpose of the current study was to make comparison of AD-SD and AB-SD patients to a non-dysphonic population on multiple parameters of vocal function including both quantitative and qualitative measurements.

Methods

Subjects

Fourteen AD-SD subjects (three males and seven females) and eight AB-SD subjects (five males and three females) served as subjects for this study. Sixty-nine subjects (25 males and 44 females) without voice complaints were selected as controls. Controls were used to provide

subjects for which similar data had been collected, in order to enhance the understanding of the peculiarities of AD-SD or AB-SD. A summary of the three groups of subjects is detailed in Table 1. The age of subjects were nearly identical between the AD-SD group and normal group. Although the AB-SD group was slightly younger, there were no statistically significant age differences found between groups. Diagnosis of AD-SD and AB-SD was made using typical qualitative clinical criteria as judged by otolaryngologists and speech pathologists with extensive experience with SD patients.

Table 1 Age in years for three groups of male and female speakers

	GROUP AGE mean (range)	MALE mean	FEMALE mean
AD-SD subjects	51.1 (24-77) n=14	35.7 n= 3	55.3 n=11
AB-SD subjects	46.5 (33-61) n= 8	51.4 n= 5	38.3 n= 3
Normal subjects	51.0 (30-92) n=69	49.3 n=25	52.0 n=44

Data Collection and Analysis

Five types of data were collected in this study: perceptual voice judgments, indirect laryngoscopic examinations, videostroboscopic laryngeal examinations, aerodynamic examinations, and acoustic recordings.

Judges

All three judges used in this study had normal vision, normal hearing and had ten or more years of experiences with voice disorders: two were otolaryngologists and one was a speech pathologist. The auditory perceptual tasks included listening to an audio-tape of randomized samples of the recorded voice of each subject to assess GRBAS values and make qualitative descriptions of the SD symptoms. Video observations consisted of assigning a stroboscopic value (three point rating scale) for symmetry, regularity, glottal closure, amplitude and mucosal wave.

Perceptual Voice Judgments

An audio recording was made using a high-quality reel-to-reel tape recorder (Nagra LVII model) and a Cardioid microphone in a sound-treated room with the subjects seated in a chair with a constant Mic-Mouth distance of 12 inches. Recording levels were consistent throughout the data collection. Perceptual judgments were made using the "GRBAS scale". The GRBAS scale was developed by the Committee for Phonatory Function Tests of the Japan Society of Logopedics and Phoniatrics. This scale uses a 4-point equal-appearing interval scale (0: normal, 1: slight impairment, 2: moderate impairment, 3: severe impairment) to rate grade (G), rough (R), breathy (B), asthenic (A), and strained (S). Scale G represents the degree of hoarseness or voice abnormality. Scale R represents a psycho-acoustic impression of the irregularity of vocal fold vibrations. Scale B represents a psycho-acoustic impression of the extent of air leakage through the glottis. Scale A denotes weakness or lack of power in the voice. And scale S represents a psycho-acoustic impression of a hyperfunctional state of phonation[18].

Additionally, judges were asked to note the presence or absence of vocal characteristics typically associated with SD. For this task the judges were provided a check list of symptoms first described by Izbedski and Dedo[19]. The examiners listened to audio-tape recordings of free conversation, the Rainbow passage, and sustained vowels. For each sample they were asked to judge the presence or absence a these vocal characteristics.

Indirect Laryngoscopic and Stroboscopic Examinations

An indirect laryngoscopic examination using a standardized observation light and stroboscopic examination were recorded on VHS video tapes using a 90 degree Wolfe rigid indirect laryngoscope and Sony VTR system.

An indirect laryngoscopic examination was conducted to evaluate laryngeal status using the 5-point rating scale (1: normal or none, 2: mild, 3: moderate, 4: marked, 5: severe). The items observed were spasmodic activity, inflammation, weakness, scarring, flaccidity, bowing, atrophy and displacement.

A stroboscopic evaluation was conducted of SD groups to evaluate status using the 3-point rating scale on symmetry (0: symmetrical, 1: asymmetrical in amplitude or phase, 2: asymmetrical in amplitude and phase), regularity (0: regular, 1: inconsistent, 2: irregular), glottal closure (0: complete, 1: inconsistent, 2: incomplete), amplitude (0: normal, 1: small, 2: zero) and mucosal wave (0: normal, 1: small, 2: zero). Judges were able to view the tape as many times as they felt were necessary to make decisions to assign a value to each parameter.

Both indirect laryngoscopy and videostroboscopy procedures were completed during sustained vowel /i/ production, and non phonatory tasks.

Vocal Function Examinations

Measurements were recorded using a facemask attached to the Nagashima PS-77 phonatory function analyzer. Airflow, frequency and intensity were recorded on sustained vowels produced at normal pitch and loudness, high pitch and low pitch, loud and soft conditions, as well as maximum phonation time (MPT).

Acoustic Recordings

The voice sample included sustained vowels /a/ and /i/ at comfortable pitch, conversational speech, and repetitions of the sentence "The blue spot is on the key". The vowel /a/ was extracted from the sustained vowel samples obtained for each subject within each of the three groups. The vowel /i/ was extracted in two ways: the SD group's vowel samples were obtained from the sustained production and the vowel "key" produced within the sentence "The blue spot is on the key"; the normal group's vowel sample was obtained from the sustained vowel. These samples were digitized and analyzed using CSpeech[20] on an IBM PC-AT.

CSpeech is a signal-processing technique that derives acoustic measure, and signal-to-noise ratio from the least mean square fit of a waveform model to the prespeech waveform. The speech waveform was sampled at a rate of 8.3 kHz and an interpolation technique was used to improve the temporal resolution of the model fit. Following Milenkovic's protocol[20], a 1,000 msec midsection samples of the sustained vowels /a/ and /i/ were analyzed for measures of mean of jitter, shimmer, and signal-to-noise ratio (SNR). Additionally, 100 msec section samples of the /i/ produced in the test sentence were also analyzed.

Statistical Analysis

The STAX statistical software program, developed for NEC PCs, was used to determine mean, sd, range and Student T-test. The p value was determined to be significant at a level of 0.05 or less.

Results

Perceptual Voice Judgments

Figure 1 shows the mean scale points for each GRBAS parameter. Grade was high for both SD groups, Breathy and Asthenic were high for the AB-SD group, and Strained was high for the AD-SD group. These scaler values indicate the characteristic of hoarseness for the AD-SD is Strained, and for the AB-SD is Asthenic and Breathy.

The vocal symptoms of AD-SD and AB-SD subjects are shown in Figure 2. In over 50% of both types of SD patients, intermittent phonation, problems in loudness control, strained hoarseness, glottal spasms and vocal prolongations are judged affirmatively. Vocal tension, strangled voice, glottal fry and choked vocal attack were distinctively characteristic of the AD-SD group, while intermittent aphonia, breathy phonation, whispered speech and periods of normal voice were distinctive characteristics of the AB-SD group.

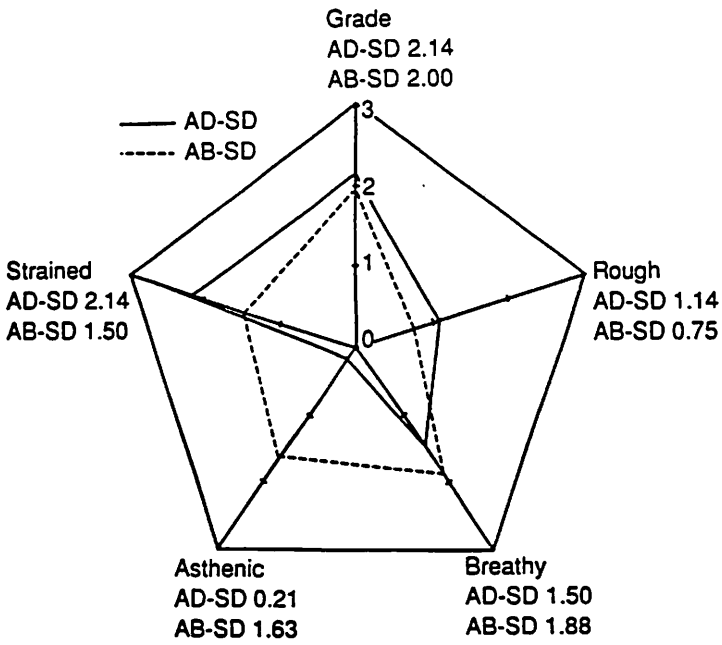

Fig.1 Graphic display of the mean scale points of the 4-point GRBAS perceptual scale (0:normal, 1:slight impairment, 2:moderate impairment, 3:severe impairment) obtained for the AD-SD and AB-SD groups of speakers.

Indirect Laryngoscopic Examinations

Figure 3 shows vocal fold status values obtained from an indirect laryngoscopic examination. Abnormal movement activity was characteristically observed for all subjects in both AD and AB groups. Other vocal status items were generally unremarkable.

Videostroboscopic Examinations

Figure 4 illustrates mean stroboscopic observation of five parameters. These results suggest that regularity, amplitude and mucosal wave were poor in the AD-SD group, while glottal closure and mucosal wave were poor in the AB-SD group.

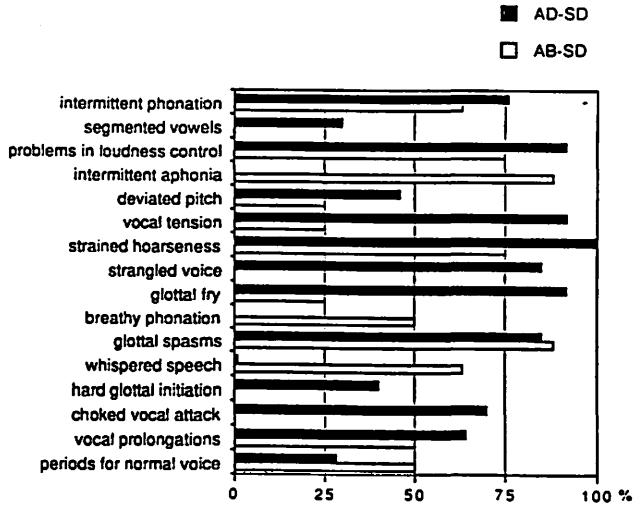

Fig.2 Summary of the different voice characteristics observed in the AD-SD group and the AB-SD group of speakers.

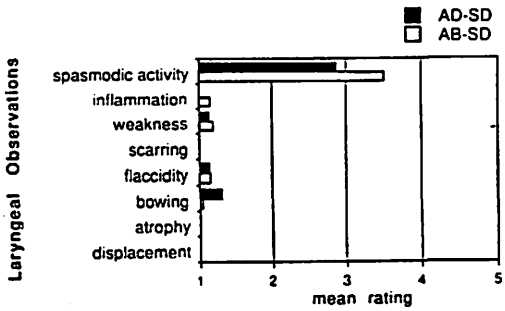

Fig.3 Mean ratings for parameters scored during indirect laryngoscopic examinations of two SD groups, using a 5-point scale (1:normal or none, 2:mild, 3:moderate, 4:marked).

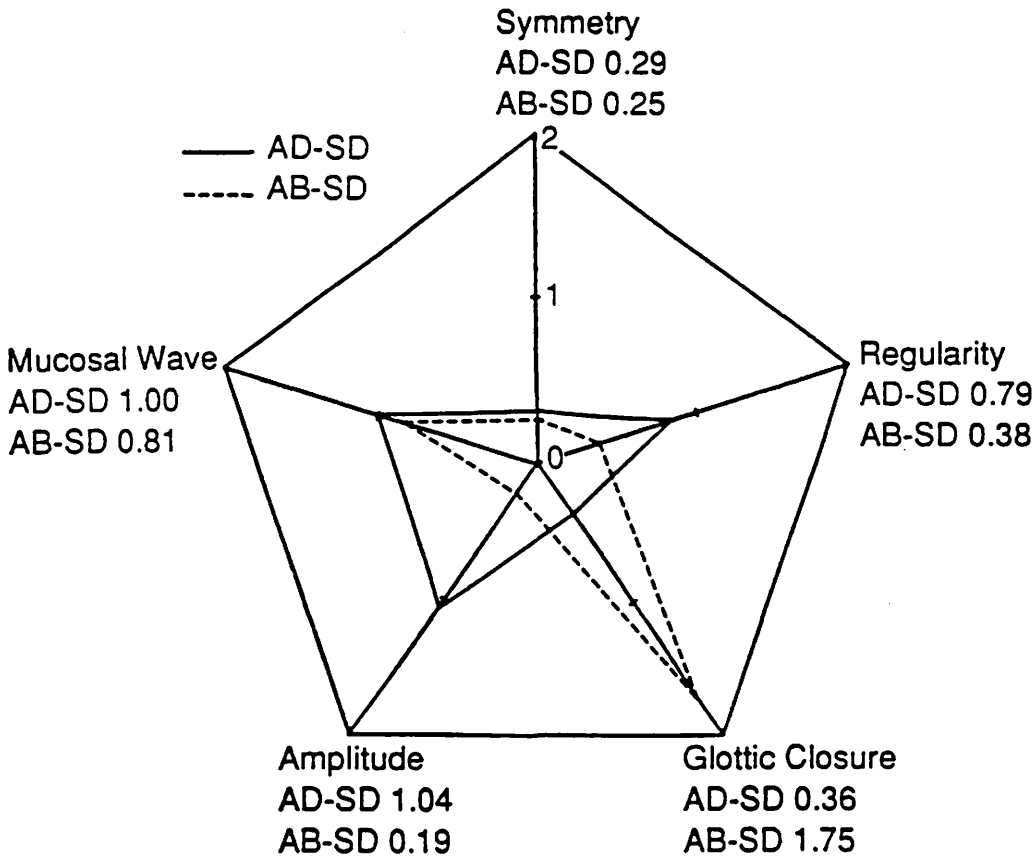

Fig.4 Graphic display of mean ratings for five stroboscopic parameters: symmetry (0:symmetrical, 1:asymmetrical in amplitude or phase, 2:asymmetrical in amplitude and phase); regularity (0:regular, 1:inconsistent, 2:irregular); glottal closure (0:complete, 1:inconsistent, 2:incomplete); amplitude (0:normal, 1:small, 2:zero); and mucosal wave (0:normal, 1:small, 2:zero).

Vocal Function Examinations

Fundamental frequency at habitual pitch, maximum high pitch and minimum low pitch for the vowel /a/ produced by the SD groups and normal group are shown in Table 2 and Figure 5. The mean fundamental frequency for the SD groups was higher than that for the normal group. Generally, the normal group had a greater pitch range than both of the SD groups. The differences between AD-SD group and normal group in male low pitch and female high pitch were supported statistically.

Intensity levels at habitual pitch, maximum high and minimum low for the vowel /a/ for each of the three groups are shown in Table 3 and Figure 6. The intensity at habitual pitch and minimum low intensity were similar across groups. Maximum high intensity was lower statistically in male and female SD groups. There were no statistically significant differences between the two SD groups.

Table 2 Fundamental frequency (Hz) at habitual, maximum high and minimum low pitch levels for the vowel /a/.

	Habitual		High		Low	
MALE	mean	sd	mean	sd	mean	sd
AD-SD subjects	187	64	373	115	113**	15
AB-SD subjects	120	24	333	124	90	9
Normal subjects	109	30	391	207	70	25
FEMALE						
AD-SD subjects	227	81	400**	82	133	44
AB-SD subjects	200	42	380	100	142	3
Normal subjects	198	26	620	214	138	29

There were statistically significant differences (**p<0.01) between the AD-SD group and the normal group.

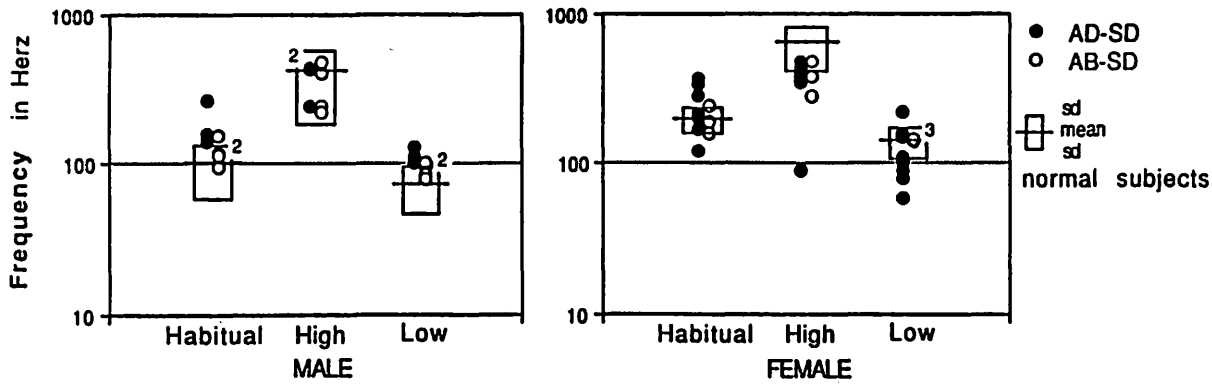

Fig. 5 Fundamental frequency (expressed in hertz displayed on log scale) obtained at habitual, maximum high and minimum low pitch levels for the vowel /a/ for three groups of speakers (normal subjects, AD-SD subjects and AB-SD subjects).

Mean airflow rate for the vowel /a/ is shown in Table 4 and Figure 7. There was considerable variability on this measure, with flows ranging from 18 to 840 ml/sec. The mean flow data was higher in the AB-SD group, and statistically lower in the female AD-SD group.

MPT of the SD groups are shown in Figure 8. The mean MPT was longer in the AD-SD group than in AB-SD group.

Table 3 Intensity levels (dB re: 0.0002 dynes/cm^2) at habitual, maximum high and minimum low levels for the vowel /a/.

	Habitual		High		Low	
MALE	mean	sd	mean	sd	mean	sd
AD-SD subjects	75.7	4.6	89.7**	6.5	61.3	2.3
AB-SD subjects	71.3	1.7	90.8**	6.6	64.8	5.9
Normal subjects	72.1	4.9	98.1	4.4	61.6	2.1
FEMALE						
AD-SD subjects	70.6	6.8	83.9**	8.9	63.1	5.0
AB-SD subjects	66.3	3.5	84.0**	9.6	61.3	1.0
Normal subjects	69.7	6.4	94.8	5.8	60.8	1.1

There were statistically significant differences (**$p<0.01$) between the SD groups and the normal group.

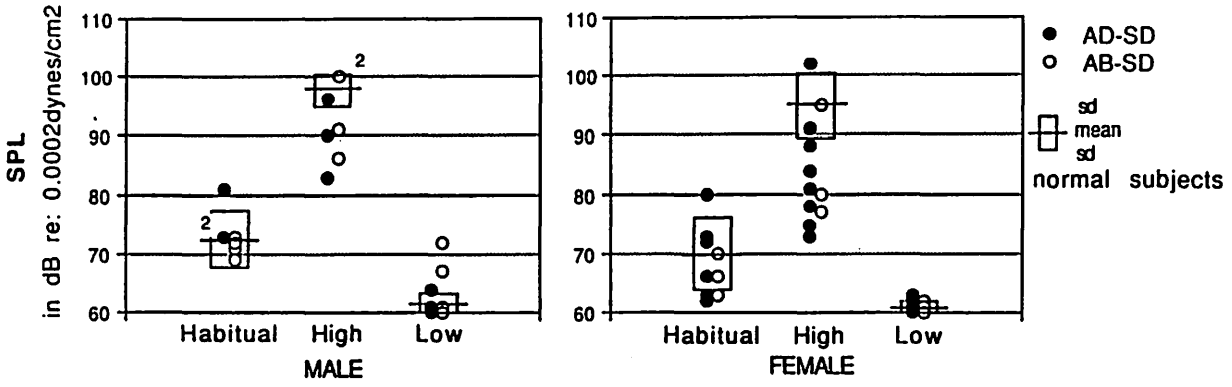

Fig. 6 Intensity levels expressed (in dB re: 0.0002 dynes/cm^2) recorded at habitual, maximum high and minimum low levels for the vowel /a/ for three groups of speakers (normal subjects, AD-SD subjects and AB-SD subjects).

Table 4 Mean airflow rate (ml/sec) for the vowel /a/.

	MALE		FEMALE	
	mean	sd	mean	sd
AD-SD subjects	157.7	62.4	79.2*	44.2
AB-SD subjects	325.5	357.5	216.7	168.1
Normal subjects	126.8	45.5	121.1	49.8

There was a statistically significant difference (*$p<0.05$) between the female AD-SD group and the female normal group.

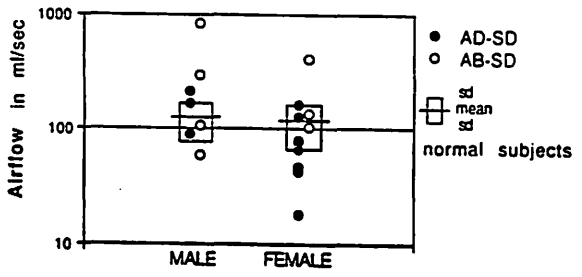

Fig.7 Mean airflow rates (expressed in ml/sec on log scale) during production of the sustained vowel /a/ for each of three groups of speakers (normal subjects, AD-SD subjects and AB-SD subjects).

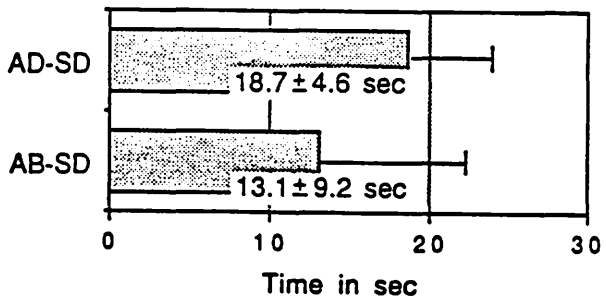

Fig.8 Maximum phonation time in seconds obtained for the vowel /a/ for the AD-SD and AB-SD groups of speakers.

Acoustic Analysis

Jitter for the sustained vowels /a/ and /i/ for each of the three group is shown in Table 5 and Figure 9. For the SD groups, Jitter for the vowel /i/ in the sentence "The blue spot is on the key" is also shown in Figure 9. Although jitter was higher in the SD groups than in the normal group, there was no significant differences except for the female AD-SD group. In the SD groups, jitter was higher on the sustained vowel /a/ than /i/, and higher in /i/ produced in the sentence than in sustained the vowel /a/.

Shimmer for the sustained vowels /a/ and /i/ for each of three groups, and /i/ in the sentence for the SD groups are shown in Table 6 and Figure 10. In males, the differences in shimmer between the groups were small for the sustained vowels, but were high in the sentences in the AB-SD group. In females, there was high variability and the mean shimmer was higher in the AD-SD group. There were no statistically significant differences found between groups. Standard deviations were high.

SNR for the sustained vowels /a/ and /i/ are shown in Table 7 and Figure 11. SNR of vowel /i/ in the sentence for the SD groups is also depicted in Figure 11. SNR was significantly lower in the female SD groups than in their normal counterparts.

Figures 12, 13 and 14 illustrate the discrepancy between sustained vowel /i/ and the vowel /i/ produced in connected discourse. Figure 12 indicates that jitter in sustained vowel /i/ was lower than the same vowel produced in a sentence. Similarly shimmer also was different for the two conditions (Figure 13), as was SNR (Figure 14).

The mean total time taken to produce the sentence "The blue spot is on the key" was measured for the SD groups (Total Time), as was the time period from sentence initiation until the first voice break (Break Time). The mean value was calculated from three tokens. Generally speaking, the total time was longer and the time prior to a voice break was shorter in the AD-SD group.

Figure 15 shows voice onset time (VOT) for "p" in the sentence "The blue spot is on the key". The mean time for the AB-SD group was shorter than for the AD-SD group.

Table 5 Jitter (msec) for the sustained vowels /a/ and /i/, and for the vowel /i/ in "The blue spot is on the key"

	/a/		/i/		/i/ in the key	
MALE	mean	sd	mean	sd	mean	sd
AD-SD subjects	0.096	0.063	0.031	0.012	0.144	0.191
AB-SD subjects	0.235	0.252	0.127	0.152	0.713	0.455
Normal subjects	0.062	0.046	0.053	0.054		
FEMALE						
AD-SD subjects	0.429*	0.381	0.197	0.251	0.414	0.390
AB-SD subjects	0.241	0.311	0.068	0.014	0.920	0.865
Normal subjects	0.079	0.173	0.040	0.074		

There was a statistically significant difference (*$p<0.05$) between the female AD-SD group and the female normal group.

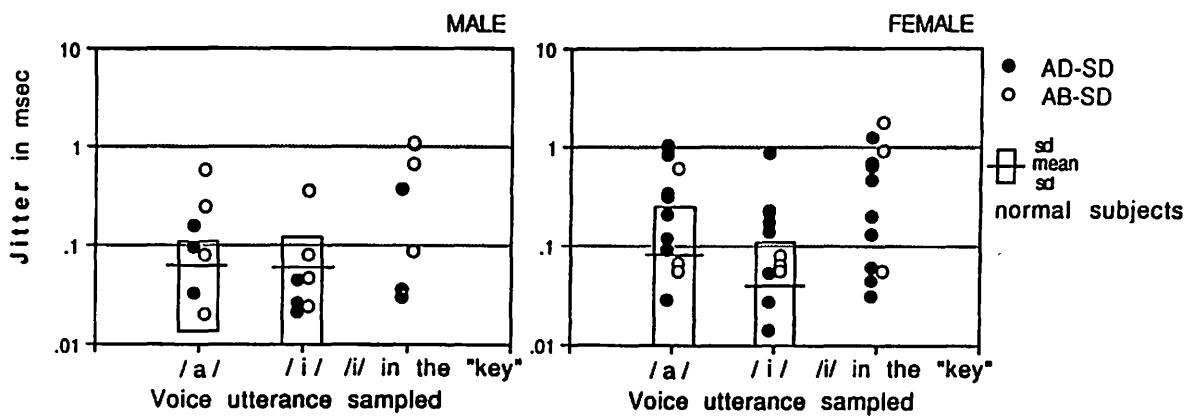

Fig.9 Jitter (expressed in msec displayed on log scale) measured during the production of sustained vowels /a/ and /i/ for each of three groups of speakers (normal subjects, AD-SD subjects and AB-SD subjects), and for the vowel /i/ produced in the sentence "The blue spot is on the key" for the AD-SD and AB-SD groups of speakers.

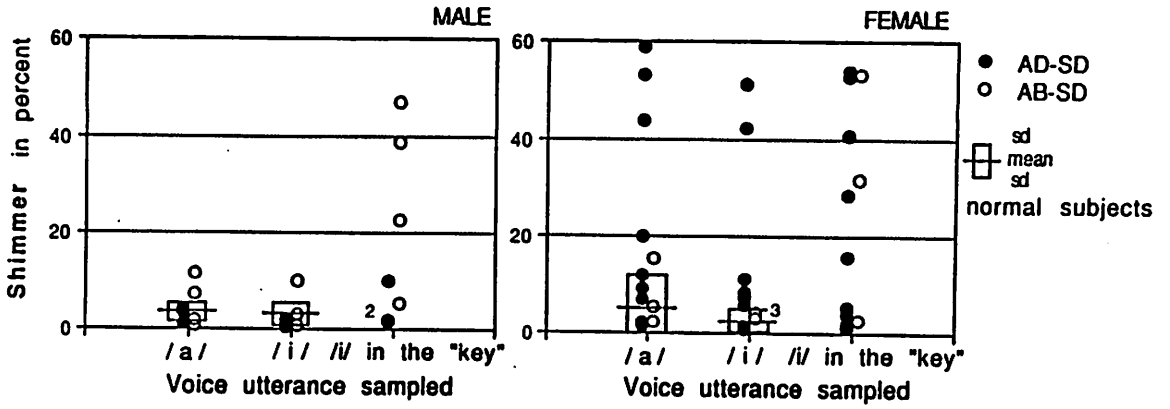

Fig.10 Shimmer (expressed in percent) measured during the production of sustained vowels /a/ and /i/ for each of three groups of speakers (normal subjects, AD-SD subjects and AB-SD subjects), and for the vowel /i/ produced in the sentence "The blue spot is on the key" for the AD-SD and AB-SD groups of speakers.

