NCVS Status and Progress Report
Volume 14/September 1999

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

Editorial and Distribution Information

Editor, Ingo Titze
Production Editor, Julie Stark
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

Mona Abaza, M.D.
Fariborz Alipour, Ph.D.
David Berry, Ph.D.
Alyson Beyler, B.S.
Florence Blager, Ph.D.
Diane Bless, Ph.D.
Myrna Burt
John Canady, M.D.
Thea Carruth, M.P.H.
Roger Chan, Ph.D.
Geron Coale, M.A.
Christy Dahl, B.A.
Charles Davis, Ph.D.
Hu Ding, M.D., Ph.D.
Mindy Ebert
Anna Edmonds, B.A.
Catherine Emerich, M.S.
Wendy Fick, B.A.
Eileen Finnegan, Ph.D.
John Folkins, Ph.D.
Charles Ford, M.D.
Steven Gray, M.D.
Elizabeth Hammond, M.D.
Henry Hoffman, M.D.
Richard Hurtig, Ph.D.
Michael Karnell, Ph.D.
David Kuehn, Ph.D.
Russel Long, M.S.
Erich Luschei, Ph.D.
Daniel McCabe, D.M.A.
Martin Milder, B.S.
Paul Milenkovic, Ph.D.
Jerald Moon, Ph.D.
Lorraine Olson Ramig, Ph.D.
Julie Ostrem, B.S.
Donald Robin, Ph.D.
Shimon Sapir, Ph.D.
Richard Schmidt, Ph.D.
Nicole Seurer, B.S.
Amy Silver, B.S.
Elaine Smith, Ph.D.
Marshall Smith, M.D.
Martin Spencer, B.A.
Julie Stark, A.A.
Brad Story, Ph.D.
Edie Swift, M.S.
Sue Ann Thompson, Ph.D.
Ingo Titze, Ph.D.
Jane VanDeWiel, B.S.
Katherine Verdolini, Ph.D.
George Woodworth, Ph.D.
Patricia Zebrowski, Ph.D.
Lynn Zimba, Ph.D.

Doctoral Students

Christine Bergan, M.A.
Renee Bogschutz, M.S.
Jay Granier, M.A.
Eric Hunter, M.S.
Melda Kündük, M.S.
Douglas Montequin, M.A.
Helen Sharp, M.S.
Susan Thibeault, M.S.

Postdoctoral Fellows

Michael Edgerton, D.M.A.
Kirrie Ballard, Ph.D.

Visiting Scholars

Anne-Maria Laukkanen, Finland
Niro Tayama, M.D., Japan

Contents

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

Index

Foreword

We have said hello and good-bye to several NCVS investigators in this reporting period. Don Robin has left Iowa to assume a position in San Diego. We wish him well. Kirrie Ballard, one of our post-doctoral fellows, had joined forces with him for a brief period at Iowa, but their on-site collaboration was unfortunately cut short. Two of their contributions are pre-printed in this volume. Shimon Sapir joined Lori Ramig in Denver for less than a year, but was very productive as a guest researcher. He is an author on four preprints featured in this volume. Hu Ding has formally left the University of Utah, but continues to work with Steve Gray at a distance. We wish the best for Shimon and Hu. Finally, to Nelson Roy, it's hello again. Having been a Ph.D. student in Wisconsin with Diane Bless, Nelson has now returned to the NCVS by joining the team at Utah with Steve Gray and Marshall Smith. Nelson is an author of two preprints in this volume.

A unique feature of this volume is an index (by author) of all of the preprints distributed by the NCVS since its organization in 1990. This includes 14 volumes. The volume and page number is given for every article. We remind the readers again that we *do not* wish these preprints to be cited as publications. This would put us in conflict with journal editors, who offer peer review of our work.

We welcome comments and suggestions on any of the manuscripts.

Ingo R. Titze, Director
September, 1999

Part I

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 14
September 1999, 1-9

An Investigation of Cricoarytenoid Joint Mechanics Using Simulated Muscle Forces

David Berry, Ph.D.
Douglas Montequin, M.A.
Roger Chan, Ph.D.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Henry Hoffman, M.D.
Department of Otolaryngology, The University of Iowa

Abstract

Rotational and translational stiffnesses were calculated for motion around the cricoarytenoid joint. These calculations were enabled through excised larynx experiments on 5 human larynxes. Known forces were applied to the arytenoid cartilage, and three markers were tracked as a function of the applied forces. Assuming rigid body motion, arytenoid translations and rotations were computed for each applied force. Translational stiffnesses were obtained by plotting force versus displacement, and rotational stiffnesses were calculated by plotting torque versus angular rotation. A major finding was that the translational stiffness along the anterior-posterior direction was three times as great the translational stiffnesses in the other two directions. This non-isotropic nature of the stiffnesses may be an important consideration for phonosurgeons who, naturally, wish to avoid subluxation of the cricoarytenoid joint in patients. The computed rotational and translational stiffnesses currently are being implemented in a 2D model of cricoarytenoid joint motion. Because 3D calculations were obtained, these data should also be useful for future 3D models of the cricoarytenoid joint. These stiffness parameters have a vital influence on pre-phonatory glottal shaping, which in turn exerts a major influence on all aspects of vocal fold vibration, including fundamental frequency, voice quality, voice register, and phonation threshold pressure.

Introduction

The pre-phonatory shape of the glottis exerts a major influence on all aspects of vocal fold vibration, including fundamental frequency, voice quality, voice register, and phonation threshold pressure. Several key that which influence glottal shape include the mechanical properties of the laryngeal tissues, muscular contractile forces, and the stiffness of several joints. The present study is an investigation of cricoarytenoid joint (CAJ) mechanics. Specifically, the mechanics of the CAJ are studied as a function of simulated muscle forces.

Many prior investigations on excised larynxes have documented the motion of the arytenoid cartilage about the CAJ (Frable, 1961; von Leden & Moore, 1961; Ardan, 1966; Maue & Dickson, 1971; Sellars & Vaugh, 1981; Sellars & Sellars, 1983; Neumann et al., 1994). Fortunately, most of these studies report similar findings. They agree that the principal motion of the arytenoid cartilage is a rocking about a roughly longitudinal axis of the cricoid cartilage, as well as a sliding motion along this axis. However, most of these studies have been qualitative in nature. Although Sellars & Vaughn (1981) and Neuman et al. (1994) are notable exceptions, former studies have been exclusively kinematic, i.e., CAJ movements were studied without regard for the forces generating such movements.

Tomographic observations (Ardran & Kemp, 1966) on living subjects have also confirmed that the arytenoid rocks on the cricoid facet during vocal fold movement. Recently, Selbie et al. (1998) conducted a quantitative study on the geometry of the CAJ facet surfaces. Based

on these geometries, they also reported that the optimal axis for movement on these surfaces corresponded to a rocking of the arytenoid on the cricoid.

The present investigation is an important extension of former CAJ studies because it quantifies and analyzes CAJ movement as a function of known muscle forces. With this information, effective rotational and translational stiffnesses are computed for the CAJ, which are critical for modeling the pre-phonatory shape of the vocal folds. For example, ongoing studies are have begun to incorporate this information into a 2D biomechanical model of CAJ motion, and future 3D models are also in the planning stages. Consequently, effective rotational and translational stiffnesses are computed for both 2D and 3D models of the CAJ. In the analysis, rigid body motion is assumed. Although rotation and translation of the arytenoid is allowed, any deformation of the arytenoid cartilage (within its physiological range of motion) is assumed to be negligible. This was also the assumption of the Selbie et al. (1998) study, which used CAJ facet geometry to infer motion of the arytenoid cartilage.

Methods

Five male larynges were obtained from the autopsy unit at the University of Iowa Hospitals and Clinics. At the time of death, the males ranged from 44-82 years in age, and 110-237 lbs. in weight. Specific data for each subject is shown in Table 1. No history of voice disorders was known for any subject and no abnormalities were apparent on any larynx.

Immediately after harvest, each larynx was refrigerated and stored in saline solution. Experiments were performed within 5-7 days of harvest. Each larynx was dissected and mounted. Initially, the trachea was shortened to approximately two inches. All extrinsic laryngeal muscles were removed, and the superior portion of the thyroid cartilage was removed to the level of the ventricular folds. During the final phase of dissection, the ventricular folds were removed to allow an unobstructed view of the true vocal folds. Throughout the dissection process, care was taken to leave as much laryngeal tissue intact as possible. However, some extraneous tissue surrounding the muscular process had to removed to place sutures, and simulate the action of laryngeal muscles.

Mounting consisted of fixing both the cricoid and thyroid cartilages. The cricoid cartilage was fixed by placing a rigid plastic tube through the opening of the trachea and into the subglottal space. Screws were placed through the cricoid cartilage to secure the tube. The hard plastic tube was clamped in order to hold the larynx in a vertical position. The thyroid cartilage was sutured to a fixed crossbar, thus fixing its position as well.

Table I.
The Five Males From Autopsy

	Age (years)	Weight (lbs)
Larynx 1	44	110
Larynx 2	56	198
Larynx 3	82	165
Larynx 4	67	196
Larynx 5	56	237
Mean	**61.0**	**181.2**
SD	**14.3**	**47.3**

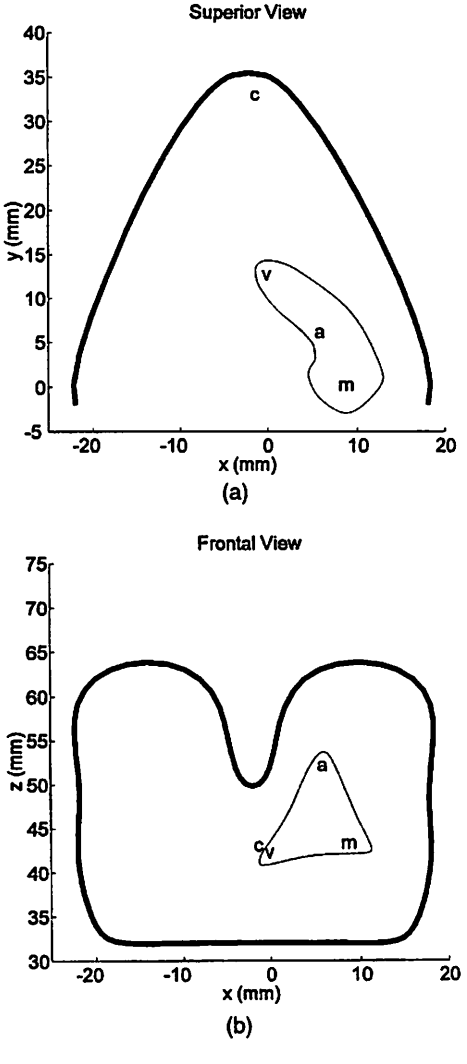


Figure 1. The three markers used to track arytenoid cartilage motion are shown from (a) superior, and (b) frontal views. The markers are denoted by the symbols "a," "v", and "m," which refer to the apex, vocal process, and muscular process, respectively. The symbol "c" indicates the location of the anterior commisure. Rough sketches of the thyroid cartilage (the darker, outer trace) and the aryteniod cartilage (the lighter, inner trace immediately surrounding the three arytenoid markers) are shown for orientation.

In order to document the motion of the arytenoid as a function of simulated muscle forces, three microsutures were placed on the arytenoid as markers. Specifically, a microsuture was placed on each of the following locations: (1) the vocal process, (2) the muscular process, and (3) the apex, as shown in Figure 1. Because the arytenoid was assumed to move as a rigid body, a minimum of three points was needed to define its motion. The positions of the three microsutures were measured before and after application of each of the simulated muscle forces using a 3D digitizer from Immersion Corporation, the MicroScribe-3DX, with a reported resolution of 0.13 mm.

In this investigation, six laryngeal muscle forces were simulated including: two components of the posterior cricoarytenoid (PCA) muscle, with separate horizontal and oblique compartments; the lateral cricoarytenoid (LCA) muscle; two components of the thyroarytenoid (TA) muscle, with separate vocalis and thyrovocalis compartments; and the interarytenoid (IA) muscle. The forces were simulated by attaching one end of a suture to the arytenoid (roughly corresponding to the muscle insertion location), coursing the suture in the direction of the muscle fibers, and attaching the remaining end of the suture to a weight and pulley system. The average directions of these applied forces (averaged over all 5 larynges) is shown in Table II.

The applied forces roughly approximated the magnitude and direction of physiological musculature forces on the arytenoid cartilage, as judged by the maximum active in laryngeal muscles (~100 kPa) multiplied by the cross-sectional areas. In a separate study, precise directions of muscle fiber bundles were quantified (Mineck et al., in press). However, because it was the goal of this study to calculate effective rotational and translational stiffnesses of the CAJ, a variety of muscle forces were applied. The applied forces and resultant displacements were decomposed into rotational and translational components, enabling the calculation of the effective stiffnesses. Through application of a large variety of muscular forces, substantial data were collected from which to infer the stiffnesses. In particular, for each of the 6 muscles, 4-5 force/weight conditions were applied, ranging from zero to a maximum weight, as shown in Table III. In general, muscles with larger cross-sectional areas had larger forces.

Table II.
Average Direction Cosines of the Applied Forces

Muscle	n_x	n_y	n_z
PCA horizontal	-0.696	-0.665	0.270
PCA oblique	-0.396	-0.657	-0.642
LCA	0.026	0.965	-0.261
TA vocalis	-0.086	0.995	0.057
TA thyrovocalis	0.116	0.985	-0.128
IA	-0.921	-0.383	0.068

Table III.
Maximum Weights Used to Simulate Various Muscles

Muscle	Weight (g)
PCA Horizontal	50
PCA Oblique	50
LCA	50
TA vocalis	100
TA thyrovocalis	80
IA	20

Theory

In our analysis, the description of arytenoid motion is performed with rigid body analysis. However, as noted previously, most of the studies of CAJ motion indicate that the arytenoid rocks/slides on the cricoid. Rocking is a specific type of rigid body movement which, qualitatively, might be described as a rotation about a moving axis. Mathematically, rocking is usually implemented as an equation of constraint. In the following analysis of empirical data, although rocking/sliding may be the principal type of motion observed, no attempt is made to *a priori* constrain the arytenoid cartilage to such motion.

Assuming the arytenoid to move as a rigid body, the current position coordinates $\boldsymbol{x}'$ of any point i on the arytenoid may be expressed as a rotation R of the initial position $\boldsymbol{x}_i$ (about the origin) followed by a translation vector $\boldsymbol{D}$:

$$\boldsymbol{x}_i' = R\boldsymbol{x}_i + \boldsymbol{D} \tag{1}$$

If, instead, one desires the rotation to occur about the center-of-mass of the arytenoid, the equation may be written as:

$$\boldsymbol{x}_i' = R(\boldsymbol{x}_i - \boldsymbol{x}_{cm}) + \boldsymbol{x}_{cm} + \boldsymbol{D} \tag{2}$$

For a 2D biomechanical model, the rotation matrix R represents a rotation θ_z about the z-axis and is written as:

$$R = \begin{pmatrix} \cos\theta_z & -\sin\theta_z \\ \sin\theta_z & \cos\theta_z \end{pmatrix} \tag{3}$$

In this 2D model, $\boldsymbol{x}'_i, \boldsymbol{x}_i, \boldsymbol{x}_{cm}$ and $\boldsymbol{D}$ are two-dimensional vectors with x and y components. In a more general three-dimensional model, these vectors are three dimensional and the total rotation R is considered to be three independent rotations about the three coordinate axes, i.e., $R = R_x R_y R_z$, where:

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{pmatrix}$$

$$R_y = \begin{pmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{pmatrix} \quad (4)$$

$$R_z = \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In this experiment, for each force applied to the arytenoid, the initial positions x_i and the final positions x'_i were measured at the three marked locations on the arytenoid (the vocal process, the muscular process, and the apex). The center-of-mass of the arytenoid x_{cm} was also estimated. Thus, for each applied force, only the rotation matrix R and the translation vector D were unknown. By defining $\bar{x}'_i$ to be the average value of the position coordinates of the three arytenoid markers after an applied force, and $\bar{x}_i$ to be the average value of the three arytenoid markers before the applied force, the two-dimensional translation vector D could be expressed, for any angle θ_z, as:

$$D = \bar{x}'_i - R(\bar{x} - x_{cm}) - x_{cm} \quad (5)$$

By computing the translation vector D in this way (see for example, Spoor & Veldpaus, 1980), Equation 2 could be minimized with respect to just one variable, θ_z. Specifically, using the FMIN routine in MATLAB 5.3 Release 11, the mean square difference between the left-hand side and right-hand side of Equation 2 was minimized as a function of θ_z. For the three-dimensional problem, Equation 5 was also used to calculate the translation vector D. However, in this case, Equation 2 required minimization with respect to three variables: θ_x, θ_y, and θ_z. This was done using the FMINS routine in MATLAB (similar to the FMIN routine, except that the FMINS routine allows optimization over more than one variable).

Each force applied to the arytenoid brought it to a new state of static equilibrium, which was quantified by measuring the new position coordinates for the three arytenoid markers. For each state, the sum of all external forces on the arytenoid was set to zero, and the sum of all external torques or moments on the arytenoid was set to zero. For the two-dimensional problem, the CAJ attachment to the arytenoid was modeled as three separate components, including: (1) a linear spring in the x-direction, (2) a linear spring in the y-direction, and (3) a linear torsional spring about the z-axis. Thus, the equations of equilibrium were written as:

$$\begin{aligned} F_{applied,x} - k_x d_x &= 0 \\ F_{applied,y} - k_y d_y &= 0 \\ (r_{attached/cm} \times F_{applied})_z - \kappa_{\theta z}\theta_z &= 0 \end{aligned} \quad (6)$$

Similarly, for the three-dimensional problem, the CAJ attachment to arytenoid was modeled as 6 separate components, with a linear spring and a linear torsional spring associated with each coordinate axis. In particular, the equilibrium equations were written as:

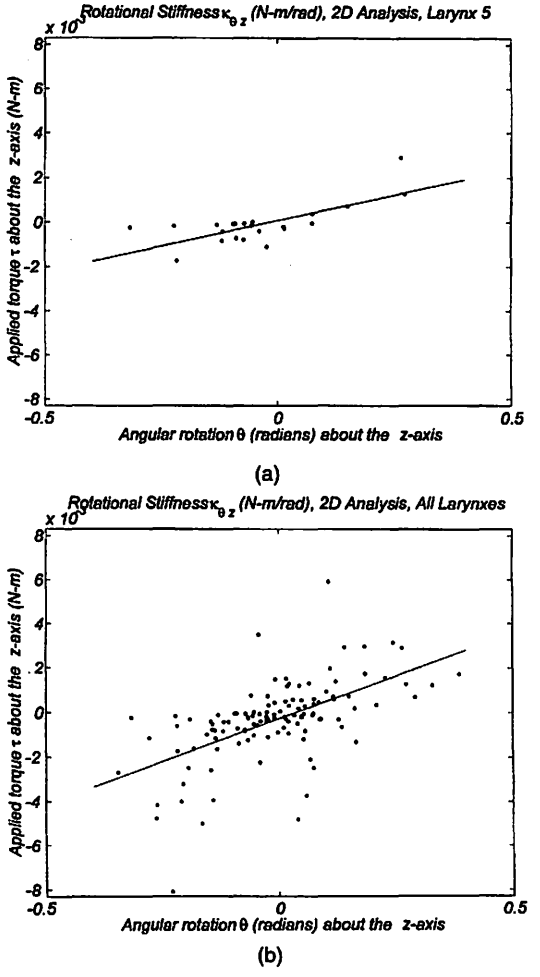


Figure 2. Plots used to calculate the rotational stiffness $K_{\theta z}$ for the 2D model utilizing data from (a) Larynx 5, and (b) all the larynxes.

$$
\begin{aligned}
&F_{applied,x} - k_x d_x = 0 \\
&F_{applied,y} - k_y d_y = 0 \\
&F_{applied,z} - k_z d_z = 0 \\
&(r_{attached/cm} \times F_{applied})_x - \kappa_{\theta x}\theta_x = 0 \\
&(r_{attached/cm} \times F_{applied})_y - \kappa_{\theta y}\theta_y = 0 \\
&(r_{attached/cm} \times F_{applied})_z - \kappa_{\theta z}\theta_z = 0
\end{aligned}
\tag{7}
$$

The vector $r_{attached/cm}$ was measured from the arytenoid center-of-mass to the suture attachment of the applied force to the arytenoid. $F_{applied}$ was the applied force vector and $F_{applied,x}$, $F_{applied,y}$, $F_{applied,z}$, the x, y, and z components of the vector, respectively. The magnitude of the total force was taken as the weight attached to the mass/pulley system. For each applied force, the direction of the force was determined by digitizing two locations: (1) the suture attachment of the applied force to the arytenoid, and a location on the suture about 1 cm away from the attachment. A unit vector capturing the direction of each force was determined by subtracting the first point from the second, and then normalizing. The center-of-mass of the arytenoid was estimated on the basis of the resultant translation and rotation of the arytenoid.

Table IV.
Translational and Rotational Stiffnesses from the 2D Analysis

	k_x (N/m)	k_y (N/m)	$K_{\theta z}$ (Nm/rad)
Larynx 1	93	250	0.0117
Larynx 2	86	225	0.0056
Larynx 3	93	311	0.0103
Larynx 4	82	331	0.0026
Larynx 5	57	146	0.0046
Mean	**82.2**	**252.6**	**0.0070**
SD	**14.9**	**73.6**	**0.0039**

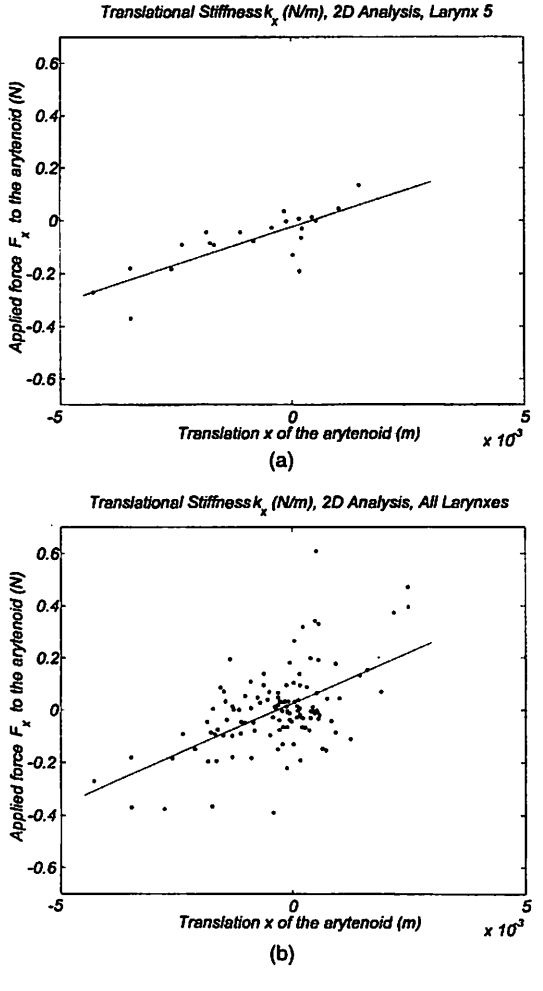


Figure 3. Plots used to calculate the translational stiffness k_x for the 2D model utilizing data from (a) Larynx 5, and (b) all the larynxes.

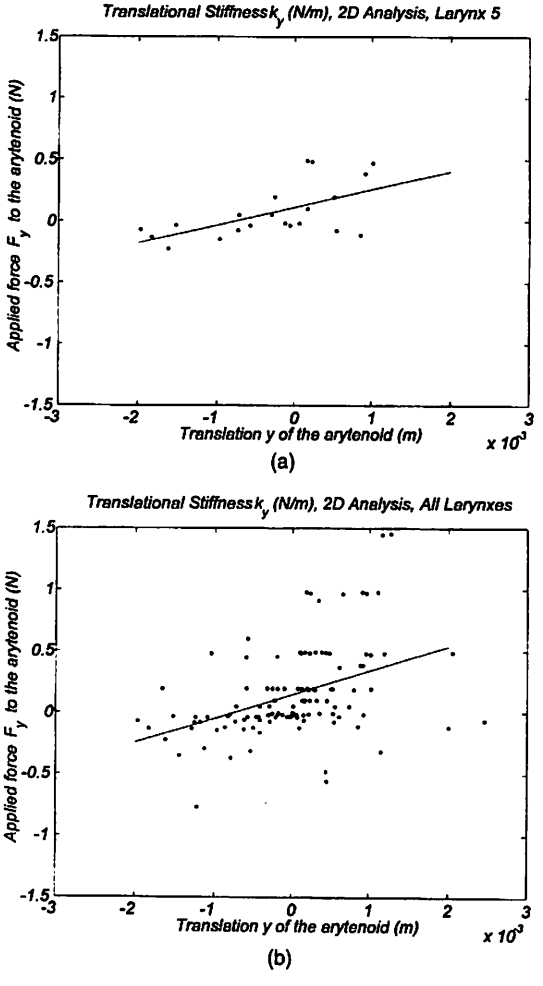


Figure 4. Plots used to calculate the translational stiffness k_y for the 2D model utilizing data from (a) Larynx 5, and (b) all the larynxes.

