NCVS Status and Progress Report
Volume 12/June 1998

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

Editorial and Distribution Information

Editor, Ingo Titze
Production Editors, Julie Stark and Julie Ostrem
Technical Editor, Martin Milder

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

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

Primary Sponsorship

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

Other Sponsorship

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

The Denver Center for the Performing Arts
Wilbur James Gould Voice Research Center
Department of Public Relations
Department of Public Affairs
Denver Center Media
Department of Development

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

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

The University of Illinois
Department of Speech and Hearing Science

NCVS Personnel

Administration

Central Office
Ingo Titze, Director
Julie Ostrem, Program Associate
Julie Stark, Secretary

Area Coordinators
Research - Ingo Titze
Training - Patricia Zebrowski
Continuing Education - Julie Ostrem
Information Dissemination - Thea Carruth

Advisory Board

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

Investigators, Affiliates and Support Staff

Fariborz Alipour, Ph.D.
Kristin Baker, Ph.D.
David Berry, Ph.D.
Alyson Beyler, B.S.
Florence Blager, Ph.D.
Diane Bless, Ph.D.
James Brandenburg, M.D.
Myrna Burt
John Butler, M.D.
John Canady, M.D.
Thea Carruth, M.P.H.
Geron Coale, M.A.
Linda D'Antonio, Ph.D.
Charles Davis, Ph.D.
Wendy Edwards, B.A.
Jeffrey Fields, B.M.
Eileen Finnegan, Ph.D.
John Folkins, Ph.D.
Charles Ford, M.D.
Curt Freed, M.D.
Amy Furness
Steven Gray, M.D.
Elizabeth Hammond, M.D.
Marilyn Hetzel, Ph.D.
Margaret Hoehn, M.D.
Henry Hoffman, M.D.
Richard Hurtig, Ph.D.
Bruce Jafek, M.D.
Michael Karnell, Ph.D.
Judith King, Ph.D.
David Kuehn, Ph.D.
Jennifer Lehnherr
Jeri Logemann, Ph.D.
Russel Long, M.S.
Erich Luschei, Ph.D.
Kathryn Maes, Ph.D.
Susan McNicholas, B.S.
Martin Milder, B.S.
Paul Milenkovic, Ph.D.
Jerald Moon, Ph.D.
Lea Ann Moriarty, B.A.
Chris O'Brien, M.D.
Lorraine Olson Ramig, Ph.D.
Julie Ostrem, B.S.
Randal Paniello, M.D.
Namrata Patil, M.D.
Annette Pawlas, M.A.
Kathe Perez, M.A.
Karen Rhue, M.D.
Donald Robin, Ph.D.
Robin Samlan, M.S.
Richard Schmidt, Ph.D.
Suzanne Segal, Ph.D.
Nicole Seurer, B.S.
Anne Smith, Ph.D.
Elaine Smith, Ph.D.
Marshall Smith, M.D.
Jennifer Spielman, M.A.
Julie Stark, A.A.
Elaine Stathopoulos, Ph.D.
Brad Story, Ph.D.
Edie Swift, M.S.
James Teslow, Ph.D.
Laetitia Thompson, Ph.D.
Sue Ann Thompson, Ph.D.
Ingo Titze, Ph.D.
Vernon Vail, B.S.
Jane VanDeWiel, B.S.
Katherine Verdolini, Ph.D.
Patricia Ward
Barbara Williams
Raymond Wood, M.D.
George Woodworth, Ph.D.
Patricia Zebrowski, Ph.D.
Lynn Zimba, Ph.D.
Nicole Zinna, B.S.

Doctoral Students

Renee Bogschutz, M.S.
Todd Brennan, M.S.
Roger Chan, B.S.
Elisa Huff, M.S.
Eric Hunter, M.S.
Melda Kündük, M.S.
Helen Sharp, M.S.
Susan Thibeault, M.S.

Postdoctoral Fellows

Michael Edgerton, D.M.A.
Katsuhide Inagi, M.D.
Aliaa Khidr, Ph.D.

Visiting Scholars

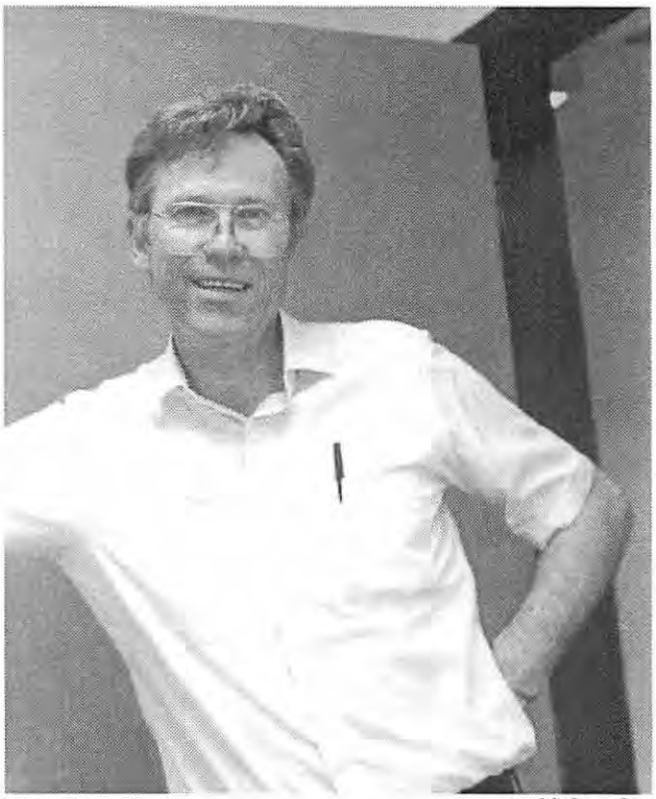
Hanspeter Herzel, Ph.D., Germany
Niro Tayama, M.D., Japan

Contents

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

Part II. Tutorial reports and updates

Foreword



Julie Ostrem photo

This volume of our Status and Progress Report is dedicated to Erich Luschei, who is retiring this summer from his university position as Professor of Speech Science at the University of Iowa. Erich has been a major player with the National Center for Voice and Speech, first as deputy director and principal investigator of a laryngeal neurophysiology project and second as the coordinator of doctoral and post-doctoral student training. When we wrote the initial grant application for the NCVS in 1989, Erich was the Michael Jordan in our team of writers. He did everything on the court and off the court, propped us up physically and emotionally, and stayed with the document right up to the final page numbering (more than 850 pages). No task was too big or too small for Erich Luschei.

Erich and several of his graduate students have contributed enormously to the research mission of the Center and to these progress reports. He taught us most of what we know about muscles and nerves. Although many of us still appear to have the orientation that the larynx and other speech organs are mere mechanical machines, Erich taught us early on that the nervous system is in charge. But he never went overboard to make claims, as others have, that the larynx on a laboratory bench (without neural and vascular attachments) is a useless object for investigation. In fact, he often questioned neurophysiologists, as vigorously as clinicians or engineers. He is not only a respecter of persons, but has a deep respect for ideas and common sense. He is a caring individual, a mentor *par excellence*.

We hope that as Erich and his wife, Nancy, travel a bit of the country and investigate new ways of focusing their energies, they will continue to give us samples of their style, philosophy, and affection that we have come to depend on.

Ingo R. Titze, Director
June, 1998

Part I

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 12
June 1998, 1-10

Estimation of Alveolar Pressure from Direct Measures of Tracheal Pressure during Speech

Eileen Finnegan, Ph.D.
Erich Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Department of Otolaryngology, University of Iowa Hospitals and Clinics
Henry Hoffman, M.D.
Department of Otolaryngology, University of Iowa Hospitals and Clinics

Abstract

The purpose of this study was to quantify upper and lower airway resistances during speech, in order to determine the extent to which modulation in tracheal pressure during phonation is due to respiratory or laryngeal adjustment. A model for understanding the relationship between respiratory drive, laryngeal resistance, and tracheal pressure is presented in this paper. Tracheal pressure and airflow were measured during sentence production in 6 subjects. Using a technique introduced in this paper, the resistance of the lower airway was measured Upper airway resistance was calculated from pressure and flow measures. The results indicate that upper airway resistances were relatively large with respect to lower airway resistances during phonation. This indicates that during phonation, the respiratory system functions as a pressure source and changes in tracheal pressure are due to alterations in respiratory drive, not to laryngeal adjustment.

Numerous investigators have measured P_t as an indicator of respiratory drive, while other investigators have attributed changes in P_t to laryngeal rather than respiratory adjustment. In excised canine larynges experiments (and in-vivo canine experiments that do not evoke phonation via midbrain stimulation), one of the controlled parameters is the aerodynamic power supplied by the lungs. In the past, some investigators have used a flow source to provide this artificial respiratory drive (Choi, Berke, Ye,, & Keiman, 1993; Choi, Ye, & Berke, 1995; Sercarz, Berke, Bielamowicz, Kreiman, Ye, & Green, 1994; Slavit & McCaffrey, 1995). With this model, modulation in tracheal pressure is achieved primarily by modulation of laryngeal resistance. More recently, researchers have used a constant pressure source (Nasri, Namazie, Kreiman, Sercarz, Gerratt., & Berke, 1994; Nasri, Namazie, Ye, Kreiman, Gerratt, & Berke, 1996; Titze, 1988; Verneuil, Kreiman, Kevorkian, Gerratt, & Berke, 1996) based on the belief that this provides a more accurate model of the respiratory system. With this model, changes in tracheal pressure are achieved by alteration in pressure provided by the pressure source. Although investigators involved in modeling of laryngeal behavior have begun to consider the respiratory system as a constant pressure source, researchers involved in human research, continue to interpret clinical data as if the respiratory system has characteristics of a flow source (i.e. as if tracheal pressure is influenced by changes in laryngeal resistance). Hirano, Ohala, & Vennard, (1969) suggested that "the importance of the vocalis muscle in changing intensity can be easily understood because this muscle changes the glottal resistance and consequently subglottic pressure" (p.626). Stathopoulos and Sapienza (1993) reported that two of the twenty subjects in their study increased P_t without increasing respiratory effort. They stated that, in these cases, increased estimated P_t was associated with changes in laryngeal parameters (i.e. decreased open quotient and increased maximum flow declination rate), but not in respiratory parameters (lung volume measures). Dromey, Ramig, & Johnson, (1995) reported that a person with Parkinson disease showed increased tracheal pressure following treatment and suggested this could be due in part to improved

valving efficiency of the larynx. There has been one study examining the degree to which tracheal pressure is determined by respiratory drive (Bouhuys, Proctor, & Mead, 1966). If P_t is considered the output pressure of the respiratory system, the source pressure of the respiratory system is the pressure produced in the alveoli of the lungs. That source pressure can be referred to as the alveolar pressure (P_a). Bouhuys et al. (1966) measured lung volume changes and esophageal pressures, using a body plethysmograph and an esophageal balloon, in 9 subjects. P_t was estimated by subtracting the static lung recoil pressure[1] from pleural pressure. Alveolar pressure was estimated by taking the difference between pleural pressure and static lung recoil pressure during phonation. They reported that, during sustained phonation, there were slight differences between subglottic pressure and estimated alveolar pressure, which increased slightly with decreasing lung volume. However, the results were not presented in a quantified manner and the data were collected during sustained phonation not during word or sentence production. The purpose of this study was to systematically quantify the extent to which changes in P_t are due to laryngeal or respiratory adjustment during phonation in the context of a sentence.

The relationship between P_t and P_a depends on the relative resistances above and below the trachea. Further, it is the relative value of these resistances that determine the extent to which changes in tracheal pressure are due to laryngeal or respiratory adjustment. In order to understand this concept, it is helpful to model the aerodynamics of the speech system as a "circuit" (see figure 1) in which alveolar pressure (P_a) is the pressure in the alveoli of the lungs and tracheal pressure (P_t) is the pressure measured below the vocal folds. Two resistances are included in the model. The resistance below P_t is the resistance offered by the lower airway, R_{law} (the resistances of the small diameter bronchioles, the bronchi, and the trachea). The resistance above P_t is the resistance offered by the upper airway, R_{uaw} (the resistance due to the vocal folds, epiglottis, tongue, and lips...). Note that during phonation, when the primary resistance of the upper airway is the larynx, R_{uaw} is essentially laryngeal airway resistance (R_{lx}). In this study, when the resistance of the upper airway is measured during sentence production, it is referred to as R_{uaw} since the resistance could be the result of changes in the position of the lips, tongue, or some other articulator in addition to any change in the larynx. If, however, the measure of upper airway resistance is made only during production of a vowel, then it is referred to as R_{lx}. In the following formulation, the more general R_{uaw} will be used. The relationship between alveolar pressure, tracheal pressure, and the upper and lower airway resistances can be stated mathematically.

P_a = pressure drop across lower airway + pressure drop across upper airway

$$P_a = (U * R_{law}) + (U * R_{uaw}), \text{ where } U = \text{airflow}$$

Rearrangement of the variables gives us,

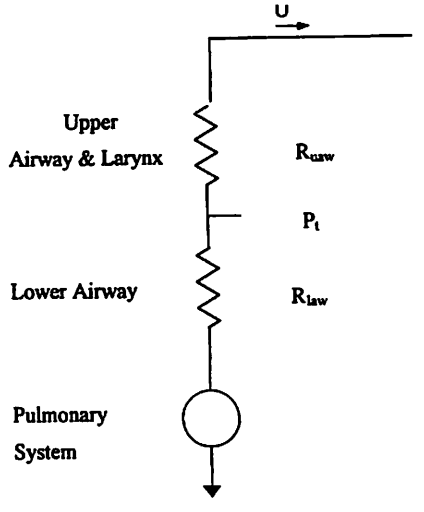
$$P_a = U (R_{law} + R_{uaw}),$$

Since $P_t = U * R_{uaw}$, we can substitute P_t / R_{uaw} for U

$$P_a = (P_t / R_{uaw}) (R_{law} + R_{uaw})$$

$$\text{so,} \quad P_t = P_a (R_{uaw} / (R_{law} + R_{uaw}))$$

Thus, the degree to which P_t approximates P_a depends on the relative value of R_{uaw} and R_{law}. The value of the ratio of these resistances, $R_{uaw} / (R_{law} + R_{uaw})$, can vary between 0 and 1. For example, if R_{law} was equal to zero, then the ratio, $R_{uaw} / (R_{law} + R_{uaw})$, would be equal to 1, and P_t would be the same as P_a. In this case, any change in P_t would be due entirely to change in P_a, and changes in R_{uaw} would have no effect on P_t. In this instance, the pulmonary system would be functioning



Model Components

U	airflow through system
P_t	tracheal (subglottal) pressure
P_a	alveolar pressure
R_{uaw}	upper airway resistance (due to larynx and supralaryngeal structures)
R_{law}	lower airway resistance (due to trachea, bronchi, and bronchioles)

Figure 1. Model of the aerodynamics of the speech system.

[1] Static lung pressure ($P_{st(l)}$) was obtained by measuring pleural pressure (P_{pl}) while the subject held his breath with the airway open. $P_{pl} = (-P_{st(l)})$. $P_{st(l)}$ was determined over a range of lung volumes.

as an ideal pressure source. At the other extreme R_{law} is infinitely large. In this second example, the ratio, $R_{uaw} / (R_{law} + R_{uaw})$, would be close to 0, and P_t would be a very small portion of P_a. Changes in P_a would have little effect on the absolute value of P_t but changes in R_{uaw} would have large effects on P_t. In this situation, the pulmonary system would be functioning as a flow source. A third example could be a case in which R_{law} and R_{uaw} were the same, then the ratio, $R_{uaw} / (R_{law} + R_{uaw})$, would be .5, and P_t would be half of P_a. The point of these examples is to illustrate that the extent to which the pulmonary system functions as a pressure source or a flow source or some combination of the two depends on the relative value of these two resistance (R_{uaw} and R_{law}). The degree to which P_t is determined by R_{lx} or P_a depends on whether the pulmonary system acts as a pressure or a flow source. It is possible to calculate R_{uaw} during speech from the tracheal pressure and airflow signals ($R_{uaw} = P_t /U$) and then if R_{law} could be determined, the $R_{uaw} / (R_{law} + R_{uaw})$ ratio could be calculated.

The primary factors determining the resistance of the lower airway (R_{law}) are the diameters of the airway passages. The trachea is a semi-rigid tube encircled by cartilaginous rings, which bifurcates to form the two bronchial tubes that extend toward the right and left lungs. Bronchioles branch off from these bronchi and with each subsequent branching become progressively smaller in diameter. The diameter of the trachea and bronchi does not change, but the diameter of the bronchioles, which do not have cartilage and are composed of smooth muscle, can change. The smooth muscles of the bronchioles are innervated by the sympathetic nervous system and when these muscles relax, the diameter of the bronchioles increases, decreasing airway resistance. Airway resistance is also increased in persons with chronic obstructive pulmonary disease (COPD) such as asthma, emphysema, or chronic bronchitis.

According to respiratory literature, during quiet breathing, R_{law} values can range from .07 to .28 kPa/(l/s). Phagoo, Watson, Silverman, & Pride (1995) compared several different methods for assessing airflow resistance in 7 normal adults. Median airway resistance measured with a body plethysmograph during panting was .13 kPa/(l/s). Median total lung resistance measured during quiet breathing using an esophageal balloon was .14 kPa/(l/s). Median total respiratory resistance measured during forced oscillation was .17 kPa/(l/s). Van Altena & Gimeno (1994) measured total respiratory resistance (which includes upper airway resistances as well as lower airway resistance) in 172 subjects using a technique involving transient interruption of airflow at the mouth during which alveolar pressure equilibrates with mouth pressure. They reported a group mean of .38 +/- .17 kPa/(l/s). Within subject variability, reported as a coefficient of variation, was 14.4 +/- 6.9%. Some investigators have found variation in total airway resistance associated with lung volume. Fisher, DuBois, & Hyde (1968) found mean respiratory resistance increased from .10 kPa/(l/s) at a lung volume of 6 liters to .35 kPa/(l/s) at a lung volume of 2 liters in 42 subjects with no history of respiratory problems.

As mentioned previously, during phonation, when the vocal tract is relatively open and upper airway resistance is primarily the resistance offered by laryngeal adduction, R_{uaw} is essentially R_{lx}. According to the speech pathology literature, during phonation, laryngeal airway resistance ranges from 2.3 to 6.6 kPa/(l/s) for men and from 1.4 to 11.2 kPa/(l/s) for women (Holmberg, Hillman, & Perkell,1988; Holmes, Leeper, & Nicholson,1994; Leeper & Graves, 1984; Stathopoulos & Sapienza, 1993;Wilson & Leeper, 1992). This suggests that R_{lx} is generally at least 10 times greater than R_{law} during phonation. According to these various estimates of R_{law} and R_{lx}, tracheal pressure would be a minimum of 90% of alveolar pressure during phonation. In this case, P_t would be within .09 kPa of P_a during loud speech produced at 1.0 kPa. Changes in R_{lx} could only increase P_t by .09 kPa, since P_t can not exceed P_a, and a large change (greater than 10 fold) in R_{lx} would be needed to increase P_t by that amount. This suggests that when the estimates of R_{law} and R_{uaw} from the literature are applied to the model, the results are consistent with the report by Bouhuys et al. (1966) that P_t is essentially the same as P_a during phonation. This suggests that the pulmonary system functions as a pressure source during phonation and that changes in P_t associated with intensity changes are due to alterations in respiratory drive rather than laryngeal adduction. However, simultaneous measures of upper and lower airway resistances during speech have not been reported.

A means for obtaining measures of alveolar pressure was developed during the course of data analysis for the study reported in a previous paper (see "Modulations in Respiratory and Laryngeal Activity Associated with Changes in Vocal Intensity During Speech" in this report). We realized that it was possible to calculate R_{law} for each subject from measures of tracheal pressure and flow, and then, to use the subjects' R_{law} value to determine P_a. The ability to measure R_{law} and calculate P_a during speech provided an opportunity to quantify the extent to which the pulmonary system acts as a pressure &/or flow source. The purpose of this study was to obtain values for both upper and lower airway resistances and to systematically quantify the extent to which P_t approximates P_a during speech and the extent to which the pulmonary system acts as a pressure source during phonation. Measures of R_{law} were obtained from tracheal pressure and airflow data. These measures were made across loudness and stress conditions to determine the consistency of the measure. Based on the individual subjects' R_{law} and the R_{uaw}, computed from tracheal pressure and

airflow, it was possible to determine the ratio of R_{uaw} to R_{law} during speech. A ratio close to 1 would indicate that the pulmonary system functions as a pressure source during phonation, and that tracheal pressure is determined primarily by alveolar pressure.

Methodology

The data analyzed for this study was initially collected for the purpose of examining changes in laryngeal and respiratory activity during change in vocal intensity. During the course of data analysis for that study, it became apparent that the data could be analyzed in a different manner to address the issue of interest in this study (i.e. the degree to which P_t approximates P_a, and the respiratory system functions as a pressure or flow source, during phonation). Therefore the subjects, procedure, and speech samples are the same as in the previous paper (see "Modulations in Respiratory and Laryngeal Activity Associated with Changes in Vocal Intensity During Speech" in this report).

Measures for Determining R_{law}: The individual subject's R_{law} was determined from P_t and airflow data collected during release of the stop plosive /p/. Figure 2a provides a sample of data collected from one subject during production of the sentence "pop took his socks off" at comfortable loudness. The top tracing is the tracheal pressure (obtained by tracheal puncture) and the bottom tracing is airflow. At the start of the record, pressure and flow gradually become less negative as the person completes inhalation prior to initiation of speech, and then move together in a positive direction as the person begins to exhale. As the lips close for production of the initial /p/ in pop, flow decreases to zero. When flow goes to zero, pressure throughout the vocal tract equalizes and tracheal pressure is equal to alveolar pressure. During the period that flow is zero, tracheal pressure (which is equal to alveolar pressure) continues to rise, indicating respiratory drive is increasing. When the seal at the lips is released, there is a dip in P_t and a peak in airflow. This sudden change in the resistance of the upper airway following a period of zero flow offers an opportunity for determining R_{law}.

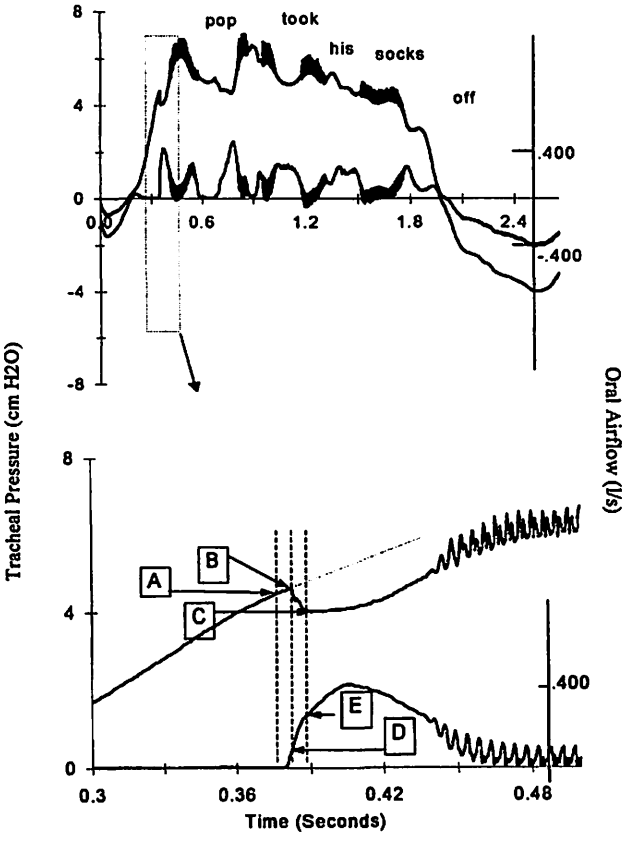


Figure 2. Method for measurement of the resistance of the lower airway (R_{law}).

As stated previously,

$$P_a = (U * R_{law}) + (U * R_{uaw})$$

Since Ruaw = P_t/U

$$P_a = P_t + (R_{law} * U).$$

Assuming that R_{law} is constant,

$$dP_a/dt = (dP_t/dt) + R_{law} * (dU/dt).$$

With rearrangement of the variables,

$$R_{law} = (dP_a/dt - dP_t/dt)/(dU/dt).$$

R_{law} was determined by calculating the change in alveolar pressure (dP_a/dt), change in tracheal pressure (dP_t/dt), and change in flow (dU/dt) during the sudden change in upper airway resistance during release of the plosive. Figure 2b provides an enlarged view of the area in the stippled box in figure 2a. The top tracing is tracheal pressure and the bottom tracing is airflow. The three vertical lines indicate two successive 10 ms. time periods. Changes in P_t and flow are measured during the 10 ms period at the start of the dip in P_t. The drop in pressure between points B and C is the change in P_t. The increase in flow between points D and E is the change in flow. Change in P_a during that same 10 ms period can not be measured. Instead, it is estimated based on the change in P_t during the 10 ms period immediately preceding the period during which measures of dP_a/dt and dU/dt were made, i.e., the increase in pressure between points A and B. Recall that during the period of zero flow prior to the dip, P_t is equal to P_a. During the drop in P_t, P_a is assumed to increase at the same rate as it was increasing during the 10 ms immediately preceding the dip. Based on these measures, (B-A)-(C-B)/(E-D), R_{law} is calculated. Lower airway resistance (R_{law}) was calculated from the /p/ in pop for 36 utterances for each subject. Mean R_{law} and standard deviation was calculated for each subject. Subject data were

analyzed to determine if R_{law} showed any systematic change with increased intensity.

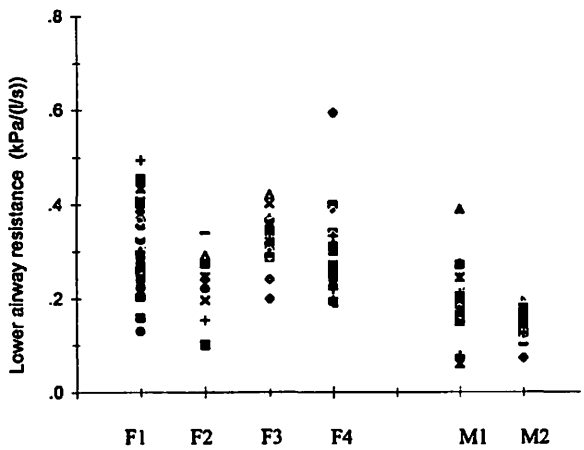
Estimation of Alveolar Pressure: The mean R_{law} values obtained for each subject were used to compute P_a from pressure and airflow ($P_a = P_t + (R_{law} * U)$).

Measures for Computation of the R_{uaw} / (R_{uaw} + R_{law}) Ratio: The tracheal pressure and flow signals collected during three repetitions of the target sentence at comfortable loudness were downsampled by a factor of 100 (resulting in 50 samples/sec instead of the original 5000 samples/sec) and then imported into a spreadsheet. Upper airway resistance (R_{uaw}) was computed by dividing P_t by flow (P_t/U) for each sample. After R_{uaw} was determined, the ratio, R_{uaw} / (R_{uaw} + R_{law}), was calculated using the R_{law} that had been obtained for that subject. The ratio was plotted as a function of time to display fluctuations in the ratio during comfortable speech.

Correlations of P_t and P_a were calculated for each subject to determine the consistency of the relationship between these two variables across conditions[2].

Results

Measures of Subjects' Lower Airway Resistance (R_{law}): Mean R_{law} values ranged from .317 to .142 kPa/(l/s) with the lowest values found for the male subjects. Figure 3 provides a scatter plot of the values calculated for R_{law} for each subject. R_{law} was calculated based on changes in pressure and flow during release of /p/ in "pop" (as illustrated in figure 2). The data collected from the four female subjects are shown on the left and from the 2 male subjects on the right of the graph. The means and standard deviations for each subject's data set are provided below the subject number on the x-axis. Five subjects' data sets contain 36 values from measures made during 4 loudness conditions, 4 stress conditions at a comfortable level, and 4 stress conditions at a loud level (1 measure/sentence X 3 repetitions X 12 conditions). One subject's data set (M2) contains 24 data points since no data was available for the loud stress condition. The data were analyzed to determine if R_{law} showed any systematic change with vocal intensity (i.e., was there any indication of change in lower airway resistance during utterances produced in a loud voice in comparison to a comfortable voice). Table 1 contains means of R_{law} obtained for each subject from data collected under different conditions. R_{law} did not vary with intensity in a consistent manner across subjects. For some subjects, R_{law} increased with vocal intensity, in others it decreased.



		F1	F2	F3	F4	M1	M2
Rlaw	mean	.307	.237	.317	.288	.185	.142
(kPa/(l/s))	sd	.085	.058	.040	.087	.054	.059
Coefficient of variation	(%)	28	24	13	30	29	42

Figure 3. Estimated R_{law} values for each subject.

Table 1.
Mean R_{law} Values (kPa/(l/s)) Calculated During 12 Different Experimental Conditions

	Levels of Loudness				COMFORTABLE, different stress				LOUD, different stress				Subject	
subject	soft	comf	loud	louder	pop	took	socks	off	pop	took	socks	off	Mean	sd
1F	.40	.40	.27	.30	.33	.40	.40	.32	.23	.20	.28	.25	.31	.09
2F	.22	.27	.26	.17	.26	.22	.10	.24	.25	.26	.28	.23	.24	.06
3F	.37	.37	.34	.28	.36	.30	.32	.30	.27	.30	.28	.31	.32	.04
4F	.22	.30	.28	.28	.29	.26	.30	.33	.36	.23	.29	.30	.29	.09
1M	.18	.18	.15	.16	.20	.19	.18	.16	.31	.22	.21	.15	.19	.05
2M	.15	.11	.15	.15	.17	.13	.16	.13					.14	.06
GROUP mean	.26	.27	.24	.22	.27	.25	.24	.25	.28	.24	.27	.25	.25	.02
SD	.10	.11	.08	.07	.07	.09	.11	.09	.05	.04	.03	.07	.08	.03

Estimation of Alveolar Pressure: Alveolar and tracheal pressure were essentially the same during phonation in these normal subjects. The mean R_{law} values obtained for each subject (listed in figure 3) were used to compute P_a. Since the tracheal pressure data were collected via tracheal puncture we obtained continuous measures of tracheal pressure, as well as airflow, and were able to compute a continuous estimation of alveolar pressure. Figure 4 provides a comparison of tracheal pressure and alveolar pressure during sustained changes in vocal intensity. Data from all 6 subjects are presented, starting at the top with subject F1 and progressing to the bottom to subject M2. Each line of data contains two tracing. The light-weight tracing is tracheal pressure. The heavier-weight tracing is estimated alveolar

[2] From this point, estimated P_a will be referred to as P_a.

pressure. These data were collected during 12 repetitions of the sentence "pop took his socks off", as indicated by the numbers at the top of the page. The first three repetitions were produced in a comfortable voice, the next three in a soft voice, and the remaining sets in a loud and louder voice. Figure 5 provides a comparison of tracheal pressure and alveolar pressure during temporary changes in vocal intensity. This figure is similar in all respects to the previous figure except this data was collected during 4 different stress conditions produced in comfortable voice. In both figures 4 and 5, the two tracing (tracheal and estimated alveolar pressure) are so similar that it is difficult to distinguish between them, particularly during phonation. This was consistent across subjects. However, between sentence productions, and particularly during inspiration, the traces separate. During inspiration when the pressures are negative, estimated alveolar pressure is always equal to or more negative than tracheal pressure. During expiration when the pressures are positive, estimated alveolar pressure is always equal to or more positive than tracheal pressure.

Ratio of $R_{uaw}/(R_{uaw} + R_{law})$ during Speech: This ratio approaches 1 during phonation in these normal subjects. As discussed in the introduction, the degree to which tracheal pressure approximates alveolar pressure is determined by the ratio of lower and upper airway resistances. In figure 6, the relationship between the relative value of the resistances and the similarity of P_a and P_t are shown. These data from subject F3 were collected during 3 repetitions of the target sentence at comfortable loudness. The five tracings displayed in this figure (starting from the top) are the ratio of $R_{uaw}/(R_{law} + R_{uaw})$, superimposed tracings of tracheal pressure and alveolar pressure, airflow, and vocal intensity. These data are divided into 6 segments. Sentence production occurs during three of the segments (segments 2, 4, and 6). The other 3 segments indicate pauses or inhalations between utterances (segments 1, 3, and 5).

As will be recalled from the explanation in the introduction, a $R_{uaw}/(R_{law} + R_{uaw})$ ratio of 1 would indicate that the resistance of the upper airway (R_{uaw}) is much larger than the resistance of the lower airway (R_{law}) and as a result, P_t is equal to P_a. A ratio of .5 indicates the resistances of the upper and lower airways are the same and as a result, P_t is half of P_a. In this figure, P_t and P_a are closer in value during speech (when the vocal fold are approximated and R_{uaw} is increased) than during breathing (when the vocal folds are abducted and R_{uaw} is decreased). P_a and P_t are closer in value during phonation than during voiceless consonants. During sentence production, the ratio of $R_{uaw}/(R_{law} + R_{uaw})$ varied between 1 and .75, and as a result, P_t remained within 100% to 75% of P_a. The lowest ratio values appeared to be coincident with increases in flow associated with production of consonants (/p, t, k, s/). As expected from the equation used to calculate P_a ($P_a = P_t + R_{law} * \text{flow}$), their values were

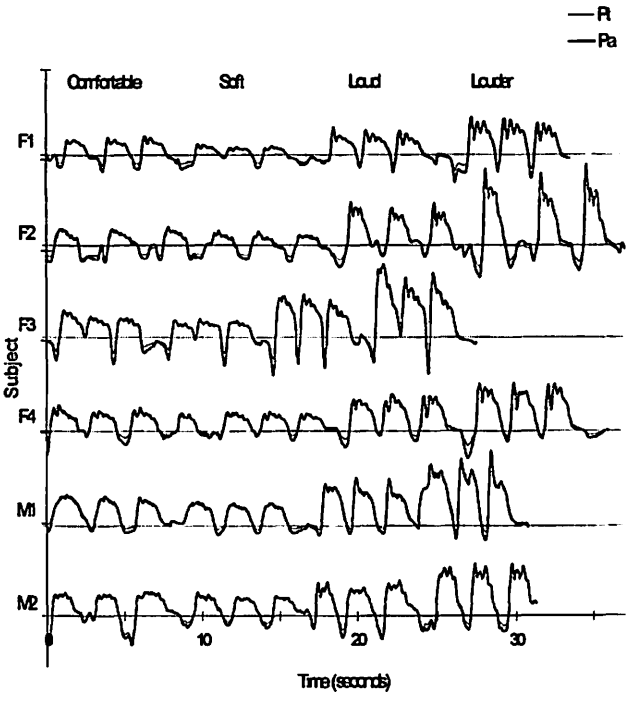


Figure 4. Tracheal and estimated alveolar pressure during sustained intensity changes for each subject.

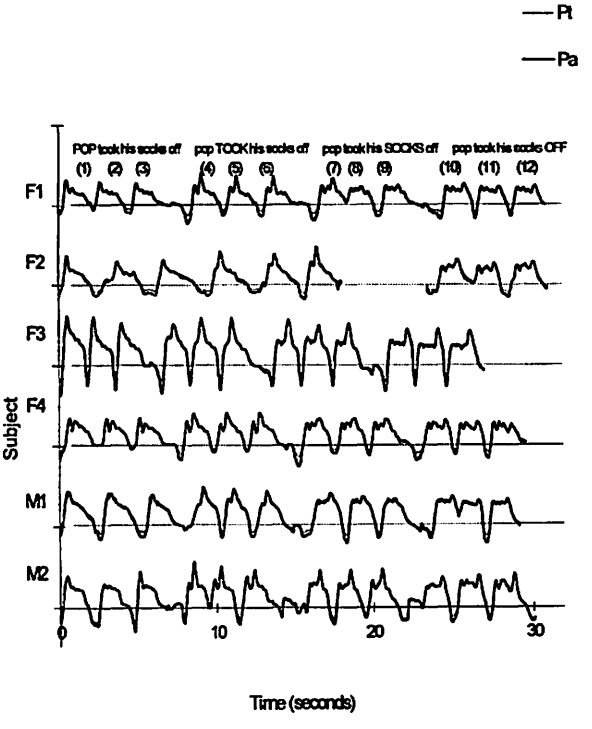


Figure 5. Tracheal and estimated alveolar pressure during temporary intensity changes for each subject.

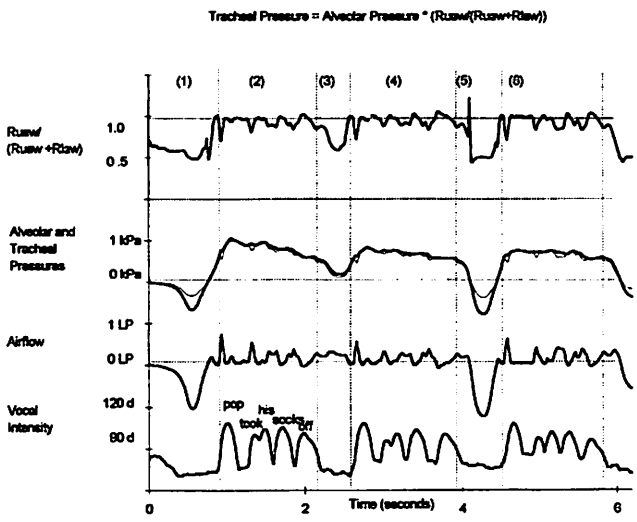


Figure 6. Illustration of the relationship between alveolar pressure, tracheal pressure, and the resistances of the upper and lower airways.

less similar when there was more flow (because R_{uaw} is lower). Between utterances, R_{uaw} was equal to or less than R_{law}, resulting in a ratio of .5, and as a result, P_t dropped to 50% of P_a.

Figure 7 contains the correlations between mean P_a and mean P_t during phonation for each subject. The degree to which P_t approximates P_a is quantified by the slope of the line. The slope is close to 1.0 for all subjects. This indicates that the ratio, $R_{uaw}/(R_{law}+R_{uaw})$, approaches 1 and that P_t is essentially the same as P_a during these productions. The strong correlations indicate that this was a highly consistent relationship. The low scatter along the entire length of the line indicates that the relationship is consistent across the range of pressures, and presumably across the range of intensities. If R_{uaw} was substantially lower during soft phonation, the ratio of upper and lower airway resistances would be smaller and this would effect the degree to which P_t approximated P_a. There was no evidence of a pressure range over which P_t increased without increase in P_a as would be expected if the change in P_t was due to alteration in R_{lx}.

Discussion

Using a technique introduced in this paper, estimates of the resistance of the lower airway were obtained and used to quantify the relationship between P_t and P_a. The model shows that the degree to which P_t approximates P_a depends on the relative resistances of the upper and lower airway. The data show that during phonation, the range of the resistance offered by the larynx (i.e. the upper airway), by subjects in this study, was typically large relative to R_{law}. Therefore, during phonation, P_t was essentially the same as P_a and changes in P_t were related to alteration in respiratory drive not laryngeal adduction. These results suggest that one role of the larynx during phonation may be to provide adequate resistance such that the pressure drop across the lower airway is minimal and the respiratory drive that originates as alveolar pressure, is relatively undiminished at the level of the vocal folds. If R_{law} is relatively constant and R_{lx} is sufficiently large relative to R_{law}, then there would be no need to further increase R_{lx} in order to maintain P_t at levels close to P_a. However, previous studies have found that R_{lx} does increase with rise in vocal intensity (Holmberg et al., 1988; Leeper & Graves, 1984; Stathopoulos & Sapienza, 1993;Wilson & Leeper, 1992). It may be that further increase in laryngeal resistance functions to prevent excessive flow as pressure increases, rather than to increases tracheal pressure. It is possible for the pulmonary system to provide a steady drive with modulation of P_t due to changes in R_{lx}. In fact, this effect is seen in the data as subjects alternated between voiced and voiceless sounds. When the vocal folds are adducted for phonation, P_t is close in value to P_a, but as the vocal folds abduct during production of

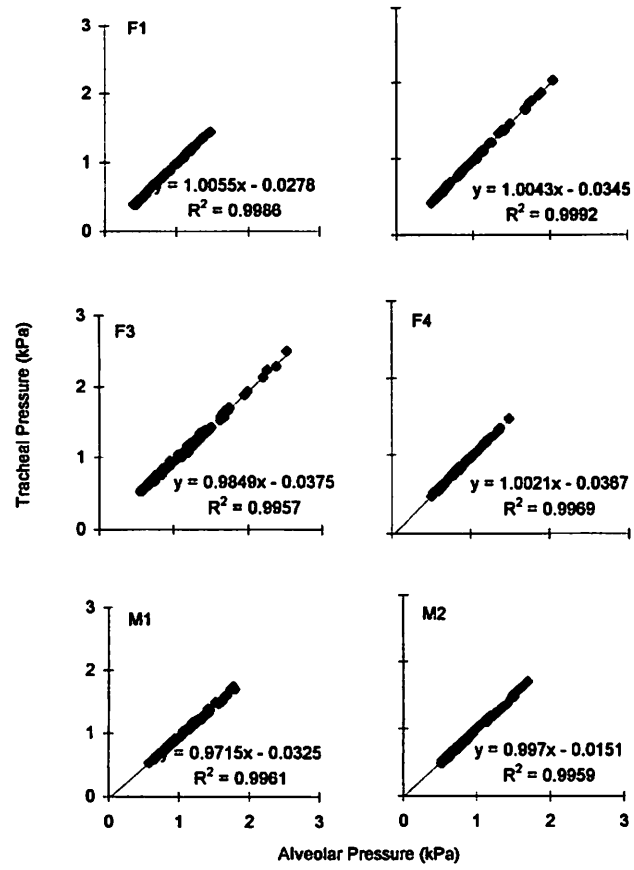


Figure 7. Correlations between tracheal pressure and alveolar pressure for each subject.

voiceless consonant, there are dips in P_t associated with increase in flow during release of the voiceless plosives. These large variations in R_{lx}, due to movement of the vocal folds, cause variations in P_t without large changes in Pa. This would be predicted by the model. However, in all subjects, once the vocal folds are adducted for voicing, further changes in R_{lx} did not produce large changes in P_t. It may be that due the magnitude of the pressures used during phonation, it is not possible to modulate tracheal pressure at the laryngeal level without producing large changes in flow not typically seen during phonation. It may be that, in the interest of controlling flow, it is preferable to control intensity via respiratory rather than laryngeal modulation. For example, this study found average P_t during unstressed comfortable speech was .8 kPa and during stressed speech was 1.35 kPa. If this change in P_t was due to modulation of a steady drive by R_{lx}, the drive (P_a) would have to be equal to or greater than 1.35 kPa. Assuming P_a was 1.4 kPa, and given an average R_{law} of .25 kPa/(l/s) (see table 2), the R_{lx} required to drop P_t to .8 kPa during the unstressed portion of the sentence would be .33 kPa/(l/s)[3]. This resistance in combination with a P_t of .8 would result in a flow of 2.4 (l/s)[4]. This is ten times greater than typical flows. This suggests that if tracheal pressure were to be modulated by laryngeal adduction, there would be large variations in phonatory flows during speech, increasing breathiness during unstressed words and decreasing the length of the utterance that could be spoken on a single breath.

The R_{law} values calculated during speech for the subjects in this study (.14 kPa/(l/s) and .32 kPa/(l/s)) were comparable to values reported in the respiratory literature. There was some indication of gender difference with lower R_{law} found in male subjects. This might be expected due to increased diameter of pulmonary tubes in taller persons. It has been suggested that females have higher R_{lx} because they have smaller larynges. Given the finding that female subjects tended to have higher R_{law}, presumably due to smaller diameter airway, it could be necessary for them to generate higher R_{lx} in order to obtain the same degree of similarly between P_t and P_a. Although there was some variability within and between subjects, in terms of R_{law} values, the important point for the purpose of this study was that even the extreme high values were relatively small in comparison to R_{uaw} during phonation.

Knowledge of the relationship between R_{uaw} and R_{law} during phonation may also provide insight into a related issue regarding reflex control of air pressure during consonant production. Some researchers have suggested that, during speech, the air pressure in the vocal tract is reflexively maintained at a constant minimum level by adjusting respiratory drive and airway resistance (Warren, 1986; Warren, Dalston, Moor, Hairfield, & Smith, 1989a; Warren, Morr, Rochet, & Dalston, 1989b; Warren, Rochet, Dalston, & Mayo, 1992; Kim, Zajac, Warren, Mayo, & Essick, 1997). Warren hypothesized that the maintenance of a constant pressure is accomplished by a servo system with continuous on-line feedback. Warren's hypothesis is that during speech, the air pressure in the vocal tract (oral and tracheal) is regulated (i.e. maintained above a certain level) by rapid active changes in respiratory drive or in vocal tract resistance. The level at which pressure is maintained is a minimum of 3 cm H_2O (.3 kPa). The regulated variable is vocal tract pressure, which is controlled by airflow and vocal tract resistance. Thus, if resistance changes, alteration in flow would be needed to maintain pressure. Other investigators (Wyke, 1983) have proposed the idea of reflexive control of phonation. Wyke hypothesizes that there is " a physiological mechanism that couples the activity of muscles like the CT and TA to the prevailing subglottal pressure as long as that pressure is greater than 3 cm H_2O. There are some very sensitive mechanoreceptors in the subglottal mucosa the frequency of which is constantly proportional to variations in tracheal pressure. These, in turn, reflexively affect the tension in the vocal fold musculature and the stiffness of the vocal folds" (p.126).

Although many reflex connections exist between the respiratory and laryngeal systems, there is limited evidence that they are normally active during phonation. An alternative explanation for the maintenance of a minimum pressure was offered by Moon, Folkins, Smith, & Luschei, (1993). Moon et al. agree that oral pressure is maintained above a minimum level, but asserts that this is due to inherent physiologic properties of the human respiratory system during speech (i.e. this is a passive not active process). They suggested that this regulation of pressure was a passive property of the human respiratory system during phonation. This suggests open-loop rather than reflexive control during phonation. Information regarding the resistances of the lower and upper airway may provide answers regarding the viability of the argument that passive properties of the aerodynamic system could be responsible for maintenance of a minimum air pressure in the vocal tract.

In a recent publication, Kim et al., (1997) measured oral pressure (P_o) and airflow during /p/ to determine response to sudden unexpected perturbations induced with the use of a valve which permitted air to escape from the oral cavity. The authors reported that the mean peak P_o dropped from .76 kPa during the normal (valve closed) condition to .47 kPa during the valve open condition. The authors attributed the maintenance of adequate, though reduced, oral pressure to an active response of the respiratory system to the perturbation. However, if the results are interpreted in terms of the model presented in this paper, with the addition

[3] Since $P_t = P_a (R_{lx} / R_{law} + R_{lx})$, given .8 kPa = 1.4 kPa * (R_{lx} / .25 kPa/(l/s) + R_{lx}), then Rlx = .33 kPa/(l/s)

[4] Flow is equal to pressure divided by resistance (U = .8 kPa/ .33 kPa/(l/s)).

of a resistor at the lips, the maintenance of an oral pressure of .47 kPa is explainable based on the resistances of the system without any active respiratory response.
During the no flow (no leak) condition P_o was equal to .76 kPa, indicating P_a was .76 kPa.

During the flow condition (with the leak) P_o will equal

$$P_o = P_a * (R_{lips} / (R_{lips} + R_{lx} + R_{law})).$$

Kim et al. provided a sample of P_o and airflow data collected during production of /p/ with a .45 cm^2 valve opening. According to this data, at peak pressure P_o = .6 kPa and airflow = .8 (l/s), indicating the resistance at the lips was .75 kPa/(l/s). From their data,

$$P_a = .76 \text{ kPa and } R_{lips} = .75 \text{ kPa/(l/s)},$$

From our study data,

$$\text{group mean } R_{law} = .25 \text{ kPa/(l/s)}$$

From some limited additional data collected during the syllable repetition task at comfortable loudness, during production of the voiceless /p/

$$R_{lx} = .35 \text{ kPa/(l/s), and}$$

Thus

$$P_o = .76 \text{ kPa} * (.75 \text{ kPa/(l/s)}/(.75 + .35 + .25 \text{ kPa/(l/s)})$$

$$P_o = .42 \text{ kPa}$$

These estimates of resistances suggest that it could be possible to explain the maintenance of an oral pressure of .47 kPa based on a passive response to changes in the resistances of the system. The high flow could be explained by the decreased resistance at the lips[5] and not to any active response of the respiratory system. Further study is needed to test this hypothesis.

The main finding of this study was not that P_t is essentially equal to P_a during speech , but rather that P_a and P_t are essentially equal when R_{uaw} is substantially larger than R_{law}, which is generally the case during phonation in normal speakers. The pulmonary system acts like a pressure source in normal speakers under normal conditions during phonation but, this is not necessarily always the case. According to the model discussed in this paper, as the upper airway resistance becomes equal to or less than lower airway resistance, the pulmonary system acts more like a flow source and less like a pressure source. There are a number of instances in which upper airway resistance is reduced. For example, during release of voiceless consonants, the resistance of the upper airway is quite low. In this situation, the drop in pressure over the lower airway is increased, P_t is reduced in comparison to P_a, and the pulmonary system appears to function as a combination of a pressure and flow source. This is likely to also be the case during respiration when the vocal folds are abducted and there is minimal upper airway resistance. In addition, the pulmonary system may not necessarily function as a pressure source in some persons with voice disorders. In some persons with voice problems, the relative values of upper and lower airway resistance could be altered due to (a) increase in R_{law} (emphysema), (b) decrease in R_{uaw} (vocal cord paralysis), or (c) lack of ability to control and coordinate the relative value of these resistance during speech, as seen perhaps in vocal tremor or spasmodic dysphonia. Measurement of upper and lower airway resistances in these persons may provide additional insight into the nature of their problems.

[5] Flow is equal to alveolar pressure divided by total resistance ($U = P_a / R_t$). In this case U = .76/1.25 = .608 l/s

Acknowledgment

This research was supported by grant no. P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders. Special thanks to Ingo Titze, Michael Karnell, Kelly Cole, and Charles A. Miller for their comments on an earlier version of this paper.

Bibliography

Bouhuys, A., Proctor, D.F., & Mead, J. (1966). Kinetic aspects of singing. Journal of Applied Physiology, 21, 483-496.

Choi, H.S., Berke, G.S., Ye, M., & Keiman, J. (1993). Function of the thyroarytenoid muscle in a canine laryngeal model. Annals of Otology, Rhinology, & Laryngology, 102, 769-76.

Choi, H.S., Ye, M., & Berke, G.S. (1995). Function of the interarytenoid (IA) muscle in phonation: in vivo laryngeal model. Yonsei Medical Journal, 36, 58-67.

Dromey, C., Ramig, L.O., & Johnson, A.B.(1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease:a case study. Journal of Speech & Hearing Research, 38(4), 751-64.

Fisher, A.B., DuBois, A.B., & Hyde, R.W. (1968). Evaluation of the forced oscillation technique for the determination of resistance to breathing. The Journal of Clinical Investigation, 47, 2045-2049.

Hirano, M., Ohala, J., & Vennard, W. (1969). The function of laryngeal muscle in regulating fundamental frequency and intensity of phonation. Journal of Speech and Hearing Research, 12, 616-628.

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

Holmes, L.C., Leeper, H.A., & Nicholson, I.R. (1994). Laryngeal airway resistance of older men and women as a function of vocal sound pressure level. Journal of Speech and Hearing Research, 37 (4), 789-799.

Kim, J., Zajac, D. J., Warren, D. W., Mayo, R., Essick, G.K. (1997). The response to sudden change in vocal tract resistance during stop consonant production. Journal of Speech, Language, and Hearing Research, 40(4), 848-857.

Leeper, H.A. Jr., & Graves, D.K. (1984). Consistency of laryngeal airway resistance in adult women. Journal of Communication Disorders, 17(3), 153-63.

Moon, J. B., Folkins, J.W., Smith, A. E., & Luschei, E. S. (1993). Air pressure regulation during speech production. Journal of the Acoustical Society of America, 94(1), 54-63.

Nasri, S., Namazie, A., Kreiman, J., Sercarz, J.A., Gerratt, B.R., & Berke, G.S. (1994). A pressure-regulated model of normal and pathologic phonation. Otolaryngology – Head & Neck Surgery, 111, 807-15.

Nasri, S., Namazie, A., Ye, M., Kreiman, J., Gerratt, B.R., & Berke, G.S. (1996). Characteristics of an in vivo canine model of phonation with a constant air pressure source. Laryngoscope, 106, 745-51

Phagoo, S.B., Watson, R.A., Silverman, M., & Pride, N.B. (1995). Comparison of four methods of assessing airflow resistance before and after induced airway narrowing in normal subjects. Journal of Applied Physiology, 79(2), 518-525.

Sercarz, J.A., Berke, G.S., Bielamowicz, S., Kreiman, J., Ye, M., & Green, D.C. (1994). Changes in glottal area associated with increasing airflow. Annals of Otology, Rhinology, & Laryngology, 103, 139-144.

Slavit, D.H., & McCaffrey, T.V. (1995). Open slope quotient: a new glottographic parameter. Journal of Voice, 9, 86-94

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

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

Van Altena, R., Gimeno, F. (1994). Respiratory resistance measured by flow-interruption in a normal population. Respiration, 61, 249-254.

Verneuil, A., Kreiman, J., Kevorkian, K., Gerratt, B.R., & Berke, G.S. (1996). Effects of driving pressure and recurrent laryngeal nerve stimulation on glottic vibration in a constant pressure model. Otolaryngology – Head & Neck Surgery, 115, 15-23.

Warren, D.W. (1986). Compensatory speech behaviors in individuals with cleft palate: a regulation/control phenomenon? Cleft Palate Journal, 23, 251-60.

Warren, D.W. (1986). Compensatory speech behaviors in individuals with cleft palate: a regulation/control phenomenon? Cleft Palate Journal, 23, 251-60.

Warren, D.W., Dalston, R., Moor, K., Hairfield, M., & Smith, L. (1989a). The speech regulating system: temporal and aerodynamic responses to velopharyngeal inadequacy. Journal of Speech and Hearing Research, 32, 566-575.

Warren, D.W., Morr, K., Rochet, A., & Dalston, R. (1989b) Respiratory response to a decrease in velopharyngeal resistance. Journal of the Acoustical Society of America, 86, 917-24.

Warren, D.W., Rochet, A.P., Dalston, R.M., & Mayo, R. (1992). Controlling changes in vocal tract resistance. Journal of the Acoustical Society of America, 91, 2947-53

Wilson, J.V. & Leeper, H.A. (1992). Changes in laryngeal airway resistance in young adult men and women as a function of vocal sound pressure level and syllable context. Journal of Voice, 6 (3), 235-245.

Wyke, B. (1983). Reflexogenic contributions to vocal fold control systems. Vocal Fold Physiology. (p.138-41).

NCVS Status and Progress Report - 12
June 1998, 11-21

Modulations in Respiratory and Laryngeal Activity Associated with Changes in Vocal Intensity During Speech

Eileen Finnegan, Ph.D.
Erich Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Department of Otolaryngology, University of Iowa Hospitals and Clinics
Henry Hoffman, M.D.
Department of Otolaryngology, University of Iowa Hospitals and Clinics

Abstract

We tested the hypothesis that an increase in respiratory drive is the primary mechanism for sustained increases in vocal intensity (i.e. an increase in overall level of loudness), whereas adjustment of laryngeal adduction is the mechanism for temporary changes (i.e. an increase due to emphatic stress). Measures of respiratory and laryngeal activity were collected from six subjects during production of a sentence at different levels of vocal intensity and with a variety of stress patterns. Tracheal pressure (P_t), obtained via tracheal puncture, provided a measure of the output of the respiratory system. Laryngeal airway resistance (R_{lx}) and laryngeal electromyography from the thyroarytenoid muscle (TA EMG) provided measures of laryngeal activity. The results of this study indicated that contrary to the outcome predicted by the hypothesis, there was no evidence that different strategies were used to control intensity based on whether the change in intensity was temporary or sustained. Respiratory drive, as indicated by P_t, and laryngeal activity, as indicated by R_{lx} and TA EMG, all tended to increase with intensity. The pressure provided by the respiratory system was more highly correlated with intensity change than were either R_{lx} or TA EMG. Finally, it was noted that respiratory drive was modulated during sentence production.

The ability to manipulate vocal intensity enables a speaker to transmit a message more effectively through use of increased intensity to highlight key points. Inability to control intensity, a characteristic of some voice disorders, attracts unwanted attention to the manner of speech and detracts from the message being communicated. Many investigators have studied the role of either the laryngeal system or the respiratory systems in control of vocal intensity. The majority of these studies have examined intensity control strategies used during sustained phonation of a vowel or during prolonged phonation of a vowel in a syllable repetition task. Relatively few studies have been done to directly assess how the laryngeal and respiratory systems function together to produce changes in the intensity of phonation during sentence production.

Researchers have frequently used tracheal pressure to provide a measure of respiratory activity. The relationship between tracheal pressure (P_t) and vocal intensity has been studied extensively during sustained phonation (Bouhuys, Proctor, & Mead, 1966; Higgins & Saxman, 1991; Isshiki, 1964;Koyama, Kawasaki, & Ogura, 1969; Tanaka & Gould, 1983; Kunze, 1964; Ladefoged & McKinney, 1963; Tanaka & Tanabe, 1986) and syllable repetition (Higgins & Saxman, 1991; Holmberg, Hillman, & Perkell, 1988; Leeper & Graves, 1984; Wilson & Leeper, 1992). The consistent finding in all of these studies has been that increase in intensity is accompanied by an increase in tracheal pressure. Early attempts to systematically quantify the relationship between P_t and intensity have generally concluded that a power relationship existed between these

two variables[1] (Isshiki, 1964; Koyama et al., 1969; Kunze, 1962;Tanaka & Gould 1983). Titze and Sundberg (1992) identified the importance of phonation threshold pressure (P_{th}), the minimum pressure needed to produce sound, in assessing the relationship between P_t and sound pressure level (SPL). They found that when phonation threshold pressure was taken into account, SPL increased at the rate of 8-9 dB per fractional excess of lung pressure over threshold $((P_l - P_{th})/P_{th})^2$. Since this fraction can be doubled more than three times, this implies a possible increase of more than 24-27 dB from an increase in tracheal pressure (Titze, 1994).

The role of the larynx in control of vocal intensity has been assessed by measuring changes in laryngeal electromyography (EMG), laryngeal airway resistance (R_{lx}), and vocal fold adduction. The findings regarding the relationship between laryngeal EMG activity and intensity have not been consistent. Some investigators have found laryngeal EMG activity increases with intensity and suggested that laryngeal muscles are active in regulating intensity (Hirano, Ohala, & Vennard, 1969; Hirano, 1987). Others have found little change in laryngeal EMG activity with intensity and suggested that intensity changes are related to changes in tracheal pressure (Faaborg-Andersen, 1957; Gay, Hirose, Strome, Sawashima, 1972). Investigators who have estimated R_{lx} during a syllable repetition task have reported that its mean value generally increased with vocal intensity (Holmberg et al., 1988; Leeper & Graves, 1984; Stathopoulos & Sapienza, 1993a; Wilson & Leeper, 1992). The relatively consistent finding of increased R_{lx} may seem at odds with the inconsistent findings regarding increase in EMG activity, if one assumes that increase in laryngeal resistance indicates increased adduction of the vocal folds. However, this may not necessarily be the case. Tanaka & Tanabe (1986) showed that at constant aerodynamic power, laryngeal resistance increased with electrical stimulation of the TA, LCA, and CT muscles in an in-vivo canine experiment. More recently, Alipour, Scherer, & Finnegan (1997) studied pressure-flow relationships using excised canine larynges and found that R_{lx} increased without change in adduction as pressure was increased and that it was possible to obtain the same R_{lx} at different levels of adduction. At constant pressure, however, increase in adduction did result in an increase in R_{lx}. Therefore, care should be taken in interpretation of changes in R_{lx}, particularly if pressure is changing at the same time. The effect of increasing vocal fold adduction on vocal intensity was studied by Titze (1988). Experimental data were collected using excised canine larynges and compared to the output of a mathematical model. Titze reported that radiated acoustic power was not affected by changes in glottal width. However, when the frequency response of the human ear was taken into consideration, differences in the perceived loudness (i.e. detection and perception of intensity) with adduction were noted. He found that phonation decreased in intensity by 4-7 dB when the abduction quotient[3] (Q_a) increased from .8 to 1 (the range over which the voice becomes very breathy) or when Q_a decreased from -.8 to -1 (the range over which the voice begins to sound pressed). In the broad range of normal voicing (i.e. for $-.8 < Q_a < .8$), change in adduction had no effect on intensity. In addition to changes in intensity associated with change in adduction, Titze studied the effect of vocal fold length on intensity. He reported that intensity decreased 5 dB, as vocal fold length was increased 10 % (from 5 % less than rest length to 5% greater than rest length), due to decrease in amplitude of vibration.

Simultaneous assessment of the contribution of both the respiratory and laryngeal mechanisms related to vocal intensity control has also been reported (Isshiki, 1964; Stathopoulos & Sapienza, 1993a). Isshiki (1964) measured tracheal pressure, laryngeal airway resistance, and vocal intensity during sustained phonation in a single subject. Similar to the findings reported by previous investigators, he found that at low pitches, both P_t and R_{lx} increased with vocal intensity, however, at high pitches, P_t increased but R_{lx} did not. He suggested that change in intensity was due to laryngeal control at low pitches (i.e. typical speaking range) and to respiratory control at high pitches (i.e. falsetto). Stathopoulos and Sapienza (1993a) assessed changes in respiratory and laryngeal activity associated with variations in vocal intensity during a syllable repetition task. They found intra-subject differences in strategies used to increase intensity. They reported the most common pattern, seen in 15/20 subjects, indicated use of both laryngeal and respiratory effort. Three of the remaining subjects demonstrated primarily respiratory effort and the other two subjects demonstrated primarily laryngeal effort. However, interpretation of the results of this study was complicated, since the parameters used as laryngeal indices (open quotient[4], maxi-

[1] These studies found that sound intensity (I) is related to subglottic pressure (Psg) by $I = Psg^x$, where x is in the range of 3 to 4.3.

[2] For example, if P_{tr} is .2 kPa and P_l is .4 kPa, then the fraction $(P_l - P_{th}/P_{tr})$ is equal to 1. If lung pressure increases to .8 kPa, the fraction increases to 2 and there is a 8-9 dB increase in intensity. If lung pressure increases to 1.0 kPa, the fraction increases to 4 and intensity will increase another 8-9 dB

[3] The abduction quotient in the ratio of the glottal half-width at the vocal processes to the amplitude of vibration of the vocal folds, and ranges from (+) 1 to (-) 1. An abduction quotient of 0 indicates the vocal processes are in contact with each other, more positive values indicate increasing abduction, while more negative values indicate increasing force of adduction.

[4] The open quotient is a measure of the glottal waveform. It is the duration of flow/period of the waveform. It ranges from 0 to 1.

mum flow declination rate[5]), could also be effected by changes in respiratory drive without change in laryngeal activity (Titze, 1992).

The studies reviewed above indicate that during sustained or prolonged phonation, tracheal pressure and laryngeal resistance increase with intensity. Since the results regarding contraction of the laryngeal adductor muscles is inconsistent, it is not clear if increases in R_{lx} were due to an active process (increased muscle contraction) or a passive result of an increase in respiratory drive. If the increase in resistance is primarily passive, then intensity is controlled primarily by respiratory drive during these tasks, as suggested by the modeling and experimental data by Titze.

It has been suggested that different strategies may be used to control intensity during speech, as opposed to during sustained or prolonged phonation. There have been several studies of the contribution of respiratory (Bouhuys et al., 1966; Hixon et al., 1973; Russell & Stathopoulos, 1988; Winkworth et al. 1994, 1995) and laryngeal activity (Hirano et al., 1969) to alteration in vocal intensity during speech. During speech, measures of tracheal pressure are difficult to obtain. The non-invasive method for estimation of P_t, described in Smitheran and Hixon (1981), requires use of a syllable repetition task and cannot be used during sentences or conversational speech. In order to measure P_t during speech, it must be measured directly or estimated from esophageal pressure, using invasive procedures. As a result, investigators have largely avoided the use of invasive procedures and relied on changes in respiratory kinematics or lung volume, rather than measurement of tracheal pressure, to provide information regarding changes in respiratory drive during ongoing speech. There are several important studies that have attempted to establish the respiratory contributions to vocal loudness by using respiratory kinematics. Early studies found that a larger proportion of vital capacity was used during loud speech (Bouhuys et al., 1966; Hixon et al., 1973; Russell & Stathopoulos, 1988) with frequent dips below functional residual capacity (Bouhuys et al., 1966; Russell & Stathopoulos, 1988). In addition, researchers found that there was a trend for subjects to initiate louder utterances at higher lung volumes using increased rib cage excursion (Hixon et al., 1973). More recently, Winkworth et al. (1994) measured variability in lung volumes associated with vocal intensity during reading. They found that, when all speech expirations were grouped for each subject into three intensity intervals, all lung volume variables were greater in the higher intensity interval when compared to the next lowest. They concluded that within each reading passage, lung volumes were significantly increased for utterances produced with greater vocal intensity. In contrast to the results found during reading, Winkworth et al. (1995) reported that during conversational speech, increases in vocal intensity were not associated with increased lung volume. They suggested that "in everyday speaking situations (when speech intensity increases are related to emphatic stress) laryngeal adjustments to increase tracheal pressure, in preference to increased expiratory effort, might be predicted as being more efficient... It is possible that periods of more sustained louder speech, ... might be associated with strategies involving respiratory adjustment to achieve the required subglottal air pressure" (p. 141). This suggestion is consistent with previous speculation by Titze (1988). He speculated that, despite the finding that changes in laryngeal adduction resulted in only modest changes in intensity in comparison to the effect of increase in tracheal pressure, "given laryngeal action is fast, this method of loudness regulation may be used frequently during speech when speed and fine control are important" (p. 237, Titze, 1988).

Limited evidence is available regarding the role of the laryngeal system in intensity control during sentence production. No measures of R_{lx} have been obtained during sentence production since direct measurement of tracheal pressure would be needed to obtain these measures. There has been one report of laryngeal EMG activity associated with intensity changes during sentence production. Hirano et al. (1969) measured TA, CT, and lateral cricothyroid (LCA) activity in 6 subjects during sentence production. The sentence was repeated five times to elicit a variety of intensity and pitch changes. In the first four repetitions, different words in the sentence were emphasized, eliciting an increase in both frequency (F_0) and intensity during the emphasized word. In the fifth production, the sentence was produced in interrogative form, to elicit a change in F_0 without change in intensity. During sentence production, all muscle EMG increased both during the emphasized word, when F_0 and intensity increased, and during the interrogative production, when F_0 increased and intensity decreased. Hirano et al. (1969) suggested decreased intensity during the interrogative could be due to falling airflow, although flow was not measured in this study.

The hypothesis to be considered in this study is that different strategies may be used during speech depending on whether an increase in vocal intensity is sustained or temporary. In this study, measures of respiratory and laryngeal activity were collected during production of a sentence at different levels of vocal intensity and with a variety of stress patterns. Tracheal pressure (P_t), obtained via tracheal puncture, provided a measure of respiratory expiratory activity. Laryngeal airway resistance and laryngeal electromyography recorded from the thyroarytenoid

[5] The maximum flow declination rate (MFDR) is the maximum rate of change in flow measured from the glottal waveform at the time when the vocal folds are closing and flow is decreasing. Increase in the MFDR is associated with faster movement of the vocal folds during closing, i.e. a quicker shut-off of flow.

muscle (TA EMG) provided measures of laryngeal activity. A secondary purpose of the study was to determine the extent to which laryngeal muscle activity was related to laryngeal resistance during speech. The activity of the thyroarytenoid muscle (TA) was of particular interest since the TA is one of the muscles most frequently targeted for treatment in disorders characterized by "over-adduction" of the vocal folds (i.e. spasmodic dysphonia and vocal tremor).

Methodology

Subjects: There were four adult female subjects (27 to 39 yr.) and two adult male subjects (25 and 31yr.). Two additional subjects were entered in the protocol but it was not possible to obtain complete P_t data, because both subjects had a greater than normal amount of tissue overlying the trachea. In these cases, the tracheal puncture was accomplished, but skin to tracheal distance caused excessive angulation of the cannula during swallowing, resulting in dislodgement. Subjects had no history of 1) voice, speech, or hearing problems, 2) surgery to the neck or chest, 3) bleeding problems, 4) use of aspirin or coumadin within the 14 days prior to the experiment day, 5) allergy to subcutaneous anesthetic agents, or 6) immobile vocal cords. Subjects spoke English as a first language and were not trained in singing, acting, or public speaking. As part of the screening physical examination of the head and neck region, the vocal cords were examined to confirm their normal mobility

Procedure: Simultaneous recordings of sound pressure level, airflow, tracheal pressure, and thyroarytenoid EMG were obtained. A small amount (0.5 cc) of 2% lidocaine with 1:100,000 epinephrine was injected just below the surface of the skin overlaying the cricothyroid ligament and bipolar hooked-wire electrodes (50 ìm diameter stainless steel wires) were inserted with a 1.5 inch, 25 gauge needle into the thyroarytenoid (TA) muscle. The ends of the electrode wires had been stripped of insulation for 1-1.5 mm. The electrode was judged to be in the TA muscle if there was EMG activity associated with sustained phonation. In addition, subjects were requested to swallow 30 cc of water and to perform valsalva maneuvers in order to obtain maximum activation of the TA muscle. Tracheal puncture was performed to measure tracheal pressure. Anesthesia was injected into the skin and tissue immediately anterior to the anterior tracheal wall. The tracheal puncture needle assembly, consisting of a 25 gauge needle inside of a 20 gauge cannula, was placed through the skin, the underlying soft tissue, and the membranous separation between tracheal rings. The insertion was made either between the first and second (mostly commonly), the second and third, or between the cricoid and the first tracheal ring. Following insertion, the needle was withdrawn, leaving the cannula in the trachea. Indirect mirror examination or flexible fiberoptic nasolaryngoscopy was used to inspect the subglottis through the open glottis to identify the correct placement of this cannula. Anesthesia of the trachea was avoided since it could alter the dynamics of the speech and breathing process studied, and was not needed if the cannula was correctly positioned. The distal end of the cannula was connected to silastic tubing (45 cm in length; .67 mm I.D.) which was attached to an external pressure transducer. This minaturized pressure transducer had been modified to have a very low volume, so even though the tubing had a small diameter, the frequency response of the transducer and cannula system was flat to a frequency of about 40 Hz. The airflow signal was transduced with a pneumotachograph attached to a circumferentially vented facemask (Glottal Enterprises, model MS 100; Rothenberg, 1977). A condenser microphone (headset model C410, AKG acoustics) with an AC power supply (model N62E, AKG acoustics), placed 6 cm in front of the subject, just outside the mask, recorded the sound pressure level. All signals were amplified, low-pass filtered at 2500 Hz., and digitized at 5000 Hz per channel by a computer system used for data analysis (WINDAQ, Dataq Inc., Akron, OH).

Calibration: Calibration of the instrumentation was performed twice, once prior to the start of data collection for the first subject and once after completion of data collection from the last subject. In order to calibrate airflow and pressure instrumentation, two flow levels were generated and monitored (0 and 1000 cc/s) using an vacuum cleaner as the flow source and a rotometer (Precision Bore Flowrator tube, FP1/2 21G10/27, F & P Co.) to monitor the flow levels. Air pressure was calibrated at two levels (0 and 10 cm H_2O) using a syringe as a pressure source and a water-filled U-tube manometer as a monitor. These signals were recorded onto the computer and used to calibrate experimental signals. In order to calibrate the microphone signal, white noise generated from the FM band of a tuner was connected to a speaker to produce a calibration signal. The microphone and a sound pressure level meter (Quest, Model IEC 651), set at A-weighting, were positioned side by side at equal distance from the speaker and signals, at intensities ranging from 60-112 dB SPL, were recorded onto the computer and used to calibrate the audio signal.

Speech Sample: Sustained increases in intensity were elicited by asking the subjects to produce three tokens of the target sentence ("Pop took his socks off") at four loudness levels (soft, comfortable, loud, and louder). In the soft condition the subject was instructed to "speak softly but do not whisper, speak as you would to someone sitting right next to you". In the comfortable condition they were instructed to "speak as you would to someone across the table". In the loud condition, the instruction was "speak as you would to someone across a quiet room" and the louder condition "speak as you would to someone across a noisy room". Temporary increases in intensity were elicited by

asking the subjects to repeat a second series of each sentence in which they produced the sentence at a comfortable loudness emphasizing different words in the sentence. The subject was provided with a model and then asked a question to elicit emphasis of the target word. For example, the investigator provided the subject with a model of "POP took his socks off", then asked "who took his socks off?". A third series was produced in a loud voice again emphasizing target words in the sentence, and a fourth series in a soft voice emphasizing target words. Table 1 provides a list of the utterances elicited from each subject. The subjects produced 3 repetitions of the sentence, in each of the 16 conditions, for a total of 48 productions per subject. The subjects were reminded to maintain an airtight seal between the mask and their face at the start of each set of data collection.

Measures: Mean measures of activity during phonation were obtained using the WINDAQ data analysis system. A 20 ms segment was selected from the center of the vowel in each word of the sentence "pop took his socks off". The mean P_t, mean flow, and RMS of the audio signal, during that 20 ms segment, were obtained. In addition, the mean level of TA EMG activity during the 20 ms segment that preceded the original segment by 40 ms was also obtained. The difference in time of measurement was to allow for contraction time, which has been shown to be approximately 40 ms (Titze, 1991). In addition, maximum levels of TA EMG activity were determined by obtaining measures of mean rectified TA EMG activity during a 10 ms. period during valsalva, swallow, and the pre-phonatory bursts associated with loud phonation. These maximum values were included in analysis and interpretation of TA EMG activity during phonation. All the measures were imported into a spreadsheet. R_{lx} was calculated from tracheal pressure and airflow (R_{lx} = pressure/flow). Correlations were determined using linear regression.

Analysis: Analysis of the data involved three basic steps (1) measurement of mean levels of P_t, flow, TA EMG, and intensity, (2) calculation of R_{lx} from P_t and flow, and (3) correlation of respiratory and laryngeal variables with vocal intensity using linear regression. The data were analyzed for evidence of use of different strategies in sustained and temporary increases in loudness. If the hypothesis were true, then we would expect respiratory variables to be highly correlated with vocal intensity during sustained increases in intensity (i.e. changes in overall loudness) and for laryngeal variables to more highly correlated with vocal intensity during temporary increases in intensity (i.e. related to emphasis).

Table 1.
List of Utterances Elicited from Each Subject

"Pop took his socks off":
1. (in a comfortable voice)
2. (in a soft voice)
3. (in a loud voice)
4. (in a louder voice)

In a comfortable voice:
5. POP took his socks off.
6. Pop TOOK his socks off.
7. Pop took his SOCKS off.
8. Pop took his socks OFF.

In a loud voice:
9. POP took his socks off.
10. Pop TOOK his socks off.
11. Pop took his SOCKS off.
12. Pop took his socks OFF.

In a soft voice:
13. POP took his socks off.
14. Pop TOOK his socks off.
15. Pop took his SOCKS off.
16. Pop took his socks OFF.

Results

Samples of Data Collected

Figure 1 provides samples of data collected from subjects in this study. The data set at the top of the page is data collected during 12 repetitions of the sentence "pop took his socks off". In this sample, subject F3 produced 3 repetitions of the sentence at each of four loudness conditions. The first three repetitions were produced in a comfortable voice, the second set of repetitions was in a soft voice, and the last two sets were in a loud and louder voice. The four signals displayed here are, from top to bottom, tracheal pressure, airflow, rectified TA EMG, and intensity. The data set at the bottom of the page is similar in all respects to the one above it, except that the subject was asked to produce 3 repetitions of the sentence in 4 different stress conditions at comfortable loudness. In the first set of 3 repetitions, the subject emphasized the word "pop". In the second set, "took" was emphasized, then "socks", and in the last set "off" was emphasized. The subject demonstrated a similar pattern of P_t, flow, and TA EMG activity during the three repetitions for each condition.

During data analysis, it was observed that mean phonatory flow varied over the course of the utterance. Flow was generally low at the start of the utterance, higher at the middle of the sentence, then decreasing at the end of the utterance. This subject exhibited mean flows that were negative at the end of the utterance, a pattern that was common across subjects. We hypothesized that these very low and negative flows could be due to transient increases in the volume of the upper airway during articulation of the sentence. Production of "pop", "socks", and "off" involve a transient downward movement of the jaw during production

of the vowel. These changes in volume could result in proportional decreases in oral pressure, effecting flow across the mask. The cause of these negative pressures is uncertain and warrants further study. However, for the purpose of this analysis, R_{lx} calculations, which depend on flow data, will be reported only for "took" and "his", the two words in the sentence produced with minimal jaw movement.

Change in Parameters with Sustained or Temporary Increase in Vocal Intensity

Figure 2 graphs the change in each parameter as a function of experimental condition for each subject. The four parameters, from the top to bottom, were vocal intensity, tracheal pressure, laryngeal airway resistance, and TA EMG activity. The experimental conditions indicated on the x-axis are, from left to right, the four changes in overall intensity (soft, comfortable, loud, and louder), the different stress conditions produced in soft voice (divided into unstressed vs. stressed words), and then the different stress conditions produced in comfortable and loud voice. The means for the first four conditions (soft, comfortable, loud, and louder) are based on 15 tokens (5 vowels/sentence X 3 repetitions of each sentence). The means of the next 3 sets of conditions (stressed and unstressed words produced in soft, comfortable, and loud levels) are based on 48 tokens for the unstressed words (4 unstressed vowels/sentence X 4 stress conditions X 3 repetitions) and 12 tokens for the stressed words (1 unstressed vowels/sentence X 4 stress conditions X 3 repetitions)[6].

Starting at the top of the graph with the display of vocal intensity as a function of experimental condition, the results indicate an increase in intensity was consistently elicited from each subject when they were asked to increase loudness or to emphasize a word. The mean vocal intensities (measured during the phonatory portion of each word), produced by the subjects as a group, during each experimental condition, are presented in table 2. Change in tracheal pressure (in kPa) as a function of experimental condition is displayed in the second part of figure 2. The results indicate subjects increased P_t when they were asked to increase overall loudness or to emphasize a word. This was a consistent finding across subjects. Mean R_{lx} (in kPa/(l/s))

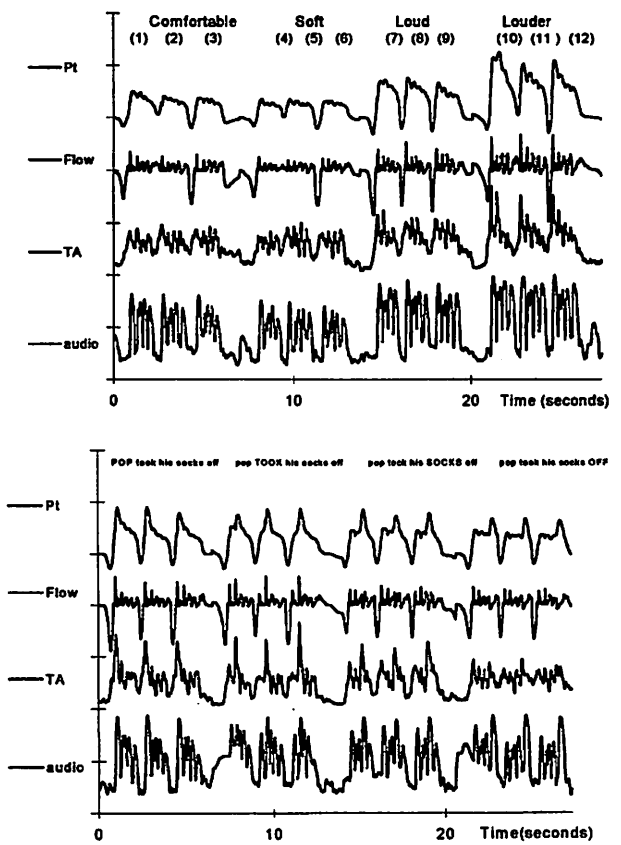


Figure 1. Samples of data collected from subject F3 during sustained and temporary changes in intensity.

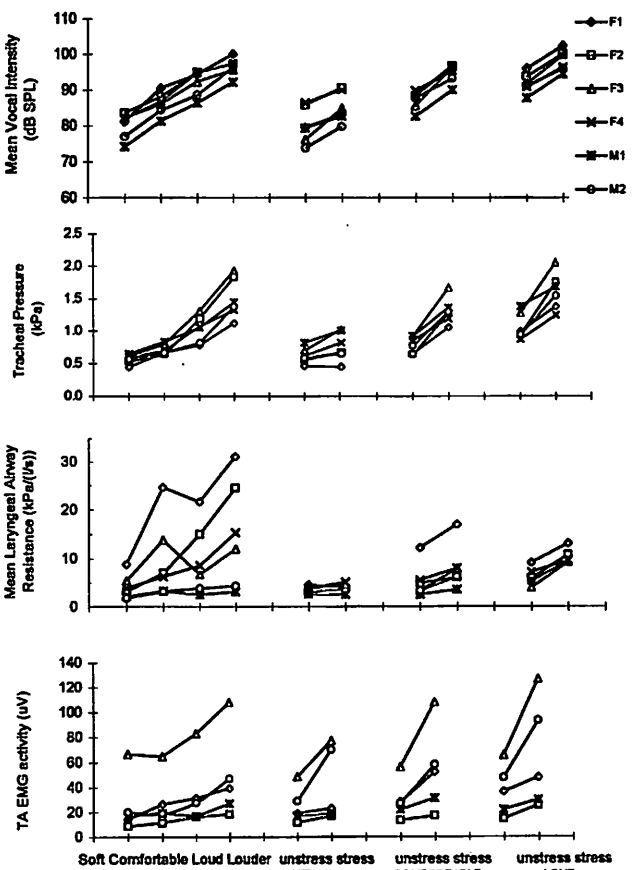


Figure 2.

[6] Since the data for calculating R_{lx} was limited to 2 of the 5 words in the sentence, the Rlx means are based on 6 tokens for the first four conditions, 9 tokens in the unstressed condition and 3 tokens in the stressed conditions.

Table 2.
Group Mean Activity Levels for Each ParameterAssociated with Each Experimental Condition

Parameter		Experimental Condition: Sustained change in Intensity				Temporary change in Intensity					
		Soft	Comf	Loud	Louder	soft		comfortable		loud	
						unstress	stress	unstress	stress	unstress	stress
Intensity	mean	80.0	86.2	91.9	96.5	80.2	85.1	87.1	94.7	91.9	98.1
(dB)	sd	2.4	2.9	3.1	2.6	3.2	2.8	3.8	2.1	3.4	1.8
P_t	mean	.48	.64	.92	1.35	.53	.63	.70	.99	.98	1.39
(kPa)	sd	.06	.06	.15	.26	.11	.13	.11	.12	.12	0.15
R_{lx}	mean	4.2	9.7	9.7	15.0	3.5	3.9	5.3	8.1	6.3	10.5
(kPa/(l/s))	sd	1.23	4.86	4.35	7.70	.68	.59	1.51	.99	2.10	1.54
TA EMG	mean	25.2	28.1	34.9	47.8	25.0	41.1	29.3	53.5	37.4	64.7
(μV)	sd	6.6	8.3	9.6	21.0	7.5	13.7	9.4	27.3	10.8	28.8

used by subjects during "took" and "his" in the different experimental conditions is displayed in the third part of figure 2. These results indicate a tendency for subjects to increase R_{lx} with increase in overall intensity and with word emphasis. Mean rectified TA EMG (in microvolts) used by subjects in the different experimental conditions is displayed in the bottom portion of figure 2. TA EMG activity was not obtained from subject F4. Increase in TA EMG activity associated with both sustained and temporary changes in intensity was consistent across subjects. Substantial changes in EMG could be seen in figure1. These data were from subject F3, the subject with the greatest EMG activity. However, the peaks in TA activity in figure 1 were associated primarily with pre and post-phonatory activity. The level of activity during phonation in all subjects was lower. When TA activity was calculated as a percent of maximum activity, it was found that the TA muscles were minimally active, 15% of maximum activity (on average), during soft phonation, increasing to 30% of maximum activity during loud phonation.

Correlations

Figure 3 shows scatter plots of the relationship between each parameter and vocal intensity for each subject, irrespective of whether the data were obtained from sustained or temporary change in intensity. These data were originally plotted for each of the four conditions to examine the relationship between intensity and each parameter when there was a sustained change in intensity or a temporary change during comfortable speech, during soft speech and during loud speech. However, since the relationship between each parameter and intensity was similar across condition (see figure 2), the data for all conditions were combined, resulting in the scatter plots provided in figure 3. The parameters are arranged by column, starting at the left with tracheal pressure and intensity in the first column, R_{lx} and intensity in the second column, and TA EMG and intensity in the third column. Subject data is arranged in rows starting at the top with subject F1 and ending at the bottom with subject M2. The scatter plots for P_t and TA EMG (columns 1 and 3) contain 240 data points (48 sentence productions X 5 vowels), while the scatter plot for R_{lx} contains 84 data points (48 productions X 2 vowels). The correlation values obtained from linear regression of these scatter plots are provided in table 3. The correlations for each condition are also provided in the table.

The consistent pattern across subjects, in all experimental conditions, was for P_t to be more highly correlated with intensity than were either R_{lx} or TA EMG. The r^2 values between P_t and intensity ranged from .79 to 1.0. The high correlation between P_t and vocal intensity across subjects can be seen in the scatter plots in figure 3, and the consistency of this pattern within each experimental condition can be seen in table 3. These high correlations found between P_t and intensity for both sustained and temporary increases in intensity are consistent with the data previously

presenting in figure 1. The scatter plots in column 2 indicate a tendency for R_{lx} to increase with intensity. The r^2 values R_{lx} and intensity were quite variable, ranging from .04 to .85. R_{lx} was more highly correlated to intensity during soft productions (group mean of .55) than during loud productions (group mean of .35). The greater degree of scatter in the plots of intensity as a function of R_{lx} indicates that there can be large variations in intensity not associated with any change in R_{lx}, over a wide range of values of R_{lx}. There was a low positive relationship between TA EMG and intensity, with r^2 values ranging from .00 to .53.

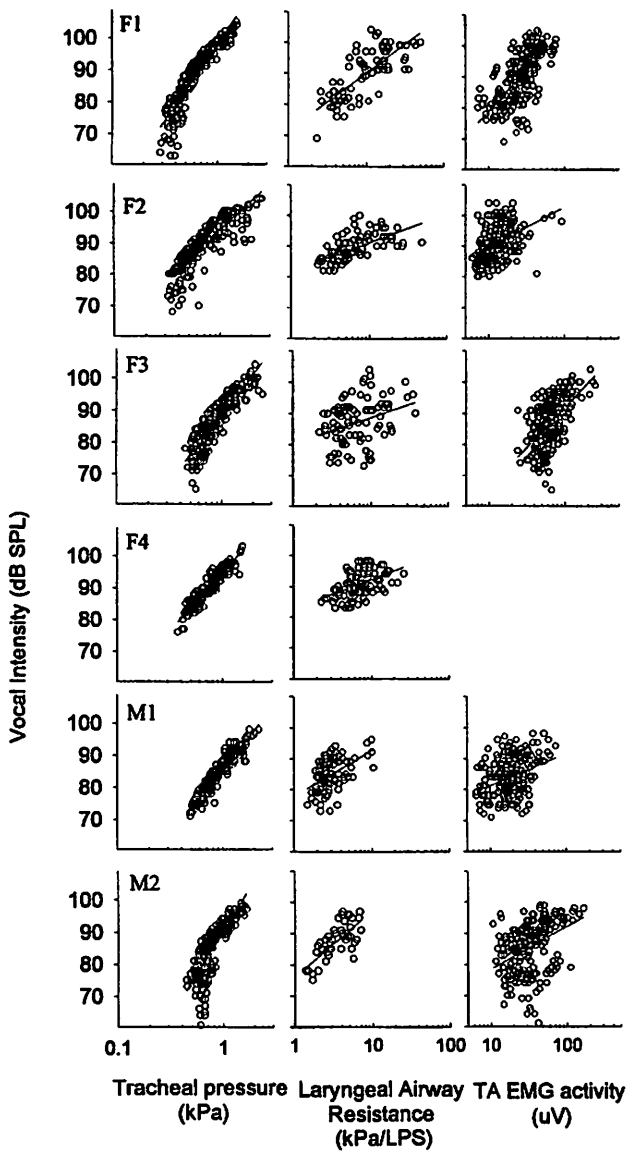


Figure 3.

Although the correlations of P_t with intensity, obtained with linear regression, were high, the data (as seen in column 1 of figure 3) was slightly curvilinear. A more linear fit was obtained by factoring threshold pressure into the equation in the manner described in Titze & Sundberg (1992). When P_{th} was assumed to be .2 kPa, the group mean increase in intensity was 9.1 dB +/- 2.1 dB per fractional excess of lung pressure over threshold $((P_t - P_{th})/P_{th})$, consistent with previous findings found during sustained phonation by Titze & Sundberg (1992). The results of this analysis are shown in figure 4.

R_{lx} as a Function of TA EMG: Correlations between TA activity and R_{lx} during sustained intensity change are shown in figure 5 (based on 84 data points/ graph). These results indicate no correlation between TA EMG activity and R_{lx}, with r^2 values ranging from .02 to .30. This lack of correlation was consistent across subjects and conditions.

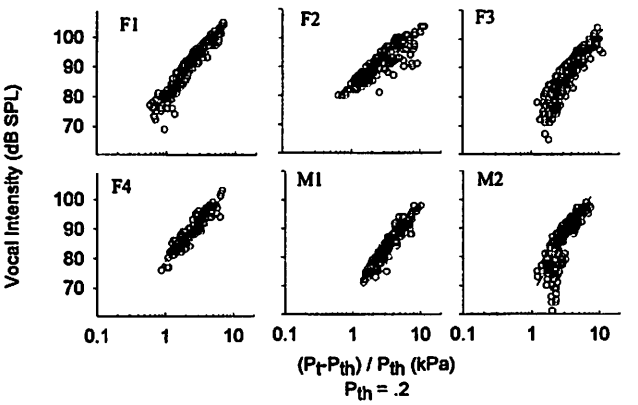


Figure 4.

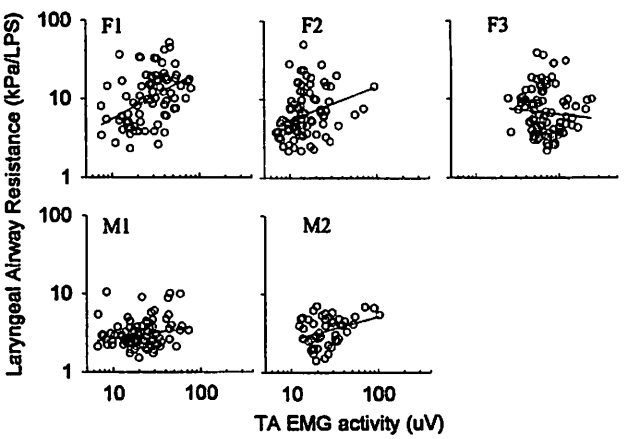


Figure 5.

Table 3.
Correlations of P_t, R_{lx}, and TA EMG with Vocal Intensity for Each Condition for Each Subject

	All data combined			Sustained Change in Intensity			Temporary Change in Intensity								
							Soft			Comfortable			Loud		
	Pt	Rlx	TA	Pt	Rlx	TA	Pt	Rlx	TA	Pt	Rlx	TA	Pt	Rlx	TA
F1	.98	.60	.53	.99	.04	.00	.97	.41	.48	1.0	.52	.25	.97	.58	.45
F2	.94	.49	.33	.93	.16	.14	.95	.37	.11	.96	.32	.05	.95	.40	.20
F3	.88	.07	.21	.90	.31	.17	.86	.83	.32	.89	.85	.48	.90	.11	.31
F4	.95	.44	*	.94	.21	*	.98	.29	*	.95	.17	*	.94	.23	*
M1	.86	.15	.12	.86	.04	.06	.85	.62	.25	.87	.20	.05	.86	.25	.13
M2	.90	.57	.34	.92	*	.08	.79	.77	.35	.91	*	.35	.91	.53	.15
mean	.92	.39	.31	.92	.15	.09	.90	.55	.30	.93	.41	.23	.92	.35	.25
sd	.04	.22	.15	.04	.12	.07	.08	.22	.13	.05	.28	.19	.04	.18	.13

Reliability

Repeat measures were made of 10% of the data for purposes of determining reliability. The same investigator made the repeat measures. The remeasurement was performed 2 months after the initial analysis. Correlations of .978 were obtained on measures of intensity, .993 for tracheal pressure, and .948 for airflow activity were obtained on measures of mean level of activity during phonation.

Discussion

The purpose of this study was to test the hypothesis that the respiratory system plays a primary role in changes in overall intensity, while adjustment of the laryngeal system is primary in short-term intensity changes related to emphasis. Contrary to the outcome predicted by the hypothesis, there was no evidence that different strategies were used to control intensity based on whether the change in intensity was temporary or sustained. It was clear, at least in the context of the sentence task used in this study, that respiratory drive, as indicated by tracheal pressure and estimated alveolar pressure, increased for both long and short-term changes in vocal intensity. The laryngeal indices, laryngeal airway resistance (R_{lx}) and electromyographic activity from the thyroarytenoid (TA EMG), also tended to increase with intensity. The pressure provided by the respiratory system was more highly correlated with intensity change than were either R_{lx} or TA EMG.

The Role of the Respiratory System: The results of this study were consistent with previous studies that have reported a correlation between P_t and intensity during prolonged vowels. A strong correlation between these two variables was found during short-term intensity changes related to emphasis as well as sustained changes in overall intensity level. This change in P_t associated with adjustment in intensity was consistent within and across subjects. Lack of correlation between lung volume measures and speech intensity during speech has been attributed to the possibility that subjects are changing intensity via laryngeal adjustment rather than altering respiratory drive. The results of this study, however, indicate respiratory drive does change during short-term increases in intensity. The respiratory system appears to be capable of producing rapid changes in pressure and this strategy was used by subjects to increase intensity for word stress. Lung volume measures provide information regarding whether the respiratory system is operating at a high volume where recoil pressures are high, but change in volume does not necessarily indicate alteration in respiratory drive (i.e. alveolar pressure). Bouhuys et al. (1966) reported that it is possible to have large changes in lung volume without change in pressure. They found that during production of a sustained tone, subjects kept tracheal pressure relatively constant even as pleural pressure increased from -.3 to 2.5 kPa and lung volume decreased from 95% to 10% vital capacity. They suggested that this constant pressure was maintained by continuously balancing recoil and muscular forces during expiration. According to Boyle's Law, at a constant temperature, the product of pressure and volume must remain constant as a given amount of gas is compressed or expanded (Ohanian, 1989). During speech, however, the respiratory system is not operating as a closed system. The amount of gas is not constant since there is

airflow. During dynamic conditions, as in speech, it is only possible to determine P_t by direct measurement.

An interesting, although unanticipated, finding of this study was the extent to which respiratory drive is modulated during speech. Previous studies have found that linguistic factors influence breathing during speech. Winkworth et al. (1994) reported that breaths are timed such that they are taken between sentences and phrases, and that deeper breaths are taken prior to the start of a sentence rather than at the start of a phrase within a sentence. They suggested that "the length and type of utterance is anticipated by the CNS"(p.553). They also reported increased initial lung volume (ILV) and lung volumes inspired associated with higher ranges of intensity and suggested that this may indicate a "degree of neural planning of respiratory activity for the intensity of the utterance" (p. 552) as well. The novel finding in our data was that respiratory drive fluctuated within the utterance. Increases of .5 kPa in respiratory drive were associated with a single stressed word in the sentence when speaking at a comfortable level (see table 4). It was not the case that a steady drive was provided by the respiratory system with modulation of tracheal pressure during phonation controlled by alterations in laryngeal resistance. Even during a sustained increase in intensity, some degree of modulation in respiratory drive was present. It has been previously suggested that constant pressure could be maintained by balancing recoil and muscular forces. The muscles associated with controlled expiration involve inspiratory (diaphragm and external intercostal) as well as expiratory (internal intercostal, rectus abdominis, external oblique and latissimus dorsi) muscles. As Bouhuys et al. (1966) speculated, it could be possible "to achieve fine control of subglottic pressure over nearly the full range of vital capacity, by use of inspiratory rib cage muscles and the hydraulic pull of [the] abdominal contents to control pleural pressure at lung volumes above functional residual capacity" (p.496).

The Role of the Laryngeal System: The findings of a tendency to increase R_{lx} with intensity and of a higher R_{lx} in female than in male subjects were consistent with results of previous investigations of R_{lx} during a syllable repetition task (Holmberg et al., 1988; Leeper & Graves, 1984; Stathopoulos & Sapienza, 1993a; Wilson & Leeper, 1992). However, the correlation between R_{lx} and intensity was weak to moderate, indicating that, during sentence production, R_{lx} could vary widely without necessarily effecting intensity. These results are consistent with the findings of Titze (1988).

An important role of the larynx is the conversion of tracheal air pressure to sound pressure. Adjusting the vocal folds in order to decrease P_{th} or increase amplitude of vibration could increase the efficiency of this conversion. It is possible to decrease P_{th} by adducting the vocal fold to decrease prephonatory width or by shortening the vocal fold to increase the thickness (Titze, 1992). In addition to decreasing P_{th}, shortening the vocal fold may also increase the amplitude of vibration by loosening the cover of the vocal fold (Titze & Talkin, 1979). Tanaka and Tanabe (1986) found that, in the in-vivo canine model, at constant aerodynamic power, contraction of TA increased resistance but not intensity. However, at constant tracheal pressure, there was a decrease in intensity with contraction of TA. This thesis found, however, that in these subjects, increase in TA EMG was correlated with an increase in intensity but not an increase in resistance. The lack of correlation between R_{lx} and TA EMG is inconsistent with suggestions that one purpose of contraction of the TA is to increase resistance, and thereby increase subglottic pressure. It could be that depending on the activity of other laryngeal muscles, contraction of the TA could shorten the vocal fold and increase amplitude of vibration without necessarily increasing resistance. This could explain the finding that although both R_{lx} and TA EMG tended to covary with intensity, they were not correlated with each other. We are unable to be certain from the results of this study, since we did not have any data regarding whether contraction of TA was associated with shortening of the muscle.

Acknowledgment

This research was supported by grant no. P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders. Special thanks to Ingo Titze, Michael Karnell, Kelly Cole, and Charles A. Miller for their comments on an earlier version of this paper.

Bibliography

Alipour, F., Scherer, R.C., & Finnegan, E.M. (1997). Pressure-flow relationships during phonation as a function of adduction. Journal of Voice, 11(2), 187-194.

Bouhuys, A., Proctor, D.F., & Mead, J. (1966). Kinetic aspects of singing. Journal of Applied Physiology, 21, 483-496.

Dromey, C., Ramig, L. O., & Johnson, A. B. (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: a case study. Journal of Speech & Hearing Research, 38(4), 751-64.

Faaborg-Andersen, K. (1957). Electromyographic investigation of intrinsic laryngeal muscles in human. Acta physiol. scand., 41, suppl.140.

Gay, T., Hirose, H., Strome, M., Sawashima, M. (1972). Electromyography of the intrinsic laryngeal muscles during phonation. Annals of Otolaryngology, 81, 401-409.

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

Hirano, M. (1987). The laryngeal muscles in singing. Neurolaryngology: recent advances. Eds. M. Hirano, J. A. Kirchner, D.M Bless. Singular Publishing Group, Inc.; San Diego. (pp. 209- 2230).

Hirano, M., Ohala, J., & Vennard, W. (1969). The function of laryngeal muscle in regulating fundamental frequency and intensity of phonation. Journal of Speech and Hearing Research, 12, 616-628.

Hixon, T.J., Goldman, M.H., & Mead, J. (1973). Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung. Journal of Speech and Hearing Research, 19, 297-356.

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

Holmes, L.C., Leeper, H.A., & Nicholson, I.R. (1994). Laryngeal airway resistance of older men and women as a function of vocal sound pressure level. Journal of Speech and Hearing Research, 37 (4), 789-799.

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

Koyama, T., Kawasaki, M., & Ogura, J. H. (1969). Mechanics of voice production. I. Regulation of vocal intensity. Laryngoscope, 79(3), 337-54.

Kunze L.H. (1964). Evaluation of methods of estimating subglottal air pressure. Journal of Speech and Hearing Research, 7, 151-164.

Ladefoged, P., & Mc Kinney, N. (1963). Loudness, sound pressure and subglottic pressure in speech. Journal of the Acoustical Society of America, 35, 454-460.

Leeper, H.A. Jr., & Graves, D.K. (1984). Consistency of laryngeal airway resistance in adult women. Journal of Communication Disorders, 17(3), 153-63.

Ohanian, H.C. (1989). Physics. New York : W. W. Norton & Co. (pp. 494-496).

Ramig, L.O., & Dromey, C. (1996). Aerodynamic mechanisms underlying treatment-related changes in vocal intensity in patients with parkinson disease, Journal of Speech and Hearing Research, 39(4), 798-807.

Rubin, H. J., LeCover, M., & Vennard, W. (1967). Vocal intensity, subglottic pressure and air flow relationships in singers. Folia Phoniatrica, 19(6), 393-413.

Russell, N.K. & Stathopoulos, E.T. (1988). Lung volume changes in children and adults during speech production. Journal of Speech and Hearing Research, 31, 146-55.

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

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

Stathopoulos, E.T. & Sapienza, C. (1993b). Respiratory and laryngeal measures of children during vocal intensity variation. Journal of the Acoustical Society of America, 94, 2531-43.

Tanaka, S. & Gould, W.J. (1983). Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects. Journal of the Acoustical Society of America, 73, 1316-21.

Tanaka, S. & Tanabe, M. (1986). Glottal adjustment for regulating vocal intensity. An experimental study. Acta Otot-Laryngologica, 102, 315-24.

Tang, J. & Stathopoulos, E.T. (1995). Vocal efficiency as a function of vocal intensity: a study of children, women, and men. Journal of the Acoustical Society of America, 97, 1885-92.

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

Titze, I. R. (1991). A model for neurologic sources of aperiodicity in vocal fold vibration. Journal of Speech & Hearing Research, 34(1), 460-472.

Titze, I.R. (1992). Phonation threshold pressure: A missing link in glottal aerodynamics. Journal of the Acoustical Society of America, 91(5), 2926-2936.

Titze, I. (1994). Principles of Voice Production. Prentice Hall, New Jersey.

Titze, I.R., Luschei, E.S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. Journal of Voice, 3, 213-224.

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

Titze, I. R., & Talkin, D. T. (1979). A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. Journal of the Acoustical Society of America, 66(1), 60-74.

Van den Berg, J.W. (1959). Direct and indirect determination of the mean subglottic pressure. Folia Phoniatria, 8, 1-24.

Wilson, J.V. & Leeper, H.A. (1992). Changes in laryngeal airway resistance in young adult men and women as a function of vocal sound pressure level and syllable context. Journal of Voice, 6 (3), 235-245.

Winkworth, A.L., Davis, P.J., Ellis, E., & Adams, R.D. (1994). Variability and consistency in speech breathing during reading; Lung volumes, speech intensity, and linguistic factors. Journal of Speech and Hearing Research, 37, 535-536.

Winkworth, A.L., Davis, P.J., Adams, R.D., & Ellis, E. (1995). Breathing patterns during spontaneous speech. Journal of Speech and Hearing Research, 38, 124-44.

NCVS Status and Progress Report - 12
June 1998, 23-31

Active and Passive Characteristics of the Canine Cricothyroid Muscles

Fariborz Alipour, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Active and passive characteristics of the canine cricothyroid muscle were investigated through a series of experiments conducted *in vitro* and compared with their counterparts in the thyroarytenoid muscle. Due to the important role of the cricothyroid muscle in phonation and respiration, its mechanical characteristics are crucial in any quantitative model. Samples from two separate portions of canine cricothyroid muscle, namely, the pars recta and pars obliqua were dissected from dog larynges excised a few minutes before death and kept in Krebs-Ringer solution at a temperature of 37 ± 1°C and a pH of 7.4 ± 0.05. The passive stress-strain characteristics were obtained by slow sinusoidal stretch and release (1 Hz). Active tetanic stress was obtained in isometric and isotonic conditions by applying field stimulation to the muscle samples through a pair of parallel-plate platinum electrodes and using a train of square pulses of 0.1 ms duration and 85 volt amplitude. Force and elongation of the samples were obtained electronically with a Dual Servo System (ergometer). The results indicate that the dynamic response of the cricothyroid muscle is almost twice as slow as that of the thyroarytenoid muscle. The average 50% tetanic contraction time for pars recta and pars obliqua were 84 ms and 109 ms respectively, in comparison to 50 ms for thyroarytenoid. The examination of force-velocity response of this muscle indicates a maximum shortening velocity of 2-3 times of its length per second, which is about half of the thyroarytenoid shortening speed. The passive properties of the pars recta and pars obliqua portions are similar to those of thyroarytenoid muscle.