Table 6 Shimmer (%) for the sustained vowels /a/ and /i/, and for the vowel /i/ in "The blue spot is on the key".

	/a/		/i/		/i/ in the key	
MALE	mean	sd	mean	sd	mean	sd
AD-SD subjects	2.93	1.66	1.15	0.76	4.41	4.74
AB-SD subjects	5.46	5.07	3.70	4.31	28.47	18.23
Normal subjects	3.93	1.62	3.16	2.39		
FEMALE						
AD-SD subjects	20.99	22.25	13.03	18.16	20.84	21.33
AB-SD subjects	7.80	6.85	3.80	0.61	29.20	25.26
Normal subjects	5.26	7.42	2.61	2.43		

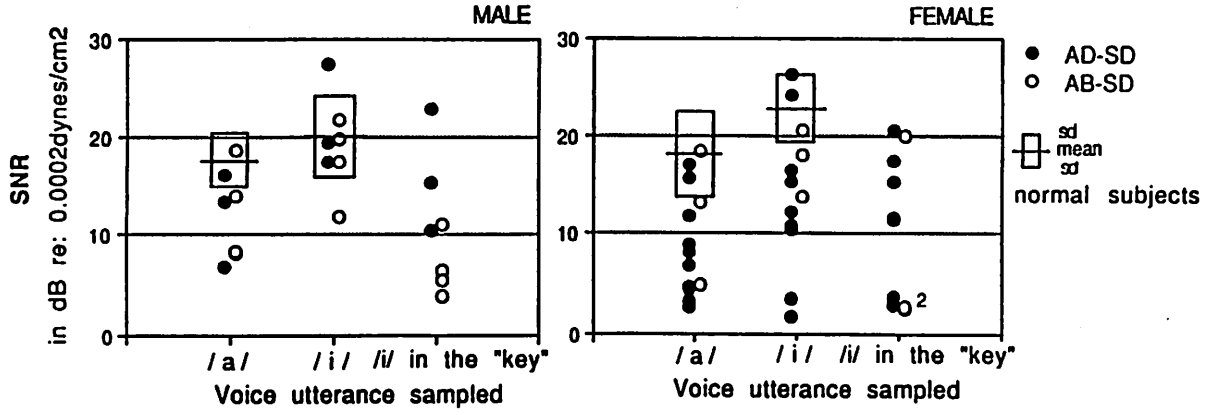

Fig.11 SNR (expressed in dB re: 0.0002 dynes/cm²) measured during the production of sustained vowels /a/ and /i/ for each of three groups of speakers (normal subjects, AD-SD subjects and AB-SD subjects), and for the vowel /i/ produced in the sentence "The blue spot is on the key" for the AD-SD and AB-SD groups of speakers.

Table 7 SNR (dB re: 0.0002 dynes/cm^2) for the sustained vowels /a/ and /i/, and for the vowel /i/ in "The blue spot is on the key".

	/a/		/i/		/i/ in the key	
MALE	mean	sd	mean	sd	mean	sd
AD-SD subjects	12.07**	4.77	21.43	5.25	16.20*	6.30
AB-SD subjects	12.28**	4.98	17.68	4.28	6.69	3.05
Normal subjects	17.73	2.74	20.14	4.08		
FEMALE						
AD-SD subjects	8.36**	5.07	14.80*	8.97	11.35	6.17
AB-SD subjects	12.17*	6.81	17.43**	3.48	8.47	9.99
Normal subjects	18.47	4.55	23.10	3.44		

There were statistically significant differences between the SD groups and the normal group (**p<0.01, *p<0.05) in the sustained vowel /a/ and /i/, and also between the male AD-SD group and the male AB-SD group (*p<0.05) in the vowel /i/ in "key".

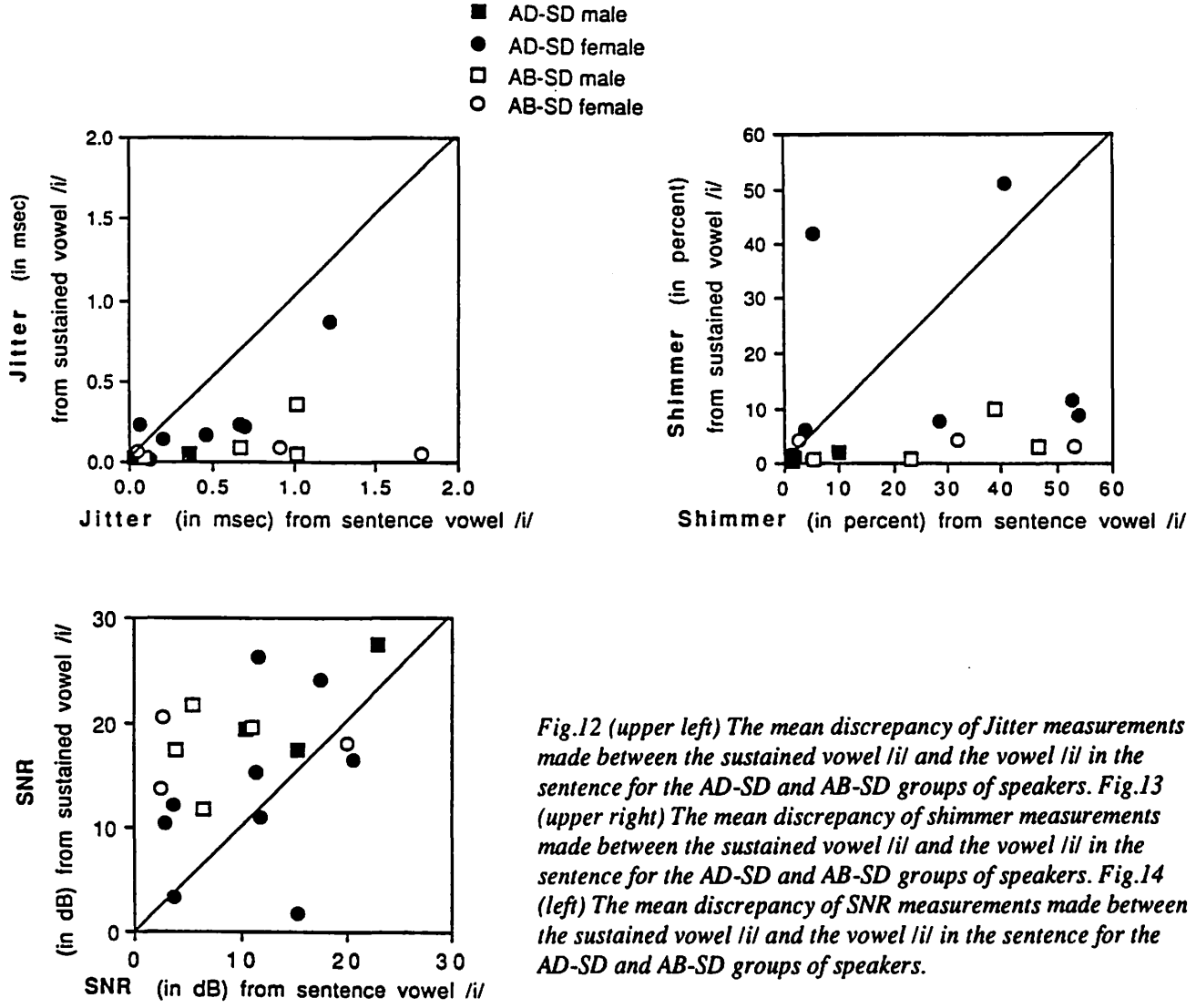

Fig.12 (upper left) The mean discrepancy of Jitter measurements made between the sustained vowel /i/ and the vowel /i/ in the sentence for the AD-SD and AB-SD groups of speakers. Fig.13 (upper right) The mean discrepancy of shimmer measurements made between the sustained vowel /i/ and the vowel /i/ in the sentence for the AD-SD and AB-SD groups of speakers. Fig.14 (left) The mean discrepancy of SNR measurements made between the sustained vowel /i/ and the vowel /i/ in the sentence for the AD-SD and AB-SD groups of speakers.

Discussion

In general, the results of this study demonstrate distinct differences between the AD-SD and AB-SD group of subjects. In all cases there appeared to be some abnormal movement patterns that were characterized in the AD-SD group by hyperadduction of the true vocal folds or of both the true and false vocal folds during phonation. In the AB-SD group the abnormal movement was demonstrated by loss of tonus of the true vocal folds, or incomplete glottic closure.

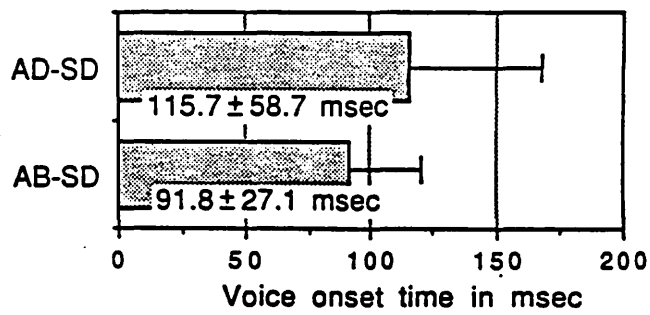

Fig.15. The voice onset time for "p" in the sentence "The blue spot on the key" for the AD-SD and AB-SD groups of speakers.

Perceptually the overall severity, as measured by "GRBAS" from the GRBAS scale, was similar for the two SD groups though the specific characteristics differed considerably. The AD-SD group most commonly had problems in strained hoarseness, problems in loudness control, vocal tension, glottal fry, strangled voice, glottal spasms, intermittent phonation, and choked vocal attack. Less frequently they exhibited intermittent aphonia, breathy phonation, whispered speech. These characteristics are similar to those reported by Ludlow[13] for four AD-SD patients. In contrast the AB-SD group exhibited intermittent aphonia, glottal spasms, problems in loudness control, strained hoarseness, intermittent phonation, and whispered speech.

Direct observation of the larynx from indirect mirror examinations and videostroboscopy revealed several interesting findings. From the indirect examination it was noted that the majority of the larynges appeared to be normal. The only exceptions were found in two of the AB-SD patients who had mild-moderate bowing. This bowing may help explain the pathogenesis of this disease. For example, in select cases there may be underlying muscle atrophy resulting in insufficient soft tissue for adequate closure without co-occurring rapid muscle fatigue. For others, the inability to sustain closure may come from an inability of the cricothyroid to sustain tension. The fact that all patients did not look similar is not surprising considering that the disease does not necessarily have a single etiology[14] nor is the severity of expression of the disease identical in all patients. Most importantly SD is considered to be a movement disorder that disrupts connected speech far more than it disrupts isolated vowel production.

Results of the stroboscopic examination support the contention that not all SD patients, even within the same type, are similar. In both groups the larynx appeared to generally be symmetrical. In the case of AD-SD the amplitude was reduced and the closed phase was long while in the AB-SD group the close phase was short, the amplitude variable, and the vibration highly irregular-frequently appearing like a flaccid flutter. Mucosal wave was abnormal in both groups, though the etiological bases was different. In the case of the AD-SD group there was extreme stiffness due to the increased hyperfunction, while in the AB-SD group there was little tension and incomplete closure creating a gap and air leak, thus making it difficult to generate sufficient subglottal pressures to help blow the vocal folds apart and create the mucosal wave.

There is limited extant data on aerodynamic and acoustic characteristics of individuals with SD. In this study we found that habitual fundamental frequency did not differ among groups. This

finding is in accordance with previous studies[16,21]. Our results could not be compared directly with Ludlow and Conner[21], because the fundamental frequency in this study was collapsed across male and female speakers. However, the data is comparable to that reported by Davis[16], who also reported no significant differences between groups, though a somewhat high fundamental frequency in the AD-SD group presumably because of increased tension. The maximal range task revealed no statistically significant differences between the AD-SD males and the other two groups. The range was constricted particulary at the upper end. Likewise, the AD-SD males exhibited a constricted range, only in this case the lower portion of the register was most affected. The intensity range found in subjects in the present study was reduced compared to the normal controls used in this study as well as those reported elsewhere[21]. Most of the SD speakers had normal airflow. The exceptions being 3/7 of the AB-SD group who had abnormally high flows and 3/11 of AD-SD group with low flows. The direction of these differences is consistent with the excessive closure patterns seen in the AD-SD cases and decreased closure seen in the AB-SD cases. What is puzzling is the large number of cases with normal flows in spite of having abnormal stroboscopic and perceptual examinations. This finding may be a function of the definition of normal; airflow has a large range of values (80-220 ml/sec) considered to be normal. Moreover, the current study does not provide longitudinal data to determine what flows the subjects exhibited prior to the onset of the disease. It may be that individual subject's flows changed as much as 100ml/sec, as a function of the disease, but the flow values still fell within normal limits.

MPT is one of the most commonly completed voice examinations. The normal time is 29.7-34.6 sec in male and 17-25.7 sec in female[22-24]. It is known that MPT is shorter in various voice disorders. In our study, mean MPT was shorter for both SD groups of reported than that previously reported for normal subjects, and MPTs for the AB-SD group was shorter than for the AD-SD group. The shorter phonation times are thought to reflect poorer glottal aerodynamic efficiency: the AB-SD patients had high airflows and wasted considerable expiratory air power at the glottis because the vocal folds failed to approximate; the AD-SD patients exhibited shorter phonation times because of the abnormal laryngeal movement patterns of hyperfunctional closure patterns cutting off phonation.

Several additional acoustic measurements are worthy of mention. Jitter and shimmer were higher and SNR was lower than normal in SD groups, though with the exception of jitter for the female AD-SD group and SNR these differences were not statistically significant. These acoustic results are difficult to compare to existing literature because of differences in populations, equipment, experimental tasks, and algorhythms used to calculate the measures. From our previous study, jitter was similar for the sustained vowels and for vowels produced in connected speech[25]. The current study showed that for a majority of SD subjects, jitter, shimmer and SNR of the vowel in the sentence were considerably worse than those of the sustained vowel.

Timing characteristics are some of the voice parameters most affected by SD. Ludlow and Connor used the total time taken to produce a sentence as a measure of change following botulinum toxin injections[17]. They found the mean sentence time was shorter following treatment. In the current investigation, the mean sentence time of the AD-SD group was longer than the AB-SD group. The mean length of time period before the occurrence of a voice break was shorter in the AD-SD group. This may mean that the AD-SD group was more severely involved than the AB-SD group. The break time in the AD-SD group did not depend on aerodynamic efficiency, because the mean MPT was longer in this group of AD-SD subjects. This is interpreted to mean that vocal coordination necessary to produce a sentence is more difficult for the AD-SD patients. A second

timing measure was the VOT. VOT was longer in the AD-SD group. These results suggest that it was difficult to begin phonation smoothly and to sustain the voice through out the sentence, particularly in the AD-SD patients. Since these timing values are clearly abnormal results suggest that timing measures may be good indicators of change following treatment.

In summary, this study demonstrated that spasmodic dysphonia has several interesting features in terms of perceptual voice judgments, stroboscopic, aerodynamic and acoustic examinations. There are differences exhibited between the AD-SD group and the AB-SD group in selected components of all of these examinations. It is concluded that objective examinations are useful supplements to subjective examinations in describing spasmodic dysphonia. Routine use of these measures to describe SD populations pre- and post-treatment may help clinicians make cross-institutional comparisons to determine the efficiency of treatment, and to better understand the manifestation of the disease ultimately helping to illuminate its pathogenesis.

References

1. Traube L. Spastishe Form der Nervosen Heiserkeit. In: Traube L, ed. Gesammelte Beitraege zur Pathologie und Physiologie. Vol 2. Berlin: Hirschwald, 1871; 674-8.
2. Arnold GE. Spastic dysphonia: I. Changing interpretations of a persistent affliction. Logos 1959; 2: 3-14.
3. Dedo HH, Izdebski K. Intermediate results of 306 recurrent laryngealnerve section for spastic dysphonia. Laryngoscope 1983; 93: 9-16.
4. Aronson AE. Psychogenic voice disorders: an interdisciplinary approach to detection, diagnosis and therapy. Philadelphia, PA: W. B. Saunders Company, 1973.
5. Dedo HH. Recurrent laryngeal nerve section for spastic dysphonia. Ann Otol Rhinol Laryngol 1976; 85: 451-9.
6. Wolfe VI, Bacon M. Spectrographic comparison of two types of spastic dysphonia. J Speech Hear Disord 1976; 41: 325-32.
7. Zwitman DH. Bilateral cord dysfunctions: abductor type spastic dysphonia. J Speech Hear Disord 1979; 44: 373-8.
8. Merson RM, Ginsberg AP. Spasmodic dysphonia: abductor type. A clinical report of acoustic, aerodynamic and perceptual characteristics. Laryngoscope 1979; 89: 129-39.
9. Hartman DE, Aronson AE. Clinical investigations of intermittent breathy dysphonia. J Speech Hear Disord 1981; 46: 428-32.
10. Aronson AE, Hartman DE. Adductor spastic dysphonia as a sign of essential (voice) tremor. J Speech Hear Disord 1981; 46: 52-8.
11. Cannito MP, Johnson JP. Spastic dysphonia: a continuum disorder. Journal of Communication Disorders 1981; 14: 215-23.
12. Izdebski K. Overpressure and breathiness in spastic dysphonia: an acoustic (LTAS) and perceptual study. Acta Otolaryngol (Stockh) 1984; 97:373-8.
13. Ludlow CL, Naunton RF, Bassich CJ. Procedures for the selection of spastic dysphonia patients for recurrent laryngeal nerve section. Otolaryngol Head Neck Surg 1984; 92: 24-31.
14. Aronson AE. Clinical voice disorders: an interdisciplinary approach. 2nd ed. New York, NY: Thieme Inc, 1985: 157-87.
15. Hartman DE, Abbs JH, Vishwanat B. Clinical investigations of adductor spastic dysphonia. Ann Otol Rhinol Laryngol 1988; 97: 247-52.

16. Davis PJ, Boone DR, Carroll RL, Darveniza P, Harrison GA. Adductor spastic dysphonia: heterogeneity of physiologic and phonatory characteristics. Ann Otol Rhinol Laryngol 1988; 97: 179-85.
17. Ludlow CL, Naunton RF, Sedory SE, Schulz GM, Hallett M. Effects of botulinum toxin injections on speech in adductor spasmodic dysphonia.Neurology 1988; 38: 1220-5.
18. Hirano M. Clinical examination of voice. Wien, Austria: Springer-Verlag, 1981: 83-4.
19. Izdebski K, Dedo HH. Spastic dysphonia. In: Darby JK ed. Speech evaluation in medicine. New York, NY: Grune and Stratton, 1981: 105-27.
20. Milenkovic P. Least mean square measures of voice perturbation. J Speech Hear Res 1987; 30: 529-38.
21. Ludlow CL, Connor NP. Dynamic aspects of phonatory control in spasmodic dysphonia. J Speech Hear Res 1987; 30: 197-206.
22. Sawashima M. Measurement of maximum phonation time. Jpn J Logoped Phoniat 1966; 7: 23-8.
23. Hirano M, Koike Y, von Leden H. Maximum phonation time and air usage during phonation: clinical study. Folia Phoniat 1968; 20: 185-201.
24. Shigemori Y. Some tests related to the air usage during phonation. Otologica (Fukuoka) 1977; 23: 138-66.
25. Leddy M, Bless DM. Jitter in sustained vowels and connected speech. ASHA 1989; 31-3: 77.

NCVS Status and Progress Report - 2
April 1992, 157-166

Botulinum Toxin Treatment of Spasmodic Dysphonia: Techniques, Indications, Efficacy

Charles N. Ford, M.D.
University of Wisconsin,
Department of Surgery, Division of Otolaryngology and Middleton VA Hospital
Diane M. Bless, Ph.D.
University of Wisconsin,
Department of Communicative Disorders, The Waisman Center, and Department of Surgery, Division of Otolaryngology
Nirav Y. Patel, M.D.
University of Wisconsin

Abstract

A group of 58 patients with spasmodic dysphonia was treated perorally or transcutaneously with botulinum toxin injection, and followed up for at least six months. Most were treated by peroral injection using indirect laryngoscopy to facilitate precise placement. An effort was made to disperse the toxin over the entire thyroarytenoid muscle to affect the greatest number of motor end-plates with the smallest effective dose. Severely symptomatic patients with primarily adductor spasmodic dysphonia had the best results. A potentiation effect is suggested by analysis of a sub-population of patients, mostly injected unilaterally, in which outcome was more favorable in those who had had prior injections. One-month post-injection studies reflect significant improvement in measured readings of a standard passage and this is consistent with the subjective and perceptual improvements noted. There were no severe complications. Peroral injection seems inadvisable in patients with mixed or abductor spasmodic dysphonia as well as in those patients with an uncontrollable gag reflex.

Introduction

Injection of botulinum toxin (BT) is considered by many to be the treatment of choice for symptomatic relief in patients with spasmodic dysphonia (SD). Symptoms of adductor SD can often be controlled by repeated injection of BT into the thyroarytenoid muscle (TAM). Early investigators used relatively large transcutaneous BT doses in an attempt to produce vocal fold paralysis[1] but subsequent studies[2] demonstrated effective relief of SD symptoms with substantially smaller amounts of BT. Since side effects--choking, aspiration, and phonatory breathiness--appeared to be dose-related, there was an incentive to determine the lowest dosage that would provide symptomatic relief without inducing side effects. Ford et al.[3] recently reported an effort to achieve an effective lower dose. In that study an indirect laryngoscopic peroral approach was used to visually direct BT placement and the minimal effective dose was titrated for each patient using incremental unilateral injections. The approach was based in part on the anatomical distribution of motor end-plates within the TAM as described by Rosen and co-workers[4]. Since BT acts at the myoneural junction and Rosen's laboratory showed that motor end-plates were diffusely distributed throughout the TAM, it seemed possible to spread the botulinum toxin over a wide area of motor end-plates within the muscle by indirect laryngoscopic guided placement. The peroral technique used by Ford et al.[3] was well tolerated by the patient and had the added advantage of not requiring laryngeal electromyography (EMG) equipment or expertise.

A review of the initial 38 patients treated in our laboratory by the peroral indirect laryngoscopic approach provided an array of determining factors associated with outcome[5]. Among this initial set of patients, efficacy could not be documented because objective measures of vocal function were not uniformly obtained at the same intervals following injection; patients returned when their symptoms recurred and that varied with each patient. Patients completed questionnaires and diaries that allowed subjective evaluation of voice results. Rank ordering of results based on this information permitted correlation of the initial demographic and evaluation data with ranked outcome. It was, therefore, possible to determine those factors that appeared to affect results.

To better assess outcome, 20 patients treated during the last year were required to return one month post-injection for complete voice evaluation, and diary entries were expanded. Although this study is continuing, our experience to date affords an opportunity to assess the safety, efficacy, indications, contraindications, and complications in a group of SD patients treated with peroral and percutaneous BT injections.

Materials and Methods

Intralaryngeal BT injection was used to treat 58 patients with symptomatic spasmodic dysphonia from 1988 to 1991 at the University of Wisconsin (Madison) Clinical Science Center. These patients had varying forms of SD and different degrees of symptoms. All patients underwent a complete otolaryngologic examination and a screening neurological evaluation including laryngeal EMG study of the vocalis and cricothyroid muscles. The first 38 patients were 26 females, 12 males, aged 24 to 92 (av = 52 years); 27 suffered adductor SD and 11 suffered abductor or mixed SD. Initially all 38 SD patients were treated by peroral injection of BT. During the past year, 20 additional patients were treated. Those with adductor SD were given the option of being treated perorally or percutaneously. Those with no preference were treated perorally, and those who exhibited excessive gagginess during indirect laryngoscopy were encouraged to be treated percutaneously. Of the 20 subsequent treatments, six patients were injected by the percutaneous method.

All patients were treated with lyophilized botulinum A toxin (recently supplied as Oculinum® by Allergan Pharmaceuticals, Irvine, California) reconstituted in sterile saline to provide a concentration of 2.5 units/0.1 mL. The mixture was drawn into a 1.0-cc tuberculin syringe fitted into either a laryngeal injector device for peroral injection into vocal folds[6] or an electrode needle for percutaneous EMG-guided injection. The peroral route of injection using indirect laryngoscopic guidance was used in 52 patients. This was done under topical 4% cocaine anesthesia[3], on an outpatient basis, with the patient seated upright. The vocal fold was penetrated on the superior surface and BT delivered to two sites in the anterior to posterior axis; BT was injected during needle insertion at each site to enhance dispersion in the superior to inferior axis. The percutaneous injection was made with a 25-gauge monopolar Teflon-coated needle connected to a BT-containing tuberculin syringe for delivery of the toxin. The needle was advanced through the cricothyroid membrane, directed laterally to avoid entering the laryngeal lumen, and guided by EMG signals as described by Blitzer[7].

During the first two years, unilateral 2.5-unit incremental doses were used to determine the smallest effective dose. Subsequently, new patients were treated with 2.5 units per side, and subsequent dosage increased of decreased depending on the results and side effects. In most instances the dose was adjusted downward to decrease the period of post-injection breathiness, but if the patient failed to achieve relief of spasmodic activity the dose was increased.

The initial 38 were studied to determine those factors associated with success or failure in peroral vocal fold injection of BT. These patients were ranked in order of success on the basis of differences in composite scores obtained from each patient's multifactorial subjective voice assessment prior to treatment and during their optimal voice after injection; the most improved patient was ranked number 1 and the least improved ranked 38. Factors considered for correlation with outcome included diagnosis, prior treatment, procedure tolerance, secondary diagnoses, age, and duration of symptoms. Severity of symptoms was estimated by measuring vocal fold diadochokinesia (VFDDK)--the number of utterances of the sound /i/ produced in one second. Other measures considered included maximum phonation time (MPT), jitter, signal-to-noise ratio (SNR), and the clinician's subjective rating of severity made during connected speech.

Vocal function measures from all 58 patients were obtained using the Nagashima PS 77 phonatory function analyzer during repeated samples of sustained vowels produced at different E pitch and loudness levels and during VFDDK tasks. A voice recording of both sustained vowel and connected speech samples was made on a Nagra tape recorder with a 6" mouth-to-microphone distance in a sound-treated environment. This audio recording was subsequently used for acoustic analysis and perceptual judgments. The perceptual judgments were rendered by three experienced speech clinicians listening to a master tape constructed of pre-treatment and post-treatment randomly recorded samples. Judges were naive as to the goals of the study and were asked to make a paired-comparison judgment as to which sample sounded better.

A full-face video recording of reading of The Rainbow Passage produced at normal pitch and loudness was used to determine three fluency measures: 1) time required for reading the passage, 2) percentage of words with breaks, and 3) percentage of words produced with abnormal voice. A Bruel and Kjaer 4914 and a Wolf 5012 laryngeal videostroboscope were used in the assessment of vocal fold motion, associated abnormal movements, and glottic closure.

Each patient was asked to rate his or her own voice on a scale indicating voice quality as perceived by the patient, effort required to speak, smoothness, breathiness, fatigue, and how others perceived the voice. They were asked to rate this information on a daily basis. The voice assessment battery was run prior to treatment and repeated at varying intervals post-treatment. During the

initial phases of this study, patients were not required to return at any specified interval and they usually elected to return at one month post-treatment for assessment. A total of 27 patients complied and they constitute the group used to assess the efficacy of this treatment program.

Results

Factors determining outcome in initial peroral BT injections

A number of factors proved unimportant in determining outcomes. Outcome rank did not appear to be related to patient age, nor did duration of symptoms prior to therapy have much influence. A number of preoperative voice measure analyses (MPT, airflow, frequency, intensity, jitter, shimmer, and SNR) did not correlate with postoperative outcome.

The primary factor determining outcome was diagnosis. The subjective rank-based outcome of the 27 patients with adductor SD was significantly better than the 11 patients with either mixed or abductor SD (Figure 1). The mean rank (1-38) of patients with adductor SD was 15.5 versus 29.4 for the others (Table I). A plot of the influence of severity of symptoms as measured by VFDDK shows that many of the best results occurred in patients with the worst VFDDK function (Figure 2 and Table II). Prior BT treatment seemed to have a favorable effect on outcome as measured by the duration of symptom-free interval after treatment; the duration of symptom relief increased with subsequent injections, with a mean increase of 4.4 weeks in this sample group (Figure 3 and Table III).