It may be of some interest to note why it would be necessary to conduct separate optimizations for both 2D and 3D models. Couldn't the 2D model be considered merely a projection of the 3D? In theory, perhaps yes. However, in practice, the optimization problem is different in 2D than in 3D. For example, in the 2D optimization, the z-data is ignored. Thus, because the 2D and 3D models optimize over distinct data sets, the end results could be slightly different.

Table V.
Translational and Rotational Stiffnesses from the 3D Analysis

	k_x (N/m)	k_y (N/m)	k_z (N/m)	$K_{\theta x}$ (Nm/rad)	$K_{\theta y}$ (Nm/rad)	$K_{\theta z}$ (Nm/rad)
Larynx 1	75	267	94	0.0044	0.0039	0.0112
Larynx 2	90	217	132	0.0078	0.0031	0.0054
Larynx 3	92	302	50	0.0042	0.0034	0.0102
Larynx 4	41	156	46	0.0038	0.0055	0.0024
Larynx 5	64	132	34	0.0018	0.0022	0.0049
Mean	**72.4**	**214.8**	**71.2**	**0.0044**	**0.0036**	**0.0068**
SD	**21.0**	**71.8**	**40.9**	**0.0022**	**0.0012**	**0.0037**

Results and Discussion

After decomposing the forces and displacements into angular and translational components, plots of force (or torque) versus displacement (or angular rotation) were made for each of the equations of Equation 6. The translational (k_x, k_y) and rotational Figure 3. Plots used to calculate the translational (K_α) were then computed by determining the slope of the best-fit line through the data with a least-square error procedure, as shown in Figs. 2, 3, and 4 respectively. Figs. 2b, 3b, and 4b show the data from all the larynxes on one plot. Although the data may appear noisy, this is not necessarily the case when the data from a single subject is used, as illustrated in Figs 2a, 3a, and 4a. Thus, much of the noise may be attributed to a relatively large inter-subject variability, which is often a common feature of bio-

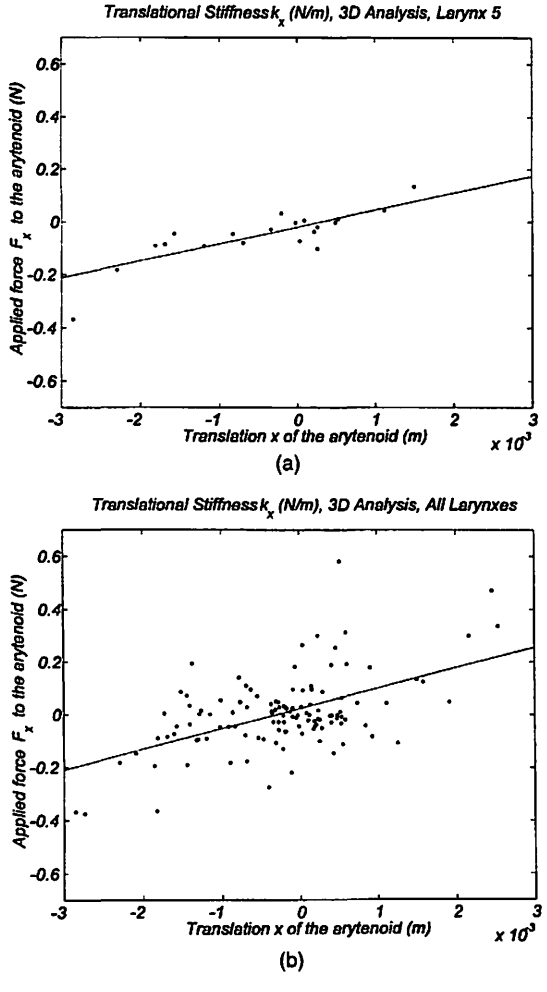


Figure 5. Plots used to calculate the translational stiffness k_x for the 3D model utilizing data from (a) Larynx 5, and (b) all the larynxes. Compare with Figure 3 from the 2D model.

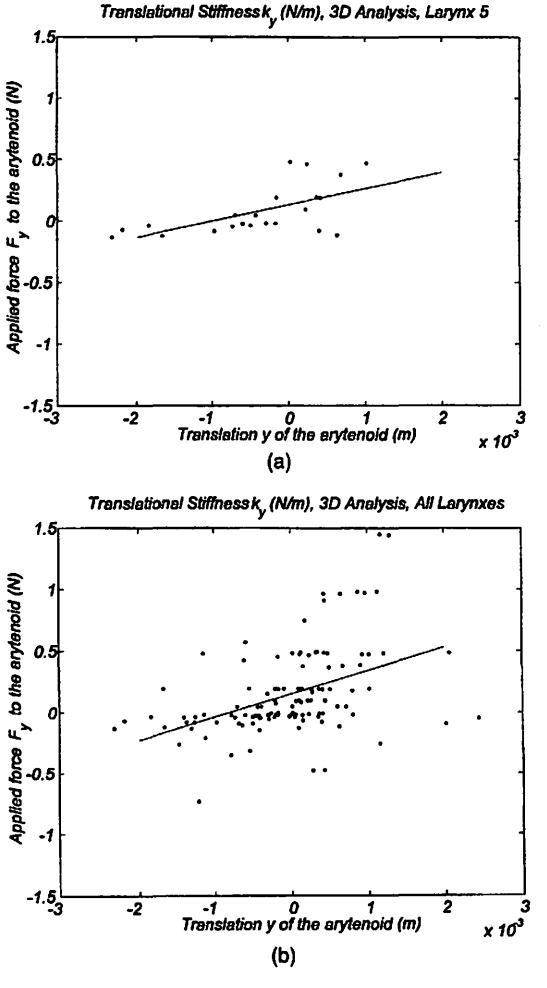


Figure 6. Plots used to calculate the translational stiffness k_y for the 3D model utilizing data from (a) Larynx 5, and (b) all the larynxes. Compare with Figure 4 from the 2D model.

mechanical data. The translational and rotational stiffnesses of each of the 5 larynxes, as well as mean values and standard deviations, are summarized in Table IV.

In these data, perhaps the most noticeable feature is that k_y, the stiffness along the y-axis (the anterior-posterior direction), is roughly three times k_x, the stiffness along the x-axis (the medial-lateral direction). This would seem to imply that, while some arytenoid translation may occur in connection with arytenoid adduction, relatively little arytenoid translation occurs in connection with lengthening and shortening of the folds. Alternately, it could simply mean that larger anterior-posterior forces components are used to generate arytenoid translation along the anterior-posterior length of the folds, about three times as large as in the medial-lateral direction. In either case, our data suggest that medial-lateral gliding is less constrained than anterior-posterior gliding.

Similarly, rotational and translational stiffnesses from the 3D analysis were calculated from the slopes of the force-displacements/torque-angular rotation in Figs 5-10. Again, composite data is shown in Figs. 5b-10b, and an example of single subject data is shown Figs. 5a-10a. The translational and rotational stiffnesses of each of the 5 larynxes, as well as mean values and standard deviations, are summarized in Table V.

In general, the three-dimensional model tells a similar story to the two-dimensional model. In particular, the values of k_x, k_y, and $K_{\theta z}$ are similar for both models, although not identical for the reason mentioned earlier (e.g., z data not included in the 2D optimization). Again, the anterior-posterior translational stiffness, k_y, is about three times greater than the stiffnesses in the other two directions. In terms of rotational stiffnesses, the stiffness about the z-axis, $K_{\theta z}$ is approximately 55-90% greater than the stiffnesses about the other two axes.

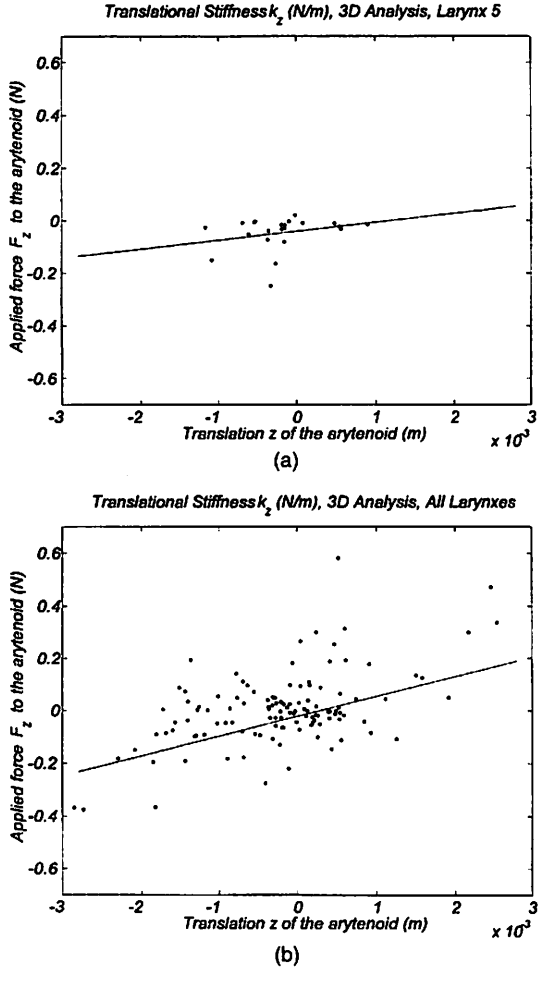


Figure 7. Plots used to calculate the translational stiffness k_z for the 3D model utilizing data from (a) Larynx 5, and (b) all the larynxes.

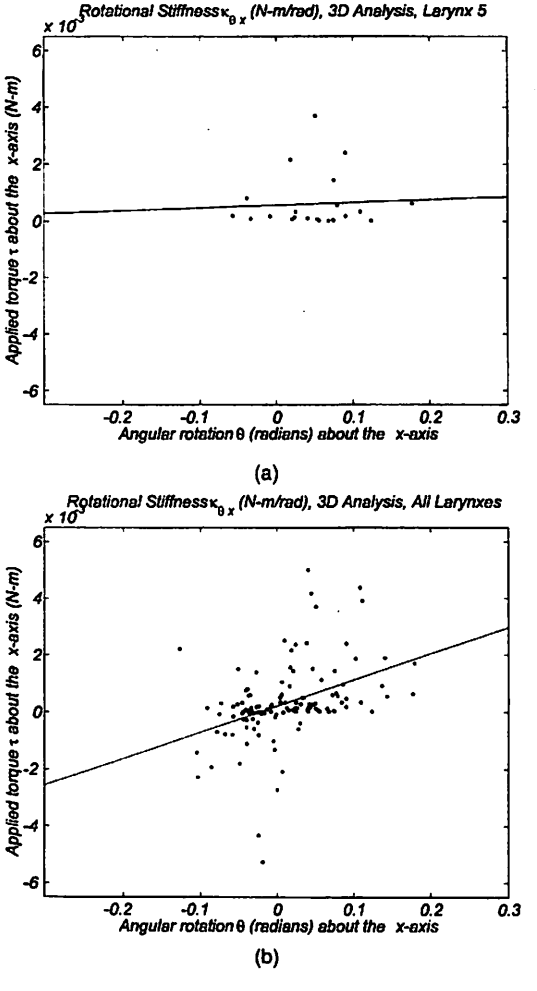


Figure 8. Plots used to calculate the rotational stiffness $K_{\theta x}$ for the 3D model utilizing data from (a) Larynx 5, and (b) all the larynxes.

Summary

Rotational and transitional stiffnesses of the CAJ were calculated for both 2D and 3D models of the CAJ. To calculate these stiffnesses, arytenoid motion was studied as a function of applied forces. Although many previous studies of arytenoid motion have been conducted, this is the first investigation to study such motion as a function of known applied forces. A major finding was that the translational stiffness along the anterior-posterior direction was roughly three times larger than the translational stiffnesses in the other two directions, suggesting that the primary constraint of the ligaments around the CAJ is to keep the arytenoid firmly attached to the cricoid during vocal fold elongation. Medial-lateral and vertical gliding are less constrained. This non-isotropic nature of the stiffnesses may be an important consideration for phonosurgeons who, naturally, wish to avoid subluxation of the CAJ in patients.

These rotational and translational stiffnesses are critical for modeling the pre-phonatory posturing of the vocal folds, which exert a major influence of many aspects of vocal fold vibration. Preliminary investigations yield reasonable results when these CAJ stiffness values are inserted into our models of vocal fold posturing. In particular, laryngeal muscles are able to both adduct and abduct the vocal folds. We note that the effective stiffnesses calculated in this investigation were computed based on a linear fit to the data. However, to implement the data with finite element models, it may be necessary to use higher-order (strain-dependent) stiffnesses. This would allow for less constraint at small displacements and greater constraint at large displacements, a condition often encountered in biomechanics to avoid injury.

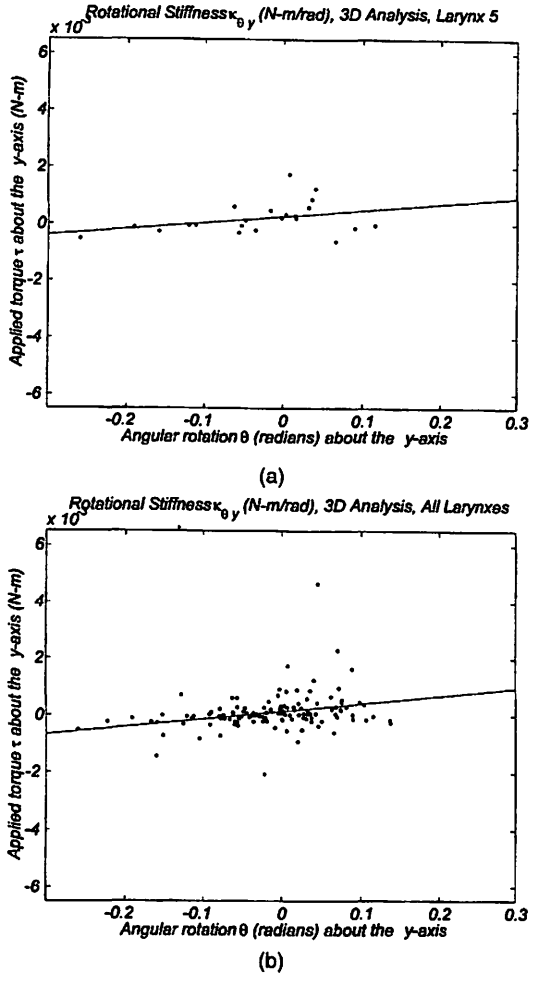


Figure 9. Plots used to calculate the rotational stiffness $\mathbf{K}_{\theta z}$ for the 3D model utilizing data from (a) Larynx 5, and (b) all the larynxes.

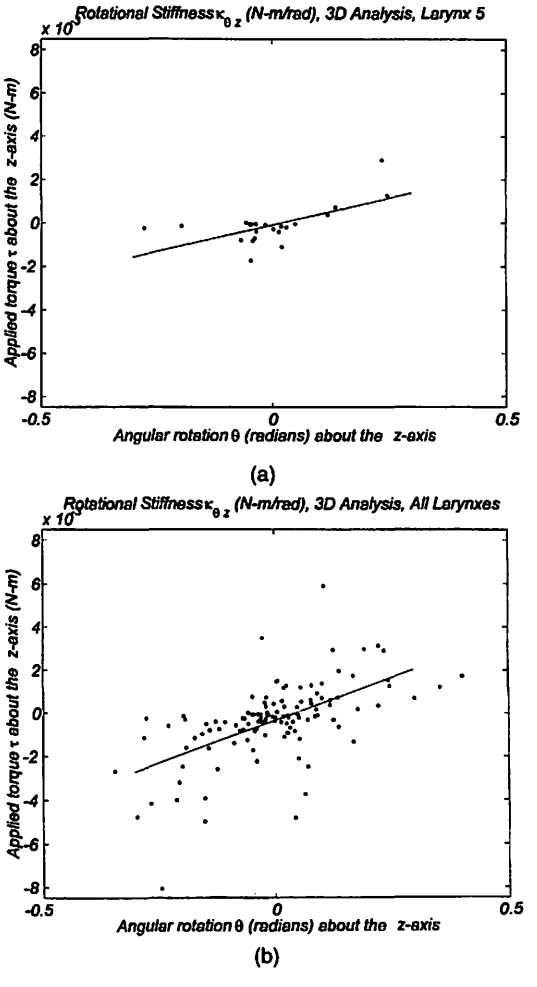


Figure 10. Plots used to calculate the rotational stiffness $\mathbf{K}_{\theta z}$ for the 3D model utilizing data from (a) Larynx 5, and (b) all the larynxes. Compare with Figure 2 from the 2D model.

Acknowledgments

This work was supported by grant no. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. The preliminary investigations of Dr. Mike Bayerl during his postdoctoral fellowship at the University of Iowa were also extremely helpful in the present study.

References

Ardan, G.M., Kemp, F.H. (1966). The mechanisms of the larynx, part I: the movements of the arytenoid and cricoid cartilages, Br. J. Radiol 39, 641-654.

Frable, M.A. (1961). Computation of motion at the cricoarytenoid joint, Arch. Ololaryng. 73, 551-556.

Maue, W.M., Dickson, D.R. (1971). Cartilages and ligaments of the adult human larynx. Arch Otolaryng 94, 432-439.

Mineck, C.W., Tayama, N., Chan, R.W., Titze, I.R. (in press). A three-dimensional anatomical characterization of the canine laryngeal abductor and adductor musculature, Ann. Otol. Rhinol. Laryngol.

Neuman, T.R., Hengesteg, A., Lepage R.P., Kaufman K.R., Woodson, G.E. (1994). Three-dimensional motion of the arytenoid adduction procedure in cadaver larynges. Ann. Otol. Rhinol. Laryngol. 103, 265-270.

Selbie, S.W., Zhang, L., Levine, W.S., and Ludlow, C.L. (1998). Using joint geometry to determine the motion of the cricoarytenoid joint, J. Acoust. Soc. Am. 103, 1115-1127.

Sellars, I.E., Sellars, S. (1983). Cricoarytenoid joint structure and function, Journal of Laryngology and Otology 97, 1027-1034.

Sellars, I.E., Vaughn, C.L. (1981). A biomechanical investigation of cricoarytenoid joint kinematics, in International Series on Biomechanics, Vol. 4A, Biomechanics VIII-A, Japan, (Human Kinetics Publishers, Champaign, IL), pp. 116-124.

Spoor, C.W., Veldpaus, F.E. (1980). Rigid body motion calculated from spatial coordinates of markers, J. Biomechanics 13, 391-393.

von Leden, H., Moore, P. (1961). The mechanics of the cricoarytenoid joint, Arch. Otolaryng. 73, 541-550.

NCVS Status and Progress Report - 14
September 1999, 11-15

The Effect of Cricothyroid Muscle Action on the Relation Between Subglottic Pressure and Fundamental Frequency During Vocal Folds Vibration

Tzu-Yu Hsaio, M.D., Ph.D.
Chia-Ming Liu, M.D.
Department of Otolaryngology, College of Medicine, National Taiwan University
Erich Luschei, Ph.D.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

The relation between subglottal pressure (P_s) and fundamental frequency (F_o) in phonation was investigated with an *in vivo* canine model. Direct muscle stimulation was used in addition to brain stimulation. This allowed the P_s-F_o slope to be quantified in terms of cricothyroid muscle activity. Results showed that, for ranges of 0-2 ma constant current stimulation of the cricothyroid muscle, the P_s-F_o slope ranged from 10 Hz/kPa to 60 Hz/kPa. These results were compared to similar slopes obtained in a previous study on excised larynges in which vocal fold length was varied instead of cricothyroid activation. A physical interpretation is given for the varying slopes in terms of the amplitude-to length ratio of the vocal folds and a dynamic stiffness.

Introduction

It is well known that the fundamental frequency (F_o) of vocal fold vibration is influenced by the subglottic pressure (P_s) during phonation (1-6). An F_o rise with P_s has been consistently observed and shown to be a nonlinear stiffness phenomenon (6). The effective dynamic stiffness of vocal folds increases with vibrational amplitude, which increases with P_s. The interesting result has been that, due to cancellations of flow and tissue nonlinearities, the relation between P_s and F_o has been nearly linear over a wide region of subglottal pressure (6). Thus, the change of F_o with P_s has been describable by a series of slopes (ranging from 10 to over 100 Hz/kPa) as reported in the literature, whether in human subjects (2-5,7,8), excised larynges (6), or an *in vivo* animals (9). Theoretically, the slope was shown to be affected by different pre-phonatory vocal fold lengths (6). More specifically, the amplitude-to-length ratio (*A/L*) determined the slope. Shorter vocal folds and greater amplitudes of vibration caused F_o to be more dependent on P_s. This has been verified in the experiment using an excised-larynx model.

The results have yet to be verified in an *in vivo* canine model with active laryngeal muscle tensions. We will show here that the degree to which F_o changes with P_s is quite dependent on CT activity, the determinant of vocal fold length. Recent reports on human subjects reveal that the effects of transglottic pressure on the F_o of phonation in different pitches are not monotonic (7,8). They attribute the results mainly to the different combinations of cricothyroid (CT) and thyroarytenoid (TA) muscles action, but these actions were not specifically controlled. Inasmuch as CT action alone can elongate and stiffen the vocal folds under constant TA action, we will control CT action separately by independent stimulation. In the *in vivo* canine evoked phonation model, which we will use, consistent phonation with similar (and constant) patterns of muscular activity in the larynx can be elicited repeatedly more than 50 times (9,10). Each repeated evoked phonation provides coordinate action pattern of the laryngeal muscles with little variability. Specific and variable CT activation can then be superimposed. Under these repeated conditions, we will show that a greater degree of CT action results in precise changes of the P_s-F_o slope. We will quantify this slope change and relate it to previous theoretical findings.

Materials and Methods

The *in vivo* evoked phonation canine model used in this experiment followed the previously reported procedures (9,10). In brief, hound-like mongrel dogs weighing approximately 20 kg were used in this study. After being anesthetized with pentobarbital (25 mg/kg intravenously) to the surgical level, the animals were placed on a stereotaxic apparatus. The cortex was exposed at about 10 mm anterior and 5 mm lateral to the ear-bar-zero. A coaxial bipolar electrode (Rhodes NE-100) was inserted into the brain, 20 mm dorsal to ear-bar – zero, using a stereotaxic manipulator.

The animal was rotated from a prone to a supine position to expose the ventral side of the neck. The surgical preparations in the neck included the exposure of the larynx and trachea, insertion of bipolar hooked-wire electrodes into CT, TA, lateral cricoarytenoid (LCA) and posterior cricoarytenoid (PCA) muscles for electromyographic (EMG) recordings. Electrical stimulation to sites in the midbrain was then delivered through the bipolar electrode to elicit vocalization. The site of the electrode was moved at 1 mm intervals to seek a proper location for elicitation of a low-pitched, stable phonation. Brain stimuli consisted of a 2 or 3 s train of 0.2ms pulses at a rate of 200Hz under low current levels (about 0.5mA). Once the site for electrical stimulation in the midbrain was selected, it was kept constant throughout the experiment.

A tracheostomy was performed for placing two low-pressure cuffed canulae into the trachea. One canula was for respiration (the caudal one) and the other was connected to an air source supplying humidified and warmed air to the larynx for phonation (the rostral one). The bilateral superior laryngeal nerves were identified and resected at the segment between the internal and external branches. Two pairs of stimulation electrodes were inserted bilaterally into the cricothyroid muscles for direct muscle fiber stimulation (Fig 1). As every train of stimuli was delivered to the midbrain, about 1.5 s of sustained phonation was obtained with a consistent pattern of the laryngeal muscle action. The evoked phonation was repeated under different levels of subglottic pressure and various degree of additional electrical stimulation to the CT muscle. The electrical stimulations to the CT, for various degrees of muscle action, were controlled by variable current stimuli of 2 ms pulses at the rate of 50 Hz during the period of evoked phonation. The current was set at 0.0, 0.5, 1.0 and 2.0 mA to simulate 4 different degrees of CT action. The constant air pressure supplying the rostral tracheal canula for phonation was controlled with a pressure regulating valve (Fairchild model 10), and was measured with a pressure transducer (Micro Switch 143PC03G). The air pressure delivered to the larynx during evoked phonation at a certain degree of CT action was increased in increments of 0.1 kPa from the lowest to highest pressure that could sustain phonation in the modal register.