Introduction

The study of voice production from a biomechanical point of view requires quantification of the active and passive properties of laryngeal muscles. Of these muscles, the thyroarytenoid muscle (TA) and the cricothyroid muscle (CT) are directly involved in pitch control. But these muscles also play an important role in the control of the vocal fold adduction and abduction (Moore and Berke 1988; Lofqvist, Baer, McGarr, and Story, 1989; Tully and Brancatisano, 1990; Titze, 1993; Nasri, Sercarz, Azizzadeh, Kreiman, and Berke 1994). In addition, the cricothyroid muscle acts as a pharyngeal dilator and glottal constrictor (Amis et al., 1992), thereby being further involved in the control of adduction (Tanaka and Tanabe, 1986). The dynamic characteristics of the CT and TA muscles are therefore key factors in any discussion and interpretation of phonatory control.

A model of vocal fold posturing (adduction and length change) requires the time course of muscle forces prior to and during phonation. Moreover, a physiologically-based model of vocal fold vibration relies on the knowledge of active and passive properties of these muscles (Titze and Talkin, 1979; Fujimura, 1981; Alipour and Titze, 1988). Thus, a better knowledge of the active and passive properties of these muscles will enhance our understanding of the normal function of the larynx. The variability in active muscle contraction has also been useful in the analysis of neurologic jitter (Titze, 1991) in phonation and in preliminary research on laryngeal pacing for vocal fold paralysis (Kojima, Omori, Nonomura, Honjo, Isshiki, and Shimizu, 1991; Goldfarb, Keane, and Lowry, 1994).

The cricothyroid muscle performs multiple function with its separate bellies (Zaretsky and Sanders, 1992). With its more vertical portion (pars recta), it elevates the cricoid ring, and with its oblique and horizontal portion (pars obliqua), it causes a retraction of the cricoid cartilage, i.e., dorsal gliding with respect to the thyroid cartilage (Arnold, 1961). These actions control the length and tension of vocal fold, and thereby the pitch. According to Vilkman (1987), 30-40% of the elongation in vocal folds are due to the gliding of the cricoid cartilage. Furthermore, rotation and gliding of the cricoid cartilage play an important role in adduction and abduction of the vocal folds (Arnold, 1961; Vilkman, 1987).

Some of the active properties of the laryngeal muscles, such as twitch contraction times and maximum tetanic forces, have been reported in the past (i.e., Faaborg-Andersen, 1957; Martensson and Skoglund, 1964; Sato and Hisa, 1966; Hast, 1966, 1967; Mardini, McCarter, Neal, Wiederhold, and Compton, 1987; Alipour, Titze and Durham, 1987; Alipour, Titze and Perlman, 1989; Rice and Cooper, 1989). For example, twitch characteristics of the cricothyroid muscle were studied by Perlman and Alipour (1988) and tetanic responses of the cricothyroid were reported by Alipour, Perlman and Titze, (1991). The two groups reported on the isometric contraction of pars recta and pars obliqua portions of this muscle during twitch and tetanic contraction. Although the isometric response data provided information on the maximum force and speed of contraction of this muscle, a dynamic model of the cricothyroid muscle also requires knowledge of its isotonic tetanic contraction, which may exhibit a different time course at less than maximum force.

The magnitude of passive and active contractile muscle forces, as well as the time-related characteristics, such as speed of contraction and contraction period, are needed in the development of theoretical models that describe the dynamic behavior in the regulation of pitch and register. Such models can also be described as constitutive equations of the muscle (Titze and Alipour, in preparation), describing its active (isometric and isotonic) and passive viscoelastic behavior. To that end, a series of in vitro experiments have been conducted in our laboratory to study the active and passive properties of the vocalis muscle and cricothyroid muscles. The canine model was used throughout these experiments because viable human muscle tissue is too difficult to obtain for controlled *in vitro* studies.

The purpose of the present study was to quantify the active and passive characteristics of different portions of the cricothyroid muscles in the isometric and isotonic state and compare them with those of the thyroarytenoid muscles. The data are presented in a format that can lead to a constitutive equation for the muscle. These data include a time function for isometric contraction, a maximum tetanic stress-strain function, an isotonic stress-strain rate function, and a passive stress-strain function. The questions which are posed in this study are:

1. What is the time course of tetanic contraction in the CT muscle?
2. What is the isotonic force-velocity characteristics of the CT muscle?
3. What is the maximum active and passive stress-strain characteristics of the CT muscle?
4. How do these characteristics compare with those of the TA muscle?

Method

The methodology for sample preparation and activating muscle tissues are discussed in previous works (Alipour et al., 1987, 1989, 1991) and will be reviewed only briefly here. Viable samples of the cricothyroid muscle tissue were dissected and prepared from fresh canine cadavers weighing 20-25 kg. To assure distinction between the muscle bellies, the most anterior bundles of the CT muscle were used for pars recta samples and the most inferioposterior bundles were used for pars obliqua samples. Samples were made from the exterior portion of each muscle 2-3 mm in depth and 3-4 mm in width (the muscle also attaches internally on the thyroid cartilage). Small pieces of thyroid and cricoid cartilage were retained at the ends of each sample for natural boundary attachments. Tevdek (2-0) polyester sutures were threaded through these cartilages. The samples were maintained in Krebs-Ringer solution and aerated with 95% oxygen and 5% carbon dioxide. The samples were mounted in a water-jacketed organ bath chamber containing the electrolyte solution (Figure 1) and were maintained at a temperature of 37° ± 1°C and pH of 7.4 ± 0.05 during the

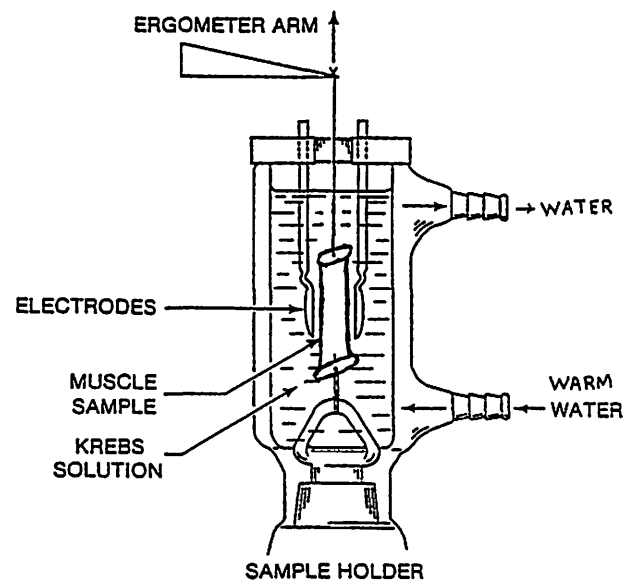


Figure 1. Schematic of sample mounting. Platinum electrodes are positioned around the sample for field stimulation. Warm water circulates in a jacket to keep the Krebs solution and the sample at the body temperature (about 37 C).

experiment. Force and elongation of the samples were controlled and measured with a dual-servo system (Cambridge Technology, model 305 ergometer). Before dissection, the longest and the shortest fibers of the sample were identified and measured to obtain an approximate average length of each sample *in situ*. When samples were then mounted vertically in the chamber, the same approximate average length could be set without any slack in the suture. This generally required an initial force of 1-3 grams on the pars recta samples and 2-5 grams on the pars obliqua samples.

Data for the active properties were obtained by applying electrical field stimulation to the muscle through a pair of parallel-plate electrodes with platinum surfaces positioned around the muscle (Figure 1). Tetanic contraction of the muscle was initiated by applying a train of square pulses of 0.1 ms duration and 85 volt amplitude at a frequency of 65 Hz transverse to the muscle fibers. This stimulation was determined to be supermaximal because any additional voltage did not increase the force of contraction. First the isometric response of each sample was recorded by setting the ergometer load to maximum (a rigid arm). This allowed the maximum tetanic force to be developed. Then by decreasing the ergometer load below the maximum tetanic force, the isotonic experiment was performed. In addition, the passive response of the muscle was obtained from separate viable samples through a sinusoidal stretch and release experiment (Alipour & Titze, 1991). The data were digitized with a 16 bit A/D converter and stored and analyzed on a VAX Station 3000 (Digital Equipment Corp).

Results

Figure 2 shows a typical isometric tetanic response of the cricothyroid pars recta muscle as a function of time. There are 7 curves, corresponding to zero through 6 mm elongation. For the type of artificial stimulation applied, each curve rises sharply to a maximum active stress in 200-300 ms and then decays slowly during the tetanus. These curves represent the greatest tetanic stresses that can be generated by the muscle because the load is theoretically infinite and stimulation is supermaximal. There is no change in the sample's length during this isometric contraction, but the experiment was repeated for different lengths. Thus, the isometric tetanic stress is a function of time and strain. But this function can be broken into two separate functions, a time dependent function and a strain dependent function. The time dependent function can be obtained by normalizing the tetanic force of each sample to its maximum tetanic

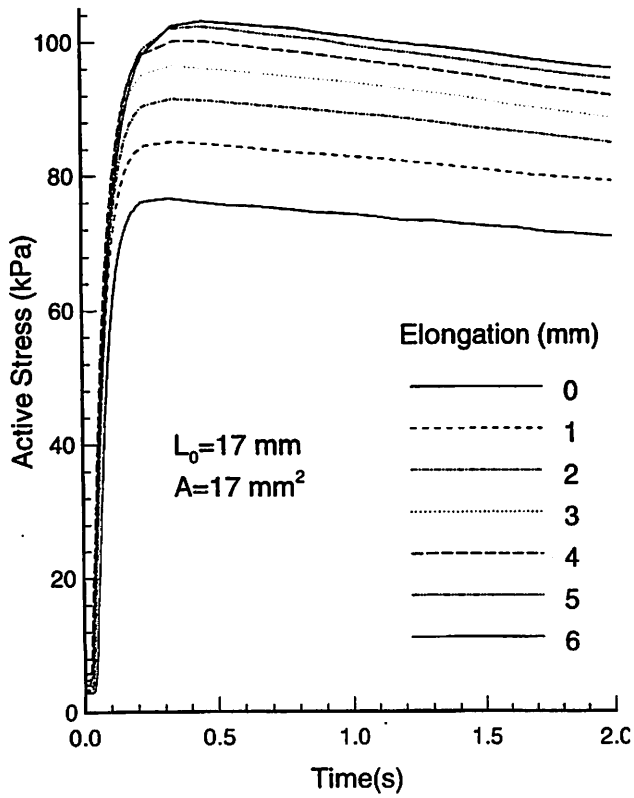


Figure 2. Typical tetanic response of cricothyroid muscle at various elongation. Stimulation was applied for about two seconds at 65 Hz.

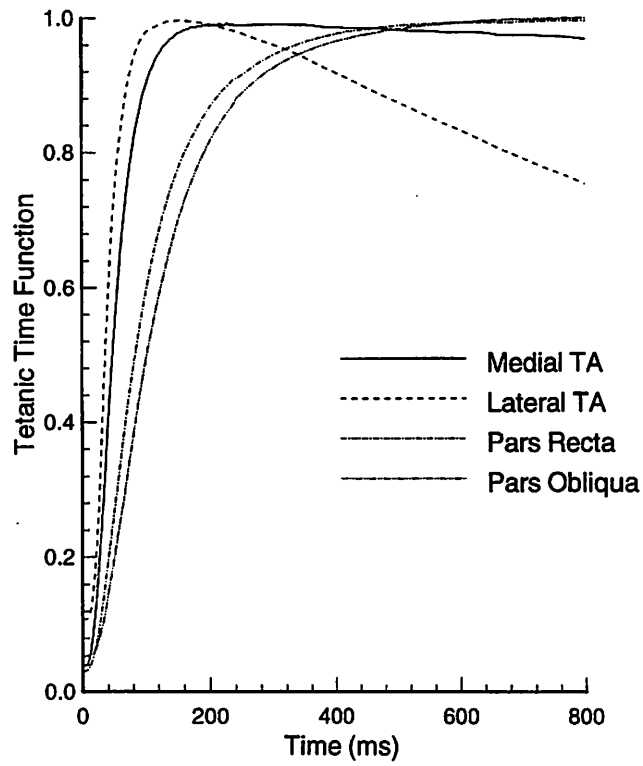


Figure 3. Averaged tetanic time functions of the thyroarytenoid and cricothyroid muscles. The active tetanic force of each sample was normalized to its peak value before averaging.

force. The strain dependent function can be obtained by the evaluation of maximum tetanic stress at various strain levels.

The normalized tetanic stress (time function) was averaged across 7 to 9 muscle samples and compared to data for the TA muscle (obtained previously). Figure 3 shows the averaged tetanic contraction (time function) of the canine TA muscle (medial and lateral portions) compared to the CT muscle (pars recta and pars obliqua portions). These curves start from a nonzero value with zero slope and rise rapidly within the first 300 ms to approach their peak values (1.0). According to Stein and Oguztoreli (1976), muscle forces behave like second order (or higher order) low-pass filters. The filter characteristics are sensitive not only to length changes, but also to the derivatives of velocity (acceleration). These dynamic properties of the muscles can be observed in the slopes of their time function. If we numerically differentiate the experimental time functions, we find some similarities between the slopes and the twitch response of these muscles (Figure 4). Stein and Oguztoreli (1976) regarded the muscle transfer function as the Laplace transform of its twitch response. In actuality, a tetanus is the summation of a series of twitches, the first of which defines the start-up and later ones contribute to the continual rise in force. These slope of the time function of Figure 3 can be modeled by an exponential equation of the form:

$$s(t) = \frac{1}{t_i}\left(\frac{t}{t_i}\right)e^{-\frac{t}{t_i}} \qquad (1)$$

where t_i is a time constant that can be determined by the fitting experimental data to equation (1). This function, which defines the speed of muscle contraction, has a maximum of $1/t_i e^{-1}$ (or $0.36788/t_i$) at $t=t_i$. Thus, the time constant defines both the location of the maximum slope and its value.

Now we can integrate equation (1) to find the tetanic time function:

$$\sigma(t) = 1-\left(1+\frac{t}{t_i}\right)e^{-\frac{t}{t_i}} \qquad (2)$$

As an example, the time constant for the medial portion of TA muscle averaged about 30 ms, and about 50 ms for the CT pars recta muscle. This time constant reflects the internal sarcomere contraction speed. As seen in Figure 3, it is highest for lateral portion of the TA and lowest for pars obliqua portion of the CT muscle.

There is no standard definition for tetanic contraction time. If we consider it to be the time required for the muscle to achieve its 100% tetanic force, it may exceeds 300 ms. Since the major task of the CT is to control vocal fold length and tension, its contraction is neither isometric nor isotonic. The load and length are both variable. Thus, tetanic force in the CT muscle may never reach its 100% maximum value. This causes the actual contraction time to be shorter than 300 ms. Using equation (2), we can calculate the contraction times for less than 100% force. For example, for 50% contraction, the required time is equal to $1.676t_i$, and for 90% it is equal to $3.887t_i$.

Table 1 shows the maximum slope and these contraction times for each muscle. The fastest muscle is the lateral TA and slowest is the CT pars obliqua. These findings are consistent with the results of twitch contraction for these muscles (Faaborg-Andersen, 1957; Martensson and Skoglund, 1964; Hast ,1966, 1967; Perlman & Alipour 1988).

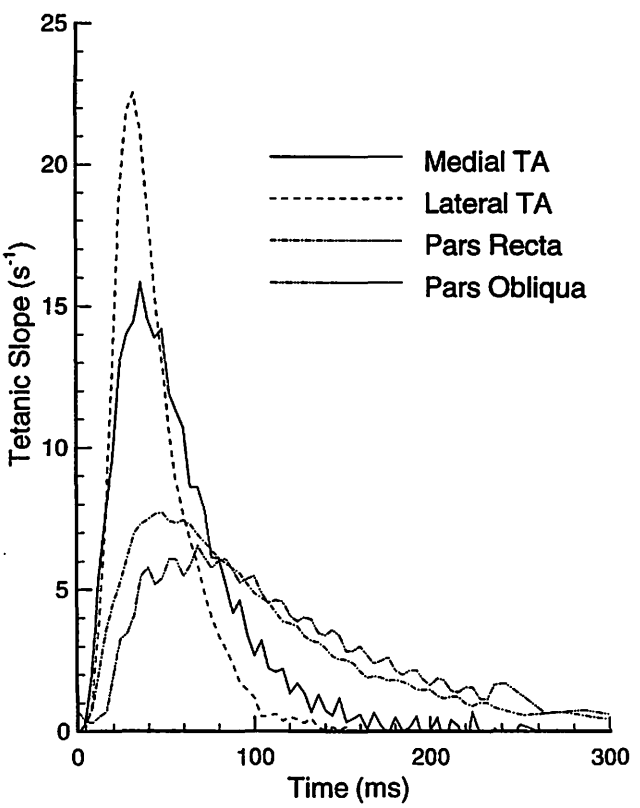


Figure 4. Slope of tetanic curves for the thyroarytenoid and cricothyroid muscles.

Table 1.
Tetanic Contraction Times of TA and CT Muscles

Muscle	Time constant, ms	Maximum Slope	contraction time, to 50%, ms	contraction time, to 90%, ms
TA, Lat	25	14.7	42	97
TA, Med	30	12.4	50	117
CT, PR	50	7.3	84	194
CT, PO	65	5.7	109	253

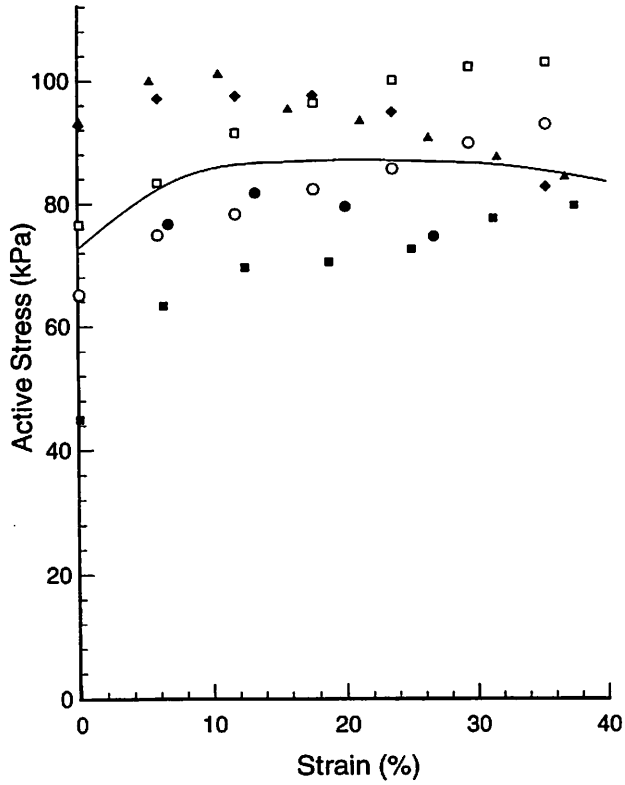


Figure 5. Active stress in samples of pars recta as a function of strain.

Figure 5 shows the maximum tetanic stress as a function of strain for 6 samples of the pars recta of the CT muscle. The average value is shown as a solid line, which can be expressed as a fourth order polynomial like:

$$\sigma_{am} = 72.5 + 236.3\varepsilon - 1400\varepsilon^2 + 3640\varepsilon^3 - 3630\varepsilon^4 \quad (3)$$

If this function is multiplied by the time function (eq. 2), the result is an average muscle characteristic in isometric contraction. The maximum active stress data reported here are a little different from those reported earlier by Alipour et al. (1991), which were based on stimulation of a sample while it was experiencing stretch and release (a quasi-isometric mode). This present study analyzed the maximum active stress for stepwise elongation, which is more compatible with the results for the thyroarytenoid muscle reported earlier (Alipour et al., 1989).

Figure 6 shows a typical isotonic response of the cricothyroid muscle. In the top graph, the negative elongation (or shortening) curves are plotted. In the bottom graph, the corresponding tetanic forces are shown. These data are obtained in a quick release experiment, in which the resistive force (load) that the muscle should overcome is lowered below its maximum tetanic force. At each level of force, the muscle contracts along the steep portion until it reaches the preset force level; then it stays constant at that level while the

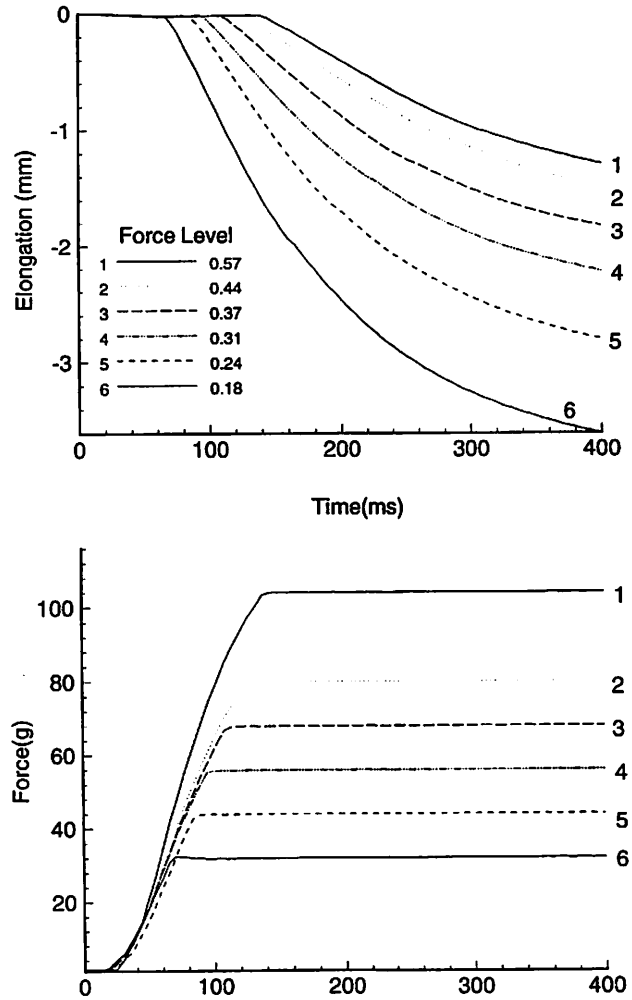


Figure 6. Typical isotonic response of the cricothyroid muscle. In a quick release experiment. As the level of force decreased from 57% (curve 1) to 18% (curve 6), the ergometer arm moved to lower positions at faster rate indication of a faster shortening.

length is decreasing. In the top figure, as the level of force decreases, the shortening curves become steeper in the initial descent. An inverse relationship between isotonic force and the shortening velocity is therefore noted in the dynamic characteristic of the skeletal muscle. A small amount of shortening, such as 1 mm, corresponds to a decrease in the tetanic force of 57%, which suggest that any laryngeal muscle with a possibility of a length decrease cannot generate its maximum tetanic force.

The maximum shortening velocity can be estimated from the slope of the initial portion of the shortening curve by linear modeling. Figure 7 (following page) shows the maximum velocity of shortening for samples of the CT muscle as a function of isotonic force. The 5 solid symbols are for the pars recta samples and the 5 hollow symbols are for the pars obliqua samples. Again, these data show that the pars obliqua portion of CT muscle is slower than the pars

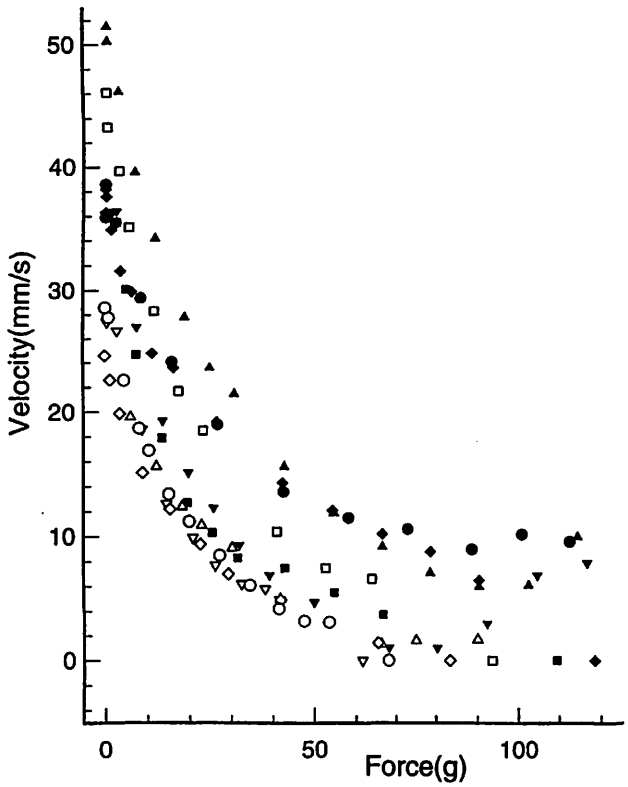


Figure 7. Isotonic force-velocity response of cricothyroid muscle. Solid symbols represent 5 pars recta samples and hollow symbols represent 5 pars obliqua samples.

recta. The data points indicate an inverse relation of the velocity and the force in this isotonic mode. At low levels of force, the velocity can reach up to 52 mm/s, which is about 2.7 times of the average sample's length per second.

When muscle models are used for pitch control and posturing, the shortening velocity is usually normalized to the sample length to yield the strain rate, and the isotonic force is normalized to the cross sectional area to obtain the active stress. Then a constitutive equation can be used to represent the dynamic behavior of these muscles. For example, one such model can be a hyperbolic equation:

$$\dot{\varepsilon} = \frac{A}{\sigma + B} \qquad (4)$$

where $\dot{\varepsilon}$ is the strain rate, σ is the active stress, and A and B are constants to be determined empirically.

Figure 8 shows combined data for samples of the thyroarytenoid muscle (squares), cricothyroid pars recta (triangles) and cricothyroid pars obliqua (circles) in normalized units. The data for each muscle are modeled with the above equation using a nonlinear least-square method (Levenberg-Marquardt). Solid lines are for the CT pars recta (PR), dash-dot lines for the CT pars obliqua (PO), and dashed lines for the TA. The parameters of the model are given in Table 2.

Table 2.
Hyperbolic Modeling of the Thyroarytenoid and Cricothyroid Mucles

Model's Parameter	Thyroarytenoid	Pars Recta	Pars Obliqua
A	41.227	22.018	15.083
B	6.186	8.919	6.254

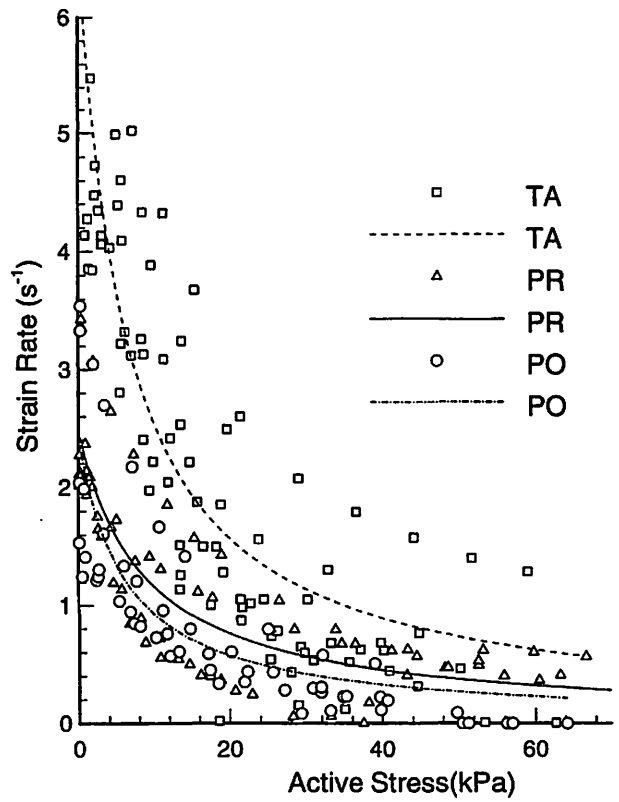


Figure 8. Isotonic stress-strain rate data and models of TA and CT muscles.

The isotonic data for the TA muscle were obtained previously by Alipour et al. (1989). The model shows that TA can generate strain rate as high as about 6.7 rest length per second while pars recta can generate strain rate of 2.5 and pars obliqua can generate strain rate of 2.4 rest length per second. These data and their models suggest that in the isotonic mode, the thyroarytenoid is almost 2.6 times faster than the cricothyroid muscle.

While the active properties of CT muscle are a major control factor in the dynamic models of the pitch control, the passive properties also play an important role. They contribute to the stiffness of the cricothyroid rotation. These passive properties are obtained by applying a slow sinusoidal stretch and release to the samples of cricothyroid muscles as described earlier by Alipour & Titze (1991). After a few cycles of stretch and release, when the resisting force stabilizes to a psudoelastic condition, data from stretch

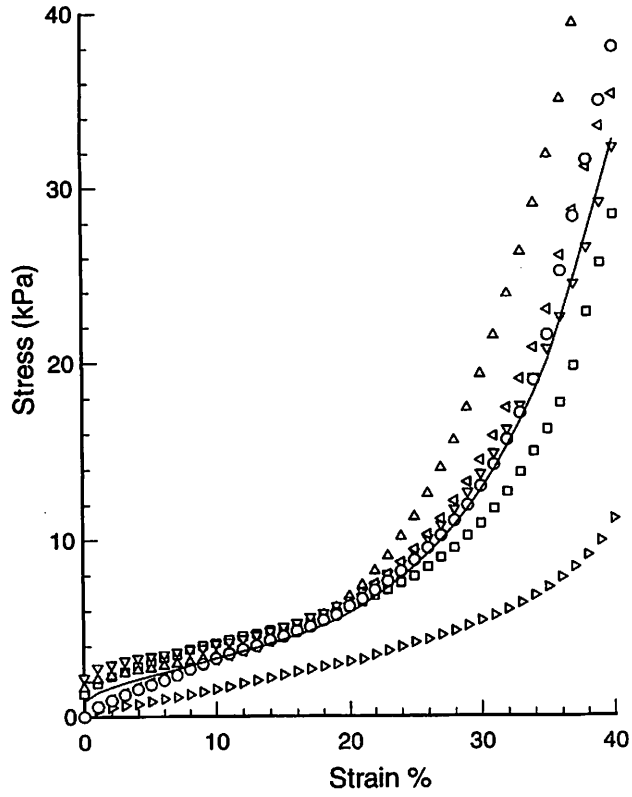


Figure 9. Passive stress-strain response of 5 samples of pars recta (symbols) with their average (solid line).

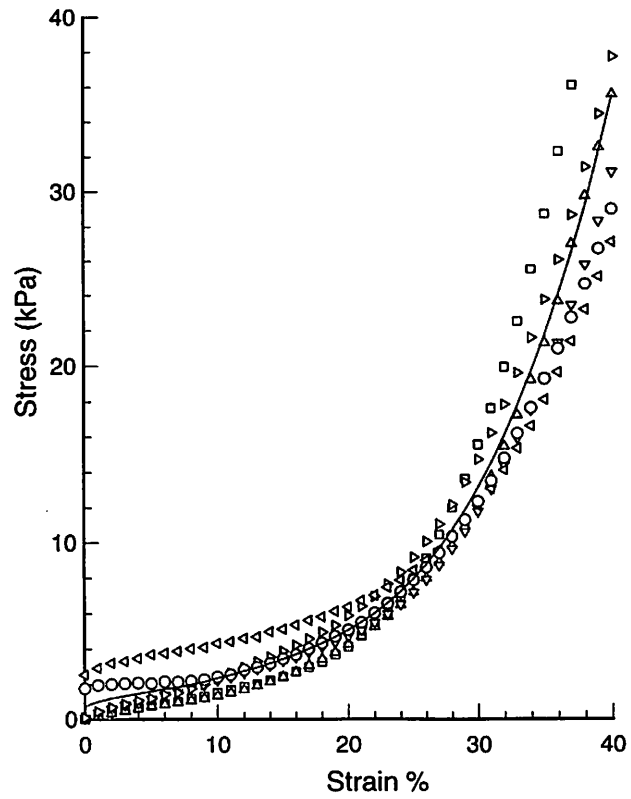


Figure 10. Passive stress-strain response of 5 samples of pars obliqua (symbols) with their average (solid line).

portion of a cycle are collected for stress-strain calculations. Figures 9 and 10 show the stress-strain relationships for samples of pars recta and pars obliqua respectively. In each case, the stress is obtained by dividing the passive force by the cross-sectional area. The strain is calculated from the ratio of elongation and the sample's original (resting) length. The symbols represent data points and the solid lines the average stress-strain curve for the samples. As noted in previous writings, the stress-strain curves appear linear at low strains and exponential at higher strains.

Figure 11 (following page) compares the average passive response of the cricothyroid muscles with those of canine vocalis muscle and mucosa (From Alipour & Titze, 1991). It appears that the CT muscle has passive properties similar to the vocalis muscle. The mucosa layer has a different response. These curves are almost linear at low strains and exponential at high strains. These curves can be modeled by polynomials or combination of the linear and exponential models. For example, a third order polynomial can be fitted to the average data within the strain range of 0-0.4 for the cricothyroid pars recta muscle as

$$\sigma = 0.3296 + 59.72\varepsilon - 384.44\varepsilon^2 + 1087\varepsilon^3 \qquad (5)$$

Similar equations can be written for pars obliqua and for various layers of the vocal fold tissue. The combination of equations (2-5), can then be used to set up a differential equation of motion for the cricothyroid muscle when a specified load or displacement is applied. This is the subject of another paper, however.

Discussion

The biomechanical properties of the cricothyroid muscle needed for quantitative modeling of pitch control, where the dynamics of rotation and gliding of the thyroid and cricoid cartilage around the cricothyroid joint remain a key problem to be solved. The response times of the muscles will help determine the speed in which pitch and register changes can be executed in phonation. In pitch changes, the cricothyroid muscle is likely to act in an isotonic mode, changing its length during contraction. For example, a vocal fold length increase of 0.7 to 3.5 mm per octave has been reported by Nishizawa, Sawashima, and Yonemoto (1988). A decrease of cricothyroid length up to 50% has been observed for high pitches (Arnold, 1961).

Sundberg (1979) has measured the maximum speed of pitch changes (over an octave or more) to be on the order

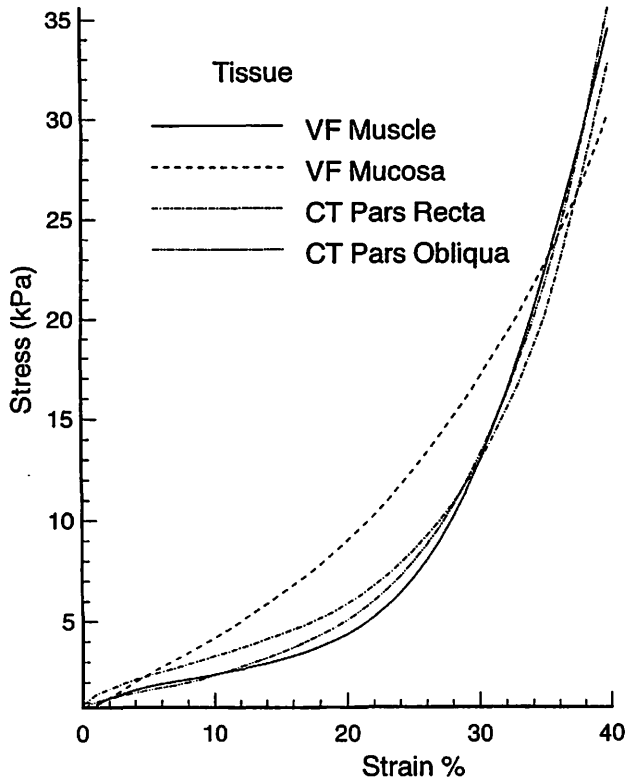


Figure 11. Average passive stress-strain curves of vocal folds muscle and mucosa and cricothyroid muscle.

of 60-100 ms. Since we have shown here that the CT muscle can change its length by about 250% in a second, a 200 ms response would be predicted for such a 50% length change. But this is a little overestimation because the TA muscle would likely oppose this length change with its own simultaneous contribution. The net length change would be less and the vocal fold stiffness would increase more rapidly because at least twice as fast. But for a paralyzed TA muscle, sole reliance on CT for pitch change may slow down the response time considerably.

For an alternate calculation, assume that the active force of the cricothyroid muscle is proportional to the fundamental frequency (Hirano, Ohala, and Vennard 1969; Arnold, 1961). The tetanic contraction time reported here for CT indicates that these muscles may take about 192 ms to reach 90% of maximum tetanic contraction and their contraction time for 50% of that maximum force is about 84 ms. This range of contraction is close to what is reported by Faaborg-Andersen (1957) on the time interval between the onset of change of action potential pattern and the onset of tone recorded by microphone in the cricothyroid muscle during phonation. Similarly, the tetanic contraction time of 117 ms for 90% activation of TA muscle is close to the latency that was reported by Shipp, Izdebski, and Morrissey (1984). The similarity passive response of the CT and TA muscles is useful in the body-cover modeling pitch control. The hyperbolic constitutive equation from the force velocity data (Eq. 3), is an essential in the dynamic modeling of pitch control.

Acknowledgment

This work was supported by research grant number P60 DC00976 from the National Institute on Deafness and Other Communication Disorder, National Institute of Health.

References

Alipour-Haghighi F, Titze IR, Durham PL. (1987). Twitch response in the canine vocalis muscle. *J. Speech Hear. Res.* 30:290-294.

Alipour-Haghighi, F. and Titze, I.R. (1988). A Finite element simulation of vocal fold vibration. In the *Proceedings of the Fourteenth Annual Northeast Bioengineering Conference*. (pp. 186-189) edited by J.R. LaCourse, IEEE, Durham, NH.

Alipour-Haghighi F, Titze IR, Perlman AL. (1989). Tetanic Contraction in Vocal Fold Muscle. *J. Speech Hear. Res.* 32:226-231.

Alipour-Haghighi, F. and Titze, I.R. (1991). Elastic Models of Vocal Fold Tissues. *J. Acoust. Soc. Am.*, 90(3):1326-1331.

Amis, T.C., Brancatisano, A., Tully, A., and Engel, L.A. (1992). Pharyngeal dilation associated with cricothyroid muscle contraction in dogs. *Journal of Applied Physiology* 73(2):762-766.

Arnold G.E. (1961). Physiology and pathology of the cricothyroid muscle. Laryngoscope 71:687-753.

Faaborg-Andersen K. (1957). Electromyographic investigation of intrinsic laryngeal muscles in humans. *Acta Physiologica Scandinavica*, (Suppl. 140) 41.

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

Goldfarb, D., Keane, W.M., and Lowry, L.D. Laryngeal pacing as a treatment for vocal fold paralysis. *Journal of Voice* 8(2):179-185, 1994.

Hast, M.H. (1966). Mechanical properties of the cricothyroid muscle. *Laryngoscope*. 76:537-548.

Hast, M.H. (1967). Mechanical properties of the vocal fold muscle. *Practica oto-rhino-laryngologica*. 29:53-56

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.

Kojima, H., Omori, K., Nonomura, M., Honjo, I., Isshiki, N., and Shimizu, Y. Electrical pacing for dynamic treatment of unilateral vocal cord paralysis. Experiment in long-denervated muscle. *Annals of Otology, Rhinology & Laryngology* 100(1):15-18, 1991.

Lofqvist, A., Baer, T., McGarr, N.S., and Story, R.S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85(3):1314-1321.

Mardini, I.A., McCarter, R.J., Neal, G.D., Wiederhold, M.L., and Compton, C.E. Contractile properties of laryngeal muscles in young and old baboons. *American Journal of Otolaryngology* 8(2):85-90, 1987.

Martensson, A., & Skoglund, C.R. (1964). Contraction properties of intrinsic laryngeal muscles. *Acta Physiologica Scandinavica*, 60:318-336.

Moore, D.M. and Berke, G.S. (1988). The effect of laryngeal nerve stimulation on phonation: a glottographic study using an in vivo canine model. *Journal of the Acoustical Society of America* 83(2):705-715.

Nasri, S., Sercarz, J.A., Azizzadeh, B., Kreiman, J., and Berke, G.S. (1994). Measurement of adductory force of individual laryngeal muscles in an in vivo canine model. *Laryngoscope* 104(10):1213-1218.

Nishizawa, N., Sawashima, M., and Yonemoto, K. (1988). Vocal fold length in vocal pitch change. In *Vocal Physiology: Voice Production, Mechanisms and Functions.* (pp. 75-82) edited by O. Fujimura. Raven Press, Ltd., New York

Perlman AL, Alipour-Haghighi F. (1988). Comparative studies of physiological properties of laryngeal muscles. Acta Otolaryngol. (Stockholm) 105:372-378.

Rice, D.H. and Cooper, D.S. Contractile properties of canine thyroarytenoid muscle reinnervated from the ansa cervicalis. *Annals of Otology, Rhinology & Laryngology* 98(2):153-156, 1989.

Sato F, Hisa Y. (1966). Mechanical properties of the intrinsic laryngeal muscles and biomechanics of the glottis in dogs. In Hast MH. Mechanical properties of cricothyroid muscle. Laryngoscope 76:489-497.

Shipp, T., & McGlone, R.E. (1971). Laryngeal dynamics associated with vocal frequency change. *Journal of Speech and Hearing Research*, 14:761-768.

Shipp, T., Izdebski, K., and Morrissey, P. (1984). Physiologic stages of vocal reaction time. *Journal of Speech & Hearing Research* 27(2):173-178.

Stein, R. B. and Oguztoreli, M.N. (1976). Does the velocity sensitivity of muscle spindles stabilize the stretch reflex? *Biological Cybernetics* 22:219-228.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics*, 7:71-79.

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

Titze I.R, Talkin D.T. (1979). A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. *J. Acoust. Soc. Am.* 66:60-74.

Titze I.R, Jiang J, Druker D.G. (1988). Preliminaries to the body-cover theory of pitch control. Journal of Voice 1:314-319.

Titze I.R. (1991). A model for neurologic sources of aperiodicity in vocal fold vibration. *Journal of Speech & Hearing Research* 34(3):460-472.

Titze, I.R. (1993). Current topics in voice production mechanisms. [Review]. *Acta Oto-Laryngologica* 113(3):421-427.

Titze, I.R. and Alipour, F. *The Myoelastic-Aerodynamic Theory of Phonation*, in preparation.

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

Vilkman E. (1987). An apparatus for studying the role of cricothyroid articulation in the voice production of excised human larynges. *Folia phoniat.* 39: 169-177.

Zaretsky, L.S. and Sanders, I. (1992). The three bellies of the canine cricothyroid muscle. *Annals of Otology, Rhinology, & Laryngology - Supplement* 156:3-16.

NCVS Status and Progress Report - 12
June 1998, 33-41

Indications of an Optimum Glottal Width in Vocal Production

David A. Berry, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Katherine Verdolini, Ph.D.
Department of Otology and Laryngology, Harvard Medical School
Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary
Voice/Speech/Swallowing Division, Beth Israel Deaconess Medical Center and Brigham and Women's Hospital
Department of Communication Sciences and Disorders, Massachusetts General Hospital Institute of Health Professions
Roger W. Chan, B.Sc.
Department of Speech Pathology and Audiology, The University of Iowa
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts

Abstract

Designed as a benefit/cost ratio in voice production, vocal economy is defined as the ratio of the acoustic output intensity to the collision intensity of the vocal folds. Vocal economy is given a physical and mathematical basis. Application is extended to quantitative laboratory studies, positioning the measure for eventual use in the clinic. Tests were conducted in two experimental situations: (1) with an excised canine larynx, which did not include a vocal tract, and (2) with a computer simulation, which included a vocal tract for the vowels /ɑ/, /i/, and /u/. Data were gathered at constant fundamental frequency (150 Hz), but subglottal pressure was varied from 1.0 - 1.6 kPa and glottal width was varied from a pressed condition to a 3 mm gap. The results for the excised larynx experiment identified a clear economy maximum at a glottal width of 2 mm. In contrast, the computer simulations produced economy maxima at around 1 mm for /ɑ/, /i/ and /u/, although the maxima were less well defined for /u/. These findings provide additional evidence in support of an optimal glottal width in vocal production, and have potential relevance for the treatment of a wide range of voice disorders, including those involving both hyper- and hypoadduction.

The general concept of an output/cost ratio has been considered important in voice physiology for decades. As an example, glottal efficiency (Schutte, 1981) has been defined as the ratio of radiated oral acoustic output power to aerodynamic input power, the latter being mean subglottal pressure times mean flow during phonation (Bouhuys, Mead, Proctor, and Stevens, 1968; Schutte, 1981; van den Berg, 1956). As another example, the ratio of AC to DC flow through the glottis during phonation has also been labeled a type of vocal efficiency (Isshiki, 1981; see also Hillman, Holmberg, Perkell, Walsh and Vaughn, 1989).

Clinically, it is also reasonable to consider a different type of output/cost relation in phonation, where cost is considered to be the literal expense to the healthy tissue. Within this framework, the output/cost ratio would be the amount of acoustic output obtained during phonation divided by the amount of trauma inflicted upon the vocal folds. Clearly, quantitative measures of phonotrauma are not available for living subjects, animal or human. However, eventual (or potential) trauma might be estimated using indirect measures of tissue stress. Impact stress is a good candidate to consider. Several lines of evidence suggest that impact stress is an important causal factor in phonotrauma, as for example with nodules, the most common clinically encountered lesions associated with voice

disorders (see for example, FitzHugh, Smith & Chiong, 1958; Nagata, Kurita, Yasumoto, Maeda, Kawasaki and Hirano, 1983). Most importantly, collision stresses, which are directed perpendicular to the tissue fibers, are the most likely source of phonotrauma (Titze, 1994a). These stresses are greatest at the midpoint of the membranous vocal fold, where nodules are known to occur (Jiang & Titze, 1994). The pathologies include obliteration of microvilli and surface desquamation of epithelial cells (Gray, Titze, and Lusk, 1987), a reduction in basal membrane adherence to the lamina propria, and abnormal fibronectin accumulation and appearance of collagen Type IV within the lamina propria (Gray, 1991; Gray, 1997; Gray, Hirano and Sato, 1993; Gray, Pignatari and Harding, 1994). Thus, impact stress should be a good indicator of potential trauma.

These considerations provided the basis for a previous report which proposed the concept of "laryngeal economy" as an approach to output/cost relations in voice production (Verdolini & Titze, 1995). The prediction in this earlier study was that maximum economy would be obtained with barely abducted folds. This prediction was based on acoustic power calculations of an inverse-filtered glottal waveform (Titze, 1994b, Chapter 9), which shows a broad maximum around an open quotient slightly greater than 0.5. While this initial concept of laryngeal economy was interesting, quantitative experimental follow-up is needed.

In this report, we begin by refining the definition of vocal economy, E_v. Next, two experimental studies are described. In the first experiment, empirical E_v curves are generated using an excised canine larynx (no vocal tract). The next experiment consists of a computer simulation study which incorporates the vocal tract shapes for /ɑ/, /i/, and /u/. The results from this second experiment might be more predictive of human results than the no-tract experiment, and could specifically assess the contribution of the vocal tract to E_v. The specific questions are: (1) Do excised larynx data confirm the presence of a maxima in the glottal vocal economy function, and if so, for which specific glottal widths? (2) Does the addition of a vocal tract influence the E_v maximum?

Definition of Vocal Economy, E_v

As previously mentioned, vocal economy was developed as a clinically-motivated output/cost ratio. With this in mind, vocal economy is defined as the ratio of the acoustic output intensity to the impact intensity of the folds (i.e., the rate at which energy per unit area is absorbed by laryngeal tissues due to collision of the left and right folds). In a previous study (Verdolini and Titze, 1995), laryngeal economy (previously "laryngeal efficiency") was defined as the ratio of radiated acoustic output power to impact stress. While the ouput/cost idea is similar in either case, most traditional efficiency and economy ratios have been defined as dimensionless quantities. In particular, where deviations are expected to vary across orders of magnitudes, as in this study, it is important to have a dimensionless quantity so that ratios can be reported in a common logarithmic scale such as dB. When so expressed, the economy ratio can be computed as the subtraction of the two intensity curves.

A physical argument will now be given to demonstrate that the impact intensity of the vocal folds is roughly proportional the square of the impact stress. This is important to the present study because impact intensity is not a directly measurable quantity, while impact stress is (Jiang & Titze, 1994; Hess, Verdolini, Bierhals, Mansmann and Gross, in press). First of all, observe that the general definition for the stored energy, E, in a small volume, V, of tissue may be expressed by the following integral:

$$E = \frac{1}{2}\int_V \sigma \cdot \epsilon \, dV \qquad (1)$$

where σ is the tissue stress and ϵ is tissue strain (or deformation). If we let V represent a small volume of tissue near the medial surface of the folds, it is reasonable to assume that the major component of stress during collision is the lateral impact stress. Thus, the integrand of Equation 1 may be expressed as $\sigma_x \epsilon_x$. According to Hooke's Law, ϵ_x may be expressed as a linear combination of the various stress components. However, because we have already assumed that the dominant component of stress is the lateral impact stress, ϵ_x may be taken to be approximately proportional to σ_x. Additional complexity is introduced when viscous losses are taken into account: (1) a time delay exists between stress and strain, and (2) a knowledge of the time-history is necessary in order to relate stress and strain. However, the integrand of Equation 1 remains proportional to the square of the impact stress (σ_x^2). Similarly, by assuming σ_x to vary sinusoidally with time, it may be shown that the average impact intensity, I_C, of the folds is roughly proportional to the square of the magnitude of the impact stress (σ_C):

$$I_C \propto \sigma_C^2 \qquad (2)$$

Expressed as the ratio of radiated acoustic intensity to the impact intensity of the folds, vocal economy, E_v, may be written mathematically as:

$$E_V = 10\log_{10}\left(\frac{I_A}{I_C}\right) = 10\log_{10}(I_A) - 10\log_{10}(I_C) \qquad (3)$$

where I_A is the acoustic intensity and I_C is the impact intensity. Both lines of Equation 3 are mathematically equivalent. Thus, the above definition of E_v is computed by subtracting the impact intensity from the acoustic intensity.

To facilitate a practical implementation of the above definition of E_v, we make several minor modifications. Empirically, acoustic intensity is inferred with a sound level meter, which measures the acoustic sound pressure level, P_A, referenced to some nominal value of pressure, P_0 (normasl 20 μPa, which is the value used throughout this study). This can be done because intensity is known to be proportional to the square of the sound pressure level. Similarly, it has just been argued that impact intensity is approximately proportional to the square of the impact stress. Like sound pressure level, impact stress is also referenced to some nominal value of stress, σ_0. Thus, vocal economy becomes:

$$E_V = 10\log_{10}\left(\frac{I_A}{I_{P_0}}\right) - 10\log_{10}\left(\frac{I_C}{I_{\sigma_0}}\right) = 20\log_{10}\left(\frac{P_A}{P_0}\right) - 20\log_{10}\left(\frac{\sigma_C}{\sigma_0}\right) \quad (4)$$

where σ_C is the magnitude of the impact stress. With this definition, E_v can be computed as the acoustic intensity measured on a sound level meter plus a logarithmic function of the impact stress. E_v can thus be viewed as a slight modification of the measured acoustic intensity, with an additive weighting factor to account for potential damage to vocal fold tissues.

Experiment 1: Excised Larynx Study (No Vocal Tract)

This first experiment was not intended to be a definitive study, but rather an exploratory investigation to introduce the concept of vocal economy to the laboratory. As such, the impact stress measurements were made on only one larynx using a stress transducer which was borrowed for the purposes of this study (see acknowledgments).

Methods

Tissue. An excised canine larynx (from a 23 kg male) was provided by a cardiac research laboratory at the University of Iowa. The animal was sacrificed for use in cardiac research, but the larynx was made available to us about an hour post-mortem.

Experimental Set-up and Procedures. The experimental set-up was developed previously (Durham, Scherer, Druker and Titze 1987). In addition, specifics pertinent to this study such as the use of the stress transducer were described in detail in several recent reports (Verdolini, Chan, Titze, Hess, and Bierhals, in press; Hess, Verdolini, Bierhals, Mansmann, and Gross, in press). In brief, the larynx was mounted on a laboratory bench. A three-pronged device, coupled to a micrometer, was used to manipulate the arytenoids and thereby the inter-vocal process width. Another micrometer was used to establish vocal fold length. During phonation trials, predetermined inter-vocal process widths were held constant with wooden shims, the thickness of which had been confirmed with a digital caliper (Mitutoyo Digimatic).

Trials were conducted with several inter-vocal process widths, including -1, -0.5, 0, 0.55, 1.1, 1.65, 2.2, and 3.3 mm. Negative widths did not imply tissue overlap, but rather a squeezing together of the tissue relative to the 0 mm width. With each width condition, discrete subglottic pressures of 1.0, 1.2, 1.4, and 1.6 kPa were delivered to induce phonation. Vocal fold length was manipulated so that fundamental frequency (f_0) would be approximately constant across all trials. Although the target f_0 was 125 Hz, the approximate norm for speech for adult males (see for example, Hollien and Shipp, 1972), the actual f_0 for most trials was 150 ± 5 Hz (range was 125-155 Hz). Vocal fold lengths (anterior commissure to tip of vocal process) used to accomplish the target frequency varied between 13.04 and 15.06 mm.

Output intensity measures were obtained in dB using a Bruel and Kjaer 2230 Sound Level meter at a constant distance of 15 cm and 45 degrees azimuth. The C-scale weighting was used to eliminate low-frequency room noise and artifacts below about 50 Hz. Fundamental frequency was measured using a Realistic 333-1063 condenser microphone with a dual beam oscilloscope.

Data Processing and Analysis. To control for signal drift, impact stress signals were first high-pass filtered at 1 kHz with a Kemo Dual Variable Filter Type VBF/4 (DC). Signals were then digitized at 10 kHz with a 12-bit CODAS/WINDAQ A/D conversion board and CODAS software. WINDAQ was used to display signals. For each phonation trial, stable segments of 20 or more cycles were visually identified, which corresponded to stable phonation as auditorily perceived. Peak-to-peak values in the AC stress signal were computed using the software, and an average value was obtained for the segment. For each trial, vocal economy E_v was calculated using the formula presented in Equation 4.

Results

Glottal Output Intensity. The acoustic intensity data were found to be unusually noisy. This may have been due to reflections, perturbations or other anomalies introduced by the presence of the pressure transducer, which was mounted above the larynx, and left in place during the measurements of acoustic intensity. For this reason, the acoustic intensity data from this experiment were averaged together with data from a similar excised larynx experiment in which such anomalies were not observed (Titze, 1988). This former study focused solely on acoustic output intensity as a function of glottal width, with no measurement of

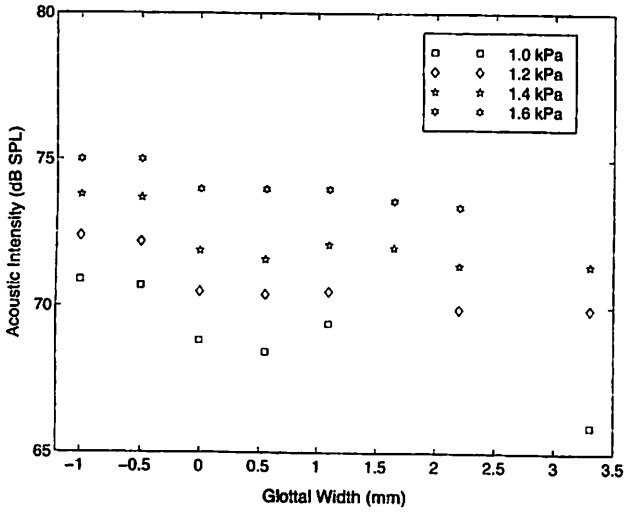


Figure 1. Glottal output power (dB SPL) as a function of glottal width (mm), for subglottic pressures of 1.0, 1.2, 1.4, and 1.6 kPa. Based on two separate excised canine larynx studies (N = 10), using a fundamental frequency of approximately 150 Hz.

impact stress. Figure 1 shows acoustic intensity as a function of glottal width, averaged across the two experiments. Over the range considered (from pressed to 3 mm abducted), there appears to be a small but distinct decrease in acoustic intensity as glottal width is increased.

Vocal Fold Impact Stress. Impact stress curves are shown in Figure 2 for 4 distinct subglottic pressures ranging from 1.0 - 1.6 kPa. Measured data points are shown as rectangular boxes, which are connected by a solid line by assuming a particular functional relationship of the data, as will be described shortly. This figure indicates a sharp monotonic decrease in impact stress, from maximal values of about 40 kPa at the most pressed glottal width (-1 mm) to minimal asymptotic values of about 1 kPa at about 1 mm glottal width and greater.

For several reasons it was deemed important to try to fit a function to these data. First of all, when it is possible to fit the data with a simple low-order function, such data-fitting often yields a smoothed version of the data, removing any high-order noise that may have been present. Second, if there are "unsampled" regions in the data, a function may help in data interpolation and extrapolation.

From inspection, it appeared that the impact stress, σ (in kPa), could be expressed in the following general form:

$$\sigma = A + B y^{C} \exp(-Dx) \qquad (5)$$

where x is the glottal width (expressed in mm), y is the subglottal pressure (expressed in kPa) and A, B, C, and D are parameters to be optimized. The exp(-Dx) was included to capture the general exponential decay of the impact stress as a function of glottal width. The exponential was multiplied

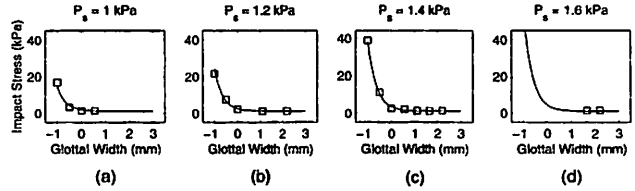


Figure 2. Vocal fold impact stress (kPa) as a function of glottal width (mm), for subglottic pressures of 1.0, 1.2, 1.4, and 1.6 kPa. Based on an excised canine larynx study (N = 1), using a fundamental frequency of approximately 150 Hz. Square boxes indicate actual data points. Solid lines represent a least-square fit of the data to the exponential function given in Equation 5.

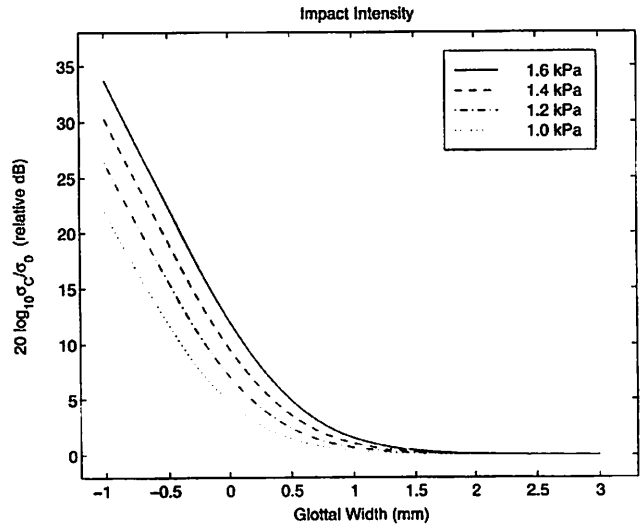


Figure 3. Impact intensity curves (derived from Figure 2 using Equation 4), which play a major role in the definition of vocal economy, E_V. E_V is obtained by subtracting the impact intensity curves from the acoustic intensity curves.

by the subglottal pressure, y, since the maximal values of impact stress for "pressed" conditions were highly dependent on the subglottal pressure. The exponent C was used to capture the most appropriate y-dependence. Finally, the constant A was used because the data did not asymptotically approach zero, as would be the case for a simple exponential function. This function explained the variance of the data remarkably well, yielding a coefficient of determination of 99.0%. The optimized values of the parameters were A = 1.16, B = 0.84, C = 3.03, and D = 2.76.

Figure 3 shows the impact intensity curves plotted on a relative dB scale. The reference value utilized for σ_0 was the asymptotic value of stress (1.16 kPa) obtained from the curves in Figure 2. Utilizing Equation 4, the curves from Figure 3 can be subtracted from the acoustic output curves (such as those found in Figure 1), to yield vocal economy, E_V.

Vocal Economy. Figure 4 shows the vocal economy results for the excised larynx experiment. Because of the tight cluster of the data across subglottic pressure conditions, just one curve was generated to approximate the data.

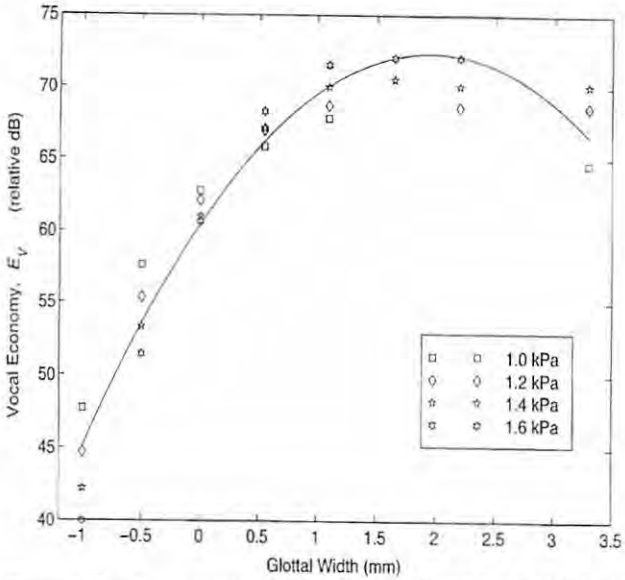


Figure 4. Vocal economy (relative dB) as a function of glottal width (mm), averaged across subglottic pressures conditions of 1.0, 1.2, 1.4, and 1.6 kPa. Based on excised canine larynx studies, using a fundamental frequency of approximately 150 Hz. A vocal economy maximum occurs at a glottal width of just under 2 mm.

The solid line represents a quadratic best-fit to the data in a least-squares sense. A vocal economy maximum occurs for a glottal width slightly under 2 mm.

Experiment 2: Computer Simulation with Vocal Tract

Methods and Results

Oral Output Intensity. In the second experiment, a simulation was conducted to examine the possible contribution of the vocal tract to E_v. Although direct measures of intra-fold impact stress recently have been obtained for awake humans (Hess, Verdolini, Bierhals, Mansmann & Gross, in press; Verdolini, Hess, Titze, Bierhals and Gross, in press), videoscopic guidance of the intraglottal sensor is required using rigid endoscopy. This precludes the possibility of non-invasively monitoring subglottal pressures, which must be accounted for in the calculation of E_v.

In this second study, subglottal, glottal, and supraglottal components were included in the model, and the resultant acoustic power was computed with varying glottal widths. A kinematic model of the vocal folds was used (Titze, 1984), in which the superior glottal half-width g at any point along the length of the folds was defined as:

$$g(y,t) = \xi_0 \left(1 - \frac{y}{L}\right) + \xi_m \sin\omega t \sin\frac{\pi y}{L} \qquad (6)$$

where ξ_0 is the posterior glottal half-width at the vocal processes, ξ_m is the amplitude of vibration, ω is the angular frequency of oscillation, t represents time, and y is an arbitrary position along the anterior-posterior length of the vocal folds ($y = 0$ corresponds to the vocal process position,

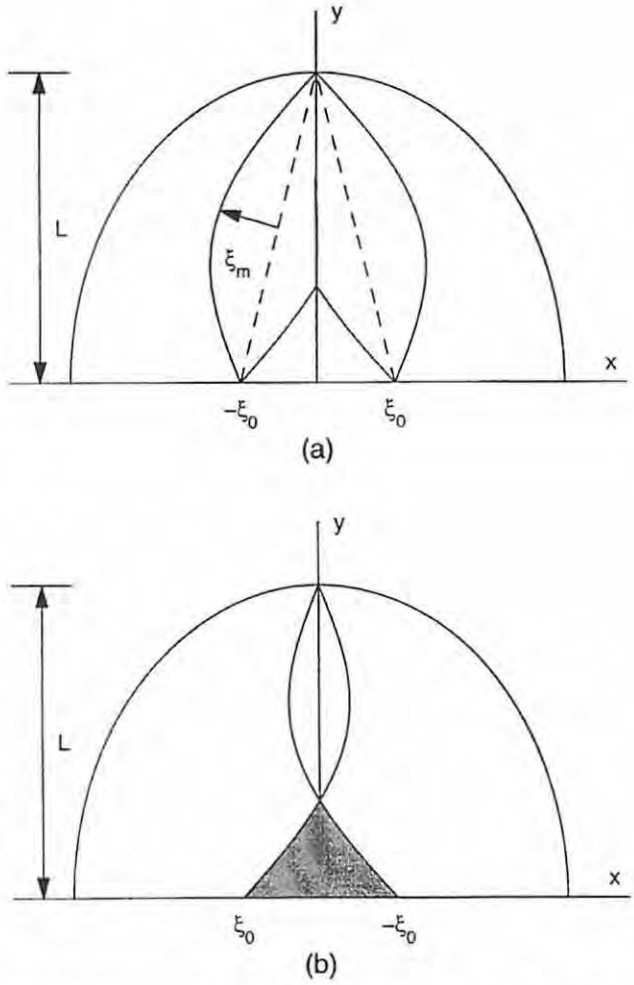


Figure 5. (a) A superior view of the vocal fold model with positive glottal half-width ξ_0. (b) A superior view of the vocal fold model with negative glottal half-width ξ_0.

and $y = L$ corresponds to the anterior commissure). In a vertical sense, variations in the glottal width were based on a traveling wave (Titze, 1984). A superior view of a vocal fold configuration with a positive ξ_0 is shown in Figure 5a, and with a negative ξ_0 in Figure 5b. As described by Titze (1984), this model incorporates a first-order nonlinear interaction between the source and the vocal tract. Acoustic wave propagation in the vocal tract and trachea is modelled using the transmission line approach (Kelly & Lochbaum, 1962; Liljencrants, 1985; Story, 1995), and the lungs are modelled as a constant pressure source. The model also implements "yielding wall" losses of the tract, as described by Story (1995).

Using this model, oral acoustic intensity was obtained as a function of glottal width for three different vocal tract configurations: /ɑ/, /i/, and /u/, as computed in the MRI (magnetic resonance imaging) investigations of Story (1995). The same subglottal pressures were used as in the preceding experiment (1.0, 1.2, 1.4, and 1.6 kPa) corresponding to amplitudes of vibration of 1.3, 1.5, 1.6, and 1.7 mm, respec-

tively. The length of the vocal folds was 1.24 cm, and the fundamental frequency was 150 Hz, similar to the frequencies in the excised larynx experiment. Oral acoustic intensity was computed relative to sound pressure level (20 μPa), as is customarily done with sound level meters.

Vocal Economy. The same data were used for vocal fold impact stress as obtained in the preceding excised larynx experiment. A more complex 3D self-oscillating model would be required in order to compute impact stress directly from the model. Also, economy was computed using the same manipulations as described previously.

Results. Figure 6 shows the oral acoustic intensity for the vowels /ɑ/, /i/, and /u/ at four different subglottal pressures. For all the vowels, the acoustic intensity is maximized for just barely adducted folds (i.e., a glottal width of 0 mm). For /ɑ/ and /i/, acoustic intensity decreases at a fairly constant slope for increasing glottal widths. This can be explained by the fact that glottal closure is not as abrupt for increasing glottal widths. Thus, the high-frequency energy would diminish for increasing glottal width, as would the overall radiated acoustic intensity because high-frequency components radiate relatively well and thus

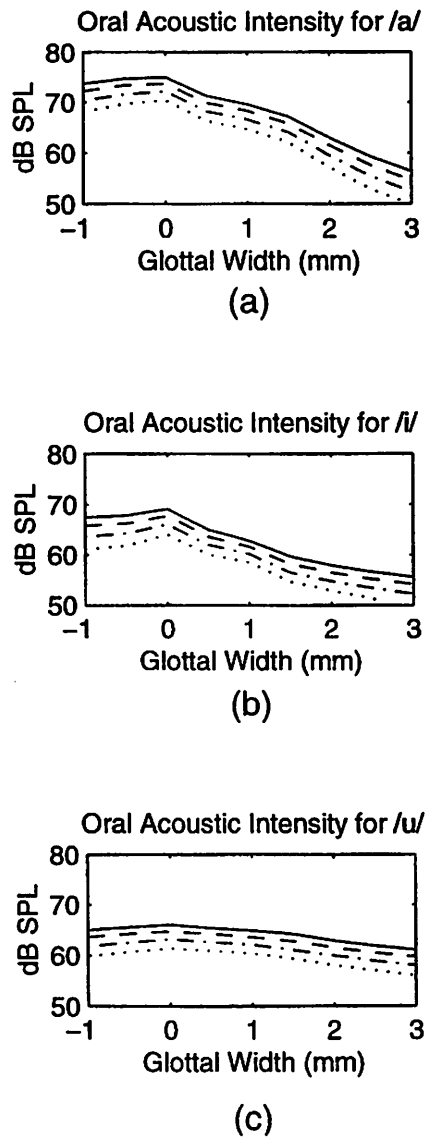


Figure 6. Oral acoustic intensity for (a) /ɑ/, (b) /i/, and (c) /u/, in dB SPL, as a function of glottal width (mm), for subglottic pressures of 1.0 (dotted lines), 1.2 (dashed-dotted lines), 1.4 (dashed lines), and 1.6 kPa (solid lines). Based on computer simulation with vocal tract, using a fundamental frequency of 150 Hz.

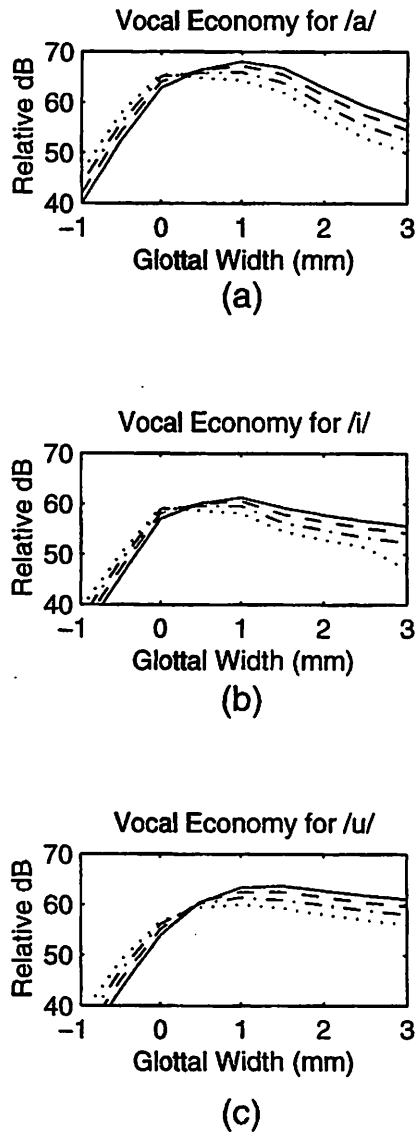


Figure 7. Vocal economy for (a) /ɑ/, (b) /i/, and (c) /u/, in relative dB, as a function of glottal width (mm), for subglottic pressures of 1.0 (dotted lines), 1.2 (dashed-dotted lines), 1.4 (dashed lines), and 1.6 kPa (solid lines). Based on computer simulation study with vocal tract, using a fundamental frequency of 150 Hz. Vocal economy maxima occur around 1 mm.

account for a substantial portion of the radiated acoustic intensity. This effect is not as pronounced for /u/ because the /u/ spectrum has less high-frequency energy initially, and thus less high-frequency energy to lose as glottal closure becomes less abrupt.

Vocal economy results across the three vowel conditions are shown in Figure 7. For both /ɑ/ and /i/, the vocal economy maximum occurred at a glottal width of about 1 mm. The vocal economy maximum for /u/ also occurred at approximately 1 mm (or slightly greater), but this maximum was not as well defined.

Discussion

The definition of vocal economy was refined, introducing it as a dimensionless output/cost ratio, i.e., acoustic intensity divided by impact intensity. With this refinement, two quantitative studies of vocal economy were performed in the laboratory. The first experimental question was whether the previously described maxima in the glottal vocal economy function (Verdolini and Titze, 1995) could be verified empirically using an excised larynx experiment. Indeed, the present experimental conditions of 150 Hz phonation trials and subglottic pressures between 1.0 and 1.6 kPa yielded clear evidence of vocal economy maxima at a glottal width of approximately 2 mm. In this experiment, the vocal fold impact intensity curves were critical in producing maxima in vocal economy. As demonstrated in earlier studies (Titze, 1988), the output intensity curves alone do not produce maxima as a function of glottal width in excised larynxes.

Although somewhat tangential, one interesting aspect of the empirical data regards the vocal fold impact stress curve itself. The curves obtained in this study show the greatest decline in impact stress (as a function of glottal width) for glottal widths less than 0 mm. These data are in general agreement with previous reports showing a monotonically decreasing function of impact stress with respect to glottal width (Jiang and Titze, 1994). However, in the earlier report, the sharp decline in impact stress seemed to occur over a substantially larger region, extending to glottal widths up to 3 mm. However, because the earlier study measured impact stress at only three glottal widths, it was not possible to make a full comparison of the two studies nor determine the form of the curve that would best fit the earlier impact stress data. Another difference between the two studies was that the present study was performed on a full larynx, whereas the former study was performed on a hemilarynx.

The second experimental question regarded the influence of the vocal tract on vocal economy. Using a computational approach, evidence suggests that the vocal tract sharpens the maxima. In contrast to the excised larynx studies, the computer simulations with a vocal tract yielded maxima in the acoustic output curves themselves for just barely adducted folds. Accounting for the influence the impact stress data, the economy maxima were shifted to a glottal width of approximately 1 mm, a configuration which yielded a comparable acoustic output, but with less trauma to the vocal folds due to impact stress.

Overall, the data provide support for the pursuit of vocal economy as output/cost ratio in voice production. Both computer simulations and excised larynx experiments suggest that there is some optimal glottal width that simultaneously maximizes acoustic output, while minimizing the impact intensity of the vocal folds. If confirmed for human subjects, vocal economy would have high clinical significance. Indeed, it would be relevant not only for patients with phonatory trauma from hyperadduction, but also for patients with hypoadducted conditions such as paralysis. The barely abducted laryngeal configuration could be seen as a general target relevant for both sets of patients, as suggested in previous work (Verdolini and Titze, 1995).

Although strong generalizations to humans are unwarranted at this point, some data in the literature on humans are in general agreement with the present findings. A voice type described as "resonant voice" appears to be produced with glottal configurations similar to those used to generate E_v maxima in the present study: barely adducted or barely abducted folds (Peterson, Verdolini-Marston, Barkmeier and Hoffman, 1994; Verdolini, Druker, Palmer and Samawi, in press). Of note, a similar glottal configuration has been described as the "flow mode" by Gauffin and Sundberg (1989).

As alluded to above, currently there are important limitations that prevent generalization of the present data to human subjects. Although excised larynx canine studies allow for greater experimental controls than possible for human subjects, the excised preparation presents some relevant differences with respect to humans. One of the most important differences is lack of innervation to the larynx. Another limitation to generalization is that, in addition to the subglottic pressure and vowel dependency of acoustic output intensity and E_v data, E_v is also frequency dependent. Some of our preliminary computations did indicate differences in the morphology of output and E_v curves when different fundamental frequencies were used (data not shown). Although the glottal widths corresponding to E_v maxima were not substantially different from those described in the present report, the full range of fundamental frequencies should be explored, and ultimately human data should be attempted.

In summary, empirical data from an excised canine larynx without a vocal tract and from a computer simulation with a vocal tract yielded quantitative evidence of maxima in the vocal economy function. In these data, the maxima were achieved with barely abducted folds (1 - 2 mm glottal widths). The findings provide additional evidence in sup-

port of an optimum glottal width in vocal production, and suggest that vocal economy is a concept that may be pursued with potential clinical relevance.

Acknowledgments

This study was supported by grant numbers P60 DC00976 and 1 K08 DC00139 from the National Institute on Deafness and Other Communication Disorders. Dr. Markus Hess and Mr. Wolfgang Bierhals are thanked for their facilitation of the project, by providing the impact stress transducer. A previous version of this paper was presented at the Twenty-Sixth Annual Symposium: Care of the Professional Voice, June, 1997, Philadelphia, Pennsylvania.

References

Bouhuys, A., Mead, J., Proctor, D., and Stevens, K. (1968). Pressure flow events during singing. Annals of the New York Academy of Science, 155, 165-176.

Colton, R.H., and Casper, J.K. (1990). Understanding voice problems: a physiological perspective for diagnosis and treatment. Baltimore: Williams and Wilkins.

Durham, P.L., Scherer, R.C., Druker, D.G., and Titze, I.R. (1987). Development of excised larynx procedures for studying mechanisms of phonation. Technical Report VABL-1, Voice Acoustics and Biomechanics Laboratory, Department of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa.

FitzHugh, G.S., Smith, D.E., and Chiong, A.T. (1958). Pathology of three hundred clinically benign lesions of the vocal cords. Laryngoscope, 68,855-875.

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

Gray, S. (1991). Basement membrane zone injury in vocal nodules. In J. Gauffin and B. Hammarberg (Eds.), Vocal fold physiology conference, pp. 21-28. San Diego: Singular Publishing Group.

Gray, S.D., Hirano, M., and Sato, K. (1993). Molecular and cellular structure of vocal fold tissue. In I.R. Titze (Ed.),Vocal fold physiology: frontiers in basic science, pp. 1-33. San Diego: Singular Publishing Group.

Gray, S.D., Pignatari, S.N., and Harding, P. (1994). Morphologic ultrastructure of anchoring fibers in normal vocal fold basement membrane zone. Journal of Voice, 8, 48-52.

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

Hess, M., Verdolini, K., Bierhals, W., Mansmann, U., and Gross, M. (in press). Endolaryngeal contact pressures. Journal of Voice.

Hillman, R.E., Holmberg, E.B., Perkell, J.S.,Walsh, M., and Vaughn, C. (1989). Objective assessment of vocal hyperfunction: an experimental framework and initial results. Journal of Speech and Hearing Research, 32, 373-392.

Hollien and Shipp (1972). Speaking fundamental frequency and chronological age in males. Journal of Speech and Hearing Research, 15, 155-159.

Isshiki, N. (1981). Vocal efficiency index. In K.N. Stevens and M. Hirano (Eds.), Vocal fold physiology, pp. 193-207. Tokyo: University of Tokyo Press.

Jiang, J.J., and Titze, I.R. (1994). Measurement of vocal fold intraglottal stress and impact stress. Journal of Voice, 8, 132-144.

Kelly, J. Jr., and Lochbaum, C. (1962). Speech Synthesis, Prac. Fourth Intern. Congr. Acoust., Paper G42, pp. 1-4.

Liljencrants, J. (1985). Dynamic line analogs for speech synthesis. Quarterly Progress and Status Report, STL-QPSR 1/1985. Speech Transmission laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden, p. 1-14.

Morrison, M.D., Nichol, H., and Rammage, L.A. (1986). Diagnostic criteria in functional dysponia. Laryngoscope, 94, 1-8.

Morrison, M.D., Rammage, L.A., Belisle, G., Pullan, B., and Nichol, H. (1983). Muscular tension dysphonia. Journal of Otolaryngology, 12, 302-307.

Nagata, K., Kurita, S., Yasumoto, S., Maeda, T., Kawasaki, H., and Hirano, M. (1983). Vocal fold polyps and nodules: a 10 year review of 1156 patients. Aurus Nasus, Larynx Supple, 10, S27-S35.

Peterson, K.L., Verdolini-Marston, K., Barkmeier, J.M., and Hoffman, H.T. (1994). Annals of Otology, Rhinology, and Otolaryngology, 103, 335-346.

Schutte, H.K. (1981). A clinical method for estimating laryngeal airway resistance during vowel production. Journal of Speech and Hearing Disorders, 46, 138-146.

Sodersten, M., and Lindestad, P.-A. (1987). Vocal fold closure in young adult normal-speaking females. Phoniatric and Logopedic Progress Report, 5, 12-19. Sweden: Karolinska Institute.

Sodersten, M., and Lindestad, P.-A., (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. Journal of Speech and Hearing Research, 33, 601-611.

Sodersten, M., Lindestad, P.-A., and Hammarberg, B. (1991). Vocal fold closure, perceived breathiness, and acoustic characteristics in normal adult speakers. In J.Gauffin and B. Hammarberg (Eds.), Vocal fold physiology: acoustic, perceptual, and physiological aspects of voice mechanisms, pp. 217-224. San Diego: Singular Publishing Group.

Story, B.H. (1995). Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract. Ph.D. dissertation. University of Iowa, Iowa City, IA.

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

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

Titze, I.R. (1994a). Mechanical stress in phonation. Journal of Voice, 8, 99-105.

Titze, I.R. (1994b). Principles of voice production. Englewood Cliffs, New Jersey: Prentice Hall.

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

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

van den Berg, J.W. (1956). Direct and indirect determination of the mean subglottic pressure. Folia Phoniatrica, 8, 1-24.

Verdolini, K., Chan, R.W., Titze, I.R., Hess, M.M., Bierhals, W. (in press). Correspondence of electroglottographic closed quotient to vocal fold impact stress in excised canine larynges, J. Voice.

Verdolini, K., Druker, D.G., Palmer, P.M., and Samawi, H. (in press). Laryngeal adduction in resonant voice. Journal of Voice.

Verdolini, K., Hess, M.M.,Titze, I.R., Bierhals, W., and Gross, M. (in press). Investigation of vocal fold impact stress in human subjects. Journal of Voice.

Verdolini, K., and Titze, I.R. (1995). The application of laboratory formulas to clinical voice management. American Journal of Speech-Language Pathology, 4, 62-69.

Verdolini-Marston, K., Burke, M.K., Lessac, A., Glaze, L., and Caldwell, E. (1995). A preliminary study on the methods of treatment for laryngeal nodules. Journal of Voice, 9, 74-85.

NCVS Status and Progress Report - 12
June 1998, 43-53

Rules for Controlling Low-Dimensional Vocal Fold Models with Muscle Activities

Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Brad Story, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Eric Hunter, M.S.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

A low-dimensional, self-oscillating model of the vocal folds is proposed which captures the two primary modes of vibration (a vertical shear mode and horizontal "compressional" mode) with a rotating plate attached to rigid bar mass. The bar mass is in turn connected to a rigid boundary with stiffness and damping elements. The model has been designed to facilitate its control by physiologically relevant parameters such as normalized muscle activation levels and lung pressure. The control system consists of an empirically derived set of rules which converts the muscle activities of the cricothyroid (CT), thyroarytenoid (TA), lateral cricoarytenoid (LCA), and posterior cricoarytenoid (PCA) muscles into physical quantities of vocal fold length/strain, adduction, convergence, mass, thickness, depth, and stiffness. Results show that the model produces oscillation for combinations of muscle activities similar those shown in previous experimental studies.

Introduction

This paper addresses an ongoing search for a link between laryngeal muscle activation and mechanical properties of simple mass-spring vocal fold models. Although finite element implementation of continuum models of the vocal folds (Alipour and Titze, 1983, 1996), the tongue (Wilhelms-Tricarico, 1995) and the velum (Berry et al, in review) offer the potential for embedding muscle contractions directly into the mathematical formulations, the appeal of simple lumped-element models of tissue remains. This is partly due to the conceptual simplicity of coupled masses and springs, and partly due to the interpretive power that nonlinear dynamics has to offer to low-dimensional models (two or three degrees of freedom).

The problem is that vocal fold parameters such as stiffness, effective mass in vibration, vocal fold length, vocal fold thickness, and rest position are not under direct control by the vocalist. Rather, perceptual dimensions such as loudness, pitch, register, and tightness are likely to govern the activation to laryngeal and respiratory muscles. Thus, to understand the oscillatory characteristics of the vocal folds in a physiologically realistic control space, we need a set of rules that transform perceptual variables to muscle activations, and a set of rules that transform muscle activations to geometrical and viscoelastic parameters of the lumped element models.

A Minimal Parameter Set

Low dimensional models of the vocal folds have been designed to capture two primary modes of tissue vibration, a vertical shear mode and a horizontal "compressional" mode (for identification and nomenclature of the modes, see Titze and Strong, 1975; Berry and Titze, 1996). The "compressional" mode does not truly represent tissue compression, because human tissue is incompressible at audio frequencies, but there is an *apparent* compression in the horizontal direction while there is an *apparent* expansion in the vertical direction, leaving the volume conserved. The one-mass model of Flanagan and Landgraf (1968) captured only the "compressional" mode, the two mass

model of Ishizaka and Flanagan (1972) captured both modes, but not in an explicit manner, whereas the translational-rotational model of Liljencrants (1991) was designed expressly to capture both the shear and "compressional" mode. The three-mass model of Story and Titze (1995) also captured both modes, but with an extra degree of freedom. Thus, if both low-dimensionality and mode separation are of primary concern, the translational-rotational model of Liljencrants is an excellent choice.

As a minimum, the following parameters are needed to describe the natural frequencies of the modes and to relate the point-mass mechanics of the tissue layers to the distributed surface pressures in the glottis:

1. A compressional stiffness (K) for the body of the vocal folds and a rotational shear stiffness (K_c) for the cover to control the natural frequencies of oscillation and (less directly, the fundamental frequency F_o).
2. The length (L), thickness (T), and depth (D) of the vocal folds to define boundary constraints and tissue surfaces.
3. The depth of the cover (D_c) to separate the body tissue layer from the cover tissue layer in terms of effective mass of vibration (not incorporated by Liljencrants, 1991).
4. The glottal half width (ξ_{02}) at the top of the vocal fold to control the adduction of the vocal process.
5. The glottal convergence (ξ_c) from the bottom of the vocal folds to the top of the vocal folds to control vocal register.
6. The mucosal upheaval point along the medial surface of the vocal folds or, correspondingly, the nodal point (z_n) of the shear mode to control upper and lower amplitudes of vibration (not incorporated by Liljencrants, 1991).

With the use of these parameters, we first begin to review the mechanics of vocal fold vibration as applied to a two degrees of freedom body-cover model. Muscle activation will be brought into the picture later.

Low-Dimensional Body-Cover Models

We begin the quantification of self-sustained oscillation with a body-cover model that has only two degrees of freedom. As illustrated in Figure 1 in coronal cross-section, the cover is represeted by a rotating plate and the body is represented by a rigid bar connected to a compressible spring. We call this the *bar-plate* version of the body-cover model. On the left of the figure is a sketch of the distribution of tissue in a continuum sense, which will be called upon later to relate continuum elastic constants to lumped-element parameters. The model has no variation perpendicular to the plane of the paper (the anterior-posterior direction, which we have called the y-direction), which is the direct consequence of chosing a rigid plate and a rigid bar. There is also a rigid connection between the point of rotation (nodal point) of the plate and the bar. A torsional restoring

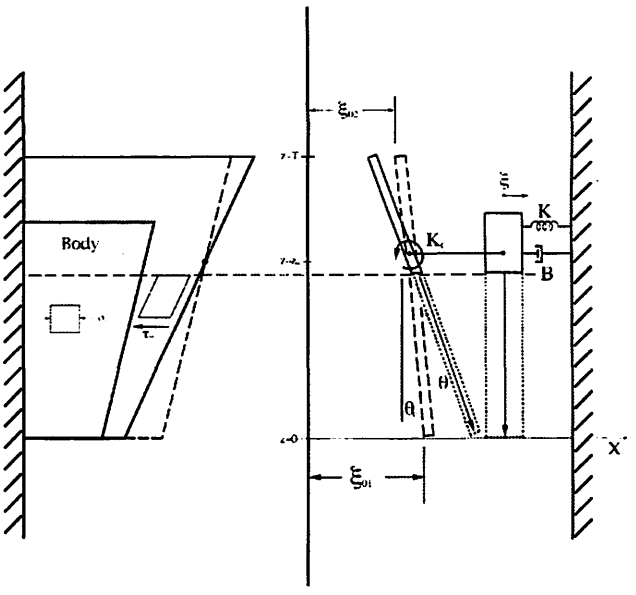


Figure 1. Sketch of bar-plate model (right) and body-cover tissue distribution (left).

force maintains an equilibrium for the plate. The z-direction is the direction of airflow and the x-direction is the direction of lateral tissue displacement.

Unless otherwise stated, we make the following assumptions throughout the body-cover derivations: 1) The two vocal folds move symmetrically with respect to the glottal midplane ($x=0$), 2) There is no vertical displacement of tissue, 3) The body of the vocal fold can extend downward with thyroarytenoid (TA) muscle contraction, thereby pushing the bottom of the cover medially and effectively moving the nodal point upward (see dotted and dashed lines), 4) The restoring properties of the cover are determined by a simple shear (torsion) constant, 5) The glottal area varies linearly from the bottom to top of the vocal folds, 6) Bernoulli flow applies from the lungs to the minimum glottal diameter, at which point jet flow continues and the pressure remains constant in the jet from flow detachment to glottal exit. Liljencrants' (1991) rules or Pelorson's (1994) rules for a slightly enlarged detachment area and a downstream detachment point will eventually be incorporated, but we are still testing the validity of those rules pulsetile flow.

Equations of Motion for Two Degrees of Freedom

Consider first the motion of the vocal fold cover. This motion can be described with one degree of freedom if we assume rotation about a nodal point z_n on the medial surface,

$$I_c\ddot{\theta} + B_c\dot{\theta} + K_c\theta = T_a \quad , \qquad (1)$$

where θ is the rotation angle, T_a is the applied aerodynamic torque, I_c is the moment of inertia for rotation of the cover,

B_c is the rotational damping, and K_c is the rotational stiffness. Since only two degrees of freedom for the entire vocal fold are allowed, the cover is assumed to move rigidly with the body mass at the nodal point z_n. (A third degree of freedom could be added by replacing the rigid rod connection at z_n with another spring, but there is no compelling reason to do so.)

The equation of motion for the body is written as

$$M\ddot{\xi} + B\dot{\xi} + K\xi = F_a \quad , \tag{2}$$

where ξ is the body displacement (also the nodal point displacement in the cover), M is the body mass, B is the body damping, K is the body stiffness, and F_a is the aerodynamic driving force on the body.

Equations 1 and 2 are coupled by the fact that both the driving torque T_a on the cover and the driving force F_a on the body are dependent on glottal flow. The glottal flow in turn depends on the glottal entry and exit areas, defined geometrically as

$$a_1 = 2L[\xi_{02} + (T - z_n)\tan\theta_o - (0 - z_n)\tan(\theta_o + \theta) + \xi] \quad , \tag{3}$$

$$a_2 = 2L[\xi_{02} + (T - z_n)\tan\theta_o - (T - z_n)\tan(\theta_o + \theta) + \xi] \tag{4}$$

where L is the length of the glottis, T is the thickness of the vocal folds, ξ_{02} is the upper pre-phonatory displacements, θ_o is the pre-phonatory convergence angle, and z_n is the nodal point as measured from the bottom of the vocal folds. Note that both θ and ξ enter these equations for the glottal entry and exit areas; hence coupling of the equations of motion is guaranteed.

Relating Lumped Constants to Elastic Moduli

For a negligible amount of fiber tension in the vocal folds, it is possible to express the rotational stiffness of the cover (K_c) and the compressional stiffness of the body (K) in terms of one measurable isotropic shear modulus μ (Chan & Titze, 1997). This can be done by referring back to continuum mechanics. The constitutive equation for transversally isotropic tissue with planar strain (Fung, 1993) can be written as

$$\sigma_x = \frac{2\mu}{(1-\nu)}\left(\frac{\partial \xi}{\partial x} + \nu\frac{\partial \zeta}{\partial z}\right) \quad , \tag{5}$$

and

$$\tau_{xz} = \mu\left(\frac{\partial \xi}{\partial z} + \frac{\partial \zeta}{\partial x}\right) \quad , \tag{6}$$

where σ_x is the normal stress in the x direction, τ_{xz} is the shear stress in the x direction, μ is the shear modulus, υ is the Poisson ratio, and ζ is the vertical displacement. Since we have already assumed ζ=0 to reduce the number of degrees of freedom of the model, the tissue must now be treated as completely compressible (υ=0). If we further assume that all the compression takes place linearly over a depth D in the body, then

$$\sigma_x = 2\mu\frac{\partial \xi}{\partial x} = 2\mu\,\xi/D \qquad (in\ body) \tag{7}$$

$$\tau_{xz} = \mu\frac{\partial \xi}{\partial z} = \mu\theta \quad . \qquad (in\ cover) \tag{8}$$

where θ is the torsional angle defined earlier. None of the above assumptions are defensible, of course, in strict continuum mechanical terms, but the lumped-element approximations are useful for later comparisons to continuum models.

The normal stress σ_x in Equation 7 is assumed to be constant throughout the body (see tissue element on left side of Figure 1). Hence σ_x can be converted to a net restoring force by multiplying the cross sectional area of the tissue LT by this stress

$$F_x = LT\sigma_x = K\xi \quad . \tag{9}$$

When Equations 7 and 8 are combined, this now yields the relation between the shear modulus of the continuum and the spring constant K,

$$K = 2\mu LT/D \quad . \tag{10}$$

Likewise, for a depth D_c in the cover, the shear stress τ_{xz} is multiplied by a shear area LD_c to get a shear (torsional) restoring force $LD_c\,\mu\theta$, which is assumed constant throughout the cover. This shear force is then multiplied by the upper moment arm $(T\text{-}z_n)/2$ and the lower moment arm $z_n/2$ to get a restoring torque,

$$T_0 = (LD_c\mu\theta)\left[\frac{T - z_n}{2} + \frac{z_n}{2}\right] \tag{11}$$

This torque is divided by the angular strain θ to get the torsional stiffness:

$$K_c = \frac{1}{2}\mu LTD_c \tag{12}$$

Note that both K and K_c are directly proportional to μ, the shear modulus in the tissue, but these spring constants also contain the geometrical factors length, thickness, and depth of the tissue layers. In particular, both stiffnesses increase linearly with vocal fold length and thickness, but rotational stiffness increases with cover depth whereas compressional stiffness decreases with body depth.

Consider now the moment of inertia I_c of the cover. It can be derived from basic principles as well,

$$I_c = \int_{z_o}^{T} z^2 dm + \int_{o}^{z_o} z^2 dm \quad (13)$$

$$= \rho L D_c \int_o^t z^2 dz \quad (14)$$

$$= \frac{1}{3}\rho L D_c T^3 \quad (15)$$

With these derivations, the natural frequency of vibration of the body is

$$F_{ob} = \frac{1}{2\pi}\sqrt{\frac{K}{M}} \quad (16)$$

$$= \frac{1}{\pi D}\sqrt{\frac{\mu}{2\rho}} \quad , \quad (17)$$

and the natural frequency of the cover is

$$F_{oc} = \frac{1}{2\pi}\sqrt{\frac{K_c}{I_c}} \quad (18)$$

$$= \frac{1}{\pi T}\sqrt{\frac{3\mu}{8\rho}} \quad . \quad (19)$$

Note that both natural frequencies are proportional to the square root of the shear modulus to density ratio, which we will define as the mucosal wave velocity,

$$c_m = \sqrt{\frac{\mu}{\rho}} \quad . \quad (20)$$

This is the velocity with which shear modes propagate in unbounded isotropic tissue. The quantities μ and ρ are easily measurable in the laboratory (Chan & Titze, 1998). They are independent of geometry and can therefore become basic building blocks for both distributed and lumped element representations.

Another point of interest is that the ratio of the two natural frequencies,

$$\frac{F_{oc}}{F_{ob}} = \frac{\sqrt{3}}{2}\frac{D}{T} \quad (21)$$

is dependent only on the geometry of the vocal fold. As thickness increases, the natural frequency of the cover decreases relative to that of the body. As depth increases, the natural frequency of the body decreases relative to that of the cover. The two frequencies are identical when T=0.866D.

The foregoing results and the natural frequencies apply only to the case when there is a negligible anterior-posterior tension in the tissue fibers (muscle and ligament). In a later section on muscle activation, K and K_c will be modified to account for the string-like tension in the fibers.

Mean Driving Pressure and Driving Torque

The assumed linear variation of the glottal area from entry to exit allows the pressure in the glottis to be integrated over the medial surface to obtain both the net driving torque on the cover and the net driving force on the body. Using ideal Bernoulli conditions for any point z upstream of the flow detachment point

$$P_s + \frac{1}{2}\rho v_s^2 = P(z) + \frac{1}{2}u^2/a^2(z) \quad , \quad (22)$$

where P_s is the subglottal pressure, v_s is the sublottal (tracheal) particle velocity, $P(z)$ is the intraglottal pressure at any point z, u is the flow, and $a(z)$ is glottal area corresponding to the intraglottal pressure. Now let P_{kd} be defined as the kinetic pressure at the point of flow detachment,

$$P_{kd} = \frac{1}{2}\rho v_d^2 \quad , \quad (23)$$

where v_d is the particle velocity at detachment. In comparison to this kinetic pressure, the tracheal kinetic pressure $\frac{1}{2}\rho v_s^2$ is assumed negligible. The intraglottal pressure is then

$$P(z) = P_s - P_{kd}a_d^2/a^2(z) \quad . \quad (24)$$

For a convergent glottis, for which the flow is fully attached (a_d=a_2), the mean intraglottal pressure over the medial surface is now obtained by integration,

$$P_g = \frac{1}{T}\int_o^T P(z)dz \quad (25)$$

$$= P_s - P_{kd}\left[\frac{a_2^2}{T}(-a^{-1})\left(\frac{da}{dz}\right)^{-1}\right]_o^T \quad . \quad (26)$$

In evaluating this integration, it is necessary to assume that the glottal area gradient

$$\frac{da}{dz} = (a_2 - a_1)/T \quad (27)$$

is independent of z, but this assumption has already been made.

Upon further evaluation of the limits in Equation (26) with the linear area gradient (27), we obtain

$$P_g = P_s - P_{k2}\frac{a_2}{a_1} \quad . \quad (28)$$

This is the mean intraglotal pressure expressed as a function of the entry area a_1 and the exit area a_2.

To obtain the exact aerodynamic torque, we must integrate the differential torque $P(z)zdz$ over the plate surface. This yields an awkward logarithmic expression that has singularities when the glottal areas go to zero. An approximate torque expression can be obtained by comput-

ing two average pressures, a lower average pressure P_l that rotates the cover counterclockwise and an upper average pressure P_u that rotates the cover clockwise. Following the same integration steps as in Equations 26-28, the average pressures are

$$P_l = \frac{1}{z_n}\int_o^{z_n} P(z)dz = P_s - P_{k2}\frac{a_2^2}{a_n a_1} \quad (29)$$

$$P_u = \frac{1}{T-z_n}\int_{z_n}^{T} P(z)dz = P_s - P_{k2}\frac{a_2}{a_n} \quad , \quad (30)$$

were $a_n=a(z_n)$ is the glottal area at the nodal point.

With these upper and lower average pressures, and with a lower moment arm $z_n/2$ and an upper moment arm $(T-z_n)/2$, the simplified expression for the aerodynamic torque on the cover is

$$T_a = Lz_n P_l\left(\frac{z_n}{2}\right) - L(T-z_n)P_u\left(\frac{T-z_n}{2}\right) \quad (31)$$

Transglottal Pressure and Coupling to the Vocal Tract

If we make the assumption that there is no pressure recovery at glottal exit due to the emergence of a jet, then the exit pressure P_2 equals the input pressure to the epilarynx tube,

$$P_2 = P_e \quad . \quad (32)$$

Using once again Bernoulli's equation up to the point of detachment,

$$P_s = P_2 + \frac{1}{2}\rho v_2^2 = P_e + P_{k2} \quad , \quad (33)$$

the kinetic pressure P_{k2} in Equations 28-30 can be replaced by the transglottal pressure $(P_s - P_e)$. This leads to important relationships between the average driving pressures on the vocal folds for a convergent glottis:

$$P_g = P_s - (P_s - P_e)\frac{a_2}{a_1} \quad . \quad (34)$$

$$P_l = P_s - (P_s - P_e)\frac{a_2^2}{a_n a_1} \quad (35)$$

$$P_u = P_s - (P_s - P_e)\frac{a_2}{a_n} \quad (36)$$

For a divergent glottis, we assume that the jet separates at glottal entry and the glottal pressure is a constant throughout

$$P_l = P_g = P_u = P_e \quad (37)$$

When there is contact between portions of the vocal fold surfaces, we also need the hydrostatic pressure,

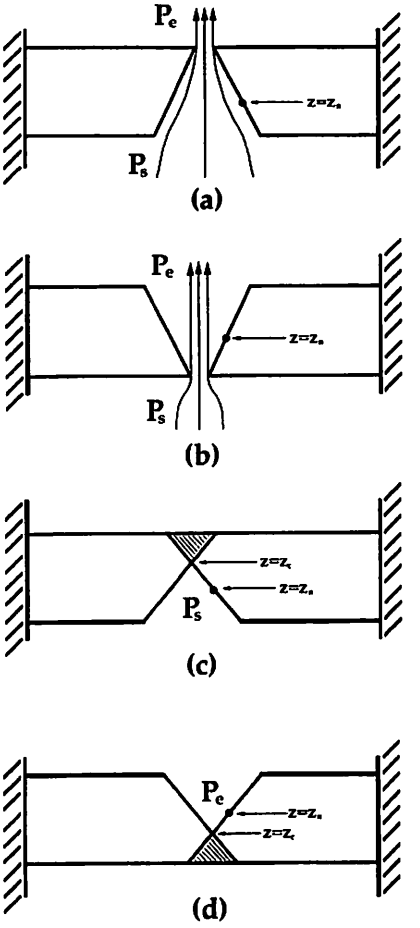


Figure 2. Four conditions for computation of driving forces and torques, (a) convergent and open, (b) divergent and open, (c) convergent and contact at top, (d) convergent and contact at bottom.

defined here as the mean between the subglottal pressure and the supraglottal (epilaryngeal) pressure,

$$P_h = (P_s + P_e)/2 \quad (38)$$

We now consider four separate conditions for the driving force and torque as determined by the glottal configuration. The conditions correspond to the sketches shown in Figure 2.

(a) For convergence and glottal opening (Figure 2):

$$a_1 > a_2 > 0 \quad (39)$$

$$F_a = LTP_g \quad (40)$$

$$T_a = \frac{1}{2}Lz_n^2 P_l - \frac{1}{2}L(T-z_n)^2 P_u \quad (41)$$

(b) For divergence, or a rectangular shape, and glottal opening (Figure 2):

$$a_2 \geq a_1 > 0 \tag{42}$$

$$F_a = LTP_e \tag{43}$$

$$T_a = \frac{1}{2}LTP_e(2z_n - T) \tag{44}$$

(c) For convergence and contact at the top between $z=z_c$ and $z=T$ (Figure 2):

$$a_1 \geq a_2 \quad and \quad a_2 \leq 0 \tag{45}$$

$$F_a = LP_s z_c + LP_h(T - z_c) \tag{46}$$

$$T_a = \frac{1}{2}LP_s z_c(2z_n - z_c) - \frac{1}{2}LP_h(T - z_n)(T + z_c - 2z_n) \tag{47}$$

(d) For divergence and contact at the bottom between $z=o$ and $z=z_c$ (Figure 2):

$$a_1 < a_2 \quad and \quad a_1 \leq 0 \tag{48}$$

$$F_a = LP_h z_c + LP_e(T - z_c) \tag{49}$$

$$T_a = \frac{1}{2}LP_h z_c(2z_n - z_c) - \frac{1}{2}LP_e(T - z_c)(T + z_c - 2z_n) \tag{50}$$

The driving torques for the partial contact cases (*c* and *d* above) were calculated by evaluating the average forces (aerodynamic and hydrodynamic) with their respective moment arms for the nodal point z_n.

The contact point z_c needs to be known for the above force and torque calculations. This point is determined by setting the medial surface displacement to zero.

$$\xi_{02} + (T - z_n)\tan\theta_o - (z_c - z_n)\tan(\theta_o + \theta) + \xi = 0 \quad , \tag{51}$$

which gives the result

$$z_c = Min\left\{T, Max\left[0, z_n + \frac{\xi_{02} + \xi + (T - z_n)\tan\theta_o}{\tan(\theta_o + \theta)}\right]\right\}. \tag{52}$$

With this contact point, which is limited to be no less than 0 and no greater than *T*, the vocal fold contact area is easily computed as

$$a_c = L(T - z_c) \qquad a_1 > 0 \qquad a_2 \leq 0 \tag{53}$$

$$= Lz_c \qquad a_1 \leq 0 \qquad a_2 > 0\,, \tag{54}$$

and the glottal area is

$$a_g = Max[0, Min(a_1, a_2)] \quad . \tag{55}$$

The corresponding glottal flow calculation was developed in Titze (1994) and is repeated here for completeness:

$$u = a_g c\{-(a_g/A^*) \pm [(a_g/A^*)^2 + 4(P_s^+ - P_e^-)/(\rho c^2)]^{\frac{1}{2}}\} \quad , \tag{56}$$

where *c* is the sound velocity in air, P_s^+ is the forward wave below the glottis, P_e^- is the backward wave above the glottis, ρ is the air density, and

$$A^* = A_s A_e/(A_s + A_e) \tag{57}$$

is the effective vocal tract area.

Development of Rules for Muscle Control

In this section we will develop rules for controlling the geometric and elastic parameters of the body-cover model by muscle activation. These rules are not in their final stage of development, but are quite functional at this point. We begin with an elongation rule.