Table I. Rank vs. AD and AB or Mixed

	NOBS	MEDIAN	MEAN RANK	Z VALUE
AD	27	14.00	15.5	-3.49
AB or Mixed	11	33.00	29.4	3.49
Overall	38		19.5	

H = 12.20; d.f. = 1; p = 0.000

This table indicates the effect of diagnosis (the particular type of SD) on outcome based on subjective ranking.

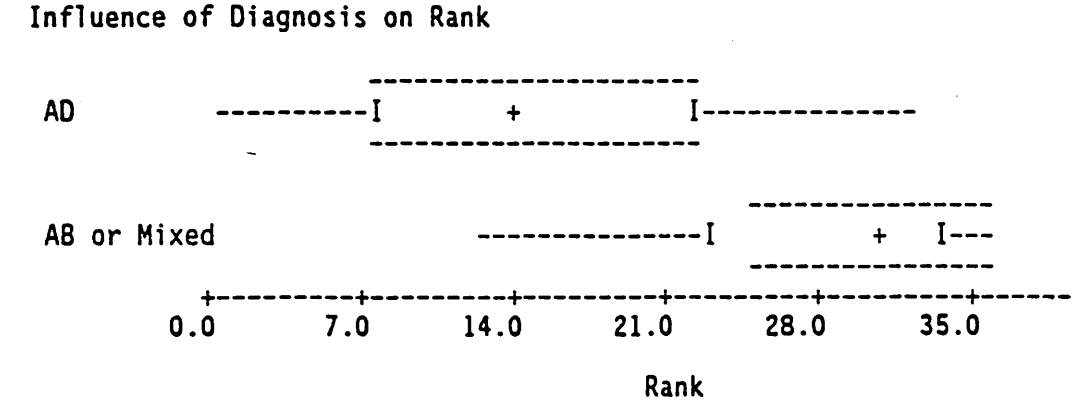

Figure 1. The outcome rank patients with adductor SD (AD) are contrasted with those patients with mixed or abductor SD (AB or mixed).

Table II. Statistical Summaries of Rank vs. Severity (The regression equation is rank = -1.94 + 5.40 per_vfddk; 31 cases were used; 7 cases contained missing values)

Predictor	C.O.V.	S.D.	t-ratio	p value
Constant	-1.941	6.670	-0.29	0.773
pr_vfddk	5.396	1.649	3.27	0.003

$s = 9.655$; $R^2 = 27.0\%$; $R^2(adj) = 24.4\%$

Analysis of Variance

SOURCE	DF	SS	MS	F	p value
Regression	1	997.89	997.89	10.70	0.003
Error	29	2703.60	93.23		
Total	30	3701.48			

These data indicate the effect of preoperative severity of symptoms (as judged by VFDDK) on outcome.

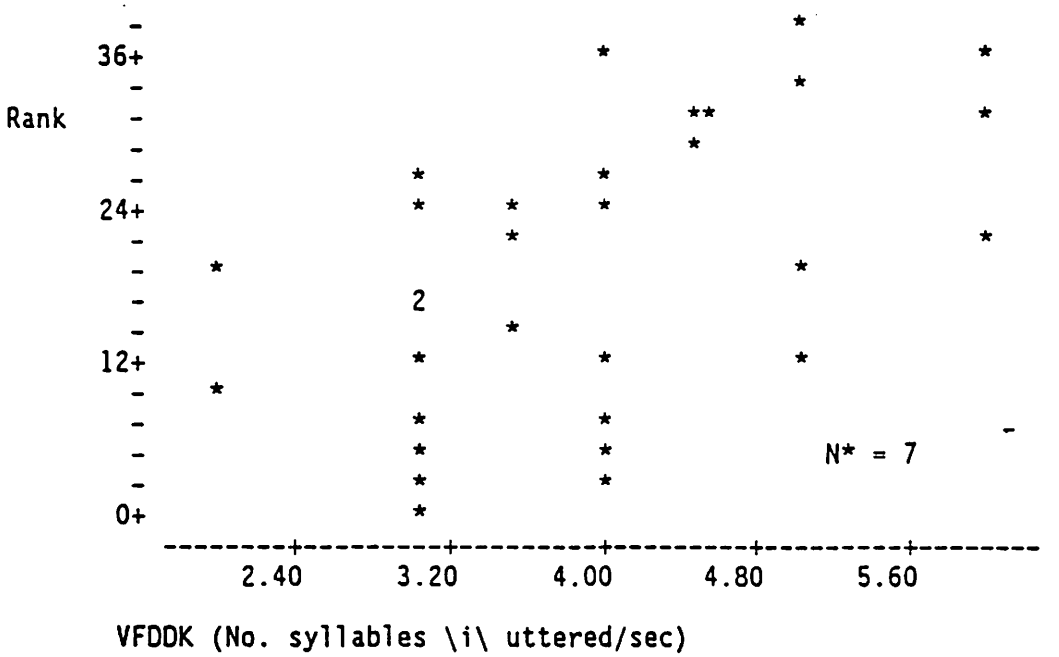

Figure 2. Rank vs. Severity of Pretreatment Symptoms (based on vocal fold diadochokinesia)

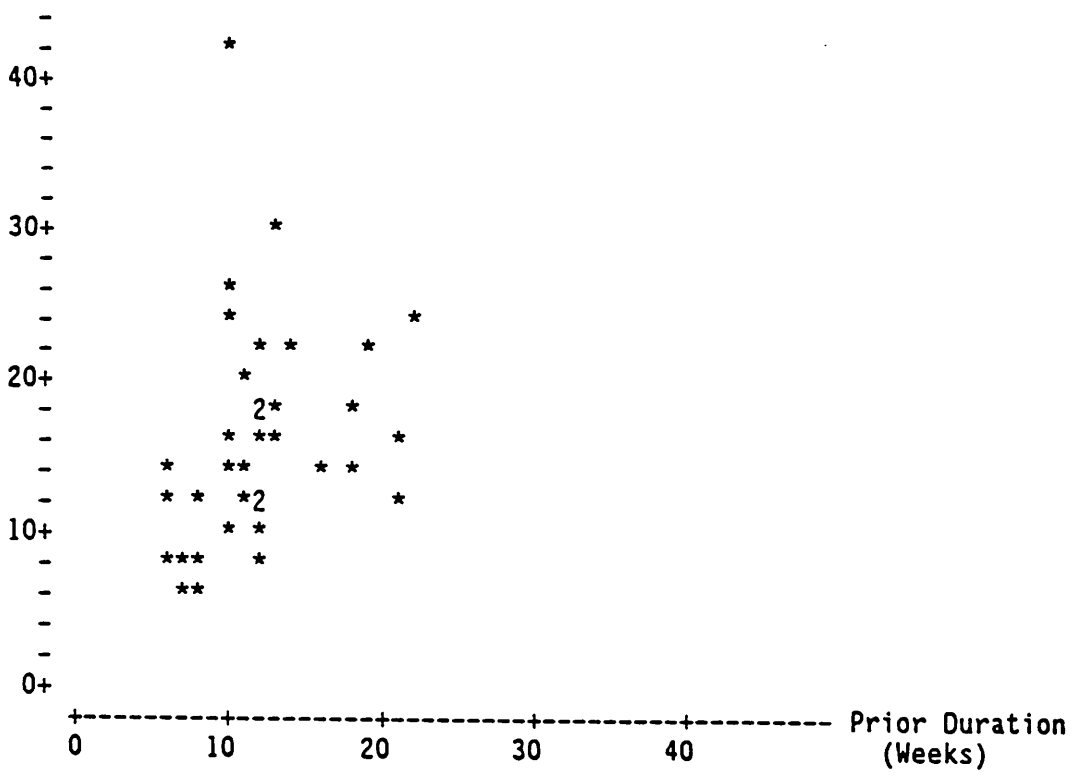

Figure 3. Duration of symptom relief of successive treatments. The plot shows the duration of any one treatment versus the duration of the previous treatment. For example, because patient x had durations of 14, 22, and 23 weeks, two points are plotted--one at (14,22) and one at 22,23).

Table III. Histogram of differences between durations of consecutive treatment

Midpoint	Count	
-10	1	*
-5	4	****
0	13	**********
5	10	**********
10	4	****
15	3	***
20	0	
25	1	*
30	0	
35	1	*

N	Mean	MEDIAN	S.D.	S.E.	WILCOXON STAT	p-value
37	4.38	3.00	8.55	1.40	450.0	0.003

WILCOXON - TEST OF MEDIAN = 0.000000 VERSUS MEDIAN N.E. 0.000000

Tolerance for peroral injection seemed to adversely affect outcome. Seven of 38 patients had hyperactive gag reflexes or aversion to intraoral manipulation; three of these were easily managed with relaxation techniques but in four cases the injection was rushed or traumatic. Since six of the seven patients in this group ranked in the bottom half (19-38), it appears that technical factors may have played a role in outcome.

Many patients had undergone extensive therapy prior to BT; such patients often did not do as well as patients who had had no prior intervention. Various forms of surgical treatment including laser surgery and recurrent laryngeal nerve section, as well as drugs and psychotherapy had been tried. Among the eight patients who had had prior surgical therapy, the three who had had nerve section ranked in the upper 50% ranking (7,9,17) whereas those who had undergone medialization procedures did poorly in the ranking (34,35,37). The particular type of SD appeared to be the dominant factor in these patients since all of the poor results were in patients with predominantly abductor SD.

Efficacy of BT injections

Efficacy was in some cases offset by side-effects. Although 43 of the 58 patients (75%) described some transient breathiness and occasional choking on liquids within a few days of injection, none suffered major complications. Two patients experienced traumatic injections as a result of gagging during injection. This resulted in slight edema of the vocal fold that resolved without treatment. One patient experienced hyperventilation during application of topical anesthetic to the larynx. One patient appeared to develop a systemic response following percutaneous injection of the vocalis; she had declined peroral injection because of gagging difficulty with previous injections. She complained of pain at the site of needle insertion during the procedure; she then became apprehensive and slightly nauseated, and complained of visual blurring. Neurological, ophthalmological, and allergic evaluations were negative except for slight pupil dilatation and decreased accommodation. She was treated with diphenhydramine (25 mg q 6 hours), the symptoms regressed overnight, and she was discharged the following day. Radioimmunoassay failed to identify IgE antibodies to BT in the patient's serum.

Efficacy studies were limited to the group of 27 patients who returned for one-month follow-up voice evaluation. There were no significant differences among the 27 for those measures made during sustained vowel production. The post-treatment recording of SNR, jitter, shimmer, and standard deviation of fundamental frequency showed little change compared to those recorded pre-treatment. For example, comparing pre-treatment to post-injection SNR, the mean change was 0.66 (S.D. = 2.5880, $p = 0.1870$).

The greatest differences attributed to BT injection were noted during conversational speech and designated readings. Both the objective timing and subjective distortion percentages measure from the connected speech samples taken from the first paragraph of the Rainbow Passage reading demonstrated significant differences (Table IV). The mean time required to read the paragraph decreased by nearly 4.5 seconds, the percentage of words judged to be abnormal decreased by 41%, and the percentage of words with voice breaks decreased by 26%.

Perceptual judgments based on tape recordings of random pre-injection and post-injection samples indicated clearly improved voices in 80% of cases, 12% were questionable or judges disagreed, and 8% judged worse. Most patients reported improved voices.

Patients' responses to questionnaires filled out during examinations formed the basis for the ratings used here. More detailed daily diaries of voice changes occurring between treatments were

kept by 12 patients. The value of this record-keeping was ambiguous, for subjects tended to rate all parameters similarly for any given day. These diaries indicated that patients felt the overall quality of their voices, breathiness, smoothness, vocal fatigue, and ease of phonation were nearly equally affected by treatment. There was little fluctuation between days of the same week but considerable fluctuation between weeks. Often the ratings on Monday predicted the subsequent rating made during the rest of the week.

Table IV. Efficacy measures comparing pre-treatment to one month post-injection based on reading first paragraph of Rainbow Passage

Function	Subjects	Mean Change	S.D.	p-value
Words in abnormal voice (%)	27	-41.37	43.4518	0.0001
Words with breaks (%)	27	-26.14	19.1573	0.0001
Time required to read	27	-4.38	9.8125	0.0284

Discussion

In any discussion of BT injection for SD, it is important to recognize that such treatment fails to address the cause of the disease or symptom complex. BT injection is a means of temporarily controlling the symptoms of SD by chemically blocking neurotransmission and altering laryngeal function. While work is continuing to elucidate the pathogenesis of SD, BT treatment offers an opportunity for patients to resume normal activities while awaiting a more lasting treatment or possible cure.

The rationale for peroral injection is based on three assumptions: 1) This approach allows precise visually controlled placement of BT so that a diffuse field of motor end-plates in the TAM can be affected with minimal dose; 2) patients may prefer this technique because it is well tolerated by most and, if dosage is reduced, there is less likelihood of dose-related side effects such as breathiness and choking; and 3) it is technique that relies on skills common to otolaryngologists and does not require EMG equipment or expertise. As with other procedures, it is important to select the proper patients for treatment.

Analysis of the data from the first 38 patients, all of whom were treated perorally, provides a patient profile of candidates suitable for this approach and suggests some contraindications. The most important factor seems to be the diagnosis. The patients with adductor SD did much better than all others (Figure 1 and Table I). Recent work by Blitzer and Brin[2] has demonstrated a technique for percutaneous injection of the posterior cricoarytenoid muscle (PCA) that has been successful in addressing the abductor SD patient; since the PCA is difficult to access using the peroral approach, patients with abductor SD should probably not be treated in this manner. Indeed the worst results, based on subjective ranking, were in the abductor group following TAM injection. Percutaneous injection also allows injection of the cricothyroid muscle, an approach the Ludlow[8] has found useful in selected abductor SD patients. Severity of symptoms appears to have an effect on

outcome based on analysis of preoperative VFDDK (Figure 2 and Table II), but patients with the worst symptoms preoperatively had the greatest room for improvement; in some severely impaired patients, the resultant voice after injection might not be better than one with milder symptoms who improved to a lesser degree.

The observation that prior BT injection was associated with improved responses (Figure 3 and Table III), based on length of symptom-free intervals, deserves special comment. Most patients returning for treatment tend to come in before symptoms are as severe as they had been prior to initial treatment, so that baselines vary. A recent report by Ludlow (presented at the State-of-the-Art Symposium on SD, Philadelphia, 1991) indicated that patients developed a tolerance to BT treatments so that results diminished over time. It is noteworthy that most of the 38 patients treated in our study were injected with minimal doses unilaterally whereas the Ludlow study was based on bilateral percutaneous injections. Our data are not conclusive enough at this time to warrant further speculation as to the reason for potentiation or tolerance; this is an issue that needs to be addressed as we continue to treat patients with BT and present alternative options for long-term management.

A number of factors appeared to have no influence on outcome; these included age and duration of symptoms at the time of presentation. Vocal function studies of sustained vowel productions made preoperatively did not seem to predict how well patients might do following BT injection. Patients who underwent extensive treatment such as surgery, drug, or psychotherapy prior to BT injection did not fare as well as others, but these patients often had diagnoses other than adductor SD. Among this group it appeared that prior recurrent laryngeal nerve section did not adversely affect outcome but the impact of other treatment may not have been a significant factor because of the confounding factor of diagnosis; in other words, patients with adductor SD did well even though they had prior nerve section, whereas patients with multiple other procedures often had abductor or mixed SD that might account for their poor results.

Gaggy patients and those with a low tolerance for indirect laryngoscopy and intra-oral manipulation tended to do poorly with the peroral injection technique. It would appear that this approach is contraindicated in such patients. We also consider patients with allergy to local anesthetics and those with active laryngitis as inappropriate candidates for peroral BT injection.

Relief of disabling voice breaks characteristic of SD is the critical factor in determining success of treatment. All 27 patients who participated in one month post-injection evaluations found it easier to talk. Analysis of voice recordings indicated that they could produce conversational speech with fewer breaks and fewer abnormal sounds. These changes, as well as the time required to read the Rainbow Passage, all indicated significant improvement at one month post-injection (Table IV). Perceptual ratings, however, showed only 80% of voices clearly improved at one month, based on comparison of tape recordings of the patient's voices. As we have found in other studies, this may reflect the failure of judges to assess effort of phonation and other subjective factors when qualitatively assessing recorded voice samples.

The use of daily diaries raises several questions about the method of rating, about a patient's ability to make judgments about voice, and about varying duration of the laryngeal dystonia symptoms. The fact that most subjects rated all parameters similarly for each day within a week leads us to suspect that their ratings were biased by overall vocal quality or effort and by the first rating of the week. We anticipated seeing frequent differences between breathiness and the other parameters. Failure to see such differences may be because breathiness is a dominant factor in the other ratings or because patients were unable to distinguish breathiness from the other features.

Summary

A group of 58 SD patients was treated with BT with no severe complications. Most were treated by peroral injection using indirect laryngoscopy to facilitate precise placement. An effort was made to disperse the BT over the entire thyroarytenoid muscle to affect the greatest number of motor end-plates with the smallest possible dose. Severely symptomatic patients with primarily adductor SD had the best results. In a subpopulation of patients, most of whom had been injected unilaterally, outcome was more favorable in those who had been previously injected with BT. One-month post-injection studies reflected significant improvement in measured readings of a standard passage; this was consistent with the subjective and perceptual improvements noted. Peroral injection seems inadvisable in patients with mixed or abductor SD as well as in those patients with an uncontrollable gag reflex.

Acknowledgements

The authors gratefully note the contributions of Nadia Tettlier, Danna Koschkee, and Dawn B. Lowery toward the completion of this project. We would like to thank Tom Cook for statistical advice and Celeste Kirk for her editorial assistance.

References

1. Miller RH, Woodson GE, Jankovic J. Botulinum toxin injection of the vocal fold for spasmodic dysphonia. Arch Otolaryngol Head Neck Surg 1987: 113:603-605.
2. Blitzer A, Brin MF. Laryngeal dystonia: a series with botulinum toxin therapy. Ann Otol Rhinol Laryngol 1991, 100: 85-89.
3. Ford CN, Bless DM, Lowery JD. Indirect laryngoscopic approach for injection of botulinum toxin in spasmodic dysphonia. Otolaryngol Head Neck Surg 1990; 103: 752-758.
4. Rosen M, Malmgren LT, Gacek RR. Three-dimensional computer reconstruction of the distribution of neuromuscular junctions in the thyroarytenoid muscle. Ann Otol Rhinol Laryngol 1983; 92: 424-429.
5. Ford CN. Indirect laryngoscopic approach for injection of botulinum toxin in spasmodic dysphonia. NIH CDC Abstracts 1990; 1: 131-133.
6. Ford CN. A multipurpose laryngeal injector device. Otolaryngol Head Neck Surg 1990:103:135-137.
7. Blitzer A. Botulinum toxin treatment of spasmodic dysphonia: a progress report. [Voice Symposium, New York, 1988; (Abstract)].
8. Ludlow CL. Clinical trials for speech disorders. NIH CDC Abstracts 1990; 1: 125-127.

NCVS Status and Progress Report - 2
April 1992, 167-172

Botulinum Neurotoxin A Injection to the Pharyngeal Constrictor Muscles After Total Laryngectomy

K. Linnea Peterson, M.D.
Henry T. Hoffman, M.D., M.S.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
The National Center for Voice And Speech, The University of Iowa
D. Van Demark, Ph.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Julie M. Barkmeier, M.A.
The National Center for Voice And Speech, The University of Iowa

Abstract

Hypertonicity of the pharyngeal constrictor muscles is considered to be a frequent cause of dysphagia as well as tracheo-esophageal puncture (TEP) speech failure following total laryngectomy. Methods to address this problem include pharyngeal plexus neurectomy at the time of laryngectomy, myotomy, and mechanical hypopharyngeal dilation after laryngectomy. We describe the use of electromyographically (EMG) guided needle electrode injection of Botulinum neurotoxin A (Botox) into the pharyngeal musculature following total laryngectomy in patients who were either not suitable candidates for or refused myotomy or dilation. Technique, dosage and results are described.

Introduction

Botulinum toxin type A has been used clinically since the early 1980's when its usefulness for the treatment of strabismus was first reported (1). Since then many clinical applications for it have been found in the treatment of muscular spasm and hypertonicity. It has been used in the fields of ophthalmology, neurology and otolaryngology in treating such disorders as blepharospasm (2, 3),

torticollis (4), cranial-cervical dystonia (2, 4) and spasmodic dysphonia (5, 6, 7). Botulinum toxin acts at the motor nerve terminals (8) blocking the release of acetylcholine and causing a chemical denervation of the muscle (9). It does not affect the electrical excitability or the conductivity of the nerve or muscle (9). This denervation is temporary and lasts on the order of ten weeks to nine months (3, 4, 6).

Increased pharyngoesophageal tonicity may be a problem after total laryngectomy as there appears to be a disturbance in the relaxation and coordination of the upper esophageal sphincter after laryngectomy (10). These patients often present with symptoms of dysphagia or they may have failed tracheo-esophageal speech (11, 12). Singer and Blom reported a 12 % incidence of TEP speech failure due to pharyngoesophageal spasm in their evaluation of 129 patients (13). To date, treatment options for this hypertonicity include pharyngeal plexus neurectomy at the time of laryngectomy, myotomy and mechanical dilation of the hypopharynx. Patients who are not good surgical candidates for or who refuse surgery may benefit from local injection of Botox in order to weaken the constrictor musculature. We believe this to be the first report of the use of Botox for this condition.

Methods

Patients initially selected for this study had undergone total laryngectomy and had symptoms of hypopharyngeal hypertonicity. They refused myotomy or dilation, were no longer considered good surgical candidates or had previously undergone cricopharyngeal myotomy which had failed. Our protocol for treatment involves titrating the dose for each patient to achieve the desired response. The sites for local injection are chosen to give a distribution of the toxin on one side of the hypopharynx. If the hypoglossal nerve has been sectioned in previous neck dissections or has otherwise been involved in disease, the side of nerve injury is preferred (See Figure 1.). The area overlying the hypopharynx is prepped with alcohol and anesthetized with one percent lidocaine with 1:100,000 epinephrine. The carotid is palpated and drawn laterally away from the injection sites. A 27 gauge teflon coated EMG needle is used for the botulinum toxin injection. The needle is placed into the constrictor musculature with verification of location by a burst of electrical activity upon swallow. The botulinum toxin is then injected. The procedure is repeated for each of two other sites. The initial dose used is five units in 0.2 cc in each of the three sites for a total of 15 units in 0.6 cc. Patients are monitored for effects of the neurotoxin and are reinjected two weeks after the initial dose if the effects are not satisfactory. Should an injection be indicated after two weeks, the dose used for the second injection is 2.5 units in 0.1 cc in each of the three sites for a total of 7.5 units in 0.3 cc. Subsequent doses used are the same as the last successful injection. If reinjection is necessary after two weeks, these doses are additive; that is the total dose of the two injections is the dose used for the next procedure.

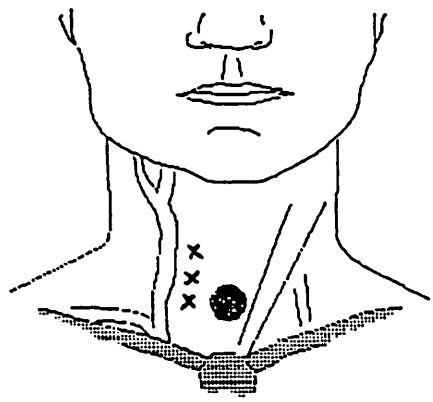

Figure 1: This figure shows the location of the three sites of injection.

Results

Case 1

This 39 year old male underwent total laryngectomy, right radical neck dissection, and left supraomohyoid neck dissection in August 1989 for a $T_4N_{3b}M_0$ squamous cell carcinoma, which was additionally treated with postoperative radiotherapy. In September of 1990 a left neck recurrence was treated with a left modified neck dissection, sparing cranial nerve XI. In April of 1991 he had a TEP created, but he failed TEP speech with the presumptive diagnosis of hypertonicity of the cricopharyngeal constrictor musculature. He was able to speak only with an electrolarynx or with the use of a catheter to produce air flow. On May 16, 1991 he underwent a trial of lidocaine injected into the hypopharnygeal musculature with positive results, substantiating the diagnosis of hypopharyngeal hypertonicity. On June 4, 1991 he received a Botox injection into the hypopharyngeal musculature: a total dose of 15 units injected into three separate sites. He was able to produce speech within a few days after injection. Two months later he had improved ease of phonation as well as increased volume and intelligibility. He was able to achieve a maximum phonation time of nineteen seconds, sustaining "ah". His rate of speech averaged 200 words per minute. Three months after injection he continued to be able to phonate with ease.

Case 2

This 55 year old male was treated with a supraglottic laryngectomy in June of 1986 for a squamous cell carcinoma of the epiglottis. In March of 1987 he developed a recurrence at the left base of tongue and underwent radiation therapy to the hypopharynx and neck after declining surgical treatment. In May of 1988 he was noted to have an ulcerative lesion of the left hypopharynx which was found to be poorly differentiated squamous cell carcinoma on biopsy. He then underwent completion laryngectomy, resection of squamous cell carcinoma of the left pharyngeal wall, left radical neck dissection, left thyroid lobectomy, and left pectoralis major myocutaneous flap reconstruction on June 10, 1988. In October of 1988 he had a tracheoesophageal puncture created. He had difficulty swallowing and controlling secretions and also failed TEP speech because of presumptive hypertonicity of the constrictor musculature. In 1990 he lost 20 pounds as his dysphagia worsened. A modified barium swallow at this time showed a gradually releasing constriction of the lower hypopharynx. His weight stabilized with dietary supplementation, but difficulty controlling his secretions continued to be a primary complaint. On June 25, 1991 he was injected with 15 units of Botox into the right hypopharyngeal musculature in a volume of 0.6 cc divided into three sites. He noted only mild to moderate improvement in control of his secretions and therefore presented for a second treatment two weeks later. At this time he was injected with 2.5 units in each of three sites in the right hypopharyngeal muscle for a total dose of 7.5 units in 0.3 cc. He did well with this injection and reported good control of secretions as well as improvement in his TEP speech. This lasted approximately 2.5 months, and he returned on October 11 for reinjection. At this point he was again having difficulty with control of his secretions. He underwent injection of Botox with a dose of 7.5 units in 0.3 cc in each of three sites for a total dose of 22.5 units. He reported satisfactory results with improved control of his secretions and greater ease of phonation, although he continued to use TEP speech on only a limited basis.

Case 3

This 73 year old woman was treated with a total laryngectomy in March of 1990 for chronic aspiration. She had a TEP created which was subsequently closed in October of 1990 because of leakage about the prosthesis. She underwent speech therapy at this time with little success. She had difficulty swallowing secondary to presumptive cricopharyngeal spasm and underwent a cricopharyngeal myotomy in March of 1991. She related that her swallowing improved for two months after which her symptoms recurred. On August 12, 1991 she presented with a two day history of being unable to swallow solids or liquids and was found to have a hypopharyngeal food bolus impaction. This was removed and a modified barium swallow at this time showed poor tongue control with a narrowing of the distal hypopharynx with build up of both liquids and solids. The constriction gradually released, substantiating the impression of cricopharyngeal hypertonicity. On August 20 the patient underwent Botox injection to the hypopharyngeal musculature under EMG guidance. She was injected with 5 units in 0.2 cc in each of three sites for a total dose of 15 units. The hypopharynx deviated to the left and the injection was made in the midline, into the right hypopharyngeal musculature. Two weeks later she reported some improvement in swallow, but continued to have difficulty and presented for reinjection. At this time she was injected with 2.5 units in 0.1 cc for a total dose of 7.5 units. She presented for follow up three weeks later with marked improvement. At this time she reported a four pound weight gain and was able to eat a variety of foods. She stated that now she could "swallow better on the right." Modified barium swallow at this time showed the pharyngoesophageal constriction having only mild effect on the swallowing reflex. She was able to swallow a range of consistencies from liquid to paste with cracker. She presented again for injection 2.5 months later and was injected with a dose of 7.5 units in each of three sites for a total dose of 22.5 units. She continues to do well one month later.