The EMG recordings from laryngeal muscles (CT, TA, PCA and LCA), the subglottic pressure, voice signals, and midbrain stimulation marker were recorded on an 8-channel DAT data recorder (TEAC RT-130D) for further analysis. For this report, only the CT changes are of interest and will be reported. The signals were analyzed using DATAQ and CODAS signal processing hardware and software on a personal computer.

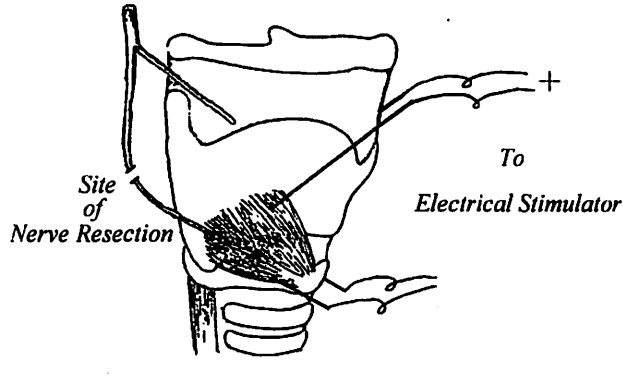


Figure 1. Lateral view of larynx showing site of electrical stimulation.

Results

Three animals in this experiment completed the collection of F_o-P_s data with at least three different degrees of CT muscle action. The subglottic pressure ranged from approximately 0.55 to 3.72 kPa, a range that is quite typical for human phonation under soft and very loud conditions. The F_o-P_s data pairs of these three animals are plotted in Figures 2, 3, 4. The different symbols represent various levels of electrical stimulation current to the CT muscle. The

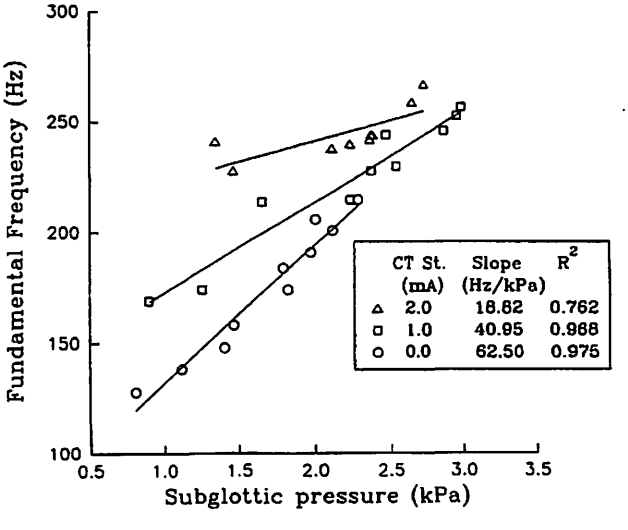


Figure 2. Fundamental frequency versus subglottal pressure for three levels of cricothyroid stimulation (canine 1).

solid lines are the linear regression for each data set. The regression coefficients (R^2) for the data sets were between 0.8 and 1.0, the better fit being for the data groups with smaller electrical current stimulation to CT. These results reflect the fact that F_o changed linearly with P_s under various degrees of CT muscle action. The slope of the regression lines ranged from 9.79 to 62.5 Hz/kPa. The values became smaller as the CT muscle action increased in each of the animals' data. The F_o in all the evoked phonations in this experiment, with or without CT stimulation, ranged from 95 Hz to 277Hz. The register was perceived (informally) to be modal in all cases, although it is sometimes difficult to assign a register to animal phonation.

Analysis and Discussion

To interpret the results, a simple mathematical model was revisited from previous investigations on excised larynges (6). Fundamental frequency was computed by a "thick string" formula with active and passive fiber properties (14),

$$F_o = \frac{1}{2L}\sqrt{\frac{\sigma_p}{\rho}\left(1 + \frac{d_a}{d}\frac{\sigma_m}{\sigma_p}a_{TA}\right)}^{1/2} \quad (1)$$

where L is the vocal fold length, σ_p is the passive tissue stress (in the anterior-posterior tissue fiber direction, σ_m is the maximum active stress (100 kPa), ρ is the tissue density (1.03 g/cm3), d is the depth of the vocal fold in the lateral direction (5 mm), $d_a = a_{TA}d$ is the active (muscular) portion of the depth, and a_{TA} is the thyroarytenoid muscle activity (ranging from 0.0 to 1.0). The vocal fold length was computed as

$$L = L_o(1 + \varepsilon) \quad , \quad (2)$$

where L_0 is the resting length and ε is the vocal fold strain, which has a static (postural component ε_s and a dynamic component ε_d. The static component was determined by the muscle activities according to a previously determined empirical relation (14),

$$\varepsilon_s = 0.2(2a_{CT} - a_{TA}) - 0.4 \quad , \quad (3)$$

were a_{CT} is the normalized activity of the CT muscle (ranging from 0.0 to 1.0. The constants were determined experimentally by recurrent and superior nerve stimulation in anesthetized dogs (14,15).

The dynamic strain was shown theoretically to have an amplitude and length dependence (6),

$$\varepsilon_d = \frac{1}{8}\pi^2\left(\frac{A}{L}\right)^2 \quad (4)$$

where A is the vibrational amplitude at the mid-membranous vocal fold. The total strain is then the summation of the two component strains

$$\varepsilon = \varepsilon_s + \varepsilon_d \quad , \quad (5)$$

which determines the passive tissue stress according to an empirically-fitted exponential curve for vocal fold tissue (16)

$$\begin{aligned} \sigma_p(\varepsilon) &= 0 && kPa \quad for\ \varepsilon < -0.5 \\ &= 0.1(\varepsilon + 0.5) && kPa \quad for\ -0.5 < \varepsilon < -0.4 \\ &= 0.1(\varepsilon + 0.5) + 35[e^{2.5(\varepsilon+0.4)} - 2.5(\varepsilon + 0.4) - 1] && kPa \quad for\ \varepsilon > -0.4 \end{aligned} \quad (6)$$

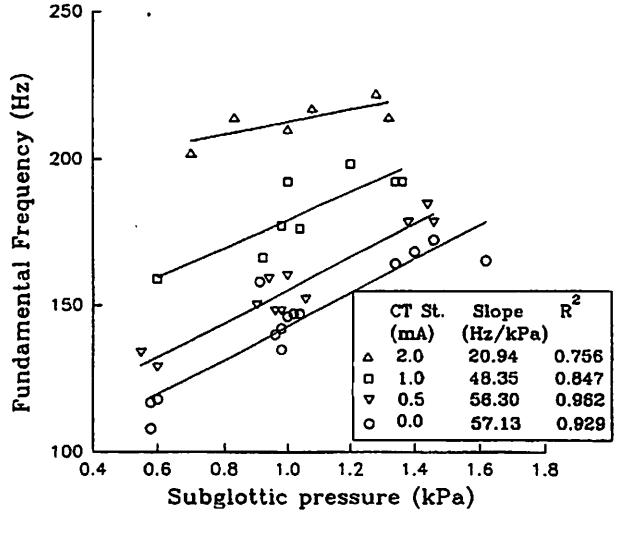


Figure 3. Fundamental frequency versus subglottal pressure for three levels of cricothyroid stimulation (canine 2).

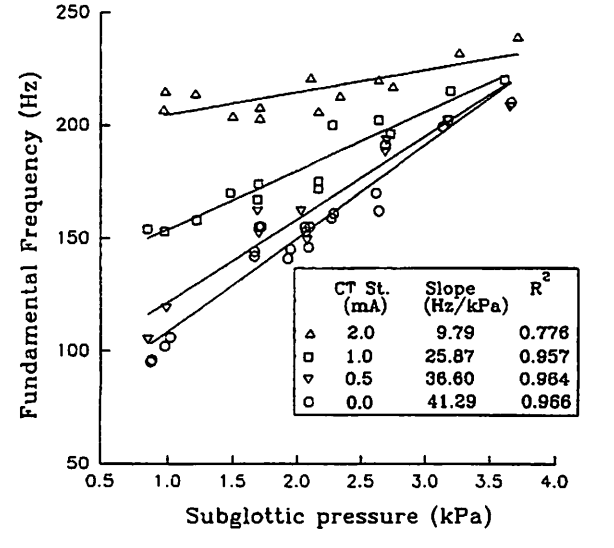


Figure 4. Fundamental frequency versus subglottal pressure for three levels of cricothyroid stimulation (canine 3).

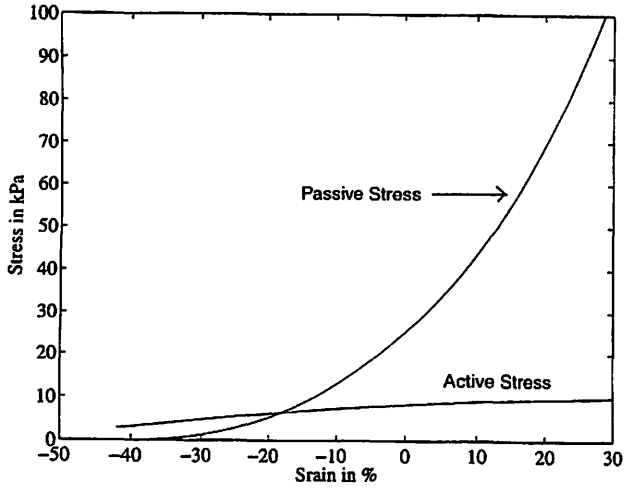


Figure 5. Stress-strain for passive and active tissue in the model.

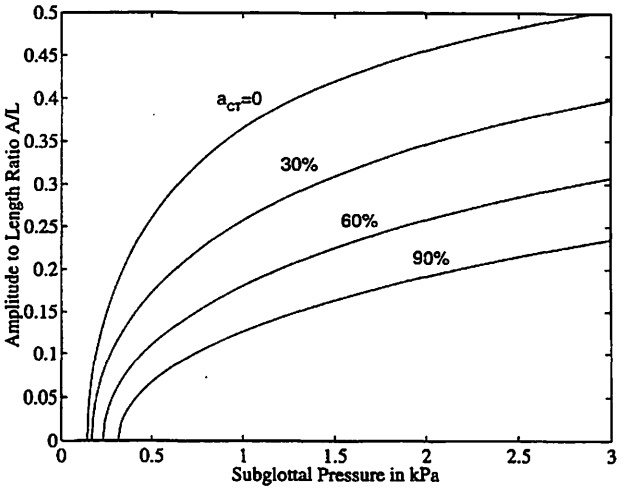


Figure 7. Amplitude-to-length ratio versus subglottal pressure for four levels of cricothyroid muscle activity a_{CT} in the model.

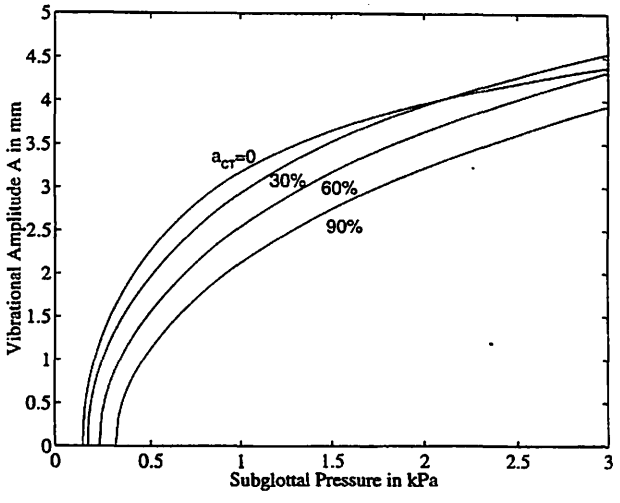


Figure 6. Amplitude of vibration versus subglottal pressure for four levels of cricothyroid muscle activity a_{CT} in the model.

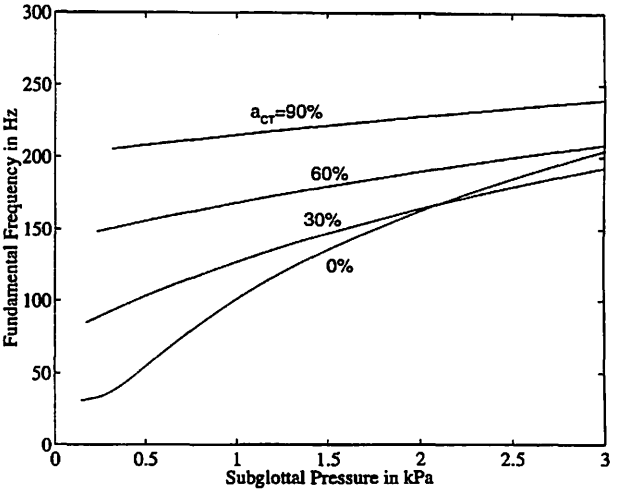


Figure 8. Fundamental frequency versus subglottal pressure for four levels of cricothyroid muscle activity a_{CT} in the model.

This exponential curve is plotted in Figure 5, along with a small active stress (for $a_{TA} = 0.1$). It represents an average of canine vocal fold tissues (mucosa and muscle) as determined by measurement (16).

The final two relations needed to complete the model are

$$A = 0.1L_o\left(\frac{P_s - P_{th}}{P_{th}}\right)^{1/2} \qquad (7)$$

where P_{th} is the phonation threshold pressure (17),

$$P_{th} = 0.14 + 0.06\left(\frac{F_o}{120}\right)^2 \; kPa \quad , \qquad (8)$$

Figure 6 shows the vibrational amplitude versus subglottal pressure for four values of a_{CT} (0%, 30%, 60%, and 90%). Thyroarytenoid muscle activity (a_{TA}) was held at a constant value of 0.1 (10%). Note that vibrational amplitude A decreases with increased a_{CT}. The intercepts on the horizontal axis are the phonation threshold pressures (P_{th}) for different values of a_{CT}. The P_{th} intercept increases with a_{CT} because F_o increases with a_{CT} according to Equation 8.

The length of the vocal fold also increases with a_{CT}, as Equations 2 and 3 indicate. This makes the amplitude-to-length ratio (A/L) highly sensitive to a_{CT}. Figure 7 shows this A/L ratio plotted as a function of P_s for different values of a_{CT}. It is this ratio that governs the slope between F_o and P_s *vis a vis* the dynamic strain in Equation 4. Finally, Figure 8 illustrates the resulting F_o versus P_s curves. For a_{CT}

= 90%, the slope is only about 10 Hz/kPa, while for $a_{CT} = 0$ the slope is 60 Hz/kPa. This agrees well with the measurements reported here, although it is difficult to assign a percent of maximum stimulation to the currents used for direct muscle stimulation in the experiment. Also, the exact value of a_{TA} was not known. Nevertheless, the similarity between the curves is compelling.

Conclusion

The relation between F_o and P_s is nearly linear when TA activity is held constant and CT activity is varied. This has been shown by using an *in vivo* canine model with active laryngeal muscle action (9,11,12,13). In this series of three animal studied, the linear relation between F_o and P_s was confirmed under different levels of CT action.The wide range of slopes is attributed to the amplitude to length ratio (*A/L*) of the vocal folds. In order to increase F_o, the vocal folds must be lengthened to increase the tension. Under this tension, the amplitude of vibration becomes smaller, dramatically lowering the *A/L* ratio. Kitajima and Tanaka (7) claimed that the F_o-P_s relation at different F_o is not monotonic. They discuss several reasons why F_o-P_s relation may have varied. They mainly attribute the results to different combinations of CT and TA action at different F_o. It is true that in human subjects the increasing F_o of chest voice needs concomitant action of CT and TA (14). Therefore, due to the different levels of TA action, the mechanical characteristics, including the length and tension of vocal folds, may not follow the presumed simple rules given here. In this study, there was a relatively constant TA action during repeated evoked phonations in the *in vivo* canine model (9,10). Under these circumstances, different degree of CT action gave lengths and stiff nesses of the vocal folds as predicted by theory.

Acknowledgment

Support for this research was provided in part by Grant NSC89-2314-B-002-232 from National Science Council, Taiwan, and Grant P60 DC00976 from the National Institutes of Health.

References

1. Isshiki N. Regulatory mechanism of the pitch and volume of voice. Oto-Rhino-Laryngology (Kyoto) 1959;52:1065-94.

2. Lieberman P, Knudson R, Mead J. Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. 1969;45:1537-43.

3. Hixon TJ, Klatt DH, Mead J. Influence of forced transglottal pressure on fundamental frequency. J Acoust Soc Am 1971;49:105.

4. Baer T. Reflex activation of laryngeal muscles by sudden induced subglottal pressure change. J Acoust Soc Am 1979;65:1271-5.

5. Rothenberg M, Mahshie J. Induced transglottal pressure variations during voicing. J Phono 1986;14:365-71.

6. Titze I. On the relation between subglottal pressure and fundamental frequency in phonation. J Acoust Soc Am 1989;85:901-6.

7. Kitajima K, Tanaka K. The effects of intraoral pressure change on F0 regulation—Preliminary study for the evaluation of vocal fold stiffness. J voice 1995;9:424-8.

8. Tanaka K, Kitajima K, Kataoka H. Effects of transglottal pressure change on fundamental frequency of phonation: Preliminary evaluation of the effect of intraoral pressure change. Folia Phoniatr Logop 1997;49:300-7.

9. Hsiao T-Y, Solomon NP, Luschei ES, Titze IR, Liu K, Fu T-C, Hsu M-M. Effect of subglottic pressure on fundamental frequency of the canine larynx with active muscle tensions. Ann Otol Rhinol Laryngol 1994;103:817-21.

10. Hsiao T-Y, Fu T-C, Tan C-T, Lee S-Y. An *in vivo* canine model for the study of phonation physiology by midbrain stimulation. J Formos Med Assoc 1994;93:475-80.

11. Hirano M. Morphological structure of the vocal cord as a vibrator and its variations. Folia Phoniatr 1974;26:89-94.

12. Isshiki N. Phonosurgery Theory and practice. Tokyo: Springer-Verlag, 1989;1-59.

13. Moore DM, Berke GS. The effect of laryngeal nerve stimulation on phonation: A glottographic study using an *in vivo* canine model. J Acoust Soc Am 1988;83:705-15.

14. Titze IR, Jiang J, Druker DG. Preliminaries to the body-cover theory of pitch control. J Voice 1988;4:314-9.

15. Titze IR, Jiang, J, Lin E. The dynamics of length change in canine vocal folds. J Voice 1997;11(3):267-276.

16. Alipour-Haghighi F, Titze IR. Elastic models of vocal fold tissues. J Acoust Soc Amer 1991;90(3):1326-1331.

17. Titze IR. The physics of small-amplitude oscillation of the vocal folds. J Acoust Soc Am 1988; 83(4):1536-1552.

NCVS Status and Progress Report - 14
September 1999, 17-26

A Three-Dimensional Anatomical Characterization of the Canine Laryngeal Abductor and Adductor Musculature

Corey Mineck, M.D.
Department of Speech Pathology and Audiology, The University of Iowa
Niro Tayama, M.D.
Department of Speech Pathology and Audiology, The University of Iowa
Department of Otolaryngology, University of Tokyo, Tokyo, Japan
Roger Chan, Ph.D.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

The biomechanics of vocal fold abduction and adduction during phonation, respiration, and airway protection has yet to be completely understood. Specifically, the rotational and translational forces on the arytenoid cartilages resulting from intrinsic laryngeal muscle contraction have not been fully described. Anatomical data on the lines of action and moment arms for the intrinsic laryngeal muscles are also lacking. This study was conducted to quantify the three-dimensional orientations and the relative cross-sectional areas of the intrinsic abductor and adductor musculature of the canine larynx. Eight canine larynges were used to evaluate the three muscles primarily responsible for vocal fold abduction and adduction: the posterior cricoarytenoid (PCA), the lateral cricoarytenoid (LCA), and the interarytenoid (IA) muscles. Each muscle was exposed and divided into discrete fiber bundles whose coordinate positions were digitized in three-dimensional space. The mass, length, relative cross-sectional area and angle of orientation for each muscle bundle were obtained, allowing for the calculations of average lines of action and moment arms for each muscle. This mapping of the canine laryngeal abductor and adductor musculature provides important anatomical data for use in laryngeal biomechanical modeling. These data may also be useful for surgical procedures such as arytenoid adduction.

Introduction

The physiology and biomechanics of vocal fold abduction and adduction during phonation, respiration, and airway protection (e.g. in swallowing) have not been studied completely. Traditionally, it has been assumed that each laryngeal muscle functions as a unit, with a single direction of action. However, recent research has demonstrated that at least some of the intrinsic laryngeal muscles are composed of different compartments that may function independently. For example, Sanders et al. showed that the posterior cricoarytenoid (PCA) muscle has two to three separate bellies that may each function as an independent unit during different types of abduction.[1,2,3,4] Sanders and his coworkers also showed that the innervation of the human PCA muscle stems from two separate nerve branches to supply two different compartments, further lending support to independent function.[3,4] Therefore, it was the purpose of this study to characterize the three-dimensional orientations and relative cross- sectional areas of the laryngeal abductor and adductor muscles for the purpose of biomechanical modeling. The canine larynx was chosen because the active contractile properties of some canine laryngeal muscles had previously been measured.[5,6,7] Because it is practically difficult to measure the active properties of viable human laryngeal muscles, the canine offers a model similar in laryngeal muscle morphology, anatomy, and possibly function.

The motion of the arytenoid cartilage on the cricoarytenoid joint (CAJ) is complex and has been examined extensively.[8,9,10,11,12,13,14] Classically, arytenoid motion on the

CAJ has been described as a rotation around a "vertical" axis.[15] However, many anatomical and vocal fold kinematic studies have suggested that the two major arytenoid motions are (1) rocking around the longitudinal axis of the CAJ facet, in a somewhat anterior-posterior direction, and (2) sliding along this axis, in a somewhat medial-lateral direction.[8,10,11,13,14] Selbie et al. showed that the rocking axis could be congruent with the classical "vertical" axis, such that the rocking motion is often perceived as a rotation under the perspective of an endoscope.[14] Current beliefs suggest that the arytenoid motion is critically determined by several major factors, including the geometry of the CAJ facets, the anatomy of the fibroelastic connective tissues in the synovial joint, and the actions of the laryngeal abductor and adductor muscles.[16]

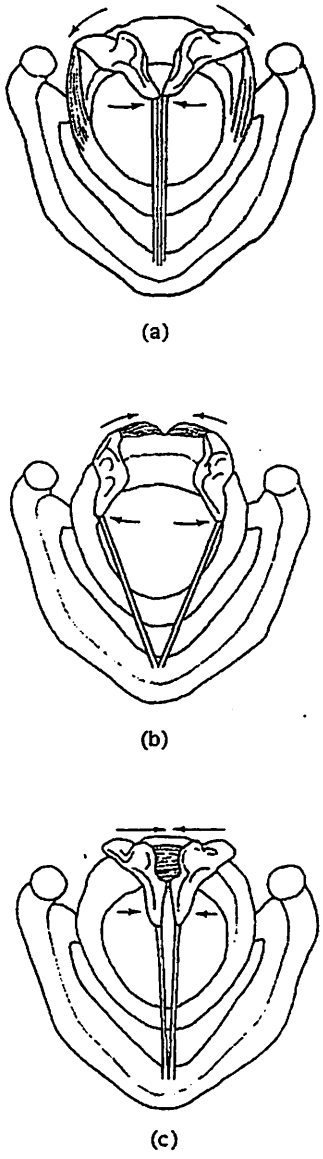


Figure 1. Classical descriptions of arytenoid motion and vocal fold abduction/adduction associated with the contraction of (a) the lateral cricoarytenoid (LCA) muscle, (b) the posterior cricoarytenoid (PCA) muscle, and (c) the interarytenoid (IA) muscle. [After Netter, FH (1997). Atlas of human anatomy (2nd edn.) East Hanover, NJ: Novartis. Used with permission]

The laryngeal abductor and adductor muscles are believed to have rapid contraction rates relative to the other respiratory and laryngeal muscles.[17] Tetanic contraction times have been estimated to be on the order of 10 msec,[17] in comparison to around 50 msec for the thyroarytenoid muscle and around 90 msec for the cricothyroid muscle.[5,7] With such rapid contraction times, these muscles play a primary role in rapid opening of the glottis during inspiration, and in rapid closure of the glottis for protection against foreign body inhalation (e.g. during swallowing).

The PCA muscle originates at the posterior surface of the lamina of the cricoid cartilage and converges on the laterally directed muscular process of the arytenoid cartilage. This muscle has been described as having three parts in canines and either two or three parts in humans. For the canine PCA, a horizontal, a vertical, and an oblique belly has been described by Sanders et al.[1,2] For the human PCA, Bryant et al. described a medial and a lateral belly,[18] but Sanders et al. divided this muscle again into a horizontal, a vertical, and an oblique portion.[4] The PCA rocks the arytenoid posteriorly so that the vocal process swings laterally, superiorly, and posteriorly, abducting the vocal fold (Figure 1, part a).