Elongation Rule

Experimentation with excised larynges and *in vivo* animal preparation (Titze, Jiang and Druker, 1987; Titze, Jiang, and Lin, 1997) has shown that vocal fold elongation can be written as

$$\epsilon = G(Ra_{CT} - a_{TA}) - Ha_{LC} \tag{58}$$

where ϵ is the longitudinal vocal fold strain (elongation divided by the cadaveric rest length), a_{CT} is the normalized cricothyroid muscle activity (ranging from 0.0 to 1.0), a_{TA} is the normalized thyroarytenoid muscle activity (same range), and a_{LC} is the normalized lateral cricoarytenoid muscle activity (same range). The constants in the equation are the gain of elongation *G*, the torque ratio *R*, and the adductory strain factor *H*. This rule has been modified from a previous rule (Titze, Jiang and Druker, 1987) to include the adductory strain of Ha_{LC} that occurs for pre-phonatory posturing. This inclusion left the nature of the equation the same, but changed the coefficients slightly. For this study, we are letting *G*=0.2, *R*=3.0 and *H*=0.2. The previous rule was for canines, whereas the current version is intended for humans, for which the range of ϵ must be higher to achieve a larger pitche range. Thus, *G* was increased from 0.1 to 0.2 for humans. With this gain, the maximum elongation is 60% when a_{CT}=1.0 and a_{TA}=a_{LC}=0. This corresponds to the maximumal superior nerve stimulation condition in the Titze, Jiang, and Lin (1996) study, for which the canine vocal folds elongated 45%. The maximum shortening occurs for a_{CT}=0 and a_{TA}=a_{LC}=1.0. For this case, the rule gives -40% and the measured canine value was -17% under maximal recurrent nerve stimulation. With full contraction

of all muscles, the rule gives 20% elongation and the measurement on dogs was 26%.

Phonation cannot be realized over the full range of elongation. For a_{LC}=0.5, which is probably typical for soft adduction, the range of ϵ is -30% to 50%, or 0.7 to 1.5 of the resting length. This approximate 2:1 range was observed by Nishizawa et al (1988).

With the above rule, the vocal fold length can be written as

$$L = L_o(1 + \epsilon) \quad , \tag{59}$$

where L_o is the cadaveric rest length (1.5 cm in males and 1.0 cm in females).

Thickness and Depth Rules

Vocal fold thickness and depth increase with vocal fold shortening. In a purely passive sense, the Poisson ratio determines this increase. For an incompressible medium, the ratio is 0.5. But there is also an active control of thickness and depth. We propose the following rules:

$$T = \frac{T_o(1 - 0.8\,a_{CT})}{1 + 0.5\epsilon} \tag{60}$$

$$D = \frac{(1 - a_{TA})D_{mo} + D_{lo}}{1 + 0.5\epsilon} \quad , \tag{61}$$

$$D_c = \frac{D_{co}}{1 + 0.5\,\epsilon} \tag{62}$$

where T_o is the resting thickness (0.45 cm for males), D_{mo} is the resting depth of the muscle (0.4 cm for males), D_{lo} is the resting depth of the deep layer of the lamina propria (considered to be part of the body, 0.1 cm in males), and D_{co} is the depth of the cover (0.2 cm in males). The denominator in all cases represents the passive (Poisson ratio) change.

The assumption is that a_{TA} reduces the vibration depth of the vocal fold, a claim made by Saito et al. (1983). For full TA contraction, the muscle depth in vibration reduces to zero, but the deep layer (ligament) remains as D. Another assumption is that CT activation thins the vocal fold, T changing from T_o to 0.2 T_o with a_{CT}. Interestingly, the ratio of the two natural frequencies remains relatively steady with co-contraction of CT and TA,

$$\frac{F_{oc}}{F_{ob}} = \frac{\sqrt{3}}{2}\,\frac{(1 - a_{TA})D_{mo} + D_{lo}}{T_o(1 - 0.8\,a_{CT})} \quad , \tag{63}$$

a point that we will explore further in the Results section.

When the TA contracts, the bottom edge is pushed medially (Figure 1), thus reducing the convergence angle (see convergence rule below). Concurrently, the nodal point z_n for the shear mode effectively moves up on the medial surface, because there is greater vibrational amplitude at the bottom than the top. Empirically, this is related to the point of mucosal upheaval (Yumoto et al., 1993). We propose the rule

$$z_n = T\,a_{TA} \tag{64}$$

With this rule, the nodal point is near the bottom for falsetto register (thin vocal folds and greatest amplitude at the top) and near the center in modal register (thick vocal folds with greater amplitude at the bottom).

Adduction Rule

The following adduction rule for the glottal half-width at the vocal fold processes is adopted:

$$\xi_{02} = 0.38L(1 - 2.0a_{LC} + 0.33a_{PC}) \quad . \tag{65}$$

This rule anchors the "just touching" (ξ_{02}=0) case arbitrarily at a_{LC}=0.5 and a_{PC}=0. For a_{LC}>0.5 and a_{PC}=0 the vocal processes are pressed together (or overlapped). On the other extreme, for a_{LC}=0 and a_{PC}=1.0, the vocal process gap to length ratio (2 ξ_{02}/L) is about 1.0 (a 45° angle at the anterior commissure).

Data by Scherer (1995) suggest that oscillation is obtained over 14% of the range of motion of the vocal process as seen by videolaryngoscopy. The above rule was based, in part, on this observation. When a_{LC}=0.4 and a_{PC} = 0, the rule gives a vocal process gap to length ratio (2 ξ_{02}/L) of 0.15, which is 15% of the range of motion, as reported by Scherer. No oscillation was found beyond this value of a_{LC} with this model. The cadaveric (rest) position yields a vocal process gap to length ratio of 0.76, which is also the value reported by Scherer.

Convergence Rule

Glottal convergence and medial surface bulging are, to a large extent, governed by activation of the thyroarytenoid muscle (Hirano, 1975). The critical issue for oscillation is that a portion of the medial surface of the vocal folds must be nearly vertical to produce a section of the glottis that is rectangular or diverging. This rectangular or divergent section is achieved over the whole vocal fold in modal register but only near the top of the fold in falsetto register. Since the medial surface has curvature in the vertical direction, it requires at least three masses stacked vertically to approximate this quadratic surface and thereby obtain a partial rectangular or divergent glottis. With the single plate model for the cover, however, bulging is not an option, which severely restricts the range of convergence under which the model can operate. We propose the following rule for convergence:

$$\xi_c = T + (0.234 - 0.667\,a_{TA}) \tag{66}$$

With this rule, the range of ξ_c/T is from 0.234 to -0.433.

Stiffness Rules

To include the longitudinal muscle fiber stiffness in the body of the vocal fold, the isotropic "compressional" stiffness $2\mu LT/D$ from Equation 10 is augmented by a "string" stiffness in the fibers,

$$K = 2\mu LT/D + (\pi^2 TD/L)\sigma_b \quad , \qquad (67)$$

where σ_b is the fiber stress in the body (including the deep layer of the lamina propria). The constant $\pi^2 TD/L$ comes from equating the string resonance frequency $(1/2L)\sqrt{\sigma_b/\rho}$ to an equivalent mass-spring resonance frequency $(1/2\pi)\sqrt{K/M}$ and realizing that $M=\rho LTD$.

The fiber stress in the body is made up of the muscle stress and the ligament stress. But stresses do not add across layers; rather, forces add for parallel fibers. Hence, we combine the stresses as forces,

$$\sigma_b = \frac{\sigma_l A_l = \sigma_m A_m}{A_l + A_m} \quad , \qquad (68)$$

where σ_l is the ligament part of the stress in the body, σ_m is the muscle stress in the body, and A_l and A_m are respective cross sectional areas. These areas can further be written as

$$\sigma_b = \frac{\sigma_l A_l = \sigma_m A_m}{A_l + A_m} \quad , \qquad (69)$$

and

$$A_m = \frac{(1 - a_{TA})D_{mo}}{1 + 0.5\epsilon} \quad , \qquad (70)$$

according to the rules established in Equations (60) and (61).

A final consideration is the composition of the muscle stress. This has an active and a passive component

$$\sigma_m = a_{TA}\sigma_{am} + \sigma_p \quad , \qquad (71)$$

where σ_{am} is the maximum active stress in the TA muscle fibers (100-200 kPa). This maximum active stress can be held constant or (more physiologically) made to vary with vocal fold elongation. The exact formula for $\sigma_{am}=\sigma_{am}(\epsilon)$ is not known yet; hence we will assume a constant of 100 kPa in simulations to follow.

The torsional stiffness can likewise be augmented to include a string-dependent term

$$K_c = \frac{1}{2}\mu LTD_c + \frac{1}{3}\pi^2 TD_c T^3 \sigma_c / L \qquad (72)$$

where σ_c is the fiber stress in the cover and the constants are once again obtained by equating the string resonance frequency to an equivalent torsional resonance frequency.

The fiber stress in the cover and ligament is purely passive, which together with the passive component of the muscle has the form

$$\sigma_i = \begin{cases} \frac{-\sigma_{0i}}{\epsilon_{1i}}(\epsilon - c_{1i}) & \text{for } \epsilon_{1i} \le \epsilon \le \epsilon_{2i} \\ \sigma_{2i}(e^{C_i(\epsilon-\epsilon_{2i})} - C_i(\epsilon-\epsilon_{2i}) - 1) + \left(\frac{-\sigma_{0i}}{\epsilon_{1i}}\right)(\epsilon - \epsilon_{1i}) & \text{for } \epsilon > c_{2i} \end{cases} \qquad (73)$$

where i signifies the particular tissue layer (i.e., cover, ligament, muscle). This combination of a linear and exponential function for the passive stress has been found to be the best match to measure vocal fold stress-strain curves (Alipour and Titze, 1991; Min, Alipour and Titze, 1995). In the above equations, σ_o is the stress when $\epsilon=0$, ϵ_1 is the strain where the linear portion goes to zero, and σ_2 and ϵ_2 are the stress and strain where the exponential portion begins. These constants are fitted individually to each stress-strain curve and are given in Table 1.

Table 1.
Constants Used in Equation (73) to Generate Stress-Strain Curves for the Cover Ligament and Muscle Layers

Layer	ϵ_1	ϵ_2	σ_o	σ_2	C_e
cover	-0.5	0.0	1300	1.36e6	1.5
ligament	-0.5	0.0	1300	45000	7.5
muscle	-0.5	0.0	3000	20000	7.4

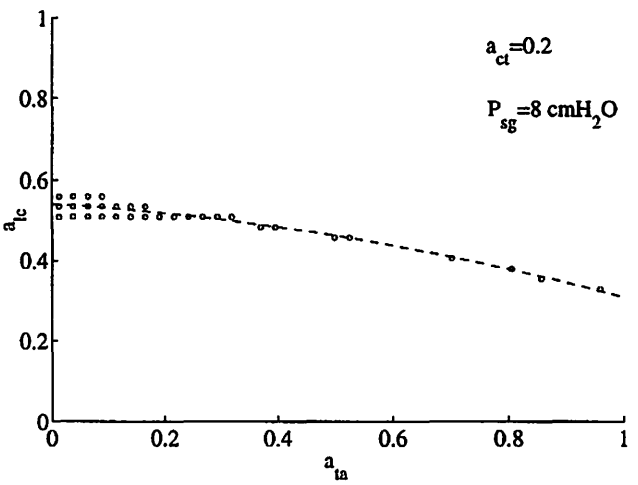


Figure 3. Adductory muscle activation plot (MAP) showing the regions of self-sustained oscillation; the activity of the lateral cricoarytenoid muscle (a_{lc}) is shown on the y-axis while the thyroarytenoid activity (a_{TA}) is shown along the x-axis. The 'o's' indicate a muscle activity pair at which oscillation is sustained. Cricothyroid activity (a_{CT}) and lung pressure were held constant at 0.2 and 8 cm H_2O, respectively.

Results

Figure 3 shows the region of oscillation of the bar-plate model in an adductory muscle activation plot (adductory MAP). The figure was created by performing separate 200 ms simulations for 1600 pairs of TA and LCA activities (40 increments between 0 and 1 for each muscle). Lung pressure and a_{CT} were held constant at 8 cm H_2O and 0.2, respectively. Existence of sustained phonation at each activation pair was determined by applying a zero-crossing detector to the last 100 ms of the demeaned glottal area signal. TA muscle activity is plotted horizontally, LCA activity is plotted vertically and each 'o' symbol denotes an a_{LC}, a_{TA} pair at which oscillation was sustained. Note that phonation is restricted to a small region of this MAP, as was pointed out by Scherer (1995). Furthermore, as TA is increased, the region of phonation becomes disconnected, showing large jumps between activation pairs where phonation was sustained. Oscillation occurs approximately along the dashed curve represented by,

$$a_{LC} = -0.15a_{TA}^2 - 0.08a_{TA} + 0.54 \quad , \quad (74)$$

suggesting that abduction at the vocal process (with release of LCA) must be counterbalanced with adduction at the bottom (with contraction of TA) in order to maintain phonation, and vice versa.

Figure 4 is a zoomed-in version of Figure 3, showing more detail of the leftmost oscillation region of the adductory MAP; the level of a_{CT} was again held constant at 0.2. The gray area encompasses the oscillation region when lung pressure is held constant at 8 cm H_2O while the thick shows how this region is shifted slightly upward when lung pressure is increased to 16 cm H_2O. Thus, at low levels of TA activity higher lung pressures require more adduction to sustain phonation while at higher levels of TA activity (approximately 0.3) the adduction can be reduced. A similar zoomed-in case is shown in Figure 5 except that a_{CT} is held constant at 0.6. The gray area and the thick line again indicate the regions of oscillation when lung pressure is 8 and 16 cm H_2O, respectively. In this case the levels of adduction remain nearly the same for both lung pressures.

Figure 6 (following page) shows a muscle activation plot (MAP) for the elongating and tensing muscles, CT and TA. The iso-lines are the natural frequency of the cover in (a), the natural frequency of the body in (b), and the ratio of the two in (c). Note that F_{oc} changes mainly with a_{CT}, whereas F_{ob} changes mainly with a_{TA}. Thus, to maintain a 1:1 entrainment between the body and cover, the two muscles must co-contract. This is shown in part (c) as the nearly diagonal line labeled "1". The diagonal nature of this MAP was established earlier by measurement of EMG activity on several subjects (Titze, Luschei and Hirano, 1989).

Figure 7 shows the same MAP again, but this time a symbol 'o' is shown at each a_{CT}, a_{TA} pair where oscillation was sustained while adduction was adjusted according to Equation 74. The numbers shown in the figure represent fundamental frequencies (F_o) of selected points. Most of the self-sustained oscillation occurs in the lower left quadrant with F_o's beginning at a low of 170 Hz and rising with either increased TA or CT activity or both. There are also many points of oscillation in the upper left quadrant where the highest F_o's are attained. This region represents high levels of activity in the CT muscle which, by Equation 58, translates into high vocal fold strains and consequently high passive stiffnesses in all vocal fold layers. The lower right quadrant represents a region of high activity in the TA muscle and comparatively low CT activity. This quadrant

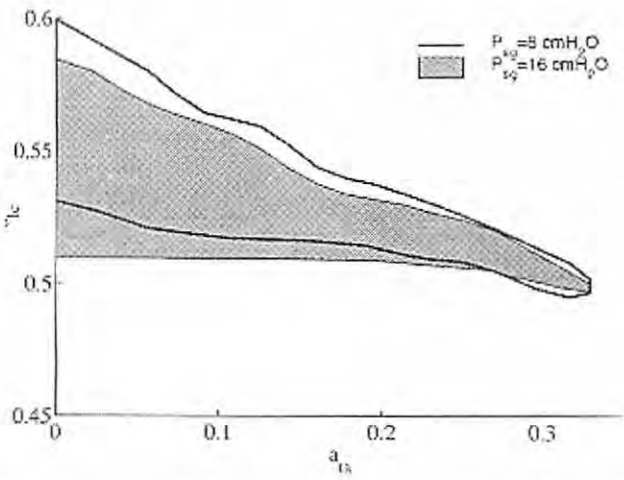


Figure 4. Zoomed-in adductory muscle activation plot (MAP) for low level of cricothyroid activity (a_{CT} = 0.2), (a) lung pressure = 8 cm H_2O, (b) lung pressure = 16 cm H_2O.

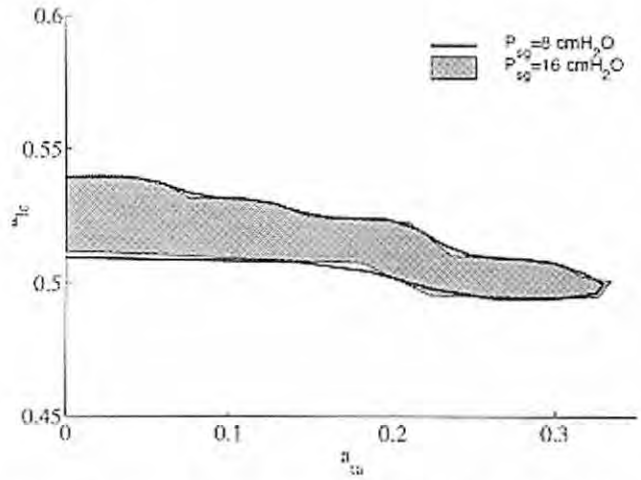


Figure 5. Zoomed-in adductory muscle activation plot (MAP) for higher level of cricothyroid activity (a_{CT} = 0.6), (a) lung pressure = 8 cm H_2O, (b) lung pressure = 16 cm H_2O.

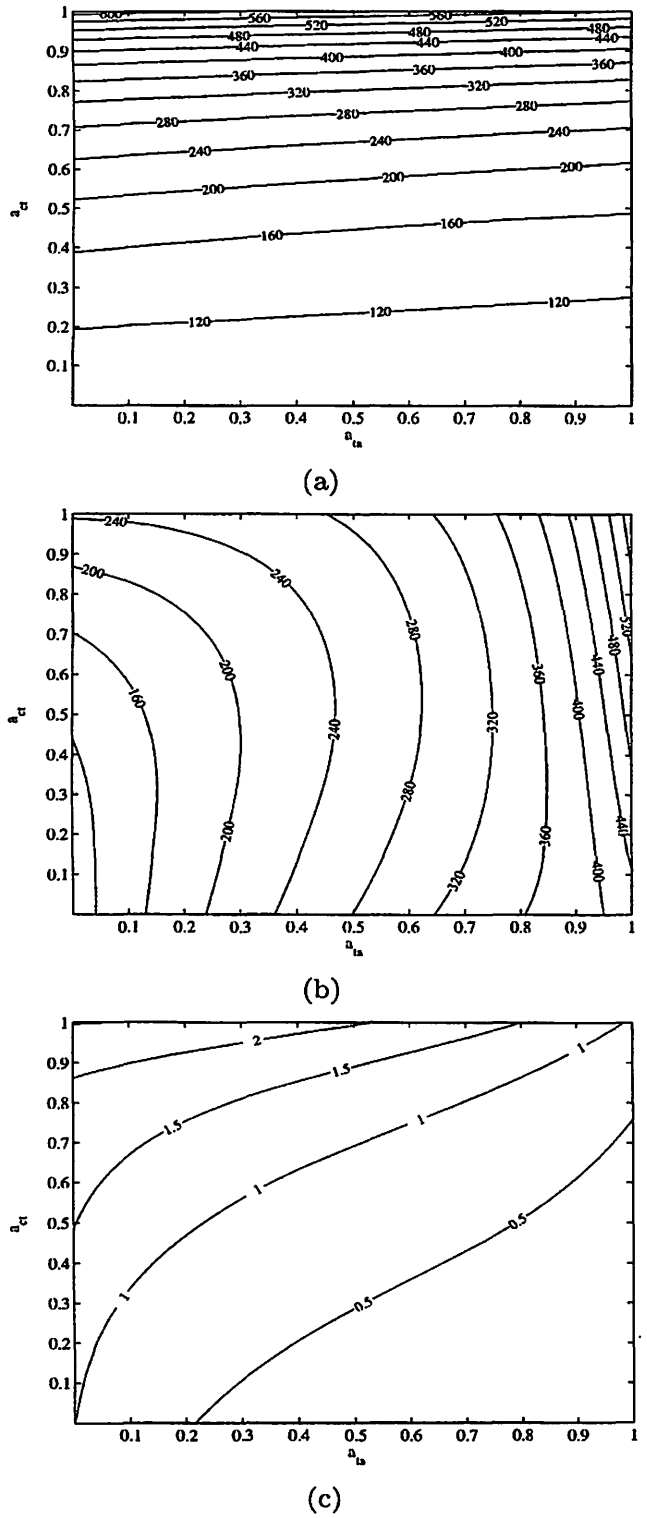


Figure 6. Elongation and tension muscle activation plots showing the effect of cricoarytenoid (a_{CT}) and thyroarytenoid (a_{TA}) activation on (a) natural frequency of the cover (F_{oc}), (b) natural frequency of the body (F_{ob}), and (c) the ratio of F_{oc}/F_{ob}. These plots are based on the rules developed in this paper and Equations 16, 18, and 21.

shows a narrow oscillation section extending throughout the full range of a_{TA} but is restricted to a range of a_{CT} between about 0.18 and 0.30. As the value of a_{TA} is increased the fundamental frequency decreases to a low of 97 Hz. With high TA contraction the active stress in the muscle is high but the vocal fold strain is decreased (see Equation 58) thereby lowering the passive stiffness in all tissue layers.

Conclusions

Low dimensional models of the vocal folds can be made to oscillate in a semi-realistic way, but rules are necessary to capture the co-variation between parameters. In particular, the medial surface of the vocal folds, which is by design a linear function if the vertical degrees of freedom are less than three, must be nearly rectangular for self-sustained oscillation. This requires a rule between the co-contraction of the lateral cricoarytenoid muscle and the thyroarytenoid muscle.

In order to incorporate a fibrous elastic restoring force as well as a nonfibrous (isotropic) restoring force, the spring constants and torsional constants need to have multiple terms that incorporate vocal fold dimensions and muscle activities in different ways. The model has shown that a co-contraction between cricothyroid and thyroarytenoid muscles produces oscillation regions (and fundamental frequencies) that are quite comparable to what has been measured.

This paper has primarily developed a low-dimensional vocal fold model and proposed physiologically-based rules for its control. However, much work remains to be done concerning the control of fundamental frequency and voice quality with these rules. In addition, refinement of both passive and active stress measurements of the vocal fold tissue is needed.

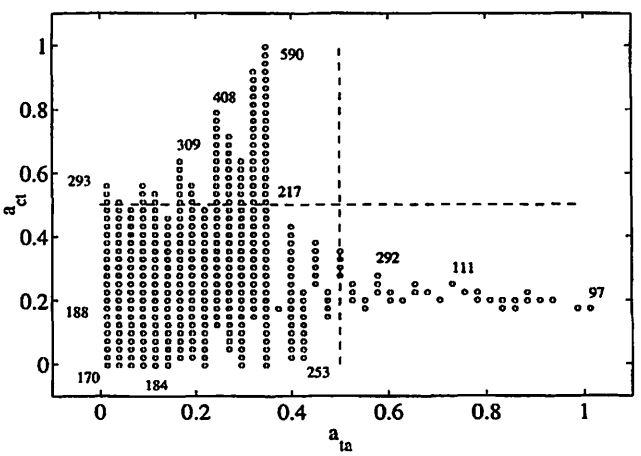


Figure 7. Elongation and tension muscle activation plots showing regions of self-sustained vocal fold oscillation as a function of cricoarytenoid (a_{CT}) and thyroarytenoid (a_{TA}) activations.

Acknowledgment

This work was supported by grant #R01 DC02532 from the National Institute on Deafness and Other Communication Disorders.

References

Alipour-Haghighi, F., Titze, I. R., and Perlman, A. L. (1989). Tetanic contraction in vocal fold muscle. *J. Speech Hear. Res., 32*, 226-231.

Alipour-Haghighi, F., & Titze, I.R. (1991). Elastic models of vocal fold tissues. *J. Acoust. Soc. Amer. 90*(3), 1326-1331.

Alipour, F., and Titze, I. (1983). Simulation of particle trajectories of vocal fold tissue during phonation. In I.Titze & R. Scherer (Eds.), *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control* (pp. 183-190). Denver Center for the Performing Arts.

Alipour, F., and Titze, I. (1996). Combined simulation of two-dimensional airflow and vocal fold vibration. In P. Davis & N. Fletcher (Eds.), *Vocal Fold Physiology: Controlling Complexity and Chaos* (pp. 17-29). San Diego: Singular Publishing Group, Inc.

Berry, D.A., Moon, J.B., Kuehn, D.P. (in review). A histologically-based finite element model of the soft palate. *Cleft Palate-Craniofacial J.*

Berry, D., and Titze, I. (1976). Normal modes in a continuum model of vocal fold tissues. *J. Acoust. Soc. Amer., 105*(5), 3345-3354.

Chan, R., and Titze, I. (1997). Dynamic shear modulus of vocal fold tissues and phonosurgical biomaterials. *J. Acoust. Soc. Amer., 101*, 3179(A).

Flanagan, J.L. and Landgraf, L.L. (1968). Self-oscillating source for vocal-tract synthesis. *IEEE Transactions on Audio and Electroacoustics, AU-16*(1), 57-64.

Fung, Y.C. (1993). *Biomechanics: Mechanical Properties of Living Tissue (2nd Ed.).* New York: Springer Verlag.

Hirano, M. (1975). Phonosurgery: Basic and clinical investigations. *Official report of the 78th annual convention of the Oto-Rhino-Laryngological Society of Japan.*

Ishizaka, K., and Flanagan, J. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *The Bell Systems Technical Journal, 51*(6), 1233-1268.

Liljencrants, J. (1991). A translating and rotating mass model of the vocal folds. *STL Quarterly Progress and Status Report, 1.* Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden.

Min, Y., Titze, I., & Alipour, F. (1995). Stress-strain response of the human vocal ligament. *Annals of Otol. Rhin. Laryng., 104*(7), 563-569.

Pelorson, X., Hirschberg, A., vanHassell, R., Wijnands, A., and Auregan, Y. (1994). Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model. *J. Acoust. Soc. Amer., 96*(6), 3416-3431.

Saito, S., Fukuda, H., Kitahara, S., Isogai, Y., Tsuzuki, T., Muta, H., Takayama, E., Fujioka, T., Kokawa, N., and Makino, K. (1983). Pellet tracking in the vocal fold while phonating-experimental study using canine larynges with muscle activity. In I.Titze & R. Scherer (Eds.), *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control* (pp. 169-182). Denver Center for the Performing Arts.

Scherer, R. (1995). Laryngeal function during phonation. In Rubin, J., Sataloff, R., Korovin, G., and Gould, W., *Diagnosis and Treatment of Voice Disorders.* New York: IGAKU-SHOIN Medical Publishers, Inc.

Story, B., and Titze, I. (1995). Voice simulation with a body cover model of the vocal folds. *J. Acoust. Soc. Amer., 97*(2), 1249-1260.

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

Titze, I., and Strong, W. (1975). Normal modes in vocal cord tissues. *J. Acoust. Soc. Am., 57*(3), 736-744.

Titze, I., Jiang, J., & Druker, D. (1988). Preliminaries to the body-cover theory of pitch control. *J. Voice, 1*(4), 314-319.

Titze, I.R., Jiang, J.J., & Lin, E. (1997). The dynamics of length change in canine vocal folds. *J. Voice 11*(3), 267-276.

Titze, I.R., Luschei, E.S., & Hirano, M. (1989). The role of the thyroarytenoid muscle in regulation of fundamental frequency. *J. Voice, 3*(3), 213-224.

Wilhelms-Tricarico, R. (1995). Physiological modeling of speech production: Methods for modeling soft-tissue articulators. *J. Acoust. Soc. Am. 97*, 3085-3098.

Yumoto et al (1993). Vocal fold vibration viewed from below. *Larynx 5*, 127-134.

NCVS Status and Progress Report - 12
June 1998, 55-63

Phonation Onset: High Speed Glottography and Modelling

Patrick Mergell, M.S.
Department of Phoniatrics and Pedaudiology, University Erlangen-Nuremberg, Germany
Hanspeter Herzel, Ph.D.
Institute for Theoretical Biology, Humboldt University Berlin, Germany
Thomas Wittenberg, M.A.
Monika Tigges, M.D.
Ulrich Eysholdt, M.D.; Ph.D.
Department of Phoniatrics and Pedaudiology, University Erlangen-Nuremberg, Germany

Abstract

Phonation onset has been discussed in the framework of dynamical systems as a Hopf bifurcation, i.e. as a transition from damped to sustained vocal fold oscillations due to changes of parameters defining the underlying laryngeal configuration (adduction, subglottal pressure, muscular activity). An analytic envelope curve of the oscillation onset has been deduced by analyzing the Hopf bifurcation in simplified vocal fold models. It is governed by a single time constant which can be identified with the physiological parameter: *phonation onset time*. This parameter reflects the laryngeal state prior to phonation and can be used as an important, and quantitative classification criterion to assess functional dysphonia in clinical diagnosis. The determination of the phonation onset time from digital high speed videos has been described in detail. A quantitative comparison of high speed measurements with model calculations indicates that relatively long (short) onset times are related to the subjective assessment *soft* (*hard*) used by physicians, voice pathologists and logopedics.

Motivation

The subjective, and auditive classification of voice quality is always influenced by the individual experience and education of the examiner. Therefore, the *standardization* of diagnostic features is mandatory (Titze, 1994b). Many *mechanisms* of vocal phenomena and voice disorders are still not understood. One consequence is the redundancy occurring frequently in classifications of voice states, i.e. some of the used diagnostic parameters depend on each other, e.g. frequency and amplitude of the vocal fold oscillation.Hence, adequate models of the larynx, are required for decoding the information about the intrinsic structure (muscle activities, subglottal pressure, tissue properties, etc.) from measurable quantities (phonation onset time, loudness, fundamental frequency, etc.). Direct measurements probing the intrinsic structure of the larynx are still intricate and even painful for the patients or impossible in vivo. The model parameters are independent and define non-ambiguously the model configuration. If the model simulations reflect the characteristic features of a given vocal phenomenon the underlying mechanism can be reconstructed. The combination of digital high speed glottography and biomechanical modelling is a powerful tool which allows the pipelining of both problems: the introduction of standards and the explanation of mechanisms. In this paper we will point out how theoretical and phenomenologic considerations of the phonation onset together with high speed measurements helped to shed light on the considered dynamic process, and to define a classification parameter which can be used as a clinical standard.

Clinical Background

The examination of the phonation onset is a characteristic feature for the diagnosis of functional dysphonia, i.e. voice disorders with no morphological changes of the vocal folds. A hyperfunction of the laryngeal muscles (pathologically high activity) is often reflected by a hard phonation onset, whereas soft and breathy onsets are typical symptoms of hypofunctional dysphonia, (pathologically low muscle tonus), (Wendler and Seidner, 1987). Even for a trained laryngologist it is very difficult to distinguish

stridently between the hyper-, and hypofunctions of the laryngeal muscle apparatus, since one extreme muscle tonus is usually compensated by the contrary tonus of another muscle. Therefore, a computer-aided and physiologically based assessment, would be a step forward to localize dysfunctions in a complex muscle system. Some conceptional ideas will be presented in this paper.

We will show that the phonation threshold pressure is intimately correlated to the phonation onset time. Verdolini et al. stated in a detailed study that the required pulmonary effort for initiating sustained phonation has to be increased with the rising pitch and with the measure of dehydration of the vocal fold tissue (Verdolini et al., 1990); (Verdolini et al., 1994b). A well-hydrated vocal fold cover is supposed to have a low tissue viscosity. In a physical model of the mucosa it has been shown that the phonation threshold pressure decreases linearly with the tissue viscosity (Titze, 1994a); (Titze et al., 1994) which accords to Verdolini's observations. Furthermore, experiments with canine hemilarynges (Jiang and Titze, 1993) revealed the same functional behavior of the phonation threshold pressure with hydration variations of the vocal fold cover tissue.

In general, the phonation threshold pressure is affected by all kind of changes and interindividual differences in the visco-elastic and aerodynamic system of the larynx. In this context, the voice range profile (VRP) (Stimmfeld, phonetogram) has often been used to analyze voice quality and to get an idea of an adequate voice training (Titze, 1992a). The VRP can be considered as a two-dimensional Hopf bifurcation diagram which separates the fundamental frequency - sound pressure level - plane into two regions: the phonation region and the aphonia region. The lower edge of the phonation region displays the individual dependence of the phonation threshold pressure on the fundamental frequency and therefore it gives a measure of the phonatory effort at a certain pitch. The explicit connection between sound pressure level and phonation threshold pressure, has already been measured and analyzed successfully with model calculations in (Titze, 1992) and (Titze and Sundberg, 1992). We will show that it is also possible to examine the lower boundary of the voice range profile measuring of the phonation onset time at different fundamental frequencies.

Digital High Speed Glottography and Motion Analysis

Since digital high speed glottography turned out to be presently the only method to view directly vocal fold vibrations and to extract the phonation onset time of each vocal fold, we discuss in this section the possibilities provided by a high speed camera in comparison to other diagnostic tools.

Organic voice disorders indicated by a morphological change of the laryngeal texture, such as cysts, polyps, granuloma or carcinoma can usually be diagnosed with the naked eye supplemented with a laryngeal mirror or an endoscopic device. In contrast, the reasons for hoarseness related to dysfunctions of the laryngeal muscles, i.e. functional dysphonia, are hidden in the complex vibratory patterns of the vocal folds. The oscillation frequencies are generally higher than 100 Hz, whereas the human eye is able to perceive maximally 40 distinct images per second. Therefore, sophisticated examination techniques are required to record complex vocal fold oscillations. The digital high speed camera system used in our research permits a direct observation of vocal fold oscillations and,consequently, the detection of important information about glottal asymmetry, aperiodic and transitory oscillation, e.g. phonation onset, the adduction and the abduction processes. These features are not entirely available from acoustic signals (AS), electroglottograms (EGG), photoglottograms (PGG), video kymography, and video stroboscopy. The description of these examination methods can be found in (Wendler and Seidner, 1987), latest studies in video kymography are presented in (Svec and Schutte, 1996).

Semi-automatic motion analysis is applied to reduce the digital image data from high speed video sequences by extracting the trajectories of three points on each vocal fold border. A detailed description of the high-speed recording system and the image processing algorithm is given in (Wittenberg et al., 1995). The resulting high speed glottograms (HGG) are the experimental counterparts of the time series obtained in computer simulations, and therefore, a direct comparison is possible. Unlike is the situation when AS are used to explore the glottal dynamics. Since the vocal tract resonator leaves its foot prints in the speech signal, inverse filtering has to be applied to recover the glottal airflow function. At this level, it is still not possible to observe the left and right vocal fold oscillations separately because the reconstructed glottal airflow carry the *integral* information of the glottal area oscillation. The possibilities of PGG and EGG are limited likewise. Stroboscopy is a powerful tool to examine symmetric, regular and stationary vocal fold vibrations. Since the time resolution of a conventional video camera is about 20 ms, its application has to fail for the analysis of the phonation onset which takes between 5 ms and 100 ms. Video kymography can be considered as the one-dimensional version of high speed cinematography which, in contrast, provides spatial information in two dimensions.

Biomechanical Model Analysis

Vocal fold models are supposed to sketch the main laryngeal properties. The model complexity, i.e. the number of degrees of freedom and the number of model parameters, is minimized in respect to the given purposes. This means, that a more detailed model design is required for issues related to the tract source coupling or to the excitation of anterior-posterior oscillation modes of the vocal folds,

whereas a simple design permits already a very good insight into the basic features of experimentally observed vocal fold vibrations, e.g. the phonation onset. The set of model parameters represents an integral description of the properties of the oscillating tissue. The detection and variation of the controlling parameters of the phonation onset in the model allows an approximate reconstruction of the complex mechanisms in the natural larynx, including its dysfunctions.

In the past, one central goal of model analysis was to determine the critical subglottal pressure for the onset of self-sustained oscillations (phonation). The change in the dynamic behaviour is based on a Hopf bifurcation. At a critical pressure value, a stable equilibrium state of the considered model, i.e. the rest position, becomes unstable and the dynamics undergo a change from a fixed point to a limit cycle attractor.

Our computer simulations have been performed with the simplified two-mass model described in (Steinecke and Herzel, 1995) which permits a widely analytic access. The critical subglottal pressure of the model is given by $P_{th} \approx 2.4$ cm H_2O for standard phonation conditions. This result is positioned at the lower end of the threshold pressure scale, based on experimental measurements ($P_{th} \approx 3...4$ cm H_2O) (Titze, 1994a). Hence, the model simulations are supposed to describe reasonably the phonation onset of a natural voice. Steinecke and Herzel analyzed the Hopf bifurcation in the simplified two-mass model by calculating numerically bifurcation diagrams (Steinecke and Herzel, 1995). Titze deduced approximatively the threshold pressure for the excitation of small amplitudes of mucosal surface waves (Titze, 1988). The simple formula predicts a linear growth of the phonation threshold pressure with the tissue damping coefficient, the fundamental frequency and the prephonatory glottal half-width at the Hopf bifurcation. Our model simulations were consistent with these theoretical predictions. Experiments with a physical model of the mucosa (Titze et al., 1994) confirmed the results in Titze, 1988.

Another phenomenon which has to be mentioned is hysteresis. In this case, the Hopf bifurcation is characterized by two critical pressures. The upper threshold pressure marks the transition from the steady state to a limit cycle and a lower one defines the minimum sustaining pressure. This means that certain model or larynx configurations exhibit intermittent oscillatory regimes characterized by sudden changes from finite amplitude oscillation to aphonia. Hysteresis has been found in recent excised larynx experiments (Berry et al., 1996) and in a physical model of the mucosa (Titze et al., 1994). Two threshold pressures have been calculated for the plank model by Lucero (Lucero and Gotoh, 1993). In Lucero, 1993, the author shows that for a divergent glottal shape the two-mass model exhibits hysteresis.

For our purpose, we assume that hysteresis is neclectable, since generally, the difference between the two critical pressures is small compared with the pressure range used for phonation.

Parameterization of the Phonation Onset

Analytical Envelope Solution

For sake of simplicity we abandon the goal of finding an explicit solution for the collective dynamics of a multi-mass model. In contrast, we simply write the solution for the oscillation amplitude of the vocal folds as a function of time t in a rough approximation

$$x(t) = r(t)sin\phi(t) \tag{1}$$

where $\phi(t)$ is the phase and r(t) corresponds to the envelope function of the oscillation. We are interested in the growth of r(t), starting from small initial elongation values.

In general the equation of vocal fold motion can be transformed into the following autonomous nonlinear differential equation (Jackson, 1989); (Verhulst, 1990):

$$\ddot{x} + x = F(x, \dot{x}) \tag{2}$$

where $F(x,\dot{x})$ is a nonlinear driving force. Since small amplitudes are considered, Eq. (2) can be expanded in Taylor series. After using polar coordinates (r,ϕ) and averaging over one oscillation period, the radial differential equation reads (Verhulst, 1990)

$$\dot{r} = a_0 + a_1 r + a_2 r^2 + a_3 r^3 + ... \tag{3}$$

where the coefficients a_i are stemming from the Taylor expansion. At this point, there are two possibilities to proceed. We can normalize Eq. (3) applying a near-identity transformation (Verhulst, 1990)

$$r = q + \alpha_2 q^2 + \alpha_3 q^3 + ... \quad , \tag{4}$$

where we adjust the coefficients a_i in a way, that the resulting equation for q becomes linear in the ideal case. This method facilitates the calculation of a precise solution of Eq. (3) depending on the degree of normalization, i.e. on the number of nonlinearities which are eliminated. In particular, it is possible to transform Eq. (3) into the normal form of the Hopf bifurcation (compare Eq.(5)).

Another way is the use of apriori knowledge about the observed phonation onset curves to simplify Eq. (3). Considering Fig. (3) we realize, that there is an asymmetry between the positive and negative envelope slope of the oscillation, which is manifested in different absolute saturation values. It can be shown that only even nonlinearities in Eq. (3) produce an asymmetry between the positive and the negative saturation value by solving the equilibrium condition ($\dot{r}=0$). Since we are interested in the phonation onset

time which is identical for the positive and the negative envelope slope, we drop all even nonlinearities and get

$$\dot{r} = ar + br^3 + [O(5)] \quad . \tag{5}$$

Neglecting the higher order terms [O(5)], the advantage of Eq. (5) in contrast to Eq. (3) is its simple solution which reads

$$r(t) = \left(\left[\frac{1}{r_0^2} + \frac{b}{a}\right] e^{-2at} - \frac{b}{a}\right)^{-\frac{1}{2}} \tag{6}$$

where $r_0 = r(0)$ is the initial elongation. In the next chapter we will discuss the general features of Eq. (6) and its applicability regarding the quantification of the phonation onset.

General Features of the Envelope Solution

For $t \to \infty$, we get the saturation elongation, i.e. the amplitude of the limit cycle attractor

$$r_\infty = \lim_{t\to\infty} r(t) = \sqrt{-\frac{a}{b}} \quad . \tag{7}$$

For small values of r ($t \to 0$) we can neglect the cubic term in Eq. (5) ($ar \ll br^3$) and we simply find an exponential law

$$r(t) \sim r_0 e^{at} \quad . \tag{8}$$

Hence, for $a<0$, the oscillation will be damped out. In this case, more energy is dissipated in the tissue than is transferred from the airflow into the vocal folds. The inverse situation is given for $a>0$, where the elongation is growing exponentially from cycle to cycle. As the elongation increases ($t \to \infty$), the nonlinear term controlled by the parameter $b<0$ in Eq. (5) becomes more and more dominant and enforces a saturation behaviour which can be written as

$$r(t) \sim r_\infty \left(1 - \frac{1}{2}\left[\left(\frac{r_\infty}{r_0}\right)^2 - 1\right] e^{-2at}\right) \tag{9}$$

It can be shown analytically (Lucke and Mergell, in preparation) for the simplified two mass model (Steinecke and Herzel, 1995) that

$$a = \beta(P_s - P_{th}) \quad , \tag{10}$$

where P_s is the subglottal pressure, P_{th} is the phonation threshold pressure for certain larynx configurations and β is a proportionality factor, see also Fig. (7). Hence, a vanishes when P_s becomes P_{th}, exponential growth occurs for $P_s > P_{th}$ and damping exists for $P_s < P_{th}$.

Here we summarize briefly that the phonation onset is considered to be based on a supercritical Hopf bifurcation characterized by an exponential growth which in turn is delimited by a cubic nonlinearity. The change in dynamics from a fixed point to a limit cycle is a steady process and the saturation elongation obeys the relation $r_\infty \sim \sqrt{a}$, Fig. (1) (Berge' et al., 1984). Negative values of b and the incorporation of higher nonlinearities in Eq. (5) lead to a subcritical Hopf bifurcation with hysteresis, Fig. (2) (Berge' et al., 1984). There are parameter regions, $\beta(P_{ms} - P_{th}) < a < 0$, where phonation intervals (limit cycle) as well as intermittent aphonia (steady state, fixed point) coexist. In this case,

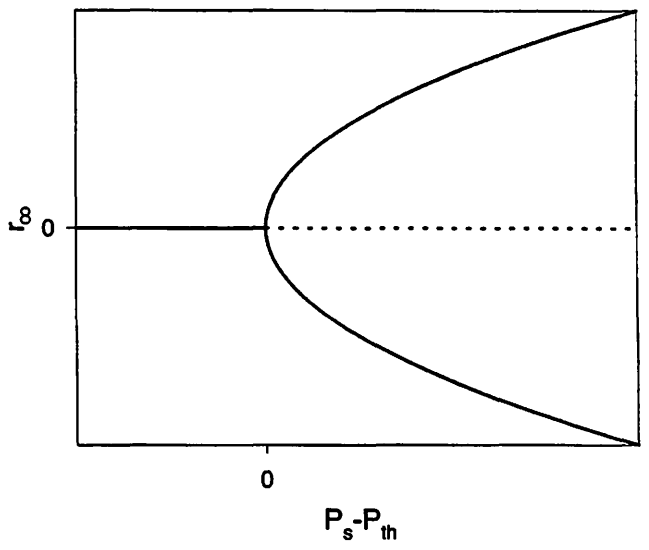


Figure 1. Note, that $P_s - P_{th} \sim a$. Hopf bifurcation based an the equation $\dot{r} = ar + br^3$, $b<0$. $r_- = \sqrt{-a/b}$ defines the saturation value of the envelope curve. At $P_s = P_{th}$ the fixed point solution is replaced by a stable limit cycle. Since the bifurcation branches start above the critical pressure P_{th}, i.e. $P_s > P_{th}$, the bifurcation is called supercritical. Solid lines indicate that the equilibrium solution is stable, dotted lines mark unstable solutions.

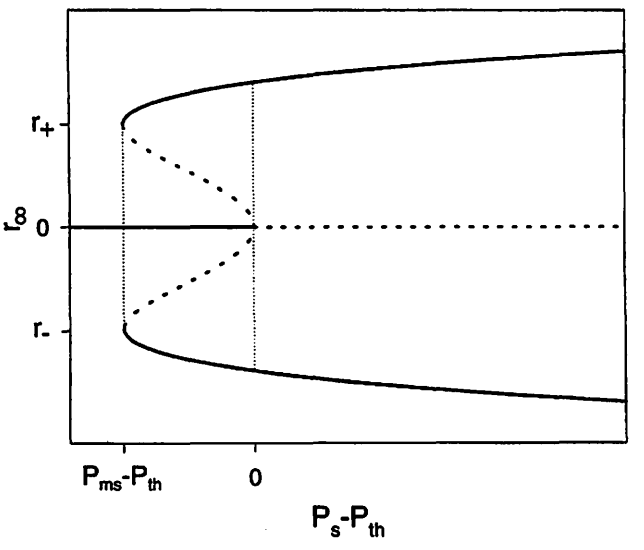


Figure 2. Hysteresis in the case of a subcritical bifurcation, where the branches exist already below the critical pressure P_{th}, i.e. $P_{ms} < P_s < P_{th}$. $P_{ms} \sim -P_{th} - (b^2/4c)$, where P_{ms} is the minimum pressure required to sustain phonation. The underlying equation is given by $\dot{r} = ar + br^3 + cr^5$, $b>0$, $c<0$. $r_\pm = \sqrt{-b/2c}$.

the threshold pressure for a transition from steady state to limit cycle is larger than the pressure value, for which phonation can just be sustained. At both thresholds, the dynamic change happens abruptly. In the intermediate region one observes jumps from one dynamic regime into the other. Since it is not possible to give a simple solution for the respective normal form and since we assume, that most of observed phonation onsets are produced in a P_s-region, where hysteresis is not essential, we forgo a more detailed discussion of the subcritical case shown in Fig. (2).

Definition of the Phonation Onset Time

By introducing the ratio $=\zeta(r_0/r_\infty)^2$, we get the parameterization from Eq. (6)

$$r(t) = r_0\left([1-\zeta]e^{-2at}+\zeta\right)^{-\frac{1}{2}} \quad . \qquad (11)$$

Note, that this solution contains beside the initial amplitude r_0 and the saturation amplitude r_∞ only a single free parameter a.

Our aim is to extract from this expression a unique definition of the phonation onset time. Here we present two ways to define the characteristic time constant from high speed video sequences considering Eq. (11) and we show that all definitions of the phonation onset time based on the presented theory are compatible.

1. By means of a nonlinear curve fit with the parameterization given in Eq. (11) one can determine r_0, r_∞ and a. In Fig. (3) this method is demonstrated. The fit has been performed to the series of elongation maxima extracted from a simulated vocal fold oscillation. Fig. (4) shows the application to high speed sequences discussed below in more detail. It will be shown, that an adequate definition of the phonation onset time is given by $\tau_{on}=a^{-1}$. If the amplitude varies strongly when saturation has been reached it is important to limit the data points taken from this region. Although it is an advantage that even noisy data can be handled in this way, the method is intricate and requires a large amount of computing time for the automatic determination from HGG.

2. A very simple and robust method is to chose two different threshold fractions, S',S, of the saturation elongation and to define the phonation onset time as the difference between the respective durations of amplitude growth. The duration of growth from the initial elongation r_0, e.g the first oscillation maximum of the phonation onset, to an arbitrary fraction S, e.g. S=70%, of the saturation elongation r_∞ is given by

$$\tau_S = \frac{1}{2a}\ln\left[\frac{\zeta^{-1}}{S^{-2}-1}\right] \quad , \qquad (12)$$

and therefore we find

$$\tau_{on} = \tau_S - \tau_{S'} = \frac{1}{2a}\ln\left[\frac{S'^{-2}-1}{S^{-2}-1}\right] = \frac{Z(S,S')}{a} \quad . \qquad (13)$$

At this point, it becomes clear that any estimation of the onset time, Eq. (13), is proportional to 1/a and the prefactor is determined by the chosen thresholds S,S'. Fulfilling the condition

$$S' = \frac{1}{\sqrt{e^2(S^{-2}-1)+1}} \quad , \qquad (14)$$

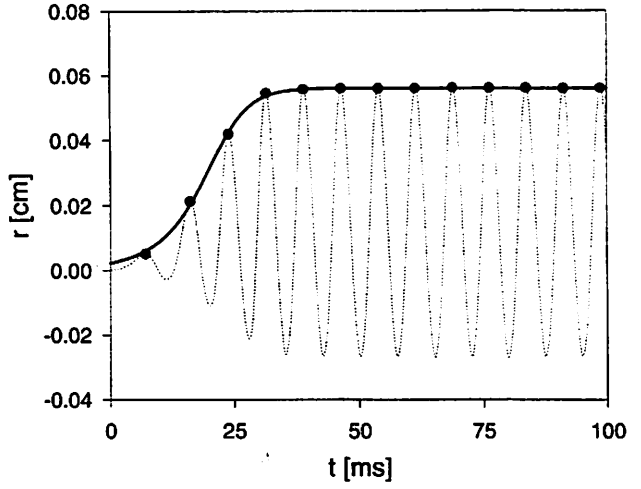


Figure 3. Simulation of a phonation onset using the simplified two-mass model by Steinecke and Herzel. The dotted curve shows the oscillation amplitude. The series of oscillation maxima which define the envelope curve are marked with filled circles. The solid line represents the performed fit to the maxima taking the parametrization from Eq. (11). $\tau_{on}=(7.1\pm0.1)$ms.

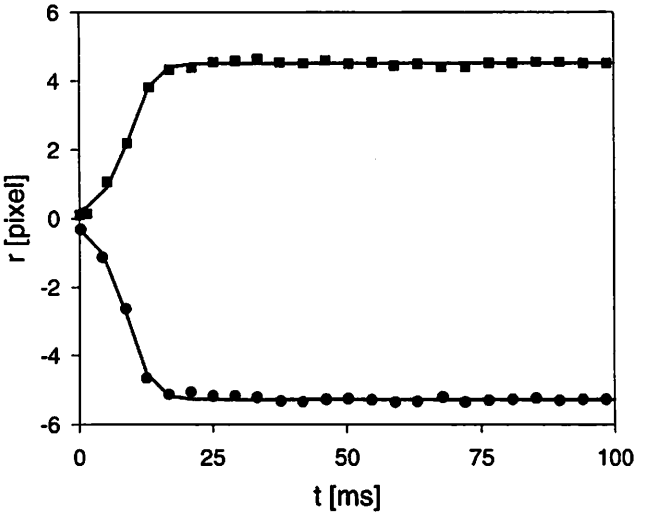


Figure 4. Fits to the onset envelope curve of left (filled quarters) and right (filled circles) vocal fold oscillations. The HGG signal of the phonation onset has been filtered with a bandpass to reduce artifacts and noise components. $\tau_{on}=(3.8\pm0.2)$ms.

where e is the Euler number, we can find values S' and S for which Z(S,S') is unity and therefore $\tau_{on}=a^{-1}$. Hence, the inverse of the parameter a can be directly identified with the phonation onset time and the dependencies on initial and saturation elongation has been eliminated. This result is a consequence of the fact, that a is invariant under scaling and time translation of the envelope curve r(t). For centering the threshold crossings around the half saturation elongation we set S'=0.5-p and S=0.5+p. Solving Eq. (14), we find p=0.178 and therefore S'=32.2%, S=67.8%.

We want to realize an automatic determination of the onset time from HGG and to introduce its standardization for an adequate comparison base in respect to different high speed sequences as well as different studies. The parameter a can be determined easily in different ways and the time interval a^{-1} can be interpreted as the duration of amplitude growth from 32.2% to 67.8% of the saturation elongation. Hence, the definition

$$\tau_{on} = \frac{1}{a} \qquad (15)$$

will serve for all further considerations. This definition has a general character since it describes the phonation onset mechanism in a universal way, i.e. as a Hopf bifurcation. Thus, it applies to many other phonation onset signals.

Results

Simulations of the Two-Mass Model

In this section the dependence of the phonation onset time on several model parameters is analyzed. Details regarding the model and parameters are given in the study by (Steinecke and Herzel, 1995). Essential physiological parameters varied in phonation are the glottal rest area a_0, the subglottal pressure P_s and the fundamental frequency f_0. It can be shown for the simplified two-mass model (Steinecke and Herzel, 1995) that for small amplitudes the parameter a is proportional to the factor P_s/a_0 (Lucke and Mergell, in preparation). The larger this factor is, the more a small perturbation of the vocal folds in rest position is amplified exponentially. Consequently, a will grow while raising the subglottal pressure and diminishing the glottal rest area (maximum adduction) whereas phonation onset time behaves inversely. These results are depicted in Fig. (5) and Fig. (6). Incomplete closure effects a raise of the phonation threshold pressure so that higher subglottal pressures are required to compensate a glottal chink which for instance is frequently found in laryngeal nerve paralysis. In Fig. (7) we demonstrate how to extrapolate the phonation onset time data in order to determine the phonation threshold pressure using Eq. (10).

The model simulations show that the phonation onset time increases systematically with the pitch f_0. This accords with several experimental and theoretical studies which are compiled in Titze, 1994a. Fig. (8) shows the Hopf bifurcation obtained by linear stability analysis and verified with model simulations. The threshold pressure is presented as a function of the parameter q, scaling the natural frequency f_{nat} of the model masses. The scaling is simply performed by

$$m_i' = \frac{m_i}{q} \quad , \quad k_i' = qk_i \quad \rightarrow \quad f_{i,nat}' = qf_{i,nat} \quad , \qquad (16)$$

where m_i and k_i are mass and stiffness coefficients. In this two-dimensional bifurcation diagram, the threshold curve separates the parameter plane into the regions of sustained phonation and aphonia. Since the threshold curve exhibits

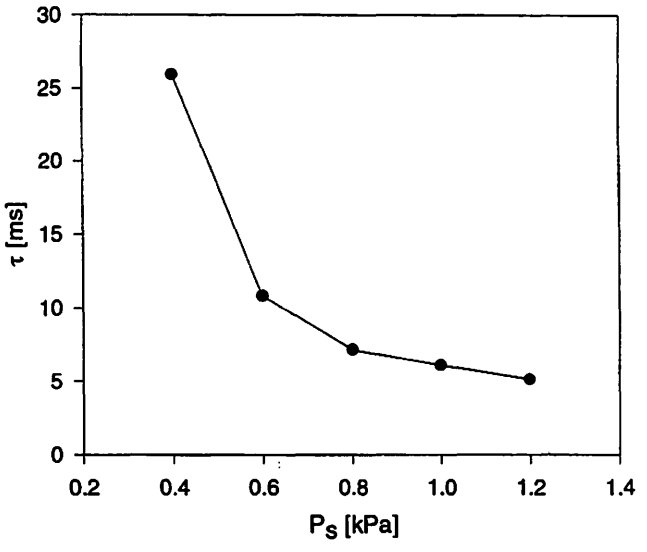


Figure 5. Phonation onset time in the two-mass model as a function of subglottal pressure. The equations and the parameters are given in Steinecke and Herzel, 1995.

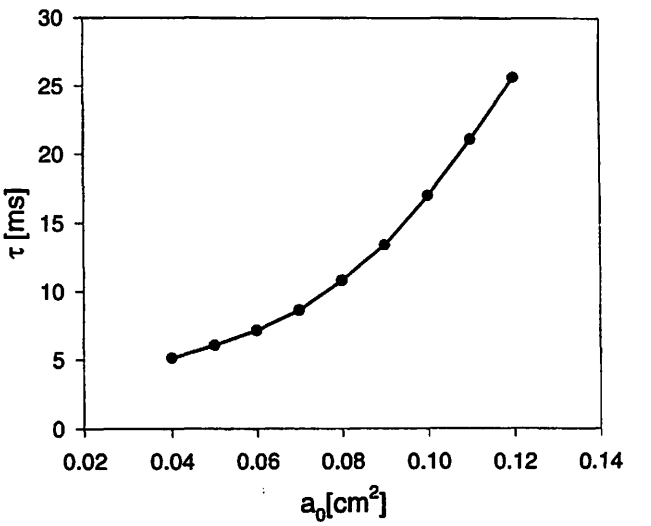


Figure 6. Phonation onset time in the two-mass mode as a function of glottal rest area.

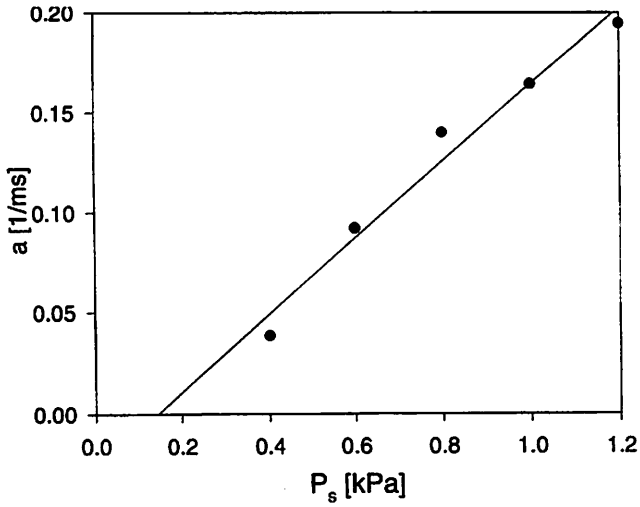


Figure 7. The parameter $a=1/\tau_{on}$ in the two-mass model as a function of the subglottal pressure. This figure suggests, that the phonation threshold pressure can be determined by the extrapolation of onset time data to a=0, where $P_s=P_{th}$.

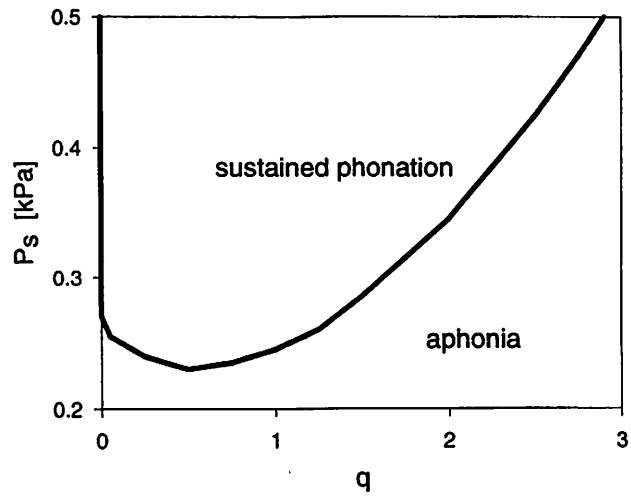


Figure 8. Two-dimensional Hopf bifurcation diagram in the two-mass model. The phonation threshold pressure curve subdivides the parameter plane into two stability regions. Aphonia corresponds to a stable fixed point, i.e. the vocal folds persist in rest position. Across the threshold curve a dynamical change occurs and sustained phonation sets in.

a minimum, we state that there exists an optimal relation between pitch and subglottal pressure for a given laryngeal configuration. This is a very important conclusion in respect to functional dysphonia. For example, in the case of hyperfunction the pitch is abnormally shifted to a higher value (Wendler and Seidner, 1987). Hence, pressure has to be increased for sustaining phonation. Near threshold pressure, sudden jumps to aphonia occur, an effect which is comparable with the break of phonation at the upper end of a glissando in the head-like register. Such aphonic episodes indicate a phonation onset with hysteresis, compare Fig. (2). As mentioned above, phonation onset time grows with an increase of the threshold pressure at fixed lung pressure. Thus, the measurement of this parameter is very informative for a detailed assessment of a certain vocal state.

Experimental Data from High Speed Glottograms

The discussed method offers a base for a standardized quantitative determination of τ_{on} from HGG and its direct linkage to a certain model configuration. In this section we will analyze three phonation onsets of healthy voices with the subjective qualities *soft, normal*, and *hard*. These examples are given to illustrate our approach. A detailed statistical study of phonation onsets in 52 subjects with functional dysphonia is published elsewhere (Wittenberg et al., 1997).

As already mentioned, semiautomatic motion analysis software was applied to extract the glottal geometry from high speed sequences (Wittenberg et al., 95). As a first step, the glottal area is separated from its laryngeal environment for each digitized image (128 pixel x 128 pixel, 0.1 mm/ pixel). Starting from equidistant reference points on the principle glottal axis, the vocal fold oscillation amplitude is

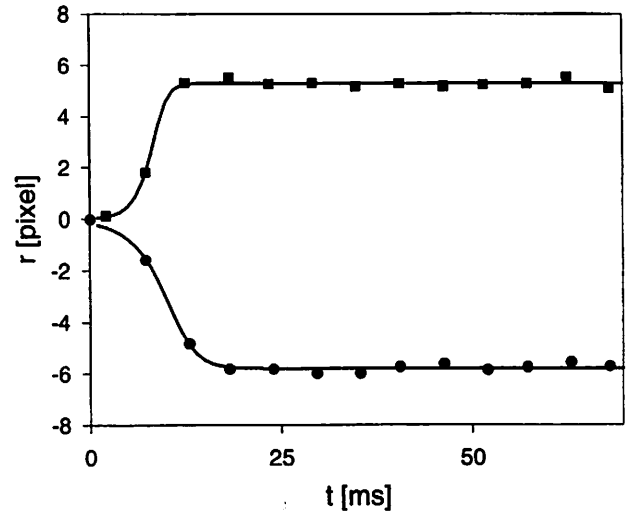


Figure 9. Phonation onset with subjective assessment hard*: $\tau_L=(1.7\pm0.6)ms$, $\tau_R=(3.3\pm0.3)ms$. $f_0=188$ Hz.*

determined via gray scale edge detection. The resulting time series of the left and the right vocal fold oscillation has been bandpass filtered to reduce the spectral components stemming from artifacts and noise. We used a Hamming window with the passband $0.9\,f_0<f<1.1\,f_0$, where f_0 is the fundamental frequency. Finally, the phonation onset time has been determined performing a nonlinear fit to the onset envelope curve.

Considering the onset diagrams, we state that data can be described excellently by the fits. As expected, the *soft* phonation onset is related to a relatively long onset time, whereas the *hard* onset characteristically has a short onset time. The two examples reflecting the relation between the phonetic classification and the corresponding numeric measure are given in Fig. (9) and Fig. (10; following page).

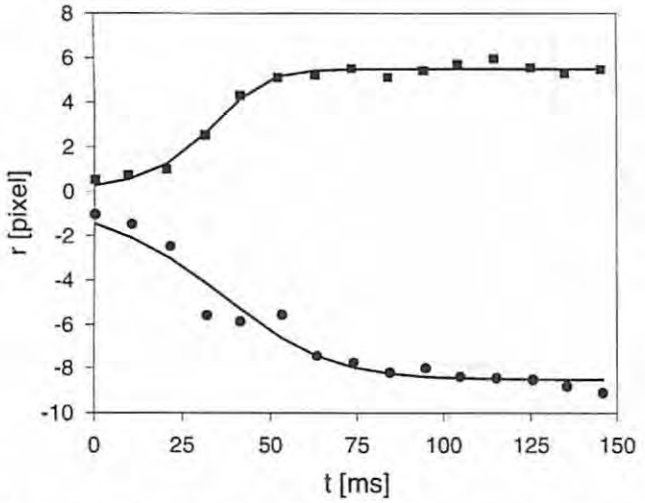


Figure 10. Phonation onset with subjective assessment soft*:* τ_L*=(12.8±1.4)ms,* τ_R*=(27.0±4.4)ms.* f_0*=101 Hz.*

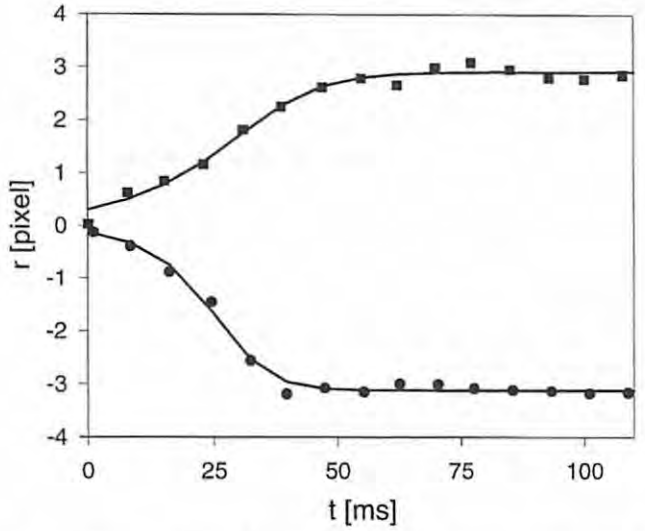


Figure 11. Phonation onset with subjective assessment normal*:* τ_L*=(15.6±1.7)ms,* τ_R*=(9.3±0.9)ms.* f_0*=131 Hz.*

The corresponding kymograms are shown in Fig. (12) and in Fig. (13). From each image frame of the corresponding high speed sequence (temporal resolution: 1922 frames/s) one single horizontal image line (128 pixel) (centered in anterior-posterior direction of the larynx and perpendicular to the principle glottal axis) has been extracted. These image lines have been put together in chronological order. The time windows have a length of 0.4 s with a linear scaling. In contrast to the hard onset, there is no prephonatic closure prior to the soft onset. This is equivalent to a large glottal rest area. In accordance to the theory discussed above, we detect a long phonation onset time. In Fig. (11) we show an example of a phonation onset with assessment *normal* with

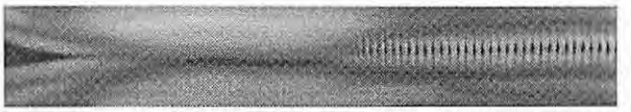


Figure 12. Phonation onset of a male subject (32 years, healthy) with subjective assessment hard. After the adduction (about 0.1s), a complete and long prephonatic closure (about 0.1s) can be observed. Even the false vocal folds, indicated by the shadows margining the zipper-like oscillation patterns, have contact with each other. The oscillation amplitude grows quickly to its saturation value.

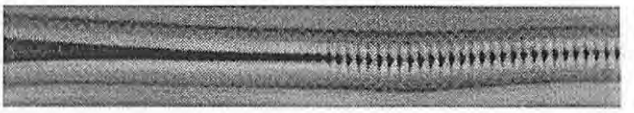


Figure 13. Phonation onset of a male subject (32 years, healthy) with subjective assessment soft. No prephonatic closure occurs and the adduction process is followed directly by a relatively long phonation onset.

an intermediate phonation onset time. But, in general, a non-ambiguous allocation of the analyzed *normal* onsets regarding the onset time scale turned out to be difficult.

Conclusion

One of the main goals of voice research is a deeper insight into voice disorders, where the complicated laryngeal processes as well as the textural or neural changes are detectable exclusively in the complex oscillation patterns of the vocal folds. On speaking of functional dysphonia, we refer to a vast array of pathologies, which have in common only the fact, that the visible laryngeal parts exhibit no morphological changes. For a precise diagnosis of these voice pathologies, a quantitative classification base is required. Thus, the combined application of high-speed camera technique, digital motion analysis, and model simulations is indispensable for the isolation of physiological parameters which categorize and quantify functional voice disorders. In this work, we isolated the central quantitative measure characterizing the phonation onset, i.e. the phonation onset time. It has been defined as the duration of amplitude growth from 32.2% to 67.8% of the saturation amplitude. The dependence of this parameter on the respective laryngeal configuration can be mapped in biomechanical models and suggests therefore the important role of the phonation onset time for the diagnosis of functional dysphonia.

Future investigations will be devoted to the systematic analysis of our vast data base from subjects with functional dysphonia, and to the automatization and optimization of the parameter extraction from high-speed video sequences.

Acknowledgments

This work has been supported by the Deutsche Forschungsgemeinschaft. We thank I.R. Titze and D. Berry at the National Center for Voice and Speech in Iowa City for many fruitful discussions.

Bibliography

P. Berg'e, Y. Pomeau and C.Vidal. *Order within Chaos.* John Wiley & Sons, Hermann (1984).

D.A. Berry, H. Herzel, I.R. Titze and B.H. Story. Bifurcations in excised larynx experiments. *Journal of Voice*, 10, 129—138 (1996).

H. Herzel, D.A. Berry, I.R. Titze, and M. Saleh. Analysis of Vocal Disorders with Methods from Nonlinear Dynamics. *Speech Hearing Res.*, 37, 1008—1019 (1994).

K. Ishizaka and J.L. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst. Techn. J.*, 51, 1233—1268 (1972).

E.A. Jackson. Perspectives of nonlinear dynamics. Cambridge University Press (1989).

J.J. Jiang and I.R. Titze. A methodological study of hemilaryngeal phonation. *Laryngoscope*, 103, 872—882 (1993).

J.C. Lucero. Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region. *J. Acoust. Soc. Am.*, 94, 3104—3111 (1993).

J.C. Lucero and T. Gotoh. On the threshold pressure and the minimum sustaining pressure in the vocal fold oscillation. *J. Acoust. Soc. Jpn.*, 14, 213—214 (1993).

I. Steinecke and H. Herzel. Bifurcations in an asymmetric vocal fold model. *J. Acoust. Soc. Am.*, 97, 1874—1884 (1995).

J.G. Svec and H.K. Schutte. Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, 10, 201—205 (1996).

I.R. Titze. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.*, 83, 1536—1551 (1988).

I.R. Titze. Acoustic interpretation of the voice range profile (Phonetogram). *J. Speech Hearing Res.*, 35, 21—34 (1992).

I.R. Titze. Phonation threshold pressure: A missing link in glottal aerodynamics. *J. Acoust. Soc. Am.*, 91, 2926—2935 (1992).

I.R. Titze and J. Sundberg. Vocal intensity in speakers and singers. *J. Acoust. Soc. Am.*, 91, 2936—2946 (1992).

I.R. Titze, R. Baken, and H. Herzel. Evidence of Chaos in Vocal Fold Vibration. In *Vocal Fold Physiology: Frontiers in Basic Science* edited by I. R. Titze, Singular Publishing Group, San Diego, 143—188 (1993).

I.R. Titze, S.S. Schmidt and M.R. Titze. Phonation threshold pressure in a physical model of the vocal fold mucosa. *J. Acoust. Soc. Am.*, 97, 3080—3084 (1995).

I.R. Titze. *Principles of Voice Production.* Prentice Hall, Englewood Cliffs, New Jersey (1994).

I.R Titze. Towards standards in acoustic analysis of voice. *Journal of Voice*, 8, 1—7 (1994).

K. Verdolini, I.R. Titze and D.G. Druker. Changes in phonation threshold pressure with induced conditions of hydration. *Journal of Voice*, 4, 142—151 (1990).

K. Verdolini-Marston, M. Sandage and I.R. Titze. Effect of hydration treatments on laryngeal nodules and polyps and related voice measures. *Journal of Voice*, 8, 30—47 (1994).

F. Verhulst. *Nonlinear Differential Equations and Dynamical Systems.* Springer-Verlag Berlin Heidelberg New York (1990).

J. Wendler and W. Seidner. *Lehrbuch der Phoniatrie.* Georg Thieme, Leipzig (1987).

T. Wittenberg, M. Moser, M. Tigges, and U. Eysholdt. Recording, processing, and analysis of digital high-speed sequences in glottography. *Machine Vision and Applications*, 8, 399-404 (1995).

T. Wittenberg, P. Mergell, M. Tigges, and U. Eysholdt. Quantitative characterization of functional voice disorders using motion analysis of highspeed video and modeling. To appear in *Proc. ICASSP-97*, Munich, Germany (1997).

NCVS Status and Progress Report - 12
June 1998, 65-69

Muscle Fiber Type Distribution in the Normal Human Levator Veli Palatini Muscle

Jerald B. Moon, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Sue Ann Thompson, Ph.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Elise Jaeckel, M.D.
Department of Otolaryngology, Kansas University Medical Center
John W. Canady, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa

Abstract

Objective

This study examined the muscle fiber type distribution within the normal adult levator veli palatini muscle.

Methods

Levator veli palatini muscle tissue was harvested from the palates of 12 (7 female, 5 male) adult non-cleft cadavers. Adjacent sections were stained for adenosine triphosphatase at pH 10.4 or 4.2. After mounting, magnifying, and photographing, Type I versus Type II fiber types were differentiated by the intensity of, or by the inhibition of, staining of matched fibers at each pH level. Type I fibers stained light at pH 10.4 and dark at pH 4.2, while Type II fibers stained light at pH 4.2 and dark at pH 10.4.

Main Outcome Measures

The number of fibers counted for each specimen ranged from 60 - 616. The numbers of Type I and Type II stained fibers appearing in each muscle tissue sample were determined and expressed as a percentage of the total number of fibers identified. A few identified fibers could not be labelled as either Type I or Type II.

Results

The overall proportion of Type I fibers, averaged across all specimens, was 59.8%. Male specimens had 67.4% Type I fibers and 31.8% Type II fibers, while female specimens had 54.4% Type I fibers and 44.4% Type II fibers.

Conclusions

Observed fiber type distributions compared favorably with those reported for other articulatory muscles, but differed slightly from previously reported distributions for normal levator veli palatini. The distributions observed in this study provide a baseline against which to relate fiber type data from the levator veli palatini of cleft palates to the functional status of the velopharyngeal mechanism.

Introduction

In their chapter on behavioral therapy for speakers with velopharyngeal impairment, Tomes et al. (1997) reviewed a number of approaches that have been attempted over the past fifty years to improve velopharyngeal function. Many previous approaches were based on assumptions regarding patients' velopharyngeal physiology and "its potential to change with therapy" (p. 530). Tomes et al. (1997) correctly pointed out that previous behavioral approaches have assumed that the individual being treated possesses the inherent anatomic and physiologic capability to improve velopharyngeal functioning for speech. This may not be a valid assumption. That is, speakers with repaired palatal clefts may suffer from physical limitations such as insufficient muscle strength or endurance and a low threshold of fatigue that would undermine behavioral approaches that focus on "simply" learning how to use the mechanism more appropriately.

The production of speech involves frequent repetitive elevations of the velum. In previous works (Kuehn and

Moon, 1994), it was shown that, in normal speakers, activation levels of the levator veli palatini occurred at a low level within the muscle's full operating range. However, patients with surgically repaired cleft palates typically used much higher activation levels, often close to the upper limit of their range, and were still mildly hypernasal (Kuehn and Moon, 1995). If the levator veli palatini is repetitively exerted near its functional limits, the possibility for fatigue exists.

All human muscles contain some combination of muscle fibers differing in their histochemical composition and susceptibility to fatigue. Type I fibers are very fatigue resistant and generally have the lowest activation thresholds (Linssen et al., 1991). While they are well suited for prolonged activity such as postural maintenance, Type I fibers are also active during phasic activity. Type II fibers (specifically IIB) are very sensitive to fatigue and have a much higher functional activation threshold. They are typically recruited when higher force outputs are required (Eriksson, 1982; Linssen et al., 1991). Velar muscle force generation, fatiguability, and muscle fiber type composition have not been addressed systematically in either the normal or repaired palatal cleft condition. Tomoda et al. (1984) did report a distribution of 46-50% Type I versus 48-54% Type II fibers in the levator veli palatini muscle of normal specimens aged 20 - 89 years. Given Kuehn's (1991) interest in velar muscle strength or endurance training and our interest in the mechanisms underlying normal and disordered velar control, further delineation of the histochemical properties of velar muscles is considered an important endeavor.

Materials and Methods

Palatal tissue was obtained from 12 cadavers (7 female, 5 male). The female donors ranged in age from 77 to 93 years, while the males ranged from 62 - 87 years. The average age of the sample group was 82.4 years. The time interval from death to tissue harvest ranged from 1 hour - fifteen minutes to 10 hours, with an average of 5 hours - 15 minutes.

The entire soft palate was removed from the cadaver using a curved (#12) scalpel blade to cut it free of the alveolus and hard palate and placed nasal side up on a gauze square soaked with saline. The levator veli palatini was identified as distinct paired bundles of muscle appearing at the cut lateral edges of the soft palate just anterior to where it was severed from the hard palate. Blocks of this muscle tissue were dissected from the palate, frozen in liquid nitrogen, and stored at -70°C until sectioning. Tissue blocks were individually mounted on slices of cork in Tissue-Tek O.T.C. compound and then cut into 10 μm sections in a cryostat at -20 degrees.

Adjacent sections were preincubated in either (1) 0.1 M Barbital buffer with 0.18 M calcium chloride at pH 10.4 for 10 minutes, or (2) 0.1 M Barbital acetate buffer for 5 minutes at pH 4.2. All slide mounted and stained sections were then incubated in Barbital buffer with calcium chloride and 0.003 M ATP for (1) 30 minutes if preincubation was at pH 10.4, or (2) 60 minutes if preincubation was at pH 4.2. The pH 10.4 sections were then placed in 2% Cobaltous chloride for 3 minutes, rinsed with dH_2O, dipped in 10% Ammonium sulfide for 10 seconds, rinsed for 5 minutes with running tap water, and dehydrated with 95% ethanol before clearing with xylene and mounting with permount. The 4.2 sections were then rinsed with 1% Calcium chloride, placed in 2% Cobaltous chloride for 3 minutes, rinsed with 0.01 M Sodium barbital and then dH_2O, dipped in 1% Ammonium sulfide for 30 seconds, and again rinsed with running tap water before dehydrating with 95% ethanol, clearing with xylene and mounting with permount.

After mounting, sections were magnified 25X and photographed. Individual fibers were identified in adjacent sections and fiber types were then differentiated on the basis of the intensity of, or by the inhibition of, staining of matched fibers at each pH level. With this staining procedure, Type I fibers which are high in oxidative enzymes were lightly stained at pH 10.4, and appeared darker at pH 4.2. Type II fibers which are high in ATPase stained lighter

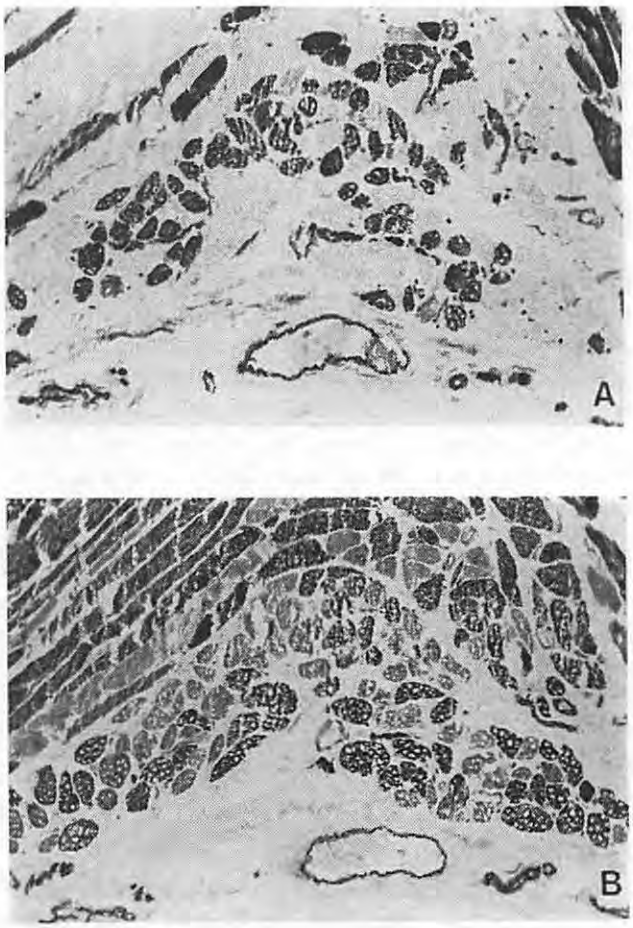


Figure 1. Adjacent sections of levator veli palatini muscle stained at pH 4.2 (A) and pH 10.4 (B).

at pH 4.2 and darker at pH 10.4. Examples of adjacent muscle sections preincubated at pH 10.4 and 4.2 are shown in Figure 1a and b. Percent distribution of Type I versus Type II fibers was calculated by dividing the number of Type I or Type Ii fibers by the total number of identified fibers. Fibers that could not be differentiated as either Type I or Type II were classified as undifferentiated.

Results

Measurement reliability was assessed by having a second judge count Type I versus Type II fibers using adjacent sections obtained from one of the twelve cadavers. A total of 97 fibers were judged. Both judged observed the same number of Type II fibers and differed by only one in their count of Type I fibers. Based on these results, reliability was judged to be adequate.

Table 1 shows the relative distribution of Type I and Type II muscle fibers for each of the 12 cadavers. The number of fibers counted in each cadaver ranged from 60 to 616. The recorded proportion of Type I fibers ranged from 31% to 75% across the 12 donors. The overall average proportion of Type I fibers in the sample was 59.8% (40.2% for Type II). The overall average proportions for male donors was 67.4% Type I (s.d. = 6.15; range = 57 - 75%) and 31.8 % Type II (s.d. = 5.42; range = 25 - 41%). For females, the proportion was 54.4 % Type I (s.d. = 14.63; range = 31-71%) and 44.4 % Type II (s.d = 14.16; range = 29 - 69%). There was no apparent relationship between proportion of Type I fibers and age. There was also no apparent relationship between the death - tissue harvest delay and fiber type distribution.

Table 1.
Distribution of Type I, Type II, and Undifferentiated Muscles Fibers

Sample #	Age	Gender	# Fibers Counted	% Type I	% Type II	% Undiff.
1	77	F	146	42	58	
2	90	F	356	70	30	
3	82	F	168	71	29	
4	82	F	616	66	33	1
5	62	M	191	57	41	2
6	81	M	250	72	28	
7	87	M	197	66	32	2
8	89	F	116	43	50	7
9	80	M	102	67	33	
10	74	M	60	75	25	
11	93	F	383	31	69	
12	92	F	198	58	42	
Average				59.83	39.17	
S.D.				13.47	12.95	

Discussion

The literature contains very few studies focussing on the histochemical makeup of the velar musculature in normal speakers. In fact, there are few studies of this kind that address any of the articulatory muscles. Our data do compare favorably with fiber type distributions recorded from masseter and the anterior portion of the medial pterygoid (both jaw closing muscles). Erickson and Thornell (1983) reported Type I fiber proportions of 62-72% and 64% for those two muscles respectively. Our overall distribution values for levator veli palatini Type I fibers are, on average, higher than those reported by Tomoda et al. (1984), while our Type II proportion is lower. Tomoda et al. harvested their muscle tissue from the region of the eustachian tube, while tissue for the present study was harvested from the palatal end of the muscle. Elder et al. (1982) addressed the issue of variations in fiber type distribution within a muscle. They determined that both the soleus and biceps brachii had significantly more fast twitch (type II) fibers at their insertions than their origins. The regional differences observed by Elder et al. (1982) might explain differences between the present data and Tomoda et al. (1984). However, a more detailed study using muscle tissue harvested from different regions of the levator veli palatini in the same cadaver is necessary to address this issue.

It is also of interest that Type I versus Type II distributions observed in the present study varied considerably across subjects. While the average Type I proportion was higher for males, a substantial amount of overlap was observed between the two groups. In addition, greater variability was observed among the female specimens than the male specimens. Between-subject variability in fiber type distribution has not been addressed specifically in the literature. However, previous researchers have reported fiber type distribution standard deviations of up to 14% (Simoneau et al., 1985; Tesch and Karlsson, 1985).

It should be noted that the subjects in the current study were all older individuals. There is obvious interest in determining fiber type distributions in younger individuals if we desire to relate muscle histochemistry to impaired function in that population. Recent studies (Grimby and Saltin, 1983; Sato et al., 1984; Lexell et al., 1986) have reported relatively stable fiber type distributions across ages ranging from young adulthood to 100 years. Based on her review of the effects of age and training on skeletal muscle physiology, Thompson (1994) concluded that loss of skeletal muscle mass with age is not associated with a preferential loss of a specific fiber type. Therefore, we might conclude that the fiber type proportions determined in the present study are representative of proportions existing in the levator veli palatini of younger individuals.

In their discussion, Eriksson and Thornell (1983) suggested that Type I fibers in the masseter and medial pterygoid are recruited at relatively low forces. Conversely, Type II fibers exhibit higher functional thresholds, generate larger forces, and may be best suited for bursts of intense, intermittent activity. When interpreting these observations in the light of other levator EMG data we have been collecting, some possible explanations for clinical observations of marginal VPI and support for levator muscle training may be offered. Previous work in our laboratory (Kuehn and

Moon, 1994) has shown that velar closure during speech produced by normal speakers is a low effort task requiring relatively little levator palatini muscle effort. That is, levator activation levels observed during speech production tend to lie at the low end of the levator activation range as recorded during a blowing task. One might argue then, that levator activity during conversational speech typically involves primarily the recruitment of fatigue resistant Type I fibers.

In a follow up study (Kuehn and Moon, 1995), levator activation levels recorded during speech produced by adults with repaired palatal clefts were related to their levator activation range. In contrast to the normal speakers studied earlier, most non-nasal speech sounds produced by these subjects were associated with near maximal levator activation. If the levator muscle is forced to work near its functional limits, recruiting fatiguable Type II muscle fibers, the possibility for fatigue obviously exists. If the threshold for fatigue is exceeded, speech would be expected to deteriorate rapidly and become excessively hypernasal. An individual with borderline velopharyngeal inadequacy may try to avoid this consequence by remaining just below the threshold for fatigue, even though he/she might then be somewhat continuously hypernasal.

This argument assumes that the velopharyngeal mechanism does fatigue when the levator muscle is forced to work near its functional limit. There are no published data that document this phenomenon either clinically or physiologically. Recent work by Kuehn and Moon (1997) has addressed this issue. In that study, normal speakers were required to reiterate the syllable /si/ 100 times while air pressure is delivered to the nasal passages, creating a load against which the levator muscle must work to elevate the soft palate. Electromyographic activity in the levator veli palatini muscle and force of velopharyngeal closure were monitored. Muscle fatigue was induced in most of the speakers at higher levels of loading (25 - 35 cm H_2O). That is, normal speakers did not show evidence of fatigue until velar loading necessitated the recruitment of Type II muscle fibers. Fatigue in these fibers reduced the muscle capacity to generate the forces needed to maintain velopharyngeal closure.

The normal speakers studied by Kuehn and Moon (1997) did not fatigue under typical non-taxing conditions where the activation levels of the levator veli palatini muscle fell far below maximum levels. Not until the muscle was "pushed" to a level closer to its physiologic limits did fatigue occur. As stated earlier, speakers with repaired palatal clefts appear to use near maximal levator activation levels routinely. If the levator vel palatini muscles of this population have a similar Type I - Type II muscle fiber distribution observed in the current study, it might be concluded that they would fatigue at lower levels of velar loading since the fatiguable fibers would be active from the outset. Little is known about fiber type distributions in the velopharyngeal musculature of this population. We are just beginning to study fiber type distributions in the levator muscles of palatal clefts. Although a couple of references to these fiber type distributions exist in the literature, they are conflicting. A predominance of both Type I and Type II fibers has been reported (Cohen et al., 1993; Schendel et al., 1994). Regardless of the fiber distribution pattern, the high levels of levator muscle activation observed by Kuehn and Moon (1995) during typical speech production would undoubtedly be associated with recruitment of a large proportion of both Type I and Type II fibers. An investigation of the fatigue properties of speakers with repaired palatal clefts in currently underway. Preliminary results suggest that these individuals do, in fact, experience velopharyngeal fatigue at much lower nasal air pressure loads compared to non-cleft speakers.

Muscle fatigue experienced by speakers with repaired palatal clefts may be related to two phenomena. First, the levator veli palatini muscle of speakers with repaired clefts may be less massive, positioned less adventitiously from a mechanical perspective, and possibly be forced to try to elevate a shorter scarred palate. In addition, the fiber type distribution of the levator muscle may be different from non-cleft speakers. Both of these conditions place the cleft speaker at a disadvantage. They are forced to contract the muscle at a much higher level, recruiting fatiguable fibers at much lower loads than normal speakers. Kuehn (1991) has introduced a therapy protocol that may produce positive physiologic, and hence perceptual effects in these speakers. The CPAP therapy protocol is an eight week intensive program of levator muscle strength/endurance training. Patients are required daily to produce a specified speech sample while wearing a nasal mask through which varying magnitudes of nasal air pressure are delivered. The magnitude of air pressure is gradually increased over the eight week period, as is the duration of time each time spent speaking under the nasal load condition.

Based on our previous electromyographic and histochemistry work, one might speculate that this type of therapy should be expected to improve velopharyngeal function. If the levator muscle can be strengthened, thereby increasing the muscles' operating range, activation levels required for typical speech production might be lowered within that range away from the threshold of fatigue. In addition, at the histochemical level, there is some evidence that strength and endurance training can result in a an increase in the proportion of non-fatiguable Type I fibers (Howald et al., 1985; Simoneau et al., 1985; Staron et al., 1991), or a conversion from Type IIB to more fatigue resistant Type IIA fibers (Allemeier et al., 1994). Current work in our laboratory is focusing on the effects of this type

of therapy on the fatigue threshold in an adult speaker with a repaired palatal cleft. Continued work in this area should be expected to shed more light on normal and disorder velopharyngeal function and our attempts to optimize it.

Acknowledgements

This manuscript is based on a paper presented at the annual meeting of the American Cleft Palate-Craniofacial Association, San Diego, 1996.

This work is supported by PHS Research Grant DC00976 from the National Institute on deafness and Other Communication Disorders.

References

Allemeier C, Fry A, Johnson P, Hikida R, Hagerman F, Staron R. Effects of sprint cycle training on human skeletal muscle. *J. Appl. Physiol.* 1994;77:2385-2390.

Cohen S, Corrigan M, Blaivas M, Kawamoto H, Markowitz B. Histochemical, morphometric and ultrastructural analysis of levator palatini muscle in human cleft palate. Paper presented at ?? Meeting, 1993.

Elder G, Bradbury K, Roberts R. Variability of fiber type distributions within human muscles. *J. Appl. Physiol.: Respirat. Eviron. Exercise Physiol.* 1982;53:1473-1480.

Eriksson P. Muscle-fibre composition of the human mandibular locomotor system: enzyme-histochemical and morphological characteristics of functionally different parts. *Swedish Dental Jour.* 1982; Suppl. 12.

Eriksson P, Thornell L. Histochemical and morphological muscle-fibre characteristics of the human masseter. The medial pterygoid and the temporal muscles. *Archs. Oral Biol.* 1983;28:781-795.

Grimby G, Saltin B. The ageing muscle. *Clin. Physiol.* 1983; 3: 209-218.

Howald H, Hoppeler H, Claassen H, Mathieu O, Straub R. Influences of endurance training on the ultrastructural composition of the different muscle fiber types in humans. *Eur. J. Physiol.* 1985;403:369-376.

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

Kuehn D, Moon J. Levator veli palatini muscle activity in relation to intraoral air pressure variation. *Jour. Speech Hear. Res.* 1994; 37: 1260-1270.

Kuehn D, Moon J. Levator veli palatini muscle activity in relation to intraoral air pressure variation in cleft palate subjects. *Cleft Palate-Cranio. Jour.* 1995; 32: 376-381.

Kuehn D, Moon J. Induced fatigue effects on velopharyngeal closure force and levator veli palatini muscle activity. Paper presented at annual meeting of american Cleft Palate-Craniofacial Assoc., New Orleans, 1997.

Lexell J, Downham D, Sjöstrom M. Distribution of different fibre types in human skeletal muscle: Fibre type arrangement of m. Vastus lateralis from three groups of healthy men between 15 and 83 years. *J. Neurol.* Sci. 1986; 72: 211-222.

Linssen W, Stegeman D, Joosten E, Binkhorst R, Merks M, Laak H, Notermans, S. Fatigue in type I fiber predominance: A muscle force and surface emg study on the relative roll of type I and type II muscle fibers. *Muscle & Nerve* 1991;14: 829-837.

Sato TH, Akatsuka H, Kuniyoshi K. Age changes in size and number of muscle fibers in human pectoral muscle. *Mech. Ageing Dev.* 1984; 28: 99-109.

Schendel S, Cholon A, Delaire J. Histochemical analysis of cleft palate muscle. *Plastic & Reconstruct. Surg.* 1994;94:919-923.

Simoneau J, Lortie G, Boulay M, Marcotte M, Thibault M, Bouchard C. Human skeletal muscle fiber type alteration with high-intensity intermittent training. *Europ. J. Appl. Physiol.* 1985;54:250-253.

Staron R, Leonardi M, Karapondo D, Malicky E, Falkel J, Hagerman F, Hikida R. Strength and skeletal muscle adaptations in heavy resistance-trained women after detraining and retraining. *J. Appl. Physiol.* 1991; 70: 631-640.

Tesch P, Karlsson J. Muscle fiber types and size in trained and untrained muscles of elite athletes. *J. Appl. Physiol.* 1985;59:1716-1720.

Thompson L. Effects of age and training on skeletal muscle physiology and performance. *Physical Therapy* 1994;74: 71-81.

Tomes L, Kuehn D, Peterson-Falzone S. Behavioral treatments of velopharyngeal impairments. In: Bzoch K, ed. *Communicative disorders related to cleft lip and palate (4th ed.).* Austin: Pro-Ed; 1997: 529-562.

Tomoda K, Morii S, Yamashita T, Kumazawa T. Histology of human eustachian tube muscles: effect of aging. *Ann. Otol. Rhinol. Laryngol.* 1984; 93: 17-24.

NCVS Status and Progress Report - 12
June 1998, 71-77

A Histologically-Based Finite Element Model of the Soft Palate

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

Abstract

A finite element model was developed for the soft palate. Initially a static two-dimensional mid-sagittal model of the velum, the model was given physical dimensions to match that of a 10-year-old boy. Biomechanical properties of the tissues were inferred based on previous histologic studies. Velar movements were induced by the influence of three extrinisic velar muscles: the levator veli palatini, the palatoglossus, and the palatopharyngeus, which were simulated as external forces acting on the velar model. Velopharyngeal "opened" and "closed" positions were simulated, as well as a variety of intermediate steps between the two configurations. The extension of the model to a time-dependent implementation was discussed. Also, velopharyngeal closure was simulated in a manner appropriate for both high and low-vowel configurations.

Introduction

While the importance of the velum as an articulator is well established, relatively little is known about its control. One step towards more fully understanding normal and abnormal velar control involves the development of a quantitative model. Quantitative models have already been developed for other articulators, including the larynx (Titze and Talkin, 1979; Alipour and Titze, 1985; Berry et al., 1994), the tongue (Kakita and Fujimura, 1977; Kakita et al., 1985; Wilhelms-Tricarico, 1995), and the lips (Muller et al., 1984). To our knowledge, this investigation represents the first attempt to develop a quantitative model of the soft palate. Development of such a quantitative model has important implications. Combined with similar quantitative models of other structures in the vocal tract (i.e., larynx, tongue, lips), a composite model of the entire vocal tract could be developed and used to further our understanding of the normal process of speech production and its control. Development of a finite element model of velopharyngeal function will greatly enhance our ability to non-invasively evaluate the effects of palatal clefting and its repair on subsequent velopharyngeal control during speech.

The focus of the efforts described in this paper has been to capture many essential features of velar function in a static, two-dimensional finite element model. Mechanical properties of the tissues were inferred based on histologic studies of the soft palate (Kuehn and Kahane, 1990; Ettema and Kuehn, 1994). Velar movements were induced by several extrinsic velar muscles, including the levator veli palatini, the palatoglossus, and the palatopharyngeus. Specifically, the muscles were simulated as external body forces acting on the velar model. The ability of the model to simulate velar opening and closing is assessed. In addition, several muscular coordination schemes are explored to achieve desired velar configurations.

Model Development

In the development of any model, one tries to capture the most essential features of the physical system while avoiding unnecessary complexities. A central aim of the present study was the development of a model to predict realistic geometries of the soft palate during velopharyngeal opening and closing gestures. The finite element method was employed to promote this type of realism, utilizing a sufficient number of elements to model the soft palate as a continuum. Because much of our traditional understanding of velar function has been developed by examining two-dimensional sagittal images of the soft palate, especially static and cineradiographic x-ray images, it was not deemed prudent to initially focus on a full three-dimensional implementation, although this is the ultimate goal. Rather, a two-dimensional midsagittal model was developed. Because the

primary movements of the soft palate can be specified as moving in a sagittal plane during normal velopharyngeal opening and closing gestures, a two-dimensional model should be able to capture many of the essential features of velar function while avoiding unnecessary complexity.

We chose to target the model for a child, employing physical dimensions to correspond to that of a 10-year-old boy. Specifically, the velar dimensions were taken to be 30 mm in anterior-posterior length and 8.3 mm in superior-inferior thickness (Subtelny, 1957). The anterior border was taken to be a fixed attachment and all other borders were assumed to be unrestrained. The initial state of the model before the effects of gravity are applied is shown in Figure 1.

Along the anterior-posterior length, the model was split into 11 sections, and along the superior-inferior thickness into 3 sections, yielding a total of 33 elements. As pointed out by Kuehn and Kahane (1990), the tissue morphology of the velum can be viewed as having 3 natural layers in the superior-inferior thickness. Although in the histological studies of Kuehn and Kahane (1990) and Ettema and Kuehn (1994) the soft palate was sectioned into 10 distinct anterior-posterior layers, the finite element model was assigned 11 layers so that the elements of the model would be as square-like as possible in their initial configuration (i.e., element length = 30 mm/11 = 2.73 mm, and element thickness = 8.3 mm/3 = 2.77 mm). This was done to promote numerical stability and convergence in the simulations. Because the number of sections assumed in the histologic studies was arbitrary (not based on any unique morphologic or histologic features of the velum), there were no concerns in mapping the results of the histologic studies directly into the finite element model utilizing linear interpolation.

The principal result of the histologic studies was a quantification of the tissue composition of each anterior-posterior layer, averaged over 10 adult specimens. These results were summarized in a figure (Ettema and Kuehn, 1994, Figure 10), with the tissue composition reported in terms of percentages. Results were reported for 6 different tissues types, including tendinous tissue, muscular tissue, connective tissue, glandular tissue, and adipose tissue. Any remaining tissues were classified as "other."

One of the most crucial decisions in the development of the finite element model was the choice of the constitutive relation describing the mechanical properties of velar tissues. Specifically, an equation was needed to predict the amount of tissue strain or stretching that would result from applied stresses. Although such biomechanical information can not be deduced directly from the histologic studies (Ettema and Kuehn, 1994), the histologic data did provide some clues when combined with biomechanical information already available on similar biologic tissues. Table 1 shows the Youngs' moduli utilized for the different tissue types. These values were based on a variety of biomechanical studies conducted on both human and animal tissues, as indicated in the footnotes of the table. However, the Young's modulus utilized for tendinous tissues (as reported in Table 1) are about an order of magnitude lower than the values reported in the literature (see the references listed in the footnotes of the table). This was necessary to obtain sufficient bending of the tissue near the fixed anterior boundary. Future studies should reveal whether, in fact, the tendinous tissues of the velum are less stiff than tendinous tissues in other parts of the body. In addition, the Young's moduli for glandular tissue and "other" tissue were not based on values from the literature, but were hypothesized relative to the stiffnesses of the other tissues reported in Table 1.

Because each anterior-posterior layer was a mixture of several tissue types, a rule had to be formulated to

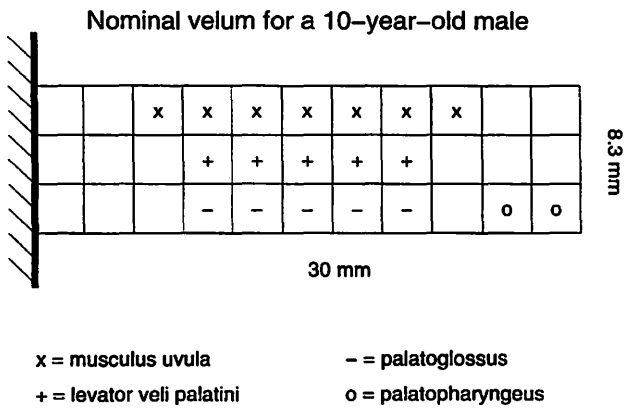


Figure 1. A schematic sagittal view of the finite element model is shown with physical dimensions to match those of a typical 10-year-old boy. Anterior is toward the left, posterior is toward the right, superior is toward the top, and inferior is toward the bottom. Selected elements are labeled as corresponding to various velopharyngeal muscles.

Table 1.
Young's Moduli for Various Tissue Types

Tissue Type	Young's Modulus (kPa)
*Tendinous Tissue	10^3
¶Muscle	10^1
†Connective Tissue	10^0
Glandular Tissue	10^{-1}
‡Adipose Tissue	10^{-2}
Other Tissue	10^{-2}

* Park (1979, p. 111); Silver (1987, p. 172).
¶ Alipour and Titze (1991).
† Park (1979, p. 112); Fung (1993, pp. 303-305).
‡ Chan and Titze (1997).

determine an effective Young's Modulus, E_{eff}, for each layer j. In the absence of a priori knowledge for generating such a rule, a linear weighting scheme was utilized:

$$E_{eff}(j) = \sum_{i=1}^{6} P_i(j) E_i \quad (1)$$

where E_i is the Young's Modulus of the i-th tissue type, and $P_i(j)$ represents the fractional percentage composition of the i-th tissue type within the j-th layer (obtained from Figure 10, Ettema and Kuehn, 1994, following a linear interpolation of the data to 11 layers rather than 10). The effective Young's modulus for the 11 anterior-posterior layers is shown in Table 2. Note that the first two layers near the fixed anterior boundary, which are composed largely of tendinous tissues, contain the highest stiffnesses in the model. Then there is a gradual exponential decay in the stiffnesses of the layers as one proceeds posteriorly.

One of limitations of this procedure is that it only differentiates tissue stiffness as a function of anterior-posterior length, but not as a function of superior-inferior thickness. In part, this is also a limitation of the histologic studies, which reported quantitative results solely as a function of anterior-posterior length. However, the studies did provide some qualitative differentiations of velar tissues as a function of superior-inferior thickness which were incorporated into the model. In particular, the locations of specific velar muscles within the model were deduced based on qualitative descriptions provided in the histologic studies (Kuehn and Kahane, 1990; Ettema and Kuehn, 1994). Although these specifications did not alter the tissue stiffnesses in the model, they did indicate the locations where forces from external muscles were exerted on the velum. It should also be noted that although it was the goal of this study to model the soft palate of a 10-year-old boy, the histologic studies were performed on adults. Although one might assume that the tissue distributions would be somewhat similar for both groups, tissue from 10-year-olds would probably have more muscle and less connective tissue than adults (Bucciante and Luria, 1934; Ettema and Kuehn, 1994).

Although the Young's modulus is a useful measure of tissue stiffness used in connection with Hooke's Law, the general applicability of Hooke's Law is known to be limited to systems in which only small tissue displacements occur. Because of the large strains experienced by velar tissues during typical gestures (e.g., in the velopharyngeal-closed position it is not uncommon for the soft palate to be stretched/strained to beyond 30% of its rest length in the velar-opened position [Simpson and Chin, 1981]), Hooke's law was avoided in favor a constitutive relationship more applicable to large strains.

Hyperelastic constitutive equations are generally considered to be the only expressions which can adequately describe large-strain deformations. Such relations involve the use of a strain energy function, W. Of all materials undergoing large strains, rubber is undoubtedly one of the most investigated and understood materials. A common strain energy function for rubber that can be directly related to the constants associated with Hooke's law is the neo-Hookean Mooney-Rivlin equation:

$$W = C_{10}(I_1 - 3) + \frac{1}{D_1}(J_{el} - 1)^2 \quad (2)$$

where I_1 and J_{el} are strain invariants (constants) related to the strain energy function W. C_{10} and D_1 are biomechanical constants which are related to the Young's modulus, E (a standard measure of tissue stiffness, values of which are reported in Table 2), and the Poisson's ratio ν (a standard measure of tissue compressibility) as follows (Saada, 1974; Abaqus User's Manual, 1994):

$$C_{10} = \frac{E}{4(1+\nu)} \quad (3)$$

$$D_1 = \frac{6(1-2\nu)}{E} \quad (4)$$

The Poisson's ratio ν was assigned a value of 0.4995 to represent a nearly incompressible tissue (0.5 corresponds to the condition of absolute incompressibility). This assumption of near-incompressible tissue is known to be valid because velar tissues are composed mostly of water.

Table 2.
Young's Moduli for each Anterior-Posterior Layer of the Model

Longitudinal Section No.	Young's Modulus (kPa)
1	100.64
2	101.15
3	61.83
4	24.36
5	6.63
6	2.50
7	2.11
8	1.69
9	1.27
10	0.76
11	0.51

Results and Discussion

Velopharyngeal Opening

To assess the ability of the finite element model to simulate velopharyngeal function, velopharyngeal opening and closing gestures are examined and discussed. To simulate velopharyngeal opening, a gravitational force was applied. As shown in Figure 1, the anterior portion of the velum (left side of the figure) is firmly attached to the hard palate. Otherwise, the velum is free to move. Under the influence of gravity, the posterior end of the velum sags, as shown in Figure 2. The posterior pharyngeal wall was positioned to produce values of velopharyngeal opening that would correspond to those reported in the literature (approximately 75% of the velar length when in the "opened" position, see Subtelny, 1957). The slope of the wall is 30° off the vertical.