Case 4

This 62 year old woman was treated with a total laryngectomy, left neck dissection sparing cranial nerve eleven in March of 1987 for a $T_3N_1M_0$ squamous cell carcinoma of the glottis. She had a TEP created in September of that year. She had difficulty with speech from that time on, although she was able to talk on the phone and in small groups. She had difficulty sustaining phonation and spoke only with great effort, her speech interrupted with intermittent breaks. She was able to phonate at an average rate of 100 words per minute. On August 20, 1991 she underwent Botox injection to the left hypopharyngeal musculature in three sites. She received a dose of 5 units in 0.2 cc in each site for a total dose of 15 units. She reported improvement in her speech within the first few days and was able to phonate with much greater ease. On return two weeks later she was noted to have increased volume and intelligibility. Her speech was fluent without breaks. Her average rate of speech at this time was 160 words per minute. She continues to do well four months after her injection.

Discussion/Conclusions

Total laryngectomy interrupts the normal pattern of both speech and swallowing (10, 12). Patients learn to communicate using alternative modes, including esophageal speech or tracheoesophageal puncture speech. The incidence of dysphagia after total laryngectomy ranges from 10 to 58 percent in different studies (15, 16). Pharyngeal muscle hypertonicity may be a cause of dysphagia and/or failure of tracheoesophageal puncture speech following total laryngectomy

(11, 12). Some surgeons prefer to do an elective myotomy at the time of laryngectomy in order to minimize the risk of this (13, 14), however elective myotomies are not universally done, nor are all myotomies successful. In addition, the prediction of dysphagia and/or tracheoesophageal puncture speech failure is difficult. Some authors advocate esophageal insufflation prior to TEP (17), however this does not predict all failures, nor does an abnormal result preclude success. At this time there are several different methods to address the problem of muscle spasm secondary to total laryngectomy, all surgical. We believe that this represents the first report of the use of the neurotoxin Botox for the treatment of pharyngoesophageal muscle hypertonicity after laryngectomy.

Patients selected for treatment in this study were those who had symptoms of presumptive hypopharyngeal muscle hypertonicity which was substantiated by a lidocaine trial, evidence of spasm on barium swallow, or in one case, by the symptoms of speech difficulty alone. Improvement in speech and/or swallowing following injection was reported in all patients. Titration of dose with reinjection was necessary in some patients, although relief of symptoms of hypertonicity was subsequently obtained in all. One patient complained primarily of dysphagia. Two patients had difficulty with tracheoesophageal speech and one had mixed symptoms involving speech and swallow. Patient number three, complaining primarily of dysphagia, reported a four pound weight gain in the first three weeks after injection. Barium swallow showed improvement with handling a range of food consistencies. The other three patients all had improvement as determined by the patients own perception as well as a speech pathologist's assessment of speech. Two patients (patients one and two) were unable to use their TEP for speech prior to the Botox injection and were able to speak with relative ease within days. The second patient complained primarily of difficulty controlling his secretions and was pleased with the results of the injection. He continued to use TEP speech on only a limited basis, but was able to phonate without significant difficulty following injection. The fourth patient had limited speech, but subsequent to the injection had increased volume of speech, improved intelligibility, and an increased rate of speech. There were no complications. Results at this stage are preliminary, however, the consistency and extent of response are significant. The results obtained in our trial suggest Botox to be a promising treatment modality for patients with symptoms related to increased pharyngoesophageal tone.

References

1 Scott, A.B.: Botulinum Toxin Injection of Eye Muscles to Correct Strabismus. Tr Am Ophth Soc, LXXIX :735-770, 1981.

2. Jankovic, J., Ford, J.: Blepharospasm and Orofacial-Cervical Dystonia: Clinical and Pharmacological Findings in 100 Patients. Ann Neurol, 13:402-411, 1983.

3. Frueh, B.R., Felt, D.P., Wojno, T.H., Musch, D.C.: Treatment of Blepharospasm with Botulinum Toxin: A Preliminary Report. Arch Ophthalmol, 102:1464-1468, 1984.

4. Jankovic, J., Orman, J.: Botulinum A Toxin for Cranial-Cervical Dystonia. Neurology, 37:616-623, 1987.

5. Miller, R.H., Woodson, G.E., Jankovic, J.: Botulinum Toxin Injection of the Vocal Fold for Spasmodic Dysphonia: A Preliminary Report. Arch Otolaryngol Head and Neck Surg, 113:603-605, 1987.

6. Blitzer, A., Brin, M.F., Fahn, S., Lovelace, R.E.: Localized Injections of Botulinum Toxin for the Treatment of Focal Laryngeal Dystonia (Spastic Dysphonia). Laryngoscope, 98: 193-197,1988.

7 Ludlow, C.L., Nauton, R.F., Sedory, S.E., Schultz, G.M., Hallett, M.: Effects of Botulinum Toxin Injections on Speech in Adductor Spasmodic Dysphonia. Neurology, 38: 1220-1225, 1988.
8 Dolly, J.O., Black, J., Williams, R.S., Melling, J.: Acceptors for Botulinum Neurotoxin Reside on Motor Nerve Terminals and Mediate its Internalization. Nature, 307:2, 457-460, 1984.
9 Sellin, L.C.: The Action of Botulinum Toxin at the Neuromuscular Junction. Medical Biology, 59: 11-20, 1981.
10. Duranceau, A., Jamieson, G., Hurwitz, A.L., Jones, R.S., Postlethwait, R.W.: Alteration in Esophageal Motility After Laryngectomy. Am J Surg, 131: 30-35, 1976.
11. Singer, M.I., Blom, E.D.: Selective Myotomy for Voice Restoration After Total Laryngectomy. Arch Oto, 107:670-673, 1981.
12. McConnell, F.M.S., Cerenko, D., Mendelsohn, M.S.: Dysphagia After Total Laryngectomy. Otolaryngologic Clinics of North America, 21:4, 1988.
13. Hamaker, R.C., Singer, M.I., Blom, E.D., Daniels, H.A.: Primary Voice Restoration at Laryngectomy. Arch Oto, 111: 182-186, 1985.
14. Wenig, B.L., Mullooly, V., Levy, J., Abramson, A.L.: Voice Restoration Following Laryngectomy: The Role of Primary Versus Secondary Tracheoesophageal Puncture. Ann Otol Rhinol Laryngol, 98: 70-73, 1989.
15. Balfe, D.M., Koehler, R.E., Setzen, M., et al.: Barium Examination of the Esophagus After Total Laryngectomy. Radiology, 143:501-508, 1982.
16. Nayar, R.C., Sharma, V.P., Arora, M.M.L.: A Study of the Pharynx After Laryngectomy. J Laryngol Otol, 98:807-810, 1984.
17. Blom, E.D., Singer, M.I., Hamaker, R.C.: An Improved Esophageal Insufflation Test. Arch Otolaryngol, 111:211-212, 1985.

Acknowledgements

This work was supported, in part, by grant No. P60 DC 00976 from the National Institute on Deafness and Other Communication Disorders.

NCVS Status and Progress Report - 2
April 1992, 173-181

Electromyographic Assessment in Spastic Dysphonia Prior to Vocalis Injection With Botulinum Toxin

Arthur A. Rodriquez, M.D.
Charles N. Ford, M.D.
Diane M. Bless, Ph.D.
Robert L. Harmon, M.D
University of Wisconsin Medical School

Abstract

Electromyographic (EMG) assessment of involuntary spasm activity in the cricothyroid (CT) and vocalis (V) muscles was correlated to clinical voice measures in 32 patients with clinically diagnosed spastic dysphonia (SD). Subjective voice rating, and quantified fluency and laryngeal diodochokinesis measures were obtained pre botulinum toxin injection (Botox). Pre Botox EMG was performed using a monopolar needle electrode. Each muscle was sequentially examined at rest, during vocal click, scale, sustained "E" at different pitches, and repeated "E" voicings. A three point EMG severity scale was used to grade each muscle. There were no significant correlations between the summed EMG severity scales for all 4 muscles, the combined EMG severity for the V of CT muscles, or the combined EMG severity for the right or left CT and V muscles and pre-Botox subjective rating, fluency, repeated sound production or post-botox change in subjective rating. EMG did discriminate between adductor and abductor patterns, but could not correctly classify mixed SD. Those with adductor SD displayed involuntary "spasm" activity in the V and CT muscles, while those with abductor SD displayed more "spasm" activity in the CT than the V muscles. Sequential EMG assessment of CT and involuntary activity in SD does not predict clinical severity of outcome following Botox injection of the V muscle.

Introduction

Spastic (spasmodic) dysphonia (SD) is a disorder of speech comprising symptoms believed related to adductor and/or abductor spasms of the vocal folds[1,2]. Electromyographic assessment has shown no evidence of a lower motor neuron disorder.[2,3] Most patients display spontaneous uncontrolled motor unit activity or "spasms" present in one or more vocal fold muscles having the characteristics of a dystonia.[1,3,4] Of those given a presumptive diagnosis of SD based on voice quality alone, some patients have tremor, myoclonus, or pyramidal and extrapyramidal disease.[1] The Vocalis (thyroarytenoid) and Cricothyroid muscles are technically the easiest to sample electromyographically, and in SD show varying degrees of "spasm" activity, both between the two muscles and from side to side.[3] The "spasms" interfere with the performance of discrete bursts of electromyographic activity normally associated with various vocal maneuvers.[3] Treatment has consisted of inducing weakness in the vocal fold musculature by neurectomy of the recurrent laryngeal nerve and, more recently, by injection of botulinum toxin into laryngeal muscles.[5-9]

The purpose of this study was to determine whether electromyographic assessment of the cricothyroid (CT) and vocalis (V) muscles in SD patients predict clinical type (adductor, mixed, or abductor), clinical severity, or response to botulinum toxin injection of the vocalis muscle.

Methods

Electromyographic activity in the bilateral CT and V muscles of 32 patients with clinically diagnosed SD was determined using a monopolar needle electrode with sequential sampling of each muscle. There were 8 males and 24 females. The mean age $\pm$ standard deviation was 53.2 $\pm$ 16.9 years. A clinical diagnosis of adductor, abductor, or mixed SD was made based on perceptual symptoms. All patients included in this study were independently evaluated by a speech pathologist and an otolaryngologist with 10 or more years of experience specializing in voice disorders. A check list of perceptual symptoms described by Izdebski and Dedo was used to note the presence or absence of vocal characteristics typically associated with abductor and adductor SD: patients exhibiting a strained hoarseness, strangled voice, glottal fry, hard glottal initiation or vocal tension were classified in the adductor SD group; patients with intermittent aphonia, whispered speech, asthenic voice quality, and voiceless vowels following voiceless consonants were classified into the abductor group.[10] Patients with both adductor and abductor characteristics were considered mixed--unless further testing revealed one set of characteristics to be compensatory for the primary valving difficulty.

Details of the technique have been published elsewhere.[11] Percutaneous placement of the 50mm monopolar electrode was performed without anesthesia and without complications. Needle placement is shown in fig. 1. To sample the CT muscle the electrode was inserted into the skin in the midline over the cricoid cartilage and aimed toward the inferior cornu of the thyroid cartilage (laterally and superiorly). Proper placement was confirmed by increasing recruitment during performance of a scale while vocalizing "E". Improper placement results in activation with neck flexion suggesting that the electrode has entered the strap musculature. To sample the V muscle, the needle was inserted in the same site as for the CT, however the needle is advanced posteriorly at a 45 degree angle with the vertical and slightly lateral through the CT muscle and the cricothyroid membrane, and then more superiorly at a 60 degree angle with the vertical while the subject vocalizes a sustained "E" to adduct the cords. Entrance into the V muscle is heralded by characteristic motor units maximally activated by a vocal click or valsalva. The vocal click is performed by approximat-

ing the arytenoids while exhaling a gently burst of air through the open mouth, thereby rapidly abducting and adducting the folds without causing vibration. During insertion, the needle is maintained in a submucous location. If the needle enters the subglottic space, noise is recorded on the electromyograph and coughing will often occur, necessitating removal of the needle.

The EMG signal was simultaneously analyzed through an electromyograph[a], and a root mean square processor[a] with display of the integrated signal (RMS-EMG) on a strip chart recorder.[b] Filter settings were 30 Hz to 8 KHz. Motor units and the interference pattern were examined at 200 uv/div., and 50 uv/div. was used to examine for spontaneous activity. A sweep speed of 10 ms/div. was used for the electromyograph display. Each muscle was examined at rest, during vocal click, during sustained "E" at different pitches, during repeated vocalization of "E", and with "E" vocalized in a scale. A three point severity scale of involuntary RMS-EMG activity (spasm) was used to grade each muscle. The absence of spasms was graded as 0, spasms in 1-3 maneuvers was graded as 1, and spasm in more than 3 maneuvers was graded as 2. An example of the RMS-EMG recording of the vocalis of a normal subject and a SD subject with spasm activity during two maneuvers is shown in figure 2.

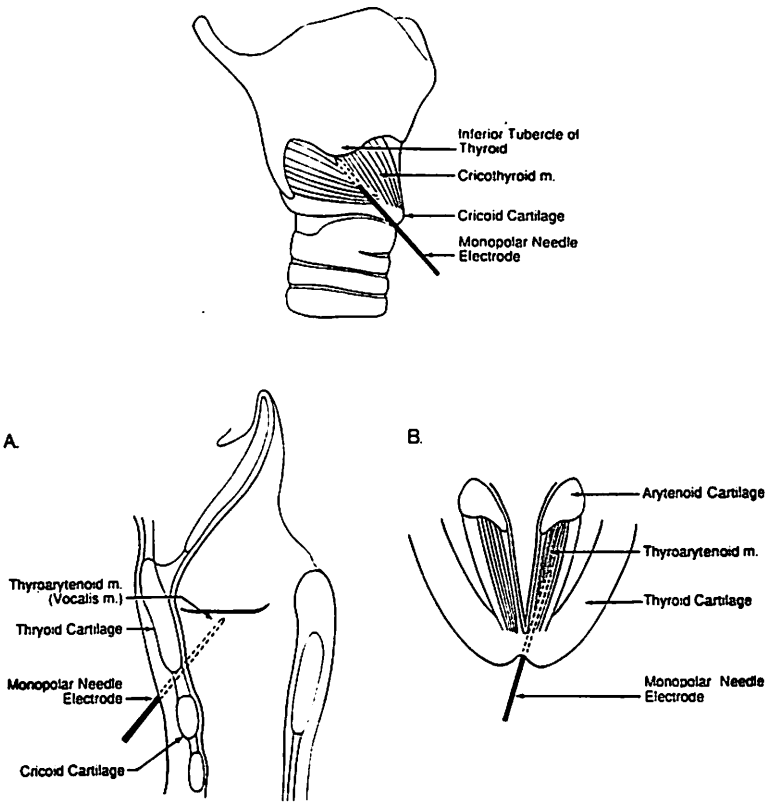

Figure 1. Technique for percutaneous electromyographic sampling of the cricothyroid and vocalis muscles.

Pearson correlations were determined between EMG severity parameters and assessments of fluency (% correct portions of a read passage), vocal fold diadodyskinesia (number of vocalizations per second), and pre-injection/post-injection change in a 1-5 subjective voice scale. Discriminant analysis was performed to assess involuntary CT and V muscle activity in predicting SD clinical type (adductor, abductor, or mixed) using Wilk's lambda. Statistical significance was accepted at $p<.05$. EMG severity parameters included each muscle's severity score, summated CT (SCT) and V (SV) scores, the sum of all severity scales for V and CT muscles (EMG SEV), summated right (R), and left (L) CT and V scores, and difference between the SCT and SV scores (SDIFF). An analysis of variance was used to determine whether there was a difference between mean SCT and SV scores across SD clinical types.

Each subject underwent injection of both V muscles following EMG and voice assessments with botulinum toxin.

[ab] Suppliers were Tracor Northern, Inc., 2551 West Beltline Highway, Middleton, WI, and Gould TA 550 Strip Chart Recorder, Gould Inc., 35129 Curtic Boulevard, Eastlake, OH.

Constant "E"

Modest increase in activity over baseline noted with significantly less than full activation.

Repeated "E"

Bursts of activity consisting of less than a full interference pattern alternating with relative inactivity between the bursts.

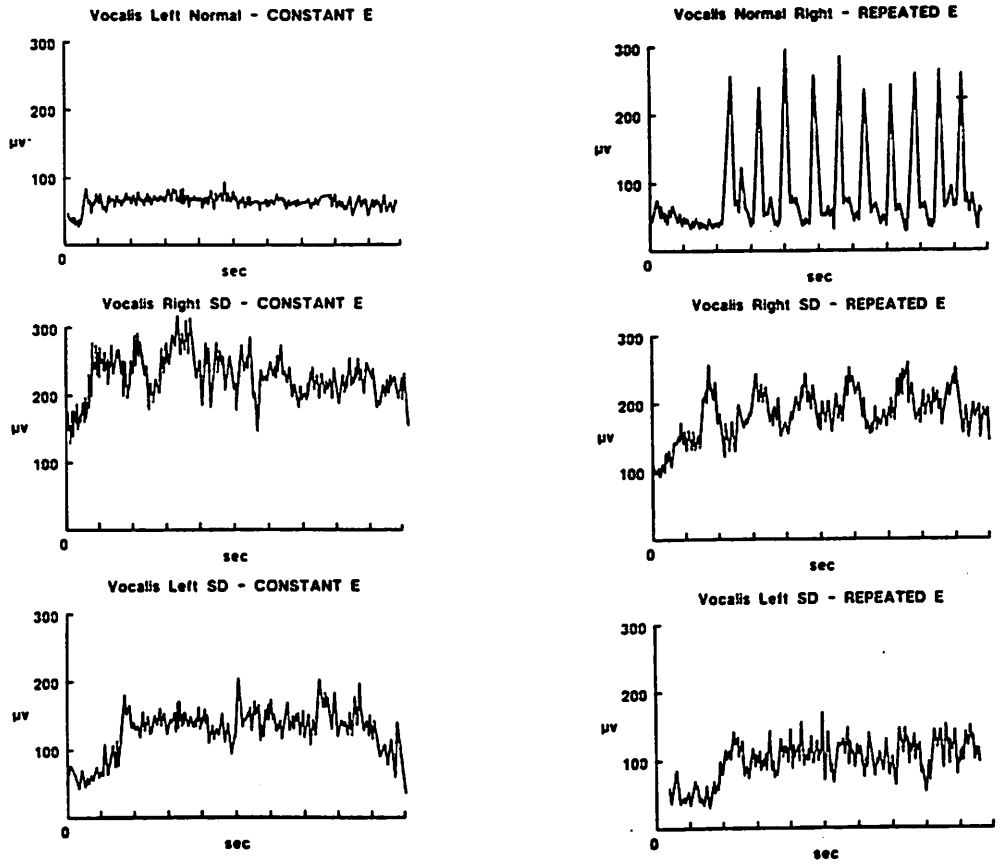

Figure 2. Example of Normal and Spastic Dysphonia RMS-EMG activity patterns during two vocal maneuvers in the vocalis.

Results

Normative motor unit action potential (MUAP) parameter data from our laboratory has been reported previously.[3] Mean MUAP amplitude± SD, and duration± SD, was 426± 194, and 3.5± 1.0 for the V and 500± 224, and 4.4± 1.6 for the CT. MUAP parameters evaluated by inspection of the electromyographic interference pattern during period of low activation were qualitatively judged to be within normal limits in all subjects. Spontaneous fibrillations or positive sharp waves were not observed.

Analysis of the RMS-EMG strip chart recordings revealed that six subjects (18.7%) had no identifiable spasms during the EMG examination. There were no significant correlations between

SCT, SV, SDIFF, R, L, EMG SEV, and fluency, vocal fold diodochokinesia, pre injection voice assessment or change in subjective voice after botulinum toxin injection.

SV, SDIFF, EMG SEV, and RV correctly predicted abductor and adductor SD ($p<.05$) (see table 1). EMG parameters were poor at discriminating mixed SD. Vocalis severity scores were higher in those with adductor SD in comparison to those with abductor SD. CT severity was greater than V activity in those with abductor SD. (see table 2)

Table 1. Discriminant analysis for variables found to be statistically significant (Wilk's lambda, $p<0.05$) in predicting clinical spastic dysphonia

Variable	SD Type	Predicted/Actual Number	% of actual SD Type
R V	1	16//21	76
	2	4/4	0
	3	0/7	0
S V	1	12/21	57
	2	4/4	100
	3	2/7	28
S DIFF	1	17/21	81
	2	3/4	75
	3	0/7	0
EMG SEV	1	12/21	57
	2	3/4	75
	3	2/7	28

(1=adductor, 2=abductor, 3=mixed, RV=right V score, SV=Σbilateral V score, S DIFF=Σbilateral CT scores–Σbilateral V scores, EMG SEV=Σbilateral CT and V scores.

Table 2. RMS-EMG summated severity scores for vocalis and cricothyroid muscles. The SV score for the abductor SD group is significantly less than the SV score for the adductor SD group ($p<.05$). There is no difference between the CT scores ($p>.05$).

EMG Summated Severity Scores for Vocalis and Cricothyroid Muscles

	Adductor SD	Abductor SD	Mixed SD
S V	$1.43 \pm .87$	0	$.8 \pm .9$
SCT	1.33 ± 1.1	$1 \pm .81$	1.43 ± 1.4

Discussion

The absence of evidence for a lower motor neuron lesion for the SD subjects in this study confirms previous reports.[1,3] None of the laryngeal muscles sampled in this study showed evidence of spontaneous fibrillations or positive sharp waves, abnormal motor unit action potential parameters, or diminished recruitment.

Vocal symptoms in patients with SD vary over time and are influenced by factors such as performance anxiety, speech context, or task specificity.[2] In our study, six subject (18.7%) had no "spasm" activity at the time they were sampled. Because a particular muscle can be monitored with this technique for only a minute or two, it seems likely that dystonic activity may be missed in some subjects, who do in fact have spasms, particularly when sustained vowels constitute the test condition. Spasm severity might also vary, leading to an underestimate or overestimate of the involuntary activity in a particular muscle. Both the failure to observe dystonic activity and the failure to grade its severity in a functional context would affect the correlation between electrodiagnostic and clinical assessments of severity. This may explain the poor correlation between involuntary activity and clinical severity or outcome after botulinum toxin injection. The sequential nature of our evaluation made it impossible to document simultaneous spasm activity in different muscles. Cocontraction during dystonic activity may be a significant factor in determining symptoms which we did not assess. A potential solution to the problems inherent in our technique would be the use of simultaneous recording from multiple muscles using fine wire electrodes. Watson and associates used fine wire hooked electrodes in the thyroarytenoids of 5 patients (two adductor and three abductor types) and additionally in the posterior cricoarytenoid of one of the patient with abductor SD.[12] They failed to find a consistent pattern of thyroarytenoid activity differences between the adductor and abductor groups. Intersubject variability in the level of V activity as a function of task was high in both

symptom groups and some subjects with abductor SD had high levels of V activity. Data from the single abductor SD subject showed inappropriately high levels of posterior cricoarytenoid activity regardless of vocal task.[12] Our finding suggest that summated vocalis "spasm" activity tends to be higher in adductor compared to abductor SD patients, but is not different in the group with mixed SD. The differences between the studies may be due to the number of subjects used, the difference in recording electrodes, the method of assessing spasm activity, or in differences in clinically categorizing abductor and mixed SD. In our study, the monopolar needle electrode was moved from time to time within the muscle to relocate it after swallowing dislodged the electrode from its location within the muscle. Additionally, we required the performance of a greater number of vocal tasks, especially requiring phonation of a scale, and repeated "E", instead of only a sustained "E", as Watson et al. required. Any spasm occurring during the production of the vocal task would have been graded in our system. Indeed, it was common to observe dystonic activity only at one point in time during the production of a particular vocal maneuver.

Our results support the notion that the voice deficit in SD is related to the particular muscle patterns which happen to have the most involuntary spasm activity. Although spasm activity was seen in the CT muscles in each group, those subjects with abductor SD did not have as much abnormal activity in the V muscle. These findings may help explain why some patients respond more favorably to botulinum toxin injections in the CT than the V. They also suggest that in abductor and adductor classified patients, electromyographic findings might be used to determine which muscles to inject for relief of symptoms. Those with mixed SD could not be classified based on summated spasm activity. Thus, those individuals who do not have clear diagnostic perceptual symptoms also do not have clearly different electromyographic findings. Presumably, the vocal fold abductor (posterior cricoarytenoid) would have demonstrated spasm activity. Unfortunately, we did not routinely sample this muscle because of the technical difficulties involved in needle insertion.

Conclusions

Sequential EMG assessment of CT and V involuntary activity in SD patients does not predict clinical severity or outcome following botulinum toxin injection.

Sequential EMG assessment of CT and V involuntary activity can discriminate between clinical abductor and adductor patterns but not correctly classify mixed SD.

Those with adductor SD displayed involuntary "spasm" activity in the V and CT muscles, while those with abductor SD displayed relatively more "spasm" activity in the CT than the V muscles.

Acknowledgements

The authors wish to thank Todd Franke for the statistical analysis and Eileen Knudtson for technical assistance.

Bibliography

1. Blitzer A, Lovelace RE, Brin MF, Fahn S, Find ME. Electromyographic findings in focal laryngeal dystonia (spastic dysphonia). Ann Otol Laryngol 1985;94:591-594.

2. Shipp T, Izdebski K, Reed C, Morrissey P. Intrinsic laryngeal muscle activity in a spastic dysphonia patient. Journal of Speech and Hearing Disorders 1984;50:54-59.
3. Rodriquez AA. Electromyographic finding in the vocalis and cricothyroid musculature in spastic dysphonia. Muscle and Nerve 1989;12:769 (abstract)
4. Cohen LG, Hallett M, Panizza M. Dystonia. AAEE course D: clinical neurophysiology of movement disorders. AAEE eleventh annual continuing education course, 1988.
5. Blitzer A, Brin MF, Fahn S, Lovelace RB. Localized injections of botulinum toxin for the treatment of focal laryngeal dystonia (spastic dysphonia). Laryngoscope 1988;98:193-197.
6. Miller RH, Woodson GE, Jankovic JJ. Botulinum toxin injection of the vocal fold for spasmodic dysphonia. A preliminary report. Arch Otolaryngol Head Neck Surg 1987;113:603-605.
7. Jankovic J, Brin MF. Therapeutic uses of botulinum toxin. New England J Med 1991;324:1186-1194.
8. Ludlow CL. Treatment of speech and voice disorders with botulinum toxin. JAMA 1990;264:2671-2675
9. Ludlow CL, Naunton RF, Sedory SE, Schulz GM, Hallett M. Effects of botulinum toxin injections on speech in adductor spasmodic dysphonia. Neurology 1988; 38:1220-1225.
10. Izdebski K, Dedo HH. Spastic dysphonia. In: Darby JK (Ed) Speech evaluation in medicine. New York, NY: Grune and Stratton, 1981:105-27.
11. Rodriquez AA, Myers BR. Laryngeal electromyography in the diagnosis of laryngeal nerve injuries. Archives of Physical Medicine and Rehabilitation 1990;71:587-590.
12. Watson BC, Schaefer SD, Freeman FJ, Dembowski J, Dondraske G, Pearle R. Laryngeal electromyographic activity in adductor and abductor spasmodic dysphonia. J Speech Hear Res 1991;34:473-482.

Part II

Summaries and tutorial reports of Training, Dissemination and Continuing Education

NCVS Status and Progress Report - 2
April 1992, 181-182

Training Update

John Folkins, Ph.D., Training Coordinator

Department of Speech Pathology and Audiology, The University of Iowa

The training program of the National Center for Voice and Speech (NCVS) provides educational opportunities for predoctoral fellows, postdoctoral fellows, and visiting scholars. The educational philosophy of the NCVS puts a high priority on open-minded scholarly inquiry and independent learning skills. All NCVS traineeships are structured in a manner flexible enough to accommodate the research training needs of students with very different goals and backgrounds.