The LCA muscle originates on the superior border of the anterior arch of the cricoid and courses posteriorly to the muscular process of the arytenoid cartilage. The LCA is believed to adduct the vocal fold by rocking the arytenoid cartilage anteriorly such that the vocal process moves medially, inferiorly, and anteriorly[19] (Figure 1b).

The IA muscle is composed of a transverse and an oblique portion. The oblique portion originates on the superior aspect of the arytenoid cartilage, crosses the midline, and inserts into the muscular process of the contralateral arytenoid cartilage. The transverse portion connects the lateral borders of the two arytenoid cartilages together. Thus, during IA contraction, the arytenoid cartilages are drawn together, adducting the vocal folds (Figure 1c).

This study was conducted to characterize the three-dimensional orientations and the cross-sectional areas of the intrinsic abductor and adductor musculature of the canine larynx. Data acquired through this study should be useful in establishing a database for three-dimensional biomechanical modeling of vocal fold posturing. Specifically, they should allow for the calculations of the lines of action and moment arms for the intrinsic laryngeal muscles during vocal fold abduction and adduction.

Table 1. Subject Information		
Canine subject	Sex	Weight
1	F	20 kg
2	M	25 kg
3	M	27 kg
4	F	26 kg
5	M	20 kg
6	F	22 kg
7	M	20 kg
8	F	20 kg

Method

Four female and four male canine larynges were excised post mortem after cardiovascular experimentation in accordance with the Institutional Animal Care and Use Committee of the University of Iowa. All canine specimens were obtained from subjects without evidence of trauma or head and neck disease (Table 1). After harvest, the larynges were either slowly or quickly frozen (using liquid nitrogen) and were stored at –20° C. Prior to dissection, each larynx was thawed overnight in a 4° C refrigerator. Immediately before dissection, larynges were further thawed in physiological saline solution (0.9%).

Each canine larynx was first dissected using a blunt instrument technique to expose the PCA, LCA, and IA muscles. Any excess fat or fascial tissue was removed in preparation for mounting and muscle bundle dissection.

The larynx was mounted on a lab bench by securing the trachea over a piece of PVC tubing using an O-clamp around the first and second tracheal rings. Pincer clamps were used to secure the cricoid cartilage in the anatomical position. The arytenoid cartilages were firmly fixed in the cadaveric position using straight pins such that any rotation or translation of the arytenoid cartilages was eliminated. The larynx was then positioned such that the posterior ridge of the cricoid lamina (the cricoid prominence) between the PCA muscles was vertical (Figure 2) .

Once the larynx was mounted, three-dimensional spatial coordinates for the cricoid cartilage, the arytenoid cartilages, and the laryngeal abductors and adductors were obtained using a MicroScribe-3DX digitizer (Immersion Corporation, Salt Lake City, UT) with the HyperSpace Modeler software (Mira Imaging, Salt Lake City, UT). The spatial resolution/accuracy of the system was 0.2 mm. In an effort to normalize the orientation across larynges, the origin for each larynx was defined as the most superior and anterior aspect of the cricoid ring in the mid-sagittal plane. Following origin definition, the coordinates of the most posterior point on the cricoid prominence (between the right and left PCA muscles) were acquired. This point was used as a reference to the origin to establish the y-axis by a translation in the negative y direction with respect to the x direction; changes in the z direction were ignored. The z-axis was established by acquiring a number of points along the cricoid prominence. The x-axis was then empirically established by its orthogonal relationships with the y- and z-axes.

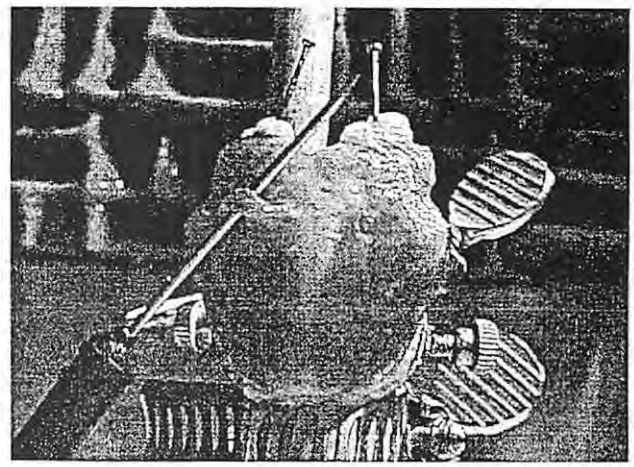


Figure 2. Posterior view of a mounted canine larynx showing the dissection (isolation) of a muscle bundle of the left PCA muscle.

Using blunt dissection instruments, an individual muscle bundle was carefully isolated and partially separated from the rest of the muscle. Next, its points of origin and insertion were visually identified (Figure 2). In all cases, the geometric center of the point of muscle bundle attachment was used as an estimate of the insertion or origin for that bundle. The coordinates of origin and insertion of each bundle were digitized before the bundle was removed using tissue forceps and iris scissors. The mass of each muscle bundle was measured using a Mettler AE100 laboratory balance with a measurement reliability of 0.1 mg (Mettler Instruments, Hightstown, NJ). Throughout the dissection, saline solution was periodically applied to the larynx to keep the tissues from drying.

The muscles of interest were organized into the following groups: left and right posterior cricoarytenoid (L PCA and R PCA, further separated into oblique and vertical portions); left and right lateral cricoarytenoid (L LCA and R LCA); and left and right interarytenoid (L IA and R IA, further separated into superior and inferior portions).

Geometrical Descriptions of Muscles

For each muscle bundle, six points were recorded (three for origin, three for insertion). The mean x, y, and z coordinates for each of the three samples were calculated to yield an average origin or insertion for each muscle bundle. The length *l* of each muscle bundle was then calculated using the Pythagorean Theorem:

$$l = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (1)$$

were x_1, y_1, z_1 are the coordinates of the origin and x_2, y_2, z_2 are those of the insertion.

An orientation vector r was defined for each muscle bundle as:

$$r = r_x i + r_y j + r_z k \quad (2)$$

where i, j, and k are unit vectors along the orthogonal axes, and that

$$r_x = x_2 - x_1 \quad (3)$$

$$r_y = y_2 - y_1 \quad (4)$$

$$r_z = z_2 - z_1 \quad (5)$$

As the mass m of each bundle was measured and the length was calculated, the cross-sectional area A was determined by:

$$A = \frac{m}{\rho l} \quad (6)$$

where ρ is the density of canine laryngeal muscle obtained previously (0.001043 g/mm^2).[20] The assumption here was that the muscle bundle has a uniform cross-sectional area along its entire course.

Resultant Vectors For Muscle Bundle Groups

The relative contribution of each muscle bundle to the total action of a muscle was estimated using cross-sectional area data. The orientation vector r for each bundle was multiplied by that bundle's cross-sectional area A. This product was defined as the scaled force vector rA. The assumption here was that all muscle fibers have equal contractile force per unit cross-sectional area. The scaled force vectors were summed and divided by the total cross-sectional area for that grouping of muscle bundles to yield an average *resultant* force vector R,

$$R \equiv R_x i + R_y j + R_z k = \frac{\sum_{i=1}^{n} r_i A_i}{\sum_{i=1}^{n} A_i} \quad (7)$$

A resultant force vector was calculated for each of the following muscle bundle groups, or *muscle portions*:

1) left posterior cricoarytenoid, oblique portion (L PCA o)
2) left posterior cricoarytenoid, vertical portion (L PCA v)
3) right posterior cricoarytenoid, oblique portion (R PCA o)
4) right posterior cricoarytenoid, vertical portion (R PCA v)
5) left lateral cricoarytenoid (L LCA)
6) right lateral cricoarytenoid (R LCA)
7) left interarytenoid, inferior portion (L IA inf)
8) left interarytenoid, superior portion (L IA sup)
9) right interarytenoid, inferior portion (R IA inf)
10) right interarytenoid, superior portion (R IA sup)

Data on the superior portion of the interarytenoid are not reported because its anatomy was found to be grossly different from that of human. Our observations showed that it did not connect the two arytenoid cartilages together. Rather, both its origin and insertion appeared to originate from the same side in the canine larynx.

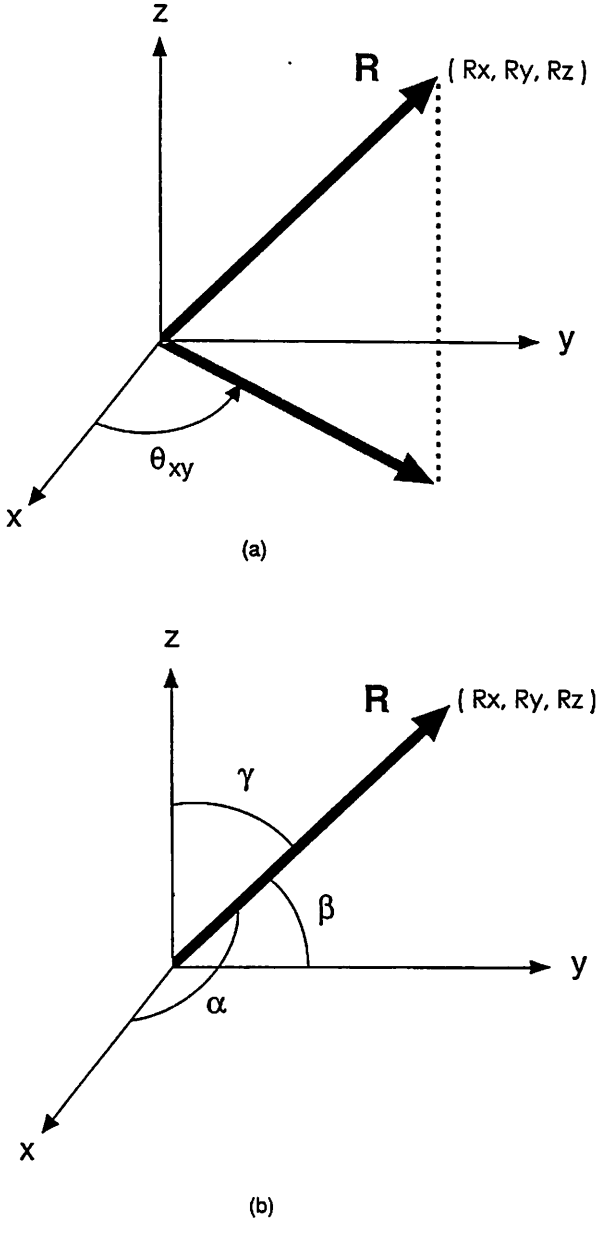


Figure 3. Three-dimensional illustration of (a) the two-dimensional projection angle θ_{xy} of the vector **R**, *(b) the angles α, β, and γ of the vector* **R**.

Planar Projection Angle Calculations

For two-dimensional planar analysis, the resultant force vector $\boldsymbol{R}$ for each muscle portion was used to calculate the projection angle in each of the three orthogonal planes. The projection angle in the xy plane was calculated using the formula (Figure 3a):

$$\theta_{xy} = \tan^{-1}\frac{R_y}{R_x} \quad (8)$$

The positive x direction was medial to lateral on the right side and the positive y direction was posterior to anterior. Angles were reported in standard fashion, with positive rotation defined as counterclockwise from the x-axis toward the y-axis.

Projection angles in the yz and xz planes were similarly calculated for each muscle portion:

$$\theta_{yz} = \tan^{-1}\frac{R_z}{R_y} \quad (9)$$

$$\theta_{zx} = \tan^{-1}\frac{R_x}{R_z} \quad (10)$$

Thus, the xy plane was horizontal, the yz plane sagittal, and the xz plane coronal.

Direction Cosines

For the purpose of three-dimensional modeling, the direction cosines were also calculated. The direction cosines were defined as the cosine of the angle between a given resultant force vector $\boldsymbol{R}$ and each of the three axes (Figure 3b),

$$\cos\alpha = \frac{R_x}{\|R\|} \quad \cos\beta = \frac{R_y}{\|R\|} \quad \cos\gamma = \frac{R_z}{\|R\|} \quad (11)$$

where $\|R\|$ is the magnitude or length of the resultant force vector.

Results and Discussion

Measurement Reliability

During the data acquisition process, there were two major sources of experimental and measurement errors. First, muscles and other soft tissues of the larynx sometimes showed slight movement and deformation under the pressure of the digitizer probe. Second, external forces applied on the larynx during muscle bundle dissection sometimes also caused slight tissue movement and deformation. Because of these errors, there was some variability in the measured coordinates across different sampled points of the same muscle bundle insertion or origin.

Experimental error or variability of the data was quantified for Canine 7, which was chosen because there was a large number of muscle bundles in each of its different muscle portions. Error in length measurement was estimated by first finding the maximum length possible based on the three origin data points and the three insertion data points for each muscle bundle. The maximum lengths were then averaged across bundles for each muscle portion. Table 2 shows the deviations between the maximum lengths and

Table 2.
Measurement Reliability

Magnitude of measurement errors by muscle portion (also in percentage errors)

	Length (mm)	Cross-sectional area (mm^2)	θ_{xy} (degrees)	θ_{yz} (degrees)	θ_{zx} (degrees)
L PCA o	0.772 (5.20%)	0.207 (0.81%)	2.97 (1.65%)	2.42 (1.35%)	4.50 (2.78%)
L PCA v	1.277 (8.81%)	0.525 (4.99%)	6.03 (3.35%)	4.34 (2.41%)	3.90 (2.16%)
L PCA (total)	0.916 (6.22%)	0.298 (0.83%)	3.79 (2.10%)	2.94 (1.63%)	4.70 (2.61%)
R PCA o	0.755 (4.99%)	0.238 (1.07%)	2.41 (1.34%)	2.77 (1.54%)	4.86 (2.70%)
R PCA v	0.767 (5.05%)	0.271 (2.55%)	3.60 (2.00%)	2.29 (1.27%)	3.66 (2.03%)
R PCA (total)	0.759 (5.01%)	0.249 (0.76%)	2.80 (1.56%)	2.61 (1.45%)	4.46 (2.48%)
L LCA	0.791 (5.47%)	0.355 (1.65%)	2.18 (1.21%)	4.56 (2.53%)	14.18 (7.88%)
R LCA	0.850 (5.91%)	0.35 (1.68%)	3.15 (1.75%)	5.66 (3.15%)	9.42 (5.23%)
L IA inf	0.718 (8.02%)	0.746 (6.60%)	4.23 (2.35%)	3.81 (2.12%)	7.44 (4.13%)
R IA inf	0.714 (7.42%)	0.919 (7.06%)	4.82 (2.68%)	5.10 (2.83%)	8.49 (4.72%)
Mean error	0.792 (6.21%)	0.486 (2.80%)	3.49 (1.94%)	4.11 (2.28%)	8.11 (4.51%)
S.D.	0.168 (1.41%)	0.240 (2.47%)	1.18 (0.65%)	1.21 (0.67%)	3.32 (1.85%)

the average lengths for the eight different muscle portions, and their percentage errors. As shown in Table 2, the mean error for all muscles was 6.2%.

The maximum possible cross-sectional area for each bundle of Canine 7 was calculated using Equation 6 based on the measured mass, the balance reliability (0.1 mg), and the average length data. The minimum possible cross-sectional area was calculated similarly but it was based on the maximum calculated length for each bundle. The difference between the maximum and the minimum for each bundle was obtained and an average was calculated for each muscle portion. Table 2 shows that the error ranged from about 1-7%. The maximum and minimum orientation angles (2-D projection angles) were calculated from the three origin and the three insertion points for each bundle to estimate the error in vector calculations. The difference between these angles was obtained for each bundle and an average error was computed for each muscle portion. Table 2 shows that the percentage error values, reported with respect to 180°, ranged from about 1-8% with most of them smaller than 3%.

Muscle Mass, Length, Cross-Sectional Area and Orientation

Table 3 shows the mass of each muscle portion for all subjects and their averages. It can be seen that the PCA muscle was always the most massive, whereas the IA had the smallest mass and the LCA was in between. The oblique portion of the PCA was consistently more massive than the vertical portion, in many cases by two to three times. The data also showed that muscles of the right and the left sides were basically symmetric to each other in terms of their mass. These results were consistent with previous classical anatomical descriptions of the laryngeal abductor and adductor muscles.

Table 4 shows the lengths of the muscle portions which were averages of individual muscle bundle lengths. Note that the lengths of the PCA were not simple averages

Table 3.
Mass of Canine Laryngeal Abductor and Adductor Muscles (in grams)

	Mass by muscle portion (g)									
Subject	1	2	3	4	5	6	7	8	Mean	S.D.
L PCA o	0.429	0.580	0.480	0.580	0.210	0.359	0.300	0.305	0.405	0.136
L PCA v	0.161	0.168	0.206	0.137	0.148	0.136	0.169	0.121	0.155	0.027
L PCA (total)	0.590	0.748	0.686	0.716	0.358	0.495	0.469	0.449	0.564	0.142
R PCA o	0.337	0.511	0.353	0.529	0.213	0.301	0.273	0.266	0.348	0.115
R PCA v	0.171	0.143	0.263	0.135	0.118	0.201	0.191	0.181	0.175	0.046
R PCA (total)	0.507	0.655	0.616	0.664	0.331	0.502	0.464	0.447	0.523	0.115
L LCA	0.273	0.351	0.402	0.466	0.184	0.346	0.268	0.269	0.320	0.089
R LCA	0.269	0.304	0.357	0.448	0.182	0.358	0.276	0.270	0.308	0.080
L IA inf	0.123	0.068	0.139	0.130	0.082	0.128	0.106	0.112	0.111	0.025
R IA inf	0.154	0.095	0.152	0.182	0.073	0.143	0.112	0.139	0.131	0.036
Total	1.916	2.221	2.351	2.607	1.210	1.972	1.695	1.686	1.957	0.439

Table 4.
Length of Canine Laryngeal Abductor and Adductor Muscles (in millimeters)

	Length by muscle portion (mm)									
Subject	1	2	3	4	5	6	7	8	Mean	S.D.
L PCA o	14.527	14.199	16.142	15.911	14.161	15.261	13.096	15.081	14.797	1.005
L PCA v	11.899	15.082	15.634	16.029	14.885	14.568	13.310	14.540	14.493	1.325
L PCA (total)	13.943	14.396	16.029	15.933	14.342	15.113	13.149	14.965	14.734	0.979
R PCA o	14.287	14.486	15.378	16.763	14.827	15.322	13.441	14.948	14.932	0.967
R PCA v	14.509	15.373	17.008	18.892	15.732	16.600	14.940	13.852	15.863	1.604
R PCA (total)	14.342	14.683	15.786	17.150	15.128	15.748	13.903	14.583	15.165	1.036
L LCA	13.259	14.993	14.768	14.918	13.913	15.394	14.431	13.976	14.457	0.701
R LCA	14.407	15.006	13.954	16.152	12.803	13.982	15.442	15.276	14.394	1.210
L IA inf	9.412	7.834	9.210	9.744	8.897	8.887	8.945	8.920	8.951	0.527
R IA inf	10.054	8.728	14.280	10.489	8.048	8.622	7.761	9.807	9.626	1.969

of those of the oblique and vertical portions, because there was always a larger number of muscle bundles in the oblique portion as evidenced by its larger mass (see Table 3). Nonetheless, the data showed that their lengths were quite similar to one another (mean differences < 1 mm). The lengths of the LCA were also close to those of the PCA, whereas IA was about 30-50% shorter.

The cross-sectional area for each muscle bundle was estimated from the measured mass and the calculated length. The average cross-sectional area for each muscle portion was calculated and results are shown in Table 5. Similar to the mass data, the PCA was consistently the largest in cross-sectional area, whereas the LCA was often about 40% smaller and the IA was about 60% smaller. Besides, the oblique portion of the PCA was again about 2-3 times larger than the vertical portion.

Table 6 shows the average projection angles and direction cosines of the eight muscle portions. The angles of orientation were averaged across all subjects. They were oriented such that the muscular process of the arytenoid cartilage was the geometric origin and a positive angle was defined as rotating in a counterclockwise direction from x to y, y to z, or z to x.

Figure 4 (following page) illustrates the angles of the two portions of the PCA on the xy, yz and xz planes. The orientation angles of the arrows represented the lines of action of the muscle portions, while the lengths of the arrows were indications of their relative force magnitudes

Table 5.
Cross-Sectional Area of Canine Laryngeal Abductor and Adductor Muscles (in mm^2) as Computed from Muscle Mass, Length, and Density [c.f. Eq. (6)]

Cross-sectional area by muscle portion (mm^2)										
Subject	1	2	3	4	5	6	7	8	Mean	S.D.
L PCA o	27.20	38.10	27.95	33.46	14.03	22.13	21.44	19.04	**25.42**	**7.86**
L PCA v	13.02	10.42	12.96	8.12	9.54	8.89	12.03	9.21	**10.52**	**1.91**
L PCA (total)	40.22	48.52	40.91	41.58	23.57	31.02	33.47	28.25	**35.94**	**8.24**
R PCA o	22.84	33.32	21.73	29.89	14.29	18.99	19.47	17.34	**22.23**	**6.41**
R PCA v	11.26	8.92	14.82	6.87	7.16	11.47	12.18	12.46	**10.64**	**2.77**
R PCA (total)	34.10	42.25	36.55	36.76	21.44	30.46	31.65	29.80	**32.88**	**6.16**
L LCA	19.48	23.03	25.83	31.78	13.11	22.16	17.79	19.37	**21.57**	**5.61**
R LCA	17.82	19.60	23.91	26.63	14.05	25.12	17.55	17.61	**20.80**	**4.41**
L IA inf	12.81	8.34	14.44	12.83	8.94	13.73	11.34	12.01	**11.30**	**2.54**
R IA inf	14.71	10.47	14.41	16.55	8.84	15.77	13.88	13.73	**13.02**	**2.90**
Total	140.14	154.20	159.05	170.13	95.96	145.26	133.68	129.73	**141.02**	**22.60**

Table 6.
Mean Projection Angles (two dimensional) and Direction Cosines (three dimensional) of Canine Laryngeal Abductor and Adductor Muscles

Mean angles of orientation by muscle portion						
	2-D projection angles (degrees)			3-D direction cosines		
	θ_{xy}	θ_{yz}	θ_{zx}	cos α	cos β	cos γ
L PCA o	-22.3	-115.5	130.7	0.739	-0.293	-0.607
L PCA v	-17.0	-97.2	157.6	0.372	-0.119	-0.920
L PCA (total)	-21.1	-108.6	139.0	0.647	-0.247	-0.721
R PCA o	-158.6	-117.6	-126.9	-0.759	-0.298	-0.579
R PCA v	-169.8	-94.3	-157.4	-0.413	-0.063	-0.909
R PCA (total)	-161.2	-107.4	-137.3	-0.666	-0.228	-0.710
L LCA	75.0	-25.6	150.9	0.238	0.869	-0.433
R LCA	99.7	-23.6	-158.6	-0.158	0.902	-0.403
L IA inf	-43.0	157.5	68.9	0.701	-0.660	0.269
R IA inf	-137.7	152.1	-64.3	-0.692	-0.627	0.358

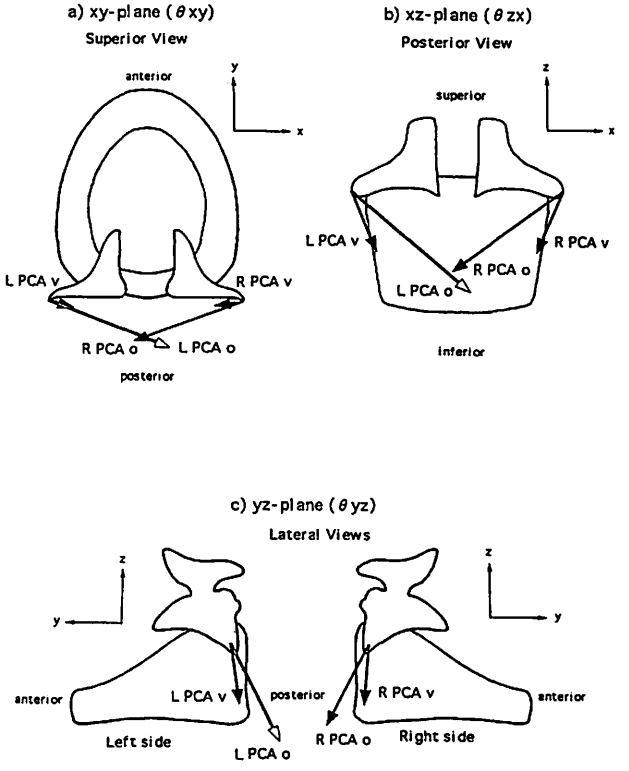


Figure 4. (a) Superior view of the resultant force vectors of the oblique and vertical portions of PCA projected onto the xy plane, (b) posterior view of the resultant force vectors of the oblique and vertical portions of PCA projected onto the xz plane, (c) lateral views of the resultant force vectors of the oblique and vertical portions of PCA projected onto the yz plane.