The influence of other muscles might have been included, in particular the influence of the palatoglossus muscle. Contraction of the palatoglossus has been shown to facilitate velopharyngeal opening in some phonetic contexts (Sitzmann and Moon, 1998). However, previous studies have also shown that muscular activation of the palatoglossus might be associated with closing gestures of the velopharyngeal port, depending on tongue position (Moon et al., 1994). A brief investigation of the influence of the palatoglossus on velar positioning during speech attempts will be presented later in the paper. However, at this point, we note that a reasonable resting velar configuration was obtained solely on the basis of the elastic properties of the tissues and the force of gravity. In this lowered position, possible restraint of the soft palate by the tongue was not considered.

With respect to technical issues, it was found that the assumption of incompressible tissue necessitated the use of second-order quadratic elements. In contrast to the smooth velar curvature shown in Figure 2, finite elements based on linear interpolation yielded a rugged velar shape.

Velopharyngeal Closing

Velopharyngeal closure is known to be positively correlated with activation of the levator muscle (Moon et al., 1994). For this reason, the dominant muscle used to produce velopharyngeal closure was the levator muscle, which was simulated as an external body force acting over the elements indicated in Figure 1. As a body force (in contrast to a point force), its influence is equally distributed across the entire volume of the indicated elements. The levator force pulled the elements in a posterior-superior direction, oriented 30° above the horizontal. To obtain velopharyngeal closure as shown in Figure 3, a total force of approximately 30 g was exerted by the levator muscle. In Figure 3, the mid-line velar length is approximately 41 mm, which represents a stretching of about 20% beyond the length of the velopharyngeal "open" position. This is well within normal ranges (Pruzansky and Mason, 1969; Simpson and Chin, 1981).

The influence of the palatopharyngeus muscle, a muscle of secondary importance in velar closure, was also included in the generation of the velar configuration shown in Figure 3. The palatopharyngeus muscle force was directed vertically downward on the two elements indicated in Figure 1. For the condition of Figure 3, the palatopharyngeus exerted a total force of approximately 3 g, one-tenth of the force exerted by the levator muscle. Although the activation of the palatopharyngeus is not necessarily posi-

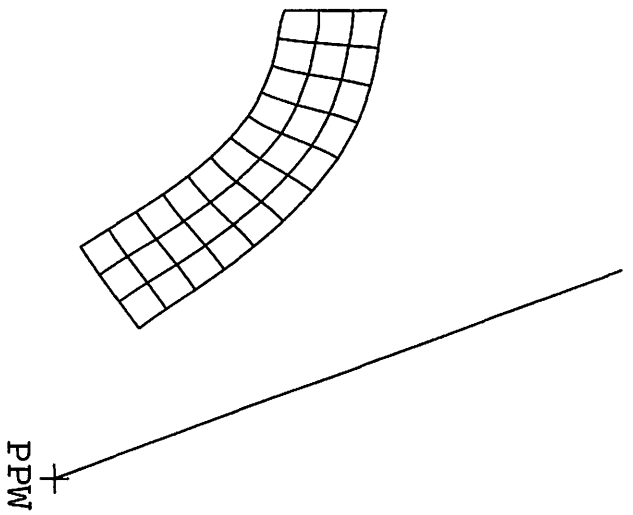


Figure 2. Deformation of the model under the influence of gravity (the velopharyngeal "open" position). The posterior pharyngeal is also displayed, and is labeled as "PPW".

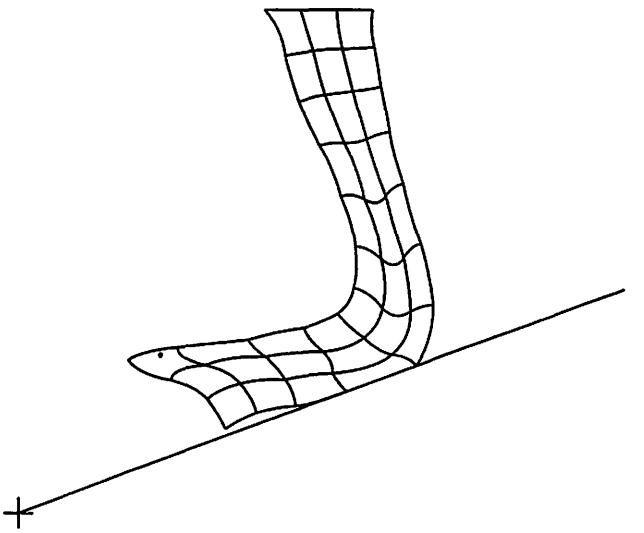


Figure 3. Velopharyngeal closure configuration obtained by employing the combined actions of the levator veli palatini and the palatopharyngeus muscles.

tively correlated with the activation of the levator muscle (Moon et al., 1994), in all the subplots of Figure 4, the magnitude of the palatopharyngeus is set at one-tenth of that of the levator muscle.

The influence of the palatoglossus muscle is also critical to velar positioning. However, because the influence of the palatoglossus may be vowel-dependent (as mentioned previously), its influence is not considered initially, but will be treated later in the paper.

Figure 4 shows a sequence of plots corresponding to increasing levels of levator force (Figure 4a corresponds to Figure 2 and Figure 4f corresponds to Figure 3). Although the velar/posterior pharyngeal wall distance decreases in roughly a linear fashion from one subplot to the next, the levator force has been increased exponentially. This relationship is illustrated in Figure 5, which shows a plot of velopharyngeal closure as a function of the levator force. The measure of velopharyngeal closure displayed in the figure is the closest anterior-posterior distance (in mm) between the velum and the posterior pharyngeal wall. The five data points in Figure 5 correspond to the five subplots of Figure 4. Figures 4e and 4f are consistent with the description provided by Kuehn (1976) indicating that the soft palate may continue its upward excursion along the posterior pharyngeal wall following its initial contact with the wall.

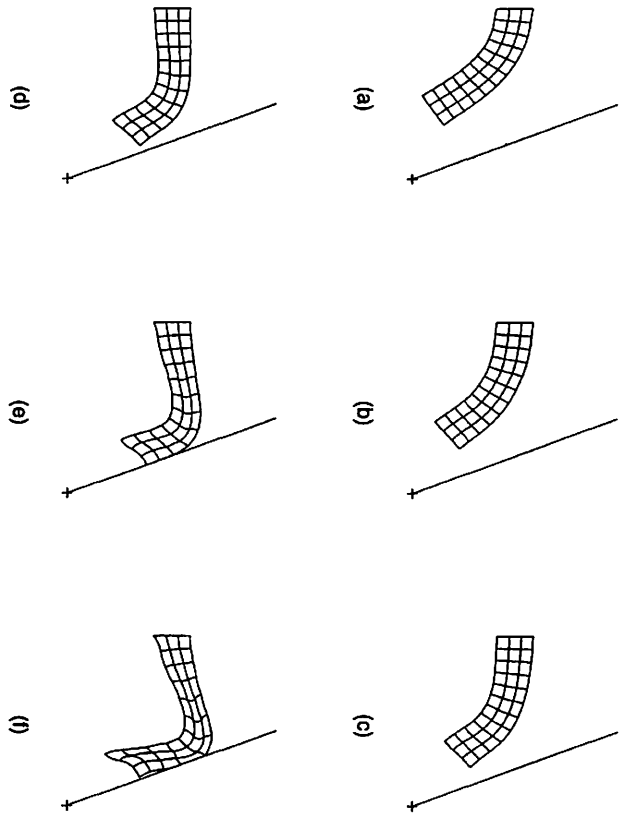


Figure 4. A sequence of plots from (a) the "opened" velopharyngeal position to (f) the "closed" velopharyngeal position.

Although only a static model is presented here, the sequence of plots in Figure 4 is illustrative of the next step to be implemented in development of the finite element model: that of time-dependence. Although a time-dependent model would include the effects of inertia, this sequence of static plots provides an initial indication that the extension of the model to the time domain should be feasible.

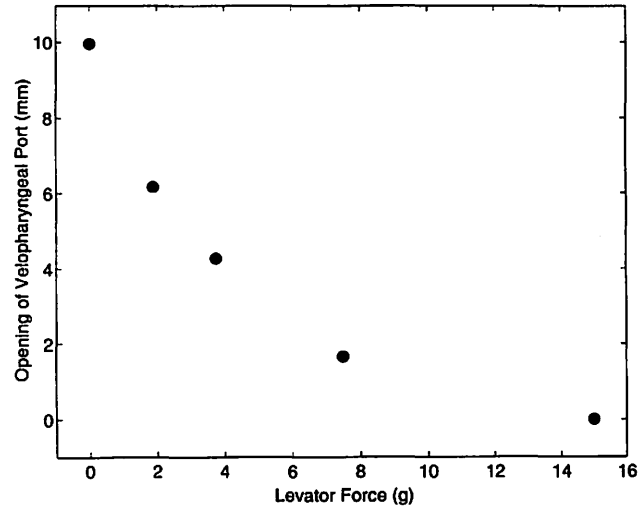


Figure 5. Velopharyngeal "opening" (the closest distance [in mm] between the velum and the posterior pharyngeal wall) expressed a function of the force exerted by the levator veli palatini muscle.

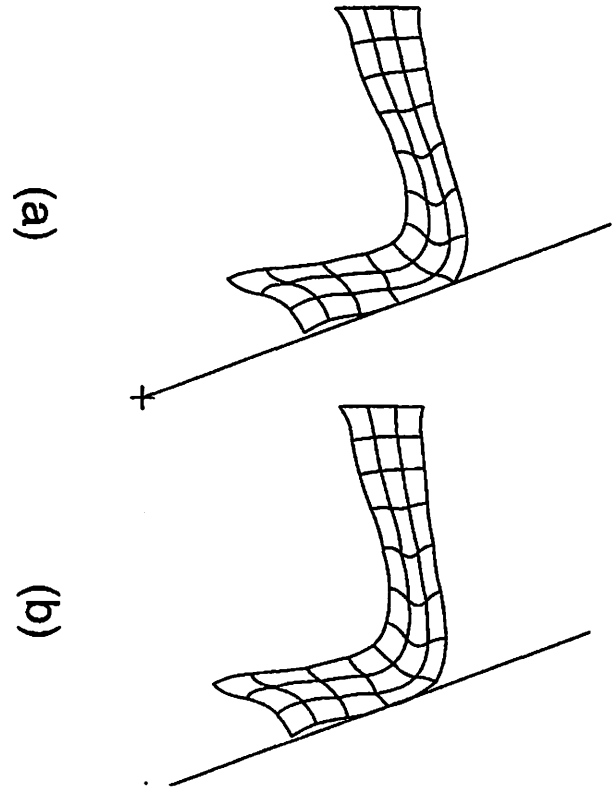


Figure 6. Velopharyngeal closure for (a) a high vowel and (b) a low vowel.

Influence of the Palatoglossus Muscle

Figure 6 shows two velar configurations, (a) one corresponding to velopharyngeal closure with no force from the palatoglossus, and (b) with a force of 4 g from the palatoglossus. As shown in Figure 1, the force of the palatoglossus is exerted on a few of the elements in the inferior 1/3 of the velum. The direction of the force is 30° left of the downward vertical. Contrasting the two configurations, (a) is reminiscent of velopharyngeal closure for a high vowel, and (b) is reminiscent of velopharyngeal closure for a high vowel, as may be observed in lateral still x-rays. That is, high vowels are associated with a more elevated velum than low vowels (Moll, 1962).

In future developments of the model, we will attempt to model this effect more directly by making the palatoglossus muscle, itself, an intrinsic part of the finite element model. The muscle will have its own elastic properties, and increased muscular activations will result in increased stiffnesses. The tongue attachment of the palatoglossus will be assigned specific locations that are vowel-dependent. However, note that it is already possible to begin examination of the muscular coordination schemes used to achieve desired velar configurations, even when the muscles are modeled as external forces.

Conclusions

Using a static two-dimensional finite element model of the soft palate, realistic velar shapes were obtained for both velopharyngeal opening and closure. Estimates of mechanical properties of velar tissues were made based on previous histologic studies. It was shown that a reasonable shape for velopharyngeal opening could be obtained solely on the basis of the elastic tissue properties of the velum and the force of gravity. It was also shown that plausible configurations for velopharyngeal closure (for both high and low vowels) could be obtained by including additional forces to simulate the influence of the levator veli palatini, the palatoglossus, and the palatopharyngeus muscles. Future studies will expand the present finite element simulation to include the muscles as an intrinsic component of the biomechanical model, rather than as external forces acting upon the velum Additionally, a time-dependent implementation of the model will be introduced which includes the influence of inertial forces. With these enhancements, more in-depth muscular coordination schemes will be explored to simulate velar function. In addition, experimental data from the laboratory will be compared with predictions from the model, especially with regard to velopharyngeal closure forces.

Acknowledgments

This work was supported by Grant no. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. The authors also thank the ICAEN computing center at the University of Iowa for generously allocating computer time and use of ABAQUS, a commercial finite element software package.

References

Abaqus/Standard User's Manual. Pawtucket RI: Hibbitt, Karlsson & Sorensen, Inc; 1994; p. 8.9.23-1.

Alipour F, Titze, IR. Simulation of particle trajectories of vocal fold tissue during phonation. In: IR Titze, RC Scherer, Eds. Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control. Denver CO: Denver Center for the Performing Arts; 1985.

Alipour, F and Titze, IR. Elastic models of vocal fold tissues. J Acoust Soc Am 1991; 90:1326-1331.

Berry DA, Herzel H, Titze IR, Krischer K. Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. J Acoust Soc Am 1994;95:3595-3604.

Bucciante L, Luria S. Transformazioni nella struttura dei muscoli voluntari dell'uoma nella senescenza. Archivo Italiano di Anatomica e di Embriologia 1934;33:110-187.

Chan RW, Titze IR. Dynamic shear modulus of vocal fold tissues and phonosurgical biomaterials. J Acoust Soc Am 1997;101:3179.

Daly, CH. The Biomedical Characteristics of Human Skin. Scotland: University of Strathclyde; 1966.

Ettema SL, Kuehn DP. A quantitative histologic study of the normal human adult soft palate. J Speech Hear Res 1994;37:303-313.

Fung YC. Biomechanics: Mechanical Properties of Living Tissues. New York: Springer-Verlag; 1993.

Kakita Y, Fujimura O. Computational model of the tongue: a revised version. J Acoust Soc Am 1977;62:S15(A).

Kakita Y, Fujimura O, Honda, K. Computation of mapping from muscular contraction patterns to format patterns in vowel space. In: V Fromkin, Ed. Phonetic Linguistics: Essays in Honor of Peter Ladefoged. New York: Academic Press; 1985.

Kuehn, DP. A cineradiographic investigation of velar movement variables in two normals. Cleft Palate J 1976;13:88-103.

Kuehn DP, Folkins JW, Linville RN. An electromyographic study of the musculus uvulae. Cleft Palate J 1988;25:348-355.

Kuehn DP, Kahane, JC. Histologic study of the normal adult human soft palate. Cleft Palate J 1990;27:26-34.

Moon J, Smith A, Folkins J, Lemke J, Gartlan M. Coordination of velopharyngeal muscle activity during positioning of the soft palate. Cleft Palate-Craniofacial J 1994;31:45-55.

Moll K. Velopharyngeal closure in vowels. J Speech Hear Res 1962;5:30-37.

Muller E, Milenkovic P, McLeod. Perioral tissue mechanics during speech production. In: C DeLisa, J Eisenfeld, Eds. Proceedings of the Second IMAC International Symposium on Biomedical Systems Modelling. Amsterdam: North-Holland; 1984.

Park JB. Biomaterials: An Introduction. New York: Plenum Press; 1979.

Pruzansky S, Mason RM. The "stretch factor" in soft palate function. J Dent Res 1969;48:972.

Saada AS. Elasticity: Theory and Applications. New York: Pergamon Press, Inc; 1974; p. 203.

Silver FH. Biological Materials: Structure, Mechanical Properties, and Modeling of Soft Tissue. New York: Springer-Verlag; 1987.

Simpson RK, Chin L. Velar stretch as a function of task, Cleft Palate J 1981;18:1-9.

Sitzmann B, Moon J. Contributions of levator veli palatini and palatoglossus muscles to velar lowering. Paper presented at annual meeting of American Cleft Palate-Craniofacial Association, Baltimore, MD; 1998.

Subtelny JD. A cephalometric study of the growth of the soft palate. Plastic Reconstruct Surg 1957;19:49-62.

Titze IR, Talkin DT. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. J Acoust Soc Am 1979;66:60-74.

Wilhelms-Tricarico R. Physiological modeling of speech production: methods for modeling soft-tissue articulators. J Acoust Soc Am 1995;97:3085-3098.

NCVS Status and Progress Report - 12
June 1998, 79-85

A Study of Age and Gender Related Elastin Distribution Changes in Human Vocal Folds

Thomas H. Hammond, B.S.
The University of Utah Medical Center
Steven D. Gray, M.D.
John Butler, M.D.
Division of Otolaryngology/Head Neck Surgery, The University of Utah Medical Center
Ruixia Zhou, Ph.D.
Elizabeth Hammond, M.D.
Department of Pathology, L.D.S. Hospital

Abstract

The composition of the lamina propria in human vocal folds has been shown to affect vocal performance. Elastin plays a significant role in the biomechanical effects of the lamina propria. We obtained 19 larynges from the State Medical Examiner from subjects whose cause of death was unrelated to the trachea and laryngeal regions. The sample contained men and women in infant, adult, and geriatric age groups. We stained the vocal folds for elastin using Verhoeff's elastic tissue stain (EVG) and studied them using an image analysis system configured for light microscopy. Distributions of elastin were measured from superficial to deep within the lamina propria (from epithelium to vocalis muscle). These elastin distributions were then compared with statistical software.

The data showed that there was a increase in elastin content from infant through geriatric stages. No gender related differences were found. Infant folds had about 23% of the elastin found in adults and geriatric patients had about 879% of the elastin found in adults. Both of these results were statistically significant ($p<<0.05$). The distributions were consistent with previous observations that the lamina propria is a layered strucutre with most of the elastin present in the intermediate layer. This layer was larger in geriatric patients than in adult and pediatric patients. We observed that the fiber diameter appeared to be larger in geriatric patients (this observation is currently being verified with electron microscopy) while smaller, spiraled fibers appeared in pediatric patients.

Introduction

The composition of the lamina propria in human vocal folds has been shown to play an important role in voice production[1]. Since the distribution of cellular and extracellular material in the lamina propria is potentially capable of affecting this performance, it is important to gain an understanding of the normal balance of these proteins so that variations from this normal state can be diagnosed. Conditions such as nodules and polyps that affect human vocal folds are known to occur in the lamina propria.

The lamina propria is an area between the epithelium and vocalis muscle. It is made up of extracellular proteins, water, and a few cells. It has been described as a trilaminar structure with superficial, middle, and deep layers. The superficial layer is recognized by the relative absence of fibrous proteins, the middle or intermediate layer is characterized by an abundance of elastin, and the deep layer is identified by its collagen content[2].

Each of these layers contribute to the overall mechanics of the lamina propria. These mechanics are best described by the cover-body theory of phonation[3]. In this theory, the cover (which consists of the epithelium, superficial and most of the intermediate layer) vibrates on a relatively stationary body (which includes the remainder of the intermediate layer, the deep layer and the vocalis muscle).

Studies from our group have characterized constituents of this important region. Hammond et. al. described gender related differences in hyaluronic acid, a glycosaminoglycan (GAG) which may contribute to a shock

absorbing affect[4]. It was found that males have significantly more hyaluronic acid than females. This study also examined the histologic distribution and ultrastructural appearance of elastin forms in the lamina propria but no gender or age related comparisons were made. The purpose of the current study is to examine possible age and gender relationships in elastin content and distribution.

Elastin is a unique protein that is responsible for the resiliency of many tissues. Elastin can be stretched to 5 times its normal length without loss of resiliency in vitro[5]. It is limited, however in its elastic properties by its association with collagen fibers[6]. In vocal folds, it has been shown to stretch to roughly two times its length[7]. This elastic character is important as it has been suggested that malfunction of the elastic properties of vocal folds could cause dysynchronous vibrations1. It is not known what exactly gives elastic fibers their resiliency, but it has been suggested that a type of molecular helical springing action could be responsible (this is not related to the coiled histologic appearance of infant elastin mentioned later)[5].

Materials and Methods

Tissue Preparation

Larynges were obtained from the State Medical Examiner within 24 hours of death. No information about the subjects were provided except for age, gender, and race. The subjects were presumed to come from an urban environment because of the geographic area served by the Medical Examiner. Larynges were examined grossly for injury. Cases with suspected laryngeal injury were excluded by the Medical Examiner. 11 male and 8 female larynges were obtained. Infant larynges ranged from 1 to 5 months old, adult larynges ranged from 30 to 39 years old, and geriatric larynges ranged from 65 to 82 years old. All subjects were of Caucasian descent although this was not a selection factor.

One fold from each larynx was extracted en bloc while the other was placed in situ in a -86C freezer for possible future use. The extracted fold was placed in Carson's Fixative for a minimum of 24 hours[8].

Histological Staining

Folds were processed routinely using alcoholic dehydration and paraffin embedment. Tissues were embedded to obtain mid membranous coronal sections of the folds (perpendicular to the vocal ligament). Paraffin blocks were cut into four micron thick sections. Sections were stained with Verhoeff's elastin tissue stain (EVG) which uses ferric chloride and iodine to stain elastin fibers black[9].

Image Analysis

The slides were analyzed using an image analysis system specially configured for histological and cytological studies. The system is run on a Dell Optiplex XMP 5100 Pentium computer (Dell Computers, Dallas, Texas, U.S.A.) with a 70 megabyte hard disk. A Microimager 1400 high resolution (1280 X 1024 X 10 bit) image scanner (XILLIX Tech Corp., Vancouver, B.C. Canada) is mounted on an Olympus BH2 microscope with a DPlanApo 20UV 20X lens and BHT 5X Photo eyepiece (Olympus Corp., Lake Success, New York, U.S.A.). The amount of staining was measured using Optimas Image Analysis Software (Version 4.1)(Bioscan Corp., Edmond, Washington, U.S.A.) developed from program macros by Dr. Ruixia Zhou. This program uses wavelength filters to evaluate tissue structures[10]. Illumination was controlled with a DC Lamp Power Supply (Olympus Corp, Lake Success, New York, U.S.A.) which is regulated through the image analysis software.

Optical interference filters are chosen to restrict certain wavelengths of light to maximize contrast within the specimen. For the EVG stain, a 630nm filter was used which accentuates the black elastin fibers while de-emphasizing the pink collagen fibers. Proper condenser settings also enhanced contrast (settings were checked frequently throughout imaging process). The general region of the specimen was found using the eyepiece of the microscope. The exact location was determined by viewing the specimen directly through the Microimager 1400 on a high resolution monitor.

The image analysis software uses a gray scale measurement to determine the amount of staining. After a properly acquired image is obtained, the computer measures the area of the image that is darker than a preset threshold gray scale value. The gray scale value is between 0 and 256 with 0 being black and 256 being white. If the preset threshold is set to 149 to measure elastin (which it was in this study), then the area of the image that has a value of 149 or lower will be measured as elastin. It is this gray scale measurement that makes the filters and contrast so important for an accurate measurement. If the average value for elastin is 149 and lower while the average value for collagen is 220 and higher, the measurements will have much less error than if the average value for elastin is 180 and the average value for collagen is 190 (collagen would have a much higher probability of being measured as elastin). There is also a function in the software to manually subtract unwanted measured areas. For example, cells often stain opaquely. These cells must be manually deleted from the measurement or they will be counted as elastin fibers. This manual deletion was done very carefully because of the variation that it can cause. Knowledge of the histologic appearance is crucial to optimal performance of this function.

The amount of elastin is distributed in a layered pattern throughout the lamina propria. Therefore, to obtain an accurate representation of the amount of elastin as well as to measure this pattern of distribution, it is necessary to measure an entire strip or swath of lamina propria from epithelial surface to vocalis muscle. This was done by taking multiple adjoining fields beginning with the superficial

most field and ending with the fields adjacent to muscular tissue. Because the border of the lamina propria adjacent to the muscle was composed of a mixture of muscle and extracellular matrix, it was difficult to determine a definitive deep border. This border was usually set at a point where the muscular tissue filled more than approximately 25% the field. The usual amount of muscular tissue in the deepest field was approximately 10-15%. The number of fields necessary to cross the lamina propria was variable (4 to 13 frames) depending on the thickness of the lamina propria and the angle of the section (a section cut more diagonally would appear to be thicker and use more frames than the same section cut more perpendicularly). All of these fields were used in determining the distribution and total amount of elastin present in a given specimen. Multiple strips were taken from each specimen to ensure that a representative measurement has been made.

Variation Strategy

Variation was the principal obstacle in making these measurements because of our limited sample size (we usually had 3 specimens per category). To reduce this variation as much as possible, several precautions were taken. First, all of the measurements were made by a single observer. Single observer reproducibility has been shown to be higher than dual observer reproducibility (the theory being that if errors are present, the same errors will be present in all the specimens which would make comparison statistics relatively accurate)[11]. Second, all of the measurements were made on the same day at one sitting which reduced human variability (again, erroneous imaging habits are more likely to change overnight than in one day) as well as equipment variability. Third, the study was done blind and random. The slides were imaged in random order and with the observer blinded to the age and sex of the patient. Using these precautions, we were able to obtain reproducible data with minimum variation from a limited sample size.

Data Analysis

Once individual measurements were made, we converted the measurements into a comparable form. First, we combined measurements from multiple strips taken from a single specimen. After this was done comparisons between categories (infant males, geriatric males, etc.) for total elastin content were made (see Figure 1).

A Kruskal-Wallis 1-Way Anova was performed on the data for total elastin content to determine if the individual categories were different enough to be statistically significant. The Kruskal-Wallis 1-Way Anova ranks all of the samples (19) as if they were all in one large group. It then analyzes the ranks of the cases according to the categories indicated (6). Individual comparisons were done for gender within infant, adult, and geriatric groups using a similar test (Wilcoxon Rank Sum Test).

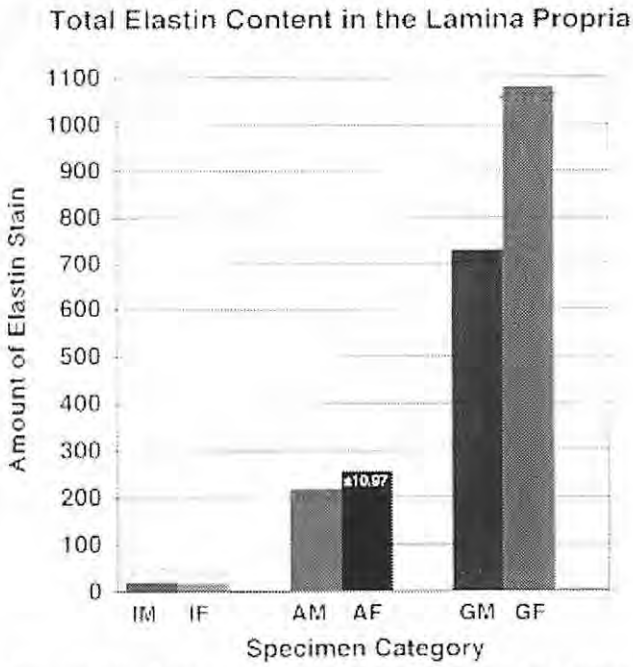


Figure 1. Total elastin content in the lamina propria. Standard deviations appear within or on top of bars. IM=infant male; IF=infant female, AM=adult male, AF=adult female, GM=geriatric male, GF=geriatric female.

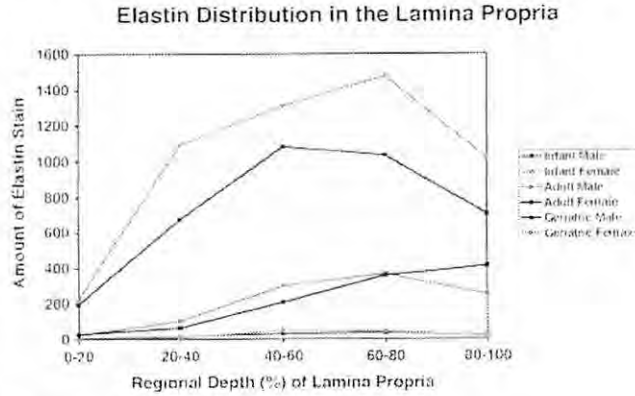


Figure 2. Elastin distribution in the lamina propria.

After making comparisons for total elastin content, we analyzed the distribution patterns for each category (see Figure 2). Because of the variation in frame number (see Image Analysis) it was necessary to normalize all of the field numbers. This normalization was done by dividing the field number by the total number of fields for that specimen. This technique was used successfully in two previous studies: Catten et. al. and Hammond et. al.[12,4]. This gave all field numbers a percentage ranking from 0 to 1 that corresponded to the relative depth across the lamina propria (0 being the most superficial and 1 being the deepest). For example: Field number 5 in a specimen with 16 total fields would have a normalized field number of 0.31 (5/16 = 0.31) meaning that its location was approximately 31% of the distance

across the lamina propria in that particular specimen. Field number 5 in a specimen with 11 total fields would have a normalized field number of 0.45 (5/11 = 0.45) meaning that it's location was approximately 45% of the distance across the lamina propria in that particular specimen.

We then arbitrarily divided the lamina propria up into 5 sections for purposes of comparison. Each section represented 20% of the lamina propria. Section 1 was the most superficial section (0-20%) while section 5 was the deepest (80-100%). We then determined the average amount of elastin staining in that particular section. In the previous example (with 16 frames), all the staining measurements from frames 4,5, and 6 (normalized numbers of 0.25, 0.31, and 0.38 respectively) were added up and divided by 3 to get the average staining per imaged field in the second 20% section of the lamina propria (20% from the epithelium to 40% from the epithelium). Frames 3 and 7 were not included because their normalized numbers were 0.19 and 0.43 respectively which lie outside of the 20%-40% range (0.20-0.40). Data was organized and analyzed using SPSS Statistical Software Version 6.0 (SPSS Inc., Chicago, Illinois, U.S.A.).

Results

We found a significant difference in the amount of elastin between all age categories (infant, adult, geriatric) and that the amount of elastin staining increases chronologically (see Figure 1). We found no significant gender difference within categories. We also made several histological observations during our study concerning elastin fiber size, orientation, and location by age range.

Infant

Infant patients had almost no measurable elastin in the lamina propria (see Figures 1,3). Cells were abundant, but the elastin present was not developed enough to stain opaquely with the EVG stain. Males and females both had similar amounts of elastin (19.8 and 19.0 respectively). Both males and females showed peak elastin distribution between 40% and 80% (see Table 1a, 1b). Although the distribution curves for infants appear normal (similar to the distribution curves of adults), we are somewhat skeptical as to their accuracy because of the difficulty in measuring such a small quantity of elastin (adult specimens had values around 500). Our observations suggest that the relative amount of total elastin present shown in the data is representative.

Adult

Adult patients had substantial elastin staining in the lamina propria (see Figures 1,4, Table 1c, 1d). Elastin distribution curves showed similar amounts of elastin (females had slightly more) as well as similar patterns in layering (see Figure 2). Both males and females had less elastin in the superficial region which increased toward deeper regions.

Geriatric

Geriatric patients had more elastin staining than either infant or adult patients in the lamina propria (see Figures 1,5, Table 1e, 1f). Females had a higher average elastin content and a smaller superficial layer. The male distribution curve showed the main increase occurring between 40-60% while in females, it occured between 20-40%

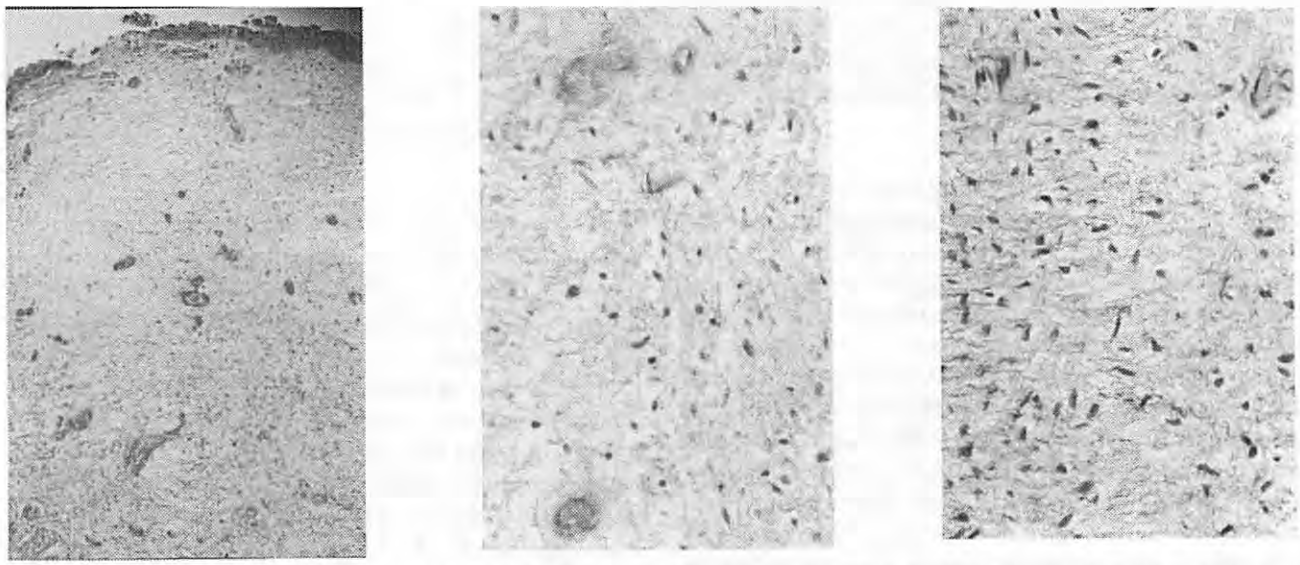


Figure 3. Infant vocal fold. Verfhoeff's elastic stain. a-left) Full photograph of lamina propria including epithelium (top) and vocalis muscle (bottom). Objective magnification=10X. b-middle) High power view of intermediate and deep lamina propria. Objective magnification=40X. c-right) High power view showing spiraled elastic fibers. Objective magnification=40X.

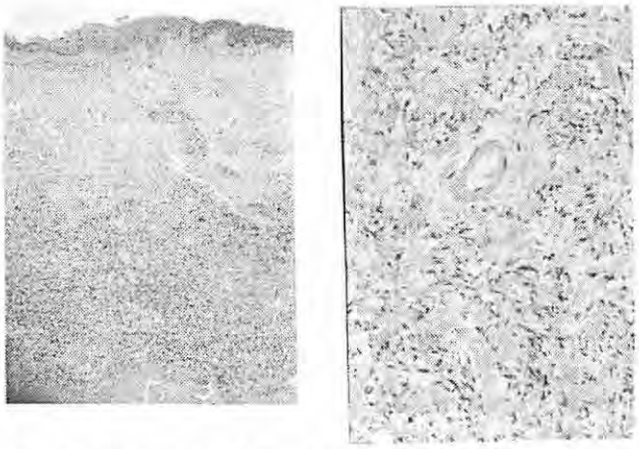


Figure 4. Adult vocal fold. Verhoeff's elastic stain. a-left) Full photograph of lamina propria including epithelium (top) and vocalis muscle (bottom). Objective magnification=10X. b-right) High power view of intermediate and deep lamina propria. Objective magnification=40X.

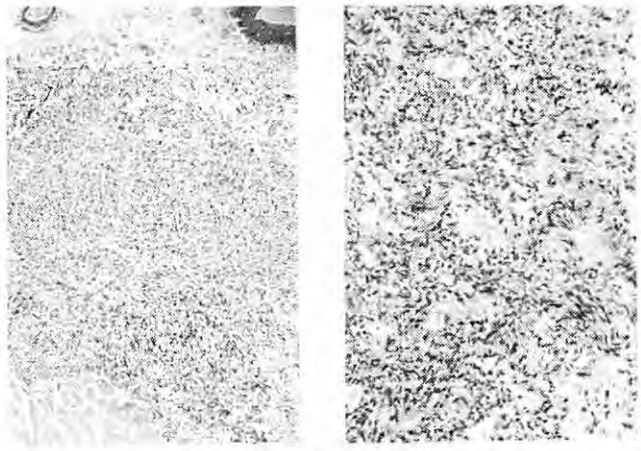


Figure 5. Geriatric vocal fold. Verhoeff's elastic stain. Compare to 3a and 4a. Note the closer proximity of the elastin towards the epithelium. a-left) Full photograph of lamina propria including epithelium (top) and vocalis muscle (bottom). Objective magnification=1-X. b-right) High power view of intermediate and deep lamina propria. Note increase in fiber

Table 1.
Age and Gender Comparison of Elastin Content in the Lamina Propria

	Elastin Content at Relative Depth (%) in the Lamina Propria				
Category	0-20	20-40	40-60	60-80	80-100
Infant Male	10.06 (10.73)*	16.20 (20.07)	31.20 (43.06	35.14 (43.51)	20.14 (26.97)
Infant Female	7.05 (4.88)	9.50 (13.44)	49.45 (69.93	41.25 (57.20)	16.70 (23.62)
Adult Male	21.73 (16.07)	101.4 (45.99)	296.67 (157.52	363.83 (215.19)	248.97 (120.85)
Adult Female	28.60 (21.08)	60.93 (8.78)	206.07 (127.11	356.03 (219.65)	409.70 (221.91)
Geriatric Male	195.33 (134.71)	670.70 (242.79)	1081.50 (361.18	1035.27 (90.30)	703.83 (114.81)
Geriatric Female	219.27 (310.35)	1091.23 (1736.98)	1306.87 (1856.19	1473.80 (1637.85)	1011.77 (818.21)

*Standard Deviations are in parenthesis

(see Figure 2). The large standard deviation seen in the geriatric females (see Table 1) was due to an extremely high elastin content of a single outlier (measurement was 2471.6) combined with a limited sample size. This outlier did, however fit the general data trends that elastin content increased with age.

Histological Observations

Two observations were made during the image acquisition of the elastin data that we feel are notable. First was a confirmation of what Hammond et. al. described as a very "abundant, tiny fibrils...observed in the superficial portion"[4]. We observed this characteristic in the majority of our specimens. Second was the difference in elastic fiber size between infant, adult, and geriatric specimens (see Figures 3,4,5). Geriatric specimens appeared to have thicker darker staining fibers than adults while infant patients appeared to have very thin coiled fibers. All patients showed a decrease in elastin staining in the area adjacent to the vocalis muscle except for adult females (histologically, adult females did have a slight decrease in elastin staining adjacent to the vocalis muscle, but the area was so small that it did not show up in the data).

Discussion

In this study we have described the amount and distribution of elastin, as represented by the EVG stain in infant, adult, and geriatric populations. We found that the amount of elastin increased with age and was not significantly different by gender The lack of significance with respect to gender may be due to small sample size.

We also found changes in the distribution patterns of elastin between age populations. This quantitative imaged data was consistent with histological evaluation of the same specimens. This study did not evaluate total thickness of the vocal fold although some comparisons of relative layer thickness of the superficial layer (how much of the total lamina propria was occupied by the superficial layer) were made.

It was evident from the overall intensity of elastin staining that the concentration of elastin progresses with maturity (see Figure 1). Middle-aged vocal folds had more staining than infant specimens and the geriatric specimens, in turn, had considerably more staining than the middle-aged adults. Even in the superficial layer, the geriatric folds showed more elastin fibers than many middle-aged folds did at their maximum concentration (see Table 1).

Although the border between the muscle and the lamina propria is ambiguous, the effect of this ambiguity on the age related differences was not significant. The reasons for this are twofold. First, all the specimens were subjected to the same random border definition. Second, the differences in elastin content by age group is large enough to achieve significance with or without the error due to border definition.

The distribution curves for elastin staining of the different age and gender categories were fairly similar (see Figure 2). The main similarity was that all the categories exhibited characteristic layers corresponding to the superficial and intermediate layer. The difference was in the relative location of those layers. Since the deep layer is usually defined by its collagen content, we did not examine gender or age related differences in this layer.

If the superficial border of the intermediate layer is defined as the location at which the elastin concentration has the largest increase, we saw that in middle-aged vocal folds, the intermediate layer began around 40-60% across the lamina propria (see Figure 2). In contrast, geriatric folds exhibited their maximum increase at a relative depth of about 20-40%. From this data we can infer that the location of the intermediate layer has been shifted superficially (in geriatric folds). This suggests that the superficial layer in geriatric specimens had been severely thinned or else infiltrated with fibrous proteins[13]. In infants, there were not enough measurable elastin fibers to define a superficial or intermediate layer.

The higher concentration of elastin and the closer position of the vocal ligament (intermediate layer) to the epithelium may explain why some geriatric folds take on appearances of sulcus vegeture or exhibit decreased mucosal wave properties. This observation poses two important questions: How does this Asuperficial shift affect biomechanical performance? What is the effect of the increase in elastin content with age?

Although there was a time that investigators in this laboratory thought the ideal material to reconstruct the superficial lamina propria would be elastin fibers, it should be noted that the superficial layer is relatively sparse in elastin fibers. Elastin fibers are most abundant in the intermediate layer where they are intimately associated with hyaluronic acid (a glycosaminoglycan) and fibromodulin (a proteoglycan). It is these three proteins (elastin, hyaluronic acid, and fibromodulin) that seem to be responsible for the biomechanical properties of the intermediate layer[14,4]. In the superficial layer, however, the biomechanical properties seem to be more dependent on the presence of water controlling proteins such as decorin and hyaluronic acid and a lack of fibrous proteins. Some immature forms of elastin have been found in the superficial layer that have less fibrillar content[4]. The precise effect of these proteins is currently under study.

Several age-related studies on vocal fold histology have been previously conducted. It has been observed that the thickness of the intermediate and superficial layers decreases with age while the thickness of the deep layer increases. It was also noted that aging caused elastic fibers in the intermediate layer to atrophy and that a decrease in elastic fiber density occurred in males[15,16]. Another study found that in the more superficial layers of the lamina propria in geriatric specimens, elastin was shown to be more resistant to degradation with elastase than elastin found in young adults and that the breakdown of elastic fibers is slowed in aged populations[2].

The findings of this study confirm the observed thinning of the superficial layer. Neither thinning of the intermediate layer nor thickening of the deep layer can be confirmed because our study did not measure collagen fiber content. Quantitation of collagen content would be necessary to define the deep layer and hence the thickness of the intermediate layer.

While this study did not include examination of elastic fiber atrophy, we did make elastin density related observations. We observed an increase in overall elastin density with age (see Figures 4b, 5b). These findings are not congruent with previous observations of a decrease in elastic fiber density[16]. The cause of this discrepancy is unclear. Differences between the studies include methodology (our study used quantitative imaging instead of qualitative observation) and sample size (our sample size was smaller).

It should be noted that an increase in elastic fiber density does not necessarily imply an increase in vocal fold elasticity. Some observations about elastin suggest that the quantity of histologically observed elastin may not correlate with functionality. Kahane observed that breakdown of elastic fibers in human vocal folds often resulted in histological Ablotch-like formations which, while contributing to the overall elastic content of the vocal fold, may not function like elastic fibers observed in middle-aged adults[16]. Similar observations have been made outside of vocal fold physiology. In Marfan's syndrome, for example, there is a molecular defect in fibrillar proteins associated with elastin, but not in elastin itself. The mechanical properties of the elastic fibers are disrupted without changing the elastin and EVG staining of the fibers. In solar elastosis of the skin,

defective elastic fibers show intense EVG staining, despite their abnormal function and ultrastructural appearance[17]. A final well known example is the changes in elastic arteries with aging. Elasticity of arteries decreases with age because of extensive cross-linking and enhanced calcium binding ability; however no histologic change in EVG staining is apparent.

Why does vocal fold elastin increase with age? One theory used to explain a similar phenomenon in aging skin suggests that decreases in elastin turnover rates may be the cause. With a decrease in turnover, extensive oxidative cross linking takes place that increases elastin content while decreasing elasticity[18,19]. This theory is congruent with one of Hirano's observations that elastin in geriatric specimens is more branched and networked together[15].

The infant patients used in this study were under 1 year old and represent a distinctive subset of patients. Our results confirm those of Hirano who found that fibrillar proteins are sparse in early postnatal stages[2]. The coiled appearance of the elastic fibers that are present in infant specimens is readily apparent histologically and raises interesting questions about elastin fiber development (see Figure 3c). Infant elastin composition has not been evaluated ultrastructurally which will be a goal of future studies.

In summary, this preliminary study showed no change in elastin fibers with respect to gender, but did show a statistically significant difference between the elastin fiber content with respect to three age groups. The geriatric population, in particular, exhibited a marked increase in elastin staining with an accompanying shift of the intermediate layer superficially. This superficial shift significantly diminished the superficial layer. We are currently studying the biomechanical effects of this shift.

Acknowledgments

This work was supported by NIH Grant #P60 DC00976 to that National Center for Voice and Speech from the National Institute on Deafness and other Communicative Disorders (NIDCD).

References

1. Kahane JC. Connective tissue changes in the larynx and their effects on voice. J of Voice 1987;1(1):27-30.

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

3. Hirano M, Sato K. Histological color atlas of the human larynx. San Diego, CA: Singular Publishing Group, 1993:44-45.

4. Hammond TH, Zhou R, Hammond E, Pawlak A, Gray SD. The intermediate layer: A morphological study of the elastin and hyaluronic acid constituents of the normal human vocal fold. J of Voice 1997;11(1): 59-66.

5. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. Cell junctions, cell adhesions, and the extracellular matrix. Molecular Biology of the Cell. 3rd Ed. New York and London: Garland Publishing, Inc., 1994:984-985.

6. Stryer L. Connective tissue proteins. Biochemistry. 3rd Ed. New York: W.H. Freeman & Company, 1988:274-275.

7. Gray SD, Hirano M, Sato K. Molecular and cellular structure of vocal fold tissue. In: Titze IR, eds. Vocal Fold Physiology. San Diego, CA: Singular Publishing Group, Inc., 1993:1-35.

8. Carson FL, Martin JH, Lynn JA. Formalin fixation for electron microscopy, a reevaluation. Am J Clin Pathol 1973;59:365-367.

9. Carson FL. Histotechnology: A Self Instructional Text. Chicago: ASCP Press, Inc., 1990:140-141.

10. Zhou R, Parker DL, Hammond EH. Quantitative peroxidase-antiperioxidase complex-substrate mass determination in tissue sections by a dual wavelength method. Analyt Quant Cytol Histol 1992;14(2):73-80.

11. Zhou R, Hammond TH, Hammond EH. Dual wavelength image analysis in quantitative immunocytochemical staining of prostate specific acid phosphatetase in tissue sections: Precision and reproducibility. Anal Quant Cyt Hist, accepted for publication.

12. Catten M, Hammond TH, Zhou R, Hammond EH, Gray SD. An analysis of cellular location and concentration in vocal fold lamina propria. Oto HNS J, accepted for publication.

13. Honjo I, Isshiki N. Laryngoscopic and voice characteristics of aged persons. Arch Oto 1980;106:149-150.

14. Pawlak A, Hammond TH, Hammond E, Gray SD: Immunocytochemical study of proteoglycans in vocal folds. Ann Otol Rhinol Laryngol 1996;105(1):6-11.

15. Hirano M, Kurita S, Sakaguchi S. Aging of the vibratory tissue of human vocal folds. Acta Oto 1989;107:428-33.

16. Kahane JC. A survey of age-related changes in the connective tissues of the human adult larynx. J Voice 1987;1(1)27-30.

17. Majno G, Joris I. Extracellular pathology. In: Cells Tissues and Disease. Blackwell Scientific Publishers, Majno, Joris (Eds.) Cambridge, MA, 256-61, 1996.

18. Jarrett A. Aging of the dermis. In: Jarrett A, eds. They Physiology and Pathophysiology of the Skin, 3rd ed. London: Academic Press, Inc., 1974, 911.

19. Tsuji T, Hamada T. Age-related changes to human dermal elastic fibres. Brit J Derm 1981;105:57-63.

NCVS Status and Progress Report - 12
June 1998, 87-94

Just Noticeable Differences for Glottal Flow Waveform Characteristics

Ronald C. Scherer, Ph.D.
Department of Communication Disorders, Bowling Green State University
Kathryn Hoberg Arehart
Communication Disorders and Speech Science Department, The University of Colorado-Boulder
Chwen Geng Guo, M.S.
Trus Joist MacMillan, Greenwood Village, Colorado
Claudio F. Milstein
Department of Speech and Hearing Sciences, The University of Arizona
Yoshiyuki Horii, Ph.D.
Communication Disorders and Speech Science Department, The University of Colorado-Boulder

Abstract

This study was motivated primarily by the need to establish the correspondences between auditory abilities and laryngeal function. Just noticeable differences (JNDs) were obtained for the open quotient and speed quotient of the glottal flow waveform. The quotients were synthesized for both the glottal flow alone and for the output pressure signal after the glottal flow signal was applied to the synthesis vocal tract for the vowel /ɑ/. Six adult men and five adult women, all singing teachers, participated as listeners. An adaptive auditory listening procedure was used to estimate JNDs for the four types of stimuli. The group average JND values were as follows. For the standard open quotient value of .6000, JND=0.0264 (sd=.010) for the glottal flow and JND=0.0344 (sd=.020) for the output pressure. For the open quotient, there was no statistically significant difference between genders or between the types of signals. For the standard speed quotient value of 2.000, JND=0.154 (sd=.043) for the glottal flow and JND=0.319 (sd=.167) for the output pressure. For the speed quotient, there was no statistically significant difference between genders, but the difference between types of stimulus (glottal flow versus output pressure) was significant ($p < .006$). The variance among the JND values was significantly larger for the output pressure stimuli compared to the glottal flow stimuli for both the open quotient and the speed quotient.

Introduction

The basic perceptual attributes of voice typically are listed as pitch, loudness, quality, and duration (1). Perceptual changes of these attributes are brought about by physical changes in the glottal flow source (for a constant vocal tract configuration). Although there have been studies of abnormal phonatory behavior dealing with small changes in synthesized signals and corresponding judgments of quality (2,3,4,5,6,7,8), apparently there have been no published studies dealing with just noticeable differences (JNDs) in the perception of normal glottal source characteristics. Such studies would determine the relationship between minimal auditory discrimination and the corresponding small changes in glottal source characteristics, leading not only to information about auditory abilities but also to inquiry into minimal vocal production needs. JND studies of glottal flow characteristics are also important in determining the necessary resolution of glottal source characteristics for highly natural speech synthesis. In addition, discrimination studies are important for the development of auditory training tapes for speech language pathologists and voice instructors for speech and singing. Creating standards for auditory abilities in discriminating sound qualities may be part of future professional training, and discrimination studies will help in that endeavor.

The purpose of this study was to determine the JNDs for the open quotient (OQ) and the speed quotient

(SQ) of the glottal volume velocity waveform. These two characteristics of the glottal flow are important to the spectrum and intensity of the voicing source (9,10,11). The OQ and SQ also are controllable to some extent by the speaker through changes in glottal adduction, vocal fold length, subglottal pressure, and vocal tract inductance (1).

Methodology

Subjects

The subjects for this study were singing teachers. Singing teachers were chosen because they spend much of their professional time listening for small changes in the voices of their students while attempting to help create small improvements in their voice quality. Six men (45 to 56 years of age) and five women (31 to 53 years of age) were subjects. The men had taught singing for 10 to 30 years, and the women for 3 to 25 years. The approximate total number of singing students, individually or in groups, over the past five years ranged from 20 to 1500 for the men, and 90 to 1000 for the women. The subjects were therefore considered to be experienced in singing instruction. The subjects were paid. Subjects passed a hearing screening in both ears at 20 dB HL (12) from 250-2000 Hz. All but three of the subjects passed the screening at 4000 Hz (see Endnote 1).

Stimuli Presented to the Subjects

The OQ and the SQ were the two characteristics of the glottal flow that were experimentally varied. The OQ is the time the flow exits the glottis during a phonatory cycle, divided by the phonatory period. The SQ is the time it takes for the flow to rise from the baseline and reach its peak, divided by the time to then reduce from the peak back to the baseline. The values of these parameters vary during speech. The OQ can vary from 1.0 during breathy voicing to about .39 for pressed voice (13). The range for the OQ for normal speakers is approximately 0.45 to 0.95 (14). The range of the SQ for normal speakers is approximately 0.50 to 2.88 (14). Consistent with these ranges, the standard values for the OQ and the SQ were chosen to be 0.6000 and 2.000, respectively.

The stimuli were obtained from a flow source model using the synthesis program SPEAK (9). The synthesized glottal flow signal is a function of the OQ, SQ, fundamental frequency, and the flow amplitude. The fundamental frequency was set at 110.39 Hz and the flow amplitude to a nominal value (the signal amplitude was held constant across all stimuli). The synthesis model used here does not include special treatment of the glottal flow pulse at the corner near glottal closure (cf. 15). The spectral characteristics of the flow pulse are affected if the corner is independently altered, a sharper corner giving rise to a flatter spectrum. In the SPEAK model used here, the corner is dependent upon both the OQ and SQ values.

Two types of stimuli were created using SPEAK. One was the glottal flow itself. The glottal flow signal was generated with a 1.0 second duration and a 50 msec Hamming tapering at the onset and offset. These files had a "buzz-like" quality to them because the flow signal was not applied to the synthetic vocal tract. The second type of stimulus was the acoustic pressure at the "lips" of the reflection-type articulatory synthesizer after the glottal flow, just described, was provided as the input to the vocal tract of the synthesizer. The vocal tract shape was specified for the /ɑ/ vowel. Thus, there were four specific stimulus categories: (1) the OQ for the glottal flow, presented to the subjects as the glottal flow signal; (2) the SQ for the glottal flow, presented to the subjects as the glottal flow signal; (3) the OQ for the glottal flow, presented to the subjects as the output pressure signal after passing through the /ɑ/-configured vocal tract; and (4) the SQ for the glottal flow, presented to the subjects as the output pressure signal after passing through the /ɑ/-configured vocal tract.

Psychophysical Method

JNDs for the glottal flow waveforms were estimated with a three-interval forced choice two-down one-up adaptive procedure (16). Each trial consisted of three 1.0 second observation (listening) intervals with an interstimulus (silent) interval of 0.5 seconds. Two of the intervals contained the standard stimulus. The third interval contained the comparison stimulus. On each trial, the order of presentation of the standard and comparison stimuli was randomized among the three intervals. The listener's task was to select the stimulus that sounded different from the other two stimuli. The subjects were given the option of viewing a monitor screen which lit up in one of three distinct locations as each of the three stimuli occurred. After responding, the listener received visual feedback regarding the correct response. Some subjects indicated that their best strategy was not to look at the monitor; others felt that the feedback helped.

For all four categories of stimuli, the initial value of the comparison stimulus was set so as to be above a listener's differential threshold, as estimated from pilot data. The value of a comparison stimulus decreased after two correct responses and increased after one incorrect response. A turnaround occurred whenever the value of the comparison went from increasing to decreasing, or vice versa. For the first turnarounds of the adaptive procedure, the value of the comparison was changed by a factor of 1.0121 for the OQ and by a factor of 1.0365 for the SQ. After the first four turnarounds, the factor by which the comparison value changed was reduced to 1.003 for the OQ and 1.009 for the SQ. Each adaptive run continued for a total of 14 turnarounds. The JND for each run was based on the geometric mean of the last 10 turnarounds (see Endnote 2).

Each listener participated in a total of four adaptive

runs for each of the four stimulus categories. The order of the 16 runs was randomized. The first run for each stimulus category was considered practice. The JND values given below for each listener were based on the average of the last three runs of each stimulus category (see Endnote 3).

The subjects listened binaurally to the stimuli using calibrated and matched (within +/- 0.5 dB) TDH-49 earphones while seated in an IAC sound booth. The stimuli were presented at constant amplification levels. The levels of the individual stimuli ranged from approximately 75 to 85 dBA. The signals differed only relative to the intrinsic changes provided by the varying OQ and SQ values.

Each session lasted approximately 45 minutes. A training session (approximately 15 minutes) preceded each session in order to acquaint (or reacquaint) the subject with the task. The training session ended when the subject said that he or she felt comfortable with the task. A maximum of two sessions were permitted in one day, with a minimum two hour rest period between sessions. All sessions for a listener were run within a nine day period.

Results

JND Values

Table I shows the OQ and SQ results for the eleven subjects. Each value listed in the table is the average over the last three sessions. Table II gives the means and standard deviations for the OQ and SQ results. The average JND value for the OQ output pressure stimuli was 0.0344 (sd=.020). That is, on average the OQ value of 0.6344 could just be discriminated from the standard OQ value of 0.6000 for the output pressure stimuli. Similarly, the average JND value for the OQ glottal flow stimuli was 0.0264 (sd=.010). Using MANOVA techniques (17), the results showed no significant difference between gender when the output pressure and flow data for the OQ were collapsed together. Furthermore, the results showed no significant difference between the output pressure and the flow stimuli when the OQ data were collapsed across gender (ref. Table III). The standard deviations for the OQ output pressure JND values were about twice as large as for the OQ flow JND values (Table II). A test for difference between variances of two related samples (18) was significant ($p < .05$).

The average JND value for the SQ output pressure stimuli was 0.319 (sd=0.167), such that listeners on average could just discriminate between the SQ value of 2.319 and the standard value of 2.000. Likewise, the average JND value for the SQ glottal flow stimuli was 0.154 (sd=0.043). The results showed no significant difference between gender when the output pressure and flow data for the SQ were

Table I.
The Mean of the Stimulus Values of the Last Three Sessions Across the Four-Categories of Stimuli for Each Subject

Subject	Open Quotient Output Pressure	Open Quotient Glottal Flow	Speed Quotient Output Pressure	Speed Quotient Glottal Flow
M1	0.6593	0.6345	2.3078	2.1831
M2	0.6125	0.6238	2.4166	2.2308
M3	0.6178	0.6275	2.1104	2.1469
M4	0.6304	0.6234	2.1636	2.1482
M5	0.6169	0.6135	2.1042	2.1123
M6	0.6552	0.6153	2.3878	2.1299
F1	0.6458	0.6420	2.3529	2.2184
F2	0.6125	0.6389	2.4297	2.1484
F3	0.6203	0.6189	2.2901	2.1010
F4	0.6460	0.6374	2.2641	2.1654
F5	0.6618	0.6157	2.6845	2.1101

Table II.
Means and Standard Deviation for the Open Quotient and the Speed Quotient Values at the JND Level

	Mean	JND	SD
Open Quotient Output Pressure			
Males (6)	.6320	.0320	.020
Females (5)	.6373	.0373	.020
Entire sample (11)	.6344	.0344	.020
Open Quotient Glottal Flow			
Males (6)	.6230	.0230	.008
Females (5)	.6306	.0306	.012
Entire sample (11)	.6264	.0264	.010
Speed Quotient Output Pressure			
Males (6)	2.248	.248	.140
Females (5)	2.404	.404	.169
Entire sample (11)	2.319	.319	.167
Speed Quotient Glottal Flow			
Males (6)	2.159	.159	.042
Females (5)	2.149	.149	.047
Entire sample (11)	2.154	.154	.043

Table III.
Statistical Results From MANAVA Procedures

Comparison between	While collapsing data across	F	Prob. of F
Open quotient:			
Male vs female	Output pressure & glottal flow	.88	.374
Output pressure vs glottal flow	Male and female	1.29	.285
Gender by factor		.03	.871
Speed quotient:			
Male vs female	Output pressure & glottal flow	2.13	.178
Output pressure vs glottal flow	Male and female	13.58	.005
Gender by factor		3.12	.111

collapsed together. However, in contrast to the results for the OQ stimuli, the results for the SQ showed a significant difference between the output pressure and the glottal flow stimuli ($p < .006$) when the gender data were collapsed together. That is, the JND value for the SQ glottal flow stimuli (0.154) was significantly different from the SQ output pressure JND (0.319). There was no significant gender interaction effect. The standard deviations for the SQ output pressure JND values were nearly four times as large as for the SQ flow JND values (Table II). A test for difference between variances of two related samples (18) was significant ($p < .001$).

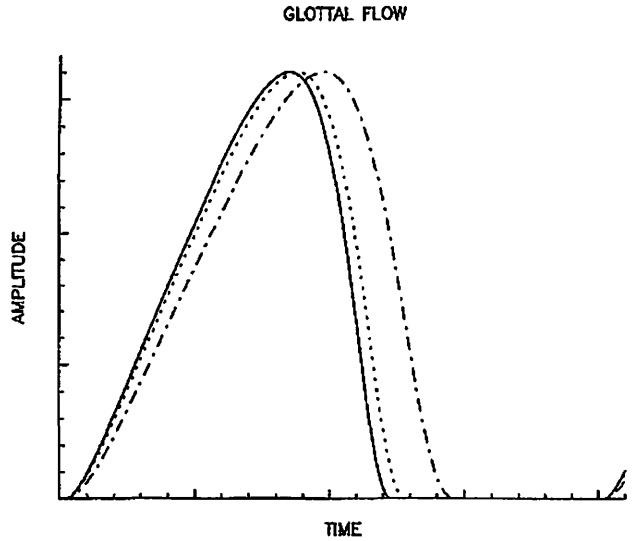


Figure 1. Glottal flow signals for open quotient. The values shown are for the standard 0.6000 (solid line), the group average JND 0.6264 (dotted line), and the largest value used in the presentations 0.7159 (dot-dashed line).

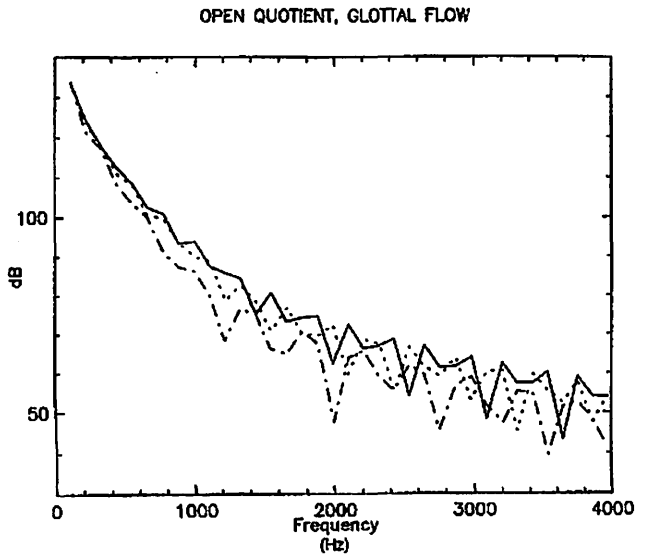


Figure 2. Spectra for the open quotient glottal flow signals. The values shown are for the standard 0.6000 (solid line), the group average JND 0.6264 (dotted line), and the largest value used in the presentations 0.7159 (dot-dashed line).

Spectral Differences, Open Quotient

Figure 1 shows glottal flow cycles for a number of OQ values. The values correspond to the standard (.6000), the group average JND for the flow stimuli (.6264), and the largest valued stimulus used for experimental (listening) comparisons (.7159). In order to show the spectral complexity contrasting these signals, the corresponding power spectra are given in Figure 2. Because the OQ value is varied, the relative amplitude of neighboring components vary, creating overlap of the group JND spectrum with the standard spectrum (see Endnote 4). Figure 3 shows the difference between the dB values for each partial of the standard signal compared to the group average JND. Differences are less

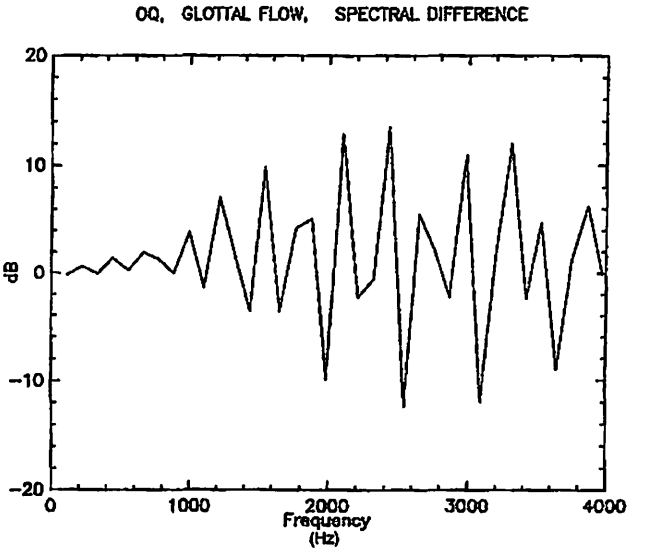


Figure 3. Spectral differences for the open quotient glottal flow between the standard and the group average JND stimulus.

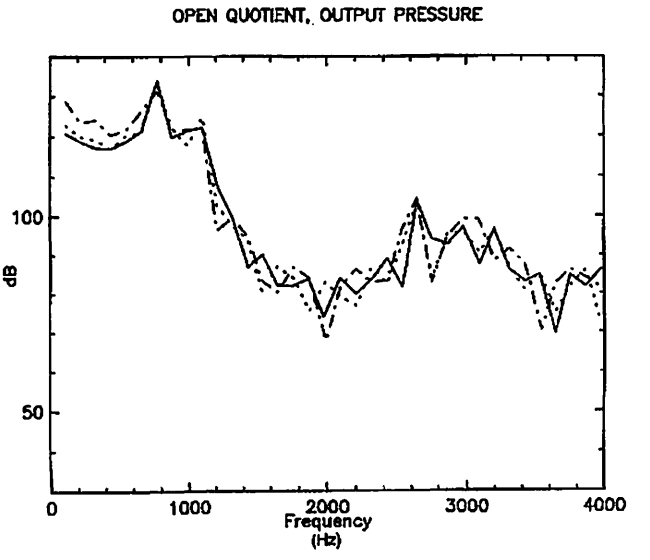


Figure 4. Spectra for the open quotient output pressure signals. The values shown are for the standard 0.6000 (solid line), the group average JND 0.6344 (dotted line), and the largest value used in the presentations 0.7159 (dot-dashed line).

than about 2 dB up to 900 Hz, then increase to approximately 13 dB near 3000 Hz. It is noted that the spectra themselves are over 40 dB down by 1000 Hz and 70 dB down by 3000 Hz (Figure 2).

Figure 4 gives the spectra for the output pressure signal for different values of the OQ. Here the formant structure for /ɑ/ is imposed upon the source signal. The values used to create Figure 4 are those from the listening tests of the output pressure, and are therefore different from those for the glottal flow stimuli. The group average JND stimulus was .6344. Again it is noted that the overlap between the standard and group average spectra is due to the varying OQ values. The spectra are higher in relative intensity up to 1000 Hz in the output pressure spectra of Figure 4 (where the spectral values stay within about 15 dB of each other) compared to the glottal flow spectra of Figure 2 (where the spectral values fall rapidly). Figure 5 shows the spectral differences between the standard and the group average JND value. The differences are comparable to those found for the glottal flow signal alone (Figure 3). That is, they vary up to about 2 dB below 900 Hz, and then increase to a maximum of about 12 dB throughout the rest of the spectral range.

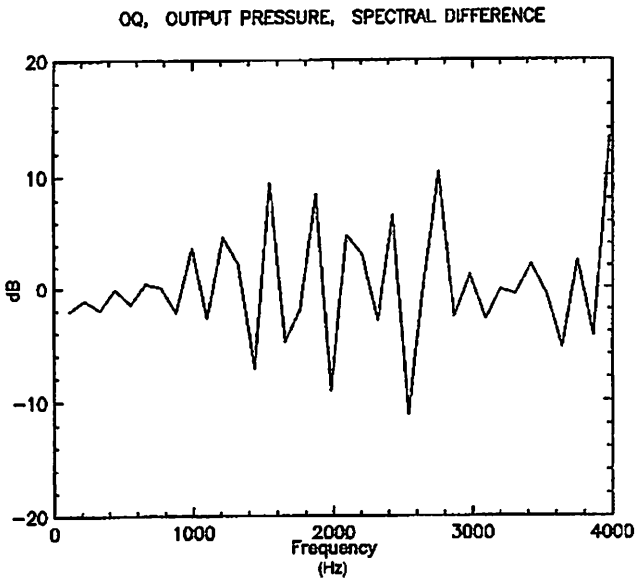


Figure 5. Spectral differences for the open quotient output pressure between the standard and the group average JND stimulus.

Spectral Differences, Speed Quotient

Figure 6 shows the glottal flow cycles for the SQ values corresponding to the standard (2.000), the group average JND for the flow stimuli (2.154), and the largest value used for experimental presentations (3.102). The spectra for these signals are shown in Figure 7. Because the OQ was held at .6000, the spectra essentially are monotonic to each other (one above the other with the standard having the lowest spectral values), rather than having the overlapping structure shown in Figure 2 for different values of the OQ. The spectral values of corresponding components do vary relative to each other, however, as shown in Figure 8 (following page). Similar to the findings in Figure 3 for the OQ flow signals, the spectral differences up to approximately 900 Hz are within 2 dB. For frequencies greater than 900 Hz, the spectral differences are typically less than 3 dB but all negative, a finding different from that seen in Figure 3 for the OQ flow. The pattern of spectral differences above about 2000 Hz in Figure 8 is similar to that of Figure 3 in that the largest differences occur at every fifth component, due to the OQ value being at or close to .6000 for all signals.

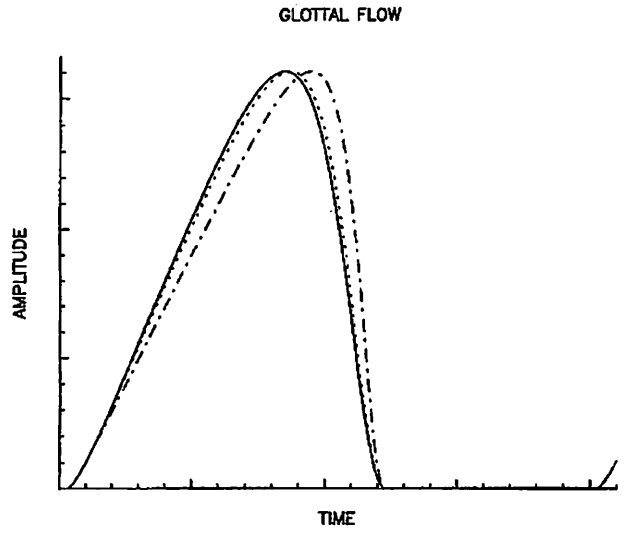


Figure 6. Glottal flow signals for speed quotient. The values shown are for the standard 2.000 (solid line), the group average JND 2.154 (dotted line), and the largest value used in the presentations 3.102 (dot-dashed line).

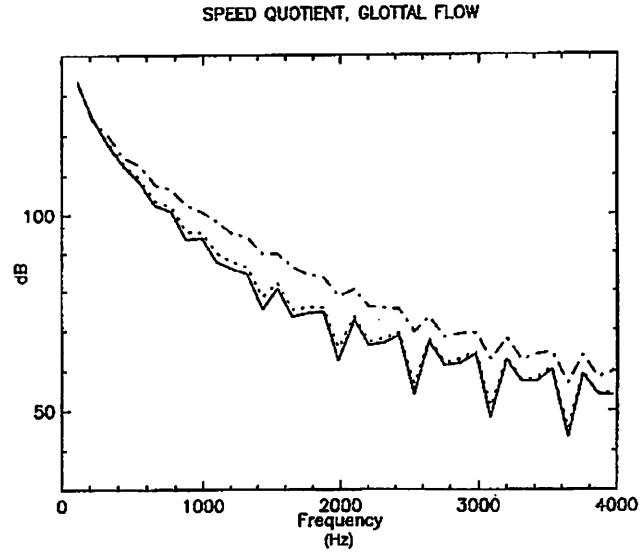


Figure 7. Spectra for the speed quotient glottal flow signals. The values shown are for the standard 2.000 (solid line), the group average JND 2.154 (dotted line), and the largest value used in the presentations 3.102 (dot-dashed line).

Figure 9 shows the spectra for the SQ output pressure signals. The JND value for this signal was 2.319 for the group average. The corresponding spectral differences are shown in Figure 10. The differences are within approximately 3 dB, similar to what was shown in Figure 8 for the flow signal. Dissimilar to Figure 8 are the observations that the imposition of the resonance structure creates some overlap of the various spectra (giving positive dB differences in Figure 10), and up to about a 3 dB difference for the lowest components with the standard being greater than the JND comparisons.

Compared to the standard OQ value of .6000, the group JND stimulus of .6344 for the output pressure was 5.73% larger. The corresponding value for the glottal flow (JND stimulus of .6264) was 4.40% larger. Compared to the standard SQ value of 2.000, the group JND stimulus of 2.319 for the output pressure was 15.95% larger. The corresponding value for the glottal flow (JND stimuli of 2.154) was 7.7% larger. Despite the larger relative increases in the value of the SQ the spectral variation between the standard and the JND stimuli is greater for the OQ as shown in Figures 3 and 5 compared to Figures 8 and 10.

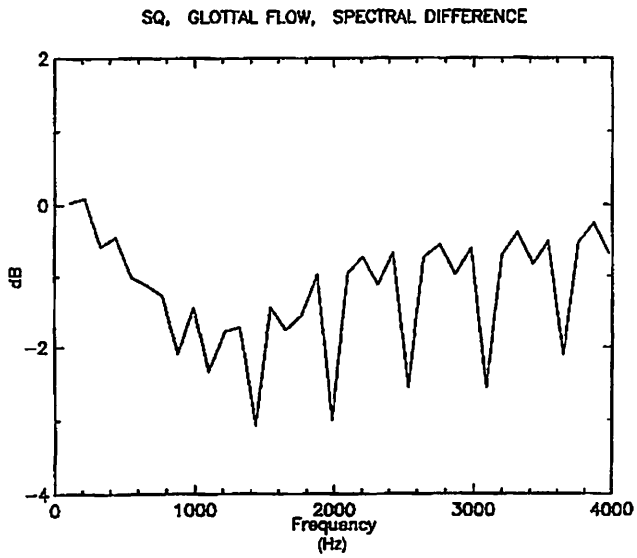


Figure 8. Spectral differences for the speed quotient glottal flow between the standard and the group average JND stimulus.

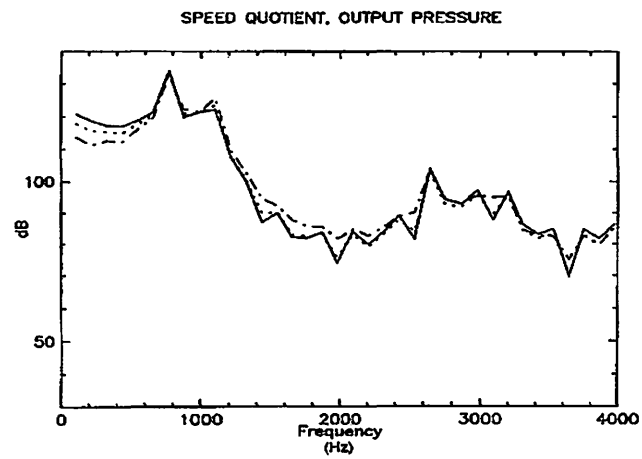


Figure 9. Spectra for the speed quotient output pressure signals. The values shown are for the standard 2.000 (solid line), the group average JND 2.319 (dotted line), and the largest value used in the presentations 3.102 (dot-dashed line).

Discussion

The output pressure signal is closer to what people would listen to during normal conversations than the glottal flow signal itself. For the output pressure stimuli, the group JND values were .0344 for the OQ and 0.319 for the SQ. These values could be considered when developing stimuli differences in training tapes for discriminating normal voice qualities, or for creating voice synthesis schemes.

Although the OQ JND value of 0.0264 for the glottal flow stimuli was not significantly different from the OQ JND value of 0.0344 for the output pressure stimuli, there was a difference for the SQ. The SQ JND of .319 for the output pressure signal was significantly larger than the JND of .154 for the glottal flow signal. While the explanation for this difference is not clear, it is interesting to consider the role of differences in the spectra of the glottal flow and output pressure stimuli. As reflected in Figure 8, all but the first two glottal flow spectral components of the standard SQ spectrum were below the components for the JND spectrum. In contrast, Figure 10 shows that there was more overlap of the output pressure standard spectrum with the JND spectrum. The JND for the glottal flow SQ may have been significantly lower, therefore, than that for the output pressure SQ because the spectral changes were essentially all in

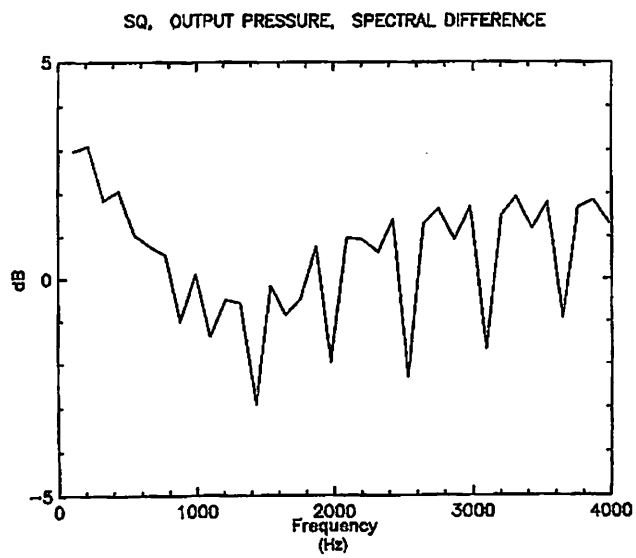


Figure 10. Spectral differences for the speed quotient output pressure between the standard and the group average JND stimulus.

the same direction for the flow signals. A more thorough analysis based on auditory models will be pursued in a subsequent paper.

The subjects listened to the four different signal categories four times. As mentioned above, the first session stimuli did not yield many of the lowest JND values, and the lowest were distributed among the remaining three sessions. This finding motivated the decision to report the JND values as the average of the last three session values. A replication of this study should determine if the JND values would decrease with more training.

The larger JND variances (Table II) found for the output pressure stimuli compared to the glottal flow stimuli suggest that the listeners had a more difficult listening experience when the vocal tract influence on the glottal flow was included. The greater complexity of the output pressure spectra may have made the judgment of small changes more difficult, wherein the listening strategies may have varied more widely among listeners.

Conclusions

This study was motivated by the need to establish correspondences between auditory abilities and laryngeal function, and to determine nominal resolutions of glottal flow characteristics for auditory training tapes and voice synthesis. Just noticeable difference (JND) values were obtained for glottal flow open quotient (for a standard value of 0.600) and the speed quotient (for a standard value of 2.000). Both the glottal flow signal itself and the glottal flow signal as input to a synthetic /ɑ/ vocal tract were used. The stimuli were synthesized using SPEAK (9). Eleven singing teachers acted as listeners. The JND values were taken as the average of the last three sessions in which the subjects heard the stimuli in an adaptive listening task. The results of this study include the following:

1. The group average JND values were: 0.0264 for the open quotient glottal flow, 0.0344 for the open quotient output pressure, 0.154 for the speed quotient glottal flow, and 0.319 for the speed quotient output pressure.

2. There were no gender differences in the JND values between the glottal flow and output pressure signals for either the open quotient or the speed quotient.

3. There were significant JND differences between the glottal flow and output pressure stimuli for the speed quotient, but not for the open quotient. The significant difference for the signal types for the speed quotient may be related to the greater spectral overlap of the output pressure signal.

4. The variance among the JND values was significantly larger for the output pressure stimuli compared to the glottal flow stimuli for both the open quotient and the speed quotient.

This study appears to be the first to determine JND values for glottal flow (glottal volume velocity waveform) characteristics per se, and as such should be replicated and expanded to include other groups of subjects and other glottal flow characteristics.

Endnotes

1. Three male subjects (M3, M4 and M5) did not pass the screening at 4000 Hz. For these subjects, thresholds at 4000 Hz were measured to be between 30 and 40 dB HL. While it is important to acknowledge that the threshold loss at 4000 Hz may have affected the performance of these three subjects, the effects may have been minimal for two reasons. First, the spectra of the signals used in this study show significant rolloff in the higher frequencies, such that the components near 4000 Hz are reduced by approximately 50-70 dB compared to the most intense spectral component. And secondly, the average ranking of the JND values across the four stimulus types (refer to Table I) for these three subjects was not high (4.5, 5.0, and 2.0, where 11 would represent the largest JND values), suggesting that there was no obvious decrement to their discrimination abilities.

2. The geometric mean for the group of 10 numbers n_1, n_2, ..., n_{10} was obtained by multiplying the numbers together, $n_1 \cdot n_2 \cdot \ldots \cdot n_{10} = M$, and then taking the 10th root of that multiplication, $M^{0.1} = JND$.

3. The decision was made to average the three last JND values for each subject for each of the stimulus categories. This followed from the finding that there were no consistent asymptotic trends after the first session. The largest JND occurred 52% of the time in the first listening session. The smallest JND value occurred 9% of the time in the first session. In the second session, the smallest JND occurred 25% of the time, and the largest 23% of the time. In the third session, the smallest JND occurred 39% of the time, and the largest 14%. And in the fourth session, the smallest JND occurred 27% of the time, and the largest in 11% of the time. Thus, the trend was not to have the best JND value occur in the fourth session, but was spread out primarily over the second, third and fourth sessions. This lack of strong asymptotic behavior may have been due to the choice of subjects, that is, experienced singing teachers who were accustomed to listening for small vocal changes, and thus may have operated near threshold with relatively little experimental training.

4. In general, changes in the open quotient (with speed quotient held constant) are expected to yield spectra that change in slope as well as in relative intensity of neighboring partials (9).

Acknowledgments

This study was supported (in part) by research grant number P60 DC 00976 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The authors are grateful to Ingo Titze for use of his SPEAK program. We also would like to thank Christopher Linker, Russel Long, Christopher Dromey, and Darrell Wong for their assistance. We appreciate the comments by Shari Campbell on an earlier draft of this manuscript.

References

1. Scherer RC. Laryngeal function during phonation. In: Rubin JS, Sataloff RT, Korovin GS, and Gould WJ, eds. *Diagnosis and Treatment of Voice Disorders*. New York: Igaku-Shoin Medical Publishers, Inc., 1995:86-104.

2. Wendahl RW. Laryngeal analog synthesis of harsh vocal quality. *Folia Phoniatrica* 1963;15:241-250.

3. Wendahl RW. Some parameters of auditory roughness. *Folia Phoniatrica* 1966a;18:26-32.

4. Wendahl RW. Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. *Folia Phoniatrica* 1966b;18:98-108.

5. Coleman RF, Wendahl RW. Vocal roughness and stimulus duration. *Speech Monographs* 1967;34:85-92.

6. Coleman RF. Effect of waveform changes upon roughness perception. *Folia Phoniatrica* 1971;23:314-322.

7. Heiberger VL, Horii Y. Jitter and shimmer in sustained phonation. In: Lass NJ, ed. *Speech and language: Advances in Basic Research and Practice, Vol 7*. New York: Academic Press, 1982:299-32.

8. Hillenbrand J. Perception of aperiodicities in synthetically generated voices. *J Acoust Soc Am* 1988;83:2361-2371.

9. Titze IR, Mapes S, Story B. Acoustics of the tenor high voice. *J Voice* 1994;95:1133-1142.

10. Fant G. Glottal source and excitation analysis. *STL-QPSR 1/1979*:85-107.

11. Fant G. The source filter concept in voice production. *STL-QPSR 1/1988*:21-37.

12. ANSI S3.6-1969. *Specifications for audiometers*. New York: American National Standards Institute, 1969.

13. Scherer R, Sundberg J, Titze I. Laryngeal adduction related to characteristics of the flow glottogram. *J Acoust Soc Am* 1989;85:S129(A).

14. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. *J Voice* 1989;3:294-305.

15. Fant G, Liljencrants J, Lin Q. A four-parameter model of glottal flow. *STL-QPSR 4/1985*:1-13.

16. Levitt H. Transformed up-down methods in psychophysics. *J Acoustical Soc Am* 1971;49;467-477.

17. SPSS Inc., 444 North Michigan Avenue, Chicago, IL.

18. Bruning JL, Kintz BL. *Computational Handbook of Statistics*. Atlanta: Scott, Foresman and Company, 1968.

NCVS Status and Progress Report - 12
June 1998, 95-113

Vocal Tract Area Functions for an Adult Female Speaker Based on Volumetric Imaging

Brad H. Story, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Ingo R. Titze, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Department of Speech Pathology and Audiology, The University of Iowa
Eric A. Hoffman, Ph.D.
Division of Physiologic Imaging, Department of Radiology, The University of Iowa College Medicine

Abstract

A collection of 3-D vocal tract shapes corresponding to vowels and consonants of American English have been acquired for a 27 year old adult female subject using magnetic resonance imaging (MRI) and electron beam computed tomography (EBCT). Each 3-D shape was condensed into a set of cross-sectional areas of oblique sections perpendicular to the centerline of the vocal tract's long axis, resulting in an "area function". Formant locations computed for each (non-occluded) area function showed reasonable similarity to those determined from the natural (recorded) speech of the imaged subject, but differences suggest that some of the imaged vocal tract shapes were articulated differently during imaging than during natural speech. A comparison is also made between area functions derived using both MRI and EBCT methods for the vowels /i/ and /ɑ/. Additionally, the area functions reported in this study are compared with those from two previous studies and demonstrate general similarities in shape but also obvious differences that can be attributed to differences in imaging techniques, image processing methods, and anatomical differences of the imaged subjects. Finally, since this set of images and subsequent area functions for female subject compliments a previous similar study concerning an adult male subject [Story, Titze, and Hoffman, J. Acoust. Soc. Am, 100(1), 1996], a limited comparison of pharyngeal and oral cavity length measures are compared across genders. Results indicate that the primary difference is that the female pharynx is about 36 percent shorter than that of the male.

Introduction

During the production of speech, information is transmitted to a listener that codes a linguistic message, but also carries with it characteristics that give generally robust cues to a speaker's identity such as gender, age, dialect, health, and emotional state. Both the linguistic and paralinguistic cues result from the physical configuration of the two primary components of the speech production system: the voice source and the vocal tract. The voice source is comprised of the vibrating vocal folds whose geometric and viscoelastic properties largely dictate the spectral content of the voiced sound. The vocal tract (during speech) is a continuously changing tube-like airway that "sculpts" the raw sound of the sound sources (periodic and sometimes turbulent) into a stream of vowels and consonants that carry along the speaker-specific qualities.

With a long term goal of using computer models to accurately simulate many different types of voice and speech qualities, much information about the physical properties of both the voice source and vocal tract are needed. The focus of this study is on the vocal tract. In particular, the interest is in using magnetic resonance imaging (MRI) to study the morphological nature of an adult female vocal tract vocal tract during the production of static speech sounds. This is a direct complement to a previous study concerning an adult male (Story, Titze, and Hoffman, 1996b). Thus, the experimental protocol and image analysis procedures used in the present study are identical to those in the previous one. As in our earlier work, this study does not address the anatomical or physiological characteristics of the articulators (lips,

jaw, velum, and tongue) explicitly, but instead attempts to understand the morphology of the vocal tract airway as a manifestation of the articulator positions for statically produced sounds. Since most speech simulation models are based on one-dimensional acoustic wave propagation, the 3-D vocal tract shape must be reduced to a set of cross-sectional areas stacked consecutively from the larynx to the lips. This representation of the vocal tract shape is generally called an "area function". An important component for speech simulation is an inventory of area functions that correspond to the vocal tract shapes used to produce human speech.

The specific aim of this paper is to present, both graphically and numerically, the vocal tract area functions derived from 3-D volumetric vocal tract imaging for one adult female subject. Like in the adult male study (Story et al.), 12 vowels, 3 nasals, and 3 plosives will be presented. In addition to using MRI, electron-beam computed tomography (EBCT) was used to acquire image sets for two vowels that were also included in the MRI-based set. The limited vocal tract data obtained from EBCT is useful for 1) measuring the volume of the teeth; this information is used to correct the MRI image sets since MRI poorly images teeth, 2) accurately measuring the length and cross-sectional area of the piriform sinuses, and 3) allows for a cross-comparison of vocal tract shapes imaged by two independent imaging systems.

While presentation of the female area functions is the primary focus of this paper, measurements of pharyngeal and oral cavity lengths will be compared to similar measurements for the previous male subject (Story et al., 1996b). The male and female data sets were not reported simultaneously in the same article because of length of time required to obtain and analyze the data as well as the amount of data reported. A limited number of female area functions available from other publications will be compared to the corresponding vocal tract shapes from this study. The format of this paper is deliberately similar to that of Story et al. (1996b) so that the new information presented here can be easily compared to that of the previous study. However, the sections comparing the MRI- and EBCT-based vowels were not included in Story et. al. but were given elsewhere (Story, Hoffman, and Titze, 1996a).

History of Female Vocal Tract Measurements

Much of the information concerning the female vocal tract consists of length measurements of the pharyngeal and oral cavities as compared to males (Holbrook and Carmody (1937), Carmody (1941), Fant (1966)), rather than actual area functions describing the vocal tract shape. A study specific to pharyngeal growth (in terms of length) for males and females ranging in age from 3 months to 16 years was performed by King (1952). By age 16, the length of the pharynx in females was about 12 percent shorter than that of a typical male. Nordstrom (1975, 1977) attempted to use the length ratios of female to male pharyngeal and oral sections from the studies cited above, to transform the well known male area functions of Fant (1960) into an equivalent set of female area functions. He found that the anatomical scaling factors poorly predicted the acoustic scaling factors measured from recorded speech, leading to the conclusion that females likely use different articulatory positions than males. This finding suggests that the female area functions cannot be considered as just a length compressed version of the male.

Using available published articulatory data, Goldstein (1980) proposed a model for the vocal tracts of growing children, both male and female. Her results also implied that females and children may use articulatory positions somewhat different than males. The success of a model such as Goldstein's is limited by the fact that it relies on deriving the area function by transforming midsagittal dimensions to cross-sectional area, rather than a direct area measurement from the 3-D shape. Empirically based transformations of this type have been proposed by Heinz and Stevens (1964), Lindblom and Sundberg (1971), and Mermelstein (1973) with the goal of converting measurements on x-ray projection images into area functions. Such transformations may be more of a limitation for computing female area functions since the rules were originally based on male subjects.

In light of this, Sundberg, Johansson, Wilbrand and Ytterbergh (1987) sought to find an accurate transformation of sagittal dimension to cross-sectional area for females. They attempted to establish the relationship between the lateral width of the pharynx and its cross-sectional area for one male and one female subject. A series of tomograms were imaged at four axial planes within the pharynx of each subject. Plots of cross-sectional area versus sagittal distance indicated a significantly different relationship between distance and area for the male and the female, suggesting that the male and female pharynx are not related by a simple scale factor.

Hogberg (1995) reported a study of the acoustic characteristics of ten French vowel area functions for one male and one female subject. The area functions were derived, using a power function transformation of sagittal distance, from sagittal x-ray tracings by Botherel et al. (1986). He found that the derived area functions produced formant frequencies that often deviated significantly from formant values measured from the recorded speech of the subject; more so for the female than for the male. He cites some potential causes for the discrepancies, including a possibly inaccurate transformation from sagittal distance to area in both the lip and larynx tube regions.

While the studies reviewed above have contributed significantly to the understanding of female speech production, they have all suffered from the lack of three-dimen-

sional vocal tract data to use as a starting point. Modern imaging techniques, especially MRI, have become a viable option for obtaining 3-D information about the human vocal tract shape for static vowels and consonants. While currently not fast enough to allow three-dimensional imaging of dynamic speech with the resolution required for our studies, MRI does offer the advantage of highly detailed three-dimensional reconstructions of each static shape. However, Crary et al. (1996) have reported imaging dynamic speech with an MRI technique that allowed them to capture three single plane images per second. Currently, the method is limited to acquisition of a midsagittal section and four axial sections which is not enough information to reconstruct accurate 3-D vocal tract shapes extending from glottis to lips. Echo planar imaging (Pykett and Rzedzian, 1987) along with other emerging methodology offers a promise that reconstruction of the dynamic airway in three dimensions may be possible in the near future.

There have been a number of vocal tract studies performed using MRI, but not all have reported area functions derived directly from a 3-D reconstruction of the vocal tract shape (see Story et al. (1996b) for a review of these studies). Of interest here is the volumetric imaging of the 3-D vocal tract shape and subsequent derivation of area functions, especially for female subjects. While Baer et al.'s (1991) study established the effectiveness and viability of using MRI to acquire information about the vocal tract, it included only male subjects. Dang et al. (1994) used MRI to acquire coronal and axial image sets used in reconstructing the nasal tract and the vocal tract for nasal consonants. Both male and female subjects were imaged but vocal tract area functions of /m/ and /n/ were only reported for a male subject. In another study, MRI-based 3-D vocal tract shapes and area functions of a male, female, and child subject for five Japanese vowels were reported by Yang and Kasuya (1994). This same data set was later used to test uniform and non-uniform length transformation schemes for converting female speech to male speech (Yang and Kasuya, 1995).

Fricative consonants of two male and two female subjects have been imaged by Narayanan (1995) and Narayanan, Alwan, and Haker (1995). These studies have provided detailed information about the constrictions and air channels that produce turbulence generated sound, as well as reconstructions of the tongue body itself. Complete area functions were given for only one male subject in Narayanan et al. (1995) but were provided for both a male and female subject in Narayanan (1995). More recently, MR imaging experiments of the liquid approximants, laterals and rhotics, have been described by Narayanan et al. (1997) and Alwan et al. (1997), respectively; the studies included both males and females. Several variant shapes (e.g. light, dark, bunched, tip up, etc) for each category were imaged and resulting 3-D reconstructions and area functions were given.

Image Acquisition and Analysis

Scanning Parameters and Protocol

Volumetric imaging of the vocal tract was performed using MRI for 22 different phoneme configurations of one female subject and also for the nasal tract and trachea. However, only a subset consisting of 12 vowels, 3 nasals, and 3 consonants will be reported at this time; the four fricatives will be reported, along with those from the male (which were also not reported in Story et al. 1996), in a forthcoming article concerning fricative production. These "phonemes" are shown in Table I with the example word that was given to the subject prior to image acquisition. For the consonants, the subject imagined that the preceding and following vowel shape was the neutral vowel. At the time of scanning, the subject (DJ) was a 27 year-old female with no history of speech or voice disorders and is native to the state of Texas in the southern United States. However, she did not have a pronounced Texas accent. The subject was 5 ft 2 in tall and weighed approximately 120 pounds. Additionally, the vowels /i/ and /ɑ/ were imaged for subject DJ using EBCT.

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

Phonetic Symbol	Example
i	heat
ɪ	hit
ɛ	head
æ	hat
ʌ	ton
ɑ	hot
ɔ	paw
o	hoe
ʊ	hood
u	who
ɜ˞	earth
l	lump
m	mug
n	nut
ŋ	sung
p	puck
t	tuck
k	cut

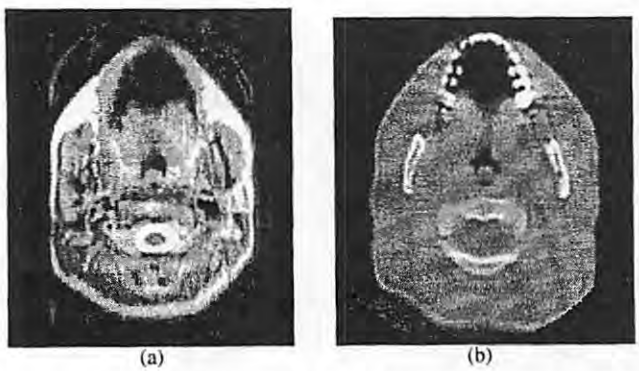
The MR images were acquired using a General Electric Signa 1.5 Tesla scanner. The image acquisition mode and pulse sequence parameters were identical to those used in Story et al. (1996b). A 24 slice series of 5 mm thick contiguous, parallel, axial sections extended from just cephalad of the hard palate down to about the first tracheal ring. The image slices were gathered in an interleaved acquisition.

All EBCT images were obtained using an Imatron C-150 scanner (Boyd and Lipton, 1983) (Imatron Corp., South San Francisco). For a single volume scan of the vocal tract, a series of 3 mm thick contiguous, parallel, axial images were acquired with the same superior/inferior extent as for the MRI. The axial images captured the full A-P (anterior-posterior) and lateral extent of the airway within each slice. The field of view (FOV) for each slice was 18 cm while the pixel matrix was 512 x 512. This provided x and y pixel dimensions in the axial plane of 0.352 mm. The time required to scan the desired 40 slice volume using the EBCT was approximately 20 seconds, which was too long to comfortably phonate without strain. Thus, the subject phonated for 10 seconds at the end of which scanning was halted to allow for respiration and then restarted. (The current version of the EBCT scanner is now able to acquire 140 slices in 15 seconds.)

The protocol for collecting the images with each of the two techniques was similar. In both cases the subject was positioned in a comfortable supine position on the patient table and was required to phonate at a comfortable pitch and loudness. More details concerning the scanning protocols can be found in Story (1995) and Story et al. (1995, 1996b).

Image Analysis

The analysis of all image sets was performed with VIDA (Volumetric Image Display and Analysis) which is a general image display and quantitation package (Hoffman et al., 1992; see also the Internet web page at http://everest.radiology.uiowa.edu).



(a) (b)

Figure 1. Series of axial slices through the oral cavity for the vowel /ɑ/: a) MRI, b) EBCT.

Figure 1 shows an axial slice through the oral cavity for the vowel/ɑ/ using both MRI and EBCT; note that the teeth are visible in the EBCT slice. The image analysis process included three main steps: 1) segmentation of the airway in each slice (like those in Fig. 1) from the surrounding tissue including an account of the space occupied by teeth (note that the dimensions of the teeth can be obtained from the EBCT images and used to correct the MR image sets (Story et al., 1996a, 1996b), 2) three-dimensional reconstruction of the airway by shape-based interpolation, and 3) determination of an airway centerline and subsequent extraction of cross-sectional areas assessed from oblique sections calculated to be locally perpendicular to the airway centerline to create an "area function". This image analysis process is the same as was used in our previous work which can be found in Story (1995) and Story et al. (1995, 1996) and also at the Internet web site- http://everest.radiology.uiowa.edu/tutor/app/vocal/vocal.html. Detailed explanations of the image analysis will not be repeated here.

Because the image sets were gathered in an interleaved fashion, any change in position of the subject's body (most importantly the head and neck) from the beginning to the end of scanning one vocal tract shape will be reflected as an offset of the two interleaved halves of the image set. A significant offset of this type will create a rippled texture in the superior/inferior dimension of the 3-D reconstructed shapes. Prior to the image analysis process for each shape, an assessment was made as to whether such an offset had occurred. This assessment was made by reformatting the axial image sets so that a midsagittal slice could be viewed. Any offset was clearly indicated in this plane if sharp discontinuities existed in the contour of the spine or other identifiable boundaries. If present, the offset was corrected by moving one half of the image set (every other slice) relative to the other until the discontinuities were as smooth as possible. This correction was performed on the vowels /ɑ/, /ɔ/, and /æ/. Even after correction, the rippled effect can be seen in the oral cavity of the /ɔ/ (see Figure 5). Such an effect was observed only in the /m/ and /n/ of the Story et al. (1996b) and no correction was applied in that case. The problem came to light only when a reconstruction of the /ɑ/, the first shape analyzed, yielded a highly rippled structure. A small amount of rippling can be seen in many of the uncorrected surface rendered shapes, usually in the pharyngeal region. The deviation in these cases was regarded as too small to be effectively corrected.

Determination of Computed and Natural Speech Formants

The formant frequencies associated with each vocal tract shape were computed with a wave-reflection analog vocal tract model (Story, 1995) which included energy

losses due to the yielding properties of the vocal tract walls, fluid viscosity, and radiation from the mouth. An acoustic side branch representing the piriform sinuses was also implemented. The dimensions of this sidebranch used in the model were an averaged version of the piriform sinus data presented in a later section. The model was sampled at a frequency of 44.1 KHz and each finite section of the area function represented a tube length of 0.396 cm. The frequency response function of each vowel-like area function was computed by injecting a one sample impulse into the glottal end of the vocal tract and Fourier transforming the output (the impulse response). The first three formants were determined with a peak picking algorithm based on parabolic interpolation (Titze et al., 1987).

For comparing the computed formant values to the natural speech of the subject, a high quality audio recording was made in which the subject produced speech sounds that corresponded to the static shapes acquired with MRI. The subject was recorded while in a supine position on the suspended floor of an anechoic chamber. Ear plugs were used to create similar acoustic feedback conditions as those experienced in the MR scanner, but no attempt was made to replicate the scanner's acoustic signal. The speech sounds were sustained for several seconds and each was repeated three times.

For each vowel, the second of the three repetitions was chosen for analysis and a sample about 1.5 to 2.0 seconds long was extracted from the middle of the total recorded production; this ignored the onset and offset of the vowel. The formant locations for the natural speech samples were determined by computing linear prediction spectra. The LPC algorithm was a 50 pole autocorrelation method (Markel and Gray, 1976). A frame-based analysis was then performed throughout the sample where the frame size was 4096 points and frames were not overlapped. For each frame, an LPC spectrum was computed and a peak picking algorithm based on a parabolic interpolation method (Titze et al., 1987) was used to find estimates for the first three resonance frequencies of the vocal tract. Finally, the mean, maximum, minimum, and standard deviation values were computed for each formant throughout the series of analysis frames.

Results and Discussion

EBCT-Based Vowel Shapes

Three projections of the surface rendered airways are shown for the two vowels /i/ and /ɑ/, along with their "raw" area functions (i.e. points are not necessarily spaced in equal increments) in Figures 2 and 3, respectively. The first picture for each case is a sagittal projection followed by a rotated and tilted view allowing for a better perspective of the 3-D shape. The third surface display for both vowels, is a perspective from the posterior of the subject. For each surface rendered airway, the most inferior point of the 3-D shape begins with the uppermost section of the trachea. Above the trachea, the airway becomes small in the region

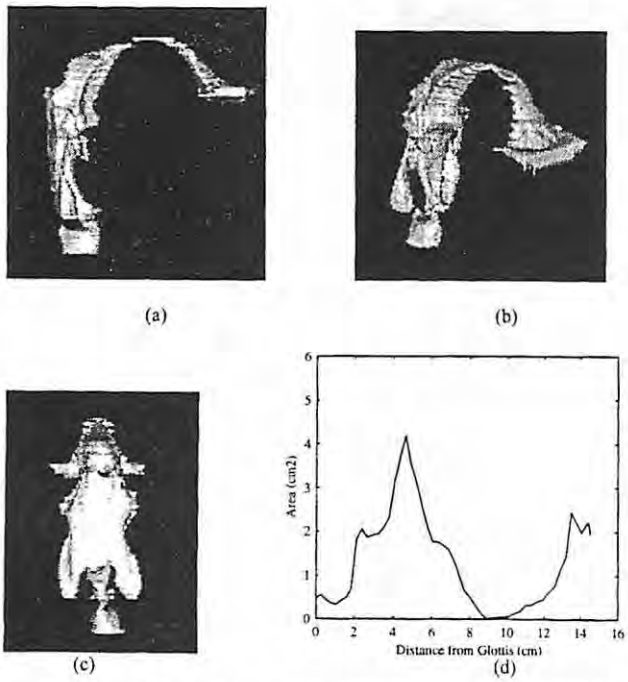


Figure 2. Shaded surface displays and area function of EBCT-based /i/ (subject DJ).

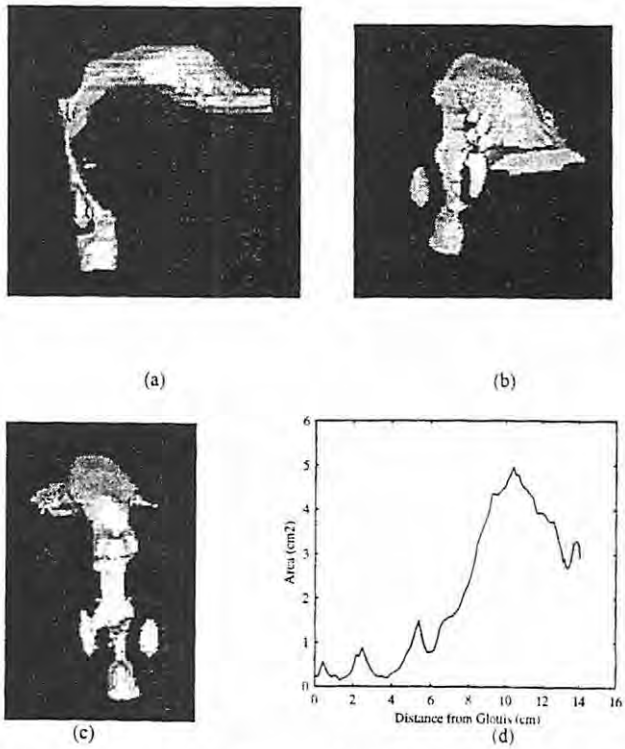


Figure 3. Shaded surface displays and area function of EBCT-based /ɑ/ (subject DJ).

of the glottis and then widens for a short distance into the laryngeal ventricle. Just above the ventricle, the airway constricts slightly as the epilaryngeal tube is formed before a widening into the lower pharynx occurs. The finger-like extensions that hang down below the pharynx are the piriform sinuses. Note that the right piriform sinus of the /ɑ/ is disconnected from the main vocal tract, thus removing itself from any acoustic consideration in this particular case. The measurements of the piriform sinuses will be presented in a later section. In each area function, the 0 cm location represents a point just above the glottis and the termination at the lips is at the right side of the graph.