The traineeships involve supervised research experiences at all stages of training. These research projects are most often done with NCVS investigators as mentors. Some projects are done in collaboration with other NCVS trainees as well as investigators. NCVS members are committed to accepting trainees with a large range of different backgrounds and introducing them to research in voice and speech.

All trainees are encouraged to do research with scientists having backgrounds from a number of different disciplines. The voice and speech training program is integrated with other areas of science; for example, we cut across research areas often associated with physiology, psychology, hydraulic engineering, otolaryngology, and music.

Our predoctoral training, postdoctoral training, and visiting scholar positions have been developed according to the plan presented in our last report. As explained in the NCVS Status and Progress Report Volume 1, page 122:

> Predoctoral trainees are anchored in the doctoral degree program in the Department of Speech Pathology and Audiology, located in the Wendell Johnson Speech and Hearing Center at The University of Iowa. All of the proposed training programs are expected to span a number of academic disciplines and take full advantage of all consortium sites.
>
> Each predoctoral student is required to develop and defend a plan of research experiences and coursework that will meet the competencies required for his or her research goals. Competencies in four tool areas are required: research design, analog instrumentation, computers, and mathematics.

Each doctoral student in the Department of Speech Pathology and Audiology is reviewed yearly by the faculty. The review includes student performance in research projects, coursework, and clinical activities. The outcome of the review is reported to the student by his or her advisor. To be considered as a candidate for an NCVS postdoctoral fellowship, the applicant must have either successfully completed a Ph.D, MD., or the equivalent. The consortium allows the postdoctoral fellows to use the clinical resources of the University of Iowa, University of Wisconsin-Madison, and University of Utah.

Like the predoctoral trainee, the postdoctoral fellow will meet with a group of five faculty members--chosen by the trainee in consultation with the faculty advisor--within the first few months of training. Prior to this meeting the trainee and advisor work out a specific statement of career goals; a description of competencies brought into the program; and a list of needed competencies, experiences and skills to prepare optimally for the stated career goals. The committee discusses these materials, makes alterations as appropriate, and approves a plan of study. When possible, planning meetings for postdoctoral trainees who are located at non-Iowa sites will be held during the regular yearly NCVS conferences.

At the end of the trainee's research experience, he or she will meet with the same five training faculty members that were involved with the planning meeting. The training experiences will be critiqued by all members of the group. A written report summarizing the meeting will be prepared by the trainee's mentor and submitted to the training director.

Visiting scholars may train at the University of Iowa, Denver Center for the Performing Arts, University of Wisconsin-Madison, or University of Utah. Training periods may vary from two week to six months. Visiting scholars are established researchers for whom the training visits are expected to introduce new scientific areas, allow them to acquire new research skills or update and enhance existing research skills. Just as importantly, visiting scholars will bring new ideas and perspectives to NCVS investigators on a continuing basis.

Because of the focus of the visiting scholarship is the mature scientist, planning meetings are deemed inappropriate. However, visiting scholars are expected to give presentations covering their planned research projects. Such seminars are open to all NCVS investigators and mentors and a format encouraging discussion is encouraged.

The beginning of the first grant year occurred during the middle of the fall semester. However, even at that time we were able to appoint one predoctoral fellow, Julie Barkmeier.

The primary activity for the first grant year was to search for the highest quality trainees and to bring such trainees in from a variety of disciplines. We feel that we were quite successful. For summer and fall of 1991, in the latter part of the first grant year, we added four additional predoctoral trainees: Kenneth Tom, Alice Smith, Brad Story, and Ann Fennell. We added two postdoctoral trainees: David Berry and Karen Linnea Peterson; and one visiting scholar, Carlin Hageman.

NCVS Status and Progress Report - 2
April 1992, 183-185

Dissemination Update

Ronald Scherer, Ph.D., Dissemination Coordinator
Recording and Research Center, The Denver Center for the Performing Arts

Verbal communication depends upon the ability to convey intended ideas and emotions. Disorders of voice and speech often prevent effective and accurate communication, and lack of adequate training of voice and speech skills limits everyday persuasiveness and the level of a culture's performance artistry. When people are better informed about how voice and speech are produced, how good health of the voice and speech organs are maintained, and how the warning signs of communication disorders are discerned, more satisfactory lives can be achieved.

The intent of the Dissemination program of the NCVS is to increase the public's awareness of voice and speech health and behavior, especially for those at vocal risk, so that more informed choices can be made to enhance communication effectiveness and avoid speech and voice problems.

A number of dissemination activities promote this goal of informing the public about speech and voice. These activities include general information projects, radio and television spots, newspaper and magazine articles, conference support, videotapes, and other projects.

General Information Activities

The National Center for Voice and Speech general flyer has been developed and is used to inform the public and voice professionals about the activities and purposes of the NCVS, especially to attract young scientists to the training program.

A Voice Card, a conveniently small plastic card with reminders of helpful voice care and use, has been developed and is in production and testing phases. The card will be available to the public as a reminder of basic voice tips to help them through the day.

During the past nine months, members of the NCVS have given more than 50 talks to over 5,000 audience members on the function and dysfunction of voice and speech. These have included talks and workshops to high school and college students and teachers, actors, singers and singing teachers, lawyers, broadcasters, cheerleaders, patients, the general public and other voice professionals.

Dr. Ingo Titze presented a lecturer recital entitled, "Voices of People and Machines," to an audience of 350 in Iowa City April 6. The lecture described how the human voice works and what researchers can learn about it by simulating it with computers.

The city of Denver has many visiting stage shows and artists. One of the nation's most active promoters is Robert Garner of The Denver Center for the Performing Arts who brings in most of the current touring broadway shows such as "Phantom" and "Cats." A voice care pamphlet on "Voice Care in the Mile-High City," has been developed and is sent to all members of the visiting shows with their pre-Denver materials. Included with this helpful information is an invitation to participate in a lecture/discussion of the voice and voice care when they are in Denver.

Similarly, the members of the Denver Center Theatre Company, a nationally prominent resident theatre company, also receive health care information included in their company handbooks.

These health care materials, giving specific advice on health problems in the dry and high Denver region and pertinent referral information, are being reviewed as models for similar materials to be developed for other regions of the country and other professional voice and speech organizations.

Newspaper Articles

A number of newspaper articles were written about Dr. Titze's research with computer voice simulation. Blurbs about the Iowa City debut of the singing computer appeared in The Des Moines Register and USA Today. In-depth articles describing voice function and the simulation of voice were written in the "Iowa City Press-Citizen," "The Daily Iowan, (Iowa City)," "The (Cedar Rapids, Iowa) Gazette," "Chicago Tribune," and two Italian newspapers. These pieces reached more than three million readers. Additionally, the "Chicago Tribune" story was offered to all 330 newspapers in the Knight-Ridder chain, with a potential audience of 72 million.

Radio and Television Spots

Dr. Steven Gray participated in a CNN radio and television spot from The University of Utah that dealt with work disability due to injury of the voice. This spot was broadcast nationally.

Dr. Titze was heard on National Public Radio's "Weekend Edition" explaining computer voice simulation; this program had approximately 1.7 million listeners. Eleven other radio stations also featured Dr. Titze and his research with simulated voice. These stations broadcast from Des Moines, Iowa; Monterey, California; Chicago, Illinois; Iowa City, Iowa; Madison, Wisconsin; Portland, Maine; Providence, Rhode Island; San Jose, California; and Montreal, Quebec, Canada. More than 4.6 million people heard these broadcasts.

Finally, Dr. Titze has been interviewed via telephone by an Italian television station. In this interview, Dr. Titze was asked about voice simulation and its research significance. His comments are scheduled to appear in the documentary program, "TG-sette."

Magazine Articles

"Health Magazine" will publish a short article on voice function (Concerning the Adam's apple and voice pitch) in the May-June, 1992, issue. Circulation is 750,000. Various members of the NCVS contributed information to this article, and the NCVS will be mentioned as a resource for further information.

"The World & I" magazine published a 3,000-word article by Dr. Ingo Titze entitled "New Dimensions in Simulated Voice." The piece discussed the interaction of the human voice and machine replication, and featured various groups within the NCVS and other voice specialists. Circula-

tion is 10,000; its intended audience is the academic community, as well as professionals and government leaders.

The May 6, 1992, issue of "The Chronicle of Higher Education" published a 200-word report about Dr. Titze's work with voice synthesis and its value in studying the physics of human sound. This publication has a circulation of more than 100,000.

An article entitled "Preserving the Actor's Voice," is in revision for "Dramatics Magazine," a publication associated with the International Thespian Society and sent to high schools. A voice teacher and composer, Dr. Robert Downard, is the author. The article is based on his association with Denver's "The Professional Voice: Use and Abuse" workshops headed by Dr. Florence Blager.

Conference Support

The NCVS has initiated material and monetary support for various voice conferences around the country. The purpose of this support is to build strong relationships between organizations and the NCVS for more effective information dissemination. The purpose also is to emphasize the "national" aspect of the National Center for Voice and Speech. This approach to enhancing dissemination will be evaluated, and a report will be given in the next Status and Progress Report.

Videotapes

Pilot video tapes have been developed at the Recording and Research Center of The Denver Center for the Performing Arts for use in the many workshops and lectures given to the public. These tapes are being tested for ease of use in lecture format and in audience response. General distribution is being considered.

Dr. Ingo Titze is developing an instructional lecture/performance videotape that will be considered for distribution.

These materials complement the educational materials being prepared through the Continuing Education project of the NCVS.

Other Projects

University cheerleaders have been a target, high-risk voice group for the voice workshop team in Denver. A short study of the effectiveness of voice care and voice usage information has been reported by Dr. Florence Blager of the NCVS at the MedArt International (1992 World Congress on Arts and Medicine), March 1992, New York City. The results of that study suggest that intervention significantly raises the awareness of vocal health in university cheerleaders and may help prevent vocal problems, but does not necessarily improve the clarity of the voice.

Other activities being implemented include the writing of short articles on voice health to be sent to business magazines and professional organization newsletter, the development of a "Graduate Seminar: The Voice--A Multi-Disciplinary Approach" to be held in Denver, Colorado, and the coordination of materials dissemination with the NIDCD Clearinghouse.

NCVS Status and Progress Report - 2
April 1992, 187-198

Rationale and Structure of a Curriculum in Vocology

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

Abstract

A proposal is presented for formal academic training in vocology, the science of vocal habilitation and voice disorders. Rationale is given for expanding current methods of intervention offered in speech-language pathology to include more of the techniques offered in theatre arts and music. Formally, this can be accomplished by establishing a specialty track (vocology) in an M.A. program. The primary focus of vocal habilitation should be to meet the vocational and recreational needs of an individual, not simply a normal voice.

Introduction

In writing this paper, I wish to acknowledge at the outset that the ideas presented here belong to many people. At several recent conventions I have sensed a stirring among voice teachers and clinicians that suggests something is lacking in the formal training of practitioners in voice. I would like to serve merely as a conduit, therefore, in getting some of the ideas on paper. Perhaps in time more logic and more clarity will adorn a formalized proposal for an expanded curriculum in voice training and rehabilitation, but for the present the ideas are simply in the brainstorming stage.

Four issues will be touched upon in this article. First, there is a conceptual issue between speech-related and nonspeech-related voice disorders. Traditional perspectives on behavioral treatment of voice disorders may have been too narrow, primarily because voice has sometimes been thought to be a mere component of speech. It will be argued that many voice problems are being treated ineffectively as speech problems because their origins have relatively little to do with speaking *per se*. Second, there is the legal issue of who is licensed to train, remediate, or modify vocal

behavior. Speech-language pathologists, voice coaches and singing teachers need to get a better understanding of their respective roles in voice management. Third, there is the educational issue of what constitutes adequate training for practitioners. Finally, there is the practical issue of improving the tools for diagnosis and therapy. Each of these issues will be examined briefly. The intent is to stimulate discussion among professionals in the vocal arts and sciences, and to organize and unify approaches to intervention, assuming that cross-disciplinary approaches are indeed the solution.

I. THE CONCEPTUAL ISSUE

As a newcomer to a department of speech pathology and audiology some years ago, I was impressed by the apparent symmetry that exists in the speech chain between production and perception. Descending pathways in the motor system and ascending pathways in the sensory systems connected to similar building blocks. Since symmetry is appealing to most physicists, I was quite pleased with the order of the system I was going to be studying.

As I learned more, however, a conceptual asymmetry between production and perception emerged for me, primarily in the way components were assembled and labeled. Consider the words *hearing* and *speech*. Hearing deals with the reception and perception of nonspeech sounds as well as speech sounds. Many hearing scientists do not study speech *per se*. They study the basic auditory mechanisms, using stimuli such as clicks, tones, and noises. I wondered, is there a comparable field of investigation on the production side? Could it possibly be identified as "sound production in humans" or "the science of human sounding"? If these labels are too long, could they simply be abbreviated as "sounding science" to parallel hearing science? Aside from the label, I asked myself if there indeed should be a science that encompasses *all* the sounds a human can produce. Probably not. It would include not only voicing, whispering, and hissing, but also whistling, tongue clicking, finger clicking, clapping, sneezing, snoring, and belching. Sound can be emitted from so many parts of the human body, much of it being incidental to communication, that *human sounding*, unlike hearing, is difficult to cast into an organized science.

Is acoustic phonetics the match to hearing science? Acoustic phonetics is a well-developed science, but it cannot be paired with hearing science because it is not defined outside of the context of speech. For a more symmetric pairing with hearing science, one might perhaps look for an organ like the ear (a transducer) that acts as a major sound source for both speech and nonspeech production.

That led me to return to *voicing*. Although the term voicing is not usually used to describe sounds made in the vocal tract other than those produced by vocal fold vibration, there is much to be said for the hearing-voicing combination. From a biological point of view, the study of nonspeech vocalizations produced by both humans and animals should perhaps receive a label, such as "voicing science" or "voice science." Here laughing, crying, moaning, singing, humming, shouting, and other forms of vocal expression can be studied both outside and within the context of speech. We know that voicing, as hearing, is innate, but there is still debate about whether speech is innate. Developmentally, voicing is usually a precursor to speech. Speech builds upon vocalization, and in early stages of development begins to include "other human soundings" that supplement the primary source of sound.

From a health care point of view, hearing and voicing are matched in that they involve treatment of two primary organs, the ear and the larynx. Otology and laryngology are long established medical specialties. For nonmedical treatment, the clinical divisions are not as clear, nor can

they perhaps be. Speech-language pathology is not to voicing what audiology is to hearing. It covers a much broader spectrum of disorders that are not associated with a single organ of the body. Neurogenic disorders, for example, usually affect the articulators as well as the larynx.

Nevertheless, care and protection of the human larynx as the primary sound production organ has not received the prominence given to care and protection of the auditory mechanism. In a typical speech-language pathology Master's program, for example, study of voice production and voice disorders lacks focus on prevention, habilitation, and care. As the term "pathology" suggests, the disorder is identified and treated, but treatment emphasis is more on remediation than mediation. In my opinion, treatment should be directed toward bringing about vocal behavior *appropriate for the social* or *professional needs of the individual*, which may be far from "normal." A substantial sector of our population vocalizes for long periods of time (in a classroom, over the telephone, with clients and customers), and in noisy environments (automobiles, airplanes, subways, factories, etc.), which suggests that a greater awareness about the vocal organ and its care under high demand may be needed.

The need for new directions in clinical voice becomes clearer when one realizes that many disorders of the larynx are not speech disorders, just as many disorders of the ear are not speech (perception) disorders. The inability to vocalize loudly (or softly), the inability to vocalize at a high (or low) pitch, the inability to vocalize for long periods of time, the inability to vocalize clearly (without sounding harsh or hoarse), the inability to manage vocal registers, and the inability to keep from straining or pressing one's voice are no more speech disorders than high frequency hearing loss, poor temporal resolution, or poor pitch discrimination. There is no question that speech production is impaired by many disorders of the larynx, just as speech perception is impaired by disorders of the ear, but focusing on speech tasks alone may not get to the root of the problem. In audiology, loss of hearing acuity is not determined solely by speech perception tests, but rather by a whole battery of speech-like and nonspeech-like auditory stimuli.

Examples of speech related voice disorders would be the inability to produce proper intonation and stress, or the inability to produce a voiced stop, a voiced fricative, or a glottal stop. This would parallel similar voiced-unvoiced distinctions or prosodic feature discrimination on the perception side. In some languages other than English, the demand on the voicing mechanism for phonetic contrast is high. Hence, there could be a considerably greater number of speech-related voice disorders when the language demands these contrasts. In English there are relatively few.

Typical voice clients are professionals who have fatigued or abused their vocal mechanisms, e.g., salespeople, teachers, actors, singers, lawyers, politicians, ministers, auctioneers, cheerleaders, aerobics instructors, coaches, stock traders, and construction workers who must communicate in noisy environments. Many of these people speak for long hours and often under considerable psychological stress. Psychogenic voice disorders may occur because voice is so intimately connected with emotions such as fear and anxiety. Other expressions of emotion, such as elation, melancholy, or despair, are almost definable in terms of vocal outputs (laughter, weeping, choking-up, etc.). Feelings of depression, loss of self-esteem, loneliness, or helplessness can apparently impact so strongly on the vocal mechanism that voicing is inhibited or altered by these emotional disturbances (Aronson, 1989).

In addition, there are patients who suffer from diseases or trauma to the larynx, some of whom must readjust their vocal productions to match different airway structures created by head or neck surgery. A key problem that most clients and patients face is how to vocalize efficiently under adverse or abnormal conditions, when extreme or unusual demands are placed on the vocal mecha-

nism. Telling a stock trader at the Chicago Mercantile Exchange, or a coach along the sidelines of a basketball floor, to limit the vocal output is like telling a boxer not to get hit, or a ballerina not to get on her toes. We need training in optimal usage of the voice that is applicable to conditions other than normal conversation in a quiet room. Unfortunately, low to moderate intensity conversation is often the target in typical voice therapy sessions. I hear very little shouting practice going on in therapy rooms. A target for vocal management should be the best voice for the specific communication need, not just a pleasant one-on-one speaking voice. Considering the percentage of the nation's population that suffers from mild to severe vocal problems (3-7%, depending on whether or not children are included; Wilson, 1987), this should become a high priority in speech-language pathology.

Consider the block diagram shown in Figure 1. Hearing science and voice science interface with speech science, which is compartmentalized into speech production and speech perception and is closely linked to language. (No significance should be attached to the fact that language science is placed above speech science.) Otology and laryngology parallel the hearing science and voice science division as medical specialties, whereas speech science interacts with many medical specialties. Finally, audiology, speech-language pathology, and a newly-proposed specialty called *vocology* (Titze, 1989) round out the clinical areas of training. Thus, one could argue for "triologies" in several related professions, but that would perhaps be pushing symmetry a little too far, especially if it were at the expense of excluding such specialties as dentistry, neurology, pediatrics, and psychology.

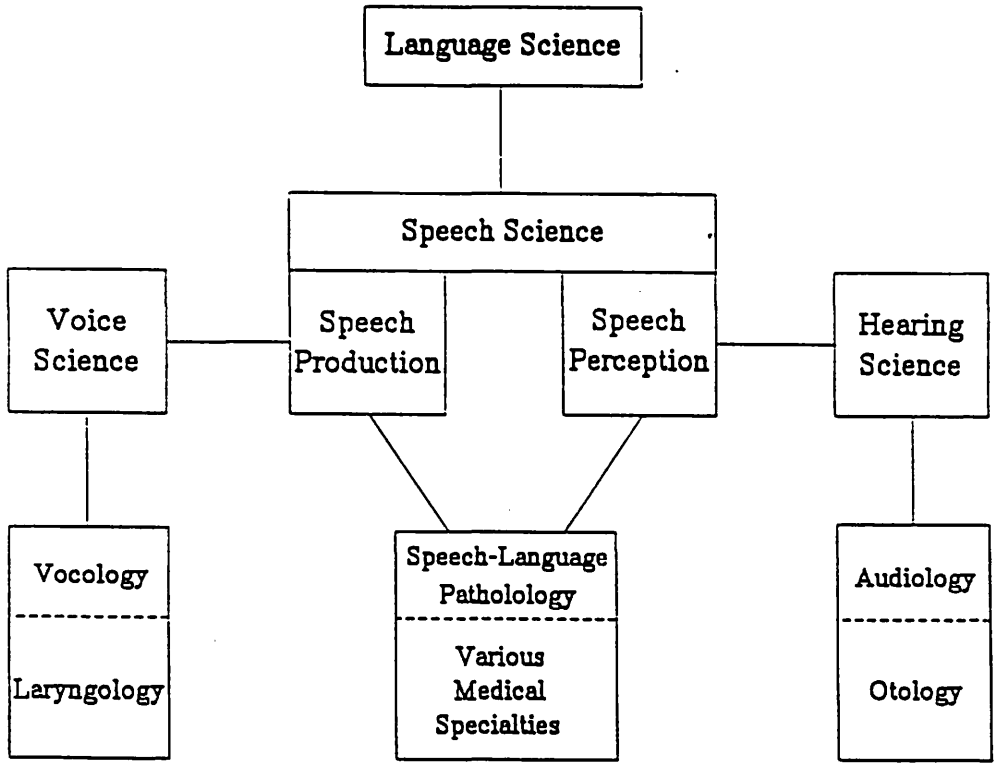

Figure 1.Organization of disciplines in relation to the newly proposed vocology specialty.

Vocology (or *vocalogy*, which was considered, but was a bit more difficult to pronounce), is defined here as *the science of vocal rehabilitation and voice disorders*. Gates (1989) has also proposed the use of this term, but for a broader specialty in treatment and care of the human voice. Vocology as defined here does not duplicate phoniatry, a medical specialty in Europe and other parts of the world that deals with treatment of voice and speech disorders. Phoniatry does not exist in this country. A vocologist in the United States would be a practitioner who is certified and licensed to carry out procedures for prevention of voice disorders, protection of the vocal mechanism, and modifying vocal behavior (including for professional purposes).

II. THE LEGAL ISSUE

Under current licensure laws in most states, a licensed speech-language pathologist can test, evaluate, habilitate, rehabilitate, and instruct in the development of voice. The phrase "instruct in the development of voice" is understood to relate to physical growth and maturation rather than development of exceptional vocal skills, such as those used in singing, acting, or public speaking.

A problem has arisen for teachers of singing, acting, and public speaking who believe they have acquired, through special training and experience, the ability to evaluate, habilitate, and train vocal behavior. Some of these voice coaches and singing teachers have no licensure or certification, but in some instances can demonstrate expertise equal or beyond that of a licensed clinician. The problem is particularly relevant when therapy is recommended for someone who is in a voice training program. If a singer or an actor develops vocal nodules, for example, who manages the treatment--the speech-language pathologist, the voice teacher, or the laryngologist? Ideally, the three work in close cooperation, but that does not always happen.

Territorial conflicts have generally been avoided by referring clients with speech-related voice disorders to the speech-language pathologist and those with singing or professional speaking disorders to the voice teacher. This approach is defensible, in principle, but it suffers from a presently unclear distinction of what constitutes one's speaking voice or one's professional voice, or what constitutes a disorder resulting from one type of vocal abuse versus another. In the previous section, the case was made that voice disorders are not always speech disorders. Many people in this country whose voices are impaired could benefit from the type of vocal behavior modification that is practiced by a skilled singing teacher or a drama coach. These specialists, who typically train singers, actors, and public speakers, have long dealt with the problems of unusual demand on the vocal organ. They listen to performers daily as they cope with highly variable conditions. They have learned how to communicate effectively in noisy environments, in large halls, in outdoor arenas, and over long periods of time. Singing and acting often involve cultured shouting, the art (and hopefully, in time, the science) of making intense and prolonged vocal productions healthy, pleasant, and efficient. Vocologists should obtain this skill. Although the typical speech-language pathologist has a better knowledge of patient management and clinical care than the typical voice coach or singing teacher, it would be difficult to make the case that the speech-language pathologist is presently more qualified to correct or habilitate vocal behavior than either of the other two, provided the knowledge base (anatomy, physiology, etc.) is roughly equivalent.

State licensure laws could be tested in a court with expert witnesses on both sides. (In some states the threat of litigation has already appeared. I was invited to attend discussions on legal issues pertaining to voice therapy on two occasions in two separate states.) I am afraid that if these threats continue, there will ultimately be more losers than winners. The growth of a profession should not

be decided in the courts. Besides, ethical codes and practices of ASHA state that "individuals shall use every resource available, including referral to other specialties as needed, to provide the best service possible," and that "individuals should establish harmonious relations with colleagues and members of other professions." Some significant dialogue is presently ongoing between the American Speech-Language-Hearing Association (ASHA) and the National Association of Teachers of Singing (NATS). The intent of this dialogue is to identify the respective turfs and to identify people who have dual expertise. Furthermore, some minimum skills have been identified by this bipartisan group to serve as a springboard for future curriculum development (see Section III). The Voice Coaches Association is also addressing the issue of training versus therapy.

III. THE EDUCATIONAL ISSUE

ASHA has traditionally resisted specialty certification. The primary argument for maintaining general training in communication disorders is that speech, hearing, voice, and language can be very interactive. The generalist can often sense this interaction and can correct an apparent misevaluation, or inappropriate management. Also, it is difficult to predict one's clientele or future work opportunities. But as more information about the disorders accumulates, specialization and fragmentation of previously formalized programs are inevitable. A key question is, what is an adequate amount of education needed to treat the spectrum of voice disorders? When that core education has been obtained, continued education during practice, combined with experience and prevailing curiosity, will then produce the necessary breadth.

Table 1 outlines a curriculum for a Master's degree in Speech Language Pathology with specialty training in vocology. Assuming a candidate enters the program with an undergraduate degree in speech-language pathology, theatre, or music, a series of core requirements would first be fulfilled. These courses are fundamental to all areas of speech-language pathology and could serve as a springboard to specialization in particular settings (public schools, hospitals, private practice) or for particular disorders (voice, language, neuropathologies, etc.). The core courses would constitute about one year of a two-year Master's program. Principles of assessment, intervention, and counseling would be taught in general terms and the broad spectrum of disorders in speech, hearing, and language would be introduced.

The second part of the training would involve a more intensive focus on voice disorders. Here one could conceive of a list of required courses in vocology and an additional list of elective courses, as shown in Table 1. Principles of Voice Production deals with basic anatomy, physiology, acoustics and biomechanics. Specific topics of study are the regulation of pitch, loudness, register, and vocal quality. Additional topics include pitch and loudness modulations in voice (vibrato, tremolo, tremor, jitter, shimmer, etc.), voice classification, theories of vocal fold vibration, and special resonance effects in subglottal and supraglottal airways.

The instrumentation course introduces the vocologist to current hardware and software for voice analysis. This includes sound spectrography, electroglottography, electromyography, voice perturbation analysis, videostrobolaryngoscopy, inverse filtering of oral flow, and various respiratory function analyses (magnetometry, plethysmography, spirometry). Valuable diagnostic training can also be received with the use of voice synthesis. By relating voice qualities such as hoarseness, breathiness, tremor, and vibrato to acoustic perturbations and modulations (jitter, shimmer, additive noise, etc.), the vocologist learns how to integrate perceptual and acoustic phenomena.

The third major area of study deals with diagnosis and treatment of voice disorders. Typical diseases of the larynx are described and a rationale is given for choosing medical, surgical, or therapeutic intervention. Neurologic diseases are contrasted with infectious diseases of the airway and disorders related to injury, vocal abuse, surgery, or fatigue. The primary emphasis in this segment of didactic coursework is to sort out the nature of the disorder and to determine the best method of treatment.

The fourth major didactic area of study in vocology is directed toward behavioral treatment of voice disorders. Here a return to older traditions is recommended. In some institutions, speech-language pathology and audiology were separated from speech and drama departments many years ago. Although this separation was important to give communication sciences and communication disorders a firm scientific footing and affiliation, some of the connections with professional voice users were severed in the process. In particular, theatre arts and music had much to offer by way of vocal training. Topics such as "voice and diction," "oral interpretation," and "public speaking" stressed not only average communication skills, but skills for particular (exceptional) needs. Emphasis was on habilitation as much as rehabilitation. For a vocologist, voice training and voice therapy are often indistinguishable.