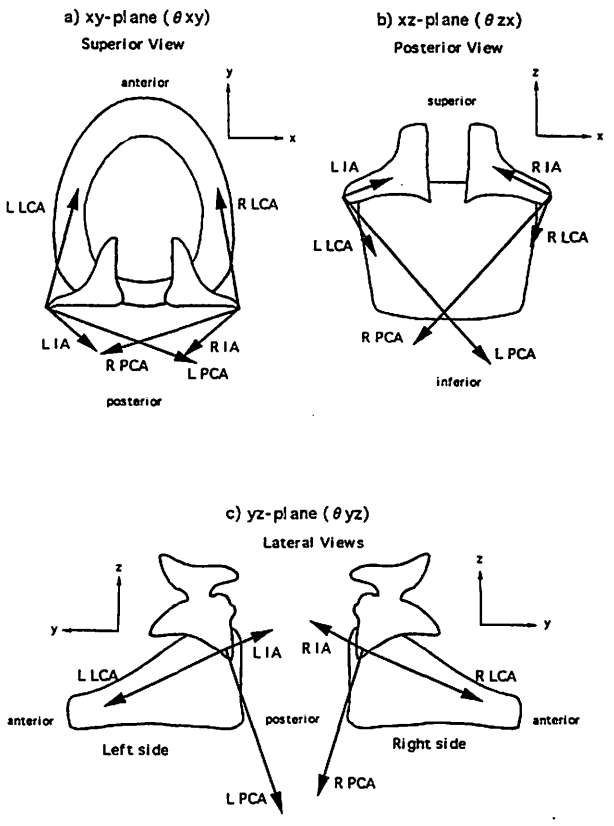


Figure 5. (a) Superior view of the resultant force vectors of the PCA, LCA, and IA projected onto the xy plane, (b) posterior view of the resultant force vectors of the PCA, LCA, and IA projected onto the xz plane, (c) lateral views of the resultant force vectors of the PCA, LCA, and IA projected onto the yz plane.

as they were scaled according to the muscle cross-sectional areas. Hence, the arrows represent projections of the resultant force vectors of the two PCA portions onto the three orthogonal planes. Not surprisingly, the oblique portion was always at a more oblique orientation than the vertical portion. It was also always stronger, especially for its component on the xy plane.

Figure 5 summarizes the resultant force vectors of all the three abductor and adductor muscles, namely LCA, PCA (the resultant of the two portions), and IA (the inferior portion). Again, the orientation angles represented their lines of action, while the lengths were estimations of the relative magnitudes of their forces projected onto the three planes. In terms of the orientation angles, the data were qualitatively consistent with classical descriptions of the muscles. The lines of action as shown in Figure 5 suggested that the LCA tends to pull the muscular process of the arytenoid cartilage anteriorly, inferiorly, and medially, thereby moving the vocal process medially and adducting the vocal fold. The PCA, on the other hand, tends to move the muscular process posteriorly, inferiorly, and medially, abducting the vocal fold. Interestingly, the IA seems to be somewhat of an antagonist of the LCA based on their orientation angles on the xy and the yz planes. Nonetheless, its superior-medial direction of action as shown on the xz plane suggested that it might move the arytenoid towards the midline, thereby adducting the vocal fold. These data on the average lines of action of the abductor and adductor muscles may be useful clinically for certain phonosurgical procedures. For example, data on the orientation angles of the LCA may serve as quantitative guidelines for clinicians to establish a more physiological direction of the sutures used in arytenoid adduction.

In terms of the relative force magnitudes, the data suggested that the PCA was always stronger than the LCA and the IA, especially on the xz plane where it was 3-4 times stronger. The LCA was slightly weaker than the PCA on the xy and the yz planes but the difference was much larger on the xz plane. The IA was always the weakest, especially on the yz plane where it was 2-3 times weaker than the other two. These findings were consistent with the cross-sectional area data as the muscle's resultant force vectors were com-

puted partly based on their cross-sectional areas. The assumption was that the maximum active stress was similar for the three different muscles, such that the relative magnitude of force generated by a muscle is proportional to its cross-sectional area.

Limitations and Suggestions for Further Studies

One limitation of the present study was that the calculations of muscle lengths and orientation angles were made assuming the muscle bundles form a straight line between the origin and the insertion points. However, this was not always the case for some muscles, especially the PCA, where muscle fibers were seen to often course around the curved laryngeal cartilage surfaces. Besides, fiber bundles on a muscle's surface were often more curved than the internal bundles. Such errors likely led to underestimations of length, overestimations of cross-sectional area, and discrepancies in estimations of the effective line of action in some of the muscles. In future experiments, increasing the number of sampling points for each bundle might help to reduce such errors. For example, by taking an extra data point at the midpoint of each muscle bundle, more accurate estimations of the muscle bundle length and its effective orientation angle can be established based on the best-fit curve for the three data points (origin, insertion, and midpoint).

During data collection, the cricoid and the arytenoid cartilages were securely fixed in the cadaveric position in order to minimize specimen movement and to assure accurate measurements of the 3-D coordinates. However, slight movements of the cricoid and the arytenoid cartilages were still sometimes observed during muscle bundle resection and data acquisition. Such movements likely introduced random errors into the 3-D coordinate data sampled by the digitizer. In future experiments, extra pins and/or other devices should be used to more securely fix the cartilages so as to eliminate these errors.

Anatomical differences between the canine and the human larynx must be acknowledged when canine data are used for modeling of the human larynx. The canine larynx is similar to the human in terms of size, morphology and basic vocal fold anatomy, but there are also some significant differences. The human IA muscle is made of a transverse and an oblique portion, both of which clearly and consistently cross the midline. For the canine IA, however, our observations showed that it can be divided into a superior and an inferior portion. The inferior portion was distinctly separated into left and right halves joined by a sheet of tendon-like connective tissue at the midline, while the superior portion stayed on the same arytenoid cartilage and extended anteriorly to insert into the structures above the arytenoid (the cuneiform and corniculate processes). Furthermore, the arytenoid cartilages of the canine are proportionately larger than those of the human.

Considering such anatomical differences between the canine and the human larynx, the three-dimensional characterization of the canine laryngeal muscles may only represent a rough approximation of the human. However, the two species exhibit sufficient similarities in basic laryngeal anatomy that the canine model remains a valuable approximation of the human larynx.

Conclusions

The posterior cricoarytenoid, lateral cricoarytenoid, and interarytenoid muscles are important in the control of vocal fold abduction and adduction. These muscles function in a well coordinated manner during phonation, respiration, and airway protection to allow for various adjustments of the anatomical orientations of the vocal folds. This study quantified the mass, length, relative cross-sectional area, and angle of orientation in three-dimensional space of these laryngeal abductor and adductor muscles in eight canine larynges. This three-dimensional anatomical characterization allowed for the calculations of average lines of action and moment arms for the muscles, providing important data for biomechanical modeling of vocal fold posturing. These data may also be useful as clinical guidelines for certain phonosurgical procedures such as arytenoid adduction.

Acknowledgments

This work was supported by a grant (P60 DC00976) from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health.

References

1. Sanders I, Jacobs I, Wu B-L, Biller HF, "The Three Bellies of the Canine Posterior Cricoarytenoid Muscle: Implications for Understanding Laryngeal Function," Laryngoscope 1993; 103: 171-177.

2. Sanders I, Rao F, Biller HF, "Arytenoid Motion Evoked by regional Electrical Stimulation of the Canine Posterior Cricoarytenoid Muscle," Laryngoscope 1994; 104: 456-462.

3. Sanders I, Wu B-L, Mu L, Li Y, Biller HF, "The innervation of the human larynx," Arch Otolaryngol Head Neck Surg 1993; 119: 934-939.

4. Sanders I, Wu B-L, Mu L, Biller HF, "The innervation of the human posterior cricoarytenoid muscle: Evidence for at least two neuromuscular compartments," Laryngoscope 1994; 104: 880-884.

5. Alipour-Haghighi F, Titze IR, Perlman AL, "Tetanic contraction in vocal fold muscle," J Speech Hear Res 1989; 32: 226-231.

6. Alipour-Haghighi F, Perlman AL, Titze IR, "Tetanic response of the cricothyroid muscle," Ann Otol Rhinol Laryngol 1991; 100: 626-631.

7. Alipour F, Titze I, "Active and passive characteristics of the canine cricothyroid muscles," J Voice 1999; 13: 1-10.

8. Sonesson B, "Die funktionelle anatomie des cricoarytenoid gelenkes," Z anat Entwick Lungsgesch 1959; 121: 292-303.

9. Frable MA, "Computation of Motion at the Cricoarytenoid Joint," Arch Otolaryngol 1961; 73: 551-556.

10. von Leden H, Moore P, "The Mechanics of the Cricoarytenoid Joint," Arch Otolaryngol 1961; 73: 541-550.

11. Ardran GM, Kemp FH, "The Mechanism of the Larynx: Part I. The Movement of the Arytenoid and Cricoid Cartilages," British J Radiol 1966; 39: 641-654.

12. Sellars IE, Sellars S, "Cricoarytenoid joint structure and function," J Laryngol Otol, 1983; 97: 1027-1034.

13 Neuman TR, Hengesteg A, Lepage RP, Kaufman KR, Woodson GE, "Three-dimensional motion of the arytenoid adduction procedure in cadaver larynges," Ann Otol Rhinol Laryngol 1994; 103: 265-270.

14. Selbie WS, Zhang L, Levine WS, Ludlow CL, "Using joint geometry to determine the motion of the cricoarytenoid joint," J Acoust Soc Am 1998; 103: 1115-1127.

15. Gray H, "Gray's Anatomy" 36th Ed, edited by PL Williams and R Warwick, Churchill Livingstone, New York, 1980.

16. Kahane J, "The cricoarytenoid joint: Perspectives gained since von Leden and Moore (1961)," J Voice, in press.

17. Martensson A, Skoglund CR, "Contraction properties of intrinsic laryngeal muscles," Acta Physiol Scand 1964; 60: 318-336.

18. Bryant NJ, Woodson GE, Kaufman K, Rosen C, Hengesteg A, Chen N, Yeung D, "Human Posterior Cricoarytenoid Muscle Compartments, Anatomy and Mechanics," Arch Otolaryngol Head Neck Surg 1996; 122: 1331-1336.

19. Sanders I, Mu L, Wu B-L, Biller HF, "The Intramuscular Nerve Supply of the Human Lateral Cricoarytenoid Muscle," Acta Otolaryngol (Stockh) 1993; 113: 679-682.

20. Perlman AL, Titze IR, "Development of an in vitro technique for measuring the elastic properties of vocal fold tissue," J Speech Hear Res 1988; 31: 288-298.

NCVS Status and Progress Report - 14
September 1999, 27-34

Characterization of the Medial Surface of the Vocal Folds: Methodology and Preliminary Findings

David Berry, Ph.D.
Martyn Clark, M.S.
Douglas Montequin, M.A.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

A methodology is developed for the quantification of the medial surface of the vocal folds in excised larynxes. Lead molds were constructed from the glottal airway of a canine larynx for three distinct glottal configurations corresponding to "pressed" folds, just barely adducted folds, and 1 mm abducted folds measured between the vocal processes. Using a high-resolution laser striping system, the 3D molds were digitally scanned. Low-order polynomials were fitted to the data, and goodness-of-fit statistics were reported. For all glottal configurations, a linear variation (flat surface) approximated the data with a coefficient of determination of 90%. This coefficient increased to roughly 95% when a quadratic variation (curvature) was included along the vertical dimension. If more than the top 5 mm or so of the folds were included (the portion usually corresponding to vibration), a cubic variation along the vertical dimension was necessary to explain a change in concavity at the conus elasticus. These data suggest the utility of a model based on a convergence and a bulging coefficient (Titze, 1989). For all glottal configurations, the convergence coefficients and bulging coefficients were computed. Because pre-phonatory conditions have a profound influence on vocal fold vibration and on the quality of phonation, such shaping parameters are highly significant. With the viability of this methodology substantiated, it is envisioned that future studies will characterize greater quantities of glottal shapes, including those of human folds.

Introduction

Initial conditions and boundary conditions have a profound impact on many nonlinear dynamical systems. In vocal fold vibration, which represents such a nonlinear system, a small change in the pre-phonatory shape of the medial surface of the vocal folds can mean the difference between a chest-like, falsetto-like, or fry-like vibration pattern, or it can mean the difference between periodic and aperiodic oscillations (Berry, Herzel, Titze & Story, 1996). For phonosurgeons who alter the glottal geometry in reconstructive procedures, pre-phonatory vocal fold shape is increasingly understood as a critical variable. For example, measurement of the angle of glottal opening and precise positioning of the arytenoid cartilage are becoming more and more common in phonosurgery (Inagi, Khidr, Ford, Bless & Heisey, 1997; Woodson & Murray, 1994; Woodson, Hengesteg, Rosen, Yeung & Chen, 1997).

Several simulation models of vocal fold vibration have been used to quantify subtleties of vocal fold shape (Titze & Talkin, 1979; Alipour & Titze, 1985; Berry, Herzel, Titze & Krischer, 1994). These models solve second-order partial differential equations to describe the resultant vibrations. In such models, the specificity of initial conditions, or pre-phonatory shape, is essential in order to make a reasonable prediction of the oscillations. A series of normal mode studies on the vocal folds has shown that phonation frequency has a remarkable correspondence with the lowest resonance frequency of the folds, as measured immediate prior to phonation (Kaneko et al., 1981; Kaneko et al., 1983). Such resonance frequencies are known to be strongly influenced by glottal geometry (Berry & Titze,

1996). However, to date, no quantitative data exists on these pre-phonatory shapes. In particular, previous MRI and CT images obtained from the vocal tract and laryngeal regions have not yielded the resolution needed for modeling the medial surface of the folds (on the order of 0.1 mm).

The focus of this paper is the development of a methodology for high-precision quantification of the medial shape of the vocal folds in excised larynxes. Once the methodological study is presented, future studies are envisioned in which greater numbers of molds will be analyzed, with more decisive conclusions drawn. The following questions will be probed: What order of polynomial must be used to approximate the vocal fold shape across a variety of glottal configurations? Does the theoretical model proposed by Titze (1989), which assumes a linear variation of vocal fold shape along the anterior-posterior length and a quadratic variation along the vertical depth, provide an adequate description of the molds? If so, what are the convergence coefficients and bulging coefficients of each of the molds, and how do these constants change as a function of glottal adduction?

Methods

A canine larynx was obtained post mortem from an experimental animal weighing approximately 20 kg. Prior to molding the airway, the excised larynx was dissected to remove the epiglottis, the ventricular folds, and all but a short section (0.5 - 1 inch) of the trachea. The interior of the larynx was coated with a silicone gel to allow easy removal of the mold. A stopper, held in place with a hose clamp, was used to block the trachea (Fig. 1a). The larynx was then mounted in a vertical orientation and the vocal folds were adducted with two-pin micrometer devices, as shown in Fig. 1b. Once a desired level of adduction was achieved, molten wax ("Tissueprep" histological wax) was injected into the glottal airway and left to harden.

In order to make a more permanent mold of the airway, the wax mold was removed and placed in dental plaster, making certain that the tracheal end of the wax mold broke the surface of the plaster (so that the wax could be removed later). After the plaster hardened, the wax was removed by placing the plaster-wax compound into a kiln and slowing heating the kiln to at least 327.5 C, which is the melting point of lead. The wax was burned away, and was replaced with liquid lead. The lead mold was allowed to cool for one hour. Subsequently, the plaster was cracked away, leaving a 3D lead mold of both the glottal and subglottal airway, as well as a 3D representation of the medial surface of the vocal folds. The moldings appeared to yield a good representation of the glottal airway. There was no evidence of tissue deformation during the wax injection, or during hardening of the plaster or the reverse lead mold.

Three molds were obtained from the larynx, each mold corresponding to a unique glottal configuration: (1) an open larynx (a glottal width of approximately 1 mm, with no micrometer-adduction devices applied), (2) arytenoids were just barely adducted, and (3) the arytenoids were pressed to the point of 1 mm beyond "just touching". Measurements of glottal width were taken between the anterior points of the vocal processes.

In preparation for analysis of the molds, it was necessary to digitize the surfaces. The physical orientation

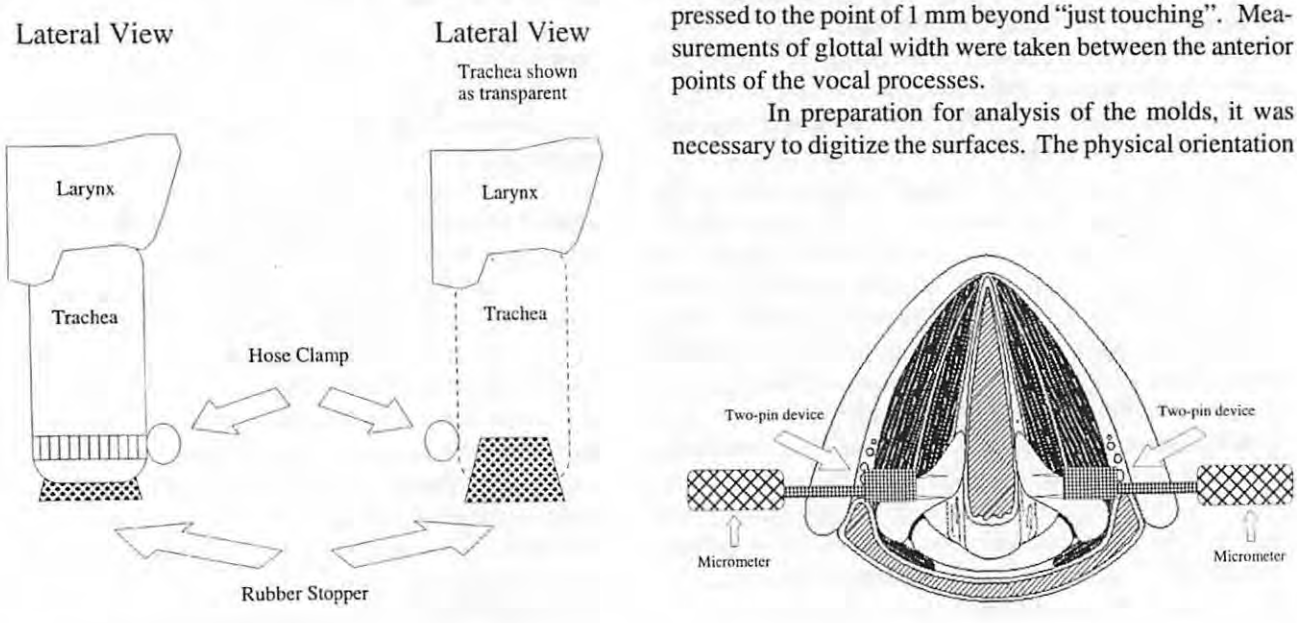


Figure 1. (a-left) A lateral view of the larynx, depicting the stopper which was placed at the base of the trachea to prevent wax leakage during the molding process. (b-right) A superior view of the larynx, depicting the 2-pin micrometer device used to control arytenoid adduction

used throughout the paper is as follows: *Z* direction, the vertical dimension; *Y* direction, the anterior-posterior length; and *X* direction, the medial-lateral dimension of the folds. The metallic composition of the lead molds facilitated digitization with a 3D laser striping system. In particular, the IMAGINE2 laser striper, built by the Department of Artificial Intelligence at the University of Edinburgh, was employed to digitize the molds (Fisher et al., 1993; Trucco et al., 1998). The system measured the *X* coordinate with an error of approximately 0.1 mm, and scanned the surfaces in increments of 0.5 mm in the *Y* and *Z* directions. Scans were taken of all of the molds from three different orientations: the left, the top, and the right, as shown in Fig. 2. Because the molds appeared to be predominantly symmetric, symmetry was assumed and only the left view was analyzed. However, if one desired to analyze the geometric asymmetries between left and right folds, the present technique could accommodate such a study (i.e., both left and right view could be analyzed).

In modeling the medial surface of the folds, it was of particular interest to know which regions of the molds corresponded to vibrating tissues. Indeed, the focus of this investigation was the pre-phonatory shape of tissue regions in which oscillations might occur. Such data directly impacts our computer models of vocal fold vibration, and are also of interest in phonosurgery. Although the non-oscillating tissue regions were not the focus of this study, they could be useful for modeling subglottal airflow.

The region where tissue vibration was possible is depicted on the lead mold shown in Fig. 3, which illustrates a medial view of the mold. The numeral "3" delineates the superior surface of the folds, which appears as a clear indentation on the mold. Above this line, the molding material spilled over the top of the glottal airway. Numeral "2" marks the inferior boundary of vibrating tissue. Palpation of the subglottal wall on the original larynx (i.e., not the mold) revealed that the depth where tissue vibration could occur varied along the length of the folds. The depth was smaller at the anterior and posterior extremities than midway along the length. The inferior boundary was roughly symmetric about the midpoint of the membranous fold, curving upward anteriorly and posteriorly, as shown in Fig. 3. The maximal depth of the region of vibration was approximately 5.6 mm on this mold.

The four vertical lines, marked with a numeral "1" show two possible choices for the anterior and posterior boundaries of the vocal fold. Presumably, the most extreme positions give the most accurate estimate of vocal fold length. This is because the most extreme positions of the mold made contact with the structures typically used to measure vocal fold length, e.g., the vocal process and the anterior commisure. However, noisy data at the extremes would sometimes necessitate a small reduction of this region to facilitate data analysis.

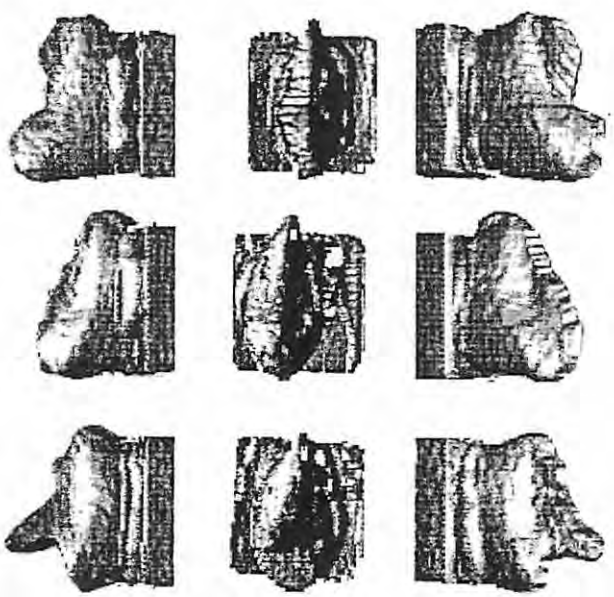


Figure 2. A surface rending of the 3D data obtained the laser striper for larynx 1. Left, superior, and right views are shown from left to right in the pictures, respectively. From top to bottom, the open larynx, arytenoids just touching, and pressed arytenoid configurations are shown.

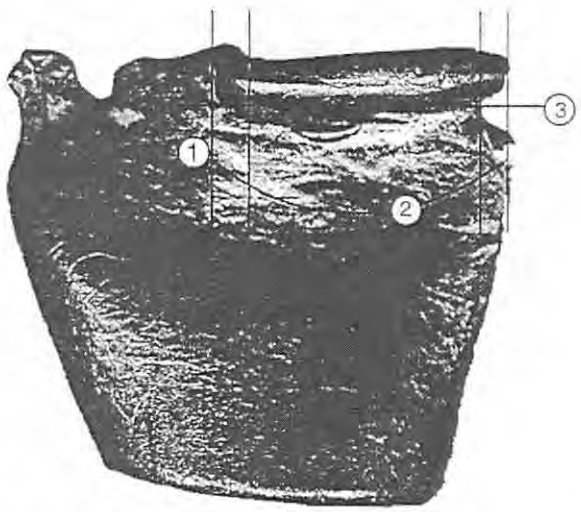


Figure 3. A side view of a sample lead mold of the glottal airway. Line "3" indicates the superior edge of the vocal fold, "2" the inferior edge where tissue vibration is possible, and "1" plausible choices for the anterior/posterior boundaries. While the most extreme markers give a more accurate indication of the true vocal fold length, the more interior markers are often used in this investigation in order to avoid noisy data near the end points.

Data Analysis, Results and Discussion

The first step of the analysis consisted in fitting low-order polynomials to the medial surface of the molds. Polynomials were expressed in the following general form:

$$X(y,z) = \sum_{i=0}^{M} \sum_{j=0}^{N} A_{ij} y^i z^j \qquad (1)$$

where A_{ij} were coefficients determined by computing a best-fit polynomial to the image data based on a least-square error procedure. In particular, the FMINS routine in MATLAB, a commercial software package, was used for this purpose.