Both the surface rendered airways and the area functions show the opposing configurations of the pharyngeal and oral cavities for these two vowels; i.e. large pharynx and constricted oral cavity for the /i/, constricted pharynx and large oral cavity for /ɑ/. However, the region just above the glottis appears quite similar for both cases. The area functions quantify this observation, indicating that the region from 0 cm to about 2 cm above the glottis begins with areas ranging from 0.25 to 0.4 cm^2 then rises slightly before decreasing to about 0.28 cm^2 for the /ɑ/ and 0.42 cm^2 for the /i/. The area begins to increase at approximately 2 cm for both vowels. This is approximately where the piriform sinuses join the main vocal tract tube.

Above the 2 cm point, the /i/ shows a sharp increase in area that reaches a value of 2.05 cm^2. Above this point, the area decreases slightly but remains mostly constant up to a point about 4 cm from the glottis. The area then increases rapidly to a peak value of 4.2 cm^2 at 4.6 cm, after which there is a steady decrease in area until a minimum value of 0.02 cm^2 is reached at the 8.9 cm location. The vocal tract is tightly constricted for a length of about 1.5 cm (from 9 cm to 10.5 cm) and then undergoes a gradual increase in area out to the lip termination which has a cross-sectional area of 1.9 cm^2.

The /ɑ/ shows an increased area above 2 cm that produces a small peak of 0.85 cm^2 before decreasing to 0.18 cm^2 at 3.8 cm. Above this location the area again rises to another small peak of 1.5 cm^2, decreases to 0.75 cm^2, and then increases into the widened oral cavity where a peak area of 4.9 cm^2 occurs 10.4 cm from the glottal end of the vocal tract. Following this peak, the area decreases gradually toward the lip termination which has an area of 2.9 cm^2.

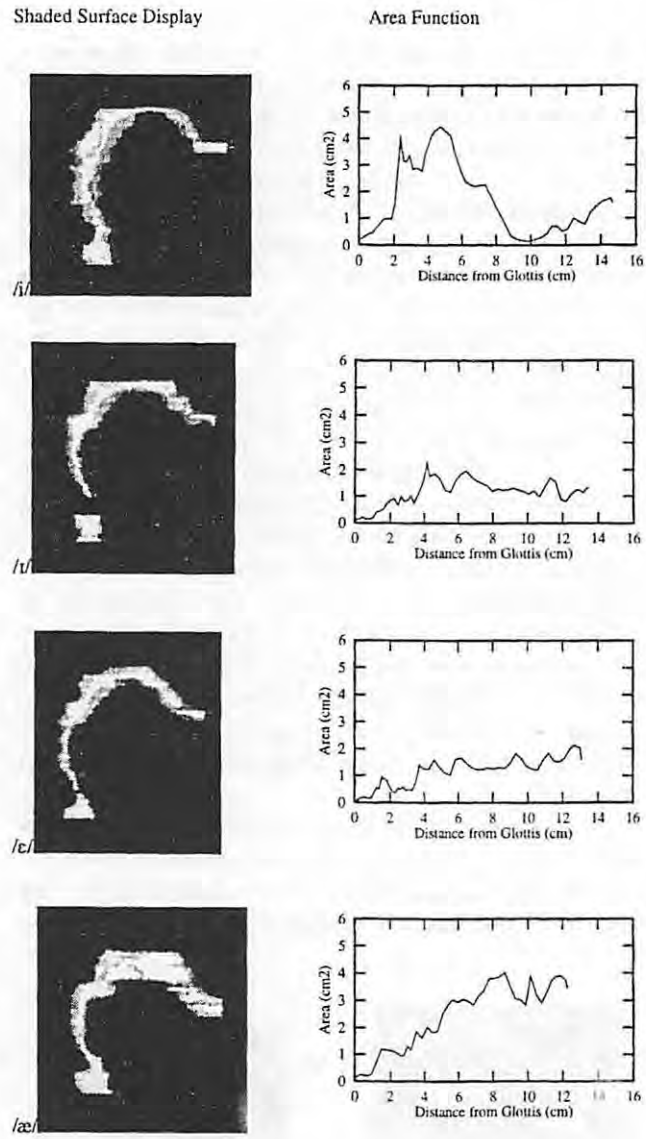


Figure 4. Surface rendered airways and "raw" area functions for /i/, /ɪ/ /ɛ/, and /æ/ (subject DJ).

MRI-Based Vowels

Sagittal projections of the surface rendered airways and their corresponding area functions, are shown in Figures 4, 5, 6, and 7 for the MRI-based vowels. As was the case for the EBCT-based vowel shapes, the most inferior point of each 3-D shape begins with the uppermost section of the trachea. Above the trachea, the airway becomes small in the region of the glottis and then widens into the lower pharyngeal section. In the MR image sets, the piriform sinuses were poorly defined and consequently were not segmented, thus the piriform extensions are not present in the surface renderings of these tract shapes. For each area function, the 0 cm location again represents a point just above the glottis and the termination at the lips is at the right side of the graph. The figures are grouped to show a progression of MRI-based shapes that first exhibit a constricted oral section, then a constricted pharyngeal region, and finally a constricted mid-section of the vocal tract.

The four surface rendered airways and area functions in Figure 4 (/i/, /ɪ/, /ɛ/, and /æ/) show a tightly constricted oral cavity and wide pharynx for the /i/ give way to increasingly larger oral cavities but shrinking pharyngeal cavities for /ɪ/, /ɛ/, and /æ/. The /ɪ/ contains a break in the

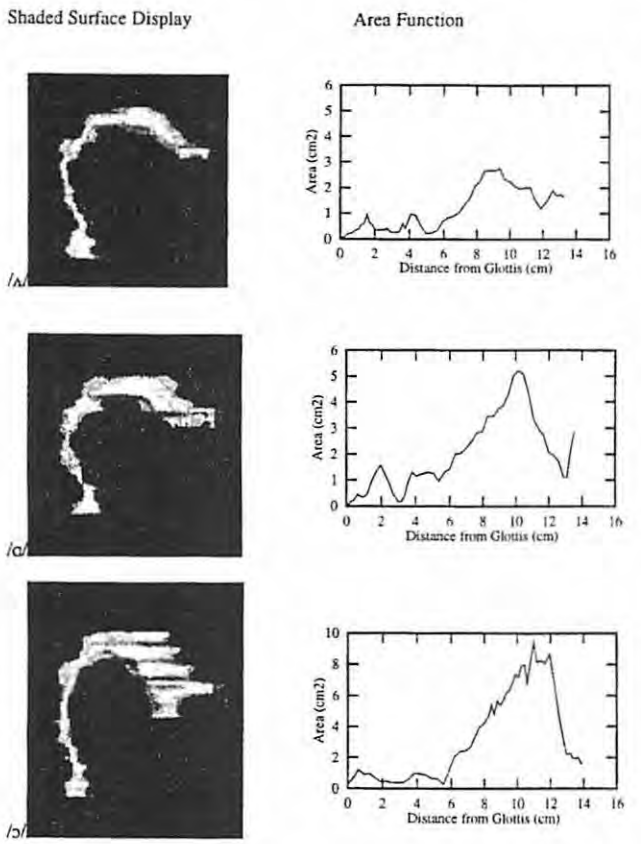


Figure 5. Surface rendered airways and "raw" area functions for /ʌ /, / ɑ/, and /ɔ/ (subject DJ).

airway in the region of the glottis because of blurring presumably caused by motion artifact. However, because these obscured slices occurred mainly in the glottal region, the area function is still believed to be complete; i.e. the 0 cm point is just above the glottis.

The specific case for the vowel /i/ indicates that the region just above the glottis increases almost linearly from an area of 0.2 cm^2 to 1.0 cm^2 at a point approximately 2 cm above the glottis. The tract reaches a maximum area of 4.5 cm^2 at a distance 5 cm from the glottis after which the area begins to decrease until a minimum of 0.14 cm^2 is reached at 9.5 cm. Following this minimum constriction, the area again rises to a value of 1.7 cm^2 at the lip termination which is 14.2 cm from the glottis.

The /ɪ/ and /ɛ/ are quite similar in all regions of the area function, especially the short section extending from 0 cm to about 1 cm. However, a careful examination of both the surface renderings and the area functions shows that the pharyngeal section (from about 2 cm to 8 cm above the glottis) for /ɪ/ is generally larger than for the /ɛ/. Above the 8 cm point, the cross-sectional area for the /ɪ/ shows a general decrease ending with a lip termination area of 1.34 cm^2. In this same region, the area for /ɛ/ tends to increase, ending at the lips with a value of 1.6 cm^2. It should also be

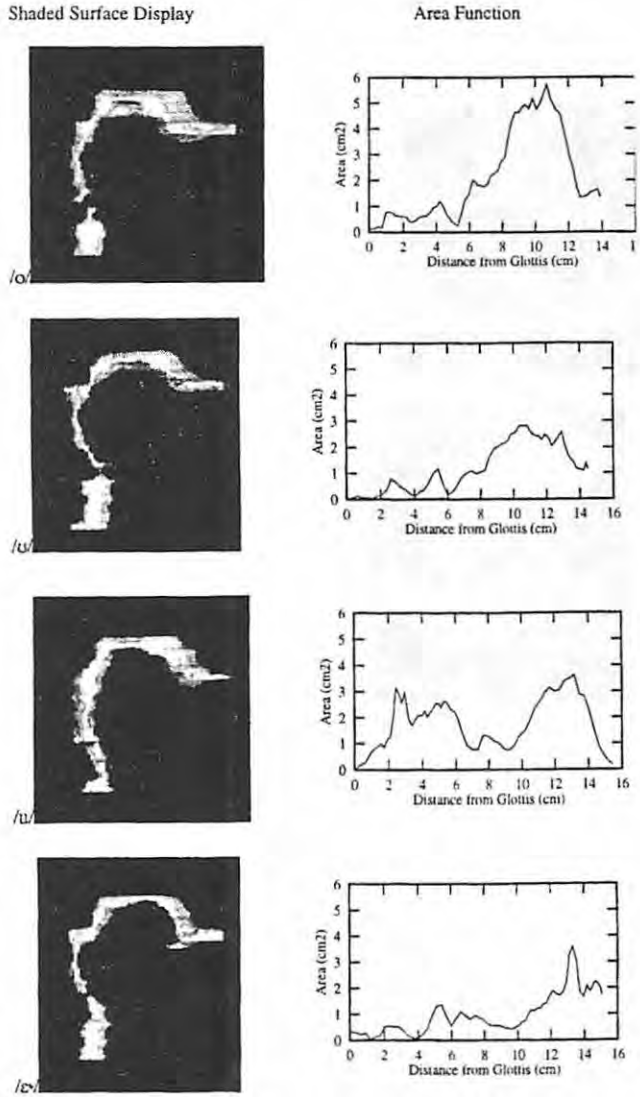


Figure 6. Surface rendered airways and "raw" area functions for /o/, /ʊ/, /u/, and /ɜ/ (subject DJ).

noted that for both vowels, nearly the entire area function is less than 2.0 cm^2. In contrast, the /i/ had a maximum area of 4.5 cm^2.

The /æ/ vowel demonstrates nearly the same characteristics as the /ɪ/ and /ɛ/ in the 0 cm to 1 cm region. Above this region the area gradually increases throughout the entire length of the vocal tract, reaching a peak value of 4 cm^2. The /æ/ maintains the characteristics of both a widened pharynx and a widened oral cavity, giving it a megaphone-like shape and also indicating that it is a transition between the front and back vowel categories.

Figure 5 shows surface renderings for /ʌ/, /ɑ/, and /ɔ/. All three show a constricted pharynx and a widened oral cavity that grows progressively larger throughout these vowels. However, the /ɑ/ has a larger lip termination area than either of the other two vowels. The area functions show

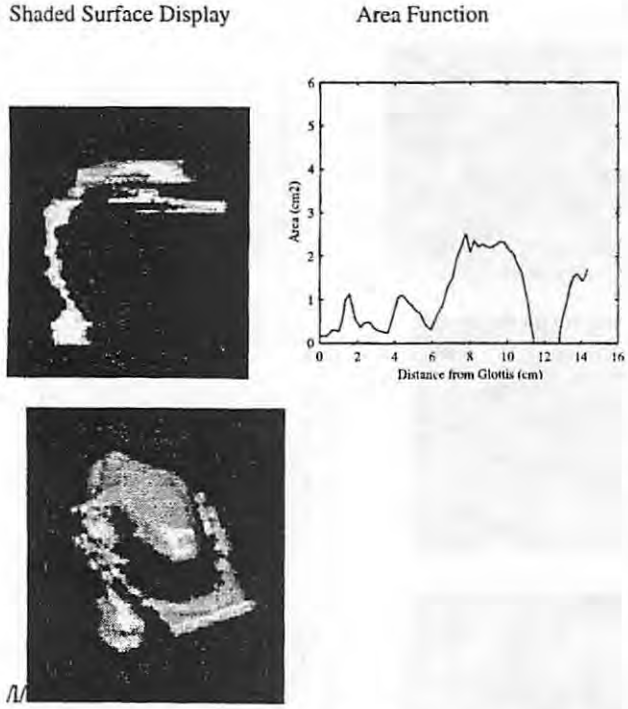


Figure 7. Surface rendered airway and "raw" area function for /l/ (subject DJ).

that the epilarynx and lower pharynx (0 cm to 6 cm above the glottis) for the /ʌ/ and /ɔ/ are both less than 1 cm^2 in area while the /ɑ/ is slightly more enlarged in this region with areas greater than 1 cm^2. Above a point about 6 cm from the glottis all three vowels have increased cross-sectional areas. The /ʌ/ reaches a peak value of 2.75 cm^2, while the /ɑ/ and /ɔ/ attain peak areas of 5.2 cm^2 and 9.4 cm^2, respectively. The maximum value of 9.4 cm^2 for the /ɔ/ vowel is the largest cross-sectional area in the entire set of area functions. Note that the vertical scale for the /ɔ/ area function extends from 0 to 10 cm^2 while all others are set to a range of 0 to 6 cm^2. The movement induced ripple mentioned previously is apparent in the oral cavity of this shape. The small, but sharp oscillations in the area function from about 8 cm to 12 cm above the glottis are due this rippling.

The airway images and area functions for /o/, /ʊ/, /u/, and /e/ are grouped together in Figure 6 because they are traditionally thought of as vowels which divide the vocal tract into two distinct chambers separated by a tight constriction. However, only the /u/ seems to exhibit this characteristic and even here the middle section is not highly constricted. A few of these vowel shapes seem to show more similarity to those in the previous figure (Fig. 5). In particular, the /o/ is similar to the/ɑ/ and the /ʊ/ similar to the /ʌ/ in the sense that the pharyngeal sections are comparably constricted and the oral cavities similarly enlarged. They are also similar in terms of absolute area measurements where the peak cross-sectional areas for /o/ and/ɑ/ are 5.7 cm^2 and 5.2 cm^2, respectively while the /ʊ/ and /ʌ/ have respective peak areas of 2.8 cm^2 and 2.75 cm^2. The main differences within these two pairs of vowels appears to be the location of the pharyngeal constrictions. For example, the /o/ has a distinct constriction at about 5.2 cm above the glottis while the /ɑ/ has comparable constriction 3 cm above the glottis. Likewise, the /ʊ/ contains pharyngeal constrictions at locations 4 cm and 6 cm from the glottis and /ʌ/ has similar constrictions at 3 cm and 5 cm above the glottis. Whether these constriction locations are enough of a difference to distinguish one vowel from another will be partially answered in a later section on acoustic modeling.

As already mentioned, the vowel /u/ demonstrates the mid-tract constriction and consequent double chamber vocal tract shape. The area function begins with a small area of 0.05 cm^2 just above the glottis. The area then increases gradually to a value of 1.1 cm^2 before increasing abruptly to 3.1 cm^2 at a distance of 2.2 cm from the glottis. Over the course of the next 4.5 cm the area falls into the range of 2.5 cm^2 but retaining the characteristic pharyngeal expansion of an /u/. The mid-tract constriction occurs from 7 cm to 9 cm with a minimum area in this region of 0.9 cm^2. The area then rises to a peak of 3.5 cm^2 before falling to 0.2 cm^2 at the lip termination.

The /e/ shape is uniquely differentiated from the other tract shapes by the presence of a small sublingual cavity. An oral constriction occurs at about 9.7 cm and the sublingual cavity at 13 cm. This cavity is often assumed to act as a sidebranch resonator, but in this case its length is small enough (less than a centimeter) that its effect as a sidebranch would be at high frequencies (greater than 10 kHz). Thus, the contribution of the sublingual cavity has been included in the area function and is the cause of the sharp area increase at the 13 cm point. The area function shows a depression occurring about 9.5-10 cm from the glottis, which is the apparent location for the oral constriction; this may be analogous to the "bunched tongue" configuration discussed by Alwan et al. (1997). While the appropriate characteristics of this vocal tract shape are present, it must be noted that the subject had difficulty in producing it. Even after several pre-imaging practice sessions she still felt that this shape was foreign to her as a phonetic element. Thus, the information regarding this tract shape must be used with some caution.

The /l/ is shown in Figure 7 from both a sagittal view and a rotated and tilted perspective to give an indication of the lateral air flow and acoustic wave path around the tongue. The contact of the tongue with the hard palate is indicated by the gap between the oral section and the lip section. The lateral pathways connecting these two cavities is, however, difficult to understand. From the surface renderings, the lateral pathways appear to connect to the vocal tract at a point in the posterior oral cavity. Such a connection point would imply that the subject pressed the

tongue against the alveolar ridge and extended the contact against the teeth posteriorly to at least the bicuspids. Such a configuration is possible but not typical of a natural production of this sound as it would have most of the oral cavity space behind the tongue acting as a sidebranch resonator. The area function was determined from the glottis to the alveolar ridge and also for the frontal cavity. The lateral pathways, which would connect these two sections, were poorly imaged and did not provide a contiguous air channel for which cross-sectional area could be measured. Thus, the area function is set to zero from 11.5 cm to 13 cm representing the distance from the main vocal tract to the frontal cavity. However, from the pathway sections that were present, an attempt was made to estimate a reasonable cross-sectional area that could serve to connect the front and back portions of the vocal tract. Using a region of interest (ROI) module in VIDA, approximately 0.10 cm^2 was determined as a possible connection area. The difficulty in imaging the lateral pathways is due to the 5 mm axial slice thickness which obscured the important detail of these small channels due to partial volume effects. The /l/ is a case where the accuracy of the reconstruction is compromised by the use only one imaging plane.

Nasals and Plosives

In keeping with the format of Story et al. (1996b), nasal and plosive tract shapes are included together because of their analogous vocal tract occlusions; i.e. /m/ and /p/ have bilabial closure, /n/ and /t/ have an alveolar occlusion, and /ŋ/and /k/ have an occlusion near the velum. Because voicing was not possible during image acquisition of the plosive data sets, no distinction between voiced and unvoiced plosives can be made; i.e. the shape for a /p/ is considered to be equivalent to /b/, /t/ for /d/, and /k/ for /g/. Figures 8 and 9 show the airway surface renderings and area functions for each of the three nasal and plosive consonants, respectively. The surface rendering of each plosive exhibits a break at the glottis caused by adduction of the vocal folds. For all six shapes, the most important feature is the location of the main vocal tract occlusion; in dynamic speech the shape of the open portion of the airway will likely be dependent on the preceeding and succeeding tract shapes but the constriction location should remain relatively consistent. For the nasals, the cross-sectional area of the open velopharyngeal port (through which air flow and acoustic waves are diverted) is also significant.

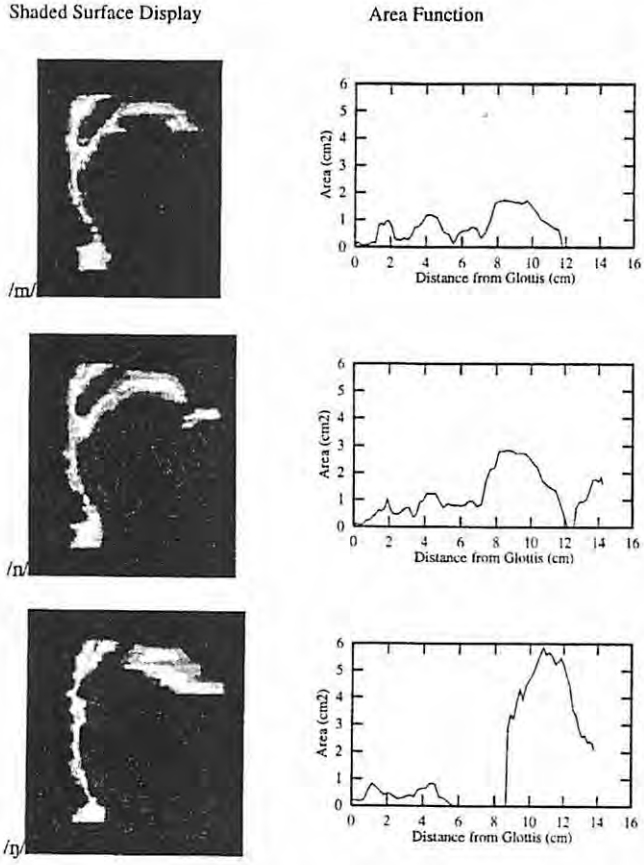


Figure 8. Surface rendered shapes and area functions for nasals (subject DJ).

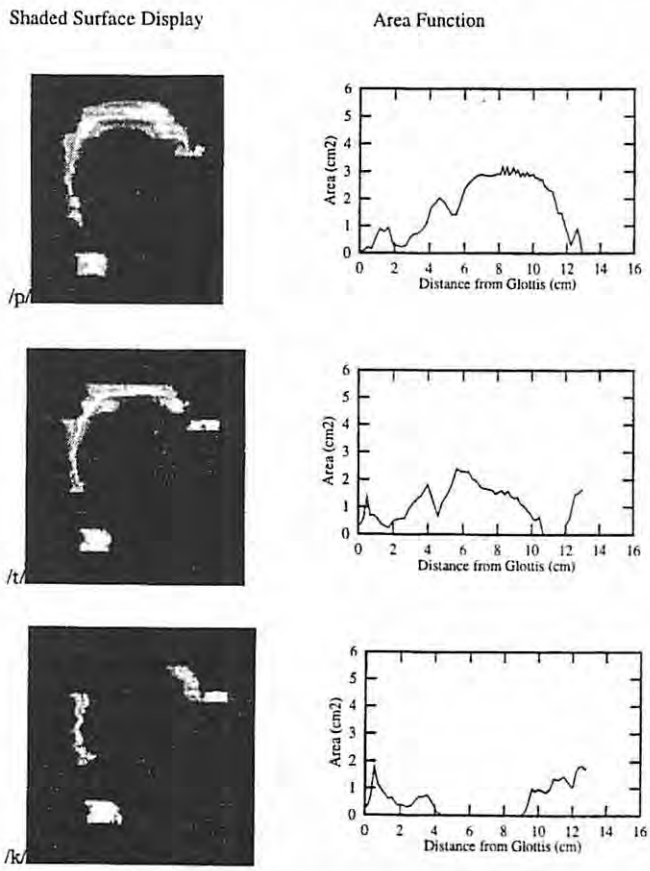


Figure 9. Surface rendered shapes and area functions for plosives (subject DJ).

The /m/ and /p/ both have a bilabial closure so that the area function becomes zero at the lip termination. Thus, a "break" in the vocal tract shape is not observed as it is for the other consonants. With the closure at the lips, the only outlet for sound during production of /m/ is through the nasal cavity, which means that the entire oral cavity becomes a side branch resonator. In the region 0 to 7 cm from the glottis, the area function for the /m/ oscillates from small areas of about 0.2-0.3 cm^2 to larger areas around 1 cm^2. Following this region, there is an increase in area to about 2 cm^2 in the oral cavity before closing down at the lips. The nasal coupling port area was measured to be 0.49 cm^2 and is located approximately 4 cm from the glottis. The /p/ is much like the /m/ in the first 2 cm of the area function, after which the area increases to a wide upper pharynx and oral cavity. From about 7 cm to 11 cm the area is relatively constant at about 3 cm^2 and then decreases rapidly toward closure at the lips.

The contact of the tongue with the hard palate and buccal walls is indicated by the break in the surface rendered shape of the /n/ and /t/. The detached portion of the airway that is downstream of the occlusion is, in both cases, acoustically inactive but is included to demonstrate the break. The /n/ is characterized by a cavity that is created from the airspace between the location of the nasal coupling (4 cm from the glottis) and the occlusion. Area functions for the /n/ and /t/ are in general quite different, with the /n/ beginning with a small area of 0.13 cm^2 just above the glottis and then increasing to about 1 cm^2 which is held more or less constant out to the 7 cm point, after which the area increases into a widened oral cavity. The occlusion occurs from 12.2 cm to 12.6 cm and then there is an increase in area out to the lip termination. The nasal coupling area was determined to be 0.31 cm^2 and is located 4 cm from the glottis. For /t/, the area function has a sharp increase in area in the epilaryngeal region, a characteristic not observed in any of the previous vocal tract shapes. Above the epilarynx the area increases to a peak of about 2 cm^2, then sharply decreases before climbing to a peak of 2.4 cm^2 at a distance of 5.6 cm from the glottis. Beyond this peak, the area declines toward the occlusion which occurs from 10.6 cm to 12.0 cm. The occlusion for /t/ occurs at a location about 2 cm closer to the glottis and is also longer than for the /n/.

The surface renderings of the /ɳ/ and /k/ both indicate an occlusion of the main vocal tract tube in the region of the velum. In fact, the /ɳ/ shows an almost straight vocal tract tube from the glottis up into the velopharyngeal tube with the tongue apparently closing off the vocal tract primarily by contacting the velum. The nasal coupling area was found to be 0.94 cm^2 and located 6.5 cm from the glottis. This coupling location is 2.5 cm higher than for the previous two nasal consonants, which is possibly caused by a lowering of the larynx during the /ɳ/ production. The area function shows a typically small epilarynx section, with an area of about 0.20 cm^2, followed by a small increase which stays relatively constant up to the nasal coupling location. The oral cavity is very large but, of course, is acoustically inactive. For the /k/, the break in the vocal tract covers a greater percentage of the total length than the /ɳ/. The break begins just higher than 4 cm above the glottis and continues for nearly 5 cm. The occlusion is apparently created by the tongue contacting both the soft palate and a portion of the hard palate. The area function shows an epilarynx similar to that observed for the /t/, where the area abruptly increases just above the glottis, this time to almost 2 cm^2. It is unclear whether such an abrupt expansion is a feature significantly different from the voiced productions or simply a result of closing off the glottal region and possibly the lowest portion of the epilarynx.

Piriform Sinuses

Measurements of the piriform sinuses were performed on the EBCT-based reconstructions of the /ɑ/ and /i/ because of the higher spatial resolution relative to MRI. Making measurements on both vowel shapes shows the differences in piriform configuration for a narrow versus wide pharynx. The piriforms can be seen in 3-D form in Figures 2 and 3 where they are the finger-like extensions that hang down below the pharynx. Figure 10 shows the area functions for the left and right sinuses for each of the vowels. The x-axis on each graph represents the distance from the entrance of each sinus (the connection point to the main vocal tract). For the /i/, the right sinus connects to the main tract 2 cm above the glottis and the left connection is at a distance of 1.8 cm. The left sinus connection for the /ɑ/

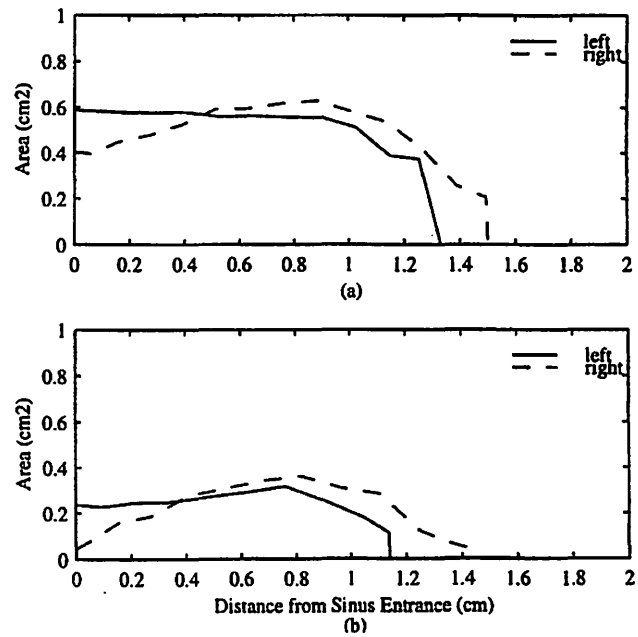


Figure 10. Area functions for left (solid) and right (dashed) piriform sinuses: a) /i/, b) /ɑ/.

occurs at 2 cm and the right sinus does not connect at all. In each case, the left sinus is shorter than the right. In general, the cross-sectional areas are nearly twice as large for the sinuses in the /i/ vowel versus those from the /ɑ/, almost certainly due to the pharyngeal widening of the /i/.

Numerical Area Functions

In this section, the area functions shown in Figs. 4-7 are presented in numerical form in order to make the data readily accessible. However, the area functions given in Figs. 4-7 were in their "raw" form, meaning that the linear distance between consecutive cross-sectional area values is not necessarily constant throughout the vocal tract (areas were separated by 0.2 to 0.4 cm of vocal tract length). The wave-reflection algorithm discussed in Section 3 requires that each area function be discretized at equal length intervals and the length of the final area function must be an even integer multiple of the length interval. Thus each "raw" area function was transformed to meet these demands by first normalizing it to a discrete length that was closest to the measured length. The wave-reflection type model dictates that the length of each finite section of the area function be equal to the speed of sound divided by two times the sampling frequency. For this study (also for Story et al. (1996b)), the length interval was chosen to be 0.396825 cm which results from a sampling frequency of 44.1 kHz and a speed of sound equal to 350 m/s. Following the length normalization, the area function is converted to the equal length interval representation by fitting it with a cubic spline and then sampling the resulting curve at the desired equally spaced intervals.

The resulting equal length interval area functions are shown in Table II and should be read by assuming that the glottal end of the vocal tract is represented by section 1 while the last section of each area function represents the mouth termination. The second row from the bottom of the table, labeled "n.c.", is the nasal coupling area, which is zero for all shapes except the nasal consonants. The last row indicates the vocal tract length for each area function. Table III gives the area functions for the two EBCT-based vowels. Like the previous table, the glottal end of the vocal tract is represented by section 1 and the last section is the lip termination.

Table II.
Equal Interval (0.396825 cm) Area Functions for 18 MRI-Based Vocal Tract Shapes

(Section 1 is the glottal end of the vocal tract and "n.c." denotes the nasal coupling.)

Sec.	/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ɝ/	/l/	/m/	/n/	/ŋ/	/p/	/t/	/k/
1	0.36	0.21	0.18	0.26	0.18	0.20	0.82	0.17	0.07	0.22	0.27	0.18	0.06	0.08	0.17	0.21	0.98	1.13
2	0.50	0.14	0.15	0.19	0.29	0.42	1.11	0.18	0.10	0.48	0.30	0.32	0.09	0.26	0.32	0.29	0.64	1.04
3	0.82	0.40	0.55	0.70	0.51	0.46	0.95	0.81	0.07	0.81	0.02	0.60	0.20	0.43	0.83	0.90	0.47	0.74
4	1.00	0.51	0.92	1.23	0.94	1.17	0.70	0.65	0.05	0.90	0.15	1.12	0.79	0.58	0.56	0.94	0.27	0.65
5	1.45	0.82	0.49	1.14	0.43	1.55	0.44	0.61	0.18	1.20	0.48	0.41	0.91	0.91	0.42	0.25	0.43	0.37
6	3.90	0.72	0.53	1.03	0.37	0.99	0.45	0.46	0.43	3.13	0.53	0.49	0.26	0.46	0.37	0.24	0.56	0.31
7	3.33	0.88	0.50	1.10	0.42	0.38	0.38	0.44	0.75	2.71	0.53	0.41	0.33	0.60	0.26	0.38	0.67	0.44
8	2.84	1.01	0.46	1.27	0.27	0.18	0.36	0.61	0.47	1.75	0.34	0.29	0.37	0.65	0.36	0.69	1.12	0.70
9	2.75	0.91	1.35	1.78	0.51	0.97	0.56	0.67	0.29	2.01	0.11	0.23	0.73	0.52	0.33	0.81	1.32	0.76
10	3.84	1.55	1.16	1.94	0.60	1.17	1.01	0.97	0.11	2.20	0.08	0.72	1.06	1.00	0.65	1.22	1.76	0.37
11	4.30	1.70	1.54	1.80	1.00	1.24	0.95	1.19	0.32	2.16	0.28	1.13	1.16	1.22	0.72	1.81	1.28	0.01
12	4.38	1.86	1.31	2.16	0.48	1.32	0.79	0.66	0.58	2.56	0.08	0.93	0.94	1.06	0.79	2.02	0.80	0.00
13	4.16	1.36	1.03	2.78	0.21	1.08	0.68	0.37	1.11	2.61	1.33	0.75	0.51	0.75	0.25	1.61	1.38	0.00
14	3.22	1.14	1.38	3.01	0.28	1.18	0.33	0.46	0.84	2.34	1.16	0.49	0.15	0.80	0.20	1.38	2.01	0.00
15	2.50	1.67	1.67	3.00	0.62	1.38	0.88	1.33	0.17	2.05	0.57	0.34	0.56	0.79	0.00	2.04	2.33	0.00
16	2.25	1.93	1.45	2.98	0.84	2.04	2.06	2.00	0.39	1.21	0.93	0.75	0.61	0.87	0.00	2.56	2.31	0.00
17	2.19	1.79	1.26	2.92	0.92	2.06	2.41	1.81	0.81	0.80	1.03	1.24	0.76	0.92	0.00	2.80	2.01	0.00
18	2.27	1.57	1.22	3.27	1.15	2.36	2.47	1.77	1.03	0.74	0.81	1.68	0.34	0.75	0.61	2.89	1.76	0.00
19	1.78	1.46	1.29	3.80	1.47	2.78	3.10	2.24	1.02	1.36	0.93	2.26	0.71	1.96	1.55	2.84	1.67	0.00
20	1.25	1.22	1.27	3.79	2.00	2.86	4.01	2.43	1.04	1.18	0.80	2.12	1.56	2.21	2.20	2.89	1.57	0.00
21	0.74	1.28	1.28	4.03	2.32	3.50	4.43	2.87	1.42	1.05	0.57	2.29	1.70	2.84	2.05	3.13	1.61	0.00
22	0.28	1.22	1.45	3.46	2.71	3.54	5.15	4.24	1.93	0.78	0.57	2.26	1.70	2.84	0.58	2.97	1.55	0.00
23	0.18	1.31	1.83	3.06	2.64	3.84	5.61	4.63	2.10	0.78	0.50	2.20	1.66	2.73	3.14	2.98	1.31	0.04
24	0.14	1.23	1.46	2.75	2.74	4.31	6.07	4.95	2.21	1.17	0.43	2.31	1.59	2.72	4.05	2.93	1.16	0.55
25	0.19	1.13	1.26	3.89	2.26	5.16	6.62	4.92	2.53	1.47	0.51	2.22	1.66	2.57	4.13	2.86	0.92	0.88
26	0.28	1.20	1.25	2.91	2.14	5.19	7.15	4.76	2.82	2.00	0.71	2.02	1.33	2.27	4.75	2.76	0.56	0.92
27	0.48	0.98	1.77	3.22	1.96	4.39	8.05	5.34	2.80	2.50	1.09	1.65	0.98	1.69	5.39	2.73	0.61	0.91
28	0.75	1.47	1.60	3.82	2.02	3.20	8.38	5.43	2.46	2.92	1.18	0.75	0.82	1.49	5.80	2.32	0.00	1.37
29	0.51	1.70	1.52	3.90	1.83	2.89	8.13	4.75	2.32	3.10	1.33	0.29	0.65	1.32	5.64	1.90	0.00	1.42
30	0.71	0.97	1.80	3.43	1.25	2.13	8.10	4.26	2.46	2.98	1.64	0.00	0.00	0.49	5.29	1.47	0.01	1.14
31	1.00	0.86	2.14		1.35	1.95	8.64	2.97	2.09	3.40	1.78	0.00		0.00	5.21	0.50	0.38	1.50
32	0.74	1.20	1.59		1.90	1.38	6.02	1.99	2.57	3.61	1.79	0.00		0.71	4.27	0.00	1.12	1.85
33	1.24	1.18			1.70	1.29	3.32	1.33	1.76	2.97	3.19	0.79		0.93	3.19			
34	1.51	1.34			1.69	2.81	2.26	1.45	1.37	2.74	3.24	1.49		1.31	2.60			
35	1.65						1.94	1.60	1.09	1.79	1.56	1.52		1.75				
36	1.59						1.58	1.39	1.16	0.87	1.93	1.71		1.58				
37										0.44	2.24							
38										0.18	1.65							
n. c.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.31	0.94	0.00	0.00	0.00
VT leng.	14.29	13.49	12.70	11.90	13.49	13.49	14.29	14.29	14.29	15.08	15.08	14.29	11.90	14.29	13.49	12.70	12.70	12.70

Four of the area functions given in Tables II and III, namely /ʊ/, /ɛ/, /m/, and the EBCT-based /i/ contain some very small areas, less than 0.1 cm^2. The locations of these areas for the /ʊ/, /ɛ/, and /m/, occur within the first four sections of the area function; i.e. the epilarynx. A check of the original image slices indicated that these small areas were not due to a mistake in the segmentation process since the airway was, in fact, very constricted in this region of the vocal tract. However, regions as small as these are represented by only a few voxels. A slight increase in the threshold value for the seeded region growing algorithm (see Story et al. 1996) could have easily included a few more voxels, which would increase the measured cross-sectional area. It was decided not to adjust the threshold value for segmentation of these small areas, but maintain a consistent threshold throughout the vocal tract. Also, this region of the vocal tract is quite susceptible to motion artifact because of close proximity to the vibrating vocal folds and gross laryngeal movement. While such small areas may be due to measurement error or movement artifact, it is shown in a later section that these two vowels produced close acoustic matches to natural speech in terms of locations for formants F1 and F2.

The small areas for the EBCT-based /i/ occur in the oral cavity constriction between 8.5 cm and 10 cm above the glottis. The are two possible contributions to the small area in this case. According to the original axial image set, the subject appeared to have produced a highly articulated /i/ (tongue positioned as high as possible without frication) and that being the case, the axial slice thickness may be larger than the thickness of the airway in that region. Thus, a slice may contain contributions from both tissue and air.

Comparison of Natural and Computed Vowel Formants

Figure 11 shows a comparison of the first three formant locations extracted from the recorded natural speech and those computed for each corresponding area function. The error bars for the natural speech formants represent the range of each formant for the series of analysis frames throughout a speech sample while the dotted line connects the mean values. The solid line passes through the computed formant values which are shown with black dots.

The computed values of the first formants show only two (/ɪ/ and /ɛ/) out of the twelve vowels falling precisely within the range of the natural speech but five others (/æ, ʌ, o, ʊ, ɛ/) lie just barely outside their respective range. The remaining five vowels (/i, ɑ, ɔ, u, l/) had first formants significantly outside the natural speech range. For seven of the vowels (/ɪ, æ, ʌ, ʊ, u, ɛ, l/), the computed second formants fell within or just slightly outside the natural speech range while the other five fell well outside the range. The computed third formants are nearly all outside the natural speech ranges except for the /o/.

Table III.
Equal Interval (0.396825 cm) Area Functions for Two EBCT-Based Vocal Tract Shapes

(Section 1 is the glottal end of the vocal tract and "n.c." denotes the nasal coupling.)

Sec.	/i/	/ɑ/
1	0.50	0.56
2	0.36	0.24
3	0.38	0.19
4	0.49	0.20
5	1.45	0.46
6	2.03	0.80
7	1.87	0.47
8	1.94	0.25
9	2.13	0.22
10	3.02	0.30
11	3.90	0.42
12	3.73	0.75
13	3.05	1.40
14	2.27	1.01
15	1.77	0.78
16	1.73	1.26
17	1.60	1.54
18	1.21	1.64
19	0.68	1.98
20	0.46	2.40
21	0.21	3.16
22	0.03	3.65
23	0.04	4.33
24	0.05	4.35
25	0.07	4.58
26	0.14	4.94
27	0.27	4.59
28	0.32	4.39
29	0.41	3.98
30	0.53	3.93
31	0.69	3.72
32	1.14	3.37
33	1.80	2.81
34	2.34	2.81
35	2.01	3.11
36	1.95	2.94
37		
38		
n.c.	0.00	0.00
VT leng.	14.29	14.29

A tabular representation of the data in Figure 11 is shown in Table IV. The percentage error for the computed formants relative to the mean natural speech formants is also provided. Additionally, the computed formants resulting from the EBCT-based /i/ and /ɑ/ are included; note that these were not included in Figure 11. Across all vowels and formants, percentage errors range from 0.4 percent to 48.6 percent with the majority below 10. The simulation of the two vowels, /ɑ/ and /ɔ/, both produced formant locations for F1 which were significantly lower than those from the natural speech. Such a decrease would suggest a centralizing effect due to fatigue of the articulatory musculature as well as possible listening fatigue of the aural system. A possible centralizing effect due to fatigue was also observed in Story et al. (1996) for a male subject. The MRI protocol, which requires the subject to produce many repetitions of a

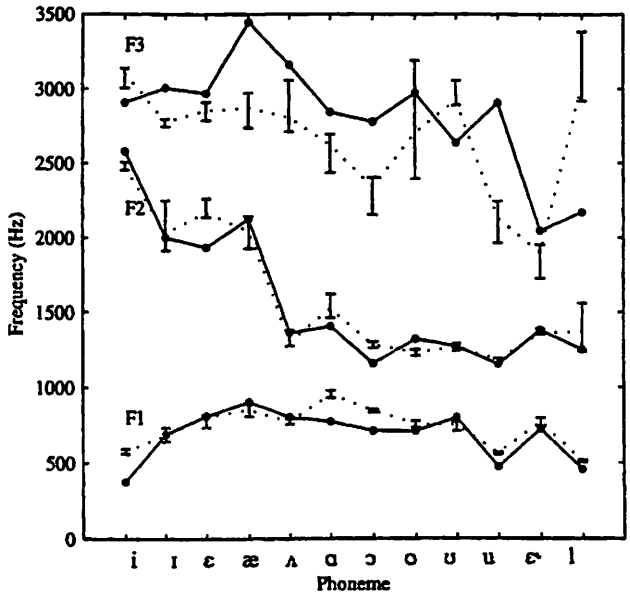


Figure 11. Comparison of formant locations for F1, F2, and F3 computed for the area functions in Table II and extracted from LPC spectra of the natural speech. The error bars depict the range of the natural speech formants while the dashed line passes through the mean values. The solid dots and connecting solid line represent the computed formant values.

Table IV.
First Three Formants From Natural Recorded Speech and Simulated Speech Based on the Area Functions Given in Tables III and IV

(The superscript "N" denotes the natural speech and "S" the simulated version. The Δ's represent the percent error of the formants from simulated speech relative to the mean value of the natural speech formants.)

	i	i-ct	ɪ	ɛ	æ	ʌ	ɑ	ɑ-ct	ɔ	o	ʊ	u	ɚ	l
$F1^N$	577	577	685	803	848	769	949	949	842	756	747	564	769	509
$F2^N$	2473	2473	2032	2164	2046	1304	1522	1522	1285	1233	1268	1184	1361	1372
$F3^N$	3107	3107	2776	2852	2871	2807	2623	2623	2339	2706	2923	2126	1898	3026
$F1^S$	378	328	689	807	900	799	769	855	709	707	795	472	721	459
$F2^S$	2582	2941	2001	1935	2123	1362	1406	1330	1159	1323	1273	1153	1381	1252
$F3^S$	2910	4620	3006	2967	3446	3161	2842	2845	2778	2972	2638	2906	2046	2170
Δ1	34.5	43.2	0.60	0.5	6.1	3.9	19.0	9.9	15.7	6.4	6.4	16.3	6.3	9.9
Δ2	4.4	18.9	1.5	10.5	3.8	4.4	7.6	12.6	9.8	7.3	0.4	2.6	1.5	8.7
Δ3	6.3	48.6	8.3	4.0	20.0	12.6	8.4	7.8	18.8	9.8	9.7	36.6	7.8	28.3

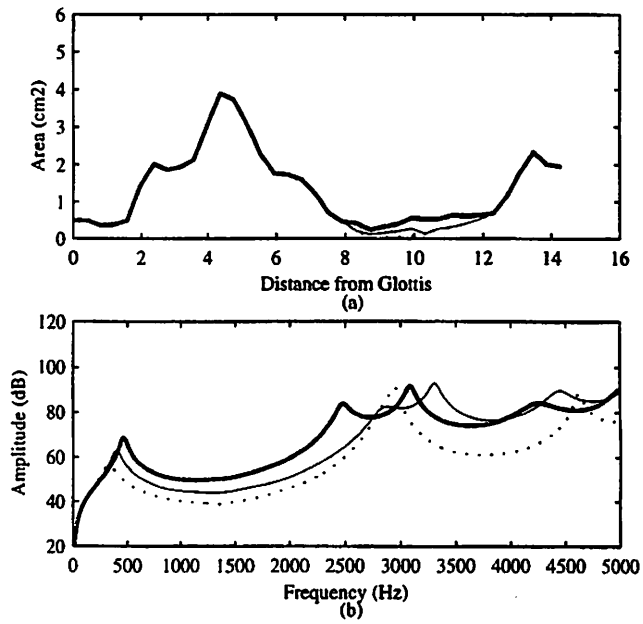


Figure 12. Modification of EBCT-based /i/ area function where the dotted line is the original, the thin line is a minor modification, and the thick line is a larger change:a) area function, b) frequency response functions.

given vowel (approx. 30 repetitions for each vocal tract shape), is a physical demanding process, thus, some fatigue is expected. The /ʊ/ and /ɛ/ area functions, whose small areas were discussed in a previous section, generated formants that were all in error by less than ten percent.

The two largest errors of 48.6 and 43.2 percent occurred in the first and third formants for the EBCT-based /i/ and interestingly the third largest error of 34.5 percent occurred in the first formant of the MRI-based /i/. This suggests that the performance of the subject for this vowel was significantly different during imaging than during the acoustic recording of her natural speech. In Section 4.5, the presence of the small areas in the EBCT /i/ were discussed as possibly being caused by an extreme version of the this vowel and, based on the computed formants, the MRI-based /i/ may also have been produced as more extreme than a comfortable production of the vowel.

Figure 12 demonstrates the effect of modifying the constriction area in the EBCT /i/ with two arbitrary area expansions. Three area functions are shown in Fig. 12a; the original as listed in Table III (dotted), a slight expansion of the constriction area (thin solid line), and a more substantial expansion of the constriction (thick solid line). The resulting frequency response functions are given in Figure 12b with corresponding line styles. A comparison of the spectra for the original area function and the modified versions indicates that the large error associated with third formant was due to the small constriction area forcing F2 and F3 to merge, thus causing the would-be F4 to be picked as F3. The slight modification of the constriction area shows the F2 and F3 peaks to be closely clustered but definitely separate. The second modification shows that the F2 and F3 peaks are both moving downward in frequency and are separated by a larger frequency range. Also note that both of the modifications move F1 upward in frequency. Thus, widening the tight constriction in this area function moves the first three formants closer to the formant locations measured from natural speech; F1, F2, and F3 values for the second modification were 464, 2482, and 3084 Hz, respectively (compare to Table IV). It is interesting that the /ɑ/, being a "corner" vowel would be centralized while its opposing corner vowel, /i/, seemed to be overarticulated.

Comparison of Area Functions from MRI and EBCT

The use of two entirely independent imaging modalities to acquire vocal tract shapes for the same two vowels allows for cross checking the validity of each measured area function; any gross measurement errors in using either method should be apparent in the comparison. It was also desired to have at least a few higher resolution image sets that could be used to understand the potential information loss resulting from the use of MRI. Another aspect of using two methods of imaging is that separate productions of the same vowel are captured at different times and on different days, thus allowing for an evaluation of measurement repeatability. Also, because the image acquisition time for EBCT is on the order of seconds rather than minutes, the EBCT-based tract shapes are much closer to being a "snapshot" of the vowel rather than a long-term average over many repetitions. Thus, EBCT avoids the potential for fatigue and consequent centralization of the tract shapes that may occur using MRI.

Figure 13 shows the /i/ and/ɑ/ area functions based on both EBCT (thick lines) and MRI (thin lines) along with their respective frequency response functions. The details of each area function have been discussed previously; the emphasis in this section will be on comparison between the two imaging methods. For both vowels, the maximum and minimum areas as well as their locations are quite comparable. For example, the /i/ has a maximum area at about 4.5 cm above the glottis for both cases and minimum in the 9 to 9.5 cm range. In addition, both versions of the /i/ have "shoulders" or humps on either side of the maximum area; these humps are 0.5 to 1 cm^2 smaller for the EBCT version but do occur at the same locations as in the MRI version.

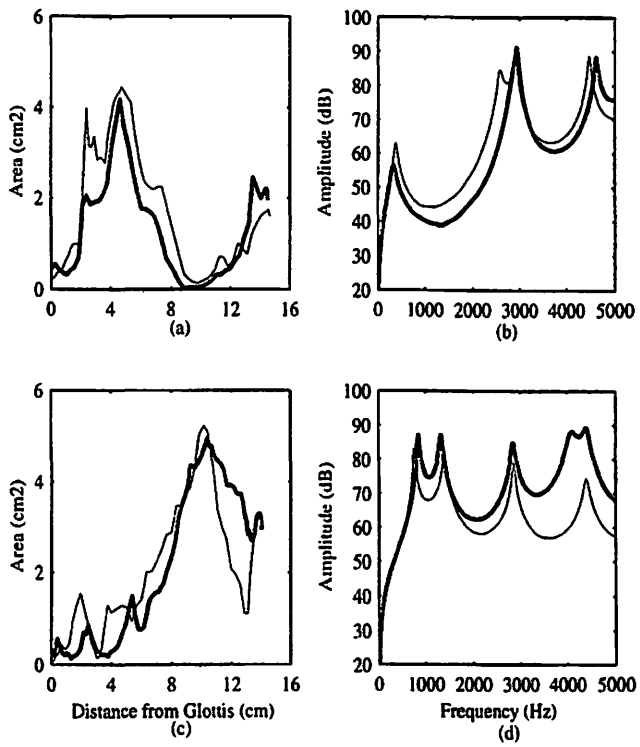


Figure 13. Comparison of /i/ and /ɑ/ acquired using MRI (thin line) and EBCT (thick line): a) area functions for /i/, b) frequency response functions for /i/, c) area functions for /ɑ/, d) frequency response functions for /ɑ/.

The /ɑ/ area function from EBCT is about 0.5 cm longer than that from MRI, resulting in a shift of the most prominent peaks and valleys toward the lips. The areas in the pharynx are larger for the MRI version than for the EBCT, but the maximum areas for both are about 5 cm^2 and occur in the 10 to 10.5 cm range. Following the maximal area, both area functions decrease to an oral cavity minimum before increasing again to the lip termination, but the MRI version does so to a greater degree.

The differences observed between the EBCT and MRI versions of the area functions are undoubtedly, in part, due to the use of independent imaging methods. However, the major difference for both vowels, we believe, can be largely explained by compensatory articulation. Notice that for the MRI version of either vowel, the pharynx is slightly greater in area than the EBCT version, but this is countered in the oral cavity by smaller areas than for the EBCT. While the frequency response functions for the MRI and EBCT area functions differ, they maintain similar formant structures appropriate for the given vowel.

Comparison of Area Functions With Previous Imaging Studies (Female Only)

In this section, the area functions reported in the present study will be compared with others reported in several previous publications. In the following discussion, whenever reference is made to "subject DJ" the reader should assume this to mean the area function(s) from the present study.

Figure 14 shows a graphical representation of the discretized area functions (given in Table II) of DJ for the vowels /i, ɑ, u, o/ superimposed with the area functions of one female subject, reported by Yang and Kasuya (1994) for the same vowels. In a very gross sense, the general shape of each vowel is similar across the two studies but large differences are apparent. For all of the vowels, the Yang and Kasuya area functions show much larger areas at the glottal end (epilarynx) than for the DJ versions. The comparison of the male area functions from Yang and Kasuya with those from Story et al. (1996b) showed the same characteristic (see page 551 of Story et al. (1996b)), indicating that this may not be a subject difference but some variation in the image acquisition or analysis process. In comparison to the DJ area functions, the cross-sectional areas throughout the vocal tract are larger for the Yang and Kasuya versions, especially in the oral cavities of the /ɑ/ and /o/ where areas exceed 9 cm^2. The exception is for /u/ where areas are

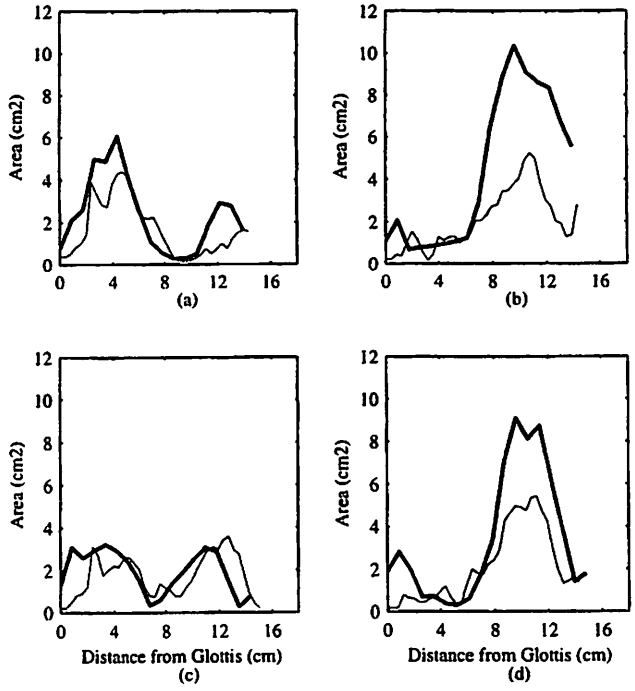


Figure 14. Comparison of the area functions for subject DJ (thin line) with Yang and Kasuya (1994) (thick line): a) /i/, b) /ɑ/, c) /u/, and d) /o/.

comparable. There are slight differences in the vocal tract lengths where they are slightly shorter for the /i/, /ɑ/, and /u/ of Yang and Kasuya while /o/ is slightly longer.

The imaging studies reported by Narayanan (1995), Narayanan et al. (1995, 1997) and Alwan et al. (1997) were concerned with images acquired from four subjects [2 males (MI, SC) and 2 females (AK, PK)] producing fricatives and liquids. However, Narayanan and Alwan (1996) presented an overview paper on imaging where area functions for the three vowels /i, ɑ, u/ were presented, in graphical form, for a male and female subject. The area functions for the female were estimated from these graphs and are shown with the DJ area functions in Figure 15a-c; (it was not indicated whether the female subject was AK or PK). Additionally, Figure 15d-e also shows estimated area functions of female subject AK for the "light /l/" (Narayanan et al., 1997) and the "syllabic /r/" (Alwan et al., 1997) with the /l/ and /ɝ/ of subject DJ. These latter two plots are not direct comparisons since subject DJ was not asked to perform any particular variant as was the case in the other two studies, but they are related closely enough that a comparison is of interest. Again, the gross shape of each area function is similar across studies, but the Narayanan/Alwan set is consistently different in that the length of each tract shape is longer and, like the Yang and Kasuya area functions, the glottal end (epilarynx) contains areas on the order of 1 to 2 cm^2 compared to a range of 0.2 to 0.4 cm^2 for DJ. Except for the /r/ and /l/, the magnitude of the areas of DJ and the Narayanan/

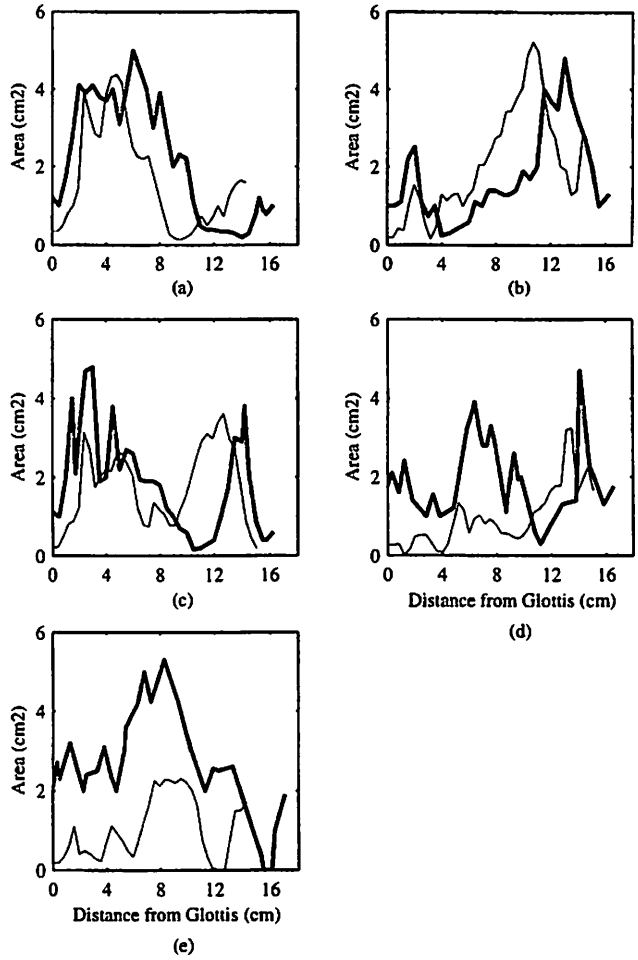


Figure 15. Comparison of the area functions for subject DJ (thin line) with estimated area function from Narayanan and Alwan (1996), Narayanan et al. (1997), and Alwan (1997): a) /i/, b) /ɑ/, c) /u/, d) /ɝ/ and "syllabic" /r/, e) /l/ and "light" /l/.

Alwan sets are essentially in the same range. The Narayanan/Alwan /r/ and /l/ are quite similar in shape to those of DJ in terms of expansion and constriction locations but the areas are typically 2 to 3 times larger.

A perplexing difference of the area functions generated in this study and those from the other studies is the dissimilarity in the epilaryngeal region. Since two different imaging methods (MRI and EBCT) were used in the present study and produced essentially the same areas in this region (see Fig. 13) there is reasonable confidence that such small areas are indeed possible. To test the effect of an increased epilaryngeal area, the frequency responses of DJ's /i/, /ɑ/, and /u/ were recomputed with the epilaryngeal area in each case increased to about 1 cm^2. In all cases, the new formants produced worse matches to the recorded speech than the originals given in Table IV, thus providing more support that these areas are reasonable. However, Yang and Kasuya (1994) also compare computed formants, based on their area

functions, with those from recorded speech and found close matches between the two, in many cases the error was less than five percent.

Comparison to the Adult Male Data of Story et al. (1996)

Much of the comparison between male and female vocal tract shapes has been concerned with ratios of pharyngeal and oral cavity lengths to the total vocal tract length. To make a connection with the previous studies of male/female vocal tract differences, length measurements and resulting ratios for the female subject are compared to similar measurements of the male subject in Story et al. (1996b).

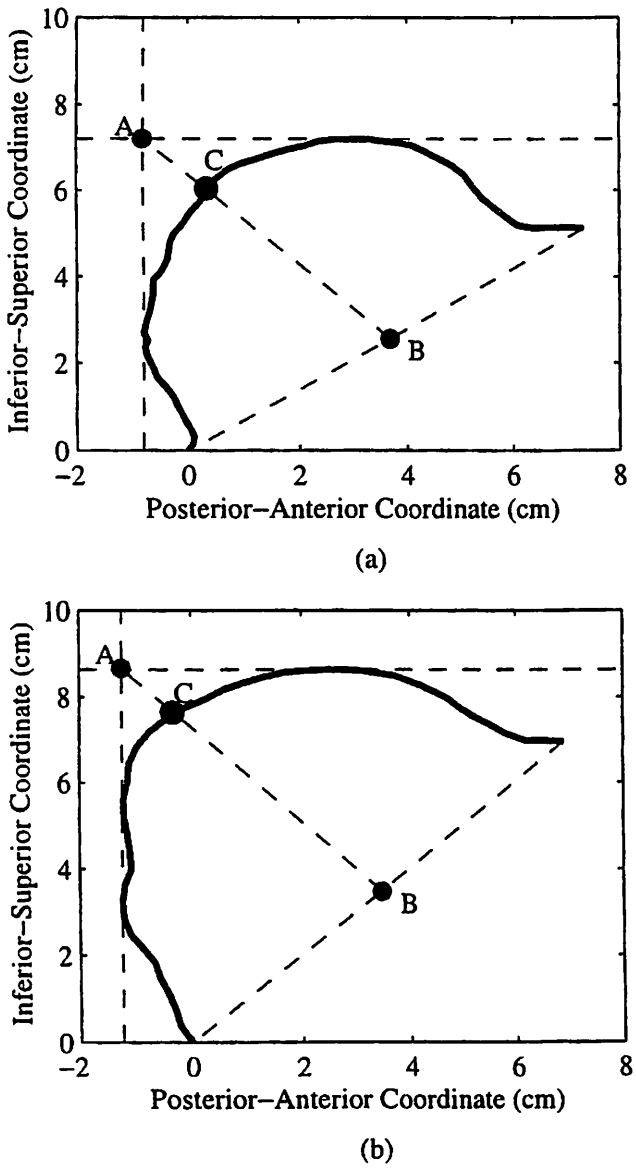


Figure 16. Demonstration of the method used to divide the vocal tract into pharyngeal and oral sections. The thick solid line is the vocal tract profile for /i/. The dashed lines divide the tract as explained in the text where C is the dividing point between pharynx and oral cavity; a) female, b) male.

However, in Story et al., length measurements were not presented, thus the data for male in this section is also previously unpublished.

Initially the tip of the uvula was to be used as an anatomical landmark to divide the tract into pharyngeal and oral sections. However, the precise location of the uvula tip was often difficult to determine because of the image slice thickness. Also, there is some question as to how the uvula location in the supine position compares with that in a upright position. Thus, the method described below was developed to provide a consistent means of dividing the vocal tract.

The length measurements of the pharynx and oral cavity for each vocal tract shape were made on a sagittal projection of the airway centerline which effectively gives a vocal tract "profile" (a 3-D centerline results from the imaging analysis process). The thick solid line in Figures 16a&b are example profiles of the vowel /i/ for female and male, respectively. The point at coordinate (0,0) is located just above the glottis and is the same point at which the area function begins; the final section is at the lips. Each profile is divided into the pharyngeal and oral sections by effectively enclosing it within a right triangle and assuming that the division occurs in the "bend" of the vocal tract. The process is as follows. First, a vertical line is drawn tangent to the most posterior point on the profile and a horizontal line is drawn tangent to the most superior point. The intersection of these two lines is denoted as point A Next, a line is drawn from the point at (0,0) (glottis) to the termination point at the

Table V.
Lengths of the Pharynx (Lp), Oral Cavity (Lo), and Complete Vocal Tract (Lt) for Female and Male Tract Shapes

(All measurements are in centimeters.)

	..... Female (DJ)			..... Male (BS)		
Phoneme	Lp	Lo	Lt	Lp	Lo	Lt
i	6.21	8.45	14.66	7.58	8.51	16.09
ɪ	5.26	8.18	13.44	8.32	8.09	16.41
ɛ	4.89	8.19	13.08	7.49	8.42	15.91
æ	3.85	8.42	12.27	8.06	8.26	16.32
ʌ	5.00	8.31	13.31	8.69	8.27	16.96
ɑ	5.03	8.52	13.55	8.68	8.62	17.30
ɔ	4.85	9.06	13.91	8.42	9.25	17.67
o	4.86	9.08	13.94	8.85	8.39	17.24
ʊ	6.03	8.43	14.46	8.69	8.70	17.39
u	6.70	8.74	15.44	9.12	8.90	18.02
ɝ	5.99	9.08	15.07	8.63	8.56	17.19
l	5.37	8.96	14.33	8.05	8.84	16.90
m	5.27	6.50	11.77	8.33	8.15	16.48
n	5.54	8.73	14.26	7.35	7.97	15.32
ŋ	5.69	8.12	13.81	8.16	8.31	16.46
p	5.29	7.62	12.91	8.15	6.78	14.93
t	4.80	8.18	12.98	7.91	8.29	16.20
k	4.44	8.38	12.82	7.96	9.27	17.24
mean	5.28	8.39	13.67	8.25	8.42	16.67

lips; this line is then bisected which is indicated by the large solid dot and is denoted as point B. A final line is extended from A to B and the point at which this line crosses the vocal tract profile is considered the division between pharyngeal and oral sections (point C). The length of each section is then computed as the cumulative length from glottis to point C for the pharynx and from point C to the lips for the oral cavity.

Table V shows the length of the pharynx, oral cavity, and complete vocal tract for the 18 tract shapes of both female and male subjects. Table VI gives the ratios of pharynx and oral cavity length to the total vocal tract length. From these two tables, it is observed that, for the female, the pharynx, with a mean length of 5.28 cm, is consistently shorter than the oral cavity length which has a mean of 8.39 cm. Table VI indicates that, on average, the pharynx accounts for 39 percent of the total tract length to 61 percent for the oral section. In contrast, the mean pharyngeal and oral section lengths for the male are nearly equal (pharynx - 8.25 cm, oral - 8.42 cm), thus each section section accounts for about 50 percent of the total length.

The total vocal tract length for the female is typically 2-3 cm shorter than the male. Such a finding is entirely expected, but it is interesting that the oral cavity lengths for both male and female are very nearly the same; 8.39 cm and 8.42 for female and male, respectively. Thus, the primary difference between the female subject in this study and the male subject of Story et al. (1996), at least in terms of vocal tract length, is that the pharynx is shorter for the female.

Table VI.
Ratios of Pharyngeal Length to Total Tract Length (Lp/Lt) and Oral Cavity Length to Total Tract Length (Lo/Lt) for Female and Male

	.....Female (DJ)		..... Male (BS).....	
Phoneme	Lp/ Lt	Lo/ Lt	Lo/ Lt	Lo/ Lt
i	0.42	0.58	0.47	0.53
ɪ	0.39	0.61	0.51	0.49
ɛ	0.37	0.63	0.47	0.53
æ	0.31	0.69	0.49	0.51
ʌ	0.38	0.62	0.51	0.49
ɑ	0.37	0.63	0.50	0.50
ɔ	0.35	0.65	0.48	0.52
o	0.35	0.65	0.51	0.49
ʊ	0.42	0.58	0.50	0.50
u	0.43	0.57	0.51	0.49
ɚ	0.40	0.60	0.50	0.50
l	0.37	0.63	0.48	0.52
m	0.45	0.55	0.51	0.49
n	0.39	0.61	0.48	0.52
ŋ	0.41	0.59	0.50	0.50
p	0.41	0.59	0.55	0.45
t	0.37	0.63	0.49	0.51
k	0.35	0.65	0.46	0.54
mean	0.39	0.61	0.50	0.50

These findings fit well with the previously reported data of Goldstein (1980), King (1952), H"ogberg (1995), and Yang and Kasuya (1995). In all of these studies the female pharynx is consistently shorter than the male.

Conclusions

MRI and EBCT have been used to volumetrically image the vocal tract airway of one female subject for 12 vowels, 3 nasals, and 3 plosives. The 3-D image sets were segmented to extract the airway which in turn was analyzed to find the cross-sectional area as a function of the distance from the glottis (along the long axis of the vocal tract). The area functions provide a speaker-specific inventory of vocal tract configurations that can be used as a starting point for studying vocal tract shape and resultant acoustic characteristics as well as providing input for a speech simulation system. The frequency response of each vowel-like area function was computed with a vocal tract model based on one-dimensional acoustic wave propagation. The computed formants were compared to those extracted from the recorded natural speech of the subject. Results indicated that the formant locations were reasonably well represented for the first two formants, but a few vowels appeared to have been centralized. The third formant was poorly matched to the recorded speech. Since the acquisition of the MR image set representing each vocal tract shape required many repetitions (approx. 30 repetitions), the area functions need to be considered as an "average" shape for a particular vowel or consonant.

The speaker-specificity of the area function set is an advantage because acoustic pressures and flows generated from speech modeling can be directly compared to experimental data collected from the same subject that was imaged. Also, the vocal tract characteristics that partially comprise the "acoustic signature" of a person are retained in this set, allowing for future study of speaker individuality. However, the disadvantage of speaker-specificity is that the reported information strictly applies to only the imaged subject and hence no general statements can be made regarding vocal tract shapes across a population. The comparisons of the area functions reported here with those from previous studies is a start in understanding differences across subjects but more imaging is needed in the future to expand the pool of data regarding vocal tract morphology.

The area function set presented here provides a direct female compliment to the adult male set presented in Story et al. (1996b). The specific vocal tract shapes, imaging methods and processing procedures were exactly the same for the both the male and female image sets. Thus, future work should include a thorough comparison of vocal tract length and area scaling between the male and female and then be compared to previous work on uniform and nonuniform scaling principles (e.g. Fant, 1975). Certainly of

interest is the issue of vowel normalization for purposes of phonetic recognition, but of equal interest is the question of speaker individuality. In particular, what contribution does the vocal tract make that allows, for example, a female to be first differentiated from a male and then differentiated from other females. A preliminary study comparing three male and female area functions was described by Story, Hoffman, and Titze (1997).

Measurements of the nasal tract, trachea, and fricative consonants will be presented in the future to augment the data set presented in this paper. Additional work is needed to collect vocal tract shape inventories for more subjects (and for languages other than American English) as that will strengthen the knowledge base pertaining to the general population as well as provide opportunity to study the individual characteristics of many people.

Acknowledgments

The authors would like to thank Steve Baker for his technical expertise and willingness to do late night scanning as well as the staff of the Division of Physiologic Imaging at the University of Iowa for allowing generous use of equipment and software. This study was supported by grant No. R01 DC02532-02 from the National Institutes on Deafness and Other Communication Disorders.

References

1. Alwan, A. A., Narayanan, S. S., and Haker, K., (1997). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The rhotics," JASA, 101(2), 1078-1089.

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

3. Botherel, A., Simon, P., Wioland, F., and Zerling, J.P., (1986). Cineradiographie des voyelles et consonnes du franôais, Travaux de l'institut de phonetique de Strasborg, Strasborg.

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

5. Carmody, F. J., (1941). "An x-ray study of pharyngeal articulation," Univ. of California Publications in Modern Philology, 21(5), 377-384.

6. Crary, M. A., Kotzur, I. M., Gauger, J., Gorham, M., and Burton, S., (1996). "Dynamic magnetic resonance imaging in the study of vocal tract configuration," J. Voice, 10(4), 378-388.

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

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

9. Fant, G., (1966). "A note on vocal tract size factors and non-uniform F-pattern scalings," STL-QPSR, 4, 22-30.

10. Fant, G., (1975). "Non-uniform vowel normalization," STL-QPSR, 2-3, 1-19.

11. Goldstein, U. G., (1980). An articulatory model for the vocal tracts of growing children, Doctoral dissertation, Department of Electrical Engineering and Computer Science, MIT.

12. Heinz, J. M., and Stevens, K. N., (1965). "On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech," 5th Congres International D'Acoustique, Liege, 7-14 September 1965, 1-4.

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

14. Hogberg, J., (1995). "From sagittal distance to area function and male to female scaling of the vocal tract," STL-QPSR, KTH, 4, 11-53.

15. Holbrook, R. T., and Carmody, F. J., (1937). "X-ray studies of speech articulations," Univ. of California Publications in Modern Philology, 20(4), pp. i-viii + 187-238.

16. King, E. W., (1952). "A roentgenographic study of pharyngeal growth," Angle Orthodontist, 22(1), 23-37.

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

18. Lindblom, B. E. F., and Sundberg, J., (1987). "Acoustical consequences of lip, tongue, jaw, and larynx movement," JASA, 50(4), 1166-1179.

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

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

21. Mrayati, M., Carre, R., Guerin, B., (1988). "Distinctive regions and modes: A new theory of speech production," Speech Communication, 7, 257-286.

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

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

24. Narayanan, S. S., Alwan, A. A., and Haker, K., (1997). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals," JASA, 101(2), 1064-1077.

25. Nordstrom, P-E, (1975). "Attempts to simulate female and infanct vocal tracts from male area functions," STL-QPSR, KTH, 2-3, 20-33.

26. Nordstrom, P-E, (1976). "Female and infant vocal tracts simulated from male area functions," J. Phonetics, 5(1), 81-92.

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

28. Pykett, I.L., and Rzedzian, R.R. (1987). "Instant images of the body by magnetic resonance," Magnetic Resonance in Medicine, 5(6), 563-571.

29. Sundberg, J., Johansson, C., Wilbrand, H., and Ytterbergh, C. (1987). "From sagittal distance to area," Phonetica, 44, 76-90.

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

31. Story, B. H., Hoffman, E. A., and Titze, I. R. (1995)."Speech simulation based on MR images of the vocal tract," SPIE Medical Imaging-Physiology and Function from Multidimensional Images, 2433, 179-190.

32. Story, B. H., Hoffman, E. A., and Titze, I. R. (1996a)."Vocal tract imaging: A comparison of MRI and EBCT," SPIE Medical Imaging - Physiology and Function from Multidimensional Images, 2709, 209-222.

33. Story, B. H., Titze, I. R., and Hoffman, E. A., (1996b). "Vocal tract area functions from magnetic resonance imaging," JASA, 100(1), 537-554.

34. Story, B. H., Hoffman, E. A., and Titze, I. R., (1997)."Volumetric image-based comparison of male and female vocal tract shapes," SPIE Medical Imaging - Physiology and Function from Multidimensional Images, 3033, 25-37.

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

36. Titze, I. R., Wong, D., Lange, R., and Story, B. H., "Comparison of three techniques for voice transformation," 1st ESCA Tutorial and Research Workshop on Speech Production Modeling - 4th Speech Production Seminar, Autrans, France, 20-24 May, 215-220.

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

38. Yang, C-S, and Kasuya, H., (1995). "Uniform and non-uniform normalization of vocal tracts measured by MRI across male, female, and child subjects," IEICE Trans. Inf. and Syst., Vol. E78-D, No. 6, 732-737.

NCVS Status and Progress Report - 12
June 1998, 115-119

Toward Occupational Safety Criteria for Vocalization

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

Abstract

Setting safe limits for sound exposure to humans has been an active part of audiology. Little has been done however, to address safety limits on the sound production side. Vocologists should become active in limiting competing noises when vocal communication is required or desired. In restaurants, public transportation vehicles, convention halls, bars and other places of business or recreation, there is often a disregard for the loudness level of conversation required. For some, there is also insufficient protection for the duration of occupational vocalization and a lack of a built in recovery time. Some suggestions for voice dosimetry measures are given.

Introduction

Much is known about the potential damage excessive exposure to sound can do to the hearing mechanism (9). Comparatively little is known, however, about the dangers of self-generated vibration in the body tissues. It is generally accepted by voice clinicians that loud and prolonged phonation can lead to voice disorders such as vocal nodules or chronic inflammation, but the precise course of self-inflicted injury by phonation is not yet known. Can vocalization expose the vocal folds to too much vibration and collision that leads to trauma and, ultimately, to irreparable damage? A laboratory investigation on canine larynges, hyperphonated *in vivo* under anesthesia, by Gray, Titze and Lusk (7) showed that damage to the vocal fold epithelium was extensive after several hours of continuous loud phonation. More recent investigations show that the basement membrane attachment of the skin to the lamina propria is a site of injury (4,5,6). In this paper, we examine some of the reasons why safe limits of vibration (phonation time) may need to be set to protect people in certain occupations. We first identify a few places where safe limits of vocalization may be exceeded by all of us, often against our wish.

Places Where Loud Music or Competing Vocalizations Tax Our Own Vocalizations

The following is a short list of places where conversations are often held at high loudness levels: restaurants, convention halls, cocktail lounges, bars, bowling alleys, pool halls, dance halls, airplanes, trains, subways, sports arenas.

In restaurants, background music is often played at high loudness levels. To guard against speaking too loudly, one can first ask the waitress to find a table away from the loudspeakers, which hopefully is also a non-smoking table. Secondly, with some courage, one can ask the waitress to go to the proprietor or manager and ask if the music could be turned down. Often the management will not comply with the request, but it is worth a try.

In convention halls, there is much less one can do, especially if the display booths and tables are only a few feet apart and a lot of people are gathered all around. Exhibitors can amplify their own voices, but that may be to the detriment of those who are unamplified. The use of sound absorbing and isolating material is the best answer.

Cocktail lounges and bars are places where many voices are potentially at risk. Often there is a high density of people in a small room. There is seldom a consideration for acoustic treatment of sound reflecting surfaces. Bar gatherings and cocktail parties can be avoided, of course, and I routinely do so, but for some people, it's a place of business; avoidance may not be the only answer.

Many recreational places present a similar problem. I went to a bowling alley recently; it was one of those days when they featured stroboscopic lights and rock and roll music of the 50's. My sole intent for going was to have a long, relaxed conversation with my brother-in-law. The music was so loud that we communicated more or less with hand signals and lip reading. I went to the proprietor and asked him to turn down the music. He did, but within five

minutes the sound level was back to where it started because someone else made the opposite request.

In public transportation venues such as airplanes, trains and subways, some people have given up on anything but brief conversations. Reading and slumbering seem to be the activities of choice.

What can be done about these situations? Vocologists need to become active in leading the anti-speaking-in-noise campaign. The anti-smoke campaign and the anti-noise campaign have led to less pollution in public places. We need to show lawmakers that loud-voice pollution carries with it a double jeopardy --- to hearing and to vocalizing.

Occupations for Which There is Little Relief

Now, let's look at *length* of exposure to ones own vibration. There are some occupations for which there is little relief from continual phonation. Probably the largest occupational group is school teachers, which constitute about 4% of the workforce (12). Elementary school teachers typically have six to seven hour days of lecturing, conversing with, and disciplining children. They do this five to six days a week, probably exceeding safe limits of tissue vibration and, more importantly, getting little time for recovery.

This is even more acute for telephone marketers because they get only an occasional five minute break between talking on the telephone. In some offices they don't even get to look up the number. Somebody feeds them the next number and they dial immediately after hanging up. A key issue is whether the conversation is closer to a monologue or a dialogue. In a dialogue, there is a chance for short-term vocal recovery, but for a monologue there is almost none.

In professions such as counseling there is more of a dialogue situation, but the length of the talking day (probably five to eight hours) is again very taxing. The same can be said for travel agents and ticket reservation agents. For these vocal endurance people, we need to become active in setting guidelines for appropriate recovery times. Let's begin by taking a trip to the athletic arena; this will give us some valuable clues.

Recovery Time Needed for Athletes

Appropriate recovery times have been worked out for athletes who stress their tissues hard. A typical professional basketball player plays every two to three days. Baseball pitchers pitch once every three to four days. Football players and soccer players play hard once a week and do lighter workouts between games. In contrast, title contending boxers compete once or twice a year. Marathon runners also don't run marathons every week. The probable reason these variable rest periods have been established by trainers for athletes is that the recovery time needed is directly proportional to the amount of localized tissue injury that has occurred. If there is acute damage to joints, ligaments, tendons, and other connective tissue, it may take days, weeks or months to be ready again. But if there is only general muscle fatigue, the recovery period is probably only a day or two.

Figure 1 illustrates a series of hypothetical curves for tissue injury and recovery period. The curves are not intended to be quantitively meaningful, but they are useful for conceptualization and may some day have quantitative validity. Let's begin by identifying a performance time. This performance time could be a 50 minute lecture, an hour recital, a two hour auction, a full day of counseling, or any other engagement that has only brief rest periods between a structured vocal delivery. Within this performance time, we can measure the actual phonation time with a voice accumulator (2).

Short-term recovery takes place whenever we stop phonating, even momentarily. This recovery is primarily a benefit to the muscles, whose chemistry gets reset for the next contraction. Meanwhile, skin cells and extracellular material in the vocal fold cover are not quickly repaired. Epithelial cells may have been heavily bombarded by vocal fold collision; consequently, they may die and be shed off. New cells will develop underneath, but that takes time. There may also be some destruction of the basement membrane region, where the skin attaches to the lamina propria, as previously mentioned. Some collagen and elastin fibers may have separated from the structural matrix of the lamina propria, now floating freely; they will eventually have to be

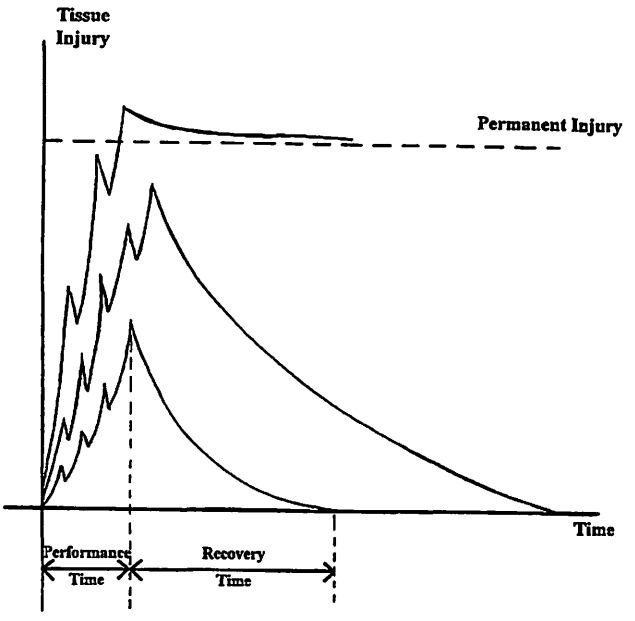


Figure 1. Hypothetical tissue injury and recovery curves for human phonation.

removed and replaced. This is accomplished by the fibroblasts, the few cells that reside within the extracellular tissue and constantly sample the state of the tissue. Old detached protein debris will be removed and reused by the fibroblasts to make new protein fibers that support the connective tissue structure. It's like worker bees constantly repairing the honeycomb structure of a beehive. The repair process may take anywhere from a few hours to as much as 72 hours to complete.

Thus, minor destruction and repair is continual, even when we don't abuse our vocal folds. The key question is: can the regenerative processes keep up with the destructive process? If not, a recovery time will be needed because an accumulation of damage will have taken place.

The obvious hypothesis is that the greater the accumulation of injury, the longer the recovery time will need to be. For a school teacher, when one lecture follows another, injury will accumulate during the day. A point can be reached where there literally is never enough recovery from day to day. Some permanent injury will then occur that is manifest as a voice pathology.

The scientific question is: what are the physiologic time constants in the above processes. In any field of science, one begins by making some empirical observations with a few simple measures.

Voice Dosimetry Measures

We propose the following dosimetry measures. They are for discussion only and will likely see considerable evolution in the next few years of application.

Accumulated Phonation Time Index:

$$I_p = \frac{phonation\ time}{8\ hour\ work\ time} \quad (1)$$

Short-Term Recovery Time Index:

$$I_{rs} = \frac{performance\ time - phonation\ time}{performance\ time} \quad (2)$$

Long-Term Recovery Time Index:

$$I_{rl} = \frac{days\ of\ rest}{days\ of\ performance} \quad (3)$$

Personalized Loudness Index:

$$I_L = \int SPL(t)dt / SPL_m \quad (SPL_m\ from\ Voice\ Range\ Profile) \quad (4)$$

Vocal Economy Index:

$$I_e = \frac{acoustic\ power\ at\ mouth}{power\ absorbed\ in\ vocal\ folds} \quad (5)$$

Some data already exist on accumulated phonation time. Holbrook (10) found that an elementary school teacher that teaches about six to seven hours a day will phonate somewhere between one to two hours. If we normalize this to an 8-hour work day, we would get a phonation index I_p of about 0.1-0.2. If the teacher has two hours of preparation time (or otherwise quiet time), then the performance time is six hours and the short-term recovery index I_{rs} is on the order of 0.6 to 0.8. The long-term recovery index I_{rl} would be about 0.7 for a five day work week.

Now consider the personalized loudness index (Equation 4). This is based on an individual's voice range profile (VRP), a plot of intensity range versus fundamental frequency. If a microphone signal can be calibrated in terms of vocal intensity, then this intensity can be integrated over some period of time and divided by the maximum intensity that has been established by the VRP. Buekers et al. (12) measured the intensities of phonation for various voice professionals with a voice accumulator. They found that sports instructors used the greatest intensities. These could be normalized to their own VRP to separate general gender or age affects.

Another index that is currently being worked on is the vocal economy index (1,14). This is a measure of acoustic power radiated at the mouth divided by the power absorbed in the vocal folds. But measurement of the power absorbed by the vocal folds is not easy. At this point, some estimates of contact stress have been obtained empirically (13), but the precise relationship between contact stress and power absorbed is still under investigation (1).

An Example of Comparative Recovery Indexes

Let's put to the test the short and long-term recovery indices, comparing vocalists with athletes. Boxers are in the ring for three minute rounds and then rest for one minute between rounds. The short-term recovery index is then

$$I_{rs} = \frac{4\ minutes - 3\ minute\ round}{4\ minutes} = 0.24 \quad , (6)$$

a relatively low number.

The long-term recovery index for a boxer who fights only twice a year is

$$I_{rl} = \frac{180\ days - 1\ day}{180\ days} = 0.99 \quad , \quad (7)$$

a relatively high number. Thus, for a boxer, the two recovery indices are near the extremes. The boxer gets very little short-term recovery but an extensive long-term recovery, suggesting that considerable tissue damage is done during the performance.

For a professional basketball player in the U.S., the numerous time-out periods have made it possible to stretch 20 minutes of actual playing time for the average player into about two hours. The short-term recovery index is then

$$I_{rs} = \frac{120\ minutes - 20\ minutes}{120\ minutes} = 0.83 \quad , (8)$$

which is much higher than for the boxer. For this reason, basketball players can be on the court every other day. Their long-term recovery index is relatively low, however

$$I_{rl} = \frac{7\ days - 3\ days}{7\ days} = 0.57 \quad , \quad (9)$$

suggesting that fatigue may set in over the long haul and an off-season is required.

Now let us look at some vocal athletes. For a telephone marketer who works eight hours and speaks two words for every one of the customer (probably a low estimate), the short-term recovery index is on the order of

$$I_{rs} = \frac{480\ minutes - 320\ minutes}{480\ minutes} = 0.3 \quad (10)$$

This is almost as low as for the boxer. In other words, telephone marketers could get beaten up vocally every day. But as Buekers et al. (2) points out, their vocal intensities can be kept low because they wear head-mounted microphones. The main problem is that they get little long-term recovery.

$$I_{rl} = \frac{7\ days - 5\ days}{7\ days} = 0.28 \quad (11)$$

Teachers have the same low long-term recovery index, but their short-term recovery index is larger (we calculated I_{rs}=0.6-0.8 for teachers who phonate 1-2 hours in a normal school day). The problem is that teachers don't generally wear head microphones, which means that their intensities are higher than those of the telephone marketers. What separates these vocal athletes is that they either miss the short-term or the long-term recovery.

Opera singers, on the other hand, have learned how to take care of themselves. In the title role of Tosca, the soprano phonates for 23 minutes (measured by a stop watch), quite similar to average playing time of a basketball player. If the opera lasts three hours, then

$$I_{rs} = \frac{180\ minutes - 23\ minutes}{180\ minutes} = 0.87 \quad , (12)$$

again quite similar to that of the basketball player. For the professional singer who rehearses or performs about three times a week, the long-term recovery index is also similar to that of the basketball player.

The Effect of Vibration on the Human Body

Do we have any hard data to suggest that vibrating tissues actually does damage? The reverse has actually been argued, that vibration is good for the body (3). On the other hand, the *Handbook of Human Vibration* (8) contains lots of data about the possible dangers of vibration on the human body. The topic is divided into whole-body vibration and hand-transmitted vibration. The first applies to people subjected to vibration in vehicles and the second to people working with power tools. There are American standards, British standards, and French standards. Continuous vibration must not exceed certain tissue acceleration limits at certain frequencies. Figure 2 illustrates one of the British standards. Note the region of acceptable and unacceptable vibration amplitudes. The in-between region is a region of concern, but not necessarily one of danger. Tissue acceleration in m^2/s is plotted against frequency. Both variables are on a logarithmic scale. Excessive vibration can affect the vascular system, the neural system, the cells, and the extracellular matrix. Zeitels (15) has claimed that vascular hemorrhages can appear on top of the vocal fold due to rapid acceleration and deceleration of blood vessels. This acceleration and deceleration is associated with the mucosal wave propagating on the surface of the vocal folds. Let's take an example. Say we're phonating at a 100 Hz, which is a low F_0 for males. To be in the acceptable tissue vibration range in Figure 2, acceleration magnitudes must be below about 10 m/s^2, or about one unit of gravitational acceleration (9.8 m/s^2). So, it's safe to be exposed to 1 g of acceleration at speech frequencies. We will now calculate what typical acceleration are in voice production.

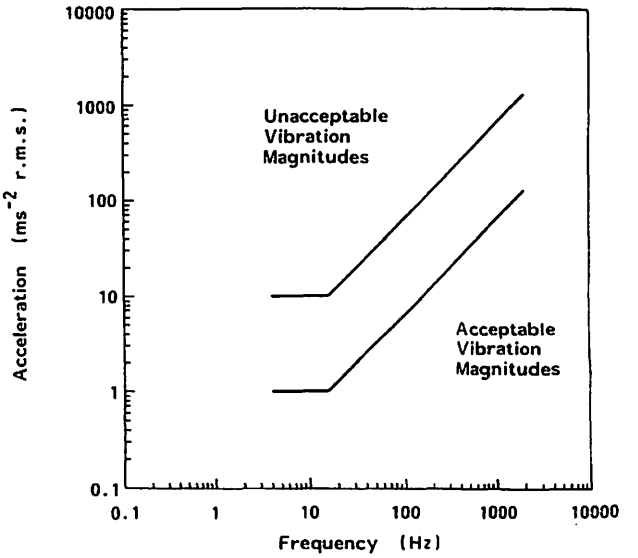


Figure 2. Recommendations for acceptable and unacceptable vibration magnitudes as a function of frequency (after Griffin, 1990).

Estimates of Vocal Fold Tissue Acceleration

If we assume perfect sinusoidal tissue vibration, the acceleration can be written as

$$a = \omega^2 A \quad , \quad (13)$$

where ω is the angular frequency and A is the vibrational amplitude. For a fundamental frequency F_o of 100 Hz and a vibrational amplitude of 2 mm, we get

$$\begin{aligned} a &= 4\pi^2 F_o^2 A \\ &= 4\pi^2 (100)^2 (0.002)\, m/s^2 \\ &\approx 790\, m/s^2 \end{aligned} \qquad (14)$$

Alternatively, if we calculate the acceleration with Newton's law, using a measured force of contact between the vocal folds and an estimation of the mass that is being decelerated in the collision process,

$$a = \frac{F}{m} = \frac{\sigma_c LT}{\rho LTD}, \qquad (15)$$

where F is the contact force, m is the mass in vibration, L is the vocal fold length, T is the vocal fold thickness, D is the vocal fold depth in vibration, σ_c is the contact stress, and ρ is the tissue density. The factor LT cancels out so that we have

$$a = \frac{\sigma_c}{\rho D} \qquad (16)$$

Jiang and Titze (11) have measured the contact stress by colliding one vocal fold with a glass plate, which contained an embedded stress transducer. They also measured the time interval over which the collision takes place. It was on the order of 1 to 2 ms. More recently, Verdolini et al. (13) have placed a pressure transducer between the vocal folds in a live human subject to measure the contact stress. In both studies the contact stress was on the order of 2-5 kPa. Taking the conservative estimate of 2 kPa in equation 16, with a tissue density of 1140 kg/m^3 and a depth of tissue vibration of 2 mm, we get

$$a = \frac{2000\, Pa}{\left(1140 \frac{kg}{m^3}\right)(0.002)\, m} \approx 877\, m/s^2 \qquad (17)$$

These values are far above the safe limit. What bails us out is the fact that vocal fold vibration is not continuous. But if we make the assumption that the effects of vibration are accumulative, then we have to conclude that for every second of phonation, we should have about 10-100 seconds of recovery. In other words, we should limit the period of phonation in the total performance time by a factor of 10-100. In terms of our recovery indices, then, ideally I_{rs} should be on the order of 0.9 to 0.99. If this index value is not reached, as in the case of the telephone marketers and teachers, a long-term recovery period must be built in. The most vocal teachers should possibly alternate with librarians or other less vocal instructors to get longer periods of recovery.

Conclusions

Preliminary calculations suggest that damage risk from tissue vibration is exceeded by professional vocalists such as telephone marketers and school teachers. The risk appears to be less for opera singers in terms of pure accumulated phonation time. The type and loudness of phonation was not considered in this calculation, but can obviously play a huge role. Vocologists are asked to become activists in finding relief for the *at risk* populations, either in terms of structuring longer recovery times or limiting the daily periods of phonation.

Acknowledgements

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

References

1. Berry D, Verdolini K, Chan R, Titze I. Indications of an optimum glottal width in vocal production. J Speech Lang Hear Res in press.

2. Buekers R, Bierens E, Kingma H, Marres EHMA. Vocal load as measured by the voice accumulator. Folia Phoniatr Logop 1995; 47: 252-261.

3. Campbell D. The Mozart Effect. New York, NY: Avon Books; 1997.

4. Courey M, Shohet J, Scott M, Ossoff R. Immunohistochemical characterization of benign laryngeal lesions. Ann Otol Rhinol Laryngol 1996; 105:6-11.

5. Gray SD. Basement membrane zone injury in vocal nodules. In: J Gauffin, B Hammarberg (Eds.), Vocal Fold Physiology. San Diego: Singular Publishing Group Inc.; 1991, 21-27.

6. Gray SD, Hammond E, Hanson DF. Benign pathologic responses of the larynx. Ann of Otol Rhinol Laryngol 1995; 104(1): 13-18.

7. Gray SD, Titze IR, Lusk RP. Electron microscopy of hyperphonated vocal cords. J Voice 1987; 1(1): 109-115.

8. Griffin MJ. Handbook of Human Vibration. San Diego: Academic Press; 1990.

9. Hammermik R, Henderson D, Salvi R. (Eds.). New Perspectives on Noise-Induced Hearing Loss. New York: Raven Press; 1982.

10. Holbrook A. Instrumental analysis and control of vocal behavior. In: M Cooper (Ed.), Approaches to Vocal Rehabilitation. Springfield: Thomas; 1977.

11. Jiang J, Titze I. Measurement of vocal fold intraglottal stress and impact stress. J Voice 1994; 8: 132-144.

12. Titze IR, Lemke JA, Montequin D. Populations in the U.S. workforce who rely on voice as a primary tool of trade. J Voice 1997; 11(3): 254-259.

13. Verdolini K, Hess M, Titze I, Bierhals W, Gross M. Investigation of vocal fold impact stress in human subjects. J Voice in press.

14. Verdolini K, Titze I. The application of laboratory formulas to clinical voice management. Am J Speech Lang Path 1995; 4: 62-69.

15. Zeitels S.M. Personal communication 1998: Voice Care Symposium, Philadelphia.

NCVS Status and Progress Report - 12
June 1998, 121-134

Toward a Theory of the Dispositional Bases of Functional Dysphonia and Vocal Nodules: Exploring the Role of Personality and Emotional Adjustment

Nelson Roy, Ph.D.
Diane M. Bless, Ph.D
The Waisman Center, The University of Wisconsin-Madison

Introduction

The human larynx is acutely responsive to sudden changes in affective states, and as such it has been labeled a "barometer of emotions." Others have described the voice as one of the most characteristic expressions of the individual -- "a mirror of personality" (Aronson, 1990; Brodnitz, 1962). Thus, when the voice becomes disordered, voice scientists and clinicians sometimes offer emotional or personality factors as likely causal explanations. However, considerable controversy surrounds when and whether such factors should be considered causal, concomitant, outcomes, or completely irrelevant. This debate is especially intense in the case of such disorders as functional dysphonia (FD) and vocal nodules (VN). This chapter will review some of the prevailing concepts, issues, controversies, and research literature surrounding the role of psychological and personality processes in individuals with functional dysphonia and vocal nodules. Moreover, a theory is offered that identifies personality factors as important considerations in the development of these voice disorders.

The extant literature is replete with speculations linking the disorders of FD and VN to psychological precursors and personality variables. It has been suggested that personality and emotional maladjustment contribute to, or are primary causes of these voice disorders, and that these voice disorders in turn create psychological problems or personality effects (Cooper, 1973; Doyle, 1994). Unfortunately, both the anecdotal and scientific literature in this area are dominated by dataless speculation and bias. These biases usually reflect the clinician-researcher's general attitudes of acceptance or willingness to consider psychological or emotional factors in voice disorders. Clinicians holding more global theories of psychosomatic causation seem more willing to connect patients' symptoms to psychosocial factors; any factors that correlate with the theory's explanatory constructs are often included as evidence for the psychogenic nature of symptoms. When the attitude toward these factors is one of skepticism, the attribution is made less often. Upon review of the literature, one must admit, as Goodstein (1958) did almost 40 years ago, that much of this literature is unverifiable in a scientific sense. His comments remain germane today, as few advancements have been made over the past several decades in our understanding of the relation between voice disorders, emotion and personality. In the next few sections we will review the available literature linking FD and VN to psychological precursors.

Functional Dysphonia

Problems in Defining Functional Dysphonia

Most authors generally agree that FD refers to a voice disturbance in the absence of visible neurological or structural pathology (Koufman & Blalock, 1982), or where the existing pathology is judged to be insufficient to account for the severity of the dysphonia. FD occurs predominantly in women and commonly follows upper respiratory infection symptoms (Gerritsma, 1991; Friedl, Friedrich & Egger, 1990; Milutinovic, 1991; Kinzl, Biebl & Rauchegger, 1988; Aronson, Peterson & Litin, 1966). It is frequently transient and varies in its response to treatment (Koufman & Blalock, 1982; Bridger & Epstein, 1989; Roy & Leeper, 1993; Fex, F., Fex, S., Shiromoto & Hirano, 1994). Because of imprecision in defining FD, few reliable prevalence and incidence statistics exist.

Functional *dysphonia* and *aphonia* are sometimes regarded as disorders represented on a continuum of vocal severity, and in some cases are believed to share a common

etiology (Aronson et al., 1966; Aronson, 1990). In *aphonia*, patients lose their voice completely and articulate in a whisper, whereas *dysphonia* suggests phonation is preserved, but disturbed in quality, pitch and/or loudness (Boone & McFarlane, 1988). Some authors warn that distinctions must be made between aphonia and dysphonia to prevent overestimation of the role of psychological factors in "dysphonia" (Friedl, Friedrich, Egger & Fitzek, 1993). Whether these vocal conditions simply represent quantitative differences along a single continuous dimension, e.g., laryngeal and extralaryngeal muscle tension, or are categorically and etiologically unique, is open for debate.

Some clinicians object to using the label of "functional dysphonia" because of its etiologic and symptomatologic ambiguity (Morrison & Rammage, 1993; Aronson, 1990; Pahn & Friemart, 1988). "Functional" implies a disturbance of physiological function rather than anatomical structure. In clinical circles, *functional* is usually contrasted with *organic* and often carries the added meaning of *psychogenic*. Stress and psychological conflict are frequently presumed to cause and/or exacerbate functional symptoms (Morrison, Nichol, Rammage, 1986; Bass, 1990; Kirmayer & Robbins, 1991). Theorists' opinions differ, however, concerning the relative contribution of psychological factors to the formation of functional voice disorders.

The role of psychological or personality processes in FD remains enigmatic partly because the term "functional" includes a continuum of medically unexplained voice disorders: psychogenic, conversion, hysterical, tension-fatigue syndrome, hyperkinetic, muscle misuse, or muscle tension dysphonia. It is therefore not understood whether FD is one disorder or many. Although the previous diagnostic labels imply some degree of etiologic heterogeneity, whether these disorders are qualitatively different and etiologically distinct remains unclear. Voice disorder taxonomies have yet to be adequately operationalized; consequently, diagnostic categories often lack clear thresholds or discrete boundaries to determine patient inclusion or exclusion. At the purely phenomological level, there may be little difference between these disorders. This nosological imprecision may account for differences found in the literature regarding the voice-psychology relationship.

Causal Models of Functional Dysphonia

The search for the link between functional voice symptoms and psychosocial factors is guided by theories of causality. Authors have described assorted psychopathological processes that may be active in voice symptom formation.

The dominant psychological explanation for medically unexplained voice loss is the concept of conversion disorder introduced by Freud. (Aronson, 1990; Butcher, 1995; Green & Mathieson, 1989; Stemple, 1984, 1993). Conversion disorder involves unexplained symptoms or deficits affecting voluntary motor or sensory function that suggest a neurological or other general medical condition (Diagnostic and Statistical Manual of Mental Disorders, DSM-IV, 1994). When the laryngeal system is involved, it is referred to as conversion dysphonia or aphonia. The voice loss, whether partial or complete, is often interpreted to have symbolic meaning. In short, patients convert intrapsychic distress into a voice symptom.

Psychological factors are judged to be associated with the voice symptoms because the onset or exacerbation of the dysphonia is preceded by conflicts or other stressors. Primary or secondary gains are thought to play an important role in maintaining and reinforcing the conversion voice disorder. Primary gain refers to anxiety alleviation accomplished by preventing the psychological conflict from entering conscious awareness. Secondary gain refers to the avoidance of an undesirable activity/responsibility and the extra attention or support conferred to the patient.

Conversion disorder has historically been associated with histrionic-hysterical (superficial, melodramatic) personality features. In some cases, individuals with conversion symptoms seem to possess a complacent or serene attitude regarding their voice symptoms. This relative lack of emotional concern, which is incongruous with their loss of voice function, has been referred to as "la belle indifference." More recent descriptions of conversion disorder deemphasize the significance of the symbolic nature of symptoms, hysterical personality traits and la belle indifference (Kirmayer & Robbins, 1991; DSM IV, 1994).

Butcher and colleagues (Butcher, Elias, Raven, Yeatman & Littlejohns, 1987; Butcher, Elias & Raven, 1993; Butcher, 1995) argue that there is little research evidence that hysterical conversion disorder, as defined by Freud, is the most common cause of voice loss unaccounted for by pathological findings. Butcher advises that the conversion label should be reserved for cases of *aphonia* where la belle indifference and lack of motivation to improve the voice coexist with clear evidence of a temporally linked psychosocial stressor. In the place of conversion, Butcher (1995) offers two alternative models to account for psychogenic voice loss. Both models minimize the role of primary and secondary gain in maintaining the voice disorder. The first is a slightly reformulated psychoanalytic model that states: "if predisposed by social and cultural bias as well as early learning experiences, and then exposed to interpersonal difficulties that stimulate internal conflict, particularly in situations involving conflict over self-expression or voicing feelings, intrapsychic conflict or stress becomes channeled into musculoskeletal tension, which

physically inhibits voice production (p.472)". The second model, based on cognitive-behavioral principles, states that "life stresses and interpersonal problems in an individual predisposed to having difficulties expressing feelings or views would produce involuntary anxiety symptoms and musculoskeletal tension, which would center on and inhibit voice production (p.473)." Both models clearly emphasize the inhibitory effects of excess laryngeal muscle tension on voice production, although via slightly different causal mechanisms.

The subject of poorly regulated laryngeal muscle tension is also a topic in the writings of Nichol and colleagues (Nichol, Morrison & Rammage, 1993; Rammage, Nichol & Morrison, 1987; Morrison & Rammage, 1993) and Aronson (1990) among others (Greene & Mathieson, 1989, Colton & Casper, 1996). In addition to acknowledging the conversion explanation for FD, Nichol and associates (1993) proposed that "tensional symptoms arise from the overactivity of autonomic and voluntary nervous systems in individuals who are unduly aroused and anxious (p.644)." They added that such overactivity leads to hypertonicity of the intrinsic and extrinsic laryngeal muscles, resulting in muscle tension dysphonias sometimes associated with adjustment or anxiety disorders, or with certain personality trait disturbances.

Another possible explanation for FD is the interaction between organic and psychogenic mechanisms. One example of this interaction is the "specificity hypothesis" advanced by Alexander (Alexander, 1950; Alexander, French & Pollock, 1968). This theory suggests that a specific stimulus (e.g., emotional conflict) elicits a distinctive response, or illness, and the organ affected (larynx) is determined by a genetic weakness or vulnerability. Milutinovic (1991) recognized the extensive etiologic overlapping of organic and functional voice disturbances and stated that "genetic factors, the state of the endocrine and neurovegetative systems, and psychological factors are significant in the development of functional dysphonia (p.179)." He suggested that psychogenic aphonia and dysphonia should be considered "phononeuroses." Since over half of his "phononeurotic" patients had documented infection of the upper respiratory airways preceding the voice disturbance, he concluded that a direct connection existed between the pathological state of the mucosa and development of FD. Milutinovic speculated that organic changes in the larynx, pharynx and nose facilitate the appearance of a functional voice problem; that is, they direct the somatization of psychodynamic conflict.

Similarly, Schalen and Andersson (1992) noted that their "psychogenic dysphonia and aphonia" patients had an abnormally high number of reported allergy/asthma symptoms (37.5%). This suggested the need for a more detailed examination of the interrelationship between psychological factors and respiratory and phonatory disorders. Likewise, Rammage, Nichol, and Morrison (1987) proposed that a relatively minor organic change such as edema, infection, or reflux laryngitis may trigger functional misuse, particularly if the individual is exceedingly anxious regarding his or her voice or health. In a similar vein, the same authors felt that anticipation of poor voice production in hypochondriacal, dependent or obsessive-compulsive individuals leads to excessive vigilance over sensations arising from the throat (larynx) and respiratory system that may lead to altered voice production.