Table 1. Proposed Vocology track in Speech-Language Pathology

Core requirements in Speech-Language Pathology
- Neural Processes and Neurogenic Disorders
- Hearing Loss and Audiometry
- Phonological Development and Disorders, Stuttering
- Psychological Issues and Counseling
- Principles of Assessment
- Principles of Intervention

Required courses in Vocology
- Principles of Voice Production
- Instrumentation for Voice Analysis
- Voice Disorders
- Voice Training and Rehabilitation
- Private Voice Lessons
- Voice Practicum (including studio observation)

Additional electives in Vocology
- Methods of Teaching Voice (Music School)
- Voice for the Actor (Drama School)
- Head and Neck Cancer (Otolaryngology)
- Cleft Palate (Otolaryngology)

Candidates for vocology should enroll in private vocal lessons each semester for the entire duration of the program. It is desirable for a voice clinician to experience the sensations associated with optimal vocal production. This can be accomplished by training one's voice under the guidance of a qualified vocal instructor. Undergraduates who contemplate vocology as a specialty should be encouraged to enroll in private voice lessons prior to entering the Master's program.

An extended Voice Practicum should include observation of vocal instruction, as well as the personal experience of training voices. This should be in addition to the current practicum offered

in the management of voice disorders. The management of vocal registration (maintaining stable and consistent vocal qualities over wide pitch and loudness ranges) and the management of optimal vocal placement (maintaining stable and consistent resonance in vowels) is routinely taught in vocal studios. Daily exercises and drills are designed to extend pitch and loudness ranges and vocal endurance. Techniques have been developed for overcoming unnecessary glottal stops, "hard" vocal attacks, breathiness, or excessive nasality.

Departments offering training in vocology should consider formal arrangements with schools of music and theatre arts to allow candidates to observe private voice lessons. This should be done without compromising existing student-teacher relationships. Observation should be across a number of voice studios because styles and approaches vary significantly among voice teachers. A sampling of these styles and approaches would allow the vocologist to determine the underlying common principles of voice training. Recommendations would be for a certain number of clock-hours of observation, spread over a minimum of three voice studios (voice teachers). Typically, students at different levels of vocal development would be observed.

A similar arrangement for observation should be made with the drama (theatre arts) school. Prolonged voice production with varying loudness, pitch, and quality is often required in the theatre. Most theatre arts departments have a person on staff who teaches vocal skills related to public speaking and drama. This resource should be embedded in the vocology curriculum.

Finally, an additional set of optional courses is listed in Table 1. Many vocal music departments offer a sequence of courses in vocal pedagogy. Although portions of these courses are a repeat of anatomy, physiology, and acoustics of the vocal instrument, a major sector is geared toward practical implementation of teaching methods. Here each member of the class brings in a pupil (a friend, a volunteer, or some assigned undergraduate voice student) to demonstrate his or her approach to training voice. The demonstrations are at a fairly elementary level, and would contrast with the more intense work that is observed in the studio. More time is given for discussion between class members, pupils, and instructor. Similarly, Voice for the Actor may be a new and stimulating experience.

A clinical fellowship in a laryngology setting is recommended. This is not a new concept, but rather similar to the CFY training already implemented in speech-language pathology and audiology. The vocologist fellow would assist in diagnostic procedures and gain experience in clinical management of voice disorders. Many major teaching hospitals and clinics have speech and hearing divisions where such training is ongoing. Some additional coursework in head and neck cancer and cleft palate may be appropriate in this period of extended training.

IV. PRACTICAL ISSUES OF VOICE EVALUATION, TRAINING, AND HABILITATION

Having argued for a clinical program that combines elements of voice training with voice therapy, there remains an issue of how evaluation and modification of vocal behavior might be approached systematically. It would be presumptuous of me to offer detailed descriptions of diagnostic or rehabilitative procedures, but an increased awareness of some guiding principles that have strengthened other disciplines may help.

First, evaluation of vocal function is in need of some standards. It is informative to quote from Hirano (1981): "In the field of otology and audiology, several techniques for the clinical examination of auditory function have been established and standardized. As a result of the standardization of these techniques, it is possible to compare the auditory function of different subjects examined at different places and/or different occasions and to monitor the results of treatments in a

reliable manner. With regard to phonation, various methods have been proposed and used by many clinicians and researchers all over the world. Unfortunately, none of these methods appears to be standardized on an international basis."

Similar statements were made at the Conference on the Assessment of Vocal Pathology (Ludlow & O'Connell-Hart, 1981) and at the more recent conference on Voice Acoustics and Dysphonia in Gotland, Sweden (Fritzell & Fant, 1986).

The field of audiology presently has more than ten standards approved by the American National Standards Institute (ANSI). They are cataloged under the heading of Bioacoustics (ANSI, 1988). These standards include specifications and methods for calibration of audiometers, methods for measurement of electroacoustical characteristics of hearing aids, methods for calibration of earphones, occluded ear simulators, earphone couplers, computation of loudness in noise, and psychoacoustical terminology. The field of Speech perception has standardized a method of measurement of monosyllabic word intelligibility, rating noise with respect to speech interference, and methods of calculating the articulation index. The field of speech production, and in particular voice production, has no ANSI approved standards. This poses a serious limitation on clinicians and researchers. Paradigms and methodologies are frequently being attacked by critical reviewers as unproved, untried, and unreliable. Reporting of results is cumbersome because procedures have to be restated in every publication, while referees find themselves repeatedly uncovering the same (or similar) technical and procedural flaws.

Because the human larynx is a delicate biological organ (like the ear), it deserves protection and functional evaluation. I believe that *vocometry* should be developed with a level of sophistication that parallels audiometry. This should include threshold of phonation measurements, threshold of "pain" measurements (in terms of vocal loudness), maximum acceptable environmental noise (for non-injurious conversation), and standardized methods for computing vocal performance indices. In addition, instruments used for vocometry (those that measure fundamental frequency, vocal intensity, voice perturbation, and spectral features, etc.) need calibration standards.

Much has been learned in recent years about the control of loudness, pitch, and vocal quality. New knowledge is also emerging steadily toward a better understanding of what constitutes normal irregularity in the vocal signal and what might be deemed pathologic. A goal of every vocologist should be to understand and be able to measure ranges of intensity, fundamental frequency, spectral slope, and perturbations in the human voice.

In teaching the control of loudness during therapy, primary, secondary, and tertiary mechanisms for intensity regulation should be clearly understood. It is known that lung pressure is the primary variable for intensity regulation (Isshiki, 1964). An increase of 6 dB per doubling of lung pressure is obtained if the spectral content of the voice signal and the fundamental frequency are unaltered. A steeper increase can be achieved, however, if some glottal and vocal tract control of loudness is combined with lung pressure variation. Increasing vocal fold adduction (from breathy voice to normal voice) can add about 2-3 dB in intensity, and tuning the vocal tract formants to the harmonics can increase the intensity by another 5-6 dB. A price is sometimes paid, however, for the use of these secondary and tertiary mechanisms. Excessive adduction may cause tissue damage, and excessive use of vocal tract tuning to boost vocal output power may distort intelligibility and render the speech unnatural. Stage performers and public speakers use these glottal valving and resonance mechanisms for altering vocal intensity (Lessac, 1967), and vocologists should be aware of their merits and shortcomings.

In managing vocal endurance, a signal-carrier guiding principle may be of use. In his book on teaching speech to the hearing-impaired child, Ling (1976) proposes a method of vocal training that follows normal speech development. The basic idea is that the breath stream is the carrier of the voice, and the voice is the carrier of articulation. According to this model, voice management should be directed toward maintaining stable and well-controlled carriers before modulations are attempted. Improper breathing is likely to disrupt phonation, and improper phonation is likely to affect articulation. Breath management, therefore, becomes a focal point for intervention when voices fatigue easily.

Another focal point of intervention for rapidly fatiguing voices is voice initiation and degree of adduction. Initiation of voicing should be able to be accomplished very gently, with slightly abducted vocal folds. Vennard (1967) has described this approach in detail. For most people, little adductory muscular effort is needed for voice initiation. Given that the vocal folds approximate reflexively as the breath stream is established, excessive voluntary adduction can lead to a pressed voice. This in turn appears to be connected with vocal fatigue if used extensively, although little scientific evidence is available. Hard glottal attacks seem to be the result of "all or none" activity of the adductors. They should not be practiced by most vocalists because of the possibility of excessive contact pressures between the folds. Sundberg and Gauffin (1978) have described a continuum of adduction of the vocal folds that varies from breathy voice to pressed voice. In the middle is a presumed optimal mode, the "flow mode." Here the vocal folds attain their greatest amplitude of vibration for a given subglottal pressure. Hence, the peak of the time-varying flow is maximized. At the same time, there is only a small leakage (if any) of air in the posterior glottis.

As a general rule, the ability to speak expressively, clearly, and over long periods of time comes with optimal usage of the voice. Although muscle training is involved, the training is more in terms of coordination and balance of groups of muscles than strengthening muscles in isolation. In a well-coordinated vocal exercise, opposing sets of muscles (agonist-antagonist pairs) appear to contract and release gradually and differentially, rather than in an all-or-none fashion. It has been shown, for example, that in a well-trained singer, the diaphragm relaxes gradually while abdominal muscles contract gradually in the early part of a sung vowel (Leanderson et al., 1988). The same principal of differential control of agonist-antagonist pairs of muscles (cricothyroid and thyroarytenoid) has been shown for a well-coordinated register change (Hirano, Vennard, & Ohala, 1970). If the term "vocal athletics" is appropriate for high performance sound production, then a comparison of prolonged and demanding speech with gymnastics or figure skating is perhaps more appropriate than a comparison with weight lifting. Efficient voice production often develops when small, seemingly insignificant, exercises are practiced repeatedly.

A vocologist should also be sensitive to personality traits that may be reflected in the human voice. It is tempting to translate the phrase "optimal usage of the voice" into "stereotypic sound." Optimal usage may differ for different speakers; hence, the approach to training and rehabilitation needs to be tailored to the individual's personality as well as to the individual's physiology. Some counseling is appropriate to establish if the vocal image portrayed by the speaker is (1) congruent with the speaker's intent, and (2) best suited for the person or the profession.

Care must be taken, of course, that voice modification does not become excessively cosmetic. If personality can be changed by altering vocal behavior, certain identifying characteristics of an individual may be altered as well. By way of an analogy, most of us would agree that crooked teeth are a correctable disorder, but what about a crooked smile or a dimple in the cheek? Similarly,

some vocal qualities (e.g., occasional creakiness or breathiness) may not reflect a pathology as much as an individuality. Frequent use of vocal fry, for example, may be indicative of a pathology for some vocalists, but harmless and attractive for others.

Summary and Conclusions

Our society places unusual demands on vocal communication for some people. Speaking in noisy, polluted, and frequently changing environments, and speaking for long periods of time, seems to have brought about a condition where an increasingly growing sector of our population is becoming vocally impaired. Voice disorders often result from heavy taxation of an organ whose primary function is not voice production, but airway protection. In combination with its susceptibility to respiratory diseases and its vulnerability to neurologic and psychogenic disorders, the larynx needs special care from a group of professionals working in concert.

Education and certification in a new field called vocology can be incorporated into traditional speech and hearing science or speech pathology and audiology departments, but close alliances should be established with theatre arts and music schools. This will limit the availability of the programs, but no immediate problem is foreseen because there is a limited market place for vocologists at present. (This likely to change as vocal awareness reaches the public schools and industry. In many secondary school systems, for example, choral conductors and speech and drama teachers are beginning to become aware of the importance of vocal health and hygiene.) Didactic coursework in basic and applied voice science needs to be developed, voice practica need to be extended to include experience with singing teachers and voice coaches, and candidates should enroll in regular private voice lessons. A clinical fellowship that includes close interaction with a laryngologist is recommended to complete the training.

Ideally, vocology would become a third arm of a "tri-ology" in the communication sciences and disorders, together (and yet distinct from) audiology and speech-language pathology. This is unlikely to occur very soon, however. Perhaps it will never occur because of strong traditions, or perhaps it will be a natural inclusion in the development of a professional doctorate. At the very least, however, vocology should become a specialty track in speech-language pathology, with requirements in SLP and minimum requirements in vocology (as outlined in this paper) being merged.

With this new specialty track, it is believed that voice training and vocal rehabilitation can be brought under a common footing. A systematic approach to training and rehabilitation should emerge that meets the challenge of increased vocal demand in our society. Thus, vocology should have as its goal not only the treatment required to restore normal vocal behavior, but also the treatment that brings about vocal behavior appropriate for the social or professional needs of every individual.

Acknowledgement

The author wishes to thank a number of colleagues at the University of Iowa for helpful critiques in writing this paper, including John Folkins, Hughlett Morris, Charles Anderson, Penelope Hall, and Amy Weiss. In addition, outside opinions were obtained from Christy Ludlow, Sandra Holley, Oren Brown, Jean Westerman-Gregg, Diane Bless, Fred Minifie, Katherine Verdolini-Marston, Ronald Scherer, and Lorraine Ramig. These opinions were extremely helpful, but were not necessarily in full agreement with the author's point of view.

References

Aronson, A. (1989). Psychogenic voice disorders. Pacific Voice Conference, San Francisco.

Catalog of American National Standards. (1988). Washington, D.C.: American National Standards Institute.

Fritzell, B., & Fant, G. (1986). Proceedings of the Conference on Voice Acoustics and Dysphonia (Gotland, Sweden). J. Phonetics, 14, 3-4.

Gates, G. (1989). Personal communication at the Pacific Voice Conference.

Hirano, M. (1981). Clinical examination of the voice. New York: Springer Verlag.

Hirano, M., Vennard, W., & Ohala, J. (1970). Regulation of register, pitch, and intensity of voice. Folia Phoniatr., 22, 1-20.

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

Leanderson, R., & Sundberg, J. (1988). Breathing for singing. J. Voice, 2(1), 2-12.

Lessac, A. (1967). The use and training of the human voice, 2nd edition. New York: Drama Book.

Ling, D.H. (1976). Speech and the hearing impaired child. Washington, D.C.: A.G. Bell Association for the Deaf.

Ludlow, D.L., & O'Connell-Hart, M. (Eds.). (1981). Proceedings of the Conference on Assessment of Vocal Pathology. Rockville, MD: ASHA Reports, No. 11.

Sundberg, T., & Gauffin, J. (1978). Waveform and spectrum of the glottal voice source. In Frontiers of speech communication research: Festschrift for Gunnar Fant, pp. 301-320. London: Academic Press.

Titze, I.R. (1989). Vocology. Ntl. Assoc. Teachers Singing Journal, 46(3), 21-22.

Vennard, W. (1967). Singing: The mechanism and technique. New York: Carl Fisher.

Wilson, D.K. (1987). Voice problems of children, 3rd edition. Baltimore, MD: Williams and Wilkins.

NCVS Status and Progress Report - 2
April 1992, 199-215

Articulatory Organizational Strategies and the Roles of Audition

Nancy Tye-Murray, Ph.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa

Abstract

This review highlights a decade of research performed at the University of Iowa with deaf talkers. The experimental findings suggest that absence of audition precludes talkers from developing typical articulatory organizational strategies, and affects their abilities to produce specific speech events. For instance, deaf talkers may not organize their articulatory behavior such that movement occurs continuously from one open posture to the next, and they may shorten their open posture steady-states. These deficits have predictable consequences for intelligibility. The data help illuminate the roles of audition in acquiring and maintaining speech; five roles are considered here.

During the past decade, we have performed several studies with deaf talkers[1] at the University of Iowa. Our goals have been to explore how the absence of audition affects speech development, and to describe the articulatory behaviors that underlie some of the characteristic articulation problems. This information might elucidate the roles of audition in learning and maintaining speech, and might lead to more effective speech teaching methods

This review highlights our work. It begins with a description of the operational model we have used for formulating experimental hypotheses. Experimental findings and implications for the roles of audition in acquiring and maintaining speech production are then presented.

[1]The term deaf will be used in reference to individuals who have a profound, prelingual, sensorineural hearing loss.

Operational Model

The speech of deaf talkers has often been described as having a great number of linguistically-defined errors. These errors have been categorized in terms of features, prosodic anomalies, and phonemic aberrations such as omissions, substitutions, and distortions (Hudgins and Numbers, 1942; Smith, 1975). These descriptors are consonant with a traditional linguistic model, and have yielded much useful information. However an exclusively linguistic frame of reference may limit, unnecessarily, investigation and remedial efforts for at least two reasons.

First, several linguistic cues may be associated with a single articulatory gesture. Conversely, a single cue may be associated with more than one gesture. As such, attempts to modify the articulatory behavior underlying an error may result in the alteration of several linguistic cues, or may affect several different articulatory gestures. The net effect may not be a reduction in the number of linguistic errors nor an enhancement of intelligibility.

The second limitation is that a linguistic vocabulary may not allow for an adequate characterization of the behavioral deficits. Some researchers have suggested that linguistic units (such as phonemes) may not correspond to units of motor processing. For instance, Folkins and Bleile (1990) propose that "speech motor strategies are organized, not to reach a series of ordered goals determined by phonological units, but to produce the speech message as a larger, holistic behavior (pg. 605)." After a talker decides what to say "using the processes of linguistic performance, there may be no need to preserve divisions attributable to linguistic units in motor control processes (pg. 605)." Although this view is controversial (see Fowler, 1986; Browman and Goldstein, 1986), it presents an important directive: research and remediation efforts should not be limited by postulated units of production, and should be influenced by models of neuromuscular control developed for motor activities other than speech.

We have not abandoned a linguistic framework entirely. However, our operational model motivates research efforts to focus upon the manner in which the general principles of deaf talkers' articulatory organization differ from those of hearing talkers. Our operational model assumes that there are general principles of articulatory organization underlying speech production. These principles encompass regulation of speech breathing, movement of the articulators coninuously from one open posture toward the next, flexion and extension of the tongue body, and alterations of articulatory behavior to achieve changes in speaking rate and syllable stress. In learning to talk, the child abstracts many (if not all) the principles of articulatory organization by listening to others speak. Once learned, these principles constrain how the child produces individual speech events. For instance, a child might say, "Keep the coop." In distinguishing between the postures for /i/ and /u/, the child extends and flexes the tongue body like typical talkers of the language.

The operational model holds implications for when and why speech might be intelligible. By assuming that information about articulatory organization is accessible through listening, the possibility arises that this information plays an important role in speech perception. The speech signals produced by hearing talkers may have a global patterning because their articulators move in a principled fashion. This global patterning unfolds throughout an entire sentence, and exists regardless of its phonetic composition. For instance, if a talker speaks quickly, no one segment of the sentence conveys speech rate. Rate is manifested throughout by the pattern of spectral change. The global patterning provides a context for listeners to perceive specific speech events. With respect to rate, a 50 ms acoustic segment in a sentence will create a different percept if listeners perceive the sentence as being spoken quickly rather than slowly. When the general principles of articulation are

aberrant, as may be true with some deaf talkers, the global patterning is absent and so intelligibility decreases. For example, a deaf and hearing talker both may forward their tongue bodies in the vocal tract and raise their jaws when producing /i/. However, when speaking the sentence in which the posture occurs, the deaf talker might move the tongue body deviantly, manage speech breathing improperly, and might not move the articulators continuously from one open posture to the next. The deaf talker's signal thus does not have the appropriate context (the global patterning in the signal) for listeners to recognize the vowel.

Experimental Findings

For clarity of presentation, our experimental findings will be considered with respect to what they imply about articulatory organizational strategies, and then with respect to how some deaf talkers may realize specific speech events. In reality, this dichotomy is not always easy to delineate. Results from a particular experiment may be divided between the two sections.

Most of our studies have involved cinefluorography, microbeam, or strain gauge. Cinefluorography yields two-dimensional x-ray moving pictures of the tongue body, jaw, lips and palate during speech production. Microbeam is similar to cinefluorography, but provides x-ray data only about the movement of gold pellets glued upon the articulators. For strain gauge, small wire transducers are taped to the surface of the lips and jaw. The transducers provide an analogue tracing of the articulatory movements that occur in a superior-inferior plane.

Principles of Articulatory Organization

One difficulty in attempting to identify deficits in the articulatory organization of deaf talkers is that no clear consensus exists about how hearing talkers orchestrate their articulatory gestures. Nonetheless, the following three attributes appear to be common among talkers with normal hearing: a) the tendency to move the articulators continuously from one open posture to the next, b) the tendency to flex and extend the tongue body abundantly, and c) the tendency to distinguish stressed from unstressed syllables by using greater movement displacements, velocities, and durations, regardless of the phonetic composition or stress pattern of the utterance.

Moving the articulators from one open posture to the next

Talkers with normal hearing may move their articulators continuously from one open posture of articulation (or "vowel steady-state") to another when speaking (Ohman, 1966; Fowler, 1986). For example, after one open posture is established and maintained, the tongue body and jaw begin moving toward the next open posture, even if the talker must form a constriction for an intervening fricative or stop consonant, as in /isa/ (Carney and Moll, 1971). Acoustically, the formant transitions of the initial vowel are influenced by the identity of the following vowel (Ohman, 1966). When the second vowel has a high frequency second formant, as with /isi/, the initial vowel-consonant transition has a higher frequency than when the second vowel has a low frequency second formant, as with /isa/.

This continuous movement from one open posture to the next may provide a "scaffolding" for speech production. That is, normal hearing talkers may achieve specific articulatory gestures by imposing modulations upon a cyclical movement from one open posture to the next.

Physiological and acoustic investigations suggest that deaf talkers do not incorporate this principle of articulatory organization into their speech production (see Tye-Murray, Zimmermann,

and Folkins, 1987, for a review). For instance, when asked to produce consonant-vowel-consonant syllables in the carrier phrase, Take a (keyword) aside, deaf talkers perform differently than hearing talkers (Rothman, 1976). The identity of the keyword vowel does not affect the formant structure of the preceding schwa, whereas it does with hearing talkers.

We have conducted two investigations to explore this issue more fully. The first involved cinefluorography (Tye-Murray, 1987) and the second, spectrographic analyses (Tye-Murray and Woodworth, 1989).

In the first study, 2 hearing and 5 deaf subjects produced syllables beginning with bilabial and palatal stop consonants. Tracings of the vocal tract were made at the time when subjects initiated closure. Hearing talkers tend to lower their jaw during a bilabial closure (as for /b/ and /p/) more when the following vowel is /a/ than /i/, probably because /a/ requires greater mouth opening than does /i/ (Sussman, MacNeilage, and Hanson, 1973). For the palatal stops /k/ or /g/, the place of vocal tract constriction is more anterior when the front vowel /i/ follows than the back vowel /u/ (as in keep versus coop) (Kent and Moll, 1972; MacNeilage and DeClerk, 1969). In this cinefluorographic study, only the hearing and the two most intelligible deaf subjects varied tongue and jaw position during the closure postures as vowel identity altered. The remaining subjects did not consistently demonstrate context effects (Waldstein and Baum, 1992, have recently reported compatible findings). This finding suggests that at least some deaf talkers do not move their articulators toward the following open posture throughout the preceding consonant closure.

Context effects on acoustic vowel and word durations were examined in the second study. Three deaf and two hearing subjects spoke words and phrases with increasing numbers of syllables (e.g., shade, shady, shadiness, the shade lingered). For hearing talkers, a syllable or vowel steady-state is shorter when other syllables follow, such that the acoustic duration of /sheId/ is shorter in the derived word shady than the kernel word shade (Lehiste, 1972). Apparently, the movement that occurs toward the second syllable limits the duration of the initial syllable in the derivative. In this study, the deaf subjects sometimes, but not always, decreased word and vowel durations for the derivatives compared to the kernel forms. The least intelligible subject rarely decreased these durations. Moreover, unlike the hearing subjects, the deaf subjects typically did not reduce vowel segments more than consonant segments. These inconsistencies and qualitative differences in achieving segmental shortening may indicate that some deaf talkers do not move their articulators continuously from one open posture to the next.

It is important to consider the consequences that this difference in articulatory organization may have for the intelligibility of deaf talkers. There are at least three. First, since the acoustic signal associated with their consonant closure postures may vary little with the identity of the following vowel, there may be less information available for vowel recognition. Second, the speech signals may sound arhythmic because the signals do not manifest cyclic-like articulatory movement (Fowler, 1979; 1983). Third, and most subtle, because gestures are not performed as modulations of oscillatory articulatory activity, the gestures may not be blended together in a way that results in intelligible speech, and there may not be the appropriate context (global patterning in the signal) for listeners to recognize specific speech events.

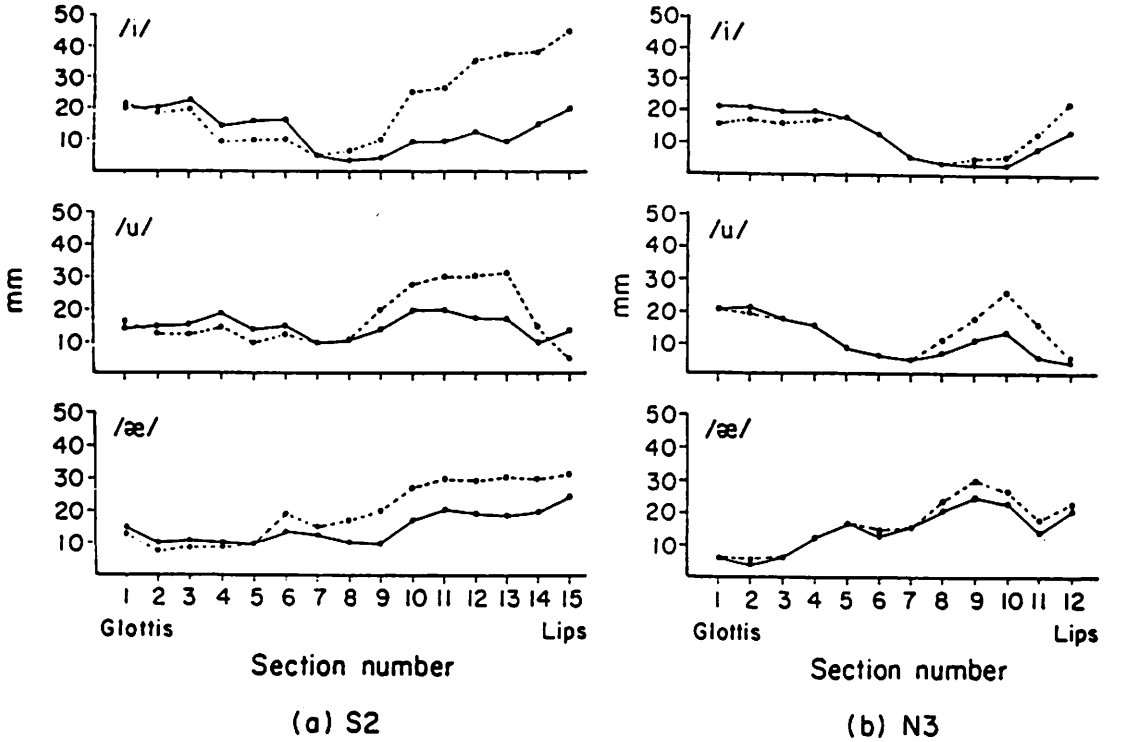

Figure 1. Vocal tract cross dimensions for a deaf (S2) and hearing (N3) talker producing /i/, /u/, and /ae/. - - - = biteblock, ____ = non-biteblock. The cross dimensions were constructed by measuring the linear distance between the tongue surface and the surface of the vocal tract at equadistant points, and then plotting the results. (From Tye, Zimmermann, and Kelso, 1983, with permission.)

Tongue activity

The second deficit in articulatory organization relates to tongue activity. Some deaf talkers may not contract and extend their tongues like hearing talkers. Two studies, one with microbeam (Tye-Murray, 1991) and another with cinefluorography (Tye, Zimmermann, and Kelso, 1983), point toward this conclusion.

In the first study, two deaf and two hearing talkers spoke simple phrases. Small gold pellets were placed upon the tongue blade (the region of the tongue that contacts the alveolar ridge during production of /t/) and the tongue dorsum (the region of the tongue that contacts the palate during production of /k/). Microbeam was used to track the pellets during opening articulatory gestures. The two tongue pellets followed similar paths for the deaf subjects but usually different paths for the hearing subjects as they spoke various syllables. Watching the two pellets displace on a visual display, it appeared that the linear distance between the two pellets increased and decreased frequently when the hearing subjects spoke, analogous to a rubber band being stretched and loosened. By contrast, the linear distance between the two pellets appeared to remain fairly constant for the deaf talkers, analogous to a tongue depressor being moved within the vocal tract cavity.