Table 1.
***COD* Percentages for the Three Molds of Larynx 1**

	Open Larynx				Just Barely Adducted				Pressed Arytenoids			
Max. Power in z (N)	Max. power in y (M)				Max. power in y (M)				Max. power in y (M)			
	0	1	2	3	0	1	2	3	0	1	2	3
0	--	31.	--	--	--	23.	--	--	--	10.	--	--
1	54.	89.	93.	--	56.	92.	92.	--	61.	83.	83.	--
2	--	92.	93.	94.	--	95.	96.	96.0	--	95.	95.	95.
3	--	93.	95.	96.	--	96.	96.	96.9	--	95.	95.	95.
4	--	93.	--	--	--	96.	--	--	--	95.	--	--

Best-fit polynomials were computed for each mold. For each glottal configuration, the values of M and N were systematically modified, thus altering the degree of the polynomials in y and z (see Eq. 1). For the linear cases ($M \leq 1$ and $N \leq 1$, excluding the case where $M = N = 1$), a unique solution was obtained using a standard linear optimization procedure. However, for higher order polynomials, results were dependent on the initial values supplied to the FMINS routine. This was because nonlinear optimization procedures do not necessarily capture a global minimum, but only a local minimum. To deal with this problem, initial guesses for higher order polynomials were based on solutions already computed for the lower order polynomials. Because of the smoothness of the surfaces of the molds, the linear polynomials already gave reasonable estimates of the surface shapes.

For every polynomial, the coefficient of determination (*COD*) was computed in order to assess its goodness-of-fit with the data. A standard term in statistics, defined in the following equation, the *COD* helped assess the goodness-of-fit of the polynomials to the empirical data:

$$COD = 1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (x_i - \bar{x})^2} \qquad (2)$$

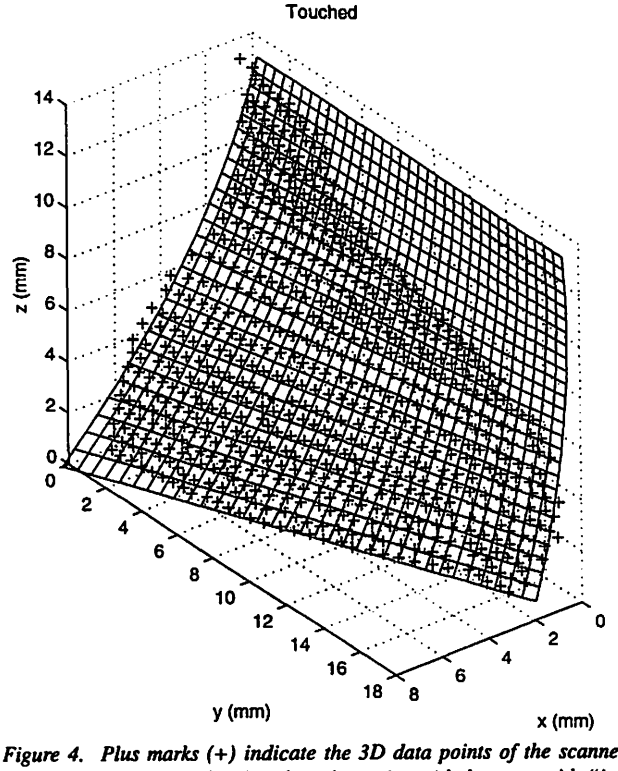


Figure 4. Plus marks (+) indicate the 3D data points of the scanned mold corresponding to the glottal configuration with the arytenoids "just touching." The mesh corresponds to the best polynomial fit for M=*1 (linear in the anterior-posterior direction) and* N=*2 (quadratic in the vertical direction).*

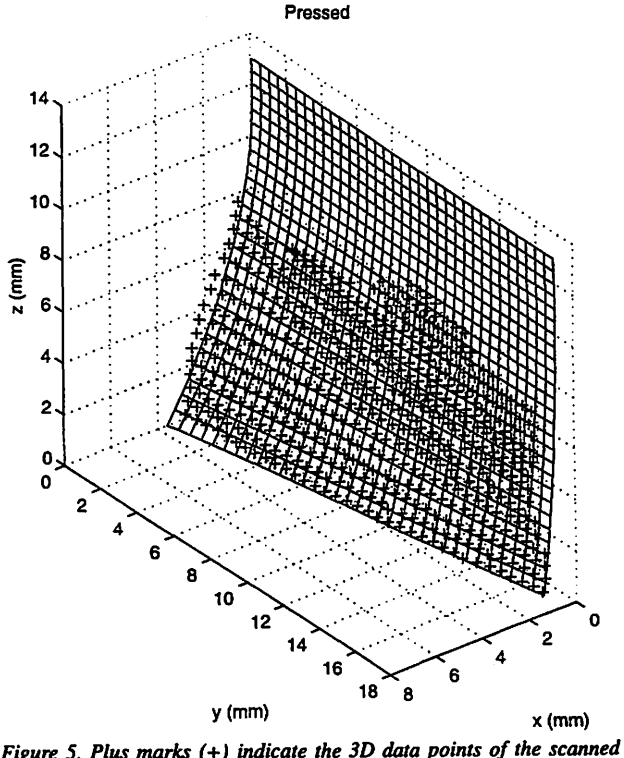


Figure 5. Plus marks (+) indicate the 3D data points of the scanned mold corresponding to the glottal configuration when the arytenoids are "pressed" together 1 mm past the "just touching" configuration. The mesh corresponds to the best polynomial fit for M=*1 (linear in the anterior-posterior direction) and* N=*2 (quadratic in the vertical direction).*

where y_i referred to predictions from the model, x_i referred to the empirical data, and $\bar{x}$ referred to the average value of the empirical data. The value of *COD* could range between zero and one. A value zero didn't capture any of the variance of the data (i.e., if the model was no better than an average value estimate), and a value of 1 gave a perfect match between model and data. With a computed table of *COD* percentages for various values of *M* and *N*, one could objectively assess the order of polynomial necessary to fit the data. The results for the three molds are shown in Table 1.

For all the glottal configurations, the data strongly suggested that the *N*=*M*=1 condition was a satisfactory condition. If either *N* or *M* was lowered, the *COD* dropped dramatically, indicating that the model could no longer adequately explain the data. For *N*=*M*=1, the *COD* ranged between 83 and 92%. An additional 3-8% percent gain could be obtained by increasing *N* to 2, suggesting that a quadratic variation in *z* might also be important in explaining the curvature of the folds. For the most part, only small gains were achieved by further increasing the values of *M* and *N*.

Using *N*=2 and *M*=1, a visual portrayal of the polynomical fit to the data is shown for the "adducted" and "pressed" conditions, as shown in Figs. 4 and 5, respectively. Notice that for the "adducted" mold, no data points were available for the superior-anterior portion of the mold. This illustrates a general limitation of the molding procedure: in order for the mold to exist in a particular region, the glottal airway needs to have a finite width (on the order of 0.1 mm). Thus, the mold did not exist (the molding material simply broke off) in any region where glottal airway had a width less than 0.1 mm. For the purpose of data analysis, such regions were zero-padded. While this was a limitation of the procedure, given that the accuracy of the measurements was on the order of 0.1 mm, this was not viewed as a serious limitation, and zero-padding these regions yielded a reasonable estimate of the "missing" data. For the "pressed" mold as shown in Fig. 5, the entire top 5 mm of the data were missing.

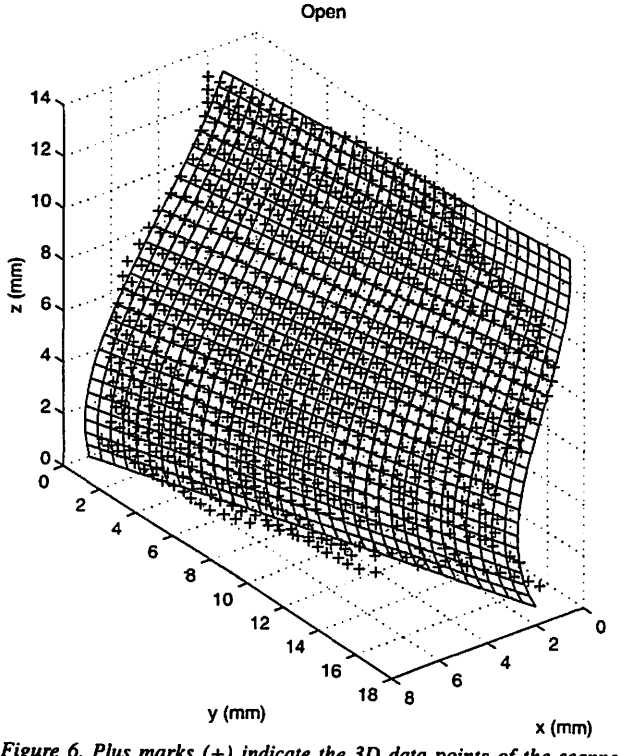


Figure 6. Plus marks (+) indicate the 3D data points of the scanned mold corresponding to the 1 mm "open" glottal configuration. The mesh corresponds to the best polynomial fit for M=*1 (linear in the anterior-posterior direction) and* N=*3 (cubic in the vertical direction).*

For the case of the separated folds there were no missing data; however, there were ambiguities associated with the data analysis. In this case, the *COD* data alone was not sufficient to make a convincing argument as to which power of *y* and *z* might best describe the curvature of the folds. For example, was a quadratic function of *z* sufficient to describe the curvature of the folds, or was a cubic function necessary? In the *y* variable, did a linear adequately describe the curvature, or was a quadratic necessary? Plotting the various models against the raw data yielded somes clues.

Fig. 6 shows the results of a cubic fit of the *z* curvature (*M*=3 and *N*=1). Although the cubic polynomial (*COD* = 93.9%) was not a great deal better than the quadratic (*COD* = 92.5%) according to the data in Table 1, physically the cubic yielded a much better representation of the mold. As can be seen in Fig. 6, the convexity of the folds changed from top to bottom (i.e., from concave to convex). A cubic polynomial was necessary to capture this variation in curvature.

Notice, however, that the polynomial fit was optimized over the top 13 mm of the airway. Although this thickness was chosen in order to capture some of the subglottal airway (which is important for modeling airflow), traditionally it is understood that only the top 3 - 5 mm of the folds are important in terms of vocal fold vibration. For the case of the cubic polynomial, the vertical inflection point was calculated in order to isolate just the "upper" curvature. The inflection point was 7.61 mm from the top anteriorly, and 8.23 mm from the top posteriorly. Midway along the vocal fold length, the inflection point was calculated to be 7.9 mm from the top. For simplicity, the inflection point was rounded off to 8 mm even. Because the mold captured the top 13 mm of the glottal airway, this inflection point corresponded to roughly the top 5 mm of the mold.

In Table 2 (following page), the *COD* percentages were calculated again using only the data corresponding to the vibrating portion of the molds (the upper 5 mm). Taking this in account, across all glottal configurations, the data suggest that a linear function in *y* and a quadratic function

Table 2.
Re-Calculation of *COD* Percentages for the Open Larynx Using Just the Upper 5 mm of the Mold

Max. Power in z (N)	Max. power in y (M)		
	0	1	2
0	0.0	44.3	--
1	29.1	89.7	90.8
2	--	96.4	98.0
3	--	96.6	--

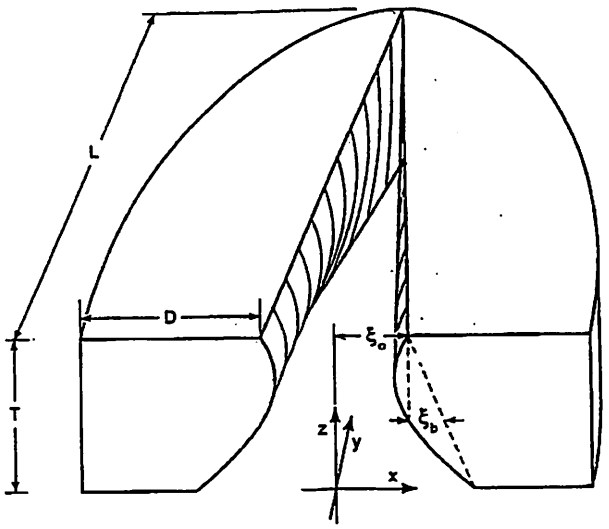
in z are sufficient to capture the curvature of the folds. This is a preliminary confirmation of the hypothesis proposed earlier by Titze (1989) that the y curvature of the pre-phonatory glottis is linear and that the z curvature is quadratic. Specifically, he hypothesized the pre-phonatory glottis could be characterized as follows:

$$X(y,z) = \left(1-\frac{y}{L}\right)\left[\xi_0+\left(\xi_c-4\xi_b\frac{z}{T}\right)\left(1-\frac{z}{T}\right)\right] \qquad (3)$$

where L is the anterior-posterior length of the folds, T is the vertical thickness, ξ_0 is the superior glottal half-width, ξ_c is the convergence coefficient describing the linear variation in z, and ξ_b is the bulging coefficient describing the quadratic variation in z. The physical meaning of these coefficients is further illustrated in Fig. 7.

Although the polynomial of Equation 3 is linear in y and quadratic z, it is obviously more restrictive than the general polynomial of Equation 1. However, one advantage of the more restrictive definition is that each parameter has a precise physical interpretation. Of course, the usefulness of this characterization must be judged by how well it matches the empirical data. To perform this evaluation, Equation 3 was fit to the same data as Equation 1. In every case, the z equals zero level was set 5 mm below the assumed top of the folds. The results of this optimization procedure are shown in Table 3.

By comparing Tables 1 and 3 (Tables 2 and 3 for the case of the open larynx), one can see that the *COD* percentages drop any where from 0.1 - 2.0%. The values of Tables 1 (Table 2 for the open larynx) which correspond to Table 3 are found in the column where M equals 1 and in the row where N equals 2. This slight drop in the goodness-of-fit did not appear to be significant. While the fit of either equation was acceptable, the direct physical interpretation of the shaping parameters from Equation 3 seemed to outweigh the slight loss in goodness-of-fit. An illustration



$$g(y,z) = \left(1-\frac{y}{L}\right)\left[\xi_o+\left(\xi_c-4\xi_b\frac{z}{T}\right)\left(1-\frac{z}{T}\right)\right]$$

Figure 7. A drawing of the pre-phonatory shaping function previously proposed by Titze (1989), illustrating the coefficients introduced in Equation 3.

Table 3.
Shaping Parameters (in mm) From Equation 3 and *COD* Percentages for the Three Molds of Larynx 1

	ξ_o	ξ_c	ξ_b	L	T	COD (%)
Open Larynx	1.71	2.75	1.03	20.0	5.00	94.4
Just Barely Adducted	0.01	1.65	0.16	18.7	4.12	94.7
Pressed Arytenoids	0.00	0.26	0.18	20.0	5.07	95.2

of how the "optimized" polynomial (for Equation 3) fit the "open" mold is shown in Fig. 8.

The shaping parameters of Table 3 appeared to be quite descriptive of the molds, at least the upper portions of the molds corresponding to the vibrating folds. The open larynx had the greatest glottal half-width ξ_o, the greatest linear convergence ξ_c (the surfaces converged medially when traversing the surfaces in a superior direction), and the greatest quadratic bulging ξ_b. For both the adducted and pressed molds, the glottal half-with was essentially zero as expected. The glottal convergence of the adducted mold was only 60% of that of the open larynx, and the glottal convergence of the pressed mold was less than 10%. In terms of glottal bulging, both the adducted folds and the pressed folds had small values, i.e., less than 20% of that of the open larynx.

One might argue whether the vocal fold length L and the thickness T should have been allowed to vary in the optimization procedure. Certainly, if the model gave large

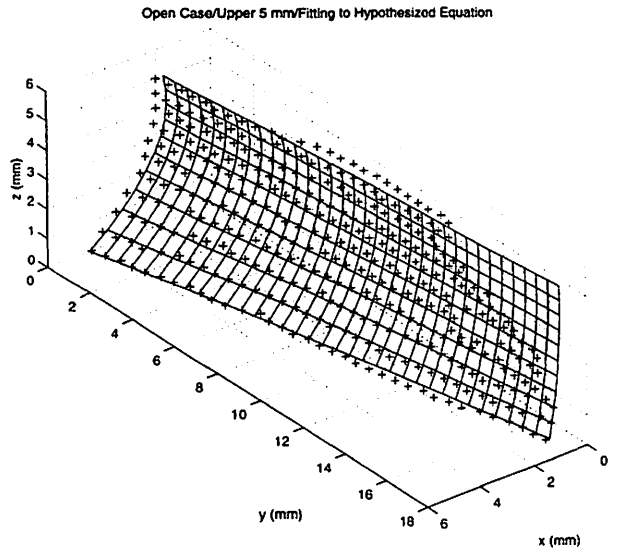


Figure 8. Plus marks (+) indicate the 3D data points of the scanned mold corresponding to the 1 mm "open" glottal configuration, focusing on just the top 5 mm or so where most tissue vibration would occur. The mesh corresponds to the best polynomial fit to Equation 3, using the optimized shaping coefficients presented in Table 3.

deviations from measured values, the technique would be suspect. There are several reasons why it was thought valid to allow such an optimization here. One reason is that the vocal fold length and effective thickness of the folds might truly vary as a function of glottal adduction. Another reason was that the data didn't necessarily extend all the way to the anterior commissure as the model assumes. For example, as indicated in our earlier description of Fig. 3, sometimes the length had to be trimmed posteriorly or anteriorly because of noisy data near these end points. By allowing the geometrical dimensions L and T to vary, an optimum alignment between model and data was allowed, while still retaining the basic form of the equation. Finally, in comparing the optimization results of Equations 1 and 3, Equation 1 had six parameters to vary (for the case where M equals 1 and N equals 2), whereas Equation 3 would have only three parameters if the values of L and T were fixed. By allowing a variation in L and T for Equation 3, both Equations 1 and 3 had a similar number parameters to optimize. Optimization results yielded comparable goodness-of-fit statistics for the two equations.

Summary and Conclusions

A technique was presented for quantifying the prephonatory shape of the vocal folds mathematically on excised larynxes. A primary limitation of this method is that the thyroarytenoid muscle is not activated in an excised larynx. However, this is a limitation that can only be overcome with imaging at higher levels of precision than are currently available on human subjects. Thus, a molding procedure was used to capture the shape of the excised larynxes. Ultimately, a lead mold was generated, which was subsequently digitized using a 3D laser-striping system with a precision of 0.1 mm. Three molds were created, digitized, and fit to low-order polynomials.

Through analysis of the coefficient of determination for various polynomial models of glottal shape, it was argued that, minimally, a linear variation in y and z was needed to capture the curvature of the folds. Further modest gains were achieved by allowing a quadratic variation in z. However, any further increase in the order of the polynomial did little to improve the correspondence between model and data. Consequently, based on the three molds analyzed in this preliminary study, it was argued that a linear variation in y and a quadratic variation in z was sufficient to capture the curvature of the folds across a range of glottal adductions.

This result is a preliminary confirmation of an earlier hypothesis by Titze (1989), in which he presents an equation (Equation 3) to describe the pre-phonatory shape of the folds based on glottal convergence and bulging coefficients. In this study, the coefficients were optimized for each of the three molds, and were enumerated in Table 3. These coefficients provided a useful description of the prephonatory shape of the vocal folds, and could be directly inserted into our models of vocal fold vibration.

Our technique has been shown to capture the prephonatory glottal shape with a precision on the order of 0.1 mm. With the technique and results presented in this study, we now have the confidence to proceed with further studies in which greater numbers of molds will be generated, digitized, and analyzed. Such investigations are of extreme importance because initial conditions and boundary conditions are know to have a profound impact on vocal fold vibration, just as they do on many other nonlinear dynamical systems.

Acknowledgments

This work was supported by grant number P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. Special acknowledgment is given to Karin Titze Cox whose earlier investigations (K.A. Titze and I.R. Titze, unpublished manuscript) yielded fruitful insights into the present study, and to Cindy Steffensmeier who manually digitized one of our initial molds using a 3D micrometer. The authors also thank the Department of Artificial Intelligence of the University of Edinburgh for use of their laser striping system.

References

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

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

Berry, D.A., Herzel, H., Titze, I.R., Story, B.H. (1996). Bifurcations in excised larynx experiments, J. Voice, 10, 129-138.

Berry, D.A., Titze, I.R. (1996). Normal modes in a continuum model of vocal fold tissues, J. Acoust. Soc. Am. 100, 3345-3354.

Fisher, R., Trucco, E., Fitzgibbon, A., Waite, M., and Orr, M. (1993). IMAGE: A 3-D Vision System, Sensor Review 13, 23-26.

Inagi, K., Khidr, A., Ford, C., Bless, D., & Heisey, D. (1997). Correlation between vocal functions and glottal measurements in patients with unilateral vocal fold paralysis. Laryngoscope, 107(6), 782-791.

Kaneko, T., Uchida, K., Suzuki, H., Komatsu, K., Kanesaka, T., and Kobayashi, N., and Naito, J. (1981). Ultrasonic observations of vocal fold vibration, in Vocal Fold Physiology, edited by K.N. Stevens and M. Hirano, (University of Tokyo Press, Japan), pp. 107-118.

Kaneko, T., Komatsu, K., Suzuki, H., Kanesaka, T., Masuda, T., Numata, T., and Naito, J. (1983). Mechanical properties of the human vocal fold—Resonance characteristics in living humans and in excised larynes, in Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control, edited by I.R. Titze and R.C. Scherer (The Denver Center for the Performing Arts, Denver), pp. 304-317.

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

Titze, K.A. and Titze, I.R. (1988). The geometry of the pre-phonatory glottis in canine larynges, unpublished manuscript.

Titze, I.R., and 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.

Trucco, E., Fisher, R., Fitzgibbon, A., Naidu, D.K. (1998). Calibration, Data Consistency and Model Acquisition with a 3-D Laser Striper, Int. J. of Computer Integrated Manufacturing 11, 292-310.

Woodson, G.E., Hengesteg, A., Rosen, C.A., Yeung, D., Chen, N. (1997). Changes in length and spatial orientation of the vocal fold with arytenoid adduction in cadaver larynges, Annals of Otology, Rhinology & Laryngology 106, 552-555.

Woodson, G.E., and Murray T. (1994). Glottic configuration after arytenoid adduction, Laryngoscope 104, 965-969.

NCVS Status and Progress Report - 14
September 1999, 35-42

Dynamic Glottal Pressures in an Excised Hemilarynx Model

Fariborz Alipour, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Ronald Scherer, Ph.D.
Department of Communication Disorders, Bowling Green State University

Abstract

During phonation, air pressures act upon the vocal folds to maintain their oscillation. The air pressures vary dynamically along the medial surface of the vocal folds, although no studies have shown how those pressure profiles vary in time. The purpose of this study was to examine time dependent glottal pressure profiles using a canine hemilarynx approach. The larynx tissue was cut in the midsaggital plane from the top to about 5 mm below the vocal folds. The right half was replaced with a Plexiglas pane with imbedded pressure taps. Simultaneous recordings were made of glottal pressure signals, subglottal pressure, particle velocity, and average airflow at various levels of adduction. The data indicate that the pressures in the glottis vary both vertically and longitudinally throughout the phonatory cycle. Pressures vary most widely near the location of maximum vibratory amplitude, and can include negative pressures during a portion of the cycle. Pressures anterior and posterior to the maximum amplitude location may have less variation and may remain positive throughout the cycle, giving rise to a new concept called dynamic bidirectional pressure gradients in the glottis. This is an important concept that may relate strongly to tissue health as well as basic oscillatory mechanics.

Introduction

Vocal fold oscillation during phonation depends upon changing tissue and aerodynamic factors throughout the cycle. Numerous models of phonation suggest that the air pressure within the glottis changes in the vertical direction through the glottis and is dependent upon glottal shape, diameter and transglottal pressure. Numerous studies have investigated these intraglottal pressures, including steady flow physical models[1-4], multi-mass computer models[5-7], computational models[8-11], and excised tissue[12,13]. It appears that longitudinal (anterior-posterior) pressures within the glottis have not been examined. The glottis is not a rectangular orifice as some modeling has assumed (for simplicity). Also excised laryngeal research has shown that the air velocity through the glottis varies depending on the longitudinal location[14-16]. Knowing how the longitudinal pressures vary, and how pressures vary in general throughout the three-dimensional glottis, are prerequisites to more complete models of phonation and associated implications for clinical and training interventions.

The primary purpose of this research was to describe longitudinal and vertical pressure variations within the glottis. In order to do so, a hemilarynx canine model was selected. For this research, pressure transducers were placed within a Plexiglas wall opposite an oscillating vocal fold. As the vocal fold vibrated due to an imposed subglottal air pressure, the pressures within the glottis varied, and those pressures on the Plexiglas wall were recorded. An array of pressure transducers was used so that pressure variations vertically as well as longitudinally could be obtained simultaneously. In this research, shims at the vocal processes were used to vary the level of adduction. Also, hot wire anemometry was used above the vocal folds to help determine timing aspects of the pressure fluctuations.