Finally, while most authors have viewed psychological factors as strongly influential in the development of FD, they have virtually ignored the possibility that such processes could be the consequence of coping with an incapacitating voice disorder (i.e., the scar hypothesis). Depression, anxiety and tension are frequent psychological concomitants of chronic illness (Dubovsky & Weissberg, 1982; Nemiah, 1961; Reiser, 1985). To our knowledge, the notion that such sequelae could be considered outcomes of a severe voice disturbance, rather than causal agents, has been investigated only once. Murry, Cannito and Woodson (1994) evaluated changes in measures of depression and anxiety in patients with spasmodic dysphonia following BOTOX injections. Reduced levels of depression and anxiety were observed one week after injection, and these reduced levels were maintained during the ensuing two-month postinjection period. In voice disorders such as FD, it is therefore unknown whether elevated depression and anxiety might be more accurately regarded as state-dependent characteristics, more suitably viewed as concomitant rather than causal.

Psychological Factors as Causal or Concomitant?

Most research studies investigating personality and/or psychological processes group both functional aphonia and dysphonia under the presumptive designation "psychogenic voice disorder," reflecting the etiological supposition. This causal inference is, in fact, often difficult to verify in either research or clinical practice. Sapir (1995) suggests that three criteria should be met to warrant the diagnosis of psychogenic voice disorder: symptom incongruity, symptom reversibility, and symptom psychogenicity.

Symptom incongruity refers to the observation that vocal symptoms are physiologically incompatible with existing or suspected disease, are internally inconsistent, and are incongruent with other speech and language characteristics. Often-cited examples of symptom incongruity are (1) complete aphonia in the context of minor vocal fold mucosal changes which are deemed insufficient to explain the severity of vocal dysfunction, or (2) aphonia in a patient who

demonstrates a normal throat clear, cough, laugh or hum, whereby the presence of such normal non-speech vocalization is at odds with assumptions regarding neural integrity and function of the laryngeal system.

Symptom reversibility refers to the *complete, sustained* amelioration of the voice disorder with short-term voice therapy (usually one or two sessions) and/or through psychological abreaction. Furthermore, maintenance of voice improvement requires no compensatory effort on the part of the patient. Sapir advises that symptom reversibility must be distinguished from spontaneous recovery, the latter being the eventual resolution of symptoms as the underlying disease process abates.

Finally, *symptom psychogenicity* refers to the finding that the voice disorder is logically linked in time of onset, course, and severity to an identifiable psychological antecedent, such as a stressful life event, or interpersonal conflict. In general, Sapir adds that "psychogenic dysphonia should be *suspected* when there is strong evidence for symptom incongruity and symptom psychogenicity, but *confirmed only* when there is unambiguous evidence of symptom reversibility (p.275)."

While this attempt to delineate criteria for establishing psychogenicity is commendable, some critics suggest that in reality, the clear demonstration that a symptom is psychogenic rarely occurs. Occasionally, symptoms will come and go in strict association with a salient psychological event or state, and this will allow a degree of confidence that the psychological factor is contributory. More commonly, potential psychosocial contributors are identified, but it cannot be precisely ascertained whether they antedated the voice problem or arose in conjunction with it. Once the voice symptom has developed, the recollection of recent life events may be biased by the tendency of both clinician and client to search for explanations and reassess the salience and negativity of events or mood states. In many cases the dysphonia has been present for a long time. This leads to difficulty in determining the precise etiological factors. Not only do patients forget important historical information, but some of the precipitating factors may have been resolved with passage of time. At the time of assessment, the patient may seem relatively free of psychological distress.

When linking a voice symptom and a psychosocial cause or context, it would seem that there is an underlying assumption that the antecedent conflict or ongoing stressors are of sufficient *intensity* and/or *duration* to account for the voice symptomatology. It is therefore not only the presence, but also the severity of stressors that makes them plausible causes of a voice disorder. An emotional event or interpersonal conflict must be viewed as sufficient to account for comparably severe somatic distress or tension. The problem here is that there are few reliable measures of the severity of a stressor. Severity of the antecedent event/conflict seems to depend on the personal meaning of events and the individual's idiosyncratic judgment of his or her ability to cope adequately with the challenge. Estimates of the intensity of social stress are thus confused with the symptomatic distress that they are intended to explain.

The identification of psychological factors that accompany or even exacerbate a symptom is customarily offered as evidence that the symptom itself is psychogenic. Skeptics suggest that such covariation still does not establish anything beyond concomitance. The covariation of symptom severity with psychological distress establishes a link but does not distinguish functional symptoms from those with obvious organic causes. Such correspondence, even if perfect, does not permit one to infer that the maladjustment caused the voice problem, or the voice problem caused the maladjustment, or that the two caused each other, or that both stem from a single common cause (Irwin, 1960). Spasmodic dysphonia (SD) for instance, *had* historically been regarded as solely psychogenic (Heaver, 1960). But, recent investigations have shown that SD appears to be a neuromotor disorder in the class of focal dystonias, whose onset of symptoms is often related to emotional upset or environmental stresses (Cannito, 1991). Patients with SD report worsening in conjunction with stressful events and improvement when these emotional and social conditions change for the better. Therefore, in actual practice the criterion of recent stressor does not distinguish functional symptoms from those of organic disease that may be aggravated by stress. In many cases, some ambiguity regarding psychogenicity remains, especially in cases where FD masquerades as SD.

The mere presence of symptom psychogenicity and incongruity raises the index of suspicion regarding the contribution of psychological factors, but as Sapir suggests, confirmation requires symptom reversibility. A clinician may have a strong suspicion that a voice disorder has psychological underpinnings; however, confirmation depends on complete sustained symptom reversal without compensatory effort. In this regard, numerous voice-facilitating techniques are available to practicing voice clinicians. However, documentation regarding which therapies are the most effective is scant. If symptoms do not remit with brief therapy, one is left to wonder whether therapy failure is related to differences in disorder, or inappropriately selected or applied techniques by clinicians who vary in levels of confidence, experience, and expectation.

To complicate matters, the same clinician who suspects a psychogenic etiology may ascribe voice therapy failure to psychological causes, such as secondary gain, or incomplete resolution of the interpersonal conflict (or persistent personality characteristics that interfere with resolu-

tion). If patients deny psychological precursors or antecedent events, the clinician may also conclude that the patient is psychologically defended and resistant to psychological insight. Thus, the same assumptions used to explain causation are then invoked to explain voice therapy failure. The circularity of this logic becomes apparent. As Irwin observed, "grant the assumptions, and the evidence is good; doubt the assumptions and there is no evidence" (Irwin, 1960, p.311). So what evidence do we have regarding the psychological and personality factors in functional dysphonia? The major results of the relevant research to date are reviewed in the next section.

Standardized Assessment of Psychological Processes in FD

Although the previous review of potential psychological mechanisms represents fascinating speculation, empirical evidence to support these explanations is lacking. Only a handful of studies exist that have used standardized instruments to assess the personality or psychological characteristics of FD patients. The existing research is generally disappointing for a variety of reasons:

(1) Most data have been collected using test instruments of unknown psychometric properties.

(2) The terminology used to classify voice disorders is often ambiguous and imprecise. Thus it is not clear whether patients with structural pathology have been included or excluded. This is especially true in the case of so called "functional" voice disorders, where some studies included subjects with vocal nodules and others did not. These problems are complicated by the inclusion of both genders, and the failure of the investigators to distinguish aphonic from dysphonic subjects. Mixing voice disorder types and genders into a single study group renders interpretation difficult.

Furthermore, recent technical advancements in observing and evaluating the larynx have improved diagnostic precision. Consequently, subjects with sulcus vocalis, vocal fold scarring, or cysts may have been previously mislabeled as having functional dysphonia or vocal nodules, thereby obscuring potential differences between and within groups.

(3) Most studies neglect to compare their findings with other voice disorder groups. This hinders interpretations of commonality versus specificity.

(4) The use of non-voice-disordered controls or normative data has been largely neglected, or researchers have selected disproportionately large comparison groups with unmatched or unspecified characteristics (Green, 1988).

(5) There are few data that have been satisfactorily analyzed using suitable statistical methods.

(6) Most investigators do not describe whether subjects were vocally asymptomatic at the time of testing. It is therefore difficult to judge whether psychological attributes reflect long-term "trait-like" characteristics, or merely represent transient reaction to the voice disorder (i.e., "state" or "scar" attributes). Information is not provided regarding the severity of vocal handicap or duration of the vocal symptoms; therefore, it is not known whether positive psychological findings represent the effects of attempting to cope with the voice disorder.

These deficits notwithstanding, a review of the major findings and interpretations of the empirical research is provided in Table 1 (following page). Direct comparison and generalization of the results is difficult because of significant methodological differences. This might partly explain the diverse results regarding the frequency and degree of hysterical personality traits (Gerritsma, 1991; Kinzl et al., 1988, Aronson et al., 1966), conversion reaction (Pfau, 1975; House & Andrews, 1987) and psychopathological symptoms (Gerritsma, 1991; Kinzl et al., 1988; Aronson et al., 1966; Pfau, 1975; House & Andrews, 1987). Despite their shortcomings, these studies have identified a general pattern of results consistent with elevated levels of anxiety, somatic complaints and introversion in the FD population. Patients have been described as socially anxious and nonassertive, with a tendency toward restraint (Gerritsma, 1991; Friedl et al., 1990). None of these researchers have attempted to explicitly integrate their findings into a coherent theory of personality structure as a vulnerability for FD. As Green (1988, p.34) states, "until more adequate research is conducted, psychological-personality variables must be considered possible etiological, consequential and therapeutic factors."

In a recent work, Roy and coworkers (Roy, McGrory, Tasko, Bless, Heisey & Ford, in press-a) attempted to shed further light on the voice-personality relationship. They described the personality/psychological characteristics of twenty-five female subjects who had received the diagnosis of FD. All subjects experienced symptom resolution following voice therapy. While vocally asymptomatic, these remitted FD subjects completed the Minnesota Multiphasic Personality Inventory (MMPI), an objective personality questionnaire (Hathaway & McKinley, 1972). When compared to a medical outpatient control group, the results showed that FD subjects scored significantly higher on 7 of 10 clinical scales, suggesting an elevated degree of emotional maladjustment. A stepwise discriminant analysis identified two clinical scales that provided valuable discriminatory information. Scale one (Hs-Hypochondriasis), which measures number and type of reported somatic complaints, and scale seven (Pt-Psychaesthenia), a measure of diffuse anxiety, discriminated the groups with 88% sensitivity and 89% specificity. The results suggested that in spite of symptom improvement after voice therapy, the FD subjects continued to exhibit poor levels of adaptive functioning, which may represent

Table 1.
Literature Review: Description of Subjects, Test Instrument(s), and Major Findings of the FD-Psychology-Personality Relationship

Authors	Subjects	Test Instrument(s)	Major Findings & Interpretations
Aronson et al. (1966)	• psychogenic aphonia/dysphonia • n = 24 F; 3 M 1 mute, 11 continuous whisper, 7 intermittent whisper/phonation, 8 continuous phonation	• Minnesota Multiphasic Personality Inventory (MMPI) • clinical psychiatric interview	• A period of acute/chronic stress antedated the onset of dysphonia in 74% of the patients. • 93% were judged to have difficulty dealing with anger. • 26% reported excessive somatic complaints. • No patient in acute psychiatric distress. • A clinical impression of hysteria was observed in less than half and 30% exhibited a conversion "V" profile. • The authors suggested that the entire group had a "hysterical flavor."
Pfau (1975)	• psychogenic aphonia/dysphonia • n = 46 F; 8 M	• German equivalent of MMPI	• Results suggested neurosis in 35% of females patients, 20% of whom were considered a hysterical reaction type. • The majority of individual profiles were either uninterpretable (37%), or considered within normal limits (28%).
Kinzl et al. (1988)	• "hyperfunctional" and "hypofunctional" aphonia • n = 22 F	• psychiatric evaluations • social support network assessments • life event histories	• Patients did not have particular personality traits in common, nor exposed to comparable conflict situations. • "Personality structures and psychopathological symptoms ranged from mild impairment to severe neurosis (p.134)." • Hysterical personality traits were frequent, but not always present. • 75% of patients have other psychosomatic functional disturbances in their histories. • Authors suggest that aphonia is a homogeneous clinical syndrome with heterogeneous personality structures and psychopathologies underlying its development.
Gerritsma (1991)	• psychogenic dysphonia/aphonia • n = 75 F; 7 M	• Wilde's Amsterdam Biographical Questionnaire	• 42% of patients scored high on neuroticism (N) and neurotic somatization (NS) scales, a pattern consistent with conversion symptoms; and, 40% scored low on the extraversion (E) scale, suggesting a tendency toward introversion.
Gerritsma (1991) cont'd	• psychogenic dysphonia/aphonia	• Social Anxiety Scale (SAS) • Wolpe-Lazarus Assertiveness Scale (WLAS)	• 4% met DSM (III) criteria for hysterical personality. • 65% of subjects were socially anxious, nonassertive or both. • Author suggests dropping the term hysterical dysphonia in favor of one of "conversion, psychogenic, or functional" aphonia.
Friedl et al. (1990)	• functional dysphonia/aphonia (20) • organic dysphonia (14) • normal control (20)	• multiple measures of personality and anxiety	• FD patients show a tendency toward restraint, and in stressful situations, the result is an intensified anxiety state. • Life events may influence the pathogenesis of FD.
Friedl et al. (1993)	• functional dysphonia/aphonia	• "empirical-psychological procedure"	• Psychological conditions were major etiologic factors in aphonia, but only partially relevant in patients with functional dysphonia
House & Andrews (1987)	• functional dysphonia/aphonia • dysphonia (65) • aphonia (4) • spastic (2)	• Present State Examination (PSE) • Bedford College Life Events & Difficulties Interview (BCLEDI)	• Authors failed to find an association between voice type and PSE score or psychiatric diagnosis. • "The majority of patients were remarkable for the apparent normality of their premorbid psychological and social functioning. Major mental illness was infrequently diagnosed and minor states of tension and anxiety predominated (33%) (p.488)." • FD is not usually found in markedly abnormal personalities and previous episodes of conversion disorder are rare.

trait-like stability. Figure 1 illustrates the differences between the FD group and the medical controls. The elevations on scales 1, 2, 3 have come to be known as the neurotic triad and are highly related to the personality dimension of neuroticism, to be discussed later. Furthermore, elevation on scale 7 measures not only diffuse anxiety, but also self-doubt regarding adequacy in interpersonal situations. When combined with elevation on Scale 0-Social Introversion (Si) (an index of a person's preference for being alone (high 0) or being with others (low 0)), a strong case can be made for

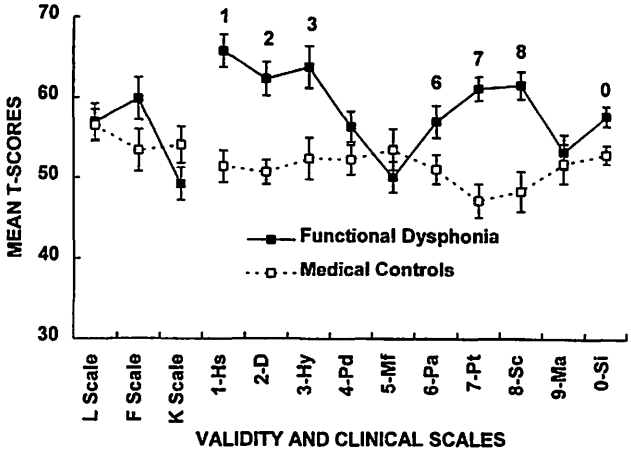


Figure 1. Composite MMPI Profiles for Functional Dysphonia and Medical Control groups using mean T-Scores (+/- standard error). Numbers placed above select scales indicate significant differences between groups.

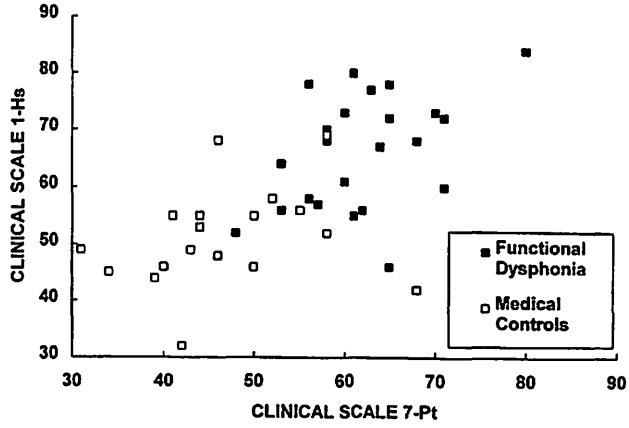


Figure 2. Scatterplot of T-Scores obtained for clinical scale 1-Hs (ordinate) and clinical scale 7-Pt (abscissa) from individual Functional Dysphonia and Medical Control subjects.

a tendency toward introversion (i.e., low E). High scorers tend to be withdrawn, socially insecure, and anxious when in contact with people. With the exception of Scale 2 (Depression), all clinical scales are viewed as assessments of character, not mood (Graham, 1987; Graham, 1990; Newmark, 1979; Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989; Duckworth & Anderson, 1995; Greene, 1989; Marks, Seeman & Haller, 1974). The scatterplot of the distribution of scores for the FD subjects versus the controls (Figure 2) clearly illustrates the discriminatory value of the two clinical scale variables (i.e. Scales 1 & 7). These data seem to support a dispositional vulnerability for the development of functional symptoms, including laryngeal problems.

Vocal Nodules

Vocal nodules are another voice disorder that potentially illustrates a link between vocal pathology and personality-psychological factors, but much less has been reported on this topic. Vocal nodules are benign callous-like lesions of the vocal folds often attributed to chronic, repetitive phonotrauma creating excessive mechanical tissue stresses and reactive histological changes. They are considered to be one of the most common manifestations of vocal hyperfunction, i.e., abuse and/or misuse of the vocal mechanism due to excessive and/or "imbalanced" muscular forces (Hillman, Holmberg, Perkell, Walsh & Vaughan, 1989), and may account for almost 4% of an otolaryngologic caseload (Nagata, Kurita, Yasumoto, Maeda, & Kawasaki, 1983). Vocal nodules tend to occur in prepubescent males and postpubescent/pre-menopausal females. In adults, at least two-thirds of patients with nodules are female (Herrington-Hall, Lee, Stemple, Niemi & McHone, 1988; Nagata et al., 1983). Surgical removal is one method of treatment; however, a more conservative approach is behavioral voice therapy that attempts to eliminate the putative cause(s) of the vocal nodule rather than the nodule itself. The short-term results of behavioral therapy or surgical excision are generally favorable (Bouchayer & Cornut, 1988; Lancer, Snyder, Jones, 1988; Murry & Woodson, 1992). But few studies have *objectively* evaluated long-term clinical outcomes. At least anecdotally, it appears that despite the efforts of surgeons and voice therapists, the lesions in some adults are resistant to therapy and/or tend to recur (Arnold, 1962; Bridger & Epstein, 1983).

For the most part, authors have attempted to distinguish vocal nodules from other voice disorders, such as functional dysphonia. However, some authors have classified VN as functional disorders and have emphasized the role of psychological precursors and predisposing personality factors (Arnold, 1962; Wilson, 1987, Aronson, 1990). One common belief is that people with VN are talkative and have aggressive tendencies (Arnold, 1962; Green, 1989; Mosby, 1970; Nemec, 1961; Toohill, 1975; Wilson, 1971; Wilson and Lamb, 1974; Withers & Dawson, 1960). Elevated levels of anxiety, reduced self-concept, emotional maladjustment, and high levels of extraversion have also been found in patients with VN (Mosby, 1970; Peter & Brandell, 1980; Toohill, 1975; Yano, Ichimura, Hoshino & Nozue, 1982).

In a recent study using the MMPI, we identified elevated levels of psychological distress and somatic complaints in a group of adult female VN patients when compared to a medical out-patient control group (Roy, McGrory & Bless, 1995). Goldman, Hargrave, Hillman, Holmberg and Gress (1996) confirmed these findings when they also identified elevated levels of anxiety, somatic complaints and voice use when compared to non-voice-disordered controls.

No differences, however, were identified between the subjects with VN and a non-pathological voice disordered control group.

Recently, White, Deary and Wilson (in press) using the General Health Questionnaire (GHQ), a measure of an individual's number and variety of health complaints, and the Eysenck Personality Questionnaire (EPQ) found no significant differences in personality traits when comparing dysphonic patients (both functional and organic) with ENT outpatient controls. They did, however, identify elevated levels of psychological distress in both voice-disordered groups and concluded that "it is not possible to identify those dysphonia patients with a major underlying psychological upset simply on the basis of laryngeal appearance and phonatory characteristics." Thus the pattern of results, although by no means definitive, suggests a trend toward elevated levels of extraversion and anxiety among subjects with vocal nodules.

Toward a Theory of the Dispositional Bases of FD & Vocal Nodules

From the previous literature review, it is apparent that the origin of common voice problems such as FD and VN is poorly understood, and likely involves the convergence of multiple factors including organic, psychological and social features. A major obstacle limiting progress in the field of voice pathology is the difficulty in conceptualizing personality-psychological processes that might contribute to the development and maintenance of these voice disorders. An effective theoretical perspective is needed to promote intuitive understanding of these disorders, reconcile divergent research findings, arouse interest in the etiology and treatment of such voice problems, and generate research designed to contrast alternative hypotheses.

A theory is proposed that couples aspects of a biological theory of personality (Eysenck, 1967; Eysenck, H. & Eysenck, M., 1987) with a neuropsychological model of the conceptual nervous system (Gray, 1975). Based on differences in personality, this theory predicts "*unique*" and "*contrasting*" signal sensitivities and behavioral response biases for individuals with FD and VN. It is proposed that definite personality traits predispose one to develop these disorders, and moderate the symptomatology and course of the voice pathology. Moreover, by virtue of its enduring nature, personality serves as a persistent diathesis, rendering an individual vulnerable for recurrence of symptoms. The following sections describe the theory's foundations and its predictions.

Defining Personality, Personality Traits, and Trait Structure

Most definitions of personality indicate that it is internal, organized, enduring--characteristic of an individual over time and situations, and related to how an individual functions in the world. The term personality implies a complex organization of systematically interrelated trait dispositions (Watson, Clark & Harkness, 1994). It is widely acknowledged that personality traits are hierarchically arranged, with specific but narrow traits at lower levels in the hierarchy, and global but broad trait dimensions or domains at the top (Goldberg, 1993; John, 1990). At the highest level of the trait hierarchy exist three stable, heritable, general personality dimensions or "superfactors." These relatively orthogonal superfactors provide for the global classification of personality traits (Digman & Takemoto-Chock, 1981; McCrae & Costa, 1986, 1987). The so-called "Big Three" dimensions have been found in a wide range of data sources, instruments, samples, and languages (John, 1990) and are commonly referred to as (1) Extraversion vs. Introversion, (2) Neuroticism vs. Stability, and (3) Constraint vs. Disinhibition.[1] These personality dimensions are typically derived using factor analytic techniques and thus, are not necessarily tied to actual psychobiological processes. However, some investigators such as Hans Eysenck (1967; Eysenck & Eysenck, 1987) and Jeffrey Gray (1982; 1987) have linked the superfactors to specific psychophysiological processes. Extraversion (E) and Neuroticism (N) play a vital role in our theory, which synthesizes Eysenck's and Gray's biological theories of personality to account for the development of functional dysphonia and vocal nodules.

Eysenck's Personality System

H. J. Eysenck (1967) developed an integrated biopsychosocial theory of personality that is based primarily on the dimensions of personality, extraversion (E), and neuroticism (N). Extraversion is related to sociable, lively, active, assertive, sensation-seeking, carefree, dominant, surgent and venturesome characteristics. Extraversion involves the willingness to engage and confront the environment, including the social environment. Extraverts (i.e., high E) tend to be dominant, sociable, and active, whereas introverts (i.e., low E) tend to be quiet, unsociable, passive, and careful. Eysenck views E as reflecting stable differences in the tonic activity level of the ascending reticular activating system, and thus cortical arousal. Introverts are thought to exhibit higher tonic levels of cortical arousal than extraverts. Extraverts are therefore expected to seek stimulation to raise their arousal to more optimal levels. On the other hand, introverts' high level of arousal is associated with the avoidance of excessive stimulation.

[1] Constraint vs. disinhibition is centered around the basic issue of impulse control. High constraint individuals are cautious, restrained, as refraining from risky adventures, and as accepting the conventions of society. These individuals plan carefully before acting and avoid situations involving risk or danger. Low constraint persons are relatively impulsive, adventurous and inclined to reject conventional restrictions (Clark, Watson & Mineka, 1994).

N, the second personality dimension, can be likened to emotionality and is related to anxious, depressed, tense, shy, moody, and emotional characteristics, as well as guilt feelings and low self-esteem. High N individuals tend to be emotionally unstable, worried, anxious or highly reactive to environmental stimuli. Eysenck proposed the visceral brain, consisting of the septum, hippocampus, cingulum, amygdala, and hypothalamus as the neurological substrate (Eysenck & Eysenck, 1985). Neuroticism magnifies response tendencies derived from E (Eysenck & Eysenck, 1978). Therefore, neurotic introverts tend to be more introverted, and neurotic extraverts tend to be more extroverted, when compared to their stable counterparts.[2]

Gray's Theory of Personality and Nervous System Function

Gray (1975, 1982, 1985) has integrated findings from learning, psychometry, pharmacology, neurophysiology, and other areas of research into a general model of personality and functional neuropsychology. He proposed a neuropsychological model of the conceptual nervous system that consists of a set of three interacting components (Figure 3): **a behavioral activation system (BAS), a behavioral inhibition system (BIS), and a nonspecific arousal system (NAS).** The BAS, referred to as the reward system, is responsive to signals of conditioned reward and nonpunishment; activity increases in the presence of such stimuli. Hence, the BAS--as a functional system-- can be conceived as associated with the attainment of motivationally significant goals (i.e. approach behavior). The BAS is considered the "go" system, and promotes the intiation of goal-directed motor behavior including approach, escape, and active avoidance.

The BIS, on the other hand, is responsible for organizing reactions to conditioned signals of punishment, signals of frustrative nonreward, novel or threat stimuli, and a class of innate fear stimuli. Nonreward refers to a context in which a reward is omitted following a response in a situation in which the response had previously been rewarded, or in which a reward for the response was anticipated. The psychological state instantiated by the occurrence of such nonrewards is called frustration.

Each of these types of input stimuli is presumed to influence common neural structures located in the septohippocampal system, with connections to the prefrontal cortex. Because the inputs of punishment and frustrative nonreward act on the same neural structures, the effects of nonreward and punishment are considered functionally equivalent. These input signals increase BIS activity, resulting in increased arousal, inhibition of ongoing behavior, and increased attention to the environment. The BIS inhibits or decelerates responses that may lead to punishment or nonreward, producing passive avoidance or extinction. In passive avoidance, an organism can avoid receiving punishment or non-reward by not performing a given action (i.e. response suppression). Gray distinguishes this type of avoidance from active avoidance, where punishment can be avoided if an organism performs a given action.

The third component of Gray's model --the nonspecific arousal system (NAS)-- serves to prepare or ready the organism to respond to BAS or BIS inputs which have motivational or emotional significance. This NAS has been linked to major changes in the functioning of the autonomic nervous system similar to the fight/flight response.

Several general points regarding Gray's theory should be recognized: (1) Reciprocal inhibitory inputs connect the two behavioral systems, such that an increase in the activity of one results in a decrease in the activity of the other. (2) An increase in the activity of either behavioral system results in heightened NAS activity via excitatory outputs from the BAS and BIS. (3) The NAS has excitatory connections affecting responses mediated by the behavioral systems, so that as NAS activity increases, the speed and vigor of behavioral responses increase proportionately.

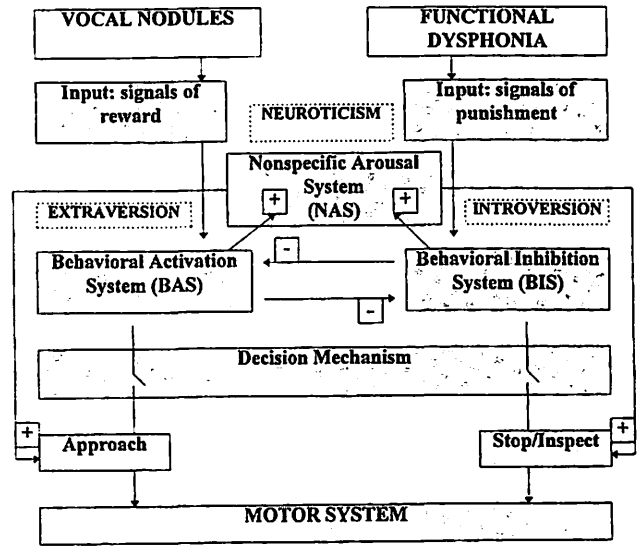


Figure 3. A theory of the dispositional bases of functional dysphonia and vocal nodules, adapted from Newman & colleagues' sythesis of Eysenck's (1967) and Gray's (1987) biological theories of personality. E and N are mapped onto the three systems within Gray's conceptual nervous system model.

[2] Eysenck also proposed a third personality dimension known as Psychoticism. (P) or "toughmindedness". This broad personality factor shares many of the impulsive features of the Constraint dimension (reversed), but also includes a predilection for antisocial and aggressive behavior. The biological substrat of P has been more problematic for Eysenck and he has suggested that it may be related to individual differences in serotonergic systems (McBurnett, 1992).

Dispositional Vulnerability in FD & VN: The Importance of Extraversion & Neuroticism

Newman and colleagues (Newman & Wallace, 1993a, b; Patterson & Newman, 1993; Wallace & Newman, 1991; Wallace & Newman, in press) proposed a synthesis of Eysenck's and Gray's theoretical formulations to account for breakdowns in self-regulatory behavior observed in disinhibited adults and children. It is this synthesis that provides the foundation for our theory of the dispositional bases of FD and VN.

Briefly, the three components of Gray's model are mapped onto Eysenck's personality dimensions of Extraversion (E) and Neuroticism (N). An individual's position on E reflects the relative strengths of the behavioral systems. For example, in extraverts the BAS is stronger than the BIS, and for introverts the BIS is the stronger of the two systems. Thus, Extraverts = BAS dominance = reward sensitive = approach behavior; Introverts = BIS dominance = punishment, threat sensitive = stop/reflect behavior. Neuroticism directly reflects the reactivity or lability of the NAS: an individual is neurotic by virtue of possessing a more reactive NAS than a stable individual. Neuroticism (N) amplifies response tendencies associated with the two behavioral systems; therefore, as N increases, extraverts tend to behave in a more extraverted manner, and introverts tend to behave in a more introverted manner. In this model, the combination of neuroticism and extraversion leads to impulsivity (disinhibition), whereas the combination of neuroticism and introversion leads to anxiety/distress (inhibition). Neurotic extraverts are highly reactive, especially to potential rewards, and initiate goal-directed behavior (e.g., approach). On the other hand, neurotic introverts are highly reactive to threatening and unexpected stimuli, and are prone to engage in BIS-mediated activities (e.g., motor inhibition, inspecting the environment for potential threats, and passive avoidance).

An important feature of this model is that BAS-dominated neurotic extraverts do not engage in BIS-mediated processes that subserve response modulation (i.e., response interruption, initiation of information gathering and processing) and therefore do not alter dominant or ongoing responses. This ultimately manifests in forms of perseverative behavior; for example, persistent approach behavior in spite of unfavorable effects. By virtue of heightened NAS activity, impulsive (neurotic extraverts) and anxious individuals (neurotic introverts) are prone to focus selectively on stimuli of direct relevance to the attainment of their goal, to the exclusion of information that is of peripheral relevance to the current response set. BIS-dominated neurotic introverts are unable to alter dominant response sets such as motor inhibition-passive avoidance, by attending to other information. This model and its presumed signal sensitivities and response biases have implications for both types of voice disorders--FD and VN--which will be described in the section that follows.

Applying the Theory to Individuals with FD & VN

This theoretical synthesis serves as the foundation for a theory describing the dispositional bases of FD and VN. Functional dysphonia represents a voice disorder in the absence of visible structural or neurological pathology. As mentioned previously, psychological processes have been identified as pathogenic; however the underlying mechanism has not been fully elucidated. Earlier we identified that these individuals were characterized by elevated scores on measures of anxiety, somatic complaints and introversion (see literature review). Excessive voice use has not been confirmed as a causative agent in this form of dysphonia. We contend that functional voice loss is related to anxiety, motor inhibition, and elevated tension states. By virtue of BIS dominance and elevated NAS, *neurotic introverts* tend to experience difficulty in evaluating and, if necessary, appropriately altering, ongoing responses or response sets in the presence of BIS inputs such as uncertainty, frustrative nonreward, punishment or potential threats. We speculate that these input signals ultimately contribute to inhibitory laryngeal motor behavior in at least two possible ways. These mechanisms are not independent or mutually exclusive, and therefore could be operating in combination within a single individual.

First, individuals who are BIS dominant will exhibit hypervigilance concerning their environment, including their internal body environment. A chief function of the BIS is to compare actual with expected stimuli. The system exercises selective control over the sensory information that reaches it, and tags it as "important" or filters it as "unimportant". Once activated, the system is assumed to heighten attentional awareness and focus on potential threats. Ambiguous sensory changes accompanying minor alterations in laryngeal function as a result of infection, edema, reflux laryngitis, or emotional states are tagged by the BIS as a mismatch, novel and perhaps threatening. In combination with increased arousal, further attentional resources are directed to the circumlaryngeal area combined with interruption of ongoing motor programs, i.e., laryngeal motor inhibition. In this control mode, the inputs of uncertainty persist, the problem is intensified, and the patient is unable to interrupt this inhibitory response pattern. The neocortical control of speech and language is unaffected, but inhibitory control by the septohippocampal system dominates, preventing normal phonation. Sustained motor inhibition without appropriate release leads to unnecessarily high muscular tonus. Partial or complete voice loss reflects the cumulative effects of heightened NAS and BIS, with resultant motor inhibition and elevated tension states.

The second pathway to vocal dysregulation emphasizes frustrative nonreward and punishment as important inputs for BIS activation. Recall that frustrative nonreward occurs in circumstances in which a reward is omitted following a response in a situation in which the response had been previously rewarded, or in which a reward for the response was anticipated. The effects of nonreward where it had been anticipated are functionally equivalent to punishment. Thus, the BIS inhibits responses that may lead to frustration. The inhibition system is activated only when the individual anticipates the omission of a reward or punishment following a response. The BIS system takes active control over vocal behavior upon receipt of stimuli which have preceded (on a previous occasion) the disruptive event (frustrative nonreward or punishment). If vocalization (speaking out) has resulted in frustrative nonreward or punishment in the past, such vocalization will be inhibited (i.e., passive avoidance). Recall that in passive avoidance an organism can avoid receiving punishment by not performing a given action. The failure to respond removes the potential for punishment. The BIS is assumed to increase passive avoidance and inhibits actions that might result in punishment.

A common theme in most writings regarding antecedents to psychogenic voice loss involves (1) a conflict over speaking out or expressing an unfavorable opinion, and/or (2) chronic stress involving job and family responsibilities, marital dissatisfaction, or long-standing communication breakdowns (Aronson, 1990; Butcher, 1995; House & Andrews, 1988). When the subject has experienced undesirable punishing or frustrating outcomes paired with previous attempts to speak out, these become conditioned inputs for the BIS. The BIS outputs remain the same regardless of the inputs. In the case of human voice production, the newer phylogenetic system (neocortex) competes with inputs from the older phylogenetic system (septohippocampal system). By way of the BIS, this older brain region may stimulate inhibition of vocalization (i.e., laryngeal "freezing" behavior), whereas the neocortex carries out its communicative intent/goal. This conflict between inhibition and activation may give rise to incomplete or disordered vocalization in a structurally and neurologically intact larynx.

As mentioned earlier, vocal nodules are benign lesions of the vocal folds thought to be caused by repetitive mucosal injury leading to histological changes and concomitant voice mutation. Excessive voice use and abuse have been implicated as causative agents. Although treatment (surgical and/or voice therapy) may improve the voice in the short term, long-term results seem less impressive. Recurrence of dysphonia is common, rather than the exception. We allege that vocal nodule development is in part a result of an information processing pattern related to the *impulsive behavior of neurotic extraverts* (i.e, BAS dominance with elevated NAS activity). In spite of the obvious harmful effects (voice change, laryngeal discomfort) of this perseverative behavior, patients with vocal nodules often appear unable to engage in appropriate response modulation (i.e., to stop vocal overuse and abuse) in the presence of salient "social" reward cues. Consequently, we reason that vocal nodule patients should score high on indices of extraversion (dominance, sociability) and neuroticism (emotional reactivity), and low on measures of constraint (i.e., reflecting impulsivity). Neuroticism serves to potentiate the signal sensitivities and response biases of extraversion leading to impulsivity.

To summarize, we connect the development of functional dysphonia and vocal nodules to personality differences related to dissimilar signal sensitivities and response biases of neurotic introverts and neurotic extraverts, respectively. We credit hyperreactivity of the BIS as a prime constituent in the pathogenesis of functional dysphonia, whereas hyperreactivity of the BAS is pathogenic in vocal nodule development. Both behavioral systems are potentiated by the NAS. This synthesis of two biological theories of personality as an explanatory model serves as a heuristic for future research. If a theory is to have scientific value, it must be open to test. One must be able to derive hypotheses from it that allow new evidence to be gathered that either refutes the theory or fails to do so, thus adding to its credibility. From this perspective, we believe that this theory and its derivations has a decided advantage over other psychoanalytically-based formulations of being open to test by a variety of approaches. We look forward to such scientific inquiry.

Assessment of these broad personality dimensions, and the traits subsumed within, is needed to help clinicians better appreciate the relation between personality, psychological factors, and voice pathology. Until the role of personality in the pathogenesis of voice disorders is better understood, *long-term* clinical outcomes for these populations may remain unsatisfactory (Bridger & Epstein, 1989; Roy, Bless, Heisey & Ford, in press-b). Improved understanding of its influence could help to explain voice therapy failure and refine treatment strategies in some cases. If personality represents a persistent vulnerability for the development, maintenance, and recurrence of certain voice pathologies, then assessment and management practices may need to be revised.

References

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders. Fourth Edition. Washington. DC.

Alexander, F. (1950). Psychosomatic Medicine: Its Principles and Applications. New York: WW Norton and Company.

Alexander, F., French, T.M., & Pollock, G.H. (1968). Psychosomatic Specificity, Vol. 1: Experimental Study and Results. Chicago: University of Chicago Press.

Aronson, A.E. (1990). Clinical Voice disorders: An interdisciplinary approach (3rd ed.). New York: Thieme.

Aronson, A.E., Peterson, H.W., & Litin, E.M. (1966). Psychiatric symptomatology in function
dysphonia and aphonia. *Journal of Speech and Hearing Disorders*, 31, 115-127.

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

Bass, C.M., (ed.) (1990). Somatization: Physical Symptoms & Psychological Illness. London: Blackwell Scientific Publications.

Boone, D.R., McFarlane, S. (1988). The Voice and Voice Therapy. 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Bouchayer, M., & Cornut, G. (1988). Microsurgery for benign lesions of the vocal folds. *Ear, Nose, and Throat Journal*, 67, 446-466.

Bridger M.M., & Epstein, R. (1983). Functional voice disorders: A review of 109 patients. *Journal of Laryngology and Otology*, 97, 1145-1148.

Brodnitz, F.S. (1962). Functional disorders of the voice. In N.M. Levin (Ed.), Voice and speech disorders: Medical aspects. Springfield: Charles C. Thomas.

Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B.(1989). MMPI-2 (Minnesota Multiphasic Personality Inventory -2): Manual of administration and scoring. Minneapolis: University of Minnesota Press.

Butcher, P. (1995). Psychological processes in psychogenic voice disorder. *European Journal of Disorders of Communication*, 30, 467-474.

Butcher, P., Elias, A., Raven, R. (1993). Psychogenic Voice Disorders and Cognitive Behaviour Therapy. San Diego: Singular Publishing Group.

Butcher, P., Elias, A., Raven, R., Yeatman, J., & Littlejohns, D. (1987) Psychogenic voice disorder unresponsive to speech therapy: Psychological characteristics and cognitive-behaviour therapy. *British Journal of Disorders of Communication*, 22, 81-92.

Cannito, M.P. (1991). Emotional considerations in spasmodic dysphonia: Psychometric Quantification. *Journal of Communicative Disorders*, 24, 313-329

Clark, L.A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology*, 103, 103-116.

Colton, R., & Casper, J.K. (1996). Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment. Baltimore: Williams & Wilkins.

Cooper, M. (1973). Modern Techniques of Vocal Rehabilitation. Springfield: Charles C. Thomas.

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders. Fourth Edition. Washington, DC.

Digman, J.M., & Takemoto-Chock, N.K. (1981). Factors in the natural language of personality: Re-analysis and comparison of six major studies. *Multivariate Behavioral Research*, 16, 149-170.

Doyle, P.C. (1994). Foundations of Voice and Speech Rehabilitation Following Laryngeal Cancer. San Diego, CA: Singular.

Dubovsky, S.L., & Weissberg, M.P. (1982). Reactions to Illness: In Clinical Psychiatry in Primary Care, 2nd edn., Baltimore: Williams & Wilkins.

Duckworth, J.C., & Anderson, W.P. (1995). MMPI and MMPI-2: Interpretation Manual for Counselors and Clinicians. Philadelphia: Taylor and Francis Group.

Epstein. S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37, 1097-1126.

Eysenck, H.J. (1967). Biological basis of personality. Springfield, IL: Thomas.

Eysenck, H.J., & Eysenck, S.B.G. (1975). Psychoticism as a dimension of personality: London: Hodder & Stoughton.

Eysenck, H.J., & Eysenck, M.W. (1985). Personality and Individual Differences. London: Plenum Press.

Eysenck, H.J., Eysenck, M.W. (1987). Personality and individual differences: A natural science approach. New York: Plenum Press.

Fex, F., Fex, S., Shiromoto, O., & Hirano, M. (1994). Acoustic analysis of functional dysphonia: Before and after voice therapy (Accent Method). *Journal of Voice*, 8, 163-167.

Friedl, W., Friedrich, G, Egger, J., Fitzek, I. (1993). Psychogenic aspects of functional dysphonia. *Folia Phoniatrica*, 45, 10-13.

Friedl, W., Friedrich, G., & Egger, J. (1990). Personality and coping with stress in patients suffering from functional dysphonia. *Folia Phoniatrica*, 42, 144-149.

Gerritsma, E.J. (1991). An investigation into some personality characteristics of patients with psychogenic aphonia and dysphonia. *Folia Phoniatrica*, 43, 13-20.

Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, 26-34.

Goldman, S.L., Hargrave, J., Hillman, R.E., Holmberg, E., & Gress, C. (1996). Stress, anxiety, somatic complaints, and voice use in women with vocal nodules: Preliminary findings. *American Journal of Speech-language Pathology: A Journal of Clinical Practice*, 5, 44-54.

Goodstein, L.D. (1958). Functional speech disorders and personality: A survey of the research. *Journal of Speech and Hearing Research*, 1, 359-376.

Graham, J.R. (1987). *The MMPI: A Practical Guide. 2nd edn.* New York: Oxford University Press.

Graham, J.R. (1990). *MMPI-2: Assessing Personality and Psychopathology*. New York: Oxford University Press.

Gray, J.A. (1975). Elements of a Two Process Theory of Learning. London: Academic Press.

Gray, J.A. (1981). A critique of Eysenck's theory of personality. In H.J. Eysenck (Ed.), *A model for personality* (pp. 246-276). New York: Springer-Verlag.

Gray, J.A. (1982). The neuropsychology of anxiety. New York: Oxford University Press.

Gray, J.A. (1985). Issues in the neuro-psychology of anxiety. In A. H. Tuma & J.D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 5-25). Hillsdale, NJ: Erlbaum.

Gray, J.A. (1987). The psychology of fear and stress (2nd ed.). New York: Cambridge Press.

Green, G. (1989). Psychobehavioral characteristics of children with vocal nodules: WPBIC ratings. *Journal of Speech and Hearing Disorders.*

Green, G. (1988). The inter-relationship between vocal and psychological characteristics: A literature review. *Australian Journal of Human Communication Disorders*, 16, 31-43.

Greene, M.C., & Mathieson, L. (1989). The voice and its disorders. 5th edn., London: Whurr Publishers.

Greene, R.L. (1989). The MMPI: An Interpretive Manual. 2nd edn. New York: Grune & Stratton.

Hathaway, S.R., & McKinley, J.C. (1972). The Minnesota Multiphasic Personality Inventory. New York: Psychological Corporation.

Heaver, L (1960). Spastic Dysphonia: A psychosomatic voice disorder. In D.A. Barbara (Ed.). Psychological and Psychiatric Aspects of Speech and Hearing. Springfield: Charles C. Thomas.

Herrington-Hall, B.L., Lee, L., Stemple, J.C., Niemi, K.R., & McHone, M.M. (1988). Description of laryngeal pathology by age, sex, and occupation in a treatment-seeking sample. *Journal of Speech and Hearing Disorders*, 53, 57-64.

Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research*, 32, 373-392.

House, A.O, & Andrews, H.B. (1987). The psychiatric and social characteristics of patients with functional dysphonia. *Journal of Psychosomatic Research*, 3, 483-490.

House, A.O., & Andrews, H. B. (1988). Life events and difficulties preceding the onset of functional dysphonia. *Journal of Psychosomatic Research*, 32, 311-319.

Irwin, J.V. (1960). Psychological implications of voice and articulation disturbances. In D.A. Barbara (Ed.). Psychological and psychiatric aspects of speech and hearing. Springfield: Charles C. Thomas.

John, O.P. (1990). The "Big Five" factor taxonomy: Dimension of personality in the natural language and in questionnaires. In L.A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66-100). New York: Guilford Press.

Kirmayer, L.J., & Robbins, J.M., (1991)(eds.). Current concepts of Somatization: Research and Clinical Perspectives. Washington: American Psychiatric Press, Inc.

Kinzl, J., Biebl, W., & Rauchegger, H. (1988). Functional aphonia: Psychosomatic aspects of diagnosis and therapy. *Folia Phoniatrica*, 40, 131-137.

Koufman, J.A., & Blalock, P.D. (1982). Classification and approach to patients with functional voice disorders. *Annals of Otology, Rhinology, Laryngology*, 91, 372-377.

Lancer, J.M., Syder D., Jones A.S., Boutillier, A.L. (1988). The outcome of different mangement patterns for vocal cord nodules. *Journal of Laryngology and Otology*, 102, 423-427.

Mans, E.J. (1994). Psychotherapeutic treatment of patients with functional voice disorders. *Folia Phoniatrica*, 46, 1-8.

Marks, P.A., Seeman, W., & Haller, D.L. (1974). *The Actuarial Use of the MMPI with Adolescents and Adults.* Baltimore: Williams & Wilkins Company.

McBurnett, K. (1992). Psychobiological approaches to personality and their applications to child psychopathology. In: Advances in Clinical Child Psychology, Vol.14, (Eds. Lahey, B.B. & Kazdin, A.E.), New York: Plenum Press.

McCrae, R.R., & Costa, P.T., Jr. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, 6, 587-597.

McCrae, R.R., & Costa, P.T., Jr. (1987). Validation of a five-factor model of personality across instruments and observers. *Journal of Personality and Individual Differences*, 6, 587-597.

Milutinovic, Z. (1991). Inflammatory changes as a risk factor in the development of phononeurosis. *Folia Phoniatrica*, 43, 177-180.

Morrison, M.D., Nichol, H., & Rammage, L.A. (1986). Diagnostic criteria in functional dysphonia. *Laryngoscope*, 94, 1-8.

Morrison, M.D., & Rammage, L. (1993). Muscle misuse voice disorders: description and classification. *Acta Otolaryngologica (Stockh).* 113:428-434.

Mosby, D.P. (1970). Psychotherapy versus voice therapy for a child with a deviant voice. A case study. *Perceptual Motor Skills*, 887-891.

Murry T., & Woodson, G., (1992). Comparison of three methods for the management of vocal fold nodules. *Journal of Voice*, 6, 271-276.

Murry, T., Cannito, M.P., & Woodson, G.E. (1994). Spasmodic dysphonia: Emotional status and Botulinum Toxin Treatment. *Archives of Otoloryngology, Head and Neck Surgery*, 120: 310-316.

Nagata, K., Kurita, S., Yasumoto, S., Maeda, T., Kawasaki, H., & Hirano, M. (1983). Vocal fold polyps and nodules: A 10 year review of 1,156 patients. *Auris-Nasus-Larynx*, 10 (Suppl.), S27-S35.

Nemec, J. (1961). The motivation background of hyperkinetic dysphonia in children: A contribution to psychologic research in phoniatry. *LOGOS*, 4, 28-31.

Nemiah, J.C. (1961). Psychological complications of physical illness. In: *Foundations of Psychopathology.* New York: Oxford University Press.

Newman, J.P., & Wallace, J.F. (1993a). Diverse pathways to deficient self-regulation: Implicatios for disinhibitory psychopathology in children. *Clinical Psychology Review*, 13, 699-720.

Newman, J.P., & Wallace, J.R., (1993b). Cognition and Psychopathy (In: Psychopathology and Cognition,), New York: Academic Press.

Newman, J.P., Wallace, J.F. (1993). Diverse pathways to deficient self-regulation: Implications for disinhibitory psychopathology in children. *Clinical Psychology Review.*

Newmark, C.S. (1979). *MMPI Clinical and Research Trends.* New York: Praeger.

Nichol, H., Morrison, M.D., & Rammage, L.A.(1993). Interdisciplinary approach to functional voice disorders: The psychiatrist's role. *Otolaryngology Head and Neck Surgury*, 108:643-647.

Pahn, J., & Friemart, K. (1988). Differential diagnosticsche und terminologische erwagungen bei sogenannten; funktionellen storungen im neuropsychiatrischen und phoniatrischen. Facgebiet. 2. Phoniatrischer Aspeckt. *Folia Phoniatrica*, 40, 168-174.

Patterson, C.M., & Newman, J.P. (1993). Reflectivity and learning from aversive events: Toward a psychological mechanism for the syndromes of disinhibition. *Psychological Review,* 4, 716-736.

Peter, F., Brandell, M.E. (1980). A sudy on the self-concept of children with vocal nodules. Paper presented at the ASHA convention.

Pfau, E.M. (1975). Psychologische Untersuchungsergegnisse sur atiologie der psychogenen Dysphonien. *Folia Phoniatrica*, 25, 298-306.

Rammage, L.A., Nichol, H., & Morrison, M.D. (1987). The psychopathology of voice disorders. *Human Communications Canada*, 11, 21-25.

Reiser, D.E. (1985). Reactions to illness. In: Reiser, D.E., & Schroder, A.K. (Eds.). *Patient Interviewing: The Human Dimension.* Baltimore: Williams & Wilkins.

Roy, N., & Leeper, N. (1993). Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: Perceptual and acoustic measures. Journal of Voice, 7, 242-249.

Roy, N., McGrory, J.J., & Bless, D.M. (1995). Psychological correlates of patients with vocal nodules. Paper presented at the American Speech and Hearing Association convention, New Orleans, LA.

Roy, N., McGrory, J.J., Tasko, S.M., Bless, D.M., Heisey, D., & Ford, C.N. (in press-a). Psychological correlates of functional dysphonia: An evaluation using the Minnesota Multiphasic Personality Inventory. *Journal of Voice.*

Roy, N., Bless, D.M., Heisey, D., & Ford, C.F. (in press-b). Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short- and long-term treatment outcomes. *Journal of Voice.*

Sapir, S. (1995). Psychogenic spasmodic dysphonia: A case study with expert opinions. *Journal of Voice*, 9, 270-281.

Schalen, L., & Andersson, K. (1992). Differential disgnosis and treatment of psychogenic voice disorder. *Clinical Otolaryngology*, 17, 225-230.

Stemple, J.C. (1984). Clinical Voice Pathology: Theory and Management. Columbus: Charles E. Merrill.

Stemple, J.C. (1993). Voice Therapy: Clinical Studies. St. Louis: Mosby Year Book.

Toohill, R.J. (1975). The psychosomatic aspects of children with vocal nodules. *Archives of Otolaryngology*, 101, 591-595.

Wallace, J.F., & Newman, J.P. (1991). Failures of Response Modulation: Impulsive Behavior in Anxious and Impulsive Individuals. *Journal of Research in Personality*, 25, 23-44.

Wallace, J.F., & Newman, J.P. (in press). Neurtocism and the Attentional Mediation of Dysregulatory Psychopathology.

Watson, D., Clark, L.A., & Harkness, A.R. (1994). Structures of personality and their relevance to psychopathology. *Journal of Abnormal Psychology*, 103 (1), 18-31.

White, A., Deary, I.J., Wilson, J.A. (in press). Psychiatric disturbance and personality traits in dysphonic patients. *European Journal of Disorders in Communication.*

Wilson, F.B. (1971). Emotional stress may cause voice anomalies in kids. *Journal of the American Medical Association*, 216, 2085.

Wilson, F.B., & Lamb, M. (1974) Comparison of personality characteristics of children with and without vocal nodules based on Rorscharch Protocol Interpretation. *Acta Symbolica*, 43-55.

Wilson, D.K. (1987). *Voice problems of children* (3rd.ed.). Baltimore, MD: Williams & Wilkins.

Withers, B.T., & Dawson, M.H. (1960). Psychological aspects...Treatment of vocal nodule cases. *Texas State Journal of Medicine*, 56,43-46.

Yano, Jl, Ichimura, K., Hoshino, T., & Nozue, M. (1982). Personality factors in the pathogenesis of polyps and nodules of the vocal cords. *Auris, Nasus, Larynx*, 9, 105-110.

NCVS Status and Progress Report - 12
June 1998, 135-143

Self-Organizing Map for the Classification of Normal and Disordered Female Voices

Daniel E. Callan
Ray D. Kent, Ph.D.
Department of Communicative Disorders, The University of Wisconsin-Madison
Nelson Roy
Department of Communication Disorders, Minot State University
Stephen M. Tasko
Department of Communicative Disorders, The University of Wisconsin-Madison

Abstract

Perceptual rating scales have been commonly used to describe voice quality. However, these scales appear to have only weak associations with acoustic measures. In addition, expert judges seem to differ in the acoustic factors upon which their perceptual ratings of pathological voices are based. One way to avoid the subjective nature of perceptual rating scales is to use objective acoustic measurements to classify voice quality. It is unlikely, however, that a single acoustic measure is sufficient for classification. Rather, multiple acoustic measures are needed. The goal of this research was to train a Self-Organizing Map (SOM) on various acoustic measures (amplitude perturbation quotient, degree of voice breaks, rahmonic amplitude, soft phonation index, standard deviation of the fundamental frequency, and peak amplitude variation) of the sustained vowel /a/ to enhance visualization of the multidimensional nonlinear regularities inherent in the input data space. The SOM was trained using 30 spasmodic dysphonia exemplars, 30 pre-treatment functional dysphonia exemplars, 30 post-treatment functional dysphonia exemplars, and 30 normal voice exemplars. After training, the classification performance of the SOM was evaluated. The results indicated that the SOM had better classification performance than that of a stepwise discriminant analysis over the original data. Analysis of the weight values across the SOM, by means of stepwise discriminant analysis, revealed the relative importance of the acoustic measures in classification of the various groups. The SOM provided both an easy way for visualization of multidimensional data, and enhanced statistical predictability at distinguishing between the various groups (over that conducted on the original data set). It is concluded that self-organizing maps provide an objective tool that can be used for the classification of voice disorders.

Introduction

One of the most persistent problems in the study of disordered voice quality is the efficient acoustic description of clinically relevant differences. Of the several studies that have attempted to relate perceptual ratings of disturbed voice quality to acoustic variables, only weak associations have been reported (Arends, Povel, Van Os, & Speth, 1990; Eskanazi, Childers, and Hicks, 1990, Wolfe, Fitch, & Cornell, 1995; Kent et al., 1994; Zwirner, Murry, & Woodson, 1991). Development of an acoustic analysis procedure that is both efficient and interpretable with respect to perceptual judgments of voice quality has been impeded by several conditions. First, acoustic variables which are the best predictors for a certain voice type may not be the best predictor for voices of a different type (Wolfe & Steinfatt, 1987). Second, expert judges may differ considerably from one another in the acoustic factors that are correlated with their perceptual ratings of pathological voices. Third, the dimensions used for perceptual ratings of voice quality can be highly interdependent. For example Kreiman, Gerratt, and Berke (1994) concluded that the "breathy" and "rough" voice qualities are related multidimensional constructs and that judgments of roughness depend strongly on concomitant breathiness but not vice versa. A similar interaction has been described for judgments of vocal pitch and roughness

(Wolfe & Ratusnik, 1988). Moderately to severely dysphonic vowels were matched with a significantly lower pitch than were mildly dysphonic vowels.

Solving the problem of acoustically characterizing disordered voice quality will likely require multiple acoustic measures that are differentially weighted. However, there are two main concerns when analyzing data with multiple dimensions: (1) it is often difficult to visualize data characterized by more than two or three dimensions, and (2) it is often difficult to characterize the weightings among the multiple dimensions of the data. The latter problem is especially challenging when no "gold standard" (a widely recognized referent) exists. One method to determine the weighting of the acoustic measures for the purpose of voice classification is based on the inherent nonlinear regularities that exist among the acoustic measures. In this regard, self-organizing maps may be used to set the weights of various acoustic measures in a manner that depends on such inherent nonlinear regularities.

There are three main characteristics of self-organizing maps that make them desirable as a tool to classify multidimensional data. (1) Self-organizing maps are able to take multidimensional data and represent it along a two dimensional surface of nodes. This provides for an easy way to visualize the relative distribution of the exemplars of the various groups across the surface of the self-organizing map. (2) In some cases, the nonlinear nature of the self-organizing map allows for better classification performance than traditional multivariate statistical techniques. (3) The relative weighting of the input dimensions responsible for defining the distribution of exemplars into groups across the self-organizing map can be easily discerned.

Self-organizing maps provide for a two-dimensional representational mapping, over an array of processing units (called nodes), of the multidimensional nonlinear regularities inherent in the input data space (Kohonen, 1995). Typically self-organizing maps are characterized by an input layer and a representational layer as shown in figure 1. The input layer can be thought of as containing a separate processing node for each of the various data parameters. The representational layer is composed of a two dimensional array of processing nodes, each of which receives a connection from all of the nodes in the input layer. Associated with each of these connections is a particular weight value. It is the relative value of the weights across the two dimensional node matrix that allows for the representation of features inherent in the multidimensional input data (Kohonen, 1995). This representational mapping is formed in a self-organizing unsupervised manner, that is, the values of the weights depend on characteristics of the input data and there is no teacher signal that modifies these weights based on some optimal target. Training of the self-organizing map occurs by initially randomizing the values of the weights of the connections from the input layer to the representational layer. Data samples are iteratively presented and the node in the representational layer is selected that has weight values closest in Euclidean distance to the input data. The weights that project to this node as well as its neighbors are adapted toward the value of the input data. The magnitude of the change is a function of a parameter called the learning rate (Kohonen, 1995). Each of the data samples is iteratively presented to the network several times until a stable representational mapping forms across the node array. One can easily visualize the distribution of exemplars, both those on which the self-organizing map was trained, and novel ones, across the two-dimensional node array. In addition, the weightings of the multiple input dimensions can be evaluated by observing the corresponding weight matrix of each of the input dimensions.

Self-organizing maps have been used to characterize several aspects of normal and disordered voice. In a series of studies carried out by Leinonen and colleagues (Leinonen, et al., 1992; Leinonen, et al., 1993; and Rihkanen et al., 1994), a self-organizing map was trained to detect characteristics of dysphonic voice by recognition of spectral composition. In order to determine how this map would characterize disordered voice, samples were taken from individuals classified with varying degrees of dysphonia using a method derived from the GRBAS perceptual rating scale (Hirano, 1981). The voices were classified based on the degree of dysphonia, roughness, breathiness, and strain. Samples of the vowel /a/ from both normal and disordered voices were presented to the self-organizing map. It was found that the trajectory pattern across the map differed for dysphonic and normal voices. It was also found that rough and breathy voices had different trajectory patterns. However, the self-organizing map was not able to distinguish the degree of pathology. In a later study conducted by Rihkanen et al., (1994), the SOM was trained using spectral vectors from both normal and pathologic voices. This method of training allowed for the relative degree of pathology to be encoded across the SOM.

Overall, the self-organizing maps used by Leinonen and colleagues are very successful in providing a two-dimensional visual representation of multidimensional properties of voice and speech that is easy to understand. One of the disadvantages of using multiple spectral dimensions as input is that it is difficult to evaluate which acoustic properties are important in the map organization that makes classification of normal and disordered voices possible.

The SOM has also been used to categorize voice disorders by using the dimensions of a perceptual rating scale as input. In the study carried out by Leinonen, et al., (1997), the categorization of various forms and degrees of dysphonia was accomplished by using perceptual ratings of pathology, roughness, breathiness, strain, and asthenia, as input dimensions to train the SOM. Five different categories of voice quality were formed across the SOM. The

categories were organized from right to left across the SOM depending on the relative degree of severity. The SOM trained on various dimensions of a perceptual rating scale was successful in categorizing both the degree of pathology as well as the ratio of breathiness and roughness across the map. However, there are several disadvantages in the use of perceptual rating scales, mainly due to their subjective nature. As was pointed out by Leinonen, et al., (1997), it would be interesting to determine if the five categories of pathology formed in the multidimensional perceptual space had corresponding acoustic patterns.

The goal of the present research is to develop an objective means of assessing voice quality, based on multiple acoustic measures, by utilizing a self-organizing map (SOM) to classify samples of normal and disordered voice in an unsupervised manner. Unsupervised self-organization of the map allows for the weightings of the input acoustic measures to be determined based on the inherent nonlinear regularities of the input data space. Unsupervised learning is particularly advantageous when no gold standard exists for classification. It is hoped that by using a self-organizing map, which is trained using various objective acoustic measures (e.g. amplitude perturbation quotient (APQ), degree of voice breaks (DVB), rahmonic amplitude (RAM), soft phonation index (SPI), standard deviation of the fundamental frequency (STD), and peak amplitude variation (VAM)) that the subjective nature inherent to perceptual rating scales will be avoided. In this study voice samples from sustained vowel /a/ productions for four groups of subjects were used to train the self-organizing map. These groups consist of individuals with normal voice, individuals with verified adductor spasmodic dysphonia, individuals with functional dysphonia before behavioral management, and the same individuals with functional dysphonia after behavioral management. Analysis of the two-dimensional self-organizing map provides an easy way to visualize and evaluate the classification of the various groups. It is hoped that the self-organizing map will provide enhanced predictability for classification over that of traditional statistical analyses. Analysis of the SOM also provides an easy way to visualize and evaluate the weighting of the various acoustic measures underlying the representational map.

Method

Subjects

The voice samples of subjects included in the study consisted of individuals with verified adductor spasmodic dysphonia, SD, (N=30), individuals with functional dysphonia pre-treatment, PR, (N=30), individuals with functional dysphonia post-treatment, PS, (N=30), (the voice samples for the PR and PS groups were taken from the same individuals before and after treatment), and individuals categorized as having normal phonation, NR, (N=30). Individuals were given a diagnosis of functional dysphonia if they had no apparent structural or neurological pathology and had a positive sustained response to behavioral management. Individuals received the diagnosis of spasmodic dysphonia following complete voice evaluation which failed to identify structural pathology sufficient to explain the nature and quality of voice symptoms. Voice symptoms were perceptually consistent with adductor spasmodic dysphonia and patients failed to respond to behavioral management. All participants were female non-smokers who spoke English as a first language. Voice samples were acquired by asking the participants to sustain the vowel /a/ at a comfortable pitch and loudness, for at least five seconds. The voice samples came from two different sources. Functional dysphonic and spasmodic dysphonic voice disordered samples were acquired at a hospital based voice pathology clinic at the University of Wisconsin-Madison. The normal and 5 of the spasmodic dysphonic voice samples were taken from the Voice Disorders Database version 1.03. compiled by the Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary (Boston, MA), Kay Elemetrics Corp., 1994. Studio level recordings with a sampling rate of 25 kHz were used for both sources.

Acoustic Measures

The central one second segment of each sustained /a/ production was extracted for acoustic analyses. The acoustic analyses included 22 measures of the MDVP (Multi-Dimensional Voice Profile) (Kay Elemetrics Corp.), using the default measurement parameters. If automated voice classification procedures are to be of practical clinical utility, they would require input variables that are easily and widely available. MDVP meets this requirement. In addition to the MDVP analysis, a cepstral analysis using CSL to acquire the amplitude and quefrency of the dominant rahmonic was also conducted. The cepstrum is a log power spectrum of a log power spectrum. The cepstrum analysis shows the dominant energy corresponding to the harmonic peaks of the spectrum. This is called the dominant rahmonic. The time axis of the cepstrum is called quefrency. The quefrency of the dominant rahmonic corresponds to the fundamental period of the signal. One of the motivating factors for including the cepstrum-based measures is the finding made by Hillenbrand et al. (1994) and Hillenbrand and Houde (1996), that suggests that cepstrum based measures are good predictors of breathiness in sustained vowels. The cepstrum measures were acquired over the central part of the down sampled signals (12500 Hz) using a 1024 point FFT with pre emphasis of .9 and a Hamming window weighting.

The acoustic parameters used to train the self-organizing map were motivated by both their apriori expectancy to capture aspects of voice quality as well as their ability to classify the four different groups as revealed by a

Table 1.
Stepwise Discriminant Analysis of the Original Data for all Four Groups (SD, PR, PS, NR) N=120

Step Entered	Variable Entered	Wilks' Lambda	df	F	df1	df2	Sig.
1	RAM	.358	116	69.4	3	116	.001
2	VAM	.273	116	35.0	6	230	.001
3	STD	.242	116	24.4	9	277.6	.001
4	APQ	.221	116	19.2	12	299.3	.001
5	SPI	.199	116	16.4	15	309.6	.001
6	DVB	.178	116	14.8	18	314.4	.001

- At each step the variable that minimizes the overall Wilks' Lambda is entered.
- Minimum significance of F to enter is .05.
- Maximum significance of F to remove is .1.

stepwise discriminant analysis. The results of the stepwise discriminant analysis are displayed in table 1. The variables entered at each step were those that minimize the overall Wilks' lambda. When shimmer was taken out of the analysis, amplitude perturbation quotient was entered in the stepwise discriminant analysis. This resulted in better classification of the four groups (68.3% with amplitude perturbation quotient entered versus 65.8% with shimmer entered). Based on apriori assumptions and the results of the stepwise discriminant analysis the following acoustic variables were selected to train the self-organizing map:

APQ - Amplitude Perturbation Quotient gives an evaluation in percent of the variability of the peak-to-peak amplitude within the analyzed voice sample at smoothing factor 11 periods (taken from Voice Disorders Database, Kay Elemetrics Corp., 1994).

DVB - Degree of Voice Breaks (in percent) the ratio of the total duration where the fundamental frequency cannot be tracked to the time of the complete voice sample (taken from Voice Disorders Database, Kay Elemetrics Corp., 1994).

RAM - Rahmonic Amplitude denotes the amplitude of the first dominant peak in the cepstral analysis which corresponds to the harmonic peak of the spectrum of the signal.

SPI - Soft Phonation Index is an average ratio of the lower-frequency to the higher frequency harmonic energy (taken from Voice Disorders Database, Kay Elemetrics Corp., 1994).

STD - Standard deviation of the fundamental frequency (in Hertz) in the vocalization (taken from Voice Disorders Database, Kay Elemetrics Corp., 1994).

VAM - Peak amplitude variation represents the relative standard deviation (in percent) of the period-to-period calculated peak-to-peak amplitude (taken from Voice Disorders Database, Kay Elemetrics Corp., 1994).

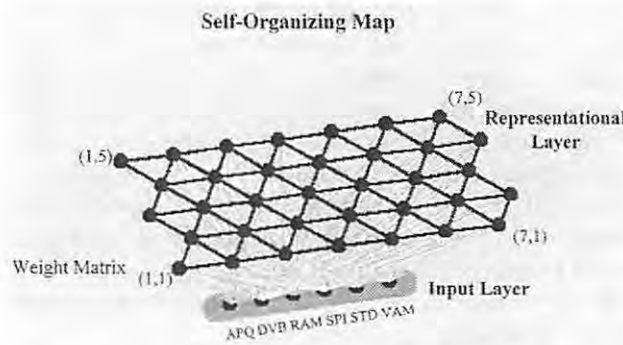


Figure 1. Figure one depicts the architecture of the self-organizing map (SOM). The input layer is composed of nodes (processing units) that encode the values of each of the input data parameters. In this case the six data parameters are amplitude perturbation quotient (APQ), degree of voice breaks (DVB), rahmonic amplitude (RAM), soft phonation index (SPI), standard deviation of the fundamental frequency (STD), and peak amplitude variation (VAM). See text for details.

Network Architecture and Training

Input: The self organizing map (SOM) was trained using 30 exemplars from each of the groups (Spasmodic Dysphonia, Pre-Treatment Functional Dysphonia, Post-Treatment Functional Dysphonia, and Normal Phonation). The six input dimensions used to train the SOM consisted of the acoustic measures of amplitude perturbation quotient (APQ), degree of voice breaks (DVB), rahmonic amplitude (RAM), soft phonation index (SPI), standard deviation of the fundamental frequency (STD), and peak amplitude variation (VAM). The entire input data space was normalized to have a mean of zero and a variance of one. Input normalization is thought to be important to ensure that the input dimensions are not disproportionately scaled. Input dimensions that vary greatly in scale may cause biases in the corresponding weight matricies of the nodes in the representational layer (Kohonen, 1995).

Architecture: The self-organizing map contained a 6 node input layer and a 7 by 5 node representational layer (See Figure 1). The dimensions of the representational layer were selected based on assumptions that the 120 exemplars would be distributed across the map in regions defining the four groups. A much larger representational layer is needed if a greater number of exemplars are used for training. Each of the nodes in the input layer encoded one of the six input dimensions, (APQ, DVB, RA, SPI, STD, VAM). The 7 by 5 node representational layer was organized in a hexagonal neighborhood lattice as recommended by Kohonen (1995) for better visual inspection of the resultant map. Each node in the representational layer has associated with it a neighborhood radius that defines the number of rings of surrounding nodes that can be affected by activation of that particular node. The size of the neighborhood radius of each of the nodes in the representational layer was altered throughout training. Each of the nodes in the representational layer had weighted connections to all six input nodes.

Training: The SOM_PAK software package developed by Kohonen, Hynninen, Kangas, and Laaksonen, (1996), was used to train the self-organizing map. The self-organizing map forms by means of projecting the probability distribution inherent in the multidimensional input space onto the two dimensional grid of nodes composing the representational layer. Map formation is accomplished by iteratively presenting input and allowing for the weighted connections from the representational layer to be corrected toward the input value. For more details regarding the self-organizing map training algorithm see Kohonen (1995), and Kohonen et al (1996).

Ten maps were trained using the same procedure but different initial random values for the connection weights. The map with the lowest overall quantization error was selected for analysis. The maps were trained in two stages: the first stage was for 2000 steps through the input data and included an initial neighborhood radius decreased from 7 to 0 and a learning rate of .07, the second stage was for 12000 steps through the input data and included a neighborhood radius decreased from 2 to 0 and a learning rate of .02. The first stage of training was carried out with a larger initial neighborhood radius and learning rate to allow for proper ordering of weight vectors of the nodes in the representational layer (Kohonen, et al, 1996). The second stage of training was carried out to fine tune the map (Kohonen, et al, 1996). After training, a Sammon Mapping (Sammon, 1969), across the six weight values associated with each of the nodes in the 7 by 5 representational layer, was conducted in order to determine the approximate Euclidean separation of the nodes in two dimensional space. The separation and position of the nodes in figures 2 and 3 is determined by the Sammon mapping. All figures were constructed using Matlab version 4.2c.1, MathWorks, Inc., 1994. All statistical analyses were conducted using SPSS version 7.5.1, SPSS, Inc., 1996.

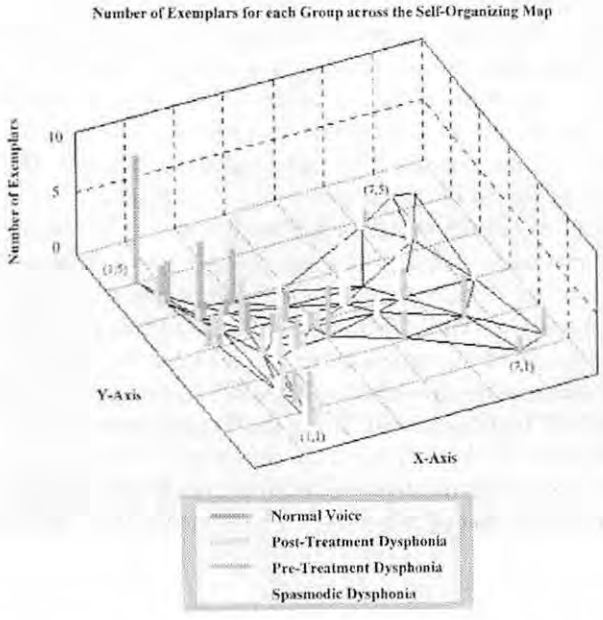


Figure 2. Figure two depicts the number of exemplars for each group across the nodes of the self-organizing map (spasmodic dysphonia = SD; pre-treatment functional dysphonia = PR; post-treatment functional dysphonia = PS; normal voice = NR). The x and y axes represent the Euclidean coordinates for each of the nodes defined by the Sammon mapping (Sammon, 1969). The node identifiers for each of the four corners of the SOM are given (1,1), (1,5), (7,5), and (7,1). See text for details.

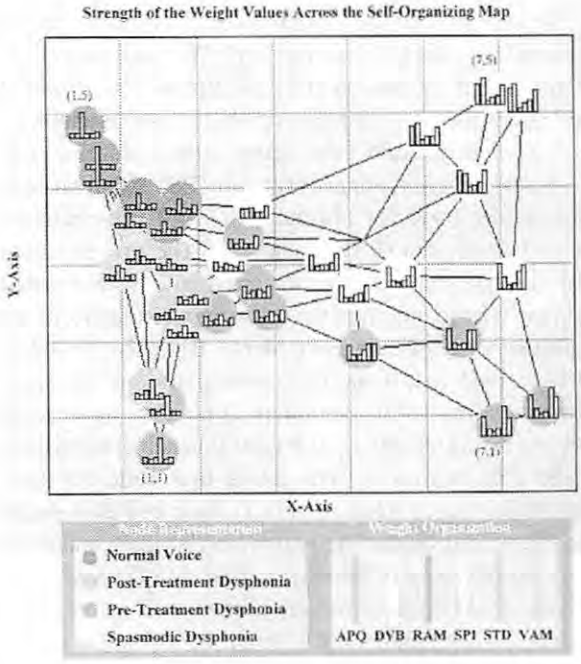


Figure 3. Figure 3 depicts the strength of the weight values across the group node representations (blue = spasmodic dysphonia SD; pre-treatment functional dysphonia = PR; post-treatment functional dysphonia = PS; normal voice = NR) of the SOM for each of the six underlying acoustic parameters (amplitude perturbation quotient (APQ), degree of voice breaks (DVB), rahmonic amplitude (RAM), soft phonation index (SPI), standard deviation of the fundamental frequency (STD), and peak amplitude variation (VAM)). The x and y axes represent the Euclidean coordinates for each of the nodes defined by the Sammon mapping (Sammon, 1969). The node identifiers for each of the four corners of the SOM are given (1,1), (1,5), (7,5), and (7,1). The small bar plot on top of each node represents the strength of the corresponding weight values for each of the underlying acoustic parameters (from left to right - APQ, DVB, RAM, SPI, STD, and VAM). The relative importance of the underlying acoustic parameters in classifying the various groups across the SOM can be determined by comparing the pattern of the corresponding weight strengths to the regions of the SOM that represent the four groups. See text for details.

Results and Discussion

The two dimensional nature of the self-organizing map (SOM) allows for an easy way to visualize the relative distribution of exemplars from each of the four groups. Figure 2 displays the number of exemplars from each of the groups that are coded by the nodes of the SOM (a numerical depiction is given in table 2). The node positions are determined by the x and y Euclidean coordinates defined by the Sammon mapping. The lines connecting the nodes in figure 2 represent the underlying hexagonal neighborhood lattice of the SOM. Nodes that are closer together on the lattice have more similar weight values than nodes that are spread farther apart.

Visual inspection of the map (shown in figure 2) indicates that there is a dense clustering of normal voice (NR) exemplars in the region of node (1,5) (compare with table 1). One third of the normal voice exemplars, 10, are represented by a single node (node (1,5)). The post-treatment functional dysphonia (PS) exemplars are clustered around nodes that code for normal (NR) voice exemplars. There is a fair degree of overlap between the nodes that code for post-treatment functional dysphonia (PS) exemplars and the nodes that code for normal (NR) and pre-treatment functional dysphonia (PR) exemplars. The pre-treatment functional dysphonia exemplars are sparsely distributed throughout a large region of the self-organizing map, as are the spasmodic dysphonia exemplars. Both the pre-treatment functional dysphonia (PR) exemplars and the spasmodic dysphonia (SD) exemplars appear to have some degree of overlap with nodes that code for each of the groups (SD and PR) as well as with nodes that code for post-treatment functional dysphonia (PS). One can also see in figure 2 that the hexagonal neighborhood lattice is much less densely concentrated in the region of the map representing spasmodic dysphonia (SD) exemplars as well as pre-treatment functional dysphonia (PR) exemplars than the region of the map representing normal voice (NR) exemplars and post-treatment functional dysphonia (PS) exemplars.

Classification Performance

In order to determine how well the self-organizing map classified the various groups, each node was given a representation corresponding to one of the four groups (SD, PR, PS or NR). Node representation is determined by the maximum number of exemplars from a group that fall on a particular node. In the case in which the maximum number of exemplars that are coded by a particular node is equal between two or more groups, or the case in which a node does not code for any exemplars, node representation is determined by means of majority voting of the group membership of the five nearest exemplars (in Euclidean space) to the node in question. The resulting node representations are displayed as large circles over the node matrix in figure 3. Comparison of the node representations given in figure 3 with the number of exemplars from each of the groups that are coded by the nodes of the SOM given in figure 2 and table 2, allows for an easy way to visualize and evaluate the classification of the exemplars from each of the four groups.

The classification performance of a stepwise discriminant analysis over the original data and the classification performance of the self-organizing map (SOM) using the same input are given in tables 3 and 4. The results indicate that the classification performance for the self-organizing map is higher than that for the stepwise discriminant analysis. The SOM classified 75.8% of the total exemplars (N=120) correctly (see table 4). Whereas, the stepwise discriminant analysis only classified 68.3% of the total exemplars correctly (see table 3). The greatest difference in classification performance between the SOM and the stepwise discriminant analysis is for the pre-treatment functional dysphonia (PR) exemplars. The SOM classified 73.3% of the 30 PR exemplars correctly (see table 4). Whereas, the stepwise discriminant analysis only classified 50% of the 30 PR exemplars correctly (see table 3). The SOM also demonstrates better classification performance for spasmodic dysphonia SD and post-treatment functional dysphonia PS exemplars over that of the stepwise discriminant analysis. The SOM classified 73.3% of both the 30 SD and the 30 PS exemplars correctly (see table 3). Whereas, the stepwise discriminant analysis only classified 66.7% of the 30 SD exemplars and 70% of the PS exemplars correctly (see table 3). In the case of classification of normal voice (NR) exemplars the stepwise discriminant analysis is slightly better than that of the SOM. The stepwise discriminant

Table 2.
Number of Exemplars for each Group Across the Self-Organizing Map

	X-NODE AXIS																											
	1				2				3				4				5				6				7			
Y-Node Axis	SD	PR	PS	NR	SD	PR	PS	NR	SD	PR	PS	NR	SD	PR	PS	NR	SD	PR	PS	NR	SD	PR	PS	NR	SD	PR	PS	NR
5	0	0	0	10	0	0	1	2	0	0	0	5	0	0	1	4	1	0	0	0	3	1	0	0	3	0	0	0
4	0	0	1	3	0	0	4	1	0	0	0	1	0	1	0	1	0	0	0	0	1	1	0	0	1	0	0	0
3	0	1	2	1	0	0	2	0	0	0	4	2	2	1	0	0	5	0	1	0	4	2	0	0	1	0	0	0
2	1	0	2	0	0	0	2	0	1	2	2	0	1	1	1	0	1	0	0	0	0	3	0	0	0	2	0	0
1	2	4	1	0	0	2	1	0	2	0	4	0	1	2	1	0	0	4	0	0	0	2	0	0	0	1	0	0

Table 3.
Classification Results Stepwise Discriminant Analysis Over Original Data

Total Correct	82 68.3%	Predicted							
N=120		SD		PR		PS		NR	
		#	%	#	%	#	%	#	%
Actual	SD N=30	20	66.7	3	10.0	7	23.3	0	0
	PR N=30	5	16.7	15	50.0	9	30.0	1	3.3
	PS N=30	1	3.3	2	6.7	21	70.0	6	20
	NR N=30	0	0	0	0	4	13.3	26	86.7

Table 4.
Classification Results Self-Organizing Map

Total Correct	91 75.8%	Predicted							
N=120		SD		PR		PS		NR	
		#	%	#	%	#	%	#	%
Actual	SD N=30	22	73.3	4	13.3	4	13.3	0	0
	PR N=30	5	16.7	22	73.3	3	10	0	0
	PS N=30	1	3.3	4	13.3	22	73.3	3	10
	NR N=30	0	0	1	3.3	4	13.3	25	83.3

Table 5.
Stepwise Discriminant Analysis of the Node Weights for Each of the Pairwise node Classification Across the SOM

	SD (N=11)							PR (N=11)							PS (N=11)						
	Loaded Vars. step	Wilks' lambda	df	F	df	sig	Class. percent	Loaded Vars. step	Wilks' lambda	df	F	df	sig	Class. percent	Loaded Vars. step	Wilks' lambda	df	F	df	sig	Class. percent
PR N=10	1. DVB	.736	19	6.8	19	.001	SD=54.5 PR=90.0														
	w/o DVB 1. VAM	.774	19	5.6	19	.029	SD=90.9														
	2. STD	.365	19	15.6	18	.001	PR=90.0														
PS N=8	1. VAM	.190	17	72.5	17	.001	SD=90.9	1. VAM	.509	14	15.4	16	.001	PR=90.0							
	2. APQ	.112	17	63.2	16	.001	PS=100	2. SPI	.381	14	12.2	15	.001	PS=87.5							
NR N=6	1. RAM	.189	15	64.2	15	.001	SD=100 NR=100	1. RAM	.179	14	65.7	14	.001	PR=100 NR=100	1. RAM	.455	12	14.4	12	.003	PS=100
															2. VAM	.234	12	18.0	11	.001	NR=100

- At each step the variable that minimizes the overall Wilks' Lambda is entered.
- Minimum significance of F to enter is .05.
- Maximum significance of F to remove is .1.

analysis classified 86.7% of the NR exemplars correctly (see table 3). Whereas, the SOM classified 83.3% of the NR exemplars correctly (see table 4).

Weighting of the Acoustic Measures

The two-dimensional nature of the self-organizing maps (SOM) provides an easy way to visualize and evaluate the weighting of the various acoustic measures underlying the representational map. It is important to note that a very low weight value may be just as important as a very high one in discriminating the various groups. The small bar plot on top of each node in figure 3 represents the strength of the corresponding weight values for each of the underlying acoustic parameters. The relative importance of the underlying acoustic parameters in classifying the various groups across the SOM can be determined by comparing the pattern of the corresponding weight values to the regions of the SOM that represent the four groups. In order to determine which acoustic parameters are responsible for classification of the four groups across the SOM, stepwise discriminant analyses were carried out over the weights for each of the pairwise group node representations (the results for each of the analyses are given in table 5). The results of the stepwise discriminant analysis for spasmodic dysphonia (SD) node representations and pre-treatment functional dysphonia (PR) node representations indicate that when degree of voice breaks (DVB) is withheld from the analysis, peak amplitude variation (VAM) and standard deviation of the fundamental frequency (STD) account for 90.9% of SD node classification and 90% of PR node classification (see table 5 and figure 3). Figure 3 indicates that SD node representations have higher VAM (6th bar) weight values relative to STD (5th bar) weight values than PR node representations.

The results of the stepwise discriminant analysis for spasmodic dysphonia (SD) node representations and post-treatment functional dysphonia (PS) node representations indicate that peak amplitude variation (VAM) and amplitude perturbation quotient (APQ) account for 90.9% of SD node classification and 100% of PS node classification (see table 5). Figure 3 indicates that SD node representations are characterized by high VAM (6th bar) and high APQ (1st bar) weight values, whereas, PS node representations are characterized by low VAM and low APQ weight values. The results of the stepwise discriminant analysis for pre-treatment functional dysphonia (PR) node representations and post-treatment functional dysphonia (PS) node representations indicate that peak amplitude variation (VAM) and the soft phonation index (SPI) account for 90% of PR node classification and 87.5% of PS node classification (see table 5). Figure 3 indicates that PR node representations have higher VAM (6th bar) weight values or higher SPI (4th bar) weight values than PS node representations.

The results of the stepwise discriminant analysis for spasmodic dysphonia (SD) node representations and normal voice (NR) node representations indicate that rahmonic amplitude (RAM) accounts for 100% of SD node classification and 100% of NR node classification (see table 5). Figure 3 indicates that SD node representations have much lower RAM (3rd bar) weight values than NR node representations. The same pattern of results exists for PR and NR node classification (see table 5 and figure 3). The results of the stepwise discriminant analysis for post-treatment functional dysphonia (PS) node representations and normal voice (NR) node representations indicate that rahmonic amplitude (RAM) and peak amplitude variation (VAM) account for 100% of PS node classification and 100% of NR node classification. Figure 3 indicates that PS node representations have combined lower RAM (3rd bar) and VAM (6th bar) weight values than PR node representations.

Based on the results of the pairwise stepwise discriminant analyses (table 5) as well as visual inspection of the weight values in figure 3, the following generalizations can be made regarding the relative importance of the acoustic parameters in determining the group node representations: SD node representations are characterized by relatively high values of VAM, APQ, and DVB, as well as low values of RAM. Most of the nodes with SD representations are characterized by relatively high values of STD and VAM, with low values of RAM. Two nodes with PR representations, in the region of node (1,1), have high SPI levels. PS node representations are characterized by mid range values of RAM and SPI, as well as low values of APQ and VAM. NR node representations are characterized by high RAM values and relatively low values of all other acoustic parameters. Overall, it appears that disordered voice (SD and PR) is characterized by greater variability in amplitude (both long term (APQ, smoothing factor 11 periods) and short term (VAM, period-to-period)), as well as fundamental frequency (STD). Voice that is considered non-pathologic (NR and PS) is characterized by a low degree of variability in amplitude (APQ and VAM) and fundamental frequency (STD), as well as a high degree of harmonic energy (RAM).

Conclusion

The self-organizing map provided an objective means of assessing voice quality, based on multiple acoustic measures. Analysis of the self-organizing map allowed for an easy way to visualize and evaluate the classification of exemplars of the various groups. The results indicate that the SOM does a fairly good job at classifying the four groups. The SOM demonstrated better classification performance over that given by a stepwise discriminant analysis of the original data. Analysis of the SOM also provided for a description of the relative contribution of the various acoustic measures underlying the representational map that are relevant in classification of the various groups. Further research needs to be conducted using a larger sample size to train the SOM. A larger sample size will generate maps that are able to classify novel exemplars accurately. The SOM obtained by this method may be a valuable complement to classifications based on auditory-perceptual ratings. It would be interesting to compare how perceptual ratings of voice quality matched up with where an exemplar was located on the SOM based on its acoustic properties.

We regard the results of this study as a promising initial step into the use of self-organizing maps with multiple acoustic measures of phonatory function. Certain limitations of this present work should be noted in future applications. First, sustained vowel phonation is a convenient sample for analysis, but it probably is not sufficient to reveal the complexity of the phonatory disorders addressed here. For example, spasmodic dysphonia can be highly task sensitive, and it is possible that somewhat different features of voice disturbance should be used with tasks such as conversation or passage reading. Still, the successful classification in this study indicates that even with the relatively simple task of sustained vowel phonation, it is possible to delineate differences in clinical populations. Second, the small number of acoustic measures used in this study may not be the best ones to use in more extensive studies of voice disorders in various clinical populations. Selection of acoustic measures to be used as input to the SOM is matter that requires further work. With a larger number of exemplars, it would be possible to increase the number of acoustic measures used as input to the SOM. Third, the SOM offers interesting possibilities for the comparison of different types of analysis, including the use of input data from derived

acoustic measures (as in the case of the present study), spectral vectors (Leinonen et al., 1992), and perceptual ratings (Leinonen et al., 1997). It would be valuable to compare the SOM results obtained with different types of input data. Such a comparison would be especially pertinent to understanding the relationships between perceptual ratings and acoustic analyses of voice qualities.

Acknowledgments

This work was supported in part by NIH research grant number 5 R01 DC 00319-11 as well as by the National Center for Voice and Speech Grant number P60 00976 from the National Institute on Deafness and Other Communication Disorders.

References

Arends, N., Povel, d. J., Van Os, E., & Speth, L. (1990). Predicting voice quality of deaf speakers on the basis of glottal characteristics. Journal of Speech and Hearing Disorders, 29, 156-170.

Eskenazi, L., Childers, D. G., & Hicks, D. M. (1990). Acoustic correlates of vocal quality. Journal of Speech and Hearing Research, 33, 298-306.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. Journal of Speech & Hearing Research. 37(4): 769-78.

Hillenbrand, J. & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality - Dysphonic voices and continuous speech. Journal of Speech & Hearing Research. 39(2): 311-321.

Hirano, (1981). Clinical Examination of Voice. Vienna, Springer, 83-84.

Kay Elemetrics Corp. (1994). Voice Disorders Database Version 1.03. Massachusetts Eye and Ear Infirmary, Voice and Speech Lab, Boston, MA.

Kent, . D., Kim, H., Weismer, G., Kent, J. F., Rosenbek, J. C., Brooks, G. R., & Workinger, M. (1994). Laryngeal dysfunction in neurological disease: Amyotrophic lateral sclerosis, Parkinson's disease, and stroke. Journal of Medical Speech-Language Pathology, 2, 157-175.

Kohonen, T. (1995). Self-Organizing Maps. Berlin, Springer, 77-127.

Kohonen, T., Hynninen, J., Kangas, J., and Laaksonen, J. (1996). SOM PAK: The Self-Organizing Map Program Package: Report A31. Helsinki University of Technology: Laboratory of Computer and Information Science.

Kreiman, J., Gerratt, B. R., & Burke, G. S. (1994). The multidimensional nature of pathologic vocal quality. Journal of the Acoustical Society of America, 96, 1291-1313.

Leinonen, L., Hiltunen, T., Laakso, M., Rihkanen, H., and Poppius, H. (1997). Categorization of voice disorders with six perceptual dimensions. Folia Phoniatr Logop, 49, 9-20.

Leinonen, I., Kangas, J., Torkkola, K., & Juvas, A. (1992). Dysphonia detected by pattern recognition of spectral composition. Journal of Speech and Hearing Research, 35, 287-295.

Leinonen, Hiltunen, T., Kangas, J., Juvas, A., and Rihkanen, H. (1993). Detection of dysphonia by pattern recognition of speech spectra. Scand. J. Logop Phoniatr., 18, 159-167.

Rihkanen, H., Leinonen, I., Hiltunen, T., and Kangas, J. (1994). Spectral pattern recognition of improved voice quality. J. Voice, 8, 320-326.

Sammon, J. (1969). A non-linear mapping for data structure analysis. IEEE Trans Comp, 18, 401-409.

Wolfe, V., Fitch, J., & Cornell, R. (1995). Acoustic prediction of severity in commonly occurring voice problems. Journal of Speech and Hearing Research, 38, 273-279.

Wolfe, V. I., & Ratusnik, D. L. (1988). Acoustic and perceptual measurements of roughness influencing judgments of pitch. Journal of Speech and Hearing Disorders, 53, 15-22.

Wolfe, V. I., & Steinfatt, T. M. (1987). Prediction of vocal severity within and across voice types. Journal of Speech and Hearing Research, 30, 230-240.

Zwirner, P., Murry, T., & Woodson, G. E. (1991). Phonatory function of neurologically impaired patients. Journal of Communication Disorders, 24, 287-300.

NCVS Status and Progress Report - 12
June 1998, 145-152

Messa Di Voce: An Investigation of the Symmetry of Crescendo and Decrescendo in a Singing Exercise

Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
George I. Shirley, D.M.A.
School of Music, The University of Michigan
Russel Long, M.S.
Elaine Stathopoulos, Ph.D.
Lorraine O. Ramig, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Linda M. Carroll, M.Phil
Teacher's College, Columbia University
Actors Studio, MFA Program, School of Dramatic Arts, The New School for Social Research
William D. Riley, M.M.
Actors Studio, MFA Program, School of Dramatic Arts, The New School for Social Research

Introduction

Messa di voce (to put or place the voice) is a vocal exercise that dates to the camerata, a group of Italian composers and writers of the 16th century who gave birth to opera. The exercise is a crescendo followed by a decrescendo at a constant pitch and vowel. The claim is that if this exercise is performed smoothly and consistently, with many repetitions, the voice will achieve its proper "placement". The Italian composer Gioacchino A. Rossini (1792-1868) gave this exercise high prominence by placing it first in a book of vocalises entitled *Gorgheggi e Solfeggi* (1825). He considered the vocalises to be fundamental to proper vocal development, reflecting the discipline imposed by the *bel canto* (beautiful singing) tradition. Rossini directed the singer to practice each exercise three times daily: the first time slowly and softly, the second time fast and soft, the third time fast and loud. The *messa di voce* exercise is one of the most difficult of all vocalises to master. It is as if Rossini was saying "Conquer this challenge successfully before you even think of attempting to do justice to the exercises that follow." He begins *messa di voce* at low C (C_3 or C_4 depending on gender) and extends it an octave and a half (to G_5 or G_6).

The Italian singing teacher Giuseppe Concone (1801-1861) also made it the first exercise in his Thirty Daily Exercises (Reprinted by Schirmer, 1894), a practice book found in many vocal studios today. Concone was a native of Turino, Italy, who made capital of his ability to impart vocal wisdom to those more gifted than he. His books of vocalises remain popular today, their rigors never out of date.

Many classical singers use the *messa di voce* exercise to improve their vocal control. It requires many combinations of airflow, lung pressure and laryngeal muscle activity. Because the exercise uses a gradual transition from *pianissimo* to *fortissimo* (and back), it challenges the singers' ability to independently change intensity while pitch, vowel, and timbre are held as constant as possible. This ability is a hallmark of many accomplished classical singers. The *messa di voce* exercise has also been touted as an exercise to improve the coordination of vocal registers. Miller (1986) states that "*Messa di voce* is the classic device for achieving mastery of a wide range of dynamic contrast. One begins at *pianissimo* level with a sustained tone, crescendoing to *fortissimo*, then decrescendoing back to

pianissimo while maintaining uniform timbre. Ideally, the whole dynamic range of *messa di voce* should be possible on every pitch within the entire vocal compass, yet it is doubtful that more than a handful of great singers have ever achieved that goal, in any generation." The singing voice is perceptually register-free when the singer is able to balance the respiratory and laryngeal forces in this exercise.

Given the value placed on this vocal exercise by singers, composers and singing teachers, we felt it appropriate to study some of its acoustic and physiologic characteristics. *Messa di voce* is well suited for laboratory study because it requires no accompaniment, is only on the order of 10 seconds long, is sung on a single vowel, and can be executed during various experimental conditions (e.g., with and without various sensors attached to the body).

Most singers find the exercise difficult at some pitches. In particular, singers who are trained in Western classical styles are concerned about "voice breaks", sudden shifts from one voice quality to another as pitch or loudness is changed. For male singers, for example, the involuntary break from falsetto register (the boyish voice) to chest register (the more mature male voice) is usually an embarrassment because it suggests lack of control. The break often happens at high pitches, particularly when intensity is increased or decreased. For some, it brings back unpleasant memories of puberty, when the voice was truly uncontrolled and unpredictable. For female singers, the break is not so catastrophic because it happens in the lower part of their pitch range, but it often distinguishes singers who can produce high intensity low notes (in chest voice) from those who cannot.

The purpose of this study was to address the following questions: (1) Is the *messa di voce* executed as intended, that is, a symmetric increase and decrease of intensity, (2) if not, can the intensity asymmetry be attributed to a non-uniform depletion of lung volume, (3) do singers vary in their ability to avoid register breaks in *messa di voce*, and (4) for those who cannot avoid register breaks, are the breaks more frequent and severe in the decrescendo portion than in the crescendo portion?

Methods

The methods employed in this study have all been used previously (Rothenberg, 1973; Smitheran and Hixon, 1981; Watson et al., 1989, 1990; Sundberg, Scherer and Titze, 1993; Stathopoulos and Sapienza, 1993; Thomasson and Sundberg, 1997). They involve placement of an airflow transducer over the mouth and nose, a pressure transducer in the oral cavity behind the lips, an electroglottograph on the neck and two inductive coil movement transducers on the upper body, one over the ribcage and one over the abdomen. With this combination of transducers, the respiratory and phonatory physiology can be inferred for the *messa di voce* exercise. More detail on the instrumentation is given below.

Subjects

Three male and three female singers were used for this study. Table 1 shows a profile of their age, voice category, and experience (private lessons and concertizing). One tenor (GS) had been soloist for the Metropolitan Opera Company in New York for 11 years. The coloratura soprano and lyric baritone (LC and WR) had sung in recitals for many years prior to becoming full-time teachers. The lyric sopranos KA and SK were accomplished graduate students at the University of Colorado Music School with recital and opera workshop experience and the tenor IT was an amateur soloist. The subjects were all volunteers in this study and completed the tasks in approximately 2 - 3 hours. Much of the time was spent in respiratory equipment calibration for each individual singer.

Table 1.
Profile of Singers Studied

Subject	Age	Gender	Category	Experience
KA	25	F	Lyric Soprano	8 years
LC	39	F	Coloratura Soprano	22 years
SK	25	F	Lyric Soprano	7 years
WR	43	M	Lyric Baritone	24 years
GS	61	M	Lirico Spinto Tenor	40 years
IT	54	M	Lyric Tenor	20 years

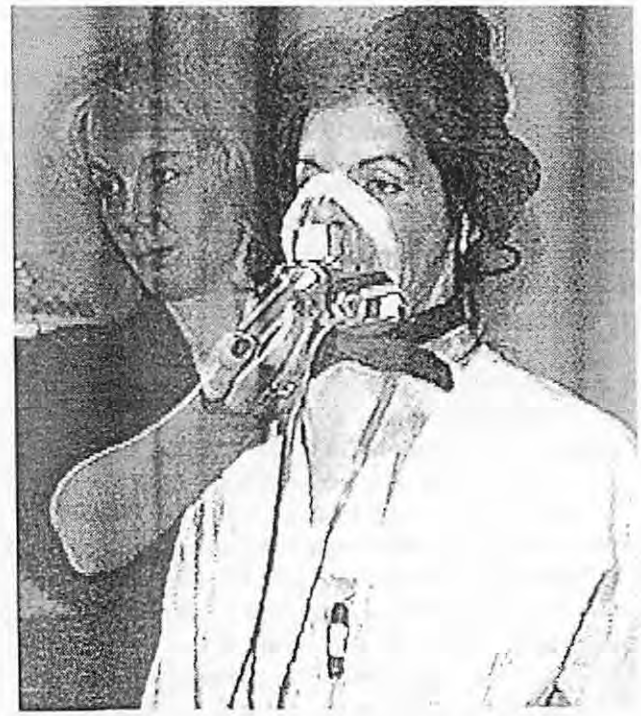


Figure 1. Photograph of subject (singer) assisted by one experimenter in keeping a tight seal on the face mask.

Instrumentation

The experiment took place in an IAC recording booth (3.15 m wide x 3.50 m long x 2.59 m high) at the Wilbur James Gould Voice Research Center in Denver, Colorado. The subjects stood upright, with inductive coils (bands) surrounding their ribcage and abdomen, the electroglottograph attached to their neck, and (for some utterances) a flow mask placed over their mouth and nose. An experimenter provided the pitches with an electronic keyboard. A second experimenter stood behind the subject to assist in maintaining pressure between the flow mask on the facial skin to avoid air leakage (Figure 1). Table 2 shows the group of signals that were recorded, the experimental condition (mask or no mask) for which the signals were analyzed, and the instrumentation used.

The sound pressure level (SPL) was determined in a continuous fashion from a microphone placed at a distance of 30 cm from the subject's mouth. This signal was converted to dB SPL (re: 0.00002 dynes/cm^2) using predetermined calibration levels from the B & K 2230 sound level meter. The EGG electrodes were held in place by a velcro collar.

The wide-band airflow signal was detected with a circumferentially-vented wire screen pneumotachograph mask and transducer, (Rothenberg, 1973). The AC peak-to-peak flow was computed directly from the oral flow signal. The air pressure transducer (Glottal Enterprises system MS1F-2) was calibrated with a regulated air flow source and a Gilmont 52274 shielded flowmeter prior to collecting data from each subject.

A glottal flow derivative was obtained by LPC inverse filtering, and the maximum flow declination rate (MFDR) was defined from the negative peaks of this waveform. Exhalation volume during each task was computed by integrating the flow signal; lung volume change was taken as the inverse of exhalation volume and normalized to each singer's maximum. The integration procedure was verified by displacing three liters of air from a Collins volume calibrator through the mask system.

The intraoral air pressure (P_o) signal was sensed with a 5 cm long polyethylene tube placed through the airflow mask. One end of the tube was placed between the participant's lips at a position approximately half-way between the corner of the mouth and mid-lip; the distal end of the tube was connected to an air pressure transducer (Glottal Enterprises). Prior to data collection, P_o was calibrated in centimeters of water (cm H_20) with a Dwyer U-tube manometer. The pressure and flow signals were amplified with a Glottal Enterprises MS1F-2 amplifier system.

Table 2.
Recording Signals, Experimental Conditions, and Instrumentation

Signal	Condition for Analysis	Instrumentation
sound pressure level	no mask	B & K 2230 sound level meter
electroglottograph	no mask	Synchrovoice electroglottograph
oral flow	mask	Glottal Enterprises flow mask
oral pressure	mask with oral occlusions	Glottal Enterprises pressure transducer
ribcage displacement	no mask	NIMS respigraph system
abdomen displacement	no mask	NIMS respigraph system

During the mask with oral occlusion tasks, each subject sang repetitions of the syllable /pa/. This syllable was selected for estimating tracheal pressure (Smitheran & Hixon, 1981). Syllable rate and number of syllables varied to allow a natural and comfortable singing style. Subjects were required to achieve velopharyngeal closure (no nasalization), to produce the syllable repetitions on one breath, to maintain a stop closure, and to stabilize the airflow during the vowel portion of the syllable. Maintenance of velopharyngeal closure was monitored from the airflow waveform; airflow was required to be zero during the stop closure for each syllable.

Ribcage and abdomen displacements were monitored with a Non-Invasive Monitoring Systems, Inc. (NIMS) Respigraph system. All maneuvers were performed in a standing position a minimum of three times. These included isovolume maneuvers performed at end expiratory level (EEL), vital capacity (VC), maximum rib cage capacity (RCC), and maximum abdominal capacity (ABC) maneuvers (Hoit & Hixon, 1986, 1987). The absolute values of the measures were not used, however, because of a non-recoverable miscalibration in the data of two male subjects. Rather, ribcage, abdomen, and lung volume measures were normalized for each subject on a scale of 0 to 100%.

Subjects stabilized their quiet breathing before each utterance. End expiratory level (EEL) was identified as the mean minimum value for three cycles of tidal volume breathing before each utterance and used as a common reference point. Lung, rib cage and abdominal volumes were normalized relative to each subjects' maxima.

All signals were recorded on a Sony PC-108M 8-channel DAT recorder. The data were digitized from the analog output of the tape recorder using a DATAQ DI-200 12-bit A/D board. EGG and oral flow were low pass filtered at 3.5 kHz, and the remaining measurements were filtered at 300 Hz. These bandwidths were sufficient to capture the highest frequency of interest in the recorded signals. The data were then analyzed using Matlab on a Macintosh 7500.

Performance Protocol

Each subject produced 27 tokens of the *messa di voce*, 3 pitches x 3 mask conditions x 3 repetitions each. The pitches were 220 Hz, 440 Hz, and 740 Hz for subject KA; 277 Hz, 622 Hz, and 830 Hz for subjects LC and SK; 185, 311, and 349 Hz for WR; 220, 370, and 440 for GS; 220, 294, and 370 for IT. These pitch ranges of about 1.5 octaves were judged to contain notes in the low, medium, and high parts

of each of the female's vocal ranges and the medium to high parts of the male's ranges.

Each token of the *messa di voce* was targeted to last about 10 seconds, 5 seconds for the crescendo and 5 seconds for the diminuendo. Not every subject followed this target exactly; hence, the data were time-normalized for ease of comparison. No metronome or other pacing device was used, primarily because it would interfere with the audio signals and with a natural way of producing the utterances.