In the second study, two prelingually deafened and two hearing subjects spoke with and without their jaws immobilized by a biteblock. The lingual contours associated with their open posture steady-states were compared between the two speaking conditions. Figure 1 shows the vocal tract cross-dimensions (following Lindblom and Sundberg, 1971) for one deaf and one hearing subject. Both subjects reached similar points of maximum constriction within the vocal tract when producing the vowels /i/, /u/, and /a/ with and without a biteblock. However, for the deaf subject, the cross-dimensional differences between the two speaking conditions were much

larger for the anterior portions of the tongue than for the hearing subject. This difference may indicate a deviant strategy for constraining tongue movement.[2]

Deviant tongue activity may have at least three consequences for intelligibility. First, it may contribute to some linguistic errors, including vowel and even consonantal errors, since deviant tongue movement may affect the acoustic signal reflecting open and closure postures, and transitional movements. Second, it might contribute to the distinctive speech quality of deaf talkers. Listeners cannot determine whether isolated vowels have been spoken by deaf or normal talkers, but perform very well when listening to sentence samples (Calvert, 1962). This suggests that articulatory dynamics factor into overall speech quality. As such, the aberrant tongue movement and how it structures the speech signal may contribute to listeners' ability to identify talkers as deaf. Finally, the difference in tongue behavior might permeate the speech signal, and decrease listeners' ability to recognize specific speech events, because there is not an appropriate context in the signal.

Stress Patterning

The problem of poor speech rhythm, so often noted in deaf talkers (Osberger and McGarr, 1983), is likely "related to a number of factors. These include faulty stress patterning (Nickerson, 1975), aberrant speech respiratory behavior (Forner and Hixon, 1977), inappropriate modulation of fundamental frequency (Martony, 1968; Sussman and Hernandez, 1979), frequent and inappropriate pauses (Boone, 1966; Heidinger, 1972; John and Howarth, 1965), and different glossal behavior (Tye-Murray, 1987) (Tye-Murray and Folkins, 1990, pg. 2682)". Two of our studies have focused on one aspect of speech rhythmicity, stress patterning. One employed strain gauge (Tye-Murray and Folkins, 1990) and the second employed cinefluorography (Tye-Murray et al., 1987).

In the first study, we considered whether stress patterning errors arise because deaf talkers are unfamiliar with the stress patterns of their language (hence, the errors should be viewed as a linguistic deficit) or whether they do not have the articulatory organization which allows them to distinguish between stressed and unstressed syllables, and to produce an intended stress pattern correctly (a motoric deficit). We posed three experimental questions: 1) Can deaf talkers produce desired stress patterns after instruction?; 2) Can they produce different articulatory behaviors for stressed and unstressed syllables and does a distinction occur for both short (e.g., TAP tap TAP tap) and long stress patterns (e.g., TAP tap tap TAP tap....)?; and 3) Does their ability to produce stress patterns change with variations in phonetic composition (e.g., PA pa PA pa versus PLOP plop PLOP plop)? If deficits in articulatory organization contribute to the stress errors, then as strings become more challenging to produce, either as a result of more intricate rhythm patterns or increased number of phonetic segments, deaf talkers might be less able to distinguish between stressed and unstressed syllables, and segmental errors might increase.

Three deaf and three hearing adults served as subjects. Strain gauge transduction of lower lip and jaw movement showed that both groups of subjects produced larger displacements and

[2]As an aside, it is interesting to note that "auditory representations" (Ladefoged, DeClerk, Lindau, and Papcun, 1972; Gay, Lindblom, and Lubker, 1981) of speech sounds are not necessary to achieve similar points of constriction when the jaw is fixed, since the congenitally deaf subject, who had no such representations, was able to do so.

longer durations for the stressed than unstressed syllables, regardless of the stress pattern. The timing of jaw activity throughout a stress pattern did not become more variable with changes in phonetic composition, and segmental errors did not increase. Listeners were consistently able to identify syllables as stressed and unstressed. These findings offer no evidence that deficits in articulatory organization limit deaf talkers in producing familiar stress patterns.

In the second study, two prelingually deaf and two hearing talkers produced strings of alternating heterogeneous syllables (e.g., tube knock tube knock...) as though speaking in time with a metronome. When hearing talkers produce them, the syllables in the strings depart systematically from acoustic isochrony (Morton, Marcus, and Frankish, 1976). The intervals between onsets of syllables are longer when the first syllable of a pair has a longer initial segment. For example, the interval between the acoustic onset of knock and tube would be longer than that between tube and knock, because /n/ typically is longer in duration acoustically than /t/. Although the acoustic onsets of the syllables are unevenly timed, talkers may evenly space their articulatory movements toward and/or during the open posture steady-states (Fowler and Tassinary, 1981).

Examination of the cinefluorographic records revealed that the deaf and hearing subjects performed similarly in timing their articulatory gestures for the syllable strings. Events associated with the establishment of open postures were evenly spaced, whereas the onsets of the syllables were not. These results corroborate the findings of Tye-Murray and Folkins (1990): when deaf talkers know a stress pattern, they seem capable of speaking it, at least, as indexed by the measures that we have employed in our investigations.

Producing Specific Articulatory Events

We have addressed two common assumptions about how deaf talkers produce specific articulatory events with our research. They are first, that deaf talkers establish an open posture with excessive jaw displacement and minimal tongue movement, and second, that deaf talkers tend to prolong segment durations.

Establishment of specific open postures of articulation

Many clinicians and researchers have hypothesized that deaf talkers establish an open posture ("vowel steady-state") with excessive jaw displacement and minimal tongue movement (Monsen, 1976; Osberger, 1987; Tye et al., 1983; Stevens, Nickerson, and Rollins, 1983; Zimmermann and Rettaliata, 1981). For instance, Ling (1976) observed, "Exaggerated movements of the jaw which lead to abnormal tongue and lip target behaviors occur with considerable frequency among hearing-impaired children (pg. 246)." Acoustic and perceptual data support this view. The second formants tend to cluster around a central value, 1800 Hz, regardless of the identity of the intended vowel (Monsen, 1976; Angelocci, Kopp, and Holbrook, 1964). This is indicative of a relatively immobile tongue body, since second formant values are determined largely by anterior/posterior tongue displacement in the vocal tract (Monsen, 1976). Perceptually, the various vowels produced by deaf talkers sound similar, and have a schwa-like quality (Hudgins and Numbers, 1942; Smith, 1975).

We have conducted three investigations that pertain to whether deaf talkers use excessive jaw displacement, and do not move their tongue bodies when producing an open posture. The first employed cinefluorography (Tye et al., 1983), the second employed strain gauge (Tye-Murray and Folkins, 1990), and the third, cinefluorography and microbeam (Tye-Murray, 1991).

In the first investigation, the cinefluorographic records indicated that the two prelingually deafened subjects had similar tongue positions for the vowels /u/ and /ae/, unlike the postlingually deafened and hearing subjects. This finding is consistent with suggestions of that deaf talkers do not move their tongue bodies much when establishing an open articulatory posture (see also Boone, 1966; Crouter, 1963; Kaiser, 1959; Stein, 1980). However, the results of our second two investigations suggest a more complex description may be appropriate for characterizing aberrations in jaw and tongue displacement patterns.

In the second investigation, deaf and hearing subjects demonstrated similar magnitudes of jaw displacement. This finding conflicts with the view of excessive jaw displacement.

In the third investigation, the trajectories (defined as displacements in an X-Y spatial plane) of the tongue body as it moved to a variety of different vowel steady-state positions were examined. Jaw displacement was also measured. Three deaf subjects and two hearing subjects participated. The deaf subjects displaced their tongues during the opening gestures, so they could not be characterized as having an immobile tongue body. Moreover, the deaf subjects did not speak with excessive jaw displacement in an absolute sense. Rather, they displaced their jaws excessively relative to their tongue body displacements. The hearing subjects displaced their tongue bodies at least twice as much as their jaws when establishing the open postures whereas the deaf subjects demonstrated similar amounts of displacement for both articulators.

In this third investigation, even though the deaf subjects displaced their tongues, the lingual movement trajectories qualitatively differed from those of the hearing subjects. Namely, the hearing subjects produced different tongue trajectory patterns for different vowel contexts; for example, they moved their tongue body anteriorly and superiorly in the vocal tract when producing words with /i/ and posteriorly and superiorly when producing words with /u/. In contrast, the deaf subjects contracted their tongue muscles such that the tongue body moved similarly for all the vowel contexts. (Dagenais and Critz-Crosby, 1992, recently reported findings compatible with these.) Figure 2 captures the distinction in tongue behavior. It shows the displacement patterns of small gold pellets that were placed upon the tongue dorsum of the hearing subjects and two of the three deaf subjects. The patterns were produced while the subjects spoke the phrase, "That's the sew". For the hearing subjects, the tongue dorsum marker shows extensive movement in both the anterior/posterior and superior/inferior planes. The extensive movement is appropriate, since the tongue body must move anteriorly and superiorly when producing the high

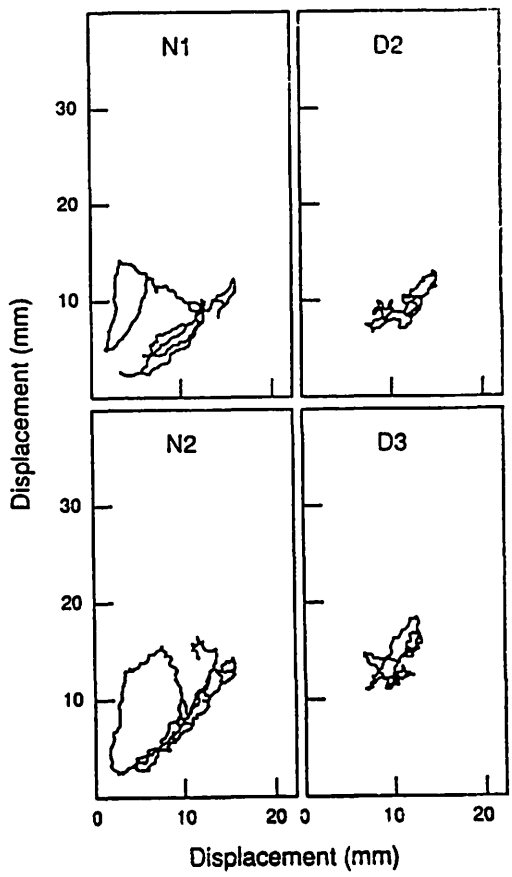

Figure 2. Displacement patterns of small pellets placed on the tongue dorsum. The effects of jaw displacement on the patterns have been subtracted. H1 and H2 have normal hearing, D2 and D3 are deaf.

front vowel /ae/ in That's, inferiorly and centrally when producing the central vowel /schwa/ in a, and anteriorly-then-posteriorly and superiorly when producing the diphthong /ou/ in sew. The lingual trajectories for the deaf subjects differ markedly from those of the hearing subjects. Whereas the pellets moved, the movements seem stereotyped and present a superior/inferior zigzag pattern.

The results of these three experiments can be related to some of the linguistic errors that have been noted for deaf talkers. Equivalent magnitudes of jaw and tongue body displacement, in conjunction with stereotypical tongue movement, may contribute to the neutral-sounding vowels produced by some deaf talkers. The measures of displacement from the third study also underscore the importance of relating tongue behavior to jaw activity, since data concerning relative displacement patterns may be critical for identifying aberrations.

We have evidence that the introduction of some hearing capability leads to the development of more distinctive tongue posturing for different vowels. Tye-Murray and Kirk (Submitted) collected spontaneous speech samples from eight deaf children on five occasions: before receiving a Cochlear Corporation Nucleus cochlear implant, 6 months after receiving the cochlear implant, 12 months, 18 months, and 24 or 36 months. The accuracy with which subjects could produce vowel and diphthongs was determined by referencing transcriptions of the spoken messages to transcriptions of the simultaneously signed messages (the children use total communication). The analyses revealed two trends: 1) with cochlear implant experience, the vowel and diphthong repertoires expanded and became more diverse and 2) vowel and diphthong production became more accurate. The investigators speculated that increased access to the first and second formants, information that is specifically coded by the cochlear implant, helped the children refine their motoric ability to distinguish between open articulatory postures.

Segment durations associated with specific speech events

The speaking rate of deaf talkers is 1.5 to 2.0 times slower than that of hearing talkers (Boone, 1966; Boothroyd, Nickerson, and Stevens, 1974; Heidinger, 1972; Hood, 1966; John and Howarth, 1965; Voelker, 1938). These studies might indicate that articulatory gestures associated with the open phases of articulation are longer in duration than those of hearing talkers (Boone, 1966; John and Howarth, 1965; Markides, 1970; Osberger and Levitt, 1979; Rothman, 1977). However, in three studies, one with cinefluorography (Tye-Murray, 1984), one with microbeam (Tye-Murray, unpublished), and one with strain gauge (Tye-Murray and Folkins, 1990), we have not found evidence to support this interpretation.

In the first study, two hearing and three deaf talkers spoke strings of syllables. Two of the deaf subjects were prelingually deafened and one was postlingually deafened. Subjects spoke the strings with a conversational and fast speaking rate. The open posture durations for the syllables did not differ between the two groups (see also Stein, 1980), although they did vary with speaking rate.

In the second study, data were collected to examine the durations of opening gestures, closing gestures, open posture steady-states, and the durations of initial consonant closures. The results appear in Figure 3. On average, the deaf subjects, particularly D3, maintained the initial consonant closure postures for a longer interval than the two hearing subjects. D2 had similar opening gesture durations as H2, and D3 had similar opening gesture durations as H1. D3 maintained the shortest open steady-state postures while the durations for D1, H1, and H2 were similar. The deaf subjects produced considerably longer closing gesture durations than either hearing subject.

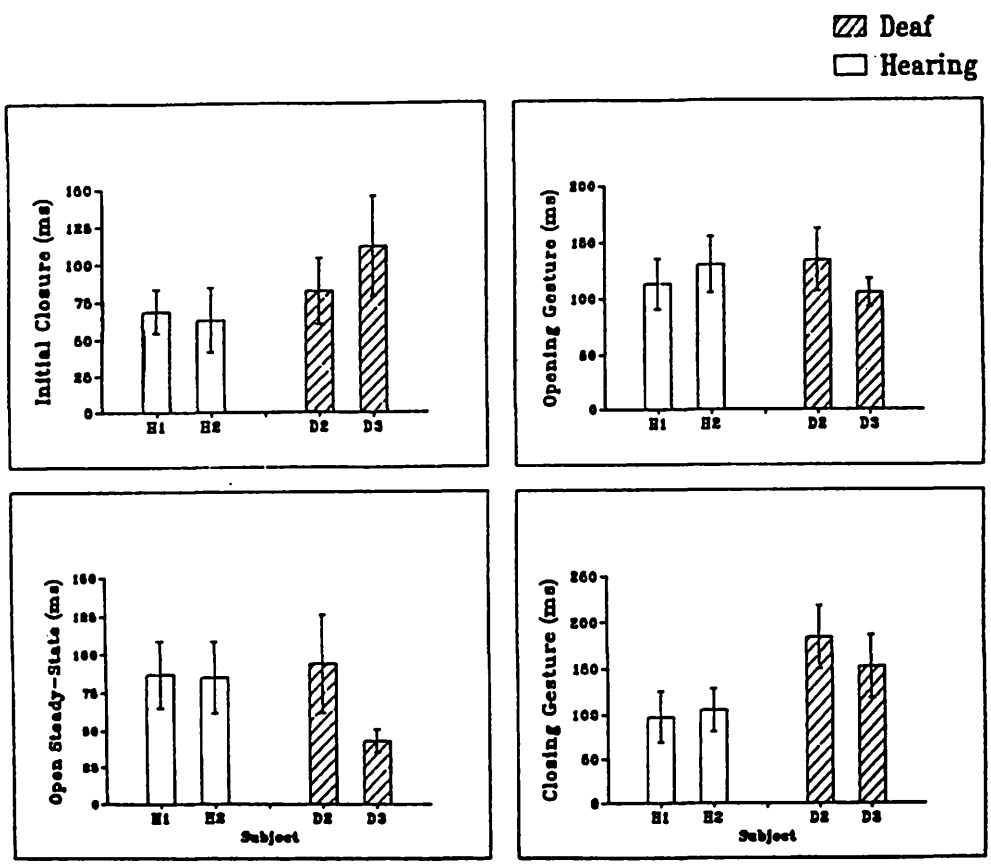

Figure 3. Movement event durations for syllables produced by deaf and hearing talkers. Data represent measures from 11 different consonant-vowel-consonant syllables (such as dean, tune, bob) produced two times each. Bars indicate standard deviations. The subjects in this investigation were those studied in Tye-Murray,1991.

Perceptually, the effect of vowel steady-state duration varies with syllabic context (Derr and Massaro, 1980; Rakerd, Verbrugge, and Shankweiller, 1980; Raphael, Dorman, and Freeman, 1975). For example, a 40 ms acoustic steady-state interval embedded in a 300 ms syllable may create a different percept than one embedded in a 150 ms syllable. For this reason, in the second study, a ratio was computed between the open posture steady-state duration and both the corresponding open cycle duration and the total duration for each syllable. Open cycle duration was the sum of the opening gesture, open steady-state, and closing gesture durations. Total duration was the open cycle duration plus the initial consonant closure duration. The results appear in Table 1. The open posture steady-states comprised a smaller fraction of the open cycle and total syllable duration for the two deaf subjects than for the two hearing subjects. For example, D3's open posture steady-state durations comprised 15% and 11% of the open cycle and the total syllable duration, respectively, as compared to 27-31% and 23-25% for the two hearing talkers. This finding indicates that the open steady-state postures may actually be too brief for some deaf talkers.

In the third study, jaw and lower lip durations as well as displacements and peak velocities during opening and closing articulatory gestures were measured for monosyllables. The measures were similar for deaf and hearing subjects, even though the deaf subjects' syllables were unintelligible. Interestingly, they vocalized for greater durations when producing syllables than did the hearing subjects, although their movement durations were similar. A group of listeners transcribed

their /t/ and /p/ segments as /d/ and /b/. Apparently, voicing that is prolonged relative to the articulatory gestures may underlie many perceptually-defined voicing errors.

An important point emerges in the analyses reported in Table1 and in the strain gauge study (Tye-Murray and Folkins, 1990). When considering segment durations (and other kinematic variables such as displacement), it is wise to do so with reference to the broader context in which the segment occurs. For example, there may be a scenario in which a deaf talker prolongs an acoustically-defined vowel steady-state. However, if other segments in the speech output are prolonged, then it might be unwise for a clinician to attempt to alter vowel duration without considering the ramifications for overall intelligibility. Another reason for considering relative measures is that absolute values may not always relate to whether an intended utterance will be perceived correctly, as indicated in the strain gauge analyses.

As a final note, these results were obtained from consonant-vowel-consonant syllables spoken in phrase or string context. It is possible that different results will surface with consonant-vowel syllables, or syllables spoken in isolation.

Table 1. Averages of the ratios computed between the open posture steady-state durations and the corresponding open cycle and syllable durations.

Subject	Open posture steady-state duration/ open cycle duration	Open posture steady-state duration/ syllable duration
(Hearing)		
H1	0.31 (0.16)	0.25 (0.14)
H2	0.27 (0.16)	0.23 (0.13)
(Deaf)		
D2	0.23 (0.16)	0.19 (0.13)
D3	0.15 (0.05)	0.11 (0.04)

Note: For each subject, values are combined for approximately 22 syllables. Standard deviations appear in parentheses.

The Roles of Auditory Information

Now that the experimental results have been reviewed, let us consider the numerous ways in which audition influences how talkers develop and maintain their speech. I will begin by making three rather disjointed observations. First, children with normal hearing acquire speech with remarkable ease and rapidity. Most three-year old children can clearly articulate sentences with five or more words, without receiving overt instruction. Second, children with minimal or no hearing must expend an enormous amount of time and effort toward attaining functional speech, a goal that most never achieve. Finally, deaf talkers often exhibit predictable intelligibility deficits, such as neutral sounding vowels and poor rhythmicity, regardless of where in the world they receive speech

instruction (e.g., United States, Hudgins and Numbers, 1942; Holland, Kaiser, 1959; Britain, Markides, 1970; Czechoslovakia, Handzel, 1956; Sweden, Martony, 1965). From these three observations, there arises an almost inescapable conclusion. Namely, children have an innate propensity for developing speech, and with rare exceptions, a propensity that is actualized only with exposure to the acoustic speech signal. Moreover, the absence of audition often leads to characteristic speech problems.

This state of affairs presents a major challenge to speech-language pathologists, teachers, parents, and deaf children, who at some point must confront such questions as, can the articulatory behaviors which underlie the intelligibility deficits be modified? And, can we ever expect more than a few deaf children to acquire functional speech? The answers to such questions are not straightforward. However, data that clarify the roles of auditory information may provide guidance about how to target remediation efforts, and may lead to more effective and expedient speech intervention procedures.

Five Roles

There are least five ways in which auditory information influences speech development. In this review, the focus has been placed upon two. First, our experimental results suggest that auditory information potentiates the development of specific strategies of articulatory organization. For example, by listening to the speech of others, children may learn how to impose modulations upon an open posture-to-open posture cycling. Second, auditory information may influence how children learn to perform specific speech gestures. For instance, they may learn to displace the tongue body relatively more than the jaw in establishing an open posture.

Our experimental results also hint toward a third way in which audition may influence speech development. Experience with the auditory signal leads young children to develop a system of phonological performance and to learn linguistic rules (Monsen, 1974). Through listening, they learn the phonemes of the language, they learn which phonemes may and may not occur in series (e.g., /tn/ does not occur in initial syllable position in English), and they learn how variations in prosody may impart different meanings to an utterance. The results of Tye-Murray and Folkins (1990) suggest that at least some stress patterning errors stem from deficits in linguistic knowledge rather than deficits in the principles of articulatory organization. When deaf talkers know a particular stress pattern, they can produce it. Similarly, the results of Tye-Murray (1991) suggest that some deaf talkers do not learn to distinguish between vowel steady-states on the basis of tongue trajectory patterns. This may be due to a deficit in linguistic knowledge. That is, deaf talkers may not distinguish between the vowels with tongue displacement because the vowels are not distinguished in their system of phonological performance. In addition or alternatively, the results may suggest a motoric problem, in line with the second role for audition considered above. For example, deaf talkers may experience difficulty in coordinating tongue movement with simultaneous jaw movement.

A fourth role that audition plays in the speech acquisition process is to provide information about the consequences of articulatory activity (Fry, 1966). Through babbling and experimentation, the child learns how various articulatory maneuvers relate to the auditory signal. These associations are probably prerequisite for the child to effect perceptually salient phonetic contrasts. For instance, by experimenting with slow and fast opening articulatory gestures, and learning the auditory conse-

quences of each, the child builds the foundation for later distinguishing between such sounds as /w/ and /b/, which are associated with slow and fast velocity movements, respectively.

Finally, auditory feedback may provide important information for monitoring and regulating ongoing speech production (see Tobey, in press, for an excellent review). Hearing talkers may use auditory feedback to calibrate their speaking behaviors and to recognize when the normal boundaries of articulation have been exceeded (Zimmermann and Rettaliata, 1981; Cowie and Douglas-Cowie, 1983). On this front, some deaf talkers may compensate by prolonging closure intervals to maximize proprioceptive information, and shortening open steady-state postures because they cannot hear the corresponding high-amplitude acoustic feedback.

Idiosyncratic Differences Amongst Talkers

Deaf talkers vary in their speaking skills, and many have idiosyncratic articulatory deficits. Thus, even though we might characterize the roles of audition, we cannot predict with accuracy how a particular deaf talker will speak. What can be assumed is that the absence of audition will inhibit the development of many articulatory behaviors that are commonly found in normal hearing talkers.

The existence of idiosyncrasies in the speech of deaf talkers suggests that latitude exists in how children compensate for deafness. As such, it is possible that teaching methods and an optimal speech environment can affect speech development. Qualities of the individual child also are influential. These qualities include how well the child uses other modalities such as vision for abstracting speech information, cognitive factors, and the child's personality and interest in communicating with speech. Very few persons have complete deafness. The presence of residual hearing can enhance speech development, especially if sensitivity exists for mid and high frequency stimuli.

Future Directions

We are currently studying how speech behaviors change after children receive cochlear implants, and how they might use the electrical signal to monitor their speaking. For instance, Tye-Murray (in press) found that some congentially deaf children respond to delayed auditory feedback after receiving a cochlear implant. She suggested that some children may develop the capability to monitor their own speech by attending to the electrical signal. We are also comparing young cochlear implant users' production and perception skills (Kirk, Tye-Murray, Fryauf-Bertschy, and Tyler, 1991). Children are asked to identify consonant-vowel stimuli using audition, and then asked to speak the items. Separate confusion matrices are constructed with their responses and analysed with Information Transmission Analysis (Miller and Nicely, 1955). Our preliminary analyses suggest that a correspondence exists between children's ability to perceive and produce such speech features as nasality and frication.

Finally, we are studying larger groups of deaf children. Our objectives are to learn more about how they produce specific articulatory events, and how deficits in linguistic knowledge contribute to the speech difficulties. Strain gauge and perceptual data are being collected.

Learning to speak without audition is a formidable task. It may be that this will change, with the advent of cochlear implants and sophisticated teaching aids, and as more information about speech production becomes available to guide teaching efforts. Time will tell, along with the coordinated efforts of researchers and clinicians.

Acknowledgments

This report was supported by research grants DC00242 and DC00976-01 from the National Institutes of Health/NIDCD. I thank Karen Iler Kirk for her editorial comments.

References

Angelocci, A. A., Kopp, G. A., & Holbrook, A. (1964). The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys. Journal of Speech and Hearing Disorders, 29, 156-170.

Boone, D. (1966). Modification of the voices of deaf children. Volta Review, 68, 686-694.

Boothroyd, A., Nickerson, R. S., & Stevens, K. N. (1974). Temporal Patterns in the Speech of the Deaf — A Study in Remedial Training. Cambridge, MA: Bolt Beranek & Newman.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. In C. Ewan & J. Anderson (Eds.) Phonology yearbook, 3 (pp. 219-252). Cambridge: Cambridge University Press.

Calvert, D. R. (1962). Deaf voice quality: A preliminary investigation. Volta Review, 64, 402-403.

Carney, P., & Moll, D. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. Phonetica, 23, 193-202.

Cowie, R., & Douglas-Cowie, E. (1983). Speech production in profound post-lingual deafness. In M. E. Lutman & M. P. Haggard (Eds.), Hearing Science and Hearing Disorders (pp. 183-230). New York: Academic Press.

Crouter, L. (1963). A cinefluorographic comparison of selected vowels spoken by deaf and hearing subjects. Unpublished master's thesis, University of Kansas, Kansas City.

Dagenais, P. A., & Critz-Crosby, P. (1991). Consonant lingual-palatal contacts produced by nor mal-hearing and hearing-impaired children. Journal of Speech and Hearing Research, 34, 1423-1435.

Derr, M. A., & Massaro, D. W. (1980). The contribution of vowel duration, Fo contour, and frication duration as cues to the /juz/-/jus/ distinction. Perception and Psychophysics, 27, 51-59.

Folkins, J. W., & Bleile, K. M. (1990). Taxonomies in biology, phonetics, phonology, and speech motor control. Journal of Speech and Hearing Disorders, 55(4), 596-611.

Forner, L., & Hixon, T. (1977). Respiratory kinematics in profoundly hearing-impaired speakers. Journal of Speech and Hearing Research, 20, 373-408.

Fowler, C. (1979). "Perceptual centers" in speech production and perception. Perception and Psychophysics, 25, 375-388.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic feet. Journal of Experimental Psychology: General, 3, 386-412.

Fowler, C.A. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics, 14, 3-28.

Fowler, C., & Tassinary, L. (1981). natural measurement criterion for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), Attention and performance, IX (pp. 521-535). Hillsdale, NJ: Lawrence Erlbaum.