The hemilarynx set-up involves using only one of the vocal folds vibrating next to a flat surface. The advantage to such an approach is bench-control of the experimental set-up, as suggested above. The vocal fold does not meet a similar vocal fold on the other side (with the associated compression between the two vocal folds during closure), but meets an immovable pane. Differences in the col-

lision mechanics may alter the normal vibratory characteristics of the vocal fold, and the extent of that difference is unknown.

Jiang and Titze[12,13] reported studies of intraglottal pressures using a hemilarynx approach. They used a single pressure transducer at various glottis locations and found that the impact pressures on the vocal fold surfaces increased with subglottal pressure and adduction. They also found that the midportion of the vocal fold received the largest impact pressures compared to locations anterior or posterior to the midportion. This is consistent with the finding of this study that the midportion location has the largest intraglottal pressure variations.

Methodology

Excised canine larynges were obtained following Cardiovascular Research experiments at the University of Iowa Hospitals and Clinics. They were quick-frozen using liquid nitrogen for storage and slowly thawed in a refrigerator prior to use. The larynges were cut in half along the midsagittal plane. Excess tissue was removed, leaving the true vocal fold, arytenoid cartilage, and the thyroid and cricoid cartilages on the retained side. The hemilarynx was mounted on a 20 mm (OD) stainless steel tubing mimicking the trachea so that the glottis was easily viewed by a camera and accessible by the equipment (see Figure 1). The tubing had a convergence plate and a half-cylindrical shape at the laryngeal level. A 60 x 60-mm quarter-inch sheet of clear Plexiglas had an array of holes drilled into it for later passage of pressure transducers (see Figure 1). The Plexiglas sheet was mounted to the stainless steel tubing. The hemilarynx was glued to the Plexiglas plate at the thyroid and cricoid cartilage ends.

Adduction was aided by the placement of metal shims of various thicknesses (0.1 - 1.0 mm) between the arytenoid cartilage posterior to the vocal processes and Plexiglas pane. A two-pronged probe was used to keep the shims in place (see Figure 2). No superior duct (vocal tract) was used in this study. Air was humidified to approximately 95-100% humidity and heated to 36-38°C using a Concha Therm III Servo Control Heater unit (RCI Laboratories).

Intraglottal air pressures were measured with miniature pressure transducers (Entran EPE-551). The outer diameter of these transducers was 2.36 mm. They fit firmly within the holes drilled into the Plexiglas sheet. The frequency response of the transducers was DC to 80 kHz. The holes in the sheet that did not receive transducers were covered with thin adhesive tape. The mean subglottal pressure at a location 10 cm below the glottis was monitored with a wall mounted water manometer (Dwyer No. 1230-8). The time-varying subglottal pressure was recorded using a piezoresistive pressure sensor (Microswitch 136PC01G1) at the same location as the manometer tap. The bandwidth for the pressure sensor was approximately 0-1 kHz. The mean flow rate was monitored with an in-line flowmeter (Gilmont model J197). Particle velocity exiting the glottis was monitored and recorded using a hot fiber-film probe (Dantec, 55R04) situated approximately 10 mm above the glottis and about 1 mm away from the Plexiglas sheet. An 8-channel Sony DAT recorder was used to record the pressure and velocity signals. Selected signals were monitored on a 4-channel Data Precision digital oscilloscope.

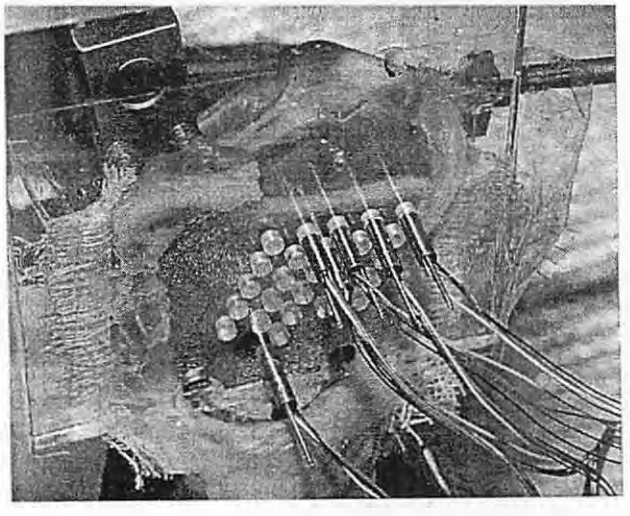


Figure 1. A side view of the Plexiglas sheet, imbedded pressure transducers (foreground), and mounted hemilarynx (on the other side of the Plexiglas).

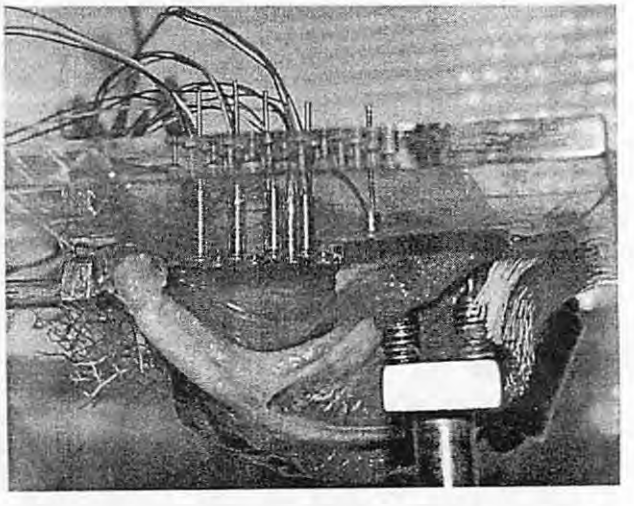


Figure 2. Top view of the mounting of the hemilarynx showing the tissue (lower in the figure), Plexiglas sheet, and pressure transducers. Also shown is the two-pronged adduction rod used to press the arytenoid cartilage against the Plexiglas.

Results

Mean subglottal pressure versus mean glottal flow: The authors recently presented data[17] showing relatively linear relationships between the mean subglottal pressure and mean glottal flow at different glottal adduction levels for excised canine larynxes. The linear relationship also appears to hold for the hemilarynges of this study. Figure 3 illustrates this for hemilarynx HL55. The slopes of the lines indicate the (differential) flow resistance values for each value of adduction. The range of flow resistance in the current study was approximately 14 to 69 cm-H_2O/LPS. The mean resistance in this study of 31 ± 14 cm-H_2O/LPS was about half that found in the earlier study[17] (64 cm-H_2O/LPS over a range of 28 to 109 cm-H_2O/LPS).

Cyclic orientation of pressure signals: In this experiment, EGG or some other signal that would indicate phase characteristics of the cycles was not feasible. To suggest the cyclic correspondence of the pressure signals, the jet velocity signal measured with an anemometry system was used. Figure 4 shows two pressure signals and the velocity signal for hemilarynx HL51. The hot wire was placed approximately 1.0 mm away from the Plexiglas pane, 11 mm above the vocal folds, and 9 mm (60%) from the anterior commissure. Pressures P3 and P4 were placed opposite the medial portion of the vibrating vocal fold, and appear to be in the lower half of the glottis during vocal fold vibration (seen from the stroboscopic record). Pressure P4 was placed about 11 mm (73%) from the anterior commissure, P3 was about 1 mm in front of the vocal process, and the vocal fold length (anterior commissure to vocal processes) was approximately 15 mm (see figure insert). For this case, the adduction shim was 0.47 mm, subglottal pressure was 28.9 cm-H_2O, and maximum excursion of the vocal folds occurred at a distance of approximately 5 mm from the anterior commissure. The peak velocity would have occurred near the time when the vocal folds were nearly maximally displaced, at which time the pressures are seen to be negative at the P3 and P4 locations. This was followed by apparent lack of closure upon the P3 and P4 taps

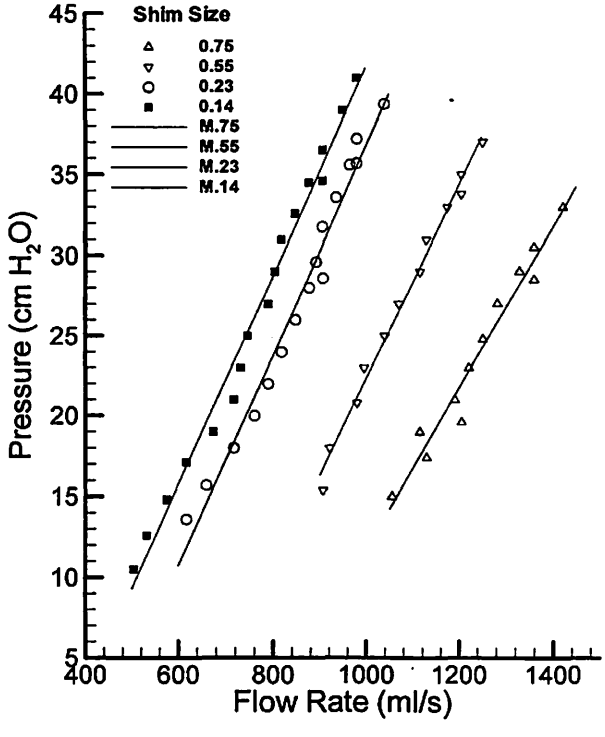


Figure 3. A typical pressure-flow relationship for a hemilarynx during phonation at various adductionn levels (shim sizes).

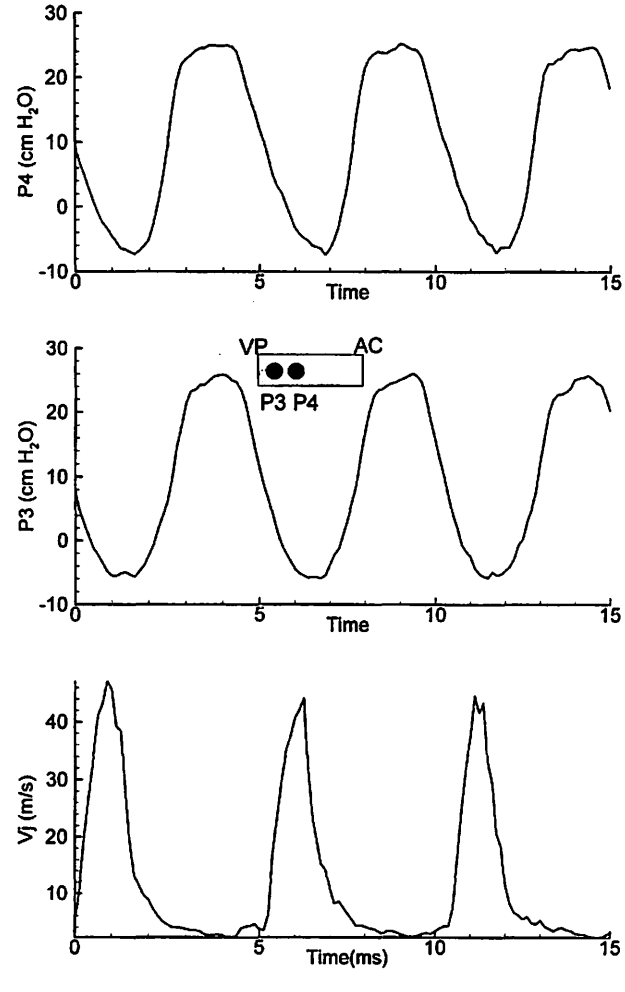


Figure 4. Typical pressure and velocity waveforms for hemilarynx HL51 during phonation. V_j is the velocity signal, P3 and P4 are pressures in the glottis. See the insert for locations of the pressure transducers. Refer to the text for details.

themselves, but movement toward closure just above those two locations, giving rise to the positive pressures registered by P3 and P4. During the maximum closure interval, the velocity was at minimum values (essentially zero) and the P3 and P4 pressures were at their maximum values (about 10 % below the mean subglottal pressure value). The stroboscopic record indicated that the closed quotient in the vicinity of P3 and P4 was approximately 30 % or 1.7 ms (obtained by counting frames). This amount of time appears to correspond to the width of the top portion of the two pressure signals, between the strongly sloped sides.

Pressure distributions in the glottis: Figure 5 shows another example of the relative timing between the velocity signal and pressure traces, but this time the pressures are distributed in two dimensions, vertically and longitudinally. The figure is for hemilarynx HL52, for a subglottal pressure of 28 cm-H_2O and shim size of 0.14 mm. Glottal length was 14 mm from the anterior commissure to the vocal processes. Pressure tap P1 was situated below the glottis about 1.5 mm anterior to the vocal processes. P2 was approximately on the lower edge of the dynamic glottis and 3 mm anterior to the vocal process. P3 was within the glottis and 8 mm from the anterior commissure, and P4 was in the glottis and 4 mm from the anterior commissure (see the figure insert). Greatest vocal fold lateral motion was located between P3 and P4 at approximately 5 mm from the anterior commissure (closer to the P4 than to the P3 position). Closure proceeded from posterior to anterior, and opening was essentially simultaneous along the superior glottal margin. This visual observation is consistent with the timing of the upswing of the pressure signals. A rise in pressure presumably reflects glottal closure above or in the vicinity of the pressure taps. P1 and P2 rise prior to P3, and P3 rises just prior to P4, reflecting the posterior to anterior closing motion. The simultaneous decrease in all pressure signals is consistent with the visual observation of simultaneous longitudinal opening of the superior glottis, such an opening releasing the aerodynamic pressures on the taps.

The velocity probe was located above the P3 location. It appears, however, that the primary effect on the velocity is from the P1 and P2 locations: the rise and fall of the velocity signal corresponds to the fall and rise, respectively, of the P1 and P2 pressure signals. This could be explained by a direction of flow that is slightly slanted from posterior to anterior, which would be consistent with the airway bend created by the arytenoid cartilage. That is, the air that bends around the arytenoid from the trachea may cause a slant to the flow, so the velocity at the fiber-film probe would be influenced by glottal activity slightly posterior at the P3 location.

Approximate closure quotients were estimated by frame counting the stroboscopic record. The estimated closure quotient for the lower glottis near P2 was 38%, and

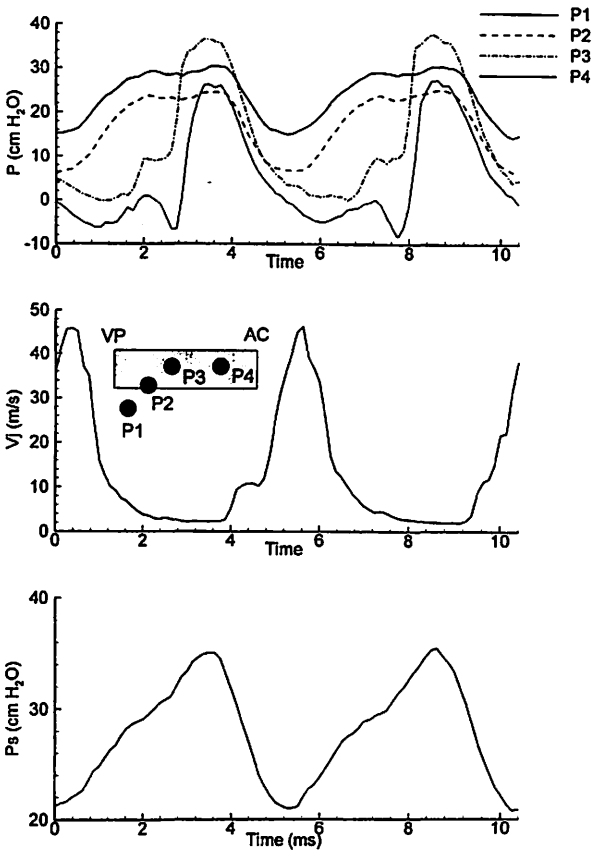


Figure 5. Simultaneous pressure and velocity waveforms for hemilarynx HL52 during phonation. Ps is the subglottal pressure, V_j is the velocity signal, and P1 through P4 are glottal pressures. See the insert for locations of the pressure transducers. Refer to the text for details.

near P3 was 21%. These correspond to 2.09 ms and 1.2 ms, which are the approximate widths of the flatter portions at the top of the P2 and P3 pressure traces, respectively.

It is noted that the P4 pressure tap shows negative pressures during the glottal open phase, and P3 nearly so. This may be consistent with the findings from other larynxes in this study that pressures appear to reach more negative values near the location of maximum oscillation. The mean pressure for the P4 location was lower than for the P3 location. The greatest negative pressures appear to be followed by a relatively fast upswing of pressure during (presumed) glottal closure. This relation, seen in figures 5 and 6, is similar to the mean glottal air pressure signal idealized by Titze[18].

Also shown in Figure 5 is the AC portion of the subglottal pressure signal (about 10 cm below the glottis). The subglottal pressure is in phase with the P1 and P2 glottal pressures and appears to be affected by subglottal resonance. Most of the subglottal signals of this study show that the subglottal pressure is in phase with the glottal pressures.

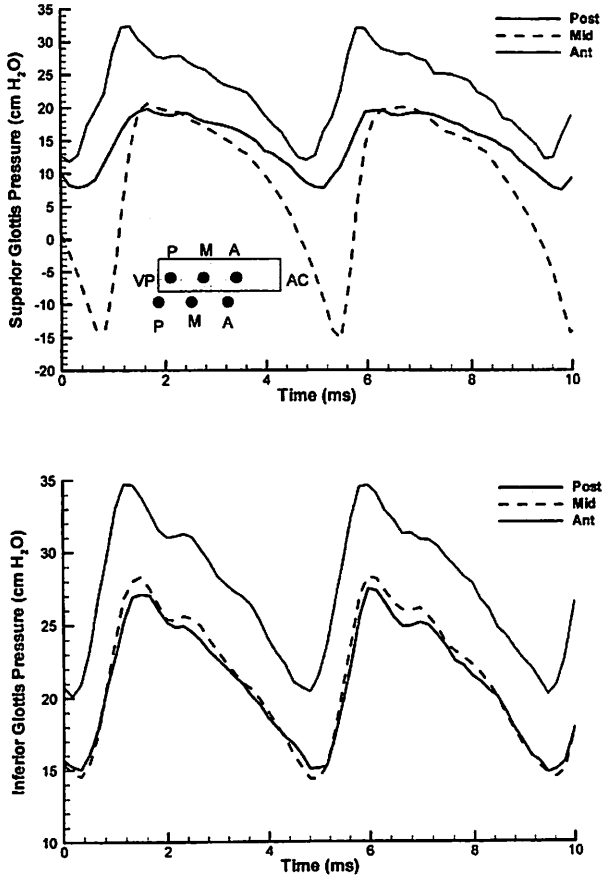


Figure 6. Pressure waveforms for two rows of three pressure transducers in the glottis, one row more inferior to the other. The top traces refer to the upper row, the bottom traces to the lower row. See insert figure and text.

Figure 6 shows glottal pressures for hemilarynx HL54. There were two longitudinal rows of three intraglottal pressure transducers for HL54 (see figure insert). The top row was within the glottis at 28%, 59%, and 89% (anterior, mid and posterior) distances from the anterior commissure to the vocal processes. The exact location of the taps during glottal vibration was not possible to determine from the stroboscopic record. The bottom row was estimated to be just below the glottis and at 39%, 72%, and 103% (anterior, mid, and posterior) distance from the anterior commissure to the vocal processes. The last mentioned pressure tap was below the vocal processes. The stroboscopic record indicated that the greatest dynamic adduction (how close the vocal fold gets to the Plexiglas pane during the oscillation) was near the anterior pressure tap, less adduction near the mid section pressure tap, and the least amount of adduction at the posterior pressure tap. Figure 6 overlaps the pressures from the three upper locations within the glottis (top trace, for a subglottal pressure of 27 cm-H_2O and a shim size of 0.75 cm). The mid glottis pressure trace indicates a negative pressure dip, whereas the other two pressure traces

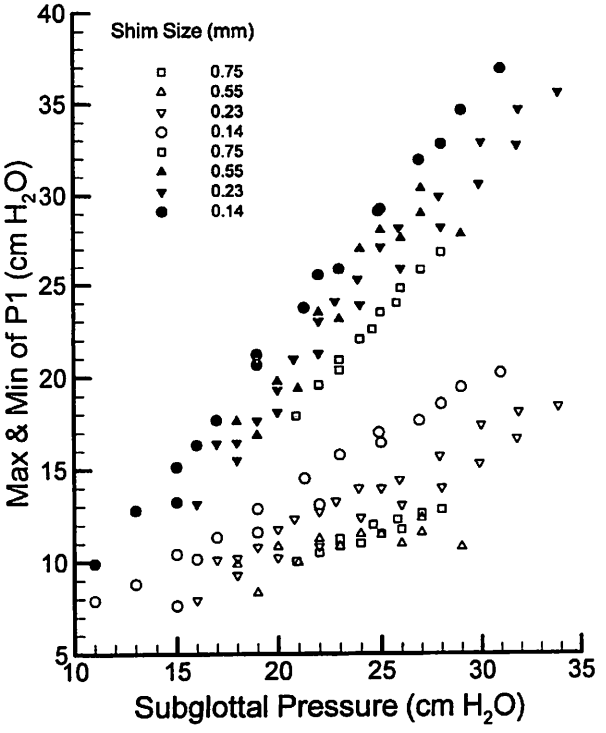


Figure 7. Maximum and minimum pressure in the hemilarynx of Figure 6 at the lower posterior pressure tap.

remain positive. The pressure variation for the mid pressure tap was about 35 cm-H_2O, which exceeded the mean subglottal pressure. Consistent with the discussion above for HL52 (Figure 5), the greatest negative pressure (and the lowest pressures for the anterior and posterior pressure taps) occur just prior the upswing in pressure during (presumed) glottal closure. The negative pressure for the mid-glottal location would suggest that the folds were aided in their return, at least in that section of the glottis. It was difficult to determine from the stroboscopic record if the pressure variation was due to any phasing differences among those three tap locations. The average pressure for the three pressure taps was largest for the anterior location, and lowest for the mid-glottis location (similar to that found for HL52).

Figure 6 also gives the pressures for the lower row of pressure taps in the Plexiglas pane for HL54. These pressures were all positive, with the mid and posterior taps giving the lowest pressures. The pressure variations were less than the mean subglottal pressure (by about half).

Figure 7 is a record of how the maximum and minimum glottal pressures at tap P1 (located at the lower posterior position of the glottis) changed as a function of the subglottal pressure and shim size for HL54 just discussed. Solid symbols represent maximum and hollow symbols represent minimum values. Both the maximum and minimum pressure values increased essentially linearly with subglottal pressure.

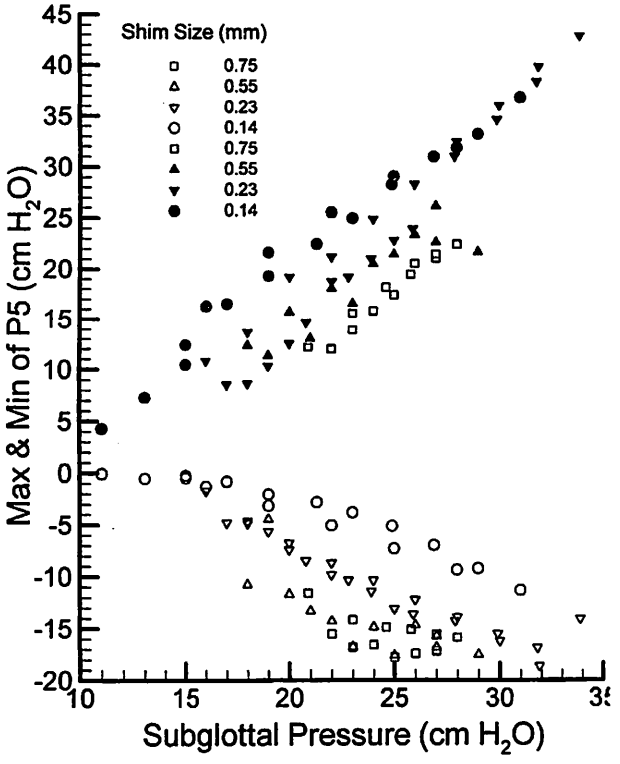


Figure 8. Maximum and minimum pressure in the hemilarynx of Figure 6 at a upper medial pressure tap.