Results and Discussion

It was expected that there might be a large variability among the subjects due to age, gender, and experience, so we decided to display the results of each individual subject rather than as an ensemble average. Figures 2 to 7 show the performance in the subject order KA, LC, SK, WR, GS, and IT. There are three repetitions (tokens) of the *messa di voce* in each of three pitch columns, progressing from low pitch on the left to medium pitch in the middle to high pitch on the right of each figure. From top to bottom, we show the following signals: SPL at 30 cm distance, the electroglottograph (EGG) envelope, peak-to-peak flow (FLOW), maximum flow declination rate (MFDR), subglottal pressure (PSUB), ribcage displacement (RIB), abdomen displacement (AB), and lung volume (LUNG VOL). These signals were normalized with respect to time for comparison purposes.

Several of the signals amplitudes were also normalized to the maximum value for each pitch. This was partly to highlight the waveform symmetry and partly because absolute values were not as reliable as expected. For this study, only the relative increases and decreases were of interest.

Sound Pressure Level

Across all subjects, tokens, and pitches, SPL at 30 cm varied from a low of 50 dB (LC on her low pitch in the diminuendo, Figure 3) to a maximum of 103 dB (KA and LC on their high pitches in the loud portion, Figures 2 and 3). This 53 dB dynamic range was never achieved by a single subject on a single note, however. The largest single note SPL range was obtained by GS (Figure 6) who decreased from 100 dB to 55 dB on his middle pitch (a 45 dB change).

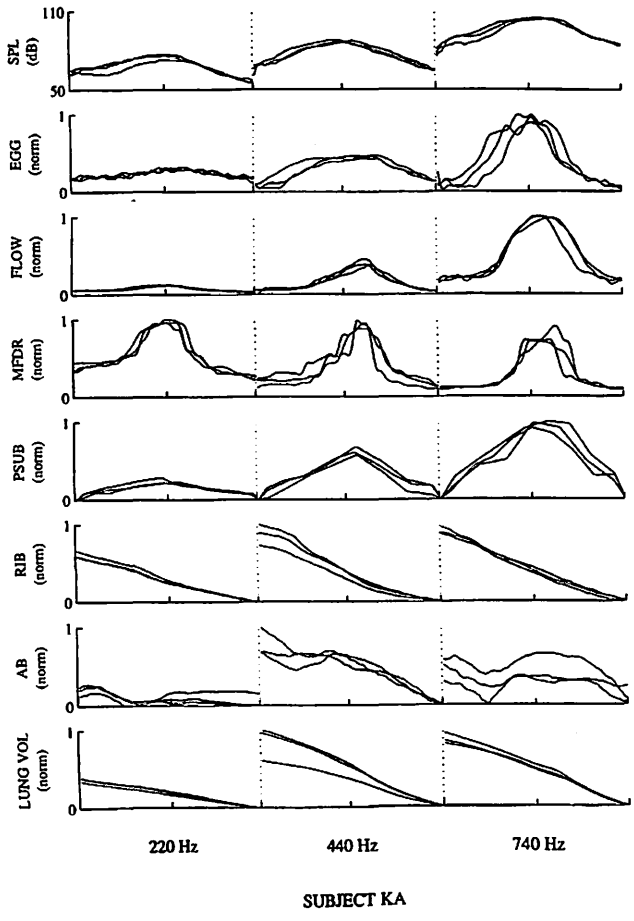


Figure 2. Acoustic and aerodynamic traces for three tokens of the messa di voce *exercise performed by subject KA on three pitches.*

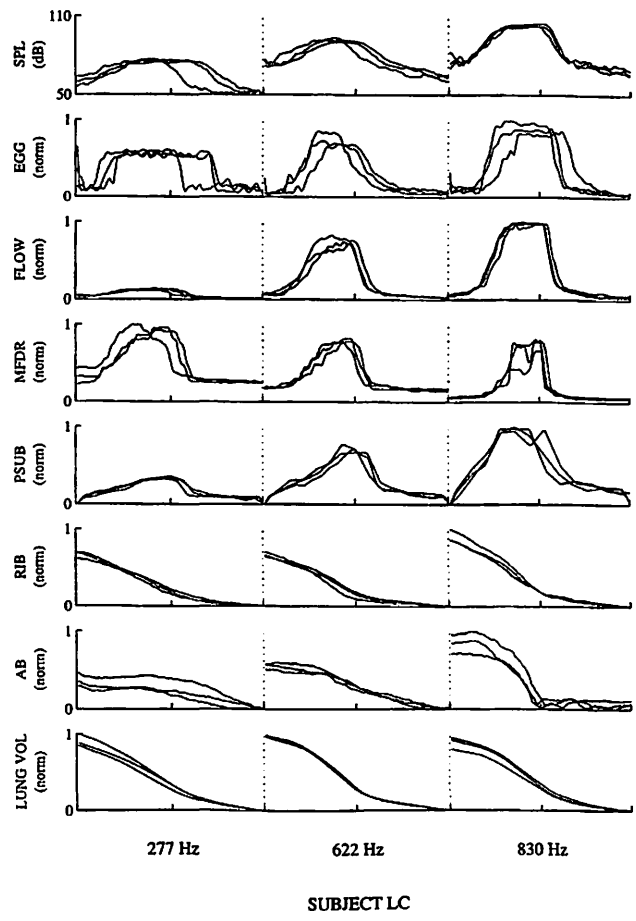


Figure 3. Acoustic and aerodynamic traces for three tokens of the messa di voce *exercise performed by subject LC on three pitches.*

Subjects KA and SK had the smallest SPL ranges, but showed the most consistent and symmetric variations over the entire set of exercises (Figures 2 and 4). Subjects WR and IT (Figures 5 and 7) had large SPL ranges on their low notes, but rather limited ranges on their middle and high notes. Perhaps most interestingly, all *messa di voce* attempts with large SPL ranges on any given note showed some asymmetry, usually with a linear or s-shaped SPL rise and an exponential-like SPL fall. This is seen on the high note of LC (Figure 3) and more clearly on the middle note of GS (Figure 6). Thus, symmetry did not appear to be of primary importance when large SPL ranges were attempted. The expected (model) shape was triangular, i.e., an increasing ramp followed by a mirrored decreasing ramp. This shape was approximated in the smaller dynamic variations of KA and SK (Figures 2 and 4), but the peak were slightly rounded.

Electroglottogram

The EGG signal was recorded primarily to assess stability of phonation. If a sudden register change were to occur, or if the larynx were to suddenly move up or down during the exercise, the EGG envelope should detect those changes. In general, we observed more fluctuations among the males than the females. Specifically, WR (the baritone) had considerable token variability on the low note (Figure 5), but much less on the middle and high note. Subject GS (the operatic tenor) had the most variable and asymmetric EGG waveform on the middle note (Figure 6), where his dynamic range was the largest (45 dB). Subject IT (the amateur tenor) had one very problematic token with large fluctuations after voice onset and before voice offset (Figure 7). These fluctuations were identified perceptually as chest-falsetto register breaks. The EGG waveform envelopes for sopranos KA and SK are basically smooth, but for LC (the coloratura soprano), the signal shows a pronounced flat top (saturation) in the loud portion of all the tokens at all pitches (Figure 3). This is also seen in other waveforms, suggesting that this subject was holding her dynamics constant near the peak for a brief moment, a deliberate part of her technique.

Peak to Peak Flow and Maximum Flow Declination Rate

The peak to peak FLOW and maximum flow declination rates (MFDRs) were most symmetric and consistent across tokens for singers KA and SK on the low and

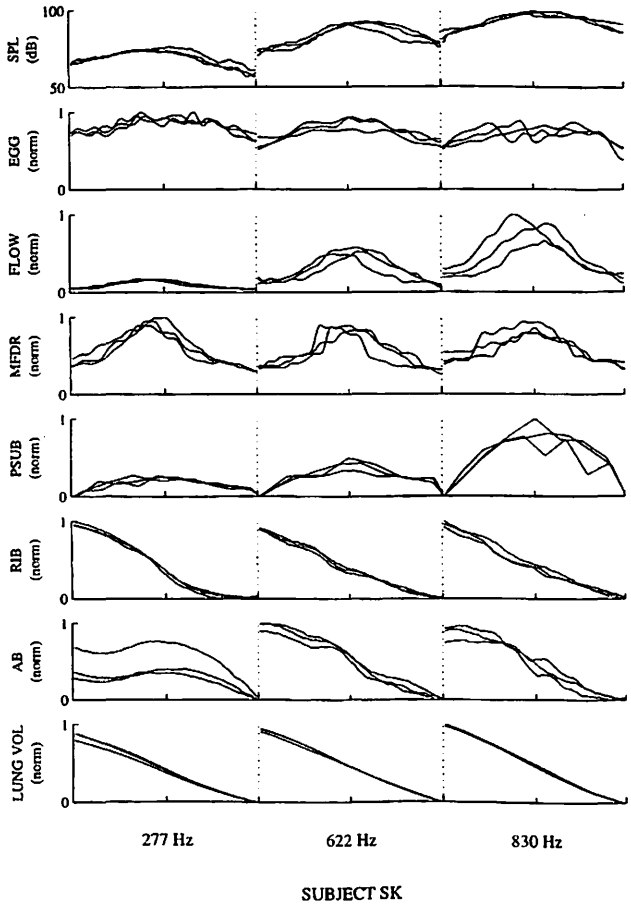


Figure 4. Acoustic and aerodynamic traces for three tokens of the messa di voce *exercise performed by subject SK on three pitches.*

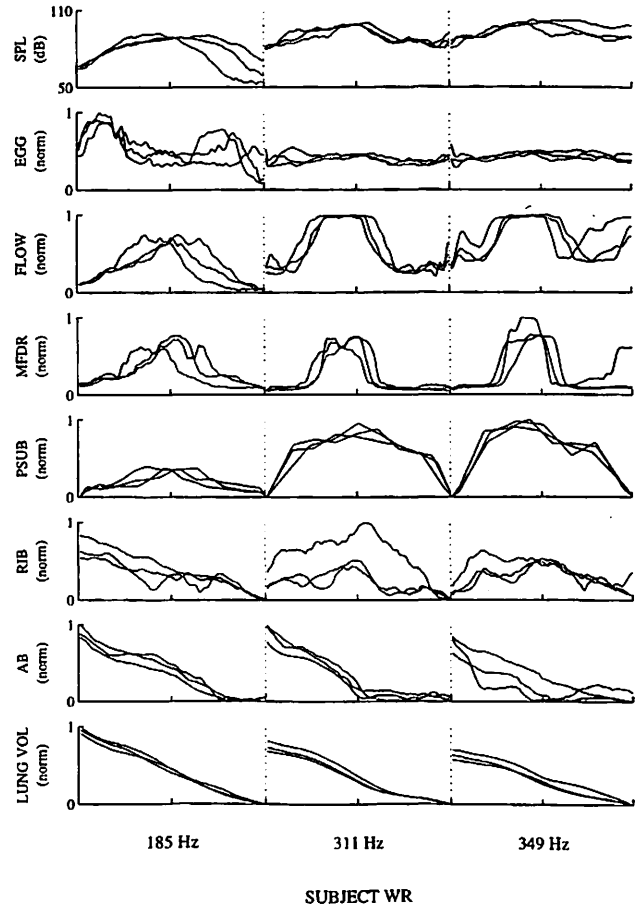


Figure 5. Acoustic and aerodynamic traces for three tokens of the messa di voce *exercise performed by subject WK on three pitches.*

middle notes (Figures 2 and 4). On the high note, KA showed a more sudden rise and fall in FLOW and MFDR, while SK skewed the FLOW deferentially for different tokens. Several singers complained about the difficulty of singing with the mask, especially at high notes when more jaw opening is needed than the mask can accommodate. Since the SPL trace, which was obtained without the mask, does not reflect this variable skewing, we suspect that his token variability is a mask artifact. Singer LC executed the exercise more square-like than ramp-like (Figure 3), with the major flow changes occurring over less than half of the length of the exercise. As mentioned above, LC tended to reach a plateau near maximum loudness.

The male singers were all more variable in FLOW and MFDR than the females. WR showed the flat-tops on FLOW that were characteristics of soprano LC (compare Figures 3 and 5 on the middle and high notes). The MFDRs of baritone WR were too variable to show clear trends. Tenor GS (Figure 6) showed the greatest token variability on the middle note for FLOW and MFDR (where he achieved the largest dynamic range), but on the low and high note the productions were more consistent. Tenor IT (Figure 7) showed large token variability on the low and middle note, but basically had a ramp-like pattern on the high note.

Subglottal Pressure

Three problems should be noted at the outset with respect to the subglottal pressure curves. First, they are poorly sampled over the exercise because several consonant-vowel syllables were needed to establish one measurement of pressure. Second, they are not from the same tokens as the remaining waveforms. Thus, only a loose correlation can be expected between the traces. Third, there may have been some lip-pressure tube leakage when the singers focused more on their productions rather than on the experimental details.

Nevertheless, the expected ramp-like subglottal pressure (PSUB) patterns were produced by soprano KA (Figure 2), especially on the middle note. Soprano LC ramped the pressure on the crescendo in a similar way, but relaxed it faster on the decrescendo (Figure 3). Soprano SK and baritone WR (Figures 4-5) showed a symmetrical rise and fall, but without a sharp peak. Tenors GS and IT showed a triangular (ramp-like) pattern as expected, but IT was more

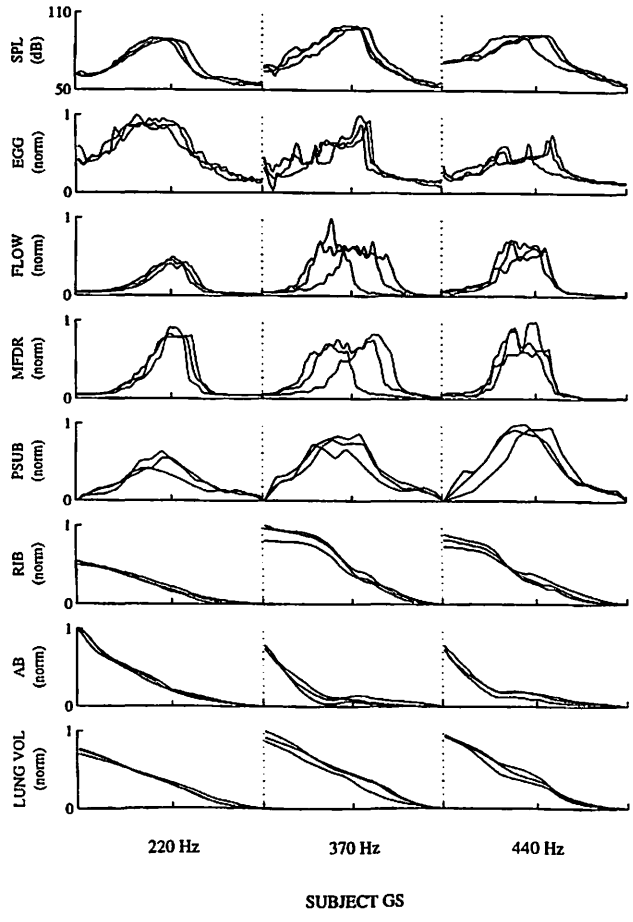


Figure 6. Acoustic and aerodynamic traces for three tokens of the messa di voce *exercise performed by subject GS on three pitches.*

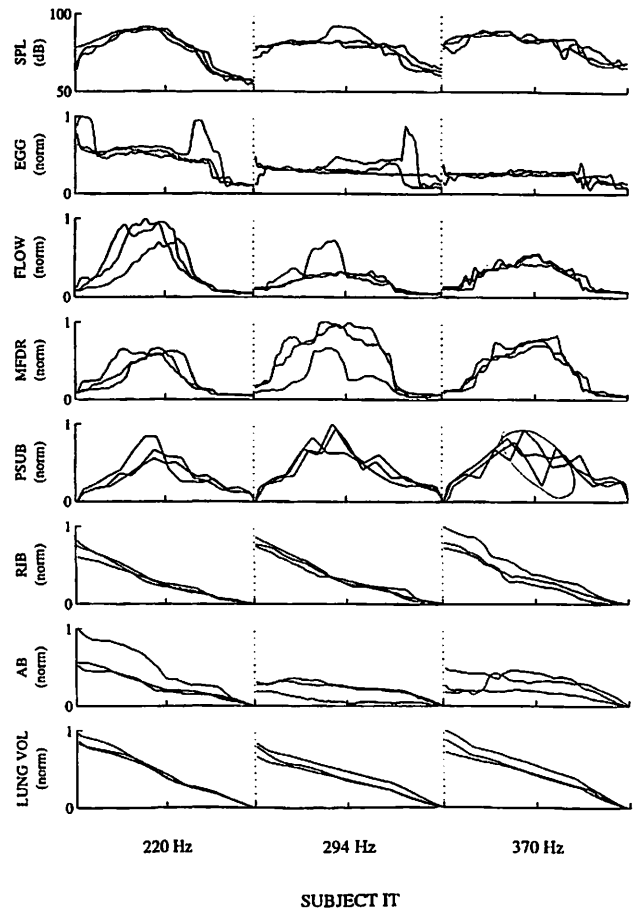


Figure 7. Acoustic and aerodynamic traces for three tokens of the messa di voce *exercise performed by subject IT on three pitches.*

variable than GS. Across all subjects, PSUB was one of the most symmetric waveshapes, with the exception of the asymmetry shown by LC.

Ribcage, Abdomen, and Lung Volume Displacements

Ribcage displacement (RIB) was a monotomically declining function for all subjects except WR. It generally mirrored the lung volume rather closely, as reported by Thomasson and Sundberg (1997) in a similar study that used complete songs as performance material. For subject WR in our study, RIB showed paradoxial movement in the crescendo. The ribs expanded first (in the inspiratory direction) and then decreased (Figure 5, middle and high notes). This initial expansion of the ribcage was accompanied by a faster decline (inward movement) of the abdomen to maintain a net lung volume decrease. Tenor GS also had a tendency to delay ribcage movement, but did not actually expand the ribs. He did, however, accelerate abdominal movement the same way that WR did. These paradoxical movements (or tendencies toward paradoxical movements) occurred only in the crescendo phase.

In general, abdomen displacement was a little more variable for most subjects than ribcage displacement. This was also found by Thomasson and Sundberg (1997). Interestingly, the largest fluctuations in abdomen displacement occurred for singer KA (Figure 2), who was otherwise stereotypic across tokens on all other waveshapes. The abdomen moved in a paradoxical direction on the high note over part of the crescendo. Soprano LC, with her abbreviated decrescendo, showed a rapid abdominal movement followed by a long period of zero displacement (Figure 3, high note). Ribcage and lung volume changes for LC had similar tendencies, but were more steadily declining. Soprano SK had a reversal of abdomen movement on her low note (Figure 4). At first AB decreased a little, then it increased, and finally it decreased again. But lung volume remained linear in this process; so did ribcage displacement. Tenor IT showed generally linear trends in all respiratory measures, but token variability was substantial.

Conclusions

We return now to the questions posed at the beginning of this paper. First, are the SPL increase and decrease symmetric in a *messa di voce* exercise? Our results have shown that SPL symmetry is not realized in a strict sense. A typical asymmetry is a delayed rise in SPL followed by an accelerated fall. This was demonstrated in two of the six singers, particularly on notes that had the largest SPL range.

The asymmetries in SPL cannot be attributed to a non-uniform expulsion of air, because the lung volume changes were generally quite linear throughout the exercise. However, substantial nonuniform variations in lung pressure and flow were seen in conjunction with variations in abdomen and ribcage displacement, suggesting that some differences in respiratory strategy seem to carry over to phonatory control. In particular, one baritone subject showed outward movement of the ribcage during the crescendo phase of the exercise while maintaining an increase in peak airflow and vocal intensity.

The asymmetries observed did not seem to be associated with register breaks (or anticipated register breaks) in any of the females. For the male singers, however, there seemed to be more laryngeal movement, judged by the EGG envelope. In particular, some register breaks occurred in both the crescendo and decrescendo portion of the exercise for the amateur tenor.

The asymmetry in the *messa di voce* was pitch dependent. Generally speaking, the higher pitches had more variability across tokens. The higher pitches also showed more cases of abbreviated decrescendos, suggesting that singers were anxious to get to the end of the exercise. Sometimes the flow and lung pressure peaks were advanced or retarded in relation to the SPL peak. This was rarely the case at a low pitch.

In summary, examination of the symmetry and controlling parameters among the varied classical voice types revealed sometimes similar, though often dissimilar, results of the subject's balance of the acoustic and aerodynamic factors. The singers did not share the same pedagogical approach to "support", which may have influenced the varied patterns of ribcage and abdominal movement. Those subjects with more performing experience tended to have unique patterns among the measured tested, while those subjects still in formal training in the academic environment showed the more expected (stereotypic) behavior with less dynamic variation. This indicates that optimization of the respiratory and vocal systems is a subject-specific process.

Acknowledgments

This work was supported for Grant No: P60-DC00976 from the National Institutes of Health. The authors thank Ronald C. Scherer and Christopher Dromey for assistance in data collection and analysis.

References

Concone, J. (1894). *Thirty Daily Exercises*. Schirmer Publications, New York.

Hoit, J., & Hixon, T. (1986). Body type and speech breathing. *J. Speech Hear. Res., 29,* 313-324.

Hoit, J. & Hixon, T. (1987). Age and speech breathing. *J. Speech Hear. Res.*, 30, 315-366.

Miller, R. (1986). *The Structure of Singing*. Schirmer Books, New York, p. 173.

Rossini, G. (1825). *Gorgheggi e Solfeggi*. Ricordi Publishing, Milan. Reprinted by Frangipani Press, Bloomington, Indiana (1986).

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. *J. Acoust. Soc. Am, 53*, 454-460.

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

Stathopoulos, E., & Sapienza, C. (1993). Respiratory and laryngeal measures of children during vocal intensity variation. *J. Acoust. Soc. Am., 94*, 5, 2531-2543.

Sundberg, J., Scherer, R., & Titze, I. (1993). Phonatory control in male singing. A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. *J. Voice, 7*(1), 15-29.

Thomasson, M. & Sundberg, J. (1997). Consistency of phonatory breathing patterns in professional operatic singers. *KTH Speech, Music and Hearing Quarterly Progress and Status Report*, 4, 29-40.

Watson, P., Hixon, T., Stathopoulos, E., & Sullivan D. (1990). Respiratory kinematics in female classical singers. *J. Voice, 4*(2), 120-128.

Watson, P., Hoit, J., Lansing, R., & Hixon, T. (1989). Abdominal muscle activity during classical singing. *J. Voice, 3*(1), 24-31.

NCVS Status and Progress Report - 12
June 1998, 153-161

Experimental Approaches to Vocal Fold Alteration: Introduction to the Minithyrotomy

Steven Gray, M.D.
Primary Care Medical Center, The University of Utah
Steven Bielamowicz, M.D.
Voice Treatment Center, George Washington University
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Heather Dove, M.A.
Speech Pathology/Rehabilitation Services, The University of Utah
Christy Ludlow, Ph.D.
National Institute on Deafness and Other Communication Disorders, The National Institutes of Health

Abstract

Treatment of challenging laryngeal disorders, such as lamina propria loss or neuromuscular dysfunction may require novel approaches and techniques. This paper discusses an evolution of experimental techniques for treatment of lamina propria loss and the use of the minithyrotomy. These techniques have been used for surgical access for lamina propria substitution as well as placement of stimulating electrodes. The minithyrotomy is tolerated well by patients, it provides access for microscopic instruments with the surgeon's hand close to the tissue of interest, avoids intralaryngeal mucosal incisions, and lines up the direction of dissection in an anterior to posterior orientation. This orientation is favorable for particular situations discussed. We will further present anatomic and physiologic concepts relevant to the surgical treatment of lamina propria dysfunction, as well as present our clinical experience. This paper is not to state how these difficult problems should be handled, but rather to present our experience in techniques that may prove useful through further development.

Surgical access to the vocal fold structures is challenging. Traditional approaches include a transoral approach, a lateral thyroplasty approach or a midline laryngofissure approach. We propose a minithyrotomy (MT) as a new approach for surgical manipulation of the vocal fold structures. The MT approach has an advantage of eliminating intralaryngeal mucosal incisions that may lead to scarring of the vocal fold epithelium to the underlying body of the vocal fold. This allows for surgical manipulation of the superficial layer of the lamina propria in a controlled fashion. Instruments can be introduced through the MT in an anterior to posterior direction and visual control is provided by flexible or rigid videoendoscopy. We have developed an MT technique for lamina propria replacement surgery. In addition, the MT provides an excellent route for placement of stimulating electrodes for vocal fold closure. The endoscopic view is essential for adequate placement of the stimulating portion of the electrode near the neuromuscular junctions of the thyroarytenoid (TA) and lateral cricoarytenoid (LCA) muscles, Both of these techniques are considered experimental, and we would like the reader to recognize that these procedures are evolving and refinements are expected. In addition, the MT can provide access for the placement of an arytenoid adduction suture. A conduit can be placed from anterior to posterior through the TA muscle. The posterior opening of the conduit can be directed to a location immediately lateral to the muscular process of the arytenoid cartilage.

In this paper, we will describe the MT approach and discuss its utility in these three procedures. We will describe a new technique of lamina propria replacement surgery, neuroprosthetic adduction implantation, and a

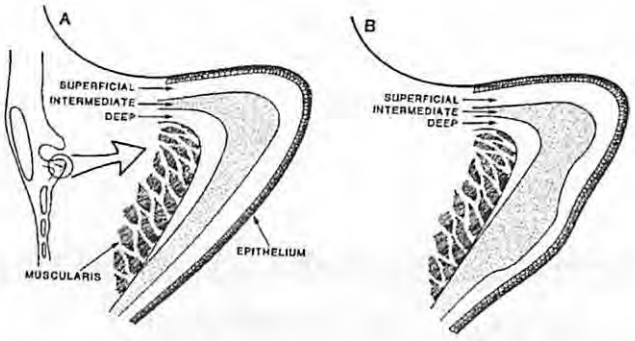
modification of the arytenoid adduction procedure. In the area of lamina propria replacement surgery, this paper will describe the infrafold anatomy, discuss aspects of infrafold configuration relevant to vocal fold oscillation, and relate those topics to possible surgical approaches. For the implantation of a laryngeal electrical stimulator, laryngeal anatomy and stimulator positions will be discussed.

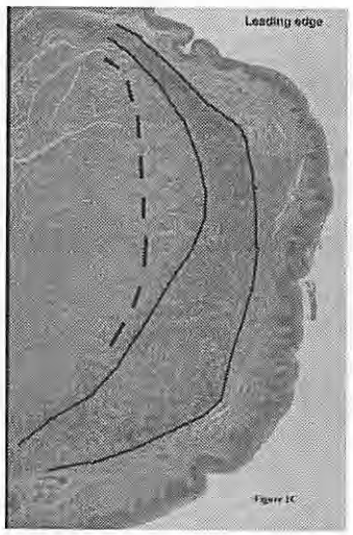
Lamina Propria Replacement Surgery

Scarring of the vocal folds is a difficult and challenging clinical disorder. Various treatment options have been proposed for treatment including speech therapy, collagen injection, medialization laryngoplasty, scar release, and fat implants[1,2,3,4,5]. Although not reported in the medialization laryngoplasty, scar release, and fat implants[1,2,3,4,5]. Although not reported in the literature, other options have been attempted by various phonosurgeons, including mucosal grafts and steroid injections. All of these procedures may have a role in treating scarred vocal folds depending on the severity and extent of the scarring. For purposes of this paper, only treatment of lamina propria loss or dysfunction by replacement or reconstruction techniques will be discussed.

It has long been recognized that the start of the mucosal wave occurs in the infrafold region. Mucosa "bunching" or "upheaval" begins approximately three millimeters below the leading edge of the vocal folds, at the level of the inferior portion of the thyroarytenoid muscle[6,7]. The exact beginning location of the mucosal wave varies since it is partially dependent on amplitude of vibration, frequency of vibration and glottic configuration. Regardless of the precise location, normal vocal production depends upon the free movement of the infraglottic vocal fold mucosa. Any impairment of the lamina propria in this region will result in vocal dysfunction. Ideally, this tissue movement should occur easily, with little energy required, allowing for mucosal movement in the vertical plane. When the vocal fold and especially the infrafold tissue have been injured or destroyed, the location of mucosal wave initiation is altered[5]. It is likely that some of our attempts to improve vocal fold scar have fallen short due to failure to adequately address and correct the infrafold region.

When considering aspects of lamina propria reconstruction, it is helpful to examine the normal infrafold anatomy. The vocal fold anatomy has been described in detail in many elegant articles[8,9]. Typically the lamina propria layers are drawn schematically in which the superficial and intermediate layers are shown to be the thickest at the leading edge of the vocal folds, with a gradual tapering of the layers as they proceed inferiorly. Histological examination of human larynges demonstrates considerable variation in the tapering of the lamina propria layers in the infrafold region. Specifically, the intermediate layer of the lamina





Figures 1A, 1B, 1C. Schematic drawing of the superficial, intermediate and deep layers of the lamina propria. Note that the traditional drawing of the lamina propria is seen in Figure 1A (upper left). However, Figure 1B (upper right) is also frequently seen. Note that the intermediate layer creates a bulge in the vocal ligament in the immediate infrafold area. The histologic correlate of 1B is seen in 1C (left) from a middle-aged man. The vocal ligament in 1C is outlined in black, whereas the dotted line represents the junction of the deep layer of the lamina propria and the vocalis muscle.

propria may remain quite thick and wide in the infrafold area and on occasion, may bulge in the infrafold area due to thickness of the extracellular matrix components. This is best illustrated by looking at Figures 1A and 1B.

Notice that the intermediate layer (outlined) may be almost as thick in the infrafold region as the leading edge. The intermediate layer is composed of elastin fibers, fibers, hyaluronic acid and fibromodulin[10,11]. These molecules give the tissue qualities of elasticity and viscosity appropriate for tissue oscillation. The effect of this infrafold bulge is under investigation, but it is conceivable that this bulge biomechanically facilitates mucosal wave upheaval, vibrational amplitude and vertical travel. Not all larynges have this infrafold thickness of the intermediate layer, and there are infrafold differences from anterior to posterior along the membranous fold. Anteriorly, the tapered layers shown in Figure 1A are seen, more posteriorly in the membranous fold, infrafold intermediate layer thickening occurs.

The shape of the infraglottis is also important. If one only augments the free edge of the vocal fold and neglects the infrafold portion then the parameters needed for tissue oscillation may actually become worse. (See Figures 2A and 2B)

The oscillation threshold pressure can be determined by using the following equation for a glottal shape, as shown schematically, in Figures 2A and 2B. In this equation, X_1 is the glottal width

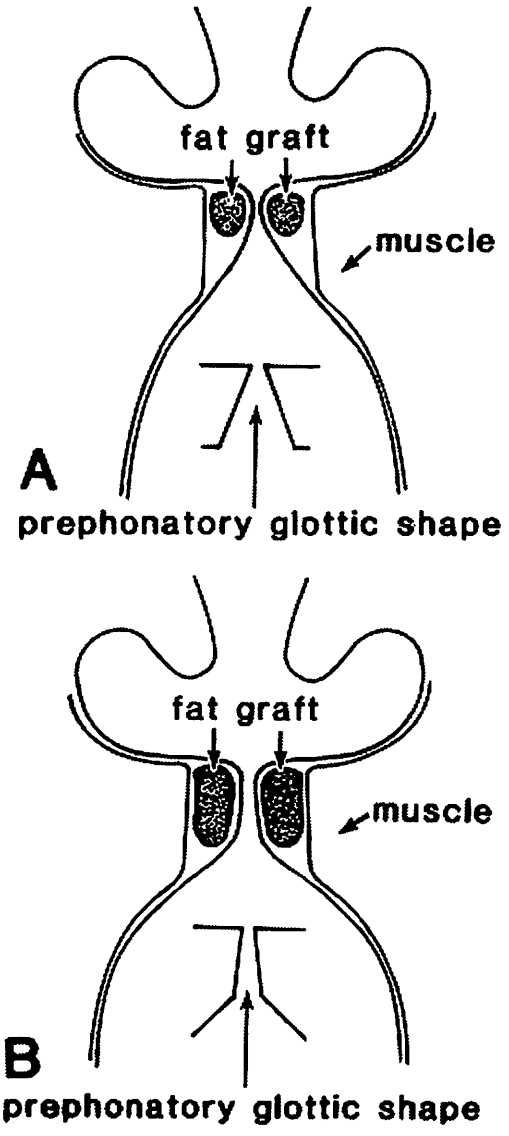
$$P_{th} = \left(\frac{2}{T}\right)(Bc)\frac{X_1^2}{X_1 + X_2}$$

at the bottom and X_2 is the glottal width at the top, T is the vocal fold thickness, B is a viscous damping coefficient in the tissue, and C is the mucosal wave velocity. Note that oscillation threshold pressure increases as X_1 increases, indicating that the "ease of phonation" is reduced by a large separation of the vocal folds at the bottom (a large convergence). The best case is a rectangular glottis, shown in Figure 2B, where $X_1 = X_2$, for which case the threshold pressure becomes

$$P_{th} = \left(\frac{2}{T}\right)(Bc)X_1$$

Note that P_{th} is lowered by decreasing the glottal width (better adduction), increasing the vocal fold thickness, lowering the viscous damping (B) and lowering the mucosal wave velocity (c).

The relevance to the repair or replacement of injured lamina propria is that restoration of the infrafold region is critical and should be addressed in surgical reconstruction of vocal fold function if we are to be successful. The infrafold region may often be neglected in our evaluation of the scarred fold because of the difficulty examining this region with traditional clinical techniques. With stroboscopy, the infrafold region is usually obscured by the leading. more medial edge of the vocal fold. Often the only indication that severe dysfunction is present is that no mucosal wave can be seen traveling up from the infrafold region[12,5]. Another option to verify the extent of scarred vocal folds is to perform the microlaryngeal infusion technique described by Zeitels[13]. Normal vocal folds infuse easily and symmetrically, whereas scarred vocal folds do not. They will show asymmetries and areas of restriction upon infusion.



Figures 2A, 2B. This figure illustrates the effect the position of any implant to the lamina propria has on glottic configuration. Figure 2A demonstrates an injected, implanted substance only at the leading edge of the vocal fold and the consequent convergent phonatory shape of the glottis (not a good shape for oscillation of tissue). Figure 2B shows the effect of the substance extending more infrafold to provide a better phonatory glottic shape conducive for tissue oscillation. The size of the implants is exaggerated in these figures for illustrative purposes.

Tissue Properties

When the lamina propria has been injured, lost, or scarred, freedom of tissue motion is impaired. If the tissue has reduced and restricted motion then it becomes stiff and unable to oscillate.

One of the properties affecting freedom of motion is viscoelasticity. As has been described, the extracellular matrix is a composition of proteins consisting of elastins, collagens, proteoglycans, glycoproteins, carbohydrates and lipids. Particularly, some of the proteoglycans affect the viscosity of the tissue[14,15]. Viscosity determines the amount of energy required for tissue oscillation. As tissue viscosity increases, the amount of energy required to overcome frictional forces for tissue oscillation increases[16]. This is manifested in the parameter B in the equation above. With injured or scarred lamina propria, the normal beneficial proteins such as hyaluronic acid and elastin material are replaced by protein components such as collagen fibers and fibronectin, which results in increased tissue viscosity.

To correct lamina propria loss or scarring, one should consider procedures which restore freedom of motion and lower tissue viscosity. While future studies may identify a more ideal injectable material, such as hyaluronic acid, currently we are investigating the use of a fat graft to

improve both freedom of motion and tissue viscosity. The construction of a cellular layer such as fat may help in maintaining freedom of tissue motion. By creating this cellular layer between the cover (epithelium and superficial layer) and body (deep layer and thyroarytenoid muscle), the fat cells may prevent fibroblast migration and resultant scarring between the two layers. The concept of surgically freeing up the scarred area sharply is not new[17]. Unfortunately, without some mechanism to prevent fibroblast migration and new scar in the surgically injured site, one would expect a short-term surgical result. By placing a layer of fat between the body and cover of the vocal fold, perhaps it would maintain a separation of the two layers.

Another reason to choose fat is that Chan and Titze showed fat has similar viscous properties to normal human lamina propria[18]. Normal human lamina propria has a dynamic viscosity of about 3-4 Pa-s at 10 Hz while fat is about 4 Pa-s. Of interest is that a proteoglycan mixture such as hyaluronic acid is around 1-2 Pa-s whereas collagen mixtures are about 10 Pa-s. Low values are less viscous (like water) while high values are more viscous (like glue). Viscosity has a linear relationship with the effort required to oscillate human vocal folds[16]. Tissue with higher viscosity will require more effort. For these reasons, fat was chosen as a possible substance for lamina propria replacement.

Surgical Approaches

We have used a transoral microlaryngoscopic approach to the scarred fold, with scar release and/or steroid injection, or insertion of a fat graft stabilized with 7-0 or 6-0 microsutures. Although our early results were often favorable with these techniques, within six months the patients began to lose the benefits of the surgery. (See clinical examples of Kenolog injections in Table 1)

Notice from table 1 that the results are quite variable. Although jitter improves in all, some of the aerodynamic measures became worse. Most importantly, the perceptual rating did not change in 2 of the 3 patients.

It is easier to understand the relapse associated with simple scar release followed by steroid injection since scar formation is probably reoccurring. It is more difficult to understand the relapse associated with transoral fat implantation. It appears that when fat is implanted transorally, it tends to migrate superiorly in the cordotomy incision. This results in the fat being in the suprafold area and also creates a convergent glottic shape, which as discussed earlier, may be detrimental to P_{th}.

We then employed a laryngofissure approach described in 1987 by Wexler's, et al. study of fat grafts to the scarred vocal folds of canine animals[19]. In this study, a laryngofissure approach was chosen because creating a tunnel pocket for fat implantation required microscopic dissection with ear instruments. Lifting the epithelium off the thyroarytenoid muscle in a continuous pocket without any epithelial tears required microscopic control of the instruments and careful tissue handling. Additionally, the ideal pocket location was primarily in the infrafold region without a significant suprafold component. A suprafold component allows implanted fat to migrate superiorly in the pocket since airflow and phonation tends to propel tissue upward. The final location of the implanted fat is critical since this affects prephonatory glottic configuration. When this technique is performed, the implanted fat is clearly visible through the epithelium and casts an evident yellow

Table 1.
Scarring: Some Results of Kenolog Injections to Vocal Fold Scar

Measures	#1 Pre-OP	#1 Post -OP	#2 Pre-OP	#2 Post-OP	#3 Pre-OP	#3 Post-OP
Mean Jitter	2.9%	0.2%	0.8%	0.4%	1.2%	0.7%
Mean SNR	13 dB	27 dB	20 dB	29 dB	22 dB	22 dB
Mean DC Flow	729 ml/second	672 ml/second	274 ml/second	202 ml/second	373 ml/second	254 ml/second
Mean MPT	6 seconds	5 seconds	11 seconds	13 seconds	7 seconds	4 seconds
Mean AC:DC	0.85	0.51	0.79	1.17	0.4	0.76
Mean Habitual Subglottic Pressure	13.7 cm H_20	9.6 cm H_20	8.4 cm H_20	5.4 cm H_20	13.3 cm H_20	9.8 cm H_20
Perceptual Rating	5	1-2	2-3	2	3	3-4

Surgeries performed in 1990-1994. Recordings performed at least 6 months post-operative. Perceptual ratings were performed by the speech pathologists (not blinded) on a 7 point scale (1=normal, 7=aphonic).

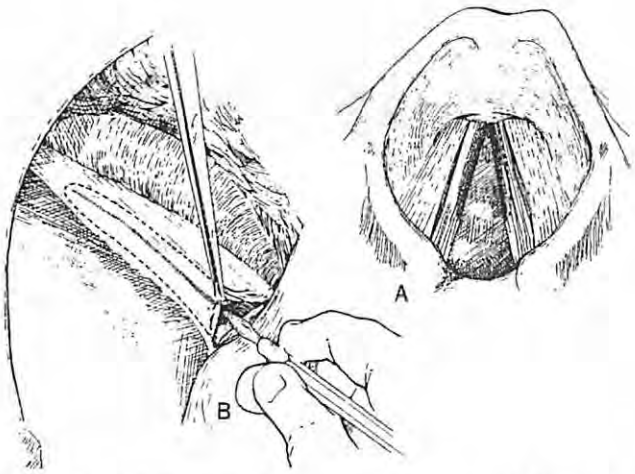


Figure 3A, 3B. Sketch of the laryngofissure approach. A) The stippled area represents the scarred lamina propria. B) Illustrates the laryngofissure approach. The ear instruments are used to free up the scarred cover from the underlying muscle. The pocket is packed with fat in a symmetric fashion. The pocket is created so that the infrafold area is included in the pocket, whereas the suprafold area is not dissected.

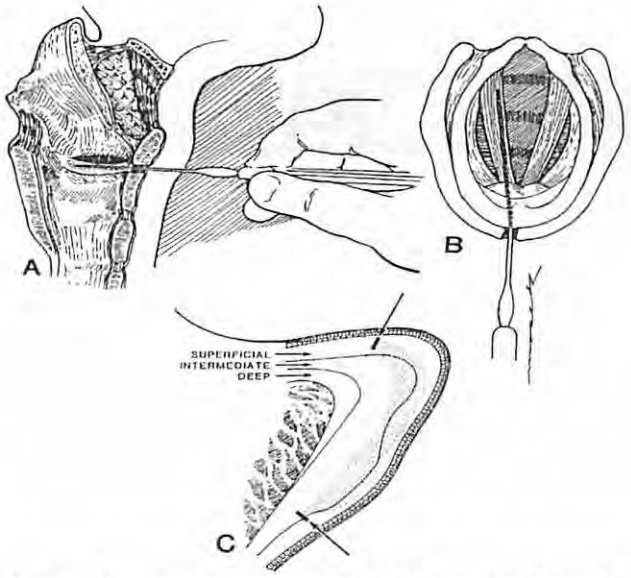


Figure 4. These three illustrations show the minithyrotomy discussed. A small burr-hole or window is created through the thyroid cartilage. Instruments are then placed through this thyroid cartilage to the infrafold edge of the vocal fold. C) The stippled area shows the location of the pocket created. Black bars are placed to indicate the limits of the pocket based upon Wexler's study and upon our patient experience.

hue to the vocal fold. If a yellow hue is not seen, the fat may have been placed too deep. If the fat is placed too deep then the viscous properties which fat may convey for vocal fold oscillation is not located superficially enough to assist tissue movement, and restricted tissue movement persists. (See Figure 3)

The laryngofissure approach appeared to satisfy the goals of the fat implantation. A pocket 3-4 mm in vertical height along the entire membranous fold and infrafold area could be packed relatively tightly with fat through an incision that could be closed to prevent fat extrusion[12]. The results from the Wexler et al. and the Jiang et al. studies were very acceptable[19,20].

We pursued fat implantation in humans based on these results, optimistic of obtaining similar functional improvement in patients with scarred vocal folds.

Methodology and Results

Laryngofissure/Lamina Propria Surgery

Using a laryngofissure approach (described above) we performed fat implantation to scarred lamina propria in three cases. We wanted to make sure that a relatively long follow-up had occurred as fat may absorb and scarring changes may take place up to two years later.

In retrospect, in our first case we should have made the pocket slightly larger and grafted more due fat due to anticipated fat resorption. The fat was resorbed more in these human patients than in the canines. However, the resorption was less than we experienced with fat injections. With each procedure, we learned to place more fat and to create a more generous pocket without anterior commissure to the vocal process. Then the needle is used again to create a second pass and then a third. Multiple passes are made until the needle can be easily passed from the anterior commissure to vocal process in the infrafold and leading edge area. This technique can be easily monitored and visualized with a microscope or a 70-degree telescope. Current suspension devices make this somewhat awkward, but we place a sticky sterile plastic wrap around our suspension device to maintain sterility.

After the small tunnels have been created, they are connected into a single pocket with ear knives (like a Tabb) or scissors, and the small fat pieces (obtained from the neck) are placed. A sharp dissection without a plane of dissection may result in accidentally tearing or cutting through the epithelium and allow fat extrusion and infection contamination. Both are possible sources of fat resorption. The pocket should include the area of scar, the free edge, and 2-3 mm inferior to the extending the superior aspect of the fold. No further surgery has been performed on the second or third patients. The first patient moved to Wisconsin after 18 months post-op and had some voice deterioration. She has had two additional fat injections into the created fat layer by Dr. Brandenburg (University of Wisconsin). Both injections have further improved her voice. The results of those injections are not presented. (See Table 2 and Table 3)

Minithyrotomy (MT)/Lamina Propria Surgery

Although the patients have all been happy with their voice quality, we were not satisfied with the need for a laryngofissure to provide exposure. From the above results fat implantation to the infrafold and leading edge of the fold was a clinically useful procedure if it could be performed

MT Patient Data				
	12/18/97	1/7/97	12/11/96	Normal Range
Mean Jitter	.64%	1.16%	2.03%	.3% or less
Mean Shimmer	1.88%	7.82%	10.04%	2% or less
Mean SNR	28.37 dB	21.6 dB	22.35 dB	30.0 dB or greater
Max Phonation Time	10 seconds	11 seconds	10 seconds	15-20 seconds

without a laryngofissure. An MT as an entrance site for the instruments while using an endoscope or microscope in the larynx for visualization was considered. This approach is not unlike other minimally invasive endoscopic surgeries where visualization and instrument access are obtained from different sites. The minithyrotomy is created by cutting or drilling about a 4 mm window in the thyroid cartilage slightly off the midline of the anterior commissure area through a 1-2 cm anterior neck incision. The site of the MT is determined by placing a needle through the thyroid cartilage at the expected site that instruments will be placed directly horizontal to the vocal fold edge. Ideally, the needle should enter the vocal fold just deep to the epithelium at the free edge of the vocal fold or slightly infrafold. By slightly moving the needle, the position of the needle can be determined viewing the tissue movement videoendoscopically.

The direction of dissection must be horizontal to the fold so that a straight pocket or tunnel can be created. If the direction is not horizontal then the pocket is larger anteriorly and smaller posteriorly and the fat graft will bunch anteriorly. Once the MT is created, then a straight pick (used in ear surgery) is placed to first create a small tunnel just under the epithelium from the scar. When the fat graft is placed, the fat should be placed so as to maintain some consistency in the density of the fat across the pocket. The fat graft should be symmetrically distributed across the vocal fold. Lumpy fat grafts will not impart symmetry.

We performed an MT on an 84 year old man with severe bowing, sulcus vergeture, and bilateral teflon granulomas. He was treated with bilateral teflon injections twice a few years earlier for severe bowing which did not improve his voice. In our evaluation he had hard firm vocal folds with no mucosal waves and also was recognized as having bilateral large sulcus vergeture. We removed the teflon from the folds through a lateral thyroplasty approach (introduced to the technique by James Netterville- Vanderbuilt University) however, the sulcus vergeture and scarring of the lamina propria from the teflon persisted. An MT was performed in June of 1996. The procedure took a little less than two hours and the patient has been happy with the result. So far his acoustic perturbation values are better post operatively and have even improved with time. His aerodynamic data are only slightly improved, and have not deteriorated over the last year. Although his voice is not normal, perceptually he has been happy with the results of the surgery (see table below).

At this point of our clinical research, we express satisfaction with the MT approach, but not with the currently available instruments for this operation. We need finer instruments that allow dissection of the epithelium from the underlying tissue. The ear instruments that were relatively adequate for a laryngofissure approach did not work as well for the endoscopic technique. It was relatively easy to create the tunnels between the scarred epithelium and the underlying tissue and muscle, but it was difficult to connect the tunnels into a single, relatively symmetric pocket with the ear knifes and scissors. The scissors were too big to fit in the pocket and the knives were not completely adequate in taking down the fibrous adhesions separating the tunnels. Smaller scissors to cut the adhesions would be an improvement. Sharp, fine micro-knives that cut from side to side would also help.

Minithyrotomy-Placement of Stimulating Electrodes

Stimulation of the adductor muscles in the larynx is currently being investigated. Dynamic adduction of the vocal folds could possibly assist in the management of patients with dysphagia and aspiration due to central nervous system dysfunction and brainstem injury including stroke and amyotrophic lateral sclerosis.

We have currently implanted twelve canines with a prototype of a functional electrical stimulation (FES) unit for dynamic vocal fold closure. We have independently devised a MT procedure for the implantation of electrodes

Table 2.
Patient Information for Fat Implantation

	Date of Surgery	Clinical Data
37-Year-Old Female	1990	Vocal fold entire length stripping (by operative report) twice for nodular thickening 5 and 4 years prior to 1990. No mucosal wave was present.
66-Year-Old Female	1992	Benign neuro tumor resulting in unilateral paralysis treated with Teflon 10 years earlier. Recurrent, large Teflon granulomas treated with many $C0_2$ laser excisions over 6 years. Eventually had severe unilateral scarring of fold.
42-Year-Old Male	1994	Anterior half of vocal fold (partial cordotomy) due to benign tumor. Had aspiration, severe dysphonia, with scarring of the remaining fold remnant.

Table 3.
Results from Fat Implantation

Measures	#1 Pre-Op	#1 Post-Op	#2 Pre-OP	#2 Post-OP	#3 Pre-OP	#3 Post-OP
Mean Jitter	UM	0.41%	2.62%	0.4%	1.3%	0.51%
Mean SNR	UM	18.4%	12.25 dB	21.8 dB	17.7 dB	27.8 dB
Mean MPT	9 seconds	14 seconds	3 seconds	5 seconds	10 seconds	10 seconds
Mean DC Flow			409 ml/second	275 ml/second	340 ml/second	418 ml/second*
Mean AC:DC			0.34	0.62	0.87	1.36
Mean Habitual Subglottal Pressure			15.53 cm H_20	8.8 cm H_20	7.65 cm H_20	5.69 cm H_20
Perceptual Rating	5	2	5	4	3-4	1

UM indicates unmeasureable. The signal was too non-periodic for the software C-speech to track for measurements. Open cells indicate that the tests were not performed. In 1990 aerodynamic measures were not obtained preoperatively.
*Patient 3 showed continued improvement in acoustic measurements for 2 years. Aerodynamic measurements improved postoperatively then at about 13 months decreased to the values reported. Earlier 9-month values included 267 ml/second for flow and 13 second MPT.
Perceptual ratings were performed by the speech pathologists (not blinded) on a 7 point scale (1=normal, 7=aphonic).

into the larynx. Our goal has been functional neuromuscular stimulation of the TA and LCA muscles. The larynx has been implanted in two different routines. Some animal larynges have been implanted with a ventricular fold and a true vocal fold electrode bilaterally. The goal of this implantation within two portions of the TA muscle was to mimic true and false vocal fold closure that is seen with a normal swallow. Other animals have been implanted with electrode combinations that include the LCA. We have noted that arytenoid closure occurs best with stimulation of the lateral aspect of the body of the arytenoid cartilage.

Our approach to implantation of the electrodes for FES has employed an MT approach with videoendoscopic guidance similar to the approach for fat grafting of the vocal folds. In this approach, the anterior larynx is exposed and the location of the anterior commissure is confirmed by placing a 22-gauge 1.5 cm hypodermic needle through the thyroid cartilage. For the true vocal fold TA electrode, the

needle is inserted approximately 3-4 mm off midline and half the distance between the superior border of the cricoid cartilage and the thyroid notch. The needle is inserted in a 15 degree lateral direction and horizontally within the inferior TA. Placement of the needle in this location is confirmed by videoendoscopy. The depth of insertion is determined by manual motion of the arytenoid cartilage while guiding the needle tip in a medial direction. A 3 mm cutting bur is then used to open the MT at the exact insertion site of the hypodermic needle. The MT burr hole is angled to accommodate the direction of insertion. Great care should be taken in creating the MT to prevent communication with the airway. After creating the MT, a 1 gauge needle loaded with the stimulating electrode is passed in a similar fashion as the hypodermic needle under videolaryngoscopic guidance. The carrier needle is removed and the stimulating electrode remains adjacent to the arytenoid cartilage body. The electrode exits through the MT and is connected to an implantable stimulating device.

The placement of the superior TA electrode within the ventricular fold occurs in a similar fashion. The hypodermic is placed through the thyroid cartilage at the midpoint of the superior and inferior border and 3-4 mm lateral to midline. The needle is directed 15 degrees laterally and 15 degrees superiorly. Visual confirmation within the ventricular fold is made by videolaryngoscopy. Again, a 3 mm cutting bur is used to make the MT and the carrier needle and electrode are inserted in a fashion similar to the hypodermic needle.

While certain adaptations will be required in humans, the basic concept of the MT will likely persist. In the canine, we have had no difficulty with edema or hematoma formation within the larynx despite placing stimulating electrodes in both true and false vocal folds bilaterally.

Minithyrotomy-Arytenoid Adduction

In addition, we have found the MT to be useful in placing an arytenoid adduction suture. This is currently our preferred initial operation for patients with a vocal fold paralysis with a lateralized arytenoid cartilage. In a recent review of primary arytenoid adductions performed for vagal paralysis after skull base surgery, 16/20 patients had complete glottic closure on videostroboscopy with excellent voice results. The remaining patients were noted to have a persistent anterior glottal gap requiring a thyroplasty as a secondary procedure. All arytenoid adductions were performed under general anesthesia immediately after skull base surgery.

Our current arytenoid adduction procedure is a modification of the procedure performed by G. Berke[21]. Initially, the midline of the thyroid cartilage is skeletonized as well as the posterior larynx. The muscular attachments of the thyrohyoid and inferior constrictor muscles are taken down. The external branch of the SLN is routinely identified and preserved. The piriform mucosa is elevated posteriorly from the undersurface of the thyroid cartilage. The muscular process is identified without disturbing the fascial insertion of the (posterior cricoarytenoid) PA and LCA muscles. The cricoarytenoid joint is not opened. A 4-0 prolene suture is placed through the fascial covering of the muscular process of the arytenoid. An MT is created with a 3 mm cutting bur as near to the anterior midline as possible and as close to the inferior border of the thyroid cartilage without jeopardizing the integrity of the inferior border of the thyroid cartilage. Superficially, the MT is angled toward the opposite side. A 16-gauge angiocath needle is passed from the MT to the lateral aspect of the body of the arytenoid cartilage with minimal force and trauma. The prolene suture is passed from posterior to anterior through the angiocath needle acting as a conduit. The angiocath in withdrawn and the suture is tied over a 4-hole microplate.

Conclusion

This paper is not an attempt to state how these difficult problems in the larynx should be handled. It is merely a report on our experimental approaches to these problems. We have found that using a MT can give a unique approach to the larynx. It is well tolerated and minimally invasive. Visualization can be obtained transorally with a video camera on an endoscope. Instrument access is excellent; however, refinements in the surgical instruments are necessary. For both lamina propria replacement surgery, implantation of stimulating electrodes and the placement of the arytenoid adduction suture, an MT approach gives a favorable direction and approach to accomplish the goals of the surgery. The MT is versatile in that it can provide access to the infrafold and leading edge area for lamina propria surgery, or to the LCA and TA for electrode placement.

The results from the laryngofissure and MT approaches to the scarred lamina propria suggest that one may operate on the lamina propria and expect voice improvement. In our small experience of medium length follow-up, we have been happy, but certainly not satisfied, with our patient outcome. They have not been as successful as the canine research demonstrated. As with all surgeries, there is a learning curve and we continue to make modifications. We expect many more modifications by others and us who are performing this type of work. The correct size of the pocket, amount of fat (or other material) placed and creating symmetry are all areas which need further work. Perhaps better substances will become available to improve tissue viscosity. Regardless of these potential changes in the future, we are starting to feel more comfortable and satisfied with the surgical approaches.

References

1. Woo P, Casper J, Colton R, Brewer D. Diagnosis and treatment of persistent dysphonia after laryngeal surgery: a retrospective analysis of 62 patients. *Laryng* 1994;104:1084-91.

2. Ford CN, Bless DM, Loftus JM. Role of injectable collagen in the treatment of glottic insufficiency: A study of 199 patients. *Ann Otol Rhinol Laryngol* 1992;101:237-47.

3. Ford CN, Bless DM. Collagen injection in the scarred vocal fold. *J Voice 1987;1:* 1 16-8.

4. Koufman JA, Isaacson G. Laryngoplastic phonosurgery. *Oto Clin North Amer 1991;24:1151-71.*

5. Benninger MS, Alessi D, Archer S, Bastian R, Ford C, Koufman J, Sataloff RT, Spiegel JR, Woo P. Vocal fold scarring: Current concepts and management. *Oto HNSJ* 1996;115(5):474482.

6. Matsushita H. The vibratory mode of the vocal folds in the excised larynx. *Folia Phoniatr* 1975;27:7-18.

7. Baer T. *Investigation of phonation using excised larynx.* Cambridge, MA: Institute of Technology; 1975. Ph.D. Thesis.

8. Hirano M. Structure of the vocal fold in normal and disease states. *Anatomical and physical study. ASHA Report* 1981;11:11-30.

9. Gray SD, Hirano M, Sato K. Molecular and cellular structure of vocal fold tissue. *Vocal Fold Physiology: Frontiers of Basic Science*, edn 1. Titze IR, ed. San Diego: Singular Publishing; 1993:1-34.

10. Hammond T, Zhou R, Hammond E, Pawlak A, Gray SD. The intermediate layer: a morphologic study of the elastin and hyaluronic acid constituents on normal human vocal folds. *J of Voice* 1997; 11(1):59-66.

11. Pawlak A, Hammond T, Hammond E, Gray SD. Immunocytochemical study of proteoglycans in vocal folds. *Ann Otol Rhinol & Laryngol* 1996; 105(1):6-11.

12. Okamura H, Yumoto E, Okamoto K. Wound healing of canine vocal folds after phonosurgery. *Ann Otol Rhinol Laryngol* 1987;96:425-428.

13. Kass ES, Hillman RE, Zeitels SM. Vocal fold submucosal infusion technique in phonomicrosurgery. *Annals Otol Rhinol Laryngol* 1996;105(5):341-347.

14. Balazs EA, Gibbs DA: The theological properties and biological function of hyaluronic acid. In Balazs EA, ed. *Chemistry and molecular biology of the intercellular matrix* (vol. 3). New York: Academic Press, 1970:1241-1253.

15. Zhu WB, Mow VC. Viscometric properties of proteoglycan solutions at physiological concentrations. In Mow VC, Ratcliffe A, Woo S. eds. *Biomechanics of diarthrodial joints.* New York: Springer-Verlag, 1990:313-344.

16. Titze IR. The physics of small-amplitude oscillation of the vocal folds. *J Acoust Soc Am* 1988; 83:1536-1552.

17. Bastian RW. Benign mucosal and saccular disorders: benign laryngeal tumors. Chapter 103, 1897-1924, *Otolaryngology Head and Neck Surgery*, 2nd Edition, Cummings CW (ed.), Mosby Year Book, 1992.

18. Chan RW, Titze IR. Viscosities of injectable biomaterials in vocal fold augmentation surgery. *Laryng*, submitted for publication.

19. Wexler D, Gray SD, Jiang J, Titze IR. Phonosurgical studies: Fat-graft reconstruction of injured canine vocal cords. *Annals Otol Rhinol Laryngol* 1989;98:668-73.

20. Jiang JJ, Titze IR, Wexler DB, Gray SD. Fundamental frequency and amplitude perturbation in reconstructed canine vocal folds. *Annals Otol Rhinol Laryngol* 1994;103(2):145-148.

21. Berke G.S. Voice disorders and phonosurgery. In: Bailey B.J., ed. *Head & Neck Surgery Otolaryngology.* Philadelphia, PA: W.B. Saunders, (1)644-657, 1993.

NCVS Status and Progress Report - 12
June 1998, 163-173

Vocal Violence in Actors: An Investigation into its Acoustic Consequences and the Effects of Hygienic Laryngeal Release Training

Karen S. Ryker, Ph.D.
Department of Theatre and Drama, The University of Wisconsin-Madison
Nelson Roy, Ph.D.
Department of Communication Disorders, Minot State University
Diane M. Bless, Ph.D.
Department of Communicative Disorders, The University of Wisconsin-Madison
Department of Otolaryngology - Head & Neck Surgery, The University of Wisconsin-Madison

Abstract

Acoustic analysis techniques were used to investigate the short-term consequences of vocally violent behavior, and to compare voice production before and after training in "hygienic laryngeal release" (HLR) techniques. Twenty-seven actors ranging in age from 17 to 48 years were audiorecorded before and after multiple productions of four vocally violent behaviors:grunting, groaning, sobbing, and shouting. Following training in HLR techniques, the experimental protocol was repeated. Audiorecordings of vowels (produced at three pitch levels: modal Fo, minimum Fo, maximum Fo) before and after vocal violence, and before and after HLR training, were analyzed using the Multidimensional Voice Program (4305, Kay Elemetrics Corp., Lincoln Park, NJ). Following vocal violence, no consistent acoustic changes were detected for voice generated at modal and minimum F; however, significant increases in both fundamental frequency range and maximum Fo were observed. Following training in HLR techniques, acoustic measures sensitive to pitch and amplitude perturbation, and non-harmonic noise, improved across pitch levels. The results also suggested that vocal training does defend the laryngeal system from undesirable changes related to vocally violent maneuvers that might surface at the extremes of an actor's pitch range. Because the HLR technique used in this investigation was multimodal, interesting questions are raised regarding which aspect of training is primarily responsible for the observed effects. Further study is required to identify such factors.

Introduction

Actors, in rehearsal and performance, frequently engage in emotionally charged behaviors, often producing voice accompanied by extreme physical exertions (as in a staged fight), or sudden emotional outbursts, such as screaming, shouting, grunting, groaning, and sobbing. These vocally violent behaviors appear to involve extremes in pitch and loudness, increases of muscular tension in the circumlaryngeal area, and explosions of air across partially closed vocal folds. Such behaviors are generally accepted to be vocally abusive, and may contribute to vocal fold mucosal injury and voice mutation.

In an attempt to preserve the voice and limit the untoward effects of these potentially damaging vocal behaviors, voice trainers have taught actors stage performance techniques. Often these techniques focus body postures and laryngeal gestures that release the voice without putting concomitant strain on the larynx. While most voice teachers maintain that healthy vocal technique can preserve a voice and prevent damage, to date there are no objective data to support this claim. Instrumentation is now available to make both quantitative and qualitative measurements of the vocal mechanism before and after training in techniques designed to "release" the voice. The purpose of this study is therefore, twofold: (1) to evaluate the short-term consequences of vocally violent behaviors, and (2) to compare phonatory characteristics before and after training in "hygienic laryngeal release" (HLR) techniques.

Although the professional voice literature is replete with references to vocal "abuse and misuse," there exists little objective information regarding what constitutes "abusive" sounds, how they are made, and what frequency, intensity, and duration of abuse is required to produce perceptible changes in voice or laryngeal tissue. Johnson (1) lists several behaviors thought to contribute to voice disturbance including loud talking, yelling, screaming, hard glottal attack, and speaking outside acceptable physiologic range. Most of these behaviors are common to actors and characterize many of their vocal exertions during performance. As described by Raphael (2)(p.87), performance therefore presents a formidable vocal task for the actor: "In order to create the illusion of illness, or duress or high emotion, the actor may modify breathing or tighten the shoulders and jaw or constrict the voice, any of which may be harmful to the vocal mechanism." This challenge is further explained by Withers-Wilson (3), who uses the example of Richard III's famous speech: "A horse! A horse! My kingdom for a horse!" Withers-Wilson summarizes,

> "As craftsman, the actor must simultaneously be saying to himself, 'Keep your throat relaxed. Call the words in your middle range. Use vowel extension. Take a breath before "A horse!" Take another breath before the next "a horse!"....The actor allows thecharacter's thoughts, feelings, manner of movement and voicing to be "housed" in his mind, heart, and body. But in performing the role, a part of the actor's consciousness must always be monitoring the execution of his character's physical and vocal expression so that he does not inflict injury upon himself or others (p. 99)."

To limit potential damage and extend longevity of the voice, several acting teachers including Berry (4), Fitzmaurice (unpublished observation), Lessac (5), Linklater (6), McCallion (7), Raphael (8), and Rodenberg (9) have developed techniques to produce potentially vocally violent behaviors in a more hygienic, less abusive manner. While each expert's approach varies, the techniques share in common muscular release, an open pharynx and jaw, abdominal breath support, focus of tone, placement of pitch, absence of glottal attack, extension of vowels and a warm-up. Some of these approaches have been used for decades, but objective assessment of their effectiveness has yet to be undertaken. Since actors must engage — often repeatedly — in vocally violent behaviors in the line of performance, information would be useful regarding how vocal violence affects vocal function, and whether training can moderate the effects of such violence. This study is the first in a series of collaborative studies between theatrical artists, voice pathologists, and voice scientists. It is part of a broader investigation that aims to evaluate acoustic, electroglottographic, and laryngeal kinesiological correlates of (a) vocally violent behaviors, and (b) the training efforts designed to modulate such behaviors.

Methods

Subjects

Twenty-seven actors (15 male, 12 female) participated in this investigation. Subjects' ages ranged from 17 to 48 years (mean 25.0 (7.1) years). As part of the larger investigation, subjects were divided into two groups based on level of acting training and experience. Group A, labeled "Novice," included 19 subjects (11 males, 8 females) with little or no professional experience and limited acting training (mean age 22.3 (4.0) years). Group B, labeled "Advanced" (mean age 32.4 (8.0) years), included 8 subjects (4 males, 4 females) with two to four years of graduate level acting and voice training, as well as varying degrees of professional experience.

All subjects were recruited from among 34 actors rehearsing and performing in summer theatre productions. Thirty subjects agreed to participate. Of these, three subjects were excluded because they did not tolerate the endoscopic evaluation. During testing and training periods, all subjects were involved in daily rehearsals and/or performance of theatrical productions.

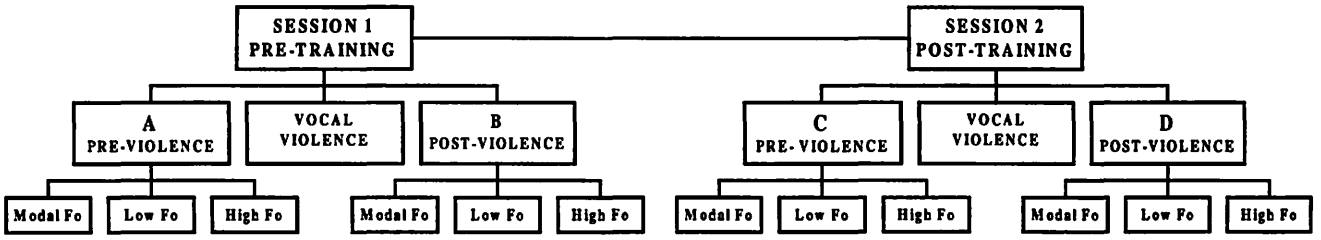
Procedures

To familiarize subjects with the test protocol, and to minimize anxiety concerning the endoscopic procedure, participants viewed a 15 minute videotape that showed two actors performing the required maneuvers with the flexible endoscope in situ. Prior to participating, each subject also completed an extensive case history questionnaire designed to identify subjects with a previous history of voice difficulties. The nature and extent of these difficulties were discussed with the second author prior to participating in the study. No subjects were excluded based upon the results of the case history questionnaire.

Each subject underwent the same assessment protocol on two separate occasions, before (session 1) and after (session 2) vocal training (Figure 1). The assessment protocol, which was divided into five sections, is briefly described below.

(1) Warm-up. For each testing session, the actor was guided in a 10 minute vocal warm-up (Appendix 1), which involved gentle muscle stretching, humming, yawning, and focus of breath and tone.

(2) Pre-test Audiorecording. Audiorecordings of each participant were made in a sound treated booth using a digital audiorecorder (PCM-2300; Sony Electronics, Park Ridge, NJ) with a headset-mounted windshielded microphone (C410; AKG Acoustics, Nashville, TN). A constant mouth-to-microphone distance of 2 cm was maintained. Recorded vocal tasks included (1) sustained productions of



(I) $VOCAL\ VIOLENCE\ EFFECT = (A + C) - (B + D)$

(II) $TRAINING\ EFFECT = (A + B) - (C + D)$

(III) $INTERACTION = (A - B) - (C - D)$

Figure 1. This schematic illustrates the time line and linear contrasts employed in this experiment. Each subject was evaluated on two separate occasions, before and after training in hygienic laryngeal release techniques (i.e., session 1 and 2). During each session, each subject was audiorecorded twice, before and after engaging in vocal violence maneuvers (i.e., pre-vocal violence [A and C] and post-vocal violence [B and D]). Sustained vowels produced at three pitch levels—modal Fo, low Fo, and high Fo—were recorded. This generated twelve sustained /a/ tokens for each subject, which were subsequently submitted for acoustic and statistical analysis. The contrasts that constituted the two main effects and their interaction are also provided to assist in the interpretation of the results.

the vowels /a/ and /i/ at comfortable pitch and loudness (modal Fo); (2) pitch glide down the scale from midrange to the lowest sustainable pitch on the vowel /a/ (minimum Fo); (3) pitch glide up the musical scale from midrange to the highest sustainable pitch on the vowel /a/ (maximum Fo); and (4) oral reading of the first paragraph of the Rainbow Passage (Fairbanks, 1960).

(3) Flexible laryngovideostroboscopy (FLVS). Flexible laryngovideostroboscopy was accomplished using a stroboscope (RLS 9100, Kay Elemetrics Corp., Lincoln Park, NJ) combined with a flexible nasopharyngolaryngoscope (FNL 10RP2, Pentax Technologies Corp., Broomfield, CO). Although the results from the flexible endoscopic examination are not reported in the current study, it is considered sufficiently important to the nature and interpretation of the data collected that the procedure be detailed here. Each FLVS examination was divided into three sections: (1) achieving a baseline through vocal tasks similar to those in the pre-test audiorecording, (2) producing four vocally violent actions, and (3) repeating the baseline vocal tasks. All vocalizations were produced within the vowel /i/ context to afford optimal visualization of the endolarynx.

Each subject's nasal passage was first anaesthetized with topical anesthesia (4% lidocaine). Electroglottography (EGG) was used to track fundamental frequency and provide a visual display of glottal contact patterns. EGG electrodes were secured over the thyroid lamina. Each subject was assessed while standing against a "leaning board." The leaning board provided constant resistance against which the actors could press to maintain postures consistent with the lower body position used in the vocally violent maneuvers. The tip of the endoscope was positioned in the midpharyngeal region to achieve *baseline* FVLS. Each subject was recorded while sustaining the vowel /i/ at comfortable pitch and loudness, during loud and quiet volumes, and during glissando from mid to highest sustainable pitch and from mid to lowest sustainable pitch.

Responding to scripted instructions read by the first author (appendix 2), each subject then produced multiple repetitions of four emotionally charged vocal behaviors: "grunt, groan, sob, and shout." Subjects were instructed to produce each target vocal behavior at stage volume. To ensure proper sampling and visualization of each vocal behavior, subjects were stopped when at least four adequately visualized tokens of the target behavior were obtained. This resulted in a session 1 mean total of 41.4 (12.9) productions per actor and a session 2 mean total of 40.2 (13.5) productions per actor. Based on a pairwise t-test comparison, these means were not significantly different.

Following completion of these vocally violent maneuvers, subjects repeated the same vocal tasks required in the baseline FLVS testing.

(4) Warm-down. Following FLVS testing, each actor was led in a 3 minute warm-down exercise (Appendix 3) directed toward extralaryngeal muscle loosening, stretching, and gentle humming.

(5) Posttest Audiorecording. The session concluded with a digital audiorecording using the identical voice stimuli employed in the earlier pretesting audiorecording (step (2) above).

Description of Training in Hygienic Laryngeal Release Techniques

Following session 1, all actors underwent two 2-hour HLR training sessions held by the first author. In the

Novice group, the number of actors attending a given training session varied from 3 to 9. The time between the two training sessions ranged from 5 to 24 days (mean 7 days). All Novice actors were then retested within 18 days of the final training session (mean 11 days, range 1-18). In the Advanced group, all 8 actors were trained together on two separate occasions. However, for this group, approximately 4 months (mean 127 days, range 124-133) elapsed between the final training session and the second testing session due to difficulties with scheduling. Total time elapsed from session 1 to session 2 ranged from 18 to 38 days (mean 28.4 [6.8] days) for the Novice group and 159 to 251 days (mean 173.3 [31.6] days) for the Advanced group.

In brief, training employed techniques largely based on the approach described by Raphael (2, 8) which involves aspects of warm-up and warm-down; muscular release of the head/neck, face, jaw, tongue, torso and limbs; appropriate breath support; use of resonant voice and placement of pitch; and diversion of vocal effort away from the larynx while clarifying the acting intent in making the sound.[1] Training in vocal hygiene and proper hydration also was provided. Thoughout each 2 hour session, actors practiced specific goals while engaging in these techniques. Demonstrations and corrective feedback were provided as required. Participants were encouraged to rehearse these techniques between training sessions.

Acoustic Analysis

Audiorecordings of the sustained vowels—produced at the three pitch levels (i.e., modal Fo, minimum Fo and maximum Fo)—were digitized at a sampling rate of 25kHz and analyzed using the Multi-Dimensional Voice Program (MDVP 4305, Kay Elemetrics Corp., Lincoln Park, NJ). The MDVP generates measures for 18 acoustic parameters: Fo (Hz) Fundamental frequency, STD (Hz) Standard deviation of Fo, Jita (us) Absolute Jitter, Jitt (%) Jitter Percent, RAP (%) Relative Average Perturbation, PPQ (%) Pitch Period Perturbation Quotient, sPPQ % Smoothed Pitch Period Perturbation Quotient, vFo % Coefficient of Fundamental Frequency Variation, ShdB (dB) Shimmer in dB, Shim (%) Shimmer Percent, APQ (%) Amplitude Perturbation Quotient, sAPQ (%) Smoothed Amplitude Quotient, vAm (%) Coefficient of Amplitude Variation, NHR Noise-to-Harmonic Ratio, VTI Voice Turbulence Index, SPI Soft Phonation, FFTR (Hz) Fo-Tremor Frequency, and ATRI (%) Amplitude Tremor Intensity Index (%).

The central 3 second segment of each sustained vowel /a/ production (modal Fo) was extracted for analysis. The central 1000 ms portion of the minimum and maximum Fo productions were also analyzed. Thus, a total of twelve vowel tokens—3 taken from each session A through D—were analyzed for each subject (Figure 1).

[1] The interested reader is referred to Raphael (2,8) for a more comprehensive description of the training goals and procedures. Also a complete training protocol summary is available from the first author upon request

Results

To answer our research questions regarding (1) the effects of short-term vocal violence and training in laryngeal release techniques, as well as (2) the effectiveness of training to moderate the effects of vocal violence on the acoustic signal, the acoustic measures were compared using a repeated measures analysis of variance (ANOVA). The measures were first log-transformed to meet the assumption of homogeneous variances required by ANOVA, and then analyzed using the SAS system's MIXED procedure with an unstructured covariance matrix (SAS Institute Inc., Cary, NC). Unlike classical repeated measures analysis of variance, this procedure allows for inclusion of subjects with some missing values.

Our experimental design permitted three contrasts to be examined: (1) pre- versus postvocal violence (i.e., the Vocal Violence Effect), (2) pre- versus posttraining (i.e., the Training Effect), and (3) the interaction of (1) and (2) (i.e., the Interaction Effect). Figure 1illustrates the experimental time line and the linear combination that constituted these contrasts. The Vocal Violence Effect compared acoustic measures before and after engaging in the vocally violent maneuvers in both session 1 and session 2. The Training Effect compared the acoustic measures before and after training in laryngeal release techniques. Finally, the Interaction Effect determined the effectiveness of training to moderate changes in the voice signal caused by vocal violence. As such, this contrast compared differences before and after vocal violence with differences before and after training; that is, the *difference* of the differences. If training was effective in diminishing immediate harm from vocal violence, the pre- versus postvocal violence difference, after training should be smaller than the pre- versus postvocal violence difference, before training. The significant results ($p<.05$) are presented separately for modal Fo, minimum Fo, and maximum Fo productions.

Modal Fo

A summary of all acoustic measures for modal Fo productions is provided in Table 1. No significant Vocal Violence Effect was detected when comparing subjects' acoustic measures before and after engaging in the vocally violent maneuvers. All acoustic measures remained relatively stable after performing such behaviors on both occasions, suggesting that short-term vocal violence had few unfavorable effects on the voice signal when produced at this comfortable pitch. Furthermore, no significant Interaction Effect was detected, suggesting that training in hygienic release techniques had no discernible effect in preventing immediate harm from vocal violence.

Table 1.
Means and (standard deviations) of acoustic measures obtained from analysis of sustained vowels produced at **modal Fo** at four times (A, B, C, D) during the experimental timeline.

Acoustic Measure	Session 1 (Pretraining)		Session 2 (Posttraining)	
	A (Pre-VV)	B (Post-VV)	C (Pre-VV)	D (Post-VV)
Fo (Hz)	165.9 (50.7)	164.4 (50.0)	168.6 (46.9)	166.5 (51.7)
STD (Hz)	1.13 (0.43)	1.21 (0.47)	1.22 (0.41)	1.35 (0.59)
Jita (us)	30.0 (18.2)	27.2 (14.1)	26.2 (13.9)	29.2 (16.0)
Jitt (%)	0.45 (0.24)	0.40 (0.17)	0.41 (0.17)	0.44 (0.19)
RAP (%)	0.27 (0.16)	0.24 (0.11)	0.24 (0.11)	0.26 (0.12)
PPQ (%)	0.25 (0.12)	0.24 (0.09)	0.24 (0.09)	0.26 (0.11)
sPPQ (%)	0.45 (0.13)	0.45 (0.11)	0.47 (0.14)	0.45 (0.16)
vFo (%)	0.70 (0.18)	0.74 (0.20)	0.73 (0.16)	0.84 (0.35)
ShdB (dB)	0.20 (0.08)	0.19 (0.06)	0.16 (0.06)	0.18 (0.06)
Shim (%)	2.33 (0.90)	2.21 (0.70)	1.89 (0.63)	2.07 (0.72)
APQ (%)	1.71 (0.74)	1.66 (0.70)	1.42 (0.50)	1.46 (0.48)
sAPQ (%)	2.69 (1.06)	2.72 (1.03)	2.32 (0.68)	2.31 (0.65)
vAM (%)	6.56 (2.66)	8.27 (3.24)	7.40 (2.47)	7.98 (4.75)
NHR	0.13 (0.02)	0.12 (0.02)	0.12 (0.02)	0.12 (0.02)
VTI	0.04 (0.03)	0.04 (0.02)	0.03 (0.01)	0.03 (0.01)
SPI	17.9 (10.4)	17.7 (10.8)	18.5 (8.30)	15.9 (0.12)
FTRI (%)	0.28 (0.11)	0.30 (0.13)	0.32 (0.13)	0.28 (0.12)
ATRI (%)	2.45 (1.53)	2.54 (1.56)	2.77 (1.55)	2.23 (1.48)

When comparing the pretraining (session 1) to the posttraining (session 2) acoustic measures however, a significant Training Effect was detected. Several acoustic measures improved following the training period, suggesting that post-training acoustic measures were better than in the pre-training condition. Four measures of short- and long-term amplitude perturbation were reduced after the HLR training interval (Figure 2). Shimmer (dB) and shimmer (%), two measures that are sensitive to short-term period-to-period variability of the peak-to-peak amplitude within the analyzed voice sample, were significantly lower in the post-training condition (session 2). In addition, amplitude perturbation quotient (APQ) and smoothed amplitude perturbation quotient (sAPQ) also were reduced after session 2. APQ provides an evaluation of the relatively short-term (smoothing factor of 11 periods), cycle-to-cycle irregularity of the peak-to-peak amplitude of the voice signal, and is the preferred measurement for shimmer in the MDVP. The smoothing reduces the sensitivity of APQ to pitch extraction errors. Smoothed amplitude perturbation quotient (sAPQ), on the other hand, provides a relatively *long-term* evaluation of the variability of the peak-to-peak amplitude (smoothing factor of 55 periods). Therefore, APQ and sAPQ differ primarily in the smoothing factor employed.

In addition to short- and long-term amplitude perturbation changes, two noise-sensitive measurements improved substantially following the training interval (Figure 3). Noise-to-Harmonic Ratio (NHR) provides an average ratio of energy of nonharmonic components in the 1.5 to 4.5 kHz frequency range to the energy of harmonic components in the 70 to 4500 Hz frequency range. This measure indexes the degree of noise in the analyzed voice signal by detecting amplitude and frequency modulations, turbulence noise, subharmonic components, and/or voice breaks. A second measure, voice turbulence index (VTI), was also lower following the training period. VTI provides an estimate of the relative energy of high- frequency noise by comparing the spectral inharmonic high-frequency energy in the 2.8 to 5.8 kHz range to the spectral harmonic energy in the 70 to 4500 Hz range. VTI makes this comparison in areas of the signal where the influence of voice breaks, subharmonic components, amplitude/frequency variations are minimal.

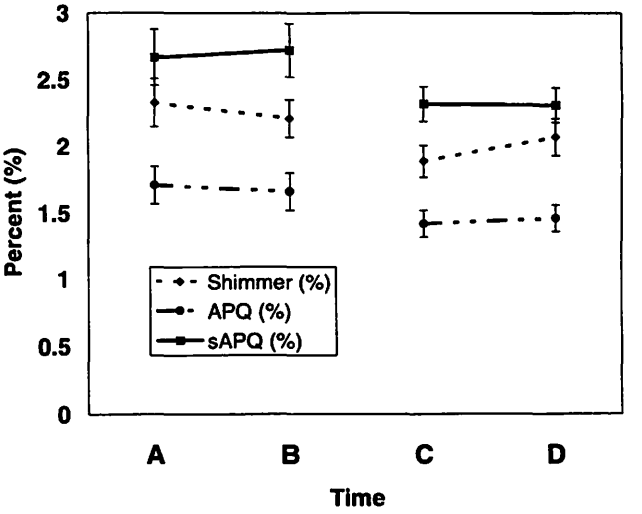


Figure 2. Comparison of amplitude perturbation measures of sustained vowels produced at Modal Fo before (A and B) and after (C and D) training revealed a significant Training Effect (p<.05).

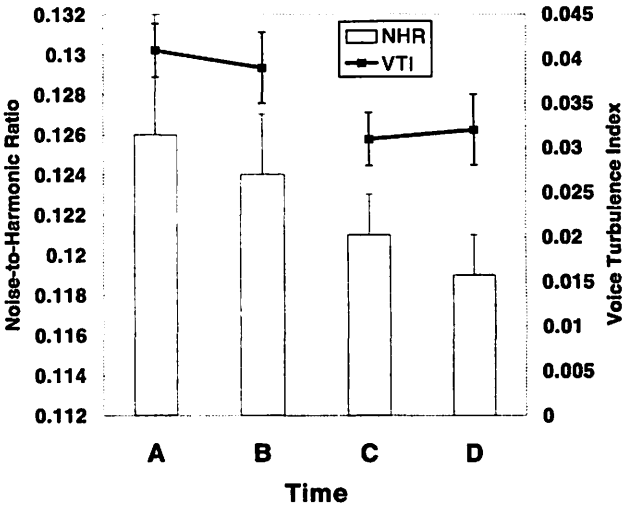


Figure 3. Comparison of noise sensitive measures of vowels produced at modal Fo before (A and B) and after (C and D) training revealed a significant Training Effect (p<.05).

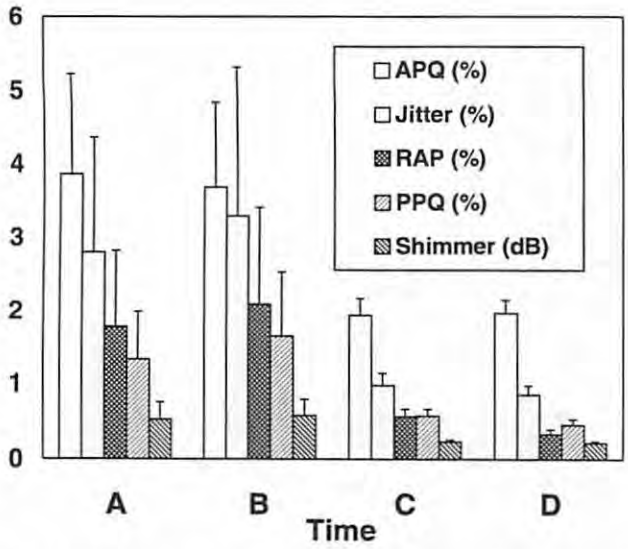


Figure 4. Comparison of frequency and amplitude perturbation measures before (A and B) and after (C and D) training revealed a significant

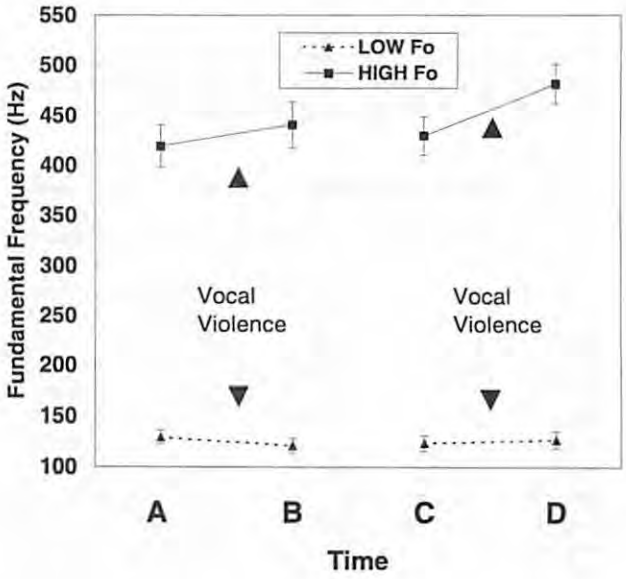


Figure 5. Comparison of fundamental frequency range before and after vocal violence revealed a significant Vocal Violence Effect ($p<.05$). Pitch range was greater immediately following vocal violence in both sessions, primarily as a result of an expansion of the uppermost pitch region. A significant interaction effect ($p<.05$) also was observed for minimum Fo productions, suggesting that training appeared to moderate the effects of vocal violence on the lowest-pitched productions.

Table 2.
Means and (standard deviations) of acoustic measures obtained from analysis of sustained vowels produced at **minimum Fo** at four times (A, B, C, D) during the experimental timeline.

Acoustic Measure	Session 1 (Pretraining)		Session 2 (Posttraining)	
	A (Pre-VV)	B (Post-VV)	C (Pre-VV)	D (Post-VV)
Fo (Hz)	129.1 (36.0)	121.0 (36.8)	123.4 (38.9)	126.8(43.0)
STD (Hz)	3.10 (5.86)	3.39 (8.69)	1.42 (0.77)	1.72 (1.02)
Jita (us)	220.6 (641.0)	257.2 (753.9)	92.1 (85.0)	73.8 (51.3)
Jitt (%)	2.80 (7.97)	3.29 (10.3)	0.98 (0.81)	0.86 (0.60)
RAP (%)	1.78 (5.30)	2.09 (6.72)	0.57 (0.50)	0.50 (0.37)
PPQ (%)	1.33 (3.36)	1.65 (4.46)	0.57 (0.47)	0.51 (0.36)
sPPQ (%)	1.09 (1.11)	1.02 (0.51)	0.84 (0.57)	0.95 (0.61)
vFo (%)	2.29 (4.36)	2.50 (5.79)	1.22 (0.75)	1.37 (0.73)
ShdB (dB)	0.53 (1.19)	0.58 (1.12)	0.23 (0.14)	0.22 (0.11)
Shim (%)	5.74 (12.6)	5.07 (10.2)	2.60 (1.61)	2.56 (1.19)
APQ (%)	3.86 (6.95)	3.77 (5.91)	1.94 (1.18)	1.97 (0.91)
sAPQ (%)	4.15 (2.62)	3.75 (2.09)	3.16 (1.36)	3.66 (1.78)
vAM (%)	7.22 (6.96)	7.01 (7.15)	5.41 (2.15)	5.45 (2.35)
NHR	0.18 (0.22)	0.17 (0.18)	0.13 (0.03)	0.15 (0.13)
VTI	0.06 (0.06)	0.52 (2.33)	0.05 (0.02)	0.07 (0.10)
SPI	8.46 (5.88)	10.1 (8.20)	8.37 (4.45)	13.5 (13.8)
FTRI (%)	0.62 (0.68)	0.44 (0.26)	0.42 (0.24)	0.57 (0.34)
ATRI (%)	2.94 (2.63)	2.67 (1.95)	2.17 (1.53)	2.70 (1.48)

Further inspection of the data showed that no significant differences were detected before or after vocal violence, or before and after training, on measurements of fundamental frequency, short- and long-term frequency perturbation, degree/number of voice breaks and/or subharmonics, or tremor.

Minimum Fo

Comparison of session 1 and session 2 identified a significant Training Effect when subjects sustained a vowel at the bottom of their pitch range (Figure 4). Several measures of phonatory instability were significantly reduced following HLR training. Specifically, pitch perturbation measures, including jitter (%), a measure of the period-to-period variability in fundamental frequency, and Relative Average Perturbation (RAP %) and Pitch Perturbation Quotient (PPQ %), measures of the variability of the pitch period at smoothing factor 3 and 5 periods respectively, all were reduced after training. In addition, two indices of amplitude perturbation—shimmer and APQ—also were significantly lower after training. No Vocal Violence Effect was detected for minimum pitch productions (Table 2).

A significant Interaction Effect also was detected for fundamental frequency (Figure 5). The change in the minimum sustained fundamental frequency following vocal violence in session 1 was greater than that seen in session 2. Subjects attained a lower minimum fundamental frequency following vocal violence in session 1, whereas after session 2 a smaller change in the opposite direction was observed. Specifically, the minimum Fo was lowered almost 6 Hz, from 128 Hz to 122 Hz, following vocal violence in session 1, but after vocal violence in session 2 a small elevation in minimum Fo was observed.

Maximum Fo

In the case of the highest sustainable pitch productions, a significant Vocal Violence Effect was observed. Following engaging in vocal violence, subjects were able to attain a higher maximum Fo regardless of training status (Figure 5). Averaged across subjects, the maximum Fo after vocal violence increased nearly 20 Hz in session 1, and 50 Hz in session 2 (Table 3).

In addition, a significant Training Effect was detected for two acoustic measures, Soft Phonation Index (SPI)

Table 3.
Means and (standard deviations) of acoustic measures obtained from analysis of sustained vowels produced at **maximum Fo** at four times (A, B, C, D) during the experimental timeline.

Acoustic Measure	Session 1 (Pretraining) A (Pre-VV)	B (Post-VV)	Session 2 (Posttraining) C (Pre-VV)	D (Post-VV)
Fo (Hz)	419.1 (106.7)	440.6 (118.5)	429.7 (96.6)	481.8 (99.7)
STD (Hz)	2.86 (1.68)	3.33 (2.07)	3.34 (1.72)	10.0 (32.5)
Jita (us)	12.3 (10.3)	14.1 (14.7)	12.9 (9.09)	13.9 (12.9)
Jitt (%)	0.05 (0.42)	0.58 (0.53)	0.55 (0.37)	0.66 (0.64)
RAP (%)	0.30 (0.25)	0.35 (0.33)	0.33 (0.23)	0.40 (0.40)
PPQ (%)	0.30 (0.24)	0.33 (0.28)	0.32 (0.21)	0.39 (0.36)
sPPQ (%)	0.36 (0.24)	0.40 (0.28)	0.39 (0.20)	0.61 (0.98)
vFo (%)	0.72 (0.55)	0.74 (0.39)	0.79 (0.50)	1.80 (5.40)
ShdB (dB)	0.08 (0.06)	0.08 (0.04)	0.12 (0.10)	0.09 (0.06)
Shim (%)	0.96 (0.65)	0.94 (0.51)	1.31 (1.18)	1.08 (0.68)
APQ (%)	0.68 (0.39)	0.73 (0.39)	0.98 (0.80)	0.83 (0.64)
sAPQ (%)	1.36 (0.66)	1.72 (0.91)	1.94 (1.20)	1.52 (0.78)
vAM (%)	3.99 (1.59)	5.78 (2.39)	6.61 (3.39)	5.99 (2.57)
NHR	0.08 (0.03)	0.08 (0.02)	0.08 (0.02)	0.20 (0.10)
VTI	0.03 (0.02)	0.03 (0.01)	0.04 (0.03)	0.03 (0.03)
SPI	14.5 (10.5)	15.7 (15.0)	9.27 (7.79)	10.4 (7.75)
FTRI (%)	0.21 (0.11)	0.26 (0.23)	0.23 (0.20)	0.27 (0.13)
ATRI (%)	1.75 (0.92)	2.70 (2.35)	3.25 (2.42)	2.04 (1.13)

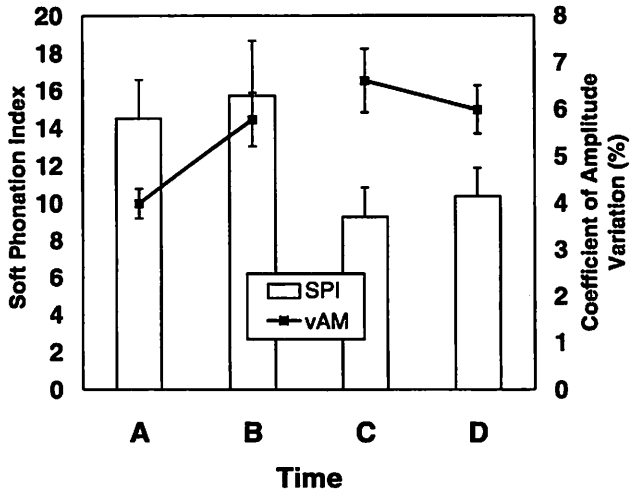


Figure 6. Comparison of soft phonation index (SPI) and coefficient of amplitude variation (vAM %) revealed a significant Training Effect (p<.05) for vowels produced at maximum Fo level. While the amplitude measure of vAM was reduced following training, an increase was observed for SPI.

and Coefficient of Amplitude Variation (vAM). SPI, an average ratio of the lower-frequency harmonic energy (70-1600 Hz) to the higher-frequency (1600-4500 Hz) harmonic energy, was reduced after training. However, vAM, a measure of the very longterm amplitude variation within the analyzed voice signal, increased after training (Figure 6).

There was also a significant Interaction Effect for vAM and sAPQ, two amplitude perturbation measures. Figure 7 shows that when pre- and posttraining results are compared, vocal violence seems to have less effect on the acoustic signal after training. Specifically, in session 1 vocal violence produced an increase in both vAM and sAPQ (i.e., A to B), whereas following training, vocal violence seems to have a

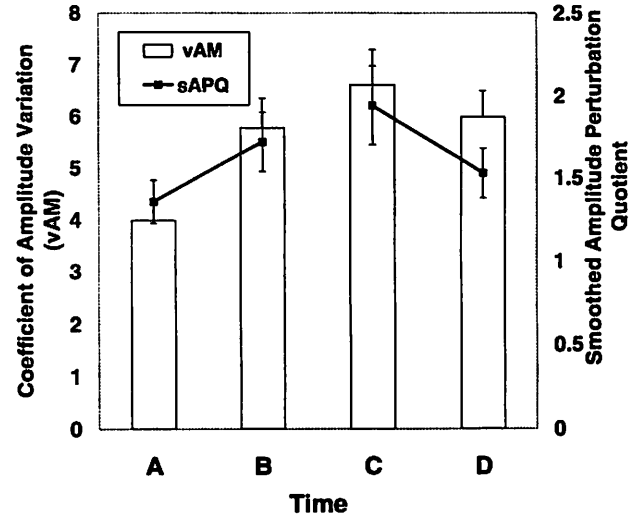


Figure 7. Comparison of amplitude perturbation measures for maximum Fo productions revealed a significant Interaction Effect (p<.05). Changes in the coefficient of amplitude variation (vAM) and smoothed amplitude perturbation quotient (sAPQ %) following vocal violence in the posttraining session (C and D) were in the opposite direction from the pattern observed

Table 4.
A summary of results obtained for each main effect and interaction for the three pitch productions and phonational frequency range (all findings significant at p <.05).

ANOVA MAIN EFFECTS & INTERACTION	MODAL Fo	MINIMUM Fo	MAXIMUM Fo	RANGE
VOCAL VIOLENCE EFFECT	NO EFFECT	NO EFFECT	Fo ↓	RANGE ↑
TRAINING EFFECT	SHIMMER APQ sAPQ NHR VTI ↓	SHIMMER APQ JITTER RAP PPQ ↓	SPI ↓ vAM ↑	NO EFFECT
INTERACTION EFFECT	NO EFFECT	Fo Change (Small & in opposite direction)	vAM sAPQ ↓	NO EFFECT

negligible impact on the acoustic signal, with these amplitude perturbation measures remaining relatively unchanged (C to D). These results suggest that training buffers the laryngeal mechanism against the effects of vocal violence.

Finally, the fundamental frequency range for each subject was derived by subtracting the highest sustained pitch production from the lowest sustained pitch (log Fo Maximum - log Fo Minimum). A significant Vocal Violence Effect was detected for fundamental frequency range. Regardless of whether vocal violence occurred before or after training, it resulted in an expansion of the pitch range, which was primarily evidenced in an elevation of the uppermost region of the pitch range (Figure 5).

Table 4 provides a complete summary of the ANOVA results for the two main effects and their interaction.

Discussion

This investigation used acoustic analysis to study (1) the acoustic consequences of (a) short-term vocal violence and (b) training in hygienic laryngeal release techniques, and (2) whether training in vocal release techniques could diminish the effects of vocal violence. We will discuss each of these research questions separately.

What are the Acoustic Consequences of Short-Term Vocal Violence?

From the results, engaging in vocally violent behaviors produced no consistent changes in the acoustic signal of voice generated at low and comfortable pitch levels (i.e., modal and minimum Fo). However, vocal violence had its greatest impact on high-pitched productions, that is, voice produced in the uppermost pitch range (maximum Fo). This effect was unexpected and contradictory to what we had anticipated. It seems that engaging in shortterm vocal violence of this type may facilitate an expansion of the uppermost region of the fundamental frequency range.

We offer several opinions to reconcile this unexpected result. One possible explanation is that in the short-term, vocally violent behaviors may provide additional physiological "warm-up" for the laryngeal mechanism. It is conceivable that the pretesting warm-up period employed in this study did not provide participants with a sufficient amount/duration of vocalizing. Consequently, the voice may have benefited from the intense vocal activity during testing. Although distinctly different vocal tasks were employed, a similar warm-up phenomenon was reported by Scherer and collegues (10) in their study of a vocally trained subject following extended reading at loud volume (80% level of her total intensity range, at one octave above her lowest sustaninable frequency). Scherer et al. referred to this as vocal warm-up or vocal adaptation. Sherman and Jensen (11) also concluded that prolonged reading over a period of 1.5 hours may be associated with a perceived increase in vocal clarity as judged by listeners. These results are contrasted with the findings of Stemple, Stanley and Lee (12) who did not identify significant change in fundamental frequency range in 10 normal speakers following prolonged voice use (2 hours). It is difficult however to directly compare our findings with these studies and others (13-15) because of significant methodological differences.

A second possible explanation for the observed expansion of the uppermost pitch range involves elevated levels of pretest anxiety. It is feasible that high levels of state anxiety may have produced undesirable laryngeal and extralaryngeal muscular tension, resulting in a restricted pitch range. After completing the test, elevated levels of anticipatory anxiety would be reduced. This reduction in anxiety may have been accompanied by reduced paralaryngeal tension, thus allowing greater flexibility of the laryngeal mechanism, especially in the uppermost pitch region.

Of equal interest is the finding that vocal violence produced no significant acoustic changes in voice generated in the mid and lowest frequencies. One potential explanation may be that the total number and intensity of vocally violent productions was not sufficiently taxing to produce observable changes. Little is known regarding the frequency, duration, and intensity of vocal abuse required to produce a perceptible change (i.e., auditory-perceptual, acoustic, or morphological) in voice. Gelfer, Andrews, & Schmidt (16) expressed similar concerns in their study comparing the effects of prolonged use (i.e., 1 hour at 80% of speaking intensity range) on the voices of trained singers and musically untrained speakers. They did not identify significant acoustic changes following prolonged use in the trained group and questioned whether the vocal task was sufficiently demanding.

It is also likely that subjects were not as vocally violent as they otherwise would have been in an authentic stage situation. The testing paradigm imposed a substantial degree of artificiality. It is recognizably more difficult to achieve authenticity when the actor endures an endoscope through the nose, and is standing in a confined position in a small examination room without aid of a theatrical situation. The presence/proximity of the examiner and the awareness of the endoscope may have created substantial apprehension resulting in more cautious-controlled vocalizations. Certainly questions of ecological validity must be addressed in future studies of this nature.

It is also conceivable that the effects of short-term vocal violence may not be immediately observable and there may be a delay in the physiological consequences of such violence. Such effects may surface only hours later. Anecdotally, subjects were instructed to report any lingering effects of the test procedure, and when contacted within 24 hours after testing, many subjects reported vocal fatigue and compromised vocal function, which had begun several hours after completion of the test. Subjects also reported that these problems resolved quickly, usually by the following morning.

Failure to detect change immediately following short-term violence may also be related to the sensitivity and resolution of the acoustic measures employed. It is possible that perhaps other phonatory function measures may be more sensitive to changes in laryngeal function; for example, aerodynamic measures such as phonation threshold pressure or glottal resistance.

What are the Acoustic Consequences Following Training in Hygienic Laryngeal Release Techniques?

The results indicated that there was a significant Training Effect for all pitch levels, evidenced in the reduction of various measures of short- and long-term phonatory

instability following training. Recall from Figure 1 that the Training effect evaluates the difference in acoustic measures between sessions 1 and 2. To clarify, this finding suggested that the acoustic signal was less perturbed and more stable following the training, regardless of whether the voice was analyzed *before or after* engaging in the vocal violence maneuvers. Because we did not employ a non-treatment, alternative treatment, or placebo control group, we are cautious when interpreting these results. Since the laryngeal release training approach is multimodal, involving practice in vocal violence techniques as well as instruction and practice in vocal hygiene, warm-up/warm-down techniques, muscular relaxation and control, breathing and alignment, effects of hydration, acting technique, and balance of hygienic technique with authentic acting of the violence, it is impossible to determine which factors are primarily responsible for the observed training effect. It is conceivable that instruction in proper hydration and vocal hygiene could partially or even wholly account for the results observed. The absence of a control group also makes it difficult to exclude potential non-training effects related to the time elapsed between sessions 1 and 2 (i.e., temporal effects), as well as the potential effects of interaction with the investigators. It is possible that during session 2, subjects' familiarity with the assessment procedure and the investigators may have lessened apprehension. This reduction in anticipatory anxiety from session 1 to session 2 may have affected voice production and hence the acoustic measures.

This study raises questions concerning whether hygienic laryngeal release techniques moderate the effects of vocal violence. The results provide modest evidence to suggest that vocal training does defend the laryngeal system from unwanted changes related to vocally violent maneuvers. For maximum Fo productions, two amplitude perturbation measures were less affected by vocal violence following training. The relative stability of these measures following vocal violence, in the posttraining session, is consistent with the assumption that training in these extreme performance techniques may serve as a buffer against undesirable laryngeal changes. This finding is of particular interest to actors, who are frequently required to operate at the extremes of their pitch range, and not just in the midrange. Modest phonatory instabilities in the furthermost limits of their range could generate disturbing consequences.

Interestingly, for the minimum Fo productions, training seemed to moderate changes in pitch following vocal violence seen in the pretraining session. Many of the same caveats mentioned previously also apply to this finding.

Caveats, Unanswered Questions, and Areas for Future Research

There are several questions that remain unanswered and as such, these results should be viewed with caution. First, our acoustic analyses were confined to sustained productions; whether prolonged vowels are sufficiently representative of the dynamic laryngeal voicing and devoicing activity that occurs in connected speech warrants further attention. Second, training was short-term and over a relatively brief time period. Whether skill acquisition and mastery could be accomplished in such a short time frame is a matter of debate. Third, can we generalize these results to theatrical settings where there are fewer situational cues to attend to specific vocally adaptive behaviors? Fourth, is one session of vocal violence sufficient to produce detectable changes in voice quality? How long, how often, and how intense must vocal violence be to induce perceptible changes in vocal health? How do these acoustic changes relate to auditory-perceptual attributes? Finally, what other constitutional and health factors influence short- and long-term vocal outcomes following repeated vocal violence?

In spite of these lingering questions, this study provides objective data suggesting that actors can be taught to use vocally violent behaviors without necessarily endangering the vocal mechanism. However, it is difficult to judge which aspect of training was responsible for the observed changes. In future studies, each component of this training should be carefully examined to identify which factors are responsible for the observed changes. Further research should attempt to identify analysis measures that are sensitive to subtle changes in phonatory function.

Acknowledgments

The authors wish to acknowledge the helpful guidance and contributions of Drs. Dennis Heisey and Glen Leverson, biostatisticians in the Department of Surgery, University of Wisconsin-Madison. A debt of gratitude is owed to Sarah Jo Burke, the Waisman Center Voice Lab staff, and the University of Wisconsin Clinical Science Center Department of Surgery, Division of Otolaryngology-Head & Neck Surgery and staff, for their generous contributions toward the completion of this work. Grateful acknowledgment is made to the members of VASTA for their support and bibliographic assistance. This research was supported in part by a University of Wisconsin-Madison Graduate School Research Grant and by the National Center for Voice and Speech through Grant P60 00976 for the National Institute on Deafness and Other Communication Disorders.