Fry, D. B. (1966). The development of the phonological system in the normal and the deaf child. In F. Smith and G. Miller (Eds.), The genesis of language: A psycholinguistic approach (pp. 187-206). Cambridge, MA: MIT Press.

Gay, T., Lindblom, B., & Lubker, J. (1981). Production of bite-block vowels: Acoustic equivalence by selective compensation. Journal of the Acoustical Society of America, 69, 802-810.

Handzel, L. (1956). Acoustic analysis of vowels in deaf children by means of the "visible speech" apparatus. Folia Phoniatrica, 8, 237-245.

Heidinger, V. A. (1972). An exploratory study of procedures for improving temporal features in the speech of deaf children. (Doctoral dissertation.) Columbia University, New York.

Hood, R. B. (1966). Some physical concomitants of the perception of speech rhythm of the deaf. Unpublished Ph.D. dissertation, Standord University.

Hudgins, C. V., & Numbers, G. C. (1942). An investigation of the intelligibility of the speech of the deaf. Genetic Psychology Monographs, 25, 289-392.

John, J., & Howarth, J. N. (1965). The effect of time distortions on the intelligibility of deaf children's speech. Language and Speech, 8, 127-134.

Kaiser, L. (1959). Speech and handling of language in normal and abnormal Dutch school children. Folia Phoniatrica, 11, 219-229.

Kent, R., & Moll, K. (1972). Cinefluorographic analysis of selected lingual consonants. Journal of Speech and Hearing Research, 15, 453.

Ladefoged, P., DeClerk, J., Lindau, M., & Papcun, G. (1972). An auditory-motor theory of speech productions. UCLA Working Papers in Phonetics, 22, 48-75.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. Journal of the Acoustical Society of America, 51, 2018-2024.

Lindblom, B., & Sundberg, J. (1971a). Neurophysiological representation of speech sounds. Papers from the Institute of Linguistics, University of Stockholm, 2, 16-21.

Ling, D. (1976). Speech and the hearing-impaired child: Theory and practice. Washington, D. C.: The Alexander Graham Bell Association for the Deaf.

MacNeilage, P. F., & DeClerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. Journal of the Acoustical Society of America, 45, 1217-1233.

Markides, (1983). The speech of hearing-impaired children. Dover, New Hampshire: Manchester University Press.

Martony, J. (1965). Studies on speech of the deaf: I. Quarterly Progress and Status Report, 3, (pp. 16-24). Stockholm: Speech Transmission Laboratories, Royal Institute of Technology.

Martony, J. (1968). On the correction of the voice pitch level for severely hard of hearing subjects. American Annals of the Deaf, 113, 195-202.

Monsen, R. (1974). Durational aspects of vowel production in the speech of deaf children. Journal of Speech and Hearing Research, 17, 386-398.

Monsen, R. B. (1976a). Normal and reduced phonological space: The production of English vowels by deaf adolescents. Journal of Phonetics, 4, 189-198.

Morton, S., Marcus, J., & Frankish, C. (1976). Perceptual centers (P-centers). Psychological Review, 83, 405-408.

Nickerson, R. (1975). Characteristics of the speech of deaf persons. Volta Review, 342-362.

Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 39, 151-198.

Osberger, M. J. (1987). Training effects on vowel production by two profoundly hearing-impaired speakers. Journal of Speech and Hearing Research, 30, 241-251.

Osberger, M. J., & Levitt, H. (1979). The effect of timing errors on the intelligibility of deaf children's speech. Journal of the Acoustical Society of America, 66, 1316-1324.

Osberger, M. J., & McGarr, N. S. (1982). Speech production characteristics of the hearing impaired. In N. Lass (Ed.), Speech and language: Advances in basic science and research, (pp. 221-283). New York: Academic Press.

Rakerd, B., Verbrugge, R. V., & Shankweiller, D. P. (1980). Speaking rate, syllable stress and vowel identity. Haskins Laboratories Status Report on Speech Research, SR-62, 149-160.

Raphael, L. J., Dorman, M. F., & Freeman, F. (1975). Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies. Journal of Speech and Hearing Research, 18, 389-400.

Rothman, H. (1976). A spectrographic investigation of consonant-vowel transitions in the speech of deaf adults. Journal of Phonetics, 4, 129-136.

Rothman, H. (1977). An electromyographic investigation of articulation and phonetic patterns in the speech of deaf adults. Journal of Phonetics, 5, 369-376.

Smith, C. R. (1975). Residual hearing and speech production in deaf children. Journal of Speech and Hearing Research, 18, 795-811.

Stein, D. M. (1980). A study of articulatory characteristics of deaf talkers. (Doctoral dissertation, University of Iowa, 1980.) Dissertation Abstracts International, 41, 1327B.

Stevens, K. N., Nickerson, R. S., & Rollins, A. M. (1983). Suprasegmental and postural aspects of speech production and their effect on articulatory skills and intelligibility. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), Speech of the hearing impaired (pp. 35-51). Baltimore: University Park Press.

Sussman, H. M., & Hernandez, M. (1979). A spectrographic analysis of the suprasegmental aspects of the speech of hearing-impaired adolescents. Audiology and Hearing Education, 5, 12-16.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production and their effect on articulatory skills and intelligibility. In I. Hochberg, H. Levitt, and & M. J. Osberger (Eds.), Speech of the hearing impaired: Research, training, and personnel preparation (pp. 35-51). Baltimore: University Park Press.

Tobey, E. (In press). Speech production. In R. S. Tyler (Ed.), Cochlear implants: Audiological foundations. San Diego: Singular Publishing Group, Inc.

Tye, N., Zimmermann, G. N., & Kelso, J. A. S. (1983). "Compensatory articulation" in hearing impaired speakers: A cinefluorographic study. Journal of Phonetics, 11, 101-115.

Tye-Murray, N. (1984). Articulatory behavior of deaf and hearing speakers over changes in rate and stress: A cinefluorographic study. (Doctoral dissertation, University of Iowa, 1984.) Dissertation Abstracts International, 45, 2128B.

Tye-Murray, N. (1987). Effects of vowel context on the articulatory closure postures of deaf speakers. Journal of Speech and Hearing Research, 30, 90-104.

Tye-Murray, N. (1991). The establishment of open articulatory postures by deaf and hearing talkers. Journal of Speech and Hearing Research, 34, 453-459.

Tye-Murray, N. (In press). Young cochlear implant users' response to delayed auditory feedback. Journal of the Acoustical Society of America.

Tye-Murray, N., & Folkins, J. W. (1990). Jaw and lip movements of deaf talkers producing utterances with known stress patterns. Journal of the Acoustical Society of America, 87, 2675-2683.

Tye-Murray, N., & Kirk, K. I. (Submitted). Relationship between the Phonetic Level Evaluation and spontaneous speech: Vowel and diphthong production by young cochlear implant users.

Tye-Murray, N., & Woodworth, G. (1989). The influence of final-syllable position on the vowel and word duration of deaf talkers. Journal of the Acoustical Society of America, 75, 629-632.

Tye-Murray, N., Zimmermann, G. N., & Folkins, J. W. (1987). Movement timing in deaf and hearing speakers: Comparison of phonetically heterogeneous syllable strings. Journal of Speech and Hearing Research, 30, 411-417.

Voelker, C. H. (1938). An experimental study of the comparative rate of utterance of deaf and normal hearing speakers. American Annals of the Deaf, 83, 274-284.

Waldstein, R. S., & Baum, S. R. (1991). Anticipatory coarticulation in the speech of profoundly hearing-impaired and normally hearing children. Journal of Speech and Hearing Research, 34, 1276-1285.

Zimmermann, G. N., & Rettaliata, B. P. (1981). Articulatory patterns of an adventitiously deaf speaker: Implications for the role of auditory information in speech production. Journal of Speech and Hearing Research, 24, 169-178.

NCVS Status and Progress Report - 2
April 1992, 217-224

Vocal Efficiency

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

Abstract

Vocal efficiency is discussed in terms of energy conversion principles. Metabolic power is computed and compared with aerodynamic power, acoustic power, and the power dissipated in vocal fold tissues during phonation. A distinction is made between vocal efficiency and glottal efficiency, which has a more restricted definition. Vocal efficiency should ultimately include estimates of muscular ergometrics and a cost/benefit criterion for increased mechanical stress due to forceful contact and collision of vocal fold tissues.

Introduction

The concept of efficiency is integrally connected with energy conversion processes. In physical systems, energy is identified in a number of forms, including mechanical, electrical, thermal, chemical, and acoustic. All of these forms and others that are less relevant for this discussion (e.g., nuclear or electromagnetic energy) exist in the human body. As a highly sophisticated energy conversion machine, the human body absorbs energy in one form and releases it in another form. How efficiently it does this depends on a number of considerations.

First, as a multi-purpose machine, the human body has more than one useful output. Since efficiency is defined, in a very generic sense, as the output obtained for a given amount of input, all of the outputs should probably be considered simultaneously if global efficiency is under consideration. Typically, however, one selects a single output of interest for efficiency calculations (e.g., the vocal output). This automatically limits the efficiency of a multi-purpose machine, because byproducts are deemed less useful than the specific output of interest. In fact, the byproducts are often deemed losses.

Consider an example that does not involve the human body, but is analogous to many of its internal components. A simple light bulb is an energy conversion system. Electric energy is the input and light energy is usually the selected output. The efficiency of the light bulb is extremely

low if all other forms of radiation (in particular infrared, or heat) are considered to be losses. If heat is included as part of useful output, however, the efficiency approaches 100%. To complicate matters, the choice of useful output may vary from season to season. Thus, in the winter the thermal byproduct may be useful, whereas in the summer it may be a definite loss. It will be seen that efficiency in voice production has similar ambiguities with regard to the usefulness of its byproducts.

Another complication is that efficiency in human performance may not be measurable in purely physical terms, but may include psychophysical dimensions. Particularly in artistic performance, the *perception* of ease, fluency, and coordination is often more relevant to the observer than the balance of raw energy. A vocalist or a ballet dancer can create illusions of ease that do not translate directly into conservation of energy. This illusion of ease may actually produce more wear and tear on the body than some apparently more effortful movement.

This brings us to yet another consideration of efficiency: preservation and longevity. A machine may be very efficient in converting energy from one form to another, but wear out its structural components in the process. This invokes the concept of economy. If the machine is replaceable, less efficiency can be weighed against greater longevity and an overall cost/performance ratio computed. If the machine is not replaceable, as the human body is thought to be in most circumstances, longevity may take precedence over short-term efficiency. Hopefully, the two go hand in hand, but there is no guarantee that they do. The entire field of ergonomics is devoted to such issues.

We now bring the above considerations into focus by discussing vocal efficiency in terms of some ergometric principles, as well as in more traditional terms.

Some Efficiency Calculations

Assume the human body were designed strictly for mechanical output (e.g., lifting, or turning a crank). Energy input would be, as always, in the form of chemical energy derived from food consumption at an average rate of 2000 kilocalories per day. Recognizing that 1 calorie is equivalent to 4.19 joules of energy and 1 day is 86,400 seconds, a simple division reveals that energy input is at an average rate of approximately 100 joules/s, or 100 watts. Thus, if the body were 100% efficient in converting chemical energy into mechanical energy (which it is not), an average human could turn a generator to keep a 100 watt light bulb burning continuously. Alternatively, a 5 kg mass could be lifted up and down 1 meter against gravity in a periodic fashion at a rate of 1 Hz. At a more realistic 10% efficiency, the continuous output for lifting would be 0.5 kg (1 lb.) and the continuous turning of a generator shaft would keep a dim 10 watt light bulb glowing.

Peak output can, of course, be much higher. An olympic weight lifter can exceed the continuous output rating several hundred fold. This suggests that the body is an excellent energy storage device, allowing great surges of power to be withdrawn over short time intervals. The reservoir of energy is replenished on a continual basis through metabolic processes.

Consider now the aerodynamic power available from the lungs. We might ask, how much power can a person produce by huffing and puffing (like the big bad wolf who blew down the houses of the three little pigs)? With the glottis and mouth wide open, assume that the entire vital capacity (about .005 m^3) were to be expelled in a one second puff with a lung pressure of 2,000 Pascals (about 20 cm H_2O). The aerodynamic power in watts would be the product of the lung pressure in Pascals and the mean flow in m^3 per second, which is 10 watts. Thus, for a few brief

moments, one could possibly blow a windmill to light the same 10 watt light bulb that could be kept glowing continuously by turning a crank with the hands. Hyperventilation and respiratory fatigue are obvious limiting factors.

In phonation, glottal resistance limits the flow to less than a tenth of the value computed for puffing. Typical mean flows are .0001-.0005 m^3/s (100-500 cm^3/s). In this range of flows, the aerodynamic power is on the order of 1 watt, unless the subglottic pressure is raised considerably above 20 cm H_2O. As a standard in voice science, it may be appropriate to compute all speech and aerodynamic powers in dB relative to 1 watt, the approximate maximum raw aerodynamic power in speech or song. As an interesting side comment, we note that this maximum aerodynamic power is about 1% of the total metabolic power of the human body.

A. Glottal Efficiency

A most interesting topic in voice science is the distribution of raw aerodynamic power among components that may be considered useful output and those that are deemed losses. A number of papers have dealt with various facets of this problem (1,6,8,44,48). One obviously useful component is the radiated acoustic power. This power can be derived from intensity measurements on human subjects (44), or it can be calculated from basic principles of acoustic radiation of sound from idealized sources (48). Both approaches yield estimates of 10^{-6} to 10^{-2} watts, depending on the source strength (peak flow), fundamental frequency, and glottal waveshape. The theoretical results from idealized sources indicate that

$$P_r = u_m^2 f_o^2 G\,, \tag{1}$$

where p_r is the radiated power, u_m is the peak AC glottal flow, f_o is the fundamental frequency, and G is a complicated function that includes a number of physical constants. Fortunately, G does not vary over as wide a range as u_m^2 or f_o^2. We see that four orders of magnitude in p_r can be obtained by realizing that u_m and f_o can both vary over about a factor of 10 (50-500 cm^3/s and 50-500 Hz for males).

It is clear from these estimates that radiated acoustic power is not a large portion of the aerodynamic power. If *glottal efficiency* E_g is defined in the traditional sense as the ratio p_r/p_a, where p_a is the aerodynamic power, then this efficiency ranges from zero to perhaps as high as 1% (44). Questions of immediate interest are: [1] where does the rest of the aerodynamic power go, [2] what does the upper limit of this glottal efficiency depend on, and [3] is this the only meaningful measure of vocal efficiency that can be defined? There are no complete answers to any of these questions, but some insights can be obtained by considering the power delivered to the vocal folds, which might yield an estimate of some of the losses.

B. Power Losses

Consider the power transferred from the airstream to the vocal folds. This is approximated by the product of the mean force against the tissue (mean pressure times surface area) and the mean velocity of the tissue.

$$Pf = P_g LT\dot{x}\,, \tag{2}$$

where Pg is the mean glottal driving pressure, L is the glottal length, T is the vocal fold thickness, and $\dot{x}$ is the mean velocity of the tissue in the lateral direction. If we assume that the mean driving pressure is on the same order of magnitude as the subglottic pressure (1 kPa, or about 10 cm H_2O), LT is on the order of 1 cm^2, and $\dot{x}$ is on the order of 1 m/s (1 mm vibrational amplitude traversed in 1 ms, a quarter period of a 250 Hz oscillation), then the power to the vocal folds is estimated to be on the order of 0.1 watt. This is an appreciable portion of the previously estimated maximum aerodynamic power (1.0 watt). More generally, the aerodynamic power can be written as

$$P_a = P_s U = P_s a_g v , \tag{3}$$

where P_s is the mean subglottic pressure, U is the mean glottal flow, a_g is the mean glottal area, and v is the mean air particle velocity. Note that the driving power of the vocal folds p_f and the power in the airstream p_a are both proportional to a surface area and a velocity. For p_f, the surface area is the medial surface of the vocal folds, where for p_a the surface area is the glottal area. The ratio LT/a_g would typically be on the order of 10:1. Conversely, the ratio of tissue velocity to air particle velocity $\dot{x}/v$ is on the order of 1:10, making the two powers of comparable size. It is clear, of course, that p_f must always be less than p_a in order to maintain energy balance and vocal fold oscillation. The power consumed by the vocal folds can be reduced by reducing the tissue viscosity, i.e., by maintaining the vocal folds in a hydrated state (13,49).

Another major consumer of aerodynamic power is air turbulence at glottal exit. Jet formation in the ventricular region causes a reduction in pressure without a concomitant increase in air particle velocity (19,43,47). The separation of the airstream from the vocal tract wall results in eddy currents, which dissipate aerodynamic energy. Although it has been shown that this is a major loss factor for steady flow conditions (43), it is not clear that pulsatile flow is subject to the same degree of energy loss. Thus, it is difficult to estimate the magnitude of the turbulent losses at this time.

Finally, viscous losses and wall vibration losses occur all along the vocal tract (above and below the glottis) as acoustic waves propagate along the airway (33). These losses contribute toward the bandwidth of the formants, but are likely to be small in comparison to the two major glottal losses discussed above.

At this point, there is an insufficient amount of knowledge about the losses to be able to predict an upper limit of vocal efficiency. Could a highly trained singer reduce tissue and air losses to a degree that 10%-50% of the aerodynamic power would be converted to radiated acoustic power? This is an interesting question that deserves some intense research.

Some Problems with Glottal Efficiency Definitions and Calculations

One of the major problems with the traditional glottal efficiency calculation is the strong dependence of E_g on fundamental frequency f_o. Dividing equation [1] by equation [3], we get

$$E_g = u_m^2 f_o^2 G / P_s U , \tag{4}$$

which shows an f_o^2 dependence. Because high frequencies are radiated much more effectively than low frequencies, the traditional efficiency calculations will generally favor high-pitched vocal productions, even though they may be forced or strained in relation to low-pitched productions. Likewise, loud productions will be favored over soft productions because E_g increases at a rate of

about 3 dB per doubling of subglottic pressure. This can be seen from equation [4] by realizing that u_m, the maximum AC flow, is roughly proportional to P_s (for more detail see 48). Some normalization procedure of E_g with respect to f_o and P_s should be considered. For example, if pre- and post-intervention measurements include vocal efficiency, the f_o and SPL levels should be specified a priori and held constant. Otherwise, an improvement in vocal efficiency may simply reflect a different choice of pitch and loudness by the singer.

Another problem with the traditional E_g calculation is that it does not include a cost/benefit ratio for vocal health and longevity. A pressed voice is usually more efficient (aerodynamically and acoustically) than a slightly breathy voice. But what price is paid for forceful adduction? An evaluation of contact stress and its resulting trauma to tissue is needed to get a more complete picture of what might be called efficient vocal production. Furthermore, the levels of muscle contraction in extrinsic and intrinsic laryngeal musculatures should be assessed in some way. Since muscles are often paired as agonists and antagonists, equivalent posturing can be obtained with high and low levels of contraction. For example, a given vocal fold length can be maintained (and also a given pitch) while cricothyroid and thyroarytenoid muscle activity both increase in some proportion (50). This can make the entire phonation process more effortful, with little or no gain in acoustic output. Adductor and abductor muscles may also be fighting each other with little effect on the shape of the glottis. The same can be said about respiratory muscles. Until some principles of muscle efficiency are included in laryngeal and respiratory activity, the picture of vocal efficiency will be incomplete.

Conclusions

As a phonation machine, the human body is very inefficient. Radiated acoustic power is between .0001% and 1% of the available aerodynamic power in phonation, which in turn is less than 1% of the metabolic power consumption of the entire body. Calculations of metabolic power for intrinsic and extrinsic laryngeal muscles (and respiratory muscles) are needed to define a more global vocal efficiency. In addition, a cost/benefit ratio is needed to assess the price a vocalist pays (in terms of tissue damage) for increased loudness and pitch, both of which increase the glottal efficiency of the voice.

References and Additional Bibliography

1. Berg, Jw. van den (1956). Direct and indirect determination of the mean subglottic pressure. Folia Phoniatr., 8, 1-24.

2. Berg, Jw. van den, Zantema, J.T., & Doornenbal, P. (1957). On the air resistance and the Bernoulli effect of the human larynx. J. Acoust. Soc. Am., 29, 626-631.

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

4. Berg, Jw. van den, & Tan, T.S. (1959). Results of experiments with human larynxes. Pract. Oto-rhino-laryngol., 21, 425-450.

5. Bouhuys, A., Proctor, D.F., & Mead, J. (1966). Kinetic aspects of singing. J. Appl. Physiol., 21, 483-495.

6. Bouhuys, A., Mead, J., Proctor, D.F., & Stevens, K.N. (1968). Pressure-flow events during singing. Ann. NY Acad. Sci., 155, 165-176.

7. Cavagna, G.A., & Margaria, R. (1965). An analysis of the mechanics of phonation. J. Appl. Physiol., 20, 301-307.
8. Cavagna, G.A., & Margaria, R. (1968). Airflow rates and efficiency changes during phonation. Ann. NY Acad. Sci., 155, 152-164.
9. Cavagna, G.A., & Camporesi, E.M. (1974). Glottic aerodynamics and phonation. In B. Wyke (Ed.), Ventilatory and phonatory control systems (pp. 76-92). London.
10. Damsté, P.H., & Lerman, J.W. (1975). An introduction to voice pathology, functional and organic. Springfield, IL: Charles C. Thomas Publishers.
11. Draper, M.H., Ladefoged, P., & Whitteridge, D. (1959). Respiratory muscles in speech. J. Speech Hear. Res., 2, 16-27.
12. Draper, M.H., & Ladefoged, P. (1960). Expiratory pressures and air flow during speech. Brit. Mod. J., 18, 1837-1843.
13. Finkelhor, B.K., Titze, I.R., & Durham, P.L. (1988). The effect of viscosity changes in the vocal folds on the range of oscillation. J. Voice, 1(4), 320-325.
14. Gould, W.J. (1971). Effect of respiratory and postural mechanisms upon action of the vocal cords. Folia Phoniatr., 23, 211-224.
15. Hirano, M., Koike, Y., & Leden, H. von (1968). Maximum phonation time and air usage during phonation (clinical study). Folia Phoniatr., 20, 185-201.
16. Hirano, M., Ohala, J., & Vennard, W. (1969). The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. J. Speech Hear. Res., 12, 616-628.
17. Hirano, M., Vennard, W., & Ohala, J. (1970). Regulation of register, pitch, and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles. Folia Phoniatr., 22, 1-20.
18. Hirano, M. (1975). Phonosurgery, basic and clinical investigations. (Partly in Japanese, translated with the help of M. Yanagida and Tj. de Graaf.) Otologica (Fukuoka), 21, Suppl. 1.
19. Ishizaka, K., & Matsudaira, M. (1972). Fluid mechanical considerations in vocal cord vibration. (SCRL Monograph No. 8). Santa Barbara, CA: Speech Communication Research Laboratory.
20. Isshiki, N. (1959). Regulatory mechanism of the pitch and volume of voice. Oto-Rhino-Laryngological Clinic, Kyoto, 52, 1065-1094.
21. Isshiki, N. (1961). Voice and subglottic pressure. Stud. Phonologica, 1, 86-94.
22. Isshiki, N. (1964). Regulatory mechanism of voice intensity variation. J. Speech Hear. Res., 7, 17-29.
23. Isshiki, N., & Leden, H. von (1964). Hoarseness: Aerodynamic studies. Arch. Otolaryngol., 80, 206-213.
24. Isshiki, N. (1965). Vocal intensity and air flow rate. Folia Phoniatr., 17, 92-104.
25. Iwata, S., Leden, H. von, & Williams, D. (1972). Air flow measurement during phonation. J. Com. Dis., 5, 67-79.
26. Kitzing, P., & Löfqvist, A. (1975). Subglottal and oral air pressures during phonation - preliminary investigation using a miniature transducer system. Md. Biol. Eng., 13, 644-648.
27. Koike, Y., Hirano, M., & Leden, H. von (1967). Vocal initiation: Acoustic and aerodynamic investigations of normal subjects. Folia Phoniatr., 19, 173-182.
28. Koike, Y., & Hirano, M. (1968). Significance of vocal velocity index. Folia Phoniatr., 20, 285-296.
29. Ladefoged, P. (1960). The regulation of subglottal pressure. Folia Phoniatr., 12, 169-175.
30. Ladefoged, P. (1963). Some physiological parameters in speech. Lang. Speech, 6, 109-119.

31. Ladefoged, P., & McKinney, N.P. (1963). Loudness, sound pressure, and subglottal pressure in speech. J. Acoust. Soc. Am., 35, 454-460.
32. Lieberman, P. (1968). Direct comparison of subglottal and esophageal pressure during speech. J. Acoust. Soc. Am., 43, 1157-1164.
33. Liljencrants, J. (1985). Speech synthesis with a reflection-type line analog. Ph.D. dissertation, Speech Communication and Music Acoustics Department, Royal Institute of Technology, S-10044 Stockholm, Sweden.
34. McGlone, R.E. (1963). An experimental study of a technique of intraesophageal pressure measurement as a method for estimating subglottic pressure. Doctoral thesis, University of Iowa, Iowa City, IA.
35. McGlone, R.E. (1967). Air flow during vocal fry phonation. J. Speech Hear. Res., 10, 299-304.
36. McGlone, R.E. (1970). Air flow in the upper register. Folia Phoniatr., 22, 231-238.
37. McGlone, R.E., & Shipp, T. (1971). Some physiologic correlation of vocal fry phonation. J. Speech Hear. Res., 14, 769-775.
38. Mead, J., Bouhuys, A., & Proctor, D.F. (1968). Mechanisms generating subglottic pressure. Ann. NY Acad. Sci., 155, 177-181.
39. Murry, Th. (1971). Subglottal pressure and airflow measures during vocal fry phonation. J. Speech Hear. Res., 14, 544-551.
40. Murry, Th., & Brown, W.S. Jr. (1971). Subglottal air pressure during two types of vocal activity: Vocal fry and modal phonation. Folia Phoniatr., 23, 440-449.
41. Perkins, W.H., & Koike, Y. (1969). Patterns of subglottal pressure variations during phonation. Folia Phoniatr., 21, 1-8.
42. Proctor, D.F. (1974). Breathing mechanics during phonation and singing. In E. Wyke (Ed.), Ventilatory and phonatory control systems (pp. 39-57), London.
43. Scherer, R. (1981). Laryngeal fluid mechanics: Steady flow considerations using static models. Ph.D. dissertation, University of Iowa, Iowa City, IA.
44. Schutte, H.K. (1980). The efficiency of voice production. Druk: Kemper/Groningen.
45. Sundberg, J. (1973). The source spectrum in professional singing. Folia Phoniatr., 25, 71-90.
46. Sundberg, J. (1977). The acoustics of the singing voice. Sci. Am., 236, 82-91.
47. Teager, H., & Teager, S. (1985). Active fluid dynamics voice production models, or There is a unicorn in the garden. In I.R. Titze & R.C. Scherer (Eds.), Vocal fold physiology: Biomechanics, acoustics, and phonatory control (pp. 387-401). The Denver Center for the Performing Arts. New York: Raven Press.
48. Titze, I.R. (1988a). Regulation of vocal power and efficiency by subglottal pressure and glottal width. In O. Fujimura (Ed.), Vocal physiology: Voice production mechanisms and functions - Vol. 2 (pp. 227-238). New York: Raven Press.
49. Titze, I.R. (1988b). The physics of small amplitude oscillation of the vocal folds. J. Acoust. Soc. Am., 83(4), 1536-1552.
50. Titze, I.R., Luschei, E., & Hirano, M. (1989). The role of the thyroarytenoid muscle in regulation of fundamental frequency. J. Voice, 3(3), 213-224.
51. Wyke, B.D. (1974a). Respiratory activity of intrinsic laryngeal muscles: An experimental study. (In B.D. Wyke (Ed.), Ventilatory and phonatory control systems, London.
52. Wyke, B.D. (1976). Laryngeal reflex mechanisms in phonation. Folia Phoniatr., 28, 528-537.
53. Yanagihara, N., & Koike, Y. (1967). The regulation of sustained phonation. Folia Phonatr., 10, 1-18.