References

1. Johnson A. Disorders of Speaking in the Professional Voice User. In: Benninger M, Jacobson B, Johnson A eds. *Vocal arts medicine: The care and prevention of professional voice disorders.* New York: Thieme, 1994.

2. Raphael B. Special Considerations Relating to Members of the Acting Profession. In: Sataloff R, ed. *Professional voice: The science and art of clinical care. New York: Raven,* 1991: 87.

3. Withers-Wilson N. *Vocal direction for the theatre.* New York, Drama Book Specialists, 1993: 99.

4. Berry C. *Voice and the actor.* New York: MacMillan Publishing Company, 1973.

5. Lessac A. *The use and training of the human voice.* California: Mayfield, 1997.

6. Linklater K. *Freeing the natural voice.* New York: Drama Book Specialists, 1976.

7. McCallion M. *The voice book.* New York: Routledge, 1988.

8. Raphael B. The sounds of violence: Vocal training in stage combat. *Theatre Topics* 1991; 1(1);73-86.

9. Rodenberg P. *The right to speak.* London: Methuen, 1992.

10. Scherer RC, Titze IR, Raphael BN, Wood RP, Ramig LA, Blager RF. Vocal fatigue in a trained and an untrained voice user. In: Baer T, Sasaki C, Harris K, eds. *Laryngeal function in phonation and respiration.* San Diego: Singular, 1991.

11. Sherman D, Jensen PJ. Harshness and oral-reading time. *J Speech Hear Dis* 1962; 27: 172-177.

12. Stemple JC, Stanley J, Lee L. Objective measures of voice production in normal subjects following prolonged voice use. *J Voice* 1995; 2: 127-133.

13. Stone RE, Sharf DJ. Vocal change associated with the use of atypical pitch and intensity levels. *Folia Phoniatr* 1973; 25: 91-103.

14. Sander EK, Ripich DN (1981). Vocal consequences of loud talking. *Paper presented at the American Speech-Language-Hearing Association Convention,* 1981.

15. Reimers Neils L, Yairi E. Effects of speaking in noise on vocal fatigue and recovery. *ASHA,* 1984; 24: 182 (A).

16. Gelfer MP, Andrews ML, Schmidt CP. Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers. *J Voice* 1991; 5: 158-167.

Appendix 1

Script used for pre-testing warm-up. All exercises were demonstrated by the first author.

Easy Warm-up (10 minutes)

(i) Shake out and loosen limbs, torso;

(ii) Massage neck muscles, cupping throat, then down into shoulders, up into skull base and scalp;

(iii) Isometrics: lace fingers together, cup hands at base of skull, press head into hands, hand into head; repeat procedure with palm to forehead, palm to each temple. In each situation, register the sensation of release as you cease pressure;

(iii) Head and neck relationship: drop head, roll to one side, add weight of hand, remove 'and float head up, repeat on other side; register sensation of release;

(iv) Shoulders: release up and down, side to side with breath engaged; loosen shoulders with diagonal arm swings;

(v) Neck lengthened, paint broad strokes in the air with head, paint broad strokes with arms;

(vi) Facial muscles: hands on face, massage flesh, blow lips loosely on /p/, massage jaw "hinges," "chew" and hold jaw, then release and register sensation of release as you cease pressure; loosen and stretch tongue;

(vii) Yawn and open-throated breathing;

(viii) "Murmur" and "slide" (7) (p.75). With lips closed as for /m/, tongue position if prepared to say /a/. Without disturbing breathing, use out-breath to make voice on descending note from middle of comfortable range, direct resonance fully into nose and closed mouth, keep back of mouth well open (jaw free) and with each outbreath, start slide a little higher in range, encourage vibrations to fill not only head, but chest, then open lips to /u/ /i/. Attend to self—any body part that still needs further attention.

Appendix 2

Script used for evoking vocally violent efforts during flexible laryngovideostroboscopy (FLVS).

Preliminary Instructions

Lean against the wall, support your upper back, head, and neck against the wall, with your hips and lower back away from the wall and your feet braced and supporting your body. Hands can hang loosely or however you choose. All vocal efforts should be made at stage volume.

(1) Grunt. The first effort is grunting. The stage acting situation is this. You are the aggressor and with each grunt, you're delivering a blow with a heavy sword, to inflict pain and to stop your enemy. With each imagined blow, grunt on /hip/. When Nelson says "go" and until he says "stop," grunt on /hip/ with that situation in mind. We'll do this several times. Any questions? Are you ready?

(2) Groan. The next effort is groaning. The stage acting situation is this. You have been fighting and are injured. You're in physical pain, but have to get your last words out before dying. These words might be something like, "This will be avenged." Groan your thoughts on /hip. When Nelson says "go" and until he says "stop," groan on /hip/ with that situation in mind. Questions? Ready?

(3) Sob. The next effort is sobbing. The stage acting situation is this. You've just lost your only child, who just drowned. Sob your grief on /hip/. When Nelson says "go" and until he says "stop," sob on /hip/ with that situation in mind. Questions? Ready?

(4) Shout. The last effort is shouting. The stage acting situation is this. You have led your troops to the wall of the fortress, and are ready for attack. You are standing below, and must summon your opponents from the fortress. Dare them to come out and fight. Shout your challenges on /hip. When Nelson says "go" and until he says "stop," shout on /hip/. Questions? Ready?

Note

Subjects were reinstructed when they did not appear to understand the requirements and were also reminded periodically to produce the efforts at stage volume. The goal was to achieve at least four productions of the target vocal maneuver that were visualized adequately with the endoscope.

Appendix 3

Script used for posttesting warm-down.

Warm-down (3 to 5 minutes)

(i) Massage neck muscles, cupping throat, then down into shoulders, up into skull base and scalp;

(ii) Isometrics (as above in warm-up)

(iii) Shoulder rotations, elbow point rotations, paint broad strokes in the air;

(iv) Yawn

(v) "Murmur" (as above)

(vi) Gentle humming, then produce quiet /u-i-u-i-u-/ on progressive pitches up and down a three- or four-tone scale.

NCVS Status and Progress Report - 12
June 1998, 175-179

Manual Circumlaryngeal Techniques in the Assessment and Treatment of Voice Disorders

Nelson Roy, Ph.D.
Department of Communication Disorders, Minot State University
Diane M. Bless, Ph.D.
Department of Communicative Disorders, The University of Wisconsin-Madison
Department of Otolaryngology- Head & Neck Surgery, The University of Wisconsin-Madison

Abstract

Poorly regulated activity of the intrinsic and extrinsic laryngeal muscles affects phonatory function and contributes to a class of disorders referred to as hyperfunctional or musculoskeletal tension voice disorders. Understanding the contribution of excessive or dysregulated laryngeal muscle activity is critical to proper diagnosis and selection of appropriate treatment(s). Manual circumlaryngeal techniques have recently received attention as useful tools in the assessment and treatment of such muscle tension voice disorders. This review will describe these manual techniques, highlight salient procedural aspects and survey the empirical literature evaluating their effectiveness.

Introduction

Without exception contemporary voice manuals cite excessive or poorly regulated activity of the intrinsic and extrinsic laryngeal muscles as significant causal factors in a variety of voice disorders [1--6]. Understanding the contribution of laryngeal and extralaryngeal muscle dysregulation is therefore critical to proper diagnosis and selection of appropriate treatments. In this regard, manual circumlaryngeal techniques have recently received attention in the clinical voice literature as potentially valuable diagnostic and treatment tools. This review article will describe these manual laryngeal techniques, highlight important procedural considerations and survey the literature which evaluates their effectiveness.

"Imbalanced" laryngeal and paralaryngeal muscle activity is the common denominator behind a class of voice disorders referred to as hyperfunctional or musculoskeletal tension voice disorders [7]. The origin of this excess muscle activity is not fully understood, but has been attributed to a variety of sources, including 1) psychological and/or personality factors which tend to induce elevated tension in the laryngeal region [1,8], 2) technical misuses of the vocal mechanism in the context of extraordinary voice demands [9--12], 3) learned adaptations following upper respiratory tract infection [13], 4) increased pharyngolaryngeal tone secondary to the laryngopharyngeal reflux reflex [14], and 5) compensation for underlying mucosal disease and/or glottic insufficiency [15]. Regardless of the presumed cause of the muscle dysregulation, most clinicians agree that restoring laryngeal muscular balance is an important therapeutic goal. Consequently, many contemporary voice therapies including progressive relaxation, Accent method, yawn-sigh, and resonant voice therapy aim to reduce excess laryngeal musculoskeletal tension, assuming that relief of tension in this region will yield a proportional improvement in voice function.

Aronson [1] speculated that therapy failure for musculoskeletal tension voice disorders may be due, at least in part, to techniques that do not yield sufficient laryngeal tension reduction. He offered that indirect (i.e., nonmanual) tension reduction techniques often fail because of the stubborn nature of excess laryngeal musculoskeletal tension. Aronson suggested that chronic posturing of the larynx in an elevated position leads to cramping and stiffness of the hyoid-laryngeal musculature and concomitant voice mutation. He described the manual laryngeal musculoskeletal tension reduction technique (also known as circumlaryngeal massage) as a direct method to assess and treat laryngeal hyperfunction syndromes. Manual repositioning (i.e., lowering) of the larynx by kneading the circumlaryngeal area was advocated as the primary treatment approach to relieve

muscle tension. Furthermore, Aronson argued that "all patients with voice disorders, regardless of etiology, should be tested for excess musculoskeletal tension, either as a primary or as a secondary cause of the dysphonia (p. 314)." Our clinical and research experience with these and related manual techniques support their clinical utility as valuable assessment and management tools.

Signs and Symptoms of Elevated Laryngeal Musculoskeletal Tension

Voice quality symptoms associated with musculoskeletal tension voice disorders can vary in severity and type, ranging from severely pressed to extreme breathiness with myriad combinations. Patients are pitch and loudness "locked," often displaying a markedly reduced dynamic range. Complete phonatory disintegration is observed early during upward pitch glides.

In cases where excess laryngeal tension has persisted for some time, patients commonly report a dull to severe ache and tightness of the anterior neck, larynx, and shoulder regions which is accompanied by increased vocal effort and fatigue, episodic anterior neck "swellings/lumps" and ear "fullness", with all symptoms intensifying with extended voice use. According to Morrison [14] the inferior bellies of the omohyoid muscles where they cross the supraclavicular fossae, are often tense and prominent during speech. General body posture may be rigid with the jaw jutting forward. Jaw, tongue and respiratory movements are restricted, reflecting the "held" nature of the voice and articulatory system.

During laryngoscopic examination certain glottic and supraglottic contraction patterns may be suggestive of excess muscle tension, however we argue that no particular voice quality or glottic/supraglottic configuration should be *definitively* identified with musculoskeletal tension voice disorders [4, 12]. Given the potential involvement of a variety of intrinsic and extrinsic laryngeal muscles -- in diverse states of relaxation and contraction -- numerous laryngeal configurations are possible. From a larygnoscopic view, what characterizes musculoskeletal tension voice disorders is dysregulation of agonist-anatagonist muscle relationships.

Focal Palpation of the Laryngeal Region

Before proceeding with manual circumlaryngeal assessment and treatment techniques, patients are educated regarding the negative effects of excessive musculoskeletal tension on voice, and the possibility that such tensions may underlie some or all of the patient's voice disorder.

At rest, musculoskeletal tension can be appraised manually by palpation of the laryngeal area to assess the degree, nature, and location of focal tenderness or muscle nodularity and/or pain. Care should be taken to avoid sustained carotid artery compression during these maneuvers. With the occiput gently supported in a neutral position, pressure is directed 1) over the major horns of the hyoid bone, 2) over the superior cornu of the thyroid cartilage, 3) within the thyrohyoid space, and 4) along the anterior border of the sternocleidomastoid muscle. During palpation, the degree of compression applied is equal to the pressure required to cause the thumbnail tip to blanch when pressed against a firm surface. When this amount of pressure is used, focal sites of tension evoke discomfort or pain. Patients may wince, withdraw (the "jump" sign) or vocalize their discomfort when trigger points are identified. The discomfort is more often unilateral than bilateral, and may radiate to one or both ears. This exquisite tenderness in response to pressure in the laryngeal region is considered abnormal and is a prominent feature of musculoskeletal tension voice disorders.

The extent of laryngeal elevation is examined by palpating within the thyrohyoid space from the posterior border of the hyoid bone to the thyroid notch. A narrowed or absent thyrohyoid space is highly suggestive of excess muscle tension whereby the larynx is suspended high in the neck. When asked to sustain a vowel at a comfortable pitch and loudness, further laryngeal elevation may occur in synchrony with voice initiation, signaling recruitment of extrinsic muscles. In some cases of muscle misuse, the larynx abnormally contracts during voicing, but returns to normal during rest. This phasic contraction pattern is in contrast with a tonic pattern where the larynx is held tightly at all times [14].

Excessive tension can also be detected in the medial suprahyoid and submental regions by palpating at rest, during voice produced at modal pitch, and then during upward pitch glide maneuvers. Taut bands and disproportionate laryngeal elevation signal excessive muscle recruitment/activity. The mobility of the larynx is also tested by attempting to maneuver it side-to-side along the horizontal plane. Resistance of lateral movement is indicative of generalized extralaryngeal hypertonicity.

Morrison (14) advocates additional manual assessment techniques to further assess restricted laryngeal mobility due to elevated laryngeal tension. He suggests palpating within the cricothyroid space to contrast the findings obtained at rest with the results obtained during upward and downward pitch glides. An absent space at rest and during these pitch maneuvers is considered atypical and undesirable. Harris and Lieberman (as reported in Morrison) proposed that fatigued cricothyroid muscles will allow the cricoid cartilage to slip forward and partly sublux the cricothyroid joint. This may be appreciated by relating the inferior midline edge of the thryoid cartilage to the cricoid arch. Finally, inferior pharyngeal constrictor tightness and

freedom of the lateral laryngeal gutters can be inferred by the facility with which the larynx can be rotated and the fingertips slid behind the posterior edge of the thyroid lamina. If the larynx is tightly held against the vertebral column by the inferior constrictors, then the lateral laryngeal gutters are not accessible. According to Morrison, if this area is sufficiently relaxed, then the movement of the arytenoid cartilage can be felt and posterior cricoarytenoid muscle contraction can be palpated during an inspiratory sniff.

Laryngeal Reposturing Maneuvers

During voicing, important information regarding inappropriate laryngeal posturing and muscle misuse can be derived by assessing the voice effects of several manipulations: 1) physically impeding elevation of the larynx by applying downward traction over the superior border of the thyroid lamina; 2) compressing the larynx by exerting anterior-to-posterior (A-P) pressure over the inferior aspect of the hyoid bone, and/or within the thyrohyoid space to reduce excess tension and stiffness of the vocal folds; and 3) applying combined medial compression and downward traction over the superior cornu of the thyroid cartilage.

Laryngeal reposturing or repositioning through brief displacement or by resisting laryngeal elevation provides valuable information regarding potential for improved voice function as well as possible causal mechanisms. By manually repositioning or stabilizing the larynx during the production of a sustained "ah", the clinician may stimulate improved voice and briefly interrupt patterns of muscle misuse. These brief moments of voice improvement are immediately identified for the patient and reinforced. They can be shaped using digital cueing or combined with tension-reduction techniques (to be discussed next). Digital cues can then be faded and the patient taught to rely on sensory feedback (auditory, kinesthetic, and proprioceptive) to maintain improved laryngeal posturing and muscle balance. Once the larynx is correctly positioned, recovery of normal voice can occur rapidly.

Manual Circumlaryngeal Massage

Skillfully applied, systematic kneading of the extralaryngeal region is believed to stretch muscle tissue and fascia, promote local circulation with removal of metabolic wastes, relax tense muscles and relieve pain and discomfort associated with muscle spasms [16]. The hypothesized physical effect of such massage is reduced laryngeal height and stiffness and increased mobility. Once the larynx is "released", and range of motion is normalized, an improvement in vocal effort, quality, and dynamic range should follow. Focal palpation and massage helps patients to become more aware of where they are holding tension. By becoming aware of these laryngeal troublespots the patient can begin to focus on relaxing them during self-massage, which can be undertaken on a daily basis.

Once the assessment procedures have been completed, the manual tension reduction technique can be undertaken according to the description of Aronson [1]. The hyoid bone is encircled with the thumb and index finger, which are then worked posteriorly into the tips of the major horns of the hyoid bone. Pressure is applied in a circular motion over the tips of the hyoid bone. The procedure is repeated within the thyrohyoid space, beginning from the thyroid notch and working posteriorly. The posterior borders of the thyroid cartilage medial to the sternocleidomastoid muscles are located and the procedure is repeated there. With the fingers over the superior border of the thyroid cartilage, the larynx is pulled downward and, at times, moved laterally. Sites of focal tenderness, nodularity, or tautness are given more attention. Gentle kneading or sustained pressure may be focused over these sites and then released. In general, the procedure is begun superficially, and the depth of massage is increased according to the degree of tension encountered and the tolerance of the patient. It may be necessary to begin the technique peripheral to the sites of intense tenderness, then gradually direct attention proximal to these sites. When excess tension is encountered, the clinician can extend the technique into the medial and lateral suprahyoid musculature. The immediate effects of massage are noticeable on the skin. Friction and circular stroking movements heighten blood circulation which is accompanied by a slight reddening and warming of the skin.

During the above procedures, the patient is asked to sustain vowels or to hum while both clinician and patient note changes in vocal quality. The patient is an active participant and encouraged to continually self-monitor the type and manner of voice produced. Some discomfort during the procedure may be unavoidable, however the clinician's goal is to achieve sufficient tension reduction without inducing reactive/reflexive muscle tension due to pain. Improvement in voice and reductions in pain and laryngeal height suggest a relief of tension. The improved voice is extended from vowels to words (usually automatic serial speech, i.e., counting, days of the week), to short phrases (e.g. "many men in the Moon," "one Monday morning"), to sentences, and paragraph recitations, and then to conversation.

Progress can plateau at any point in the treatment sequence. Improvements with these techniques often proceed in a nonlinear fashion. Once sufficient tension has been released and the patient assumes a more normal laryngeal posture, progress can be swift, with complete amelioration of the dysphonia. Signs of improvement should be observed within the first session. Some patients may require an extended, intensive treatment session or several sessions,

depending on patient tolerance and rate of progress. Patients may report focal laryngeal tenderness persisting up to 48 hours following intense circumlaryngeal massage. Generally, if changes do not occur within two treatment sessions, it is unlikely that extralaryngeal muscle tension is the primary or sole explanation for the observed dysphonia. In cases of benign mucosal disease, the residual dysphonia following circumlaryngeal massage may more accurately reflect the untoward effects of the vocal fold mucosal changes on voice production.

Review of Literature

In a series of articles, Roy and colleagues have evaluated the clinical utility of manual techniques with a variety of functional voice disorders [17--20]. Recently Roy, Bless, Heisey and Ford [17] reported the immediate and long-term effects of manual circumlaryngeal therapy for twenty-five female patients with functional dysphonia. Functional dysphonia was defined as a voice disturbance in the absence of structural or neurological pathology. Each patient received a single treatment using the manual laryngeal musculoskeletal tension reduction procedure as described by Aronson [1]. Perceptual, acoustic and interview techniques were used to assess vocal function before and after treatment. Subjects demonstrated consistent improvement across perceptual and acoustic indices of vocal function immediately after treatment and during the follow-up period. Based on perceptual ratings, ninety-six percent of patients were rated as improved, with almost two-thirds of all patients achieving normal voice return following the single treatment session.

Although subjects demonstrated consistent improvement across perceptual and acoustic indices of vocal function, the results of data gathered during follow-up interviews suggested that the 68% of patients who initially responded to manual therapy reported infrequent, partial and self-limiting recurrences early in the follow-up phase (i.e., less than 2 months post-treatment). These authors suggested that for some patients, superior results may be found when manual laryngeal techniques are combined with supportive counseling and/or more frequent clinical support.

The authors also admitted that it would be presumptuous to attribute changes in voice exclusively to a reduction in laryngeal muscle tension, however they did provide some indirect evidence to bolster Aronson's contentions concerning the relationship between voice improvement and laryngeal tension reduction. Prior to treatment, 88% of subjects reported pain and/or tenderness in the laryngeal region during palpation. During and immediately following the manual procedure, the majority of subjects reported gradual reduction and then amelioration of laryngeal pain. This finding is consistent with Aronson's clinical descriptions (Aronson [1], p.314). Furthermore, during the treatment phase, subjects typically progressed through stages of decreasing dysphonia until symptoms gradually remitted, which is also consistent with Aronson's accounts (p.315). Whether these findings represent a steady reduction in laryngeal tension, as Aronson maintains, remains open for debate.

Given the potential for musculoskeletal tension disorders to coexist with and mimic neurological spasmodic dysphonia, Roy, Ford & Bless [18] described using manual laryngeal techniques to assist in distinguishing muscle tension dysphonia from spasmodic dysphonia. The authors suggest that during palpation, the presence of laryngeal pain and tenderness, stiffness and elevation should raise the index suspicion for the existence of excessive musculoskeletal tension. The influence of abnormal muscular tensions (i.e., nondystonic) can be verified by the patient's positive response to diagnostic therapy using laryngeal reposturing techniques including circumlaryngeal massage. By identifying neurological SD patients who also present with generalized laryngeal hypertonicity, the authors proposed that clinicians could also interpret suboptimal responses to botulinum toxin treatments and prospectively advise a therapy regimen combining manual laryngeal tension reduction with botulinum injections.

In many cases excessive extralaryngeal tension superimposed on underlying vocal fold pathology can complicate the clinical presentation. Tasko, Roy and Harvey [20] presented data confirming the usefulness of manual circumlaryngeal techniques in the assessment and treatment of a variety of benign mucosal disorders, including vocal fold polyps and nodules. Over a single treatment session, statistically significant improvements in (a) listener ratings of severity of dysphonic symptoms and (b) jitter, shimmer and signal-to-noise ratio values were observed. Although as a group, the patients did not achieve perceptually normal voice, the results clearly indicated that the patients' dysphonic symptoms were not due wholly to the deleterious effects of the additive lesions on vocal fold vibration. The authors concluded that by removing the effects of imbalanced muscular forces, the clinician is afforded a more precise (i.e., purer) representation of the voice disorder. Only by defining the contribution of generalized hyperfunction to the perceived dysphonia can optimal management of these disorders be accomplished.

Conclusions

Manual techniques, including circumlaryngeal massage, augment the voice practitioner's diagnostic and treatment armamentarium. It is exceedingly important for voice diagnosticians to appreciate the effects of poorly regulated laryngeal and paralaryngeal tension. Such recognition can avoid unnecessary medical or surgical interven-

tion. These manual techniques can assist in identifying and recognizing the contribution of laryngeal and extralaryngeal hypertonicity to the perceived voice disturbance.

Acknowledgments

This work was supported in part by the National Center for Voice and Speech through Grant P60 00976 from the National Institute on Deafness and Other Communication Disorders.

References

1. Aronson AE: **Clinical Voice Disorders: An Interdisciplinary Approach.** ed. 3. New York, NY: Thieme, 1990.

2. Boone DR, McFarlane SC: **The voice and voice therapy**, ed. 5, Englewood Cliffs, NJ: Prentice Hall, 1994.

3. Colton RH, Casper JK: **Understanding voice problems: A physiological perspective for diagnosis and treatment** ed. 2. Baltimore: Williams & Wilkins, 1996.

4. Morrison MD, Rammage LA: **The management of voice disorders.** San Diego: Singular Publishing Group, 1994.

5. Stemple JC: **Voice Therapy: Clinical Studies.** St. Louis: Mosby Year Book, 1993.

6. Titze IR: **Principles of Voice Production.** Englewood Cliffs, NJ: Prentice Hall, 1994.

7. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughn C: **Objective assessment of vocal hyperfunction: an experimental framework and initial results.** *J Speech Hear Res* 1989, 32:373--392.

8. Rammage LA, Nichol H, Morrison MD: **The psychopathology of voice disorders.** *Hum Commun Canada* 1987, 11:21--25.

9. Koufman JA, Blalock PD: **Vocal fatigue and dysphonia in the professional voice user: Bogart-Bacall syndrome.** *Laryngoscope* 1988, 98:493--499.

10. Morrison MD, Nichol H, Rammage LA: **Diagnostic criteria in functional dysphonia.** Laryngoscope 1986, 94:1-8.

11. Morrison MD, Rammage LA, Gilles MB, Pullan CB, Hamish N: **Muscular tension dysphonia.** J Otolaryngol 1983, 12:302--306.

*12. Morrison MD, Rammage L: **Muscle misuse voice disorders: description and classification.** *Acta Otolaryngol (Stockh)* 1993, 113:428--434.

13. Koufman JA, Blalock PD: **Classification and approach to patients with functional voice disorders.** Ann Otol Rhinol Laryngol 1982, 91:372--377.

14. Morrison MD: **Pattern recognition in muscle misuse voice disorders: How I do it.** *J Voice* 1997, 11:108--114.

15. Roy N: **Ventricular dysphonia following long-term endotracheal intubation: A case study.** *J Otolaryngol* 1994, 23:189--193.

16. Peifang C: **Massage for the treatment of voice ailments.** *J Trad Chinese Med* 1991, 11:209--215.

*17. Roy N, Bless DM, Heisey D, Ford CN: **Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short- and long-term treatment outcomes.** *J Voice* 1997, 11:321--331.

18. Roy N, Ford CN, Bless DM: **Muscle tension dysphonia and spasmodic dysphonia: The role of manual laryngeal tension reduction in diagnosis and treatment.** *Ann Otol Rhino Laryngol* 1996, 105:851-856.

19. Roy N, Leeper HA: **Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: Perceptual and acoustic measures.** *J Voice* 1993, 7:242--249.

20. Tasko SM, Roy N, Harvey S: **Using manual laryngeal tension reduction techniques with benign mucosal disorders.** Paper presented at the convention of the American Speech-Language-Hearing Association, New Orleans, LA, 1994.

NCVS Status and Progress Report - 12
June 1998, 181-185

Posterior Wall Augmentation for Treatment of Velopharyngeal Insufficiency

Steven D. Gray, M.D.
Department of Surgery - Division of Otolaryngology, University of Utah School of Medicine
Judith Pinborough-Zimmerman, Ph.D.
Utah State Department of Health
Michael Catten, M.D.
The University of California San Francisco

Abstract

Velopharyngeal insufficiency can be treated surgically with various operations. This paper describes using a superiorly based folded pharyngeal flap for posterior wall augmentation to treat velopharyngeal insufficiency. This is a retrospective study which indicates that a folded flap to augment the posterior wall is likely as effective as other surgical techniques to treat small velopharyngeal gaps. Patients selected for this procedure had very good velar motion. Postoperative nasometry scores improved by an average of 18 over preoperative scores. Additionally, a correlation was found between age and nasometry improvement following surgery. Younger patients did better. Patients who developed velopharyngeal insufficiency due to adenoidectomy did well. Syndromic patients did not do as well when treated with this type of operation.

Introduction

Velopharyngeal inadequacy refers to the inability to separate the oral cavity from the nasal cavity during certain speech tasks, blowing and swallowing. In normal speech production, most air and sound is directed through the mouth rather than the nose. Separation of the nasal and oral cavity occurs from a combination of muscles in the velopharyngeal region. This may consist of elevation superiorly and posteriorly of the velum (soft palate), medial motion of the lateral pharyngeal walls, and occasionally some anterior motion of the posterior pharyngeal wall[1,2]. This should result in complete channeling of air flow through the mouth during most speech tasks. Velopharyngeal closure is expected on vowels and consonants with the exception of "m", "n" and "ng". Hypernasality is a term used to describe the perception of excessive nasal resonance during speech. Velopharyngeal insufficiency may occur from congenital palatal abnormalities, adenoidectomy, and other conditions causing structural problems of the velopharyngeal port.

Many operations have been designed to improve velopharyngeal closure. The two most common operations are the pharyngeal flap and the pharyngoplasty. Pharyngeal flaps have an extensive history in the repair of velopharyngeal insufficiency and have been proven effective[3,4]. Undesirable consequences of the pharyngeal flap surgery have been related to an obstructed velopharyngeal port and include apnea, nasal obstruction and hyponasality[3,5,6]. These occasional problems partially led to the popularity of the sphincter pharyngoplasty[7,8,9]. In the last two decades, this procedure has become increasingly popular as it appears that sleep apnea, hyponasality and nasal obstruction are less common[10,11]. As one would expect, the more obturation or obstruction that is required for treatment of the velopharyngeal insufficiency, the more likely sleep apnea, nasal obstruction and hyponasality may occur. In patients where minor obturation is required perhaps a procedure which creates minor obturation is desirable.

Surgeries designed to augment the posterior pharyngeal wall have been selected mainly for those cases where minor obturation is required. Surgeries designed to augment the posterior pharyngeal wall have been proposed throughout the century and are often advocated because of

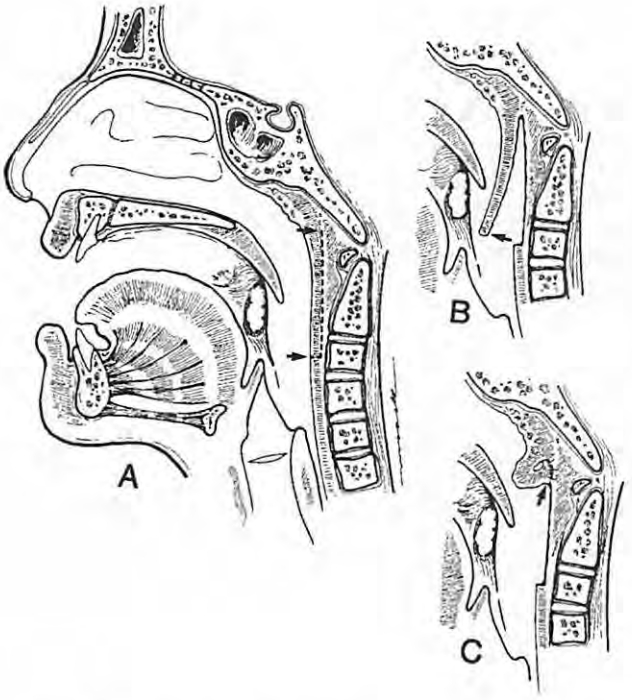


Figure 1. The relatively simple but effective procedure to augment the posterior pharyngeal wall is demonstrated. The section of the posterior pharyngeal wall to be elevated is seen in A, and the inferior level is shown to be at about the level of the inferior pole of the tonsil. Note that in B, the constrictor muscle is raised with the flap and is elevated quite superiorly. In C, the flap has been folded and sutured into place. Even though the flap was elevated higher, the fold and the maximal area of augmentation occur a little lower. Although not shown in C, sutures are often placed on the folded flap laterally to close some of the dead space and to create more stability.

their ease and simplicity[12,13,14,15,16,17]. Unfortunately, many of these surgeries have lost favor because they failed to meet expectations with respect to surgical results. Materials used to augment the posterior pharyngeal wall have ranged from cartilage, fat and bio-implantable materials[13,14,15,17]. The superiorly based pharyngeal flap, which is buckled or folded upon itself had the location of the velopharyngeal gap has gained interest. This study is an evaluation of the selection criteria used for posterior wall augmentation and the results obtained from this operation.

Methodology

A retrospective study was performed on patients who had received posterior wall augmentation by folding or buckling a superiorly based pharyngeal flap. Entry criteria were based upon those who had received the operation and at least 2 years had passed since the operation. Patients had received preoperative, and in most cases a postoperative nasometry score and speech assessment. The postoperative nasometry was performed at least 3 months following the

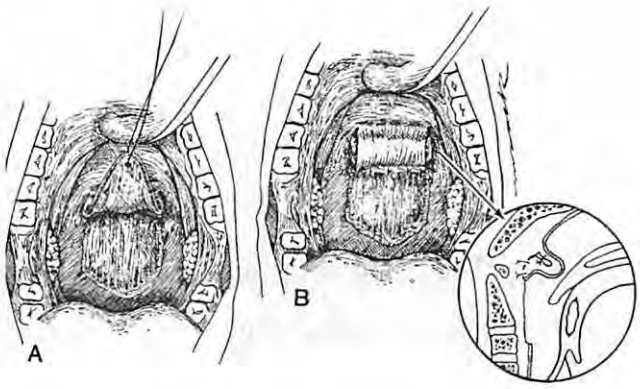


Figure 2. A) The superiorly based pharyngeal flap being elevated from the posterior pharyngeal wall. B) The fold in the flap which is then sutured back to the pharyngeal wall.

operation. Criteria evaluated were age of the operation, nasometry scores, preoperative velopharyngeal closure patterns and ratings[18], postoperative nasometry and speech assessment. Nasometry was obtained using the Zoo passage. Speech assessments were performed by the speech pathologists caring for craniofacial patients in their state (Utah, Idaho, Montana and Wyoming). These were speech pathologists with extensive clinical experience in velopharyngeal disorders. The evaluations were done by a single listener with knowledge that surgery was performed. Since the study is retrospective, the speech pathologist was not aware that their ratings would be eventually used for a study. The same speech pathologist did the ratings in each patient. Speech assessments included a hypernasality rating by the listener on a scale of 0-5[10]. Zero is normal nasality, 5 is severe hypernasality.

Analysis of the nasometry data was performed by using the paired t-test since the data was interval. The original data from the speech assessment was analyzed using the Wilcoxon Signed Ranks Test. A regression analysis was performed to correlate the preoperative and postoperative nasometry change with the variable of age.

The procedure was performed by raising a superiorly based posterior pharyngeal wall flap which includes the mucosa and the constrictor muscle. The flap is raised up to, and slightly above, the level of the velopharyngeal defect. The flap is then folded such that the folded bulge is at the appropriate level and fills in the defect present in the velopharyngeal closure. This site of flap placement is determined based on the preoperative nasoendoscopy. See Figure 1 and Figure 2. This procedure is described in other publications[19].

Table 1.

Age	Condition	Nasometry*		Hypernasality, listener rating (0-5)+		Shape of remaining gap++	Velopharyngeal Movement Rating	Posterior Wall (Irregular Surface)	Previous Surgeries
		Preoperative	Postoperative	Preoperative	Postoperative				
8	post-adenoid VPI	41	16.8	2	0	circular, intermittent touch-closure	9/1/1/0	Yes	T & A
12	post-adenoid VPI, occult submucous cleft	not obtained	not obtained	1	0	tiny, circular gap	9/4/5/0	No	adenoidectomy
4	cleft-palate	38.5	12.5	2	0	tiny, central slit	9/0/0/0	No	cleft-palate repair
5	submucous cleft	38.4	19.4	3	0	circular, central gap	8/4/4/0	No	none
14	post adenoidectomy, VPI	not obtained	not obtained	3	0	circular, touch-closure, bubbles only	9/4/4/0	No	T & A
10	none	27	22	1	0	coronal slit, tiny gap	9/2/2/0	No	none
5	post-adenoid	49	19	4	0	sagittal square, asymmetric	6/2/6/0	No	T & A
6	post-adenoid	56	22	5	0	Coronal, slight touch-closure, occasional competence	9/3/3/0	Yes	T & A
11	Microcephaly, ataxia, motor delay, severe SN hearing loss,	44	43	4	2	coronal gap, more open laterally, closure centrally	9/1/1/0	Yes	T & A, previous sphincter pharyngoplasty
15	cleft-palate	50		4	0	tiny gap, inconsistent touch-closure, bubbles	9/1/1/0	No	cleft-palate repair
7	submucous cleft	73	35	5	0	small, central gap	9/4/4/0	No	none
13	post adenoid VPI, Sotos Syndrome	47	39	2	1	tiny, coronal gap, no visible gap, bubbles	9/2/2/0	No	T & A
4	post adenoidectomy, VPI, submucous cleft	44	never returned for follow-up	3	never returned for follow-up	no gap, touch-closure, bubbles	9/2/2/0	Yes	T & A
8	none	40	43**	4	4	touch-closure, no gap, just bubbles	9/2/2/0	Yes	none

** Revision operation obtained normal speech, see text
* Zoo Passage
++ Anterior wall rated first, followed by right pharyngeal wall, left pharyngeal wall and posterior wall
+ Hypernasality scored 0-5, 0 is normal nasal resonance, 5 is severe hypernasality

Results

The intention of this study is to be a descriptive study as to the patient selection and their results. Data for this study is found in table 1. As it can be seen, that patients selected for this operation were those with very minimal velopharyngeal gaps. With the exception of two patients, all patients had an anterior wall rating of 9 out of a possible 10. In most cases, the anterior pharyngeal wall touched the posterior pharyngeal wall, but did not seal it completely to pressurized air. Patients with large velopharyngeal gaps were not selected for this operation. Nasometry scores normalized after the surgical procedure in 6 out of 10 patients. The mean preoperative nasometry score was 45 and the mean postoperative nasometry score was 27. The mean change was 18. See table 2.

Hypernasality listener ratings performed by a speech pathologist were normal following the operation in 10 out of 13 patients. The median change in score was 3. See table 2. There were three patients who failed to improve their speech to normal postoperatively. One patient (last on table 1) was clearly a technical error in the operation as the folded flap was above the area of velar closure. This was surgically revised and the postoperative result from the revision surgery has a nasometry score of 25 and is assessed as having normal speech. The other two patients both had syndromes, Sotos Syndrome and an unnamed syndrome of microcephaly, poor motor tone and sensorineural hearing loss. Both patients are currently judged as still being hypernasal, although both patients and their parents feel their speech is normal enough and wish no further therapy.

The shape of the preoperative gap, such as coronal or circular, did not seem to be a factor in predicting success or failure. It is interesting to note that one of the failures had a closure pattern which resulted in openings at the lateral pharynx and a touch-closure pattern centrally. This had the shape of a figure eight laid sideways. Indeed, postoperatively this patient still leaked in the lateral pharynx. No other patients in this study had a similar closure pattern. No occasions of sleep apnea, nasal obstruction or hyponasality were present. We specifically tested for hyponasality using the nasometer and nasal consonants and speech assessment. Nasometry scores were normal for nasal consonants in all of those who had postoperative nasometry.

A regression analysis showed a significant correlation between age and the reduction in nasometry scores following surgery. The correlation may be confounded because the two patients who were syndromic were both older. Despite that, it does appear that there remains a

Table 2.

Nasometry (pre/post change)
N=10
Mean Change + SE (Standard Error)
(-18.22 + 4.60)
Paired t-test P=0.003

Mean Pre Score + SE
(45.39 + 3.91)

Mean Post Score + Standard Error
(27.17 + 3.66)

Hypernasality
N=12
Median Change -3
Wilcoxon Signed Ranks Test
P=0.005

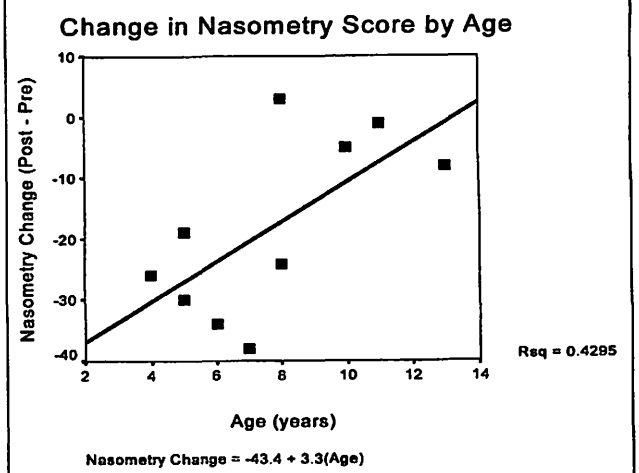
Regression Analysis:
Nasometry Change (Post-Pre) = 43.4 + 3.3*(Age)
R^2=0.43
P=0.04

correlation between age of surgical correction and improvement in nasometry scores. Younger patients did better following surgery. Older patients did not tend to experience the same degree of improvement.

Discussion

It appears that this descriptive study supports the use of autogenous posterior wall augmentation in select patients. This study does not define the patients who should be selected for this operation, as such a study design was not performed. This study does indicate who we selected for the operation and would support the use of posterior wall augmentation in those patients with very small velopharyngeal gaps. It is interesting to note that this population often includes those that have developed VPI following an adenoidectomy. Seven of the 14 patients in this study had post-adenoidectomy VPI. Some of these patients had no palatal defect associated with their post-adenoidectomy VPI. Retrospectively, we were curious to see how many of these had an irregular surface of the posterior pharyngeal wall as a cause of their VPI according to Ren, Isberg and Henningsson's experience[20]. In this group of patients, some patients did have an irregular surface, but we did not encounter any that had a large lump of adenoids which we felt would be responsible for lack of velopharyngeal closure. Based on this retrospective study, augmentation of the posterior wall is a good procedure for correcting post-adenoidectomy VPI if the velopharyngeal gap size is very small.

Table 3.



It is noteable that two of the patients, who did not improve significantly, were syndromic. Both have some delayed motor skills and both had speech delay. In patients that are syndromic, a posterior pharyngeal wall augmentation may not be the ideal procedure. The patient who did not improve due to technical error deserves comment. This patient should of had a good outcome. Even though the surgery was performed with the nasoendoscopy tape in the room at the time of the operation, it is clear that we misjudged where the level of the flap was placed. The flap provides obturation of the velopharyngeal port only over a distance of perhaps 1 cm on a vertical plane. Therefore, this operation requires precision as to where the bulge is placed. As opposed to other pharyngeal flaps, which may migrate due to scar contracture, we have not found migration to be a major problem with these folded flaps. Some atrophy of the folded flap may occur and therefore the flap may shrink a little after placement, rarely considerable atrophy may occur and it behooves the surgeon to avoid compromising the blood flow to this flap as it is sutured in place.

Age as a variable in VPI correction is a topic not often written about, but it is occassionally verbally discussed. This study may offer support that correcting deficiencies in velopharyngeal closure earlier rather than later may lead to a better speech result. The younger patient's speech mechanisms may be more pliable and adaptable to the surgical change in the velopharyngeal structure. Perhaps the older patient's perception of normal speech becomes more difficult to change.

Caution regarding the findings of any retrospective study is appropriate. Additionally, the speech assessments may have been biased by the knowledge that the patient's had received surgical treatment. The paper reports our experience with folded pharyngeal flaps and therefore does not address which procedure is best, nor does it answer well

which patients should be selected for a particular procedure. Our speech results, biased as they may be due to this study design, are similar to those reported using other procedures for VPI correction[3,10]. The folded flap procedure is more simple in design, operative time is probably less, and less obstruction appears to occur than experienced in other procedures. In our practice, only 17% of the patients from 1991-1995 were treated with the folded flap. The majority of the procedures used for VPI correction remains a sphincter pharyngoplasty or pharyngeal flap, but there may be a role for the folded flap in selected patients.

Acknowledgment

Acknowledgment is given to Karen Myer, M.S. for biostatistical assistance.

References

1. Skolnick ML, McCall GN, Barnes M: The Sphincteric Mechanism of Velopharyngeal Closure. *Cleft Palate J* 1973;10:286.

2. Croft CB, Shprintzen RJ, Rakoff SJ: Patterns of Velopharyngeal Valving in Normal and Cleft Palate Subjects: A Multiview Videofluroscopic and Nasoendoscopic study. *Laryngoscope* 1985;91:265.

3. Morris HL, et al: Clinical Results of Pharyngeal Flap Surgery: The Iowa Experience. *Plast and Reconstructive Surg* 1995; 652-662.

4. Shprintzen RJ, et al: A Comprehensive Study of Pharyngeal Flap Surgery: Tailor-Made Flaps. *Cleft Palate J* 1979;16:46.

5. Valnicek SM, et al: Perioperative Complications of Superior Pharyngeal Flap Surgery in Children. *Plast and Reconstr Surg* 1994;93(5):954-958.

6. Orr WC, Levine NS, Buchanan RT: Effect of Cleft Palate Repair and Pharyngeal Flap Surgery on Upper Airway Obstruction During Sleep. *Plast and Reconst Surg* 1987;80:226.

7. Orticochea M: Construction of a Dynamic Muscle Sphincter in Cleft Palates. *Plast and Reconstruc Surg* 1968;41:323.

8. Jackson I: Discussion: A Review of 237 Cleft Palate Patients Treated with Dynamic Muscle Sphincter. *Plast and Reconstruct Surg* 1983;71:187.

9. Riski JE, et al: Evaluation of the Sphincter Pharyngoplasty. *Cleft Palate Craniofacial J* 1992;29(3):254-61.

10. Witt PD, et al: Sphincter Pharyngoplasty: A Preoperative and Postoperative Analysis of Perceptual Speech Characteristics and Endoscopic Studies of Velopharyngeal Function. *Plast and Reconst Surg* 1994;93:1155-1168.

11. Witt PD, et al: Management of the Hypodynamic Velopharynx. *Cleft Palate Craniofacial J* 1995;32(3):179-87.

12. Gersuny R: Ueber Eine Subcutane Prostheses. *Ztsch Heilk* 1900;21:199.

13. Eckstein H: Paraffin for Facial and Palatal Defects. *Dermatol Ztschr* (Basel) 1904; 11:772.

14. Blocksma R: Correction of Velopharyngeal Insufficiency by Silastic Pharyngeal Implant. *Plast and Reconst Surg* 1963;31:268.

15. VonGaza W: Transplanting of Free Fatty Tissue in the Retropharyngeal Area in Cases of Cleft Palate. Lecture - German Surgical Society, April 9th, 1926.

16. Ward PH: Uses of Injectable Teflon in Otolaryngology. *Arch Otolaryngol* 1968;97:637.

17. Smith JK, McCabe BF: Teflon Injection in the Nasopharynx to Improve Velopharyngeal Closure. *Ann Otolaryng Rhinol and Laryng* 1977;86:559.

18. Golding-Kushner KJ, et al: Standardization of the Reporting of Nasopharyngoscopy and Multiview Videofluroscopy: A Report from an International Working Group. *Cleft Palate J* 1990;27:337-347.

19. Gray SD, Pinborough-Zimmerman J: Diagnosis and Treatment of Velopharyngeal Incompetence. Facial Plastic Surgery Clinics of North America. Philadelphia (PA) 1996:405-12.

20. Ren YF, Isberg A, Henningsson G: Velopharyngeal Incompetence and Persistent Hypernasality after Adenoidectomy in Children without Palatal Defect. *The Cleft Palate Craniofacial J* 1995;3 2:476-482.

21. Skolnick ML, McCall GN, Barnes M: The Sphincteric Mechanism of Velopharyngeal Closure. *Cleft Palate J* 1973;10:286.

NCVS Status and Progress Report - 12
June 1998, 187-195

Generalized Motor Program and Parameterization Accuracy in Apraxia of Speech and Conduction Aphasia

Heather M. Clark, Ph.D.
Department of Language, Reading, and Exceptionalities, Appalachian State University
Donald. A. Robin, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

The present study examined three aspects of motor programming (generalized motor program (GMP) accuracy, temporal parameterization accuracy, and amplitude parameterization accuracy) in subjects with apraxia of speech (AOS) or conduction aphasia (CA) and normal speaking participants. Subjects were presented with a movement pattern on a monitor that they were required to produce with the jaw, after the target pattern had been removed from view. Analyses examined differences in relative (parameterization) and absolute (GMP) timing and amplitude between the target and actual movement. Examination of individual subject performance revealed inter-subject variability within the AOS group, with two of the four subjects demonstrating unimpaired GMP accuracy but poor parameterization accuracy, while the other two subjects exhibited the opposite pattern, impaired GMP accuracy but normal parameterization. No clear pattern of deficit was noted for the subjects with CA. Results are discussed with respect to motor control theories of AOS and CA.

Apraxia of speech (AOS) has been the subject of controversy for the past several decades. Although historically the debate focused on whether AOS was primarily linguistic or motoric in nature, more recently the focus of research has been on characterizing the nature of the underlying motoric impairment (e.g., McNeil, Weismer et al., 1990; Robin et al., 1989). A predominant conceptualization of AOS is that the behavioral characteristics stem from a "motor programming" impairment (Darley, 1969; Kent & Rosenbek, 1983). From this perspective, individuals with AOS have difficulty developing and/or executing motor programs that specify the temporal and/or spatial goals of the articulators for perceptually adequate speech. However, an operational definition of the "motor program" has been absent and the systematic examination of motor programming in subjects with AOS has been nonexistent.

Schmidt (1975, 1988) developed a motor programming model for learned movement patterns which includes the notion of *generalized motor programs* (GMP). GMPs are open-loop controlled movement structures which form the basis for related groups of actions. Schmidt operationally defined the GMP as the relative timing and/or forces of an action which can be examined by measuring the time or amplitude relations among kinematic landmarks of a movement pattern (e.g., Young & Schmidt, 1991). Schmidt defines *parameters* as the absolute timing and forces of actions that serve to scale GMPs for individual movement patterns. Because a single GMP may be activated for several related actions, a finite number of GMPs are necessary to complete an infinite number of potential actions, as different parameters are assigned for each variation of the movement pattern. For example, within this framework, a single GMP may be activated during all instances of a jaw-closing gesture. However, the speed and amplitude of the gesture is varied by assigning different parameters to the GMP for each specific action.

Within Schmidt's model, different aspects of motor programming might be disrupted in AOS. For example, individuals with AOS may have difficulty developing or executing the GMP. Alternatively, their impairment may stem from an inability to accurately parameterize an intact GMP. Finally, a combination of these two factors may underlie their motor control difficulties. Based on definitions of AOS provided by Darley et al. (1975) and current data, a breakdown in GMP appears to be a promising

explanation for AOS since the extant literature provides examples of potential disruptions of relative timing during speech tasks (e.g., Kent & McNeil, 1987; Kent & Rosenbek, 1983, Seddoh et al., 1996), as well as inaccurate motor planning during non-speech tasks (Hageman et al., 1994). However, preliminary evidence that parameterization may also be disrupted in AOS was provided by Hageman et al., 1994.

In recent studies, the production of speech and non-speech movements of subjects with AOS has been compared to that of subjects with conduction aphasia (CA) (Robin et al., in preparation; McNeil & Adams, 1991; McNeil et al., 1994; McNeil, Weismer et al., 1990; Odell et al., 1991 a, b; Seddoh et al., 1996). Individuals with CA demonstrate speech sound errors similar to those displayed by individuals with AOS (Odell et al., 1991a, b; McNeil & Kent, 1990), however, the errors produced by these individuals are generally believed to result from an underlying phonological impairment associated with aphasia. McNeil and Kent (1990) proposed that some of the speech errors observed in CA may also be attributed to breakdowns in motor control. This argument is based primarily on observations of the acoustic signal or articulator movements obtained during speech samples (McNeil & Adams, 1991; McNeil et al., 1994; McNeil, Liss et al., 1990) during which subjects with AOS and CA exhibit performance deficits implicating breakdowns in motor control. However, during non-speech motor control tasks, subjects with CA do not reliably show performance deficits (Hageman et al., 1996; McNeil, Weismer et al., 1990).

The present study represents a preliminary examination of GMP and parameterization accuracy in subjects with AOS or CA during a non-speech oral motor task. It was predicted that all subjects with AOS would exhibit impaired GMP accuracy (although some subjects with AOS might also exhibit reduced parameterization accuracy). It was further predicted that subjects with CA would exhibit GMP and parameterization accuracy similar to that of non-brain damaged controls.

Method

Subjects

Two experimental subject groups were recruited, AOS and CA. A subject was included in the AOS group if he/she made exhibited effortful trial-and-error posititioning of the articulators, dysprosody, error inconsistency across repetitions, and notable difficulty in initiating utterances (Kent & Rosenbek, 1983). In addition, each subject had to exhibit these symptoms as a result of a single left hemisphere lesion, and to be at least one year post-onset (see Table 1). All subjects exhibited some level of concomitant aphasia as measured by the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983), the Western Aphasia Battery (WAB) (Kertesz, 1982), or the Multilingual Aphasia Examination (MAE) (Benton & Hamsher, 1989), but not to the extent that it interfered with their ability to complete the task. Subjects who exhibited oral weakness or incoordination were not included in this experimental group. Four subjects met these inclusion criteria. The mean age of the subjects with AOS was 69.25 (range 60-77). All subjects in the AOS group exhibited non-fluent speech.

The CA group included subjects who were at least one year post-onset left hemisphere stroke and exhibited frequent sound substitutions in the presence of fluent speech. These subjects exhibited aphasia as measured by standardized aphasia batteries including the BDAE and the WAB or the MAE. Subjects with CA all demonstrated inordinate impairment of repetition with the relative prevalence of phonemic paraphasias. Subjects with aphasia who exhibited weakness, incoordination, or groping movements of the articulators were not included in this experimental group. Four subjects met the inclusion criteria. The mean age of the subjects with CA was 54.25 (range 25-71). All of the subjects with CA exhibited fluent speech.

In addition to the two experimental groups, four non-brain damaged subjects were recruited to serve as controls. The mean age of the control subjects was 67.25

Table 1.
Subject Information

Subject Information

Subject	Age	Sex	Time Post-Onset	Site of Lesion	Spontaneous Speech	Auditory Comprehension	Repetition	Naming	Reading Comprehension
AOS1	60	M	5 years	L MCA	Nonfluent	1	4	1	1
AOS2	77	M	15 months	L basal ganglia	Nonfluent	3	4	NA	3
AOS3	69	M	3 years	L hemisphere	Nonfluent	2	4	3	2
AOS4	71	M	21 years	l. basal ganglia and insula	Nonfluent	2	4	NA	3
CA1	70	M	6 years	l. inferior half of precentral gyrus and superior half of inferior frontal gyrus	Fluent	2	3	3	2
CA2	69	M	7 years	L basal ganglia and insula	Fluent	2	4	2	2
CA3	53	M	9 years	L inferior parietal lobe, insula	Fluent	3	4	3	2
CA4	25	M	2 years	L hemisphere	Fluent	2	3	2	2
C1	69	M	n/a	n/a	n/a	n/a	n/a	n/a	n/a
C2	75	M	n/a	n/a	n/a	n/a	n/a	n/a	n/a
C3	68	F	n/a	n/a	n/a	n/a	n/a	n/a	n/a
C4	57	M	n/a	n/a	n/a	n/a	n/a	n/a	n/a

1= normal, 2 = mild impairment, 3 = moderate impairment, 4 = severe impairment, NA = not able to be assessed, n/a = not applicable.

Lesion information was obtained from chart review or from the data reported by Seddoh et al. (1996) for subjects AOS4, CA1, CA2, and CA3, who participated in both experiments. Severity ratings were determined by a speech-language pathologist and a board certified neuropsychologist based on the test results, clinical interview, chart review, and perceptual judgements of speech.

(range 57-75). Although the control group appears to better match the AOS group than the CA group, it should be noted that the mean age of the CA group was greatly decreased by a 25 year old subject. Also, the mean performance of the control subjects is similar to that of forty non-brain damaged subjects who completed a similar experiment (Clark & Robin, in preparation). Thus, the apparent age differences between the control group and the CA group were not deemed critical for the purposes of this experiment. All subjects passed a vision screening, during which they were asked to identify a letter and a set of numbers displayed on the screen used during the experimental task (see below). One nonverbal subject with AOS responded to yes/no questions about the visual display to verify adequate vision.

Apparatus

Subjects were seated twenty-four inches from a computer monitor on which the movement target patterns were displayed. Subjects' heads were stabilized in all planes with a wall-mounted cephalostat to reduce whole-head movement artifact using well standardized methodology (Muller & Abbs, 1979). Inferior-superior labiomandibular movements were transduced with a strain gauge. The amplified signal was digitized at 200 Hz by a Metrabyte Dash 16 analog-to-digital converter and stored directly to the hard drive on a personal computer.

Task

The experimental task was modeled after Wulf et al. (1993), with the modification that labiomandibular movements were targeted instead of upper limb movements. Subjects were required to produce labiomandibular opening-closing movements with specific spatiotemporal goal movement patterns. Four target movement patterns were utilized, each with the same relative timing and amplitudes, but with different absolute movement times (see Figure 1).

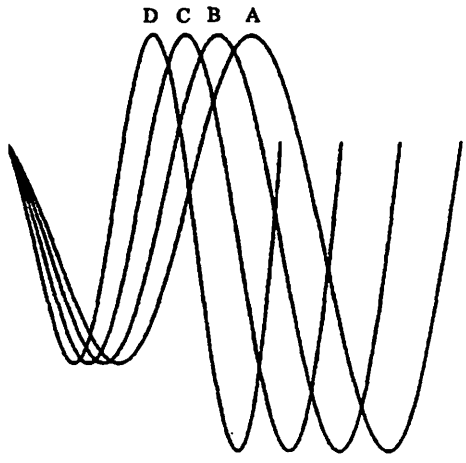


Figure 1. Target movement patterns.

The total target movement times for targets A, B, C, and D were 1238, 1073, 908, and 743 ms respectively.

Before each trial, one of the target movement patterns was displayed for four seconds. The specific target wave was identified by a letter (A, B, C, or D) displayed in the upper left corner of the screen. When the target was removed, the subject began with the lower lip in a slightly opened position, and then produced a sequence of opening-closing-opening-closing movements to produce the target pattern, attempting to match the target in both space and time. During the movement, the screen remained blank.

After a two second interval, during which the subject's movements were transduced, digitized, and recorded, knowledge of results (KR), was presented by superimposing the subject's actual movement trace (displacement over time) with that of the target pattern. The target pattern was displayed in white (on a black background), and the subject's pattern was displayed in blue. The root-mean-square deviation (RMS error) between the two patterns was calculated and displayed in the upper right hand corner of the screen. KR remained on the screen for 5 seconds.

Procedure

Subjects were provided with verbal and written instructions for the task. Following the instructions, the subjects performed ten familiarization trials using Version B. The subject's performance was discussed after each familiarization trial to ensure the subject knew how to interpret the KR.

Practice trials of Versions A, B, and D were presented in blocks of six trials, in which each version was performed six times before a switch to another version. The order of task versions was randomized, with the restriction that each version appeared once in each three-block sequence. The relative KR frequency was systematically reduced across practice trials. Specifically, in the first three blocks, KR was presented on five of the six trials, during the next six blocks, KR was presented on four of the six trials, and the during the final six blocks, KR was presented on three of the six trials. This yielded an average KR frequency of 63%.

Each subject completed a total of 90 practice trials. Following a rest period and recalibration, the subjects completed the experimental task, which consisted of ten retention trials of each practiced version A, B, and D, as well as ten transfer trials on version C, with the order of the task versions randomized. No KR was provided during these trials. Version C differed from the other target patterns only in absolute movement time (908 ms), which was between the movement times of the practice targets. The entire experiment, including familiarization trials and the rest period, ranged from 90 to 110 minutes.

Data Analysis

In the manner described by Wulf et al. (1993), temporal and amplitude scaling was conducted to determine the accuracy of GMP and amplitude and timing parameters. For each trial, the target wave was temporally rescaled (compressed or expanded) from .4 to 1.74 in increments of .02 with RMS error re-calculated at each increment. The scaling which resulted in the smallest RMS error was termed the "temporal scaling factor" and was the measure of temporal parameterization accuracy. Within this scheme, movement patterns produced too slowly, resulting in longer movement durations, result in temporal scaling factors less than 1.0, as the pattern must be compressed (i.e., multiplied by a scaling factor less than 1.0) to best match the target pattern. Conversely, movement patterns produced too quickly produce temporal scaling factors greater than 1.0, as these patterns must be expanded to best match the target. A scaling factor of 1.0 would indicate that the subject produced a movement of the exact total duration of the target pattern.

Next, the amplitude scaling factor was obtained. The variance in amplitudes of the position-time samples in the target pattern was divided by the variance in amplitudes of the temporally-rescaled trace. The square root of this ratio was deemed the amplitude scaling factor and was the measure of amplitude parameterization accuracy. An amplitude scaling factor greater than 1.0 indicated the subject produced the movement with amplitudes smaller than the target pattern, while an amplitude scaling factor less than 1.0 indicated the subject produced the movement with amplitudes greater than the target pattern.

Once the temporal and amplitude scaling factors were computed, the factors were applied to the movement trace and RMS error was again calculated. This residual RMS error was taken as the measure of GMP accuracy. That is, after correcting for errors in *absolute* timing and amplitude (by applying the scaling factors), any difference remaining between the target and the actual movement trace results from differences in *relative* timing and amplitude, which is specified by the GMP. Thus, GMP accuracy was defined as the residual RMS error calculated after the temporal and amplitude scaling factors were applied.

In order to justify the use of the analysis protocol described above, it was necessary to verify that the movements produced during the task were programmed with a single GMP (Wulf et al., 1993). This was achieved by computing correlations among temporal landmarks selected from the velocity-time function of the movement traces (Young & Schmidt, 1991). The assumption is that since movements programmed with the same GMP differ only in absolute timing, but not relative timing, the correlation between the landmarks of a single structural unit (GMP) should be near 1.0. However, if a movement is governed by two or more GMPs, the correlations between landmarks which fall under the governance of separate GMPs should fall towards zero.

Average correlations between contiguous landmarks averaged from .51 to .97, except for the correlation between Landmark 1 and Landmark 2, for which the average correlation was .157. The observed pattern of correlations was consistent with the interpretation that the lip movement was generated by one GMP (with the exception of the first landmark, which is from movement onset to the first velocity peak.). Data analysis proceeded based on several factors. First, Schmidt (personal communication) proposed this pattern of correlations likely represents "start-up" variability associated with initiating a movement. That is, the motor delays (Heuer, 1988; Heuer et al., 1995) during the initiation of a movement may be more variable than those associated with the remainder of the movement. Although it might be expected that subjects with AOS would exhibit significant start-up variability, given the difficulty these subjects exhibit initiating utterances, the unit structure exhibited by the AOS subjects was not different from that of the other groups, or that of young normal subjects (Clark & Robin, in preparation).

Secondly, because the displacement trace corresponding to the start of the movement to the first landmark was a very small portion of the entire movement (less than 10% of the total movement time or approximately 50 ms), it was reasoned that even if this portion of the movement was governed by a different GMP, including that portion in the analysis of the remainder of the movement would minimally effect the results of the analysis procedure.

Results

GMP Accuracy

Figure 2 depicts the residual RMS error for the individual experimental subjects, as well the mean performance of the control subjects. The error bars reflect intra-subject variability for the experimental subjects and inter-subject variability for the control subjects. Note that the ordinate represents error, so greater accuracy is reflected by lower RMS error values. Subjects AOS1 and AOS2 exhibited highly accurate GMPs, particularly in the retention and transfer conditions. In contrast, the remaining subjects with AOS exhibited lower GMP accuracy across conditions. Across subjects, the subjects with CA exhibited more consistent GMP accuracy, as well as accuracy levels similar to the controls, particularly in the retention and transfer conditions.

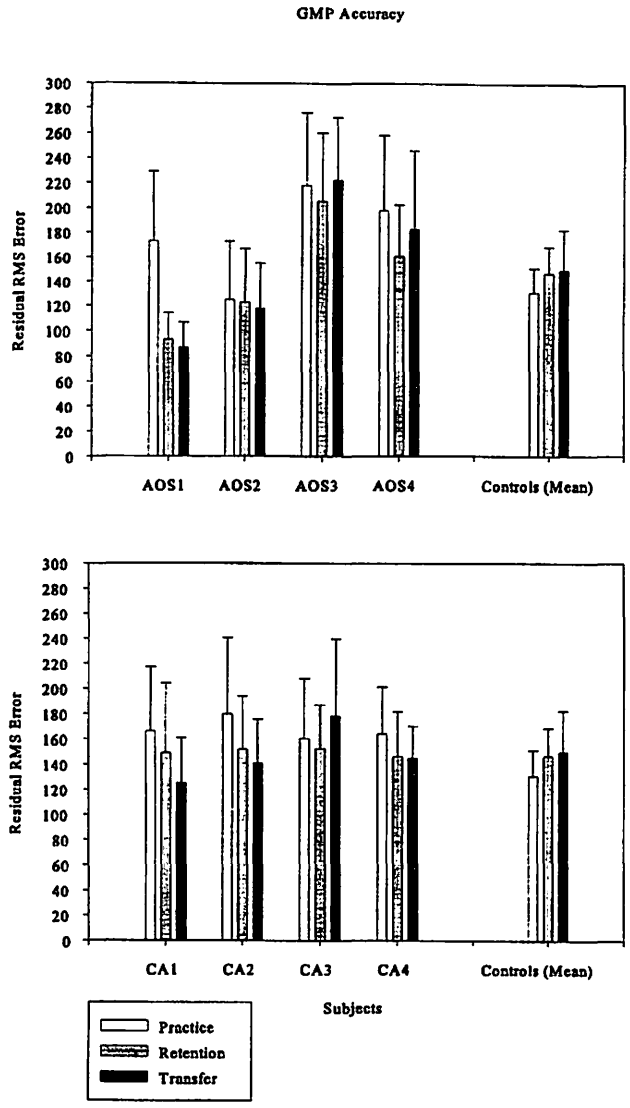


Figure 2. GMP Accuracy for Individual Subjects. Mean and standard deviation for individual subjects. The mean and standard deviation of the control group are provided for comparison.

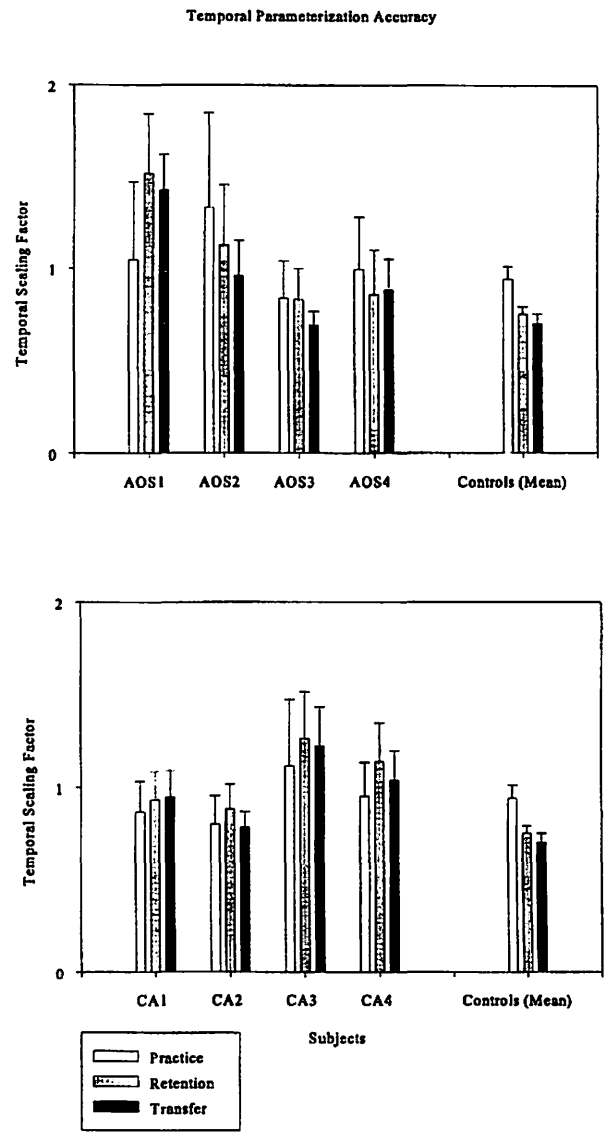


Figure 3. Mean temporal parameterization accuracy for individual subjects. Mean and standard deviation of temporal scaling factor for individual subjects.

Temporal Parameterization Accuracy

Figure 3 illustrates temporal parameterization accuracy. Recall that a scaling factor of 1.0 reflects perfect temporal parameterization accuracy, while a scaling factor greater than 1.0 indicates the subject performed the movement more quickly than the target, and a scaling factor less than 1.0 indicates the subject performed the movement more slowly than the target pattern. Subjects AO3 and AOS4 exhibited temporal parameterization accuracy equal to or greater than that of the controls in most conditions. Subjects AOS1 and AOS2 exhibited high performance variability across conditions: Subject AOS1 exhibited normal temporal parameterization accuracy only in the practice condition, while subject AOS2 was inaccurate only in the practice condition.

The subjects with CA all tended to err to the same degree although not necessarily in the same direction. Subjects CA1 and CA2 tended to produce the movements too slowly as evidenced by temporal scaling factors less than 1.0. This pattern was also demonstrated by the control subjects. In contrast, temporal scaling factors greater than 1.0 indicate that subjects CA3 and CA4 tended to produce the movements more quickly than the targets.

Amplitude Parameterization Accuracy

Amplitude scaling factors greater than 1.0 indicate that the subject produced movements smaller in amplitude than the target pattern, while scaling factors greater than 1.0 indicate the subject produced movements greater in ampli-

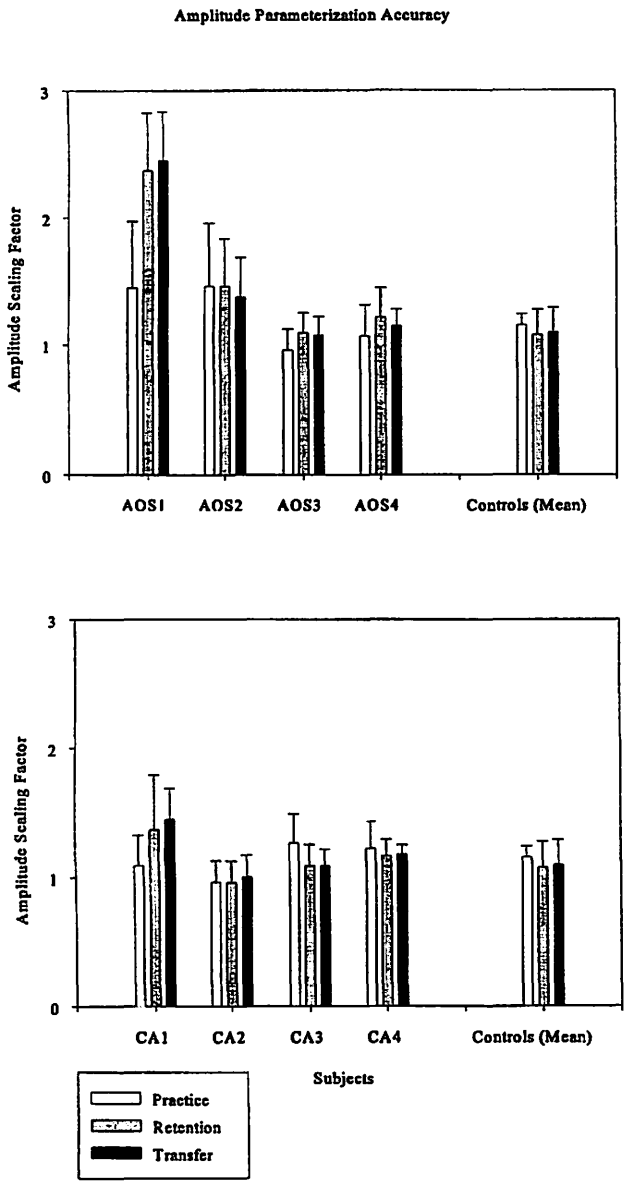


Figure 4. Mean amplitude parameterization accuracy for individual subjects. Mean and standard deviation of amplitude scaling factor for individual subjects.

tude than the target pattern. Movements perfectly matching the target in amplitude resulted in a scaling factor of 1.0. It is clear from Figure 4 that subject AOS1 exhibited very poor amplitude parameterization accuracy. This subject consistently produced movements much smaller in amplitude than the target pattern, particularly in the retention and transfer conditions. Subject AOS2 also produced movements smaller than the target, while the remaining subjects exhibited amplitude parameterization accuracy similar to that of the controls.

Subject CA1 also produced movements which were smaller in amplitude than the target patterns, particularly in the retention and transfer conditions. Each of the remaining subjects with CA exhibited amplitude parameterization accuracy in the range exhibited by the controls.

Discussion

The present experiment examined GMP and parameterization accuracy of subjects with AOS or CA during a non-speech oral motor task. Based on the prominent conceptualization of AOS as a motor programming disorder and previous evidence of disruption in relative timing of movements, it was predicted that all subjects with AOS would exhibit reduced GMP accuracy. Further, it was predicted that, in addition to GMP inaccuracy, subjects with AOS might also exhibit impaired parameterization. Subjects with CA were predicted to perform similarly to controls on all of the measures. *None* of these predictions were realized in the current study.

Performance of Subjects with AOS

While three subjects with AOS exhibited reduced GMP accuracy in at least one condition, subjects AOS1 and AOS2 exhibited high GMP accuracy in at least two conditions. Interestingly, it was only these two subjects who exhibited reduced parameterization accuracy. Thus, even though some subjects with AOS exhibited reduced GMP accuracy (AOS prediction one), no subjects simultaneously exhibited GMP and parameterization inaccuracy (AOS prediction two).

Thus, an apparent dissociation was observed; subjects with AOS exhibited impairments in GMP or parameterization, but not both. Several potential explanations arise for this finding. First, it is possible that the observed patterns reflect performance trade-offs. That is, within the context of an impaired motor programming system, subjects may have only enough processing resources to correctly program either the GMP or the parameters, but not both. Thus, the different performance patterns may simply reflect different resource allocation strategies in which subjects with AOS, reaching the limits of their capacity, were forced to choose some aspects of motor programming to which they would attend, but not all programming processes. Since neither GMP or parameterization accuracy was emphasized in the present study, subjects may have been equally likely to prioritize either of these aspects of control. However, given that the GMP defines the shape of the movement, it may be inefficient for subjects to attend more to the parameter than the GMP. Thus, those subjects who had accurate parameterization but inaccurate GMPs may have been utilizing a maladaptive strategy in that changing the parameter is relatively trivial compared to modification of the program that defines the shape of skilled actions. It is interesting to note that young non-brain damaged subjects may also adopt this strategy (Clark & Robin, in preparation), suggesting that while this strategy appears maladaptive, it may be preferred by many subjects, irrespective of the presence of apraxia. However, in the person with AOS, this strategy may be particularly disruptive since these subjects may not have the

flexibility to maintain overall performance accuracy in the face of impaired motor programming.

Important support for viewing the data as reflective of a performance trade-off across conditions is provided by the performance of subject AOS1 (see Figures 2-4). During the practice trials, AOS1 exhibited reduced GMP accuracy with relatively accurate amplitude parameterization (note that temporal parameterization accuracy was relatively stable across conditions). In contrast, during the retention and transfer conditions, GMP accuracy improved but at the cost of decreased amplitude parameterization accuracy. It is remarkable that within a single subject, performance trade-offs were observed. These trade-offs suggest that control strategies were utilized to cope with an impaired motor control system.

It may also be the case that performance differences within the AOS group represent subtypes of AOS. Subtypes of AOS have been proposed based on site of lesion (Kertesz, 1984; Square-Storer & Apeldoorn, 1991) and behavioral characteristics (Hough & DeMarco, 1996). Unfortunately, inadequate lesion data for the subjects studied here precludes differentiation based on specific site of lesion. Likewise, Hough and DeMarco developed their subtypes based on the ability of subjects to maintain phonological representations, skills which were not addressed in the present study.

It is clear from these data that the conceptualization of AOS as a deficit in "the motor program" or the GMP only is not accurate. High level motor programming is a complex process in which the GMP needs to be realized and then set according to specific parameters. AOS appears to involve the entire process of motor programming and not only one process (the GMP) within the programming of events. Such an impairment requires speakers with AOS to produce skilled movement patterns under great resource demand resulting in an increased susceptibility to breakdown.

Performance of Subjects With CA

The final prediction, that subjects with CA would perform normally on all the measures, was nullified by the performance of subject CA1. This subject exhibited reduced amplitude parameterization accuracy in the retention and transfer conditions. This finding suggests that while GMP or parameterization impairments do not appear to be characteristic of CA, such deficits may be present in some patients. Our findings suggest that the present paradigm may be sensitive to detection of these subtle motor control deficits. Thus, notions about motor control anomalies underlying the speech characteristics of CA appear inaccurate. Rather one might view the motor control deficits reported in some patients with CA as concomitant to the linguistic disorder and not necessarily as a core part of the problem.

Conclusions

The current study is the first to operationally define and then examine motor programming ability in subjects with AOS, even though Darley (1969) defined the disorder over twenty-five years ago. Interestingly, not all subjects with AOS exhibited a defective GMP, as operationally defined in this study. However, before dismissing Darley's assertions, it is necessary to explore how the current experiment's operational definition relates to Darley's original conceptualizations.

Darley's definition focused on the programming of "positioning" of articulators and "sequencing of muscle movements" (p.639). This particular definition relates most closely to Schmidt's (1975, 1988) concept of GMP *plus* amplitude parameterization, both part of the high level programming process. Thus, the present view of AOS refines Darley's notion and does not focus on the GMP only but suggests that the GMP and/or parameterization may be impaired in AOS. It is also clear that amplitude parameterization as well as temporal parameterization, to which Darley does not refer, may also be impaired.

Several questions are raised by the present findings. One of the most prominent issues is if and how impairments of GMP and/or parameterization differentially affect speech production. That is, how does the speech of subjects who exhibit primarily GMP errors differ from the speech of subjects with primarily parameterization errors? This is a particularly relevant question since researchers disagree about the degree to which speech and non-speech oral movements are related (e.g., Weismer & Liss, 1981; Folkins et al., 1995). Additional information about the relationships among GMP and parameterization accuracy of speech and non-speech movements and speech accuracy would contribute greatly to our understanding of oral motor control.

Other questions which warrant addressing include: Are GMP/parameterization impairments differentially associated with specific sites of lesion? Do subjects with different underlying impairments (GMP versus parameterization) respond differently to speech treatment? Further, are different types of treatment more efficacious when applied to subjects who exhibit specific underlying impairments, and if so, what is the nature of that treatment? And finally, what are the implications for each of the above questions if differences in GMP and parameterization accuracy primarily reflect resource allocation strategies rather than actual differences in competence?

Replication of the current study with a larger number of subjects with a broader range of severity levels, while controlling for potential confounds (e.g., visual memory deficits) is needed. Precise lesion information as well as exacting speech performance measures will aid in the under-

standing of the contribution of specific impairments of GMP and/or parameterization to speech performance, as well as assist in the possible development of more comprehensive classifications of subgroups of AOS. Treatment efficacy studies which address the issue of underlying impairment would also contribute to our understanding of how motor control is affected in AOS. Finally, additional study of the GMP and parameterization accuracy in subjects with CA will lead to a better understanding of the underlying impairment of this disorder as well and allow for estimations of how frequently motor control anomalies co-exist with linguistic disorders in these patients.

Acknowledgments

The authors wish to thank Martin Milder for his assistance in writing the computer program. We also extend thanks to Dick Schmidt for his assistance in many phases of this project and to Gaby Wulf for her willingness to share her programs with us. We appreciate the thoughtful feedback regarding an earlier version of this manuscript provided by Margaret Rogers, Julie Wambaugh, and an anonymous reviewer. This work represents a portion of a doctoral dissertation at the University of Iowa by the first author under the direction of the second author. The authors thank the other members of the committee John Folkins, Jerry Moon, Dick Schmidt, and Trisha Zebrowski.

This work was support by NIDCD Center Grant # DC00976.

References

Benton, A.L., and Hamsher, K. (1989) *Multilingual Aphasia Examination* (2nd ed.). (AJA Associates, Iowa City).

Clark, H. M., and Robin, D. A. (in preparation) Effect of reduced feedback on generalized motor program and parameterization learning.

Darley, F. L. (1969) *Nomenclature of expressive speech-language disorders*. Paper presented to Academy of Aphasia meeting, Boston, Massechusetts.

Darley, F. L., Aronson, A. E., and Brown, J. R. (1975) *Motor Speech Disorders* (W. B. Saunders Company, Philadelphia).

Folkins, J. W., Moon, J. B., Luschei, E. S., Robin, D. A., Tye-Murray, N., and Moll, K. L. (1995) What can non-speech tasks tell us about speech motor disabilities? *Journal of Phonetics*, 23, 139-147.

Goodglass, H. and Kaplan, E. (1983) *The Assessment of Aphasia and Related Disorders* (2nd ed.). (Kea Febiger, Philadelphia).

Hageman, C. F., Robin, D. A., Moon, J. B., and Folkins, J. W. (1994) Oral motor tracking in normal and apraxic speakers. *Clinical Aphasiology*, 22, 219-229.

Heuer, H. (1988) Testing the invariance of relative timing: comment on Gentner (1987). *Psychological Review*, 95, 552-557.

Heuer, H., Schmidt, R. A., and Ghodsian, D. (1995) Generalized motor programs for rapid bimanual tasks: a two-level multiplicative-rate model. *Biological Cybernetics*, 73, 343-356.

Hough, M. S., DeMarco, S. (1996) Phonemic retrieval in apraxia of speech: Is there more than one type of impairment? In D. A. Robin, K. M. Yorkston, and D. R. Beukelman (Eds.) *Disorders of Motor Speech: Assessment, Treatment and Clinical Characterization* (Paul H. Brookes, Baltimore).

Kent, R. D., and McNeil, M. R. (1987) Relative timing of sentence repetition in apraxia of speech and conduction aphasia. In J. H. Ryalls (Ed.) *Phonetic Approaches to Speech Production in Aphasia and Related Disorders* (College-Hill, Boston), pp. 181-220.

Kent, R. D., and Rosenbek, J. C. (1983) Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research*, 26, 231-249.

Kertesz, A. (1982) *Western Aphasia Battery*. (Grune & Stratton, New York).

Kertesz, A. (1984) Subcortical lesions and apraxia of speech. In J. C. Rosenbek, M. R., McNeil, and A. E. Aronson (Eds.) *Apraxia of Speech: Physiology--acoustics--linguistics--management* (College-Hill Press, San Diego), pp.73-90.

McNeil, M. R., and Adams, S. (1991) A comparison of speech kinematics among apraxic, conduction aphasic, ataxic dysarthric, and normal geriatric speakers. *Clinical Aphasiology*, 19, 279-294.

McNeil, M. R., Hashi, M., and Southwood, H. (1994) Acoustically derived perceptual evidence for coarticulatory errors in apraxia and conduction aphasic speech production. *Clinical Aphasiology*, 22, 203-218.

McNeil, M. R., and Kent, R. D. (1990) Motoric characteristics of adult aphasic and apraxic speakers. In G. R. Hammond (Ed.) *Cerebral Control of Speech and Limb Movements* (Elsevier Science Publishers, North Holland), pp. 349-386.

McNeil, M. R., Liss, J. M., Tseng, C., and Kent, R. D. (1990) Effects of speech rate on the absolute timing and relative timing of apraxic and conduction aphasic sentence production. *Brain and Language*, 38, 135-158.

McNeil, M. R., Weismer, G., Adams, S., and Mulligan, M. (1990) Oral structure non-speech motor control in normal, dysarthric, aphasic, and apraxic speakers: Isometric force and static position control. *Journal of Speech and Hearing Research*, 33, 255-268.

Muller, E. M. and Abbs, J. H. (1979) Strain gauge transduction of lip and jaw motion in the midsagittal plan: refinement of a prototype system. *Journal of the Acoustic Society of America*, 65, 481-486.

Odell, K., McNeil, M. R., Rosenbek, J. C., and Hunter, L. (1991a) Perceptual characteristics of vowel and prosody production in apraxic, aphasic, and dysarthric speakers. *Journal of Speech and Hearing Research*, 34, 67-80.

Odell, K., McNeil, M. R., Rosenbek, J. C., and Hunter, L. (1991b) A perceptual comparison of prosodic features in apraxia of speech and conduction aphasia. *Clinical Aphasiology*, 19, 295-306.

Robin, D. A., Bean, C., and Folkins, J. W. (1989) Lip movement in apraxia of speech. *Journal of Speech and Hearing Research*, 32, 512-523.

Robin, D. A., Hageman, C. F., Moon, J. B., Woodworth, G., Clark, H. M, and Folkins, J. W. (in preparation) Oral motor control ability of persons with apraxia of speech or conduction aphasia: A visuomotor tracking study.

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

Schmidt, R. A. (1988) *Motor Control and Learning*. (Human Kinetics, Champaign, IL.).

Seddoh, S. K., Robin, D. A., Hageman, C., Sim, H., Moon, J. B., and Folkins, J. W. (1996) Durational control in apraxia of speech versus conduction aphasia. *Journal of Speech and Hearing Research*, 39, 590-603.

Square-Storer, P., and Apeldoorn, S. (1991) An acoustic study of apraxia of speech with different lesion loci. In C. A. Moore, K. M. Yorkston, and D. R. Beukelman (Eds.) *Dysarthria and Apraxia of Speech: Perspectives on Management* (Paul H. Brookes, Baltimore.) pp. 216- 227.

Weismer, G., and Liss, J. (1981) Reductionism is a dead-end in speech research: Perspectives on a new direction. In C. A. Moore, K. M. Yorkston, and D. R. Beukelman (Eds.) *Dysarthria and Apraxia of Speech: Perspectives on Management* (Paul H. Brookes, Baltimore.) pp. 15-28.

Wulf, G., Schmidt, R. A., and Deubel, H. (1993) Reduced feedback frequency enhances generalized motor program learning but not parameterization learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1134-1150.

Young, D. E., and Schmidt, R. A. (1991) Motor programs as units of movement control. In N. I. Badler, B. A. Barsky, and D. Zeltzer (Eds.) *Making the Move: Mechanic Control and Animation of Articulated Figures* (Morgan Kaufman San Mateo, California) pp. 129-155.

Part II

Tutorial reports and updates for Training, Information Dissemination, and Continuing Education

NCVS Status and Progress Report - 12
June 1998, 197-198

Training Update

Patricia Zebrowski, Training Coordinator
Department of Speech Pathology and Audiology, The University of Iowa

The NCVS Training Program currently provides tuition and stipend support to three predoctoral trainees and one postdoctoral fellow. The training board is currently searching for a second postdoc. Qualified candidates with a wide variety of backgrounds will be considered: speech science, speech pathology, physiology, biology, engineering, neuroscience, psychology, medicine, vocal music or another related field.

Current Predoctoral Trainees

Renee Bogschutz, M.S., CCC-SLP

Renee completed her undergraduate and graduate degrees in speech-language pathology at Eastern New Mexico University (ENMU). Her area of research interest during her educational experience at ENMU was professional issues in speech-language pathology. She has two years of clinical experience in child and adult neurogenic communication disorders. Currently, she is a doctoral student in the Department of Speech Pathology and Audiology at The University of Iowa. Renee is particularly interested in clinical decision-making in speech-language pathology and motor speech disorders. She is working under the direction of Dr. Donald Robin.

Eric Hunter, M.S.

Eric earned his bachelor's and master's degrees in Physics from Brigham Young University. In his program, he strongly emphasized speech acoustics, graduating with a master's degree in 1997. In the Fall 1997, he came to The University of Iowa to become a speech scientist and pursue his doctoral degree in the Department of Speech Pathology and Audiology. Currently, Eric is working under the direction of Dr. Ingo Titze with various computer models of the larynx.

Helen Sharp, M.S.

Helen earned a bachelor of science degree in Human Kinetics from The University of Guelph, Ontario, and a master's degree in speech-language pathology at The University of Pittsburgh. She has clinical experience as a speech-language pathologist at Loma Linda University Medical Center, working with pediatric rehabilitation, head and neck cancer, tracheostomy, ventilator, dysphagia and cleft and craniofacial anomaly patients. Through her clinical work, Helen became interested in medical ethics and pursued this interest by enrolling as a fellow in the program in clinical medical ethics at the University of Chicago. She stayed in Chicago for an additional year as a senior fellow, participating in ethics consultations and developing staff education at a county long-term care facility in the Chicago suburbs. In 1996, Helen enrolled as full-time doctoral student at The University of Iowa. She recently completed her coursework and is developing her dissertation project in the area of head and neck cancer under the direction of Dr. Jerald Moon.

Postdoctoral Fellowship

Michael Edgerton, DMA

Mike earned a Doctorate of Musical Arts degree in Music Composition at The University of Illinois. He has also taught composition at Yonsei University and served as a visiting scholar at Hanyang University in Seoul, South Korea. He is currently ready to begin a third year of postdoctoral study at The University of Wisconsin-Madison, working with Dr. Diane Bless. His current work involves research on overtone singing, the double source phenomenon of Tibetan chant, vocal violence, ingressive phonation and a new template of vowel and consonant articulation which examines the temporal characteristics of the entire upper vocal tract and which artistically amounts to a novel and pragmatic performance and compositional tool.

Former NCVS Trainees

Eileen Finnegan completed her dissertation entitled "Modulations in Respiratory and Laryngeal Activity associated with Changes in Vocal Intensity" under the direction of Dr. Erich Luschei. She received her Ph.D. in May 1998. She has accepted a position as an assistant research scientist in the Department of Otolaryngology at The University of Iowa Hospitals and Clinics.

Elisa (Mordue) Huff is completing her dissertation in the area of adult neurogenic disorders under the direction of Dr. Donald Robin.

John Nelson completed his dissertation entitled, "Refinement of the Leaky Integrator Model for Temporal Auditory Processing Utlizing Sinuisoidal and Duty-Cycle Modulation Functions," and graduated in May 1998. He is currently an Assistant Professor in the Department of Communication Disorders at the University of Texas, Austin.

Phyllis Palmer graduated with her doctoral degree in May 1998. Her dissertation was entitled, "Contributions of Individual Muscles to the Submental Surface Electromyogram During Swallowing." She was mentored by Dr. Erich Luschei. She is currently working as a clinical speech pathologist in the Four Corners area of the southwestern United States.

Lucrezia Tomes completed her postdoctoral research under the NIDCD's mechanism to promote recruitment of individuals with disabilities into biomedical research careers. During her one and one-half year fellowship at The University of Illinois, she developed a standardized process of rating hypernasality severity using synthesized samples of speech. She worked primarily with Drs. David Kuehn and Brad Story. Lou recently accepted an academic position at The University of Wisconsin-Whitewater in the Department of Communicative Disorders.

NCVS Status and Progress Report - 12
June 1998, 199-200

Information Dissemination Update

Thea Carruth, Information Dissemination Coordinator
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts

The goal of the dissemination project is to distribute information to the general public about care of the voice as well as prevention, detection, and treatment of voice and speech disorders. Methods for reaching the public include the use of media (electronic and print), educational presentations (workshops, seminars, lectures and performances), interactive exhibits and shared information with other professional organizations dedicated to voice care, speech and training.

Public Service Announcements

A third, television Public Service Announcement entitled "One Voice, One World" was produced and distributed in 1998 and presents sounds and images ranging from Native American and Gregorian chant to a rap artist, and refers to the therapeutic benefits of chant. This PSA was designed as a follow-up to the first two and builds upon the success of the second PSA entitled "The Voice Doctor." Voice Doctor won a Telly Award in the Television Public Service Announcement category and received a nomination for the Heartland Regional Emmy Award in the Television Public Service category. Again this year, television stations were selected from medium-sized markets with either with a high number of performers (e.g.. Austin and Louisville) or, a proximity to a voice research center (e.g. Baltimore). Orlando was chosen because of the high number of vacationers in the area. The phone number for the Denver site of the NCVS appears on the screen at the end of the spot. Calls from outside the Denver area are referred to an appropriate agency in the viewer's area. Prior to mailing the video tapes to 57 television stations, telephone calls were placed to each of the Public Service Directors in an effort to reestablish rapport to answer any questions.

Pavarobotti, the singing robot, made a live appearance in February of 1998 at Clapp Recital Hall on The University of Iowa campus. He demonstrated his tenor technique at the "Vocal Extravaganza," performing with Dr. Ingo Titze and many talented vocalists of the NCVS. In addition to the singing and theatrical performances, NCVS investigators and students set up ten "hand-on" demonstrations to add a scientific dimension to the evening's activities. These interactive displays explained how the voice works; what help is available when the voice is damaged; and current research in oral communication. More than 300 people attended "Vocal Extravaganza."

Third Distribution of Public Service Announcement

MARKET	NUMBER OF TELEVISION HOUSEHOLDS
Denver CO	1,142,160
Cedar Rapids/Iowa City IA	376,160
Salt Lake City UT	616,700
Nashville TN	749,060
Minneapolis/St. Paul MN	1,410,630
Seattle WA	1,468,730
New Orleans LA	615,180
Atlanta GA	1,567,300
Tucson AZ	322,800
Boston MA	2,104,900
Sacramento CA	1,100,000
Cleveland OH	1,142,000
*Pittsburgh PA	1,141,000
Phoenix AZ	1,017,000
Orlando FL	998,000
Baltimore MD	980,000
Portland OR	933,000
Indianapolis IN	925,000
Kansas City MO	780,000
*Louisville KY	533,000
Austin TX	417,000

Source: February 1997 A.E. Nielsen report or #1996 Medium Market Guide, as indicated. Note: A "television household" is a measurement used by the A.E. Nielsen company in measuring ratings and is the standard by which television markets are ranked. The measurement represents the number of households within a city, not the number of occupants or televisions.

A video of a similar performance was shot in the Denver Center for the Performing Arts studio by Denver Center Media, and has since been presented by Dr. Titze at conferences in Provo, Utah, and Berlin, Germany.

We have expanded robot's repertoire, and his voice has been improved with recent advances in vocal fold and tract modeling. Phonetically-based lip shapes have been incorporated into the newest animation. Our goal is to make the process of creating new material for the robot more automated. We have also discussed a simplified robot, possibly a "head only" edition, that is easier and less expensive to take on the road.

Pavarobotti continues to capture the public's attention whenever and wherever he performs. There are current plans to include a version of the robot in the Gould Voice Research Center's elementary school play, "How Coyote Found Voice" this Fall. Pavarobotti is also scheduled to perform for the Whittaker Foundation in San Diego this August.

Web Site

A World Wide Web site on the Internet for the WJ Gould Voice Research Center includes: background on research projects undertaken as part of the NCVS, Information on educational materials and workshops, and research training opportunities. During the past twelve months, the site has received 18,312 hits. The URL of the site is http://web1.dcpa.org.

Workshops and Performances/Voice Care

In addition to the Vocal Extravaganza noted above, community voice workshops continue to be one of the NCVS's most visible outreach efforts. In the Denver metropolitan area, workshops are conducted once each quarter at either the Denver Center for the Performing Arts or the Arvada Center for the Performing Arts (in suburban Denver). Attendance exceeds 100 people per workshop at each location and the topic alternates between the singing voice and the speaking voice. The content includes: how sounds of the voice are produced, vocal hygiene, how to maintain a healthy voice and when to seek medical attention. Workshops are interactive and include exercises and techniques. Evaluations are collected at the end of each workshop and feedback incorporated into subsequent workshops. In March and October of 1997, Michael Edgerton conducted a total of five workshops in extended vocal techniques for the University of Wisconsin reaching a total of 550 people.

An elementary school-aged classroom experience entitled: *How Coyotes Found Voice* was written and piloted in the Denver Public School system this year (ten performances in five schools). The performances focus on the mischievous exploits of the character, Coyote, one of the most recognized figures in Native American folklore. In this play/lesson, Coyote must save the voices of all of the other animals, and in the process, learns how the voice itself is produced and how it can be properly cared for.

The story of Coyote is told through words and music by two performers and incorporates selected students into the program, thereby helping to demonstrate how the children themselves might better understand the miracle of speech, and how the human body works. The play was written to be performed either with or without the singing robot.

An excerpt from the beginning of "Coyote"

ACTOR (as storyteller) *"The People, who inhabited the world long before the humans came, would argue among themselves about who'd come first. Wind - rushing, blustering, battering - boasted that he'd come first. River - gurgling, roiling, rolling, lapping - argued that she'd been first. 'Not so,' spit sand in his dry, dusty voice. 'I was first.' On and on they argued, and while they never came to an agreement, all the People believed that before they'd come in their bodies, they'd been alive from forever as songs of the Great Spirit...The People...never changed their belief that their world was an intricate weaving of the songs of all things, like the brightly colored blankets Spider wove..."*

NCVS Status and Progress Report - 12
June 1998, 201-203

Continuing Education Update

Julie Ostrem, Continuing Education Coordinator
Department of Speech Pathology and Audiology, The University of Iowa

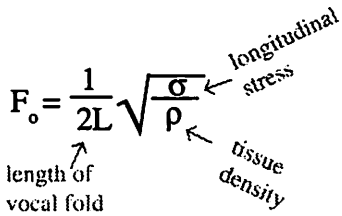
Advances in communications technology as well as traditional means of teaching provide NCVS investigators with a variety of ways to distribute information about new research to speech-language pathologists, otolaryngologists and other voice educators. More and more, NCVS investigators are exploiting the World Wide Web as an inexpensive, easily accessible and readily updatable conduit for teaching.

Over the past year, the bulk of NCVS CE resources have been directed toward developing a website describing normal voice production. Entitled "How Humans Speak, Sing, Squeak and Squeal," the website is best used in conjunction with Dr. Ingo Titze's textbook, *Principles of Voice Production.*

From the homepage, users may navigate between four main areas: Tutorials, Equation Exploder, Testing Center or Library.

The majority of teaching is done in the Tutorials section. Twenty-five mini-courses describe important selected topics in understanding how humans produce voice. Special effort is made to describe basic scientific principles in easily understandable language and to make these discussions clinically relevant. A glossary is hyperlinked to designated words, so that the user can easily find definitions of unfamiliar terminology. The topics of the 25 mini-courses are listed in the table to the right. As this report is submitted, 16 of the 25 tutorials have been written.

A second area of the website is called the Equation Exploder. This section assists users unaccustomed to thinking of physical properties in terms of mathematical equations. Terms of equations are put into words, for example, an equation explaining fundamental frequency:



Tutorials
"How Humans Speak, Sing, Squeak and Squeal"

A review of basic acoustics and physics concepts
Instrumentation in the Voice Lab
Biomechanics
- •Physical Laws Governing Voice Production
- •Biomechanical Properties of the Vocal Anatomy

Breathing
- •Lung Pressure and Power
- •Fluid Flow in Tubes

Oscillation
- •Models of Vocal Fold Vibration
- •Normal Modes of Vibration
- •Phonation Threshold Pressure

Sound Generation and Propagation
- •Graphing Phonation (waves and spectra)
- •Movement of sound from a source

Source-Filter Theory of Vowels
- •Rules for Modifying Vowels
- •How the Vocal Tract Filters Sound
- •The Spectrograph and Spectral Analysis

Voice Classification and Life Span
- •Factors Influencing Fo
- •Voice Changes Throughout Life

Control of Fo (Pitch)
- •The Cover and Body-Cover Models of Pitch Control

Intensity and Efficiency
- •Three Levels of Intensity Control (lungs, glottis, vocal tract)
- •A Discussion of Vocal Efficiency
- •The Singers' Formant

Registers
- •Voluntary register changes
- •Involuntary register changes
- •Voice qualities

Fluctuations and Perturbations
- •Physical Assessment of Fluctuations
- •Training the Ear: Tweaking Fluctuations in Voice

Just as the name implies, the Testing Center provides multiple-choice questions and automated scoring so that the user may ascertain how well s/he understood presented material.

The Library, not yet developed, is designed to provide a current bibliography of voice- and speech-related research articles.

The final nine tutorials for the educational software project will be written by Fall 1998. In the following year, animations, quick time movies, audio samples and digital photographs will be added as appropriate to enhance learning. Drs. Brad Story and Ingo Titze have already begun planning an exercise whereby the user can incrementally add fluctuations (such as jitter and shimmer) to simulated voice samples and hear the results. These audio samples will allow the vocologist to better train his/her ear for clinical observations.

Because it is still under construction, "How Humans Speak, Sing, Squeak and Squeal" is not linked to other web sites, nor have we requested search engines (such as Yahoo and Alta Vista) to index our site. Stay tuned. In approximately one year, we will share the URL with our colleagues (such as this report's readership) to solicit feedback. These comments and suggestions will be integrated, and the completion date is scheduled for January 2000.

Other investigators are using the Internet as well. Dr. Michael Karnell continues to manage a very active list server for professionals interested in voice and voice disorders. It is co-sponsored by the American Speech-Language-Hearing Association Special Interest Division 3 and The University of Iowa Department of Otolaryngology-Head and Neck Surgery. It continues to provide a valuable avenue for clinicians and others to share ideas about voice disorders, clinic instrumentation, therapies, and other relevant topics. Currently, there are well over 500 subscribers.

At the same time, the general NCVS website is undergoing a major overhaul. In addition to a facelift to the homepage, many of the NCVS publications as well as new information bytes will be added. Sections will be made for the general public interested in voice, for practitioners who routinely work with voice clients, and for prospective trainees interested in careers in oral communication.

While the Internet provides an exciting way to teach, NCVS researchers continue to utilize traditional means of providing continuing education: written materials, conferences, and video. More than 35,000 speech-language pathologists, otolaryngologists and other voice professionals were reached by these methods in the past year.

Bodymind & Voice: Foundations of Voice Education, co-published by The VoiceCare Network, NCVS and The Voice Center of Fairview, was released in December 1997. The 650-page volume integrates practical, scientific, medical-therapeutic and philosophical foundations of voice education. Its readership includes music educators, choral conductors, singing teachers, church musicians, singers, students and speech trainers. It was written in the language of those with little or no background in the voice and learning sciences, but also contains sections "For Those Who Want to Know More." *Bodymind & Voice* features 17 authors, approximately 100 "Do This" experiences for the reader, over 25 pitch patterns, more than 150 illustrations (including 44 color videostroboscopic images), and 54 chapters organized into five books. Major topic areas are: human learning; vocal mechanisms; vocal health; lifespan voice development; and voice education methods. As of June 1998, more than 350 books have been distributed.

The following is from a review of *Bodymind & Voice: Foundations of Voice Education* by Reid Spencer, Memorial University of Newfoundland. It will appear in the forthcoming issue of *Canadian Music Educator,* Vol. 39/3:

"...before I had finished the first chapter, let alone the first *Book of Bodymind & Voice: Foundations of Voice Education,* my respect had been earned, and my attention, through all five books, never once wavered. At the risk of overstatement, I believe this to be the most important new publication in Vocal Pedagogy in the last twenty years."

An earlier NCVS produced publication, *Lee Silverman Voice Treatment: A Practical Guide for Treating The Voice and Speech Disorders in Parkinson Disease,* has now been distributed to more than 1,100 speech-language pathologists. Dr. Lorraine Ramig and her colleagues, Annette Pawlas and Stefanie Countryman, wrote the guidebook for speech-language pathologists so that they may administer this effacacious method to their clients with Parkinson disease. In addition to the book, the group continues to present two-day workshops to teach the method to speech-language pathologists in numerous sites. Presentations over the past year have been given in: Boston; Canton and Dayton, Ohio; Chicago; Dallas and Tyler, Texas; Denver; Edison, New Jersey; Milwaukee; Towson, Maryland; New York City; Portland; Tucson; and West Palm Beach in the U.S. International sites are Sweden, Belgium and Germany. The publication has been so well-received that speech-language pathologists in Europe have contacted Dr. Ramig to request permission to translate the guidebook into other languages.

Dr. Katherine Verdolini has written a second, expanded edition of the booklet, *A Vocologist's Guide to Voice Therapy and Treatment.* Nearly all of the first press run of 10,000 copies have been distributed to speech-language pathologists and students interested in vocology. The original guide presented side by descriptions of seven clinical

voice therapy techniques and four popular voice training techniques from theatre. The revised guide will include the same information with updated references as well as a listing of the disorders treated by these therapy methods. A second guide is also planned for lay persons newly diagnosed with voice disorders. This booklet will contain simplified explanations of the therapies, but will also include a comprehensive section on common voice disorders. The guide for voice clients will also contain a simple description of how the voice works.

Dr. Ingo Titze continues to contribute bi-monthly columns about voice research to *Journal of Singing*, the official publication for members of the National Association of Teachers of Singing. These essays translate scientific research into practical and understandable concepts relevant to choral conductors, music educators and voice trainers. Columns since the last report are: *Noise in the Voice; Normal Modes of Vibration, Are the Corner Vowels Like Primary Colors?; Five Ingredients of a Physiologically Gifted Voice,* and *On the Springiness and Stickiness of Vocal Fold Tissues.*

The fifth in the biennial series of Phonosurgery Symposia is scheduled for July 9-11, 1998, on the University of Wisconsin-Madison campus. Meeting chairs are Drs. Diane Bless and Charles Ford. Approximately 75 physicians, speech-language pathologists and other allied health professionals have registered as this report is submitted. The three-day meeting was created to provide continuing education to otolaryngologists, speech pathologists, and professional voice users. The series is designed to improve clinical practice, to help practitioners make decisions about appropriate treatment, and to showcase "How I do it" from experts in vocology and phonosurgery. Specific objectives are:

- How to treat special populations: pediatrics, geriatrics, functional disorders, singers and patients with other medical problems;
- How to apply specific behavioral management techniques;
- How and when to apply specific surgical techniques;
- How to avoid pitfalls when applying theory to practice;
- How to interpret and when to use voice laboratory results;
- How to identify particular issues that have an impact on financial considerations such as billing, CPT codes, and third party payments; and
- How to improve your practice with outcome measures.

Program sponsors are the Division of Otolaryngology-Head and Neck Surgery, Department of Surgery, and Continuing Medical Education, University of Wisconsin Medical School; UW-Extension; the NCVS; and Meriter Hospital-Park.

Dr. Patricia Zebrowski collaborated on a newly-released videotape for speech-language pathologists, teachers, parents and physicians. The tape, entitled "The School-Age Child Who Stutters," was produced and is sold by The Stuttering Foundation of America. More than 10,000 copies have now been distributed.