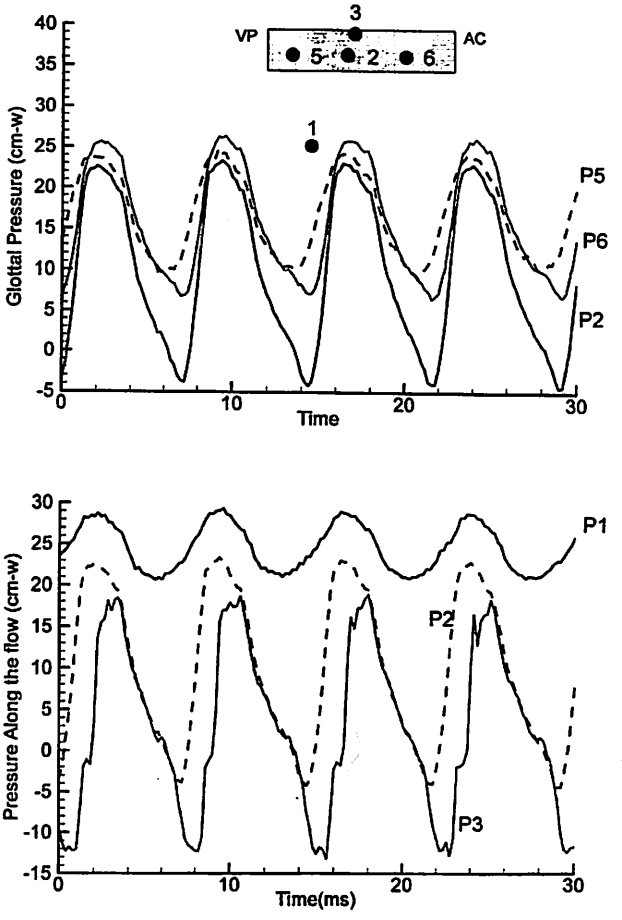


Figure 9. Pressure waveforms within the glottis in the vertical direction (lower traces) and along the length of the glottis (upper traces).

Figure 8 shows similar maximum and minimum values for the pressure at tap P5 (located at the upper medial position of the glottis) for the same hemilarynx. Unlike the pressures at P1, the maximum pressure values increased essentially linearly with subglottal pressures while the minimum pressures decreased with a linear trend. This indicates an increase of the AC pressures at this mid-glottis location as subglottal pressure increased, with increases in both the maximum positive and maximum negative values of phonatory cycle. This would be consistent with the increase of amplitude of oscillation at this point.

Figure 9 shows pressure tracings for a vertical and longitudinal array of transducer taps for hemilarynx HL57. There were three pressure taps along the medial vocal fold region, and three in the vertical direction (see figure insert). Glottal length was 15 mm and subglottal pressure was 23 cm-H_2O, with a 0.5 mm shim. Pressure taps in the glottal region from the anterior commissure were 13 mm, 10 mm, and 5 mm for taps P5, P2, and P6, respectively. Thus these taps indicate posterior, mid, and anterior tap locations. Tap P1 was 15 mm below the top of the vocal folds, and represented a subglottal pressure location. Tap P3 was located 8 mm from the anterior commissure, just above and anterior to tap P2, and represented a superior glottal tap position. The stroboscopic record indicated that the glottis closed in the posterior to the anterior direction, but opened simultaneously along the glottis. There appeared to be full closure over P2 and P3, almost (but not quite) closure over P6, and lack of closure at the posterior tap, P5. The closed quotient for the P2 location was about 19%, and 13% for the P3 location. The maximum amplitude of glottal motion was approximately at the P2 location. Figure 9 once again indicates that the mid glottal pressure tap was associated with negative air pressures during a portion of the cycle, while the anterior and posterior locations were not associated with negative pressures (upper figure), a finding consistent with the results seen in Figures 5, 6, and 8. The pressure fluctuation at the mid glottis location was larger than those for the anterior and posterior pressure taps, a finding also consistent with earlier observations. The lower figure, showing pressure traces for the vertical set of taps, suggests that the upper glottal pressure tap (P3) was associated with greater negative pressure values than the lower glottal tap (P2). One possibility may be that the upper tap P3 was influenced more by possible rarefaction pressures during closure than P2. The subglottal pressure P1 appears similar to the subglottal pressure tap in Figure 5 for HL52 in that it shows relatively smaller AC variations.

Discussion

The hemilarynx was chosen for this study of dynamic intraglottal pressures because of the difficulty in securing pressure transducers or pressure responsive materials to the vocal folds, and the ease of attaching small pressure transducers within a rigid sheet against which the vocal fold could vibrate. The lack of two vocal folds may alter the vibratory characteristics of the one vocal fold due to the unnatural wall against which it must come into contact. Reasonable vocal fold motion can be obtained, however[12]. The primary assumption made in this study is that the aerodynamic pressures that are present on the flat wall opposite the vibrating vocal fold are similar to the pressures on the vocal fold itself at the same vertical (axial) levels. This assumes that pressures are relatively constant across the cross section of the glottis, even for asymmetric glottal shapes. Recent studies with steady flow through rigid models with oblique glottal shapes (Scherer & Shinwari, in review) suggest that the pressures may not be the same on the two sides. But those studies do not take into consideration the unsteadiness of the flow and the moving walls, for which the unsteady pressures may dominate (Alipour et al., in review). We will make the assumption here that the pressures recorded from the Plexiglas sheet do reflect to a reasonable extent the pressures acting on the vocal folds at the axial positions indicated.

A relatively consistent and interesting finding in this study is that the intraglottal pressures near the location of maximum vocal fold motion reach lower or more negative values compared to more anterior and posterior glottal locations. Also, those pressures near the greatest excursion have the greatest variation in value but do not reach the largest positive pressure. This suggests that the glottal tissue associated with maximum glottal motion receives the greatest change in pressure and the greatest negative aerodynamic closure forces (and probably the greatest restoring tissue force due to the largest excursion). Therefore, those tissues may undergo the greatest external (and internal) stresses. Furthermore, the case with hemilarynx HL57 (Figure 9) suggests that more downstream (superior) glottal positions may accentuate the negative pressure swing, although the explanations for two locations may differ. That is, negative pressures for the upstream glottal sections may be due to Bernoulli forces, acceleration of the flow, and flow separation effects[4,19]. This happens when the glottis takes on diverging shapes during the maximum opening and during closing. The negative pressures in the upper portion of the glottis may be due to rarefaction of the pressure as the vocal folds come together[18] and vocal fold curvature[4,8]. The delay of the pressure trace of P3 compared to P2 in the lower figure of Figure 9 is consistent with this view because of the tissue motion delay of the P3 (superior glottal) location compared to the P2 (lower glottal) location.

The large pressure swings and negative pressures at the location of greatest lateral vocal fold motion, and the positive pressures within the more anterior and more posterior glottal locations, suggest a relatively strong (dynamic) pressure gradient along the longitudinal axis of the glottis. Figure 9 for HL57 illustrates this well. For example, the upper traces of that figure at the time of 10 ms indicate that the pressures along the longitudinal glottis near closure are all about the same value. But then as the glottis opens, the pressure differences between the locations increase, and at time 14 ms there is about a 14 cm-H_2O difference in the pressure at the maximum excursion versus the anterior and posterior positions. Indeed, the pressure difference is 14 cm-H_2O between two locations that are 3 to 5 mm apart, yielding an intraglottal *bidirectional pressure gradient* of about 3 to 5 cm-H_2O per millimeter of distance. The results for the hemilarynx HL54 seen in Figure 6 suggest a similar conclusion. For that larynx, the maximum vibration was closer to the mid position pressure tap than to the anterior tap. The pressure taps of the upper traces were 4.1 mm apart. The pressure gradients at a time of about 5 ms were about 4.9 cm-H_2O per mm of distance for the mid location to the posterior location, and 6.6 cm-H_2O per mm of distance for the mid location to the anterior location. These pressure gradients vary throughout the cycle, as shown by the figure.

The bidirectional pressure gradients along the longitudinal direction of the vocal folds suggest important mechanistic consequences. The location of the largest pressure fluctuations and greatest negative pressures in the glottis suggested in this study is consistent with the location of tissue damage at the midmembranous portion of the glottis. It is also consistent with the intraglottal impact studies[13,20,21].

The bidirectional pressure gradient in the glottis also suggests a protective mechanism for the vocal fold tissue. The records of this study indicate that the air pressures in the anterior and posterior glottal regions may be positive over the entire cycle. Positive pressures would apply repelling forces to the two glottal walls. The consequence of this may be to protect the anterior and posterior glottal sections from undue collisions forces. Thus, the dynamic bidirectional pressure gradients may help guarantee vocal fold oscillation while protecting most of the tissue from undue contact forces.

The pressure signals at different locations within the glottis appear to allow an approximation to the relative closed quotients for those different locations. The estimations of this study appear to be consistent with the conclusion that the relatively flat tops to the pressure signals indicate when there is contact near that location. An array, then, of pressure transducers may help to map the dynamic progression of contact of the vocal folds.

Conclusions

This study used a hemilarynx model to study dynamic pressure profiles in the glottal region. A Plexiglas pane with imbedded pressure transducers was used as the flat side opposite hemilarynx canine tissue. The results of this study strongly suggest that the pressures vary along the vertical dimension of the glottis during the phonatory cycle. The range of pressure change at one location may exceed the mean subglottal pressure, and the minimum value may be lower for more superior locations in the glottis. The pressures also vary greatly along the longitudinal aspect of the glottis, with greatest pressure variations near the location of the maximum amplitude of motion. The pressures near the maximum amplitude location may be negative at the same time the pressures are positive at more anterior and posterior locations. Furthermore, the pressures within the glottis may remain positive except for the region near the maximum vibratory motion. The phenomenon of longitudinal pressure differences in the glottis is called here dynamic *bidirectional pressure gradients.* This finding may have a strong relation to the health of the tissue because the negative pressures would tend to pull the tissue together, but the positive pressures away from the maximum amplitude location would help to soften the collisions between the vocal folds.

Acknowledgments

The authors thank Heather Hughes, Karin Cox and Daniel McCabe for their assistance in data collection. This work was supported by NIDCD grant No DC03566-02.

References

1. van den Berg Jw, Zantema JT, Doornenbal PJr. On the air and the Bernoulli effect of the human larynx. *J Acoust Soc Am* 1957;29(5):626-631.

2. Gauffin J, Binh N, Ananthapadmanabha TV, Fant G. Glottal geometry and volume velocity waveform. In Bless DM and Abbs JH (Eds.) *Vocal Fold Physiology: Contemporary Research and Clinical Issues.* College-Hill Press, San Diego 1983:194-201.

3. Scherer RC, Titze IR, Curtis JF. Pressure-flow relationships in two models of the larynx having rectangular glottal shapes. *J Acoust Soc Am,* 1983;73(2):668-676.

4. Scherer RC, Shinwari D. Intraglottal pressure profiles for symmetric and oblique glottal angles. *J Acoust Soc Am* 1999:(in review).

5. Ishizaka K, Flanagan JL. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal,* 1972;51(6):1233-1268.

6. Titze IR. The human vocal cords: A mathematical model, part II, *Phonetica* 1974;29:1-21

7. Story BH, Titze IR. Voice simulation with a body-cover model of the vocal folds. *J Acoust Soc Am* 1995;97(2):1249-1260.

8. Guo CG, Scherer RC. Finite element simulation of glottal flow and pressure. *J Acoust Soc Am* 1993;94(2):688-700.

9. Alipour F, Patel VC. Steady flow through modeled glottal constriction. *Journal of Engineering, Islamic Republic of Iran,* 1994;7(1):13-18.

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

11. Liljencrants J. Analysis by synthesis of glottal airflow in a physical model. *TMH-QPSR* 1996;2:139-142, Royal Institute of Technology, Stockholm.

12. Jiang JJ, Titze IR. A methodological study of hemilaryngeal phonation. *Laryngoscope* 1993;103:872-882.

13. Jiang JJ, Titze IR. Measurement of vocal fold intraglottal pressure and impact stress. *Journal of Voice* 1994;8:132-144.

14. Berke GS, Moore DM, Monkewitz PA, Hanson DG, Gerratt BR. A preliminary study of particle velocity during phonation in an in vivo canine model. *Journal of Voice,* 1989;3(4):306-313.

15. Alipour F, Scherer RC. Pulsatile airflow during phonation: an excised larynx model. *J Acoust Soc Am* 1995;97(2):1241-1248.

16. Bielamowicz S, Berke GS, Kreiman J, Gerratt BR. Exit jet particle velocity in the in vivo canine laryngeal model with variable nerve stimulation. *Journal of Voice,* 1999;13(2):153-160.

17. Alipour F, Scherer RC, Finnegan EM. Pressure-flow relationship during phonation as a function of adduction. *Journal of Voice.* 1997;11:187-194.

18. Titze IR. The physics of small-amplitude oscillation of the vocal folds. *J Acoust Soc Am* 1988;83:1536-1552.

19. Pelorson X, Hirschberg A, van Hassel RR, Wjnands APJ, Auregan Y. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *J Acoust Soc Am* 1994;96(6):3416-3431.

20. Hess, MM, Verdolini K, Bierhals W, Mansmann U, Gross, M. Endolaryngeal contact pressures. *Journal of Voice.* 1998;12:50-67.

21. Verdolini K, Hess MM, Titze IR, Bierhals W, Gross M. Investigation of vocal fold impact stress in human subjects. *Journal of Voice* 1999;13:184-202

NCVS Status and Progress Report - 14
September 1999, 43-46

Irregular Vocal Fold Vibration - High-Speed Observation and Modeling

Patrick Mergell, M.S.
Dept. of Phoniatrics and Pedaudiology, University Erlangen-Nuremberg, Erlangen, Germany
Hanspeter Herzel, Ph.D.
Institute for Theoretical Biology, Humboldt University Berlin, Berlin, Germany
Monika Tigges, M.D.
Ulrich Eysholdt, M.D., Ph.D.
Dept. of Phoniatrics and Pedaudiology, University Erlangen-Nuremberg, Erlangen, Germany
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Irregularities in vocalization signals reveal the dynamical complexity of the vocal organs in humans, mammals, and birds (Nowicki and Capranica, 1986; herzel, 1996; Fee et al., 1998; Mergell et al., 1999; Wilden et al., 1998). Irregular vocal fold vibrations are mostly perceived as a rough voice sound. They are observed in newborn cries (Sirvio and Michelsson, 1976; Mende et al., 1990),in conversational speech (Dolansky and Tjernlund, 1968) and especially in patients with vocal fold lesions, paralysis and other voice disorders (Herzel et al., 1994). Vocal instabilities can be induced either by dysfunctions of the central neural control or by pathologically changes of the mechanical properties intrinsic to the larynx. The understanding of the mechanisms leading to vocal instabilities is essential for the standardization and the objectivation of the clinical diagnosis as well as for voice training concepts.

There are strong indications that many vocal instabilities are manifestations of bifurcations and low-dimensional attractors of the highly nonlinear voice source. Indeed, subharmonics, coexistence of two independent fundamental frequencies (biphonation) (Ishizaka and Isshiki, 1976; Isshiki et al., 1977; Wong et al., 1991; Smith et al., 1992), and chaos (Steinecke and Herzel, 1995)have been found in biomechanical simulations of the vocal fold vibrations. So far, however, no quantitative comparison of observations and modeling of irregular vocal fold dynamics was achieved.

One important focus of this study is the conceptualization of the quantitative description of vocal irregularities due to laryngeal paralyses by combining digital high-speed cinematography and biomechanical modeling. We report on direct observations of non-stationary asymmetric vocal fold oscillations in a female patient. Our results show that a slowly decreasing tension of the healthy vocal fold, with nearly constant tension of the left paralysed vocal fold, induces transitions between different phase-locked episodes. Simulations with an asymmetric two-mass model are presented, which are in quantative agreement with the observations.

Modern high-speed CCD cameras provide the temporal and spatial resolution required for the analysis of asymmetric irregular vocal fold vibrations. Pattern recognition software is applied in order to evaluate the recorded digital image sequences. We used subsequent image processing and data reduction methods for obtaining time series of vocal fold oscillations (Fig. 1; following page). From a digital high-speed sequence we extract one single horizontal scan line from each frame intersecting the vocal folds at half vocal fold length. The resulting gray scale array (kymogram) visualizes the change of the distance from one to the other vocal fold during phonation. By means of a subsequent binary segmentation, the space between the vocal folds, i.e. the glottis (black-coded) is raised from the glottal environment (white-coded). The upper and lower separation lines between the black and white regions correspond to the left and right vocal fold vibrations. We call the extracted time series of the left and right vocal fold vibrations high-speed glottograms (HGG) (Eysholdt et al., 1996; Wittenberg et al., 1995). The extracted oscillation episode is characterized by a 2:3 phase-locking as it can be detected by count-

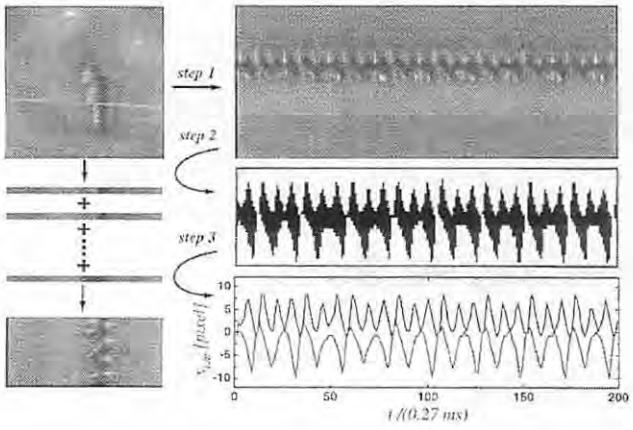


Figure 1. Image processing and data reduction. The analyzed digital high speed video has been recorded with the CAMSYS+ 128 camera system at an image rate of 3704/s (Bloss et al., 1993). The first step of image processing is the extraction of subsequent image scan lines from the high speed video sequence. The method is shown on the left side of the figure. The resulting gray-scale array (kymogram, top right) visualizes the trajectories of the mid-glottal edge points of the left and the right vocal folds. The second step is the binary segmentation of the kymogram. For that purpose, the gray-scale histogram is analyzed in order to detect the glottal components and to determine the corresponding gray-scale threshold. Above this threshold all pixels are coded white, below it all pixels are coded black as it can be seen on the middle right bitmap. The last step is the detection of the black-white boundary which corresponds to the oscillation amplitudes of the vocal folds. As final results one obtains the high-speed glottograms shown in the lower right diagram. The oscillation pattern of this phonation episode exhibits a 2:3 phase-locking.

ing the number of oscillation maxima of each vocal fold during one period Fig. 1). Such an irregular oscillation pattern is typical for laryngeal paralyses (Herzel et al., 1994). Due to lacking muscle contraction of the paralysed vocal fold, the corresponding eigenfrequencies are detuned with respect to the healthy vocal fold. As a consequence, the oscillations become desynchronized and asymmetric vibration rhythms occur.

More details of the dynamics during the complete sample length emerge from the spectrograms of the left and right vocal fold oscillations (Fig. 2). As a direct consequence of the asymmetric vocal fold tensions, the spectra reveal the coexistence of two fundamental frequencies f_{right} and f_{left} which are related to the corresponding vocal fold eigenfrequencies. The frequency of the right vocal fold f_{right} decreases monotonously from about 275 Hz to 220 Hz during phonation and $f_{left} \approx 325$ Hz remains approximately constant.

At a given point of time, the peak positions of all other spectral components are identical to the positive linear combinations $f_{mn} = |mf_{left} + nf_{right}|$ where m and n are integer numbers. These spectra are related to toroidal and entrained oscillations as found in many dynamical systems (Glass and Mackey, 1988). The laryngeal asymmetry can be characterized by the coefficient Q, which is the ratio of

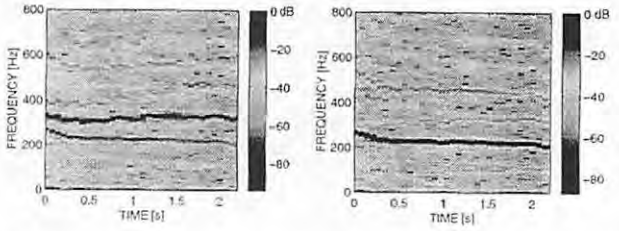


Figure 2. Spectrograms of non-stationary asymmetric HGG and corresponding computer simulations. The spectrograms show subsequent short-time spectra of signal segments. The spectral amplitude is encoded using a color scale. In this way the temporal evolution of spectral components can be visualized. Left/right graph: Spectrogram of the observed left/right vocal fold oscillation. At a certain point of time, the peak positions of the spectral components can be found at positive linear combinations of the two fundamental frequencies f_{left} and f_{right}.

the lower and the higher fundamental frequency. Consequently, Q=1 for symmetric conditions, 0<Q<1 for asymmetric conditions. If the asymmetry coefficient Q is close to a rational number one obtains a discrete spectrum as an indicator of phase-locking (entrainment). Consider for example the 2:3 entrainment at about 2 seconds (Fig. 2) with a stack of frequency peaks at multiples of $f_{left} - f_{right}$. An irrational value of Q indicates the existence of two independent frequencies f_{left} and f_{right} corresponding to so-called toroidal vocal fold oscillations, i.e. biphonation.

The non-stationarity of the measured time series originates mainly from the time-varying asymmetry coefficient. Its time dependence can be expressed with a good accuracy by an exponential decay formula[1] yielding an asymmetry coefficient which decreases from about 0.82 to 0.64. This suggests that there are several regions of phase-locking, e.g. 4:5, 3:4 and 2:3.

The experimentally measured asymmetry is the key for appropriate biomechnical modeling. For that purpose, we have used a model which originates from the work by Ishizaka and Isshiki (1976). It has been simplified by Steinecke and Herzel (1995) in order to focus on the essential features inducing bifurcations of the vocal fold dynamics. Each vocal fold is represented by two coupled oscillators. The driving Bernoulli force which is influenced by the lung pressure and the time-varying glottal geometry induces self-sustained oscillations. In order to adjust the standard model to the specific patient, the higher of the two fundamental frequencies and the glottal area at rest have been increased, effective tissue damping has been decreased and a relatively high lung pressure has been entered into the underlying dynamic equations. However, the crucial adjust-

[1] We used the fitting formula

$$Q_{exp}(t) = \frac{f_{right}(t)}{f_{left}(t)} = a\left(1 + be^{-t/\tau}\right) \quad , a = 0.64, b = 0.28, \tau = 0.98s \qquad (1)$$

to describe the experimental data, where a, b and τ are fitting parameters.

ment was performed by using the experimentally determined asymmetry coefficient. All these pathological parameter changes are typical for many cases of laryngeal paralyses.

An overview of the various nonlinear phenomena occurring in the recorded high speed video sequence can be obtained from differentiated amplitude contours of the vocal fold oscillations (Fig. 3). For direct comparison, these diagrams show the differences of subsequent oscillation maxima of the HGG and of the simulated time series plotted over time. Obviously, the characteristic features of the experimental data can be described authentically by the model simulations. The structural similarity of the differentiated amplitude contours is most evident in the regions around 0.5 s and 2 s where phase-locking appears indicated by a relatively high order of the data sequence. In these time intervals the right-left comparison of the number of contour branches clearly exhibits a 3:4 relation ($Q \approx 3/4$) around 0.5 s which is more pronounced in the model contours and a 2:3 relation ($Q \approx 2/3$) around 2 s. Moreover, the simulations also reveal narrow intervals of 4:6 and 6:9 relations between 1 s and 1.5 s and a 4:5 relation around 0.2 s which can hardly be detected in the HGG contours. Nevertheless, the model data guide the eyes to episodes of phase locking as it can be verified by zooming into the corresponding time intervals of the HGG.

Exemplarily, we show several 2:3 phase-locked oscillation cycles for direct comparison of the experimental and simulated time series (Fig. 3b). The corresponding oscillatory patterns are very similar to each other. Although the observed vocal fold oscillations are non-stationary and irregular, the simplified two-mass model allows to reproduce their temporal characteristics apart from small deviations of phase and modulation amplitude adequately.

The presented method can be applied for a vast array of voice disorders. The observed vocal fold oscillations are representative for the class of frequently occurring pathologies characterized by laryngeal asymmetry without morphological changes, increased glottal opening, and abnormally increased lung pressure. Moreover, the non-stationarity of the oscillations is a common feature of pathological phonation, since a patient with laryngeal lesion tries to compensate the rough voice sound via auditory feedback and additional motor control.

At present, the combined application of high-speed laryngoscopy, image processing, signal analysis, and biomechanical modeling as it has been proposed in this study is a highly efficient way to investigate mechanisms leading to vocal instabilities. A quantitative diagnosis of voice disorders can be given in terms of model parameters which encompass informations about the myoelastic, geometrical, and aerodynamical properties of the laryngeal system. Once a laryngeal configuration is changed pathologically, the central nervous system is temporarily not able to handle the anomalous position in the parameter space. During voice training, compensatory strategies are learned helping to find the way out of the instability regions. If the laryngeal injury cannot be compensated, adequate phonosurgery is indicated for readjusting the crucial parameters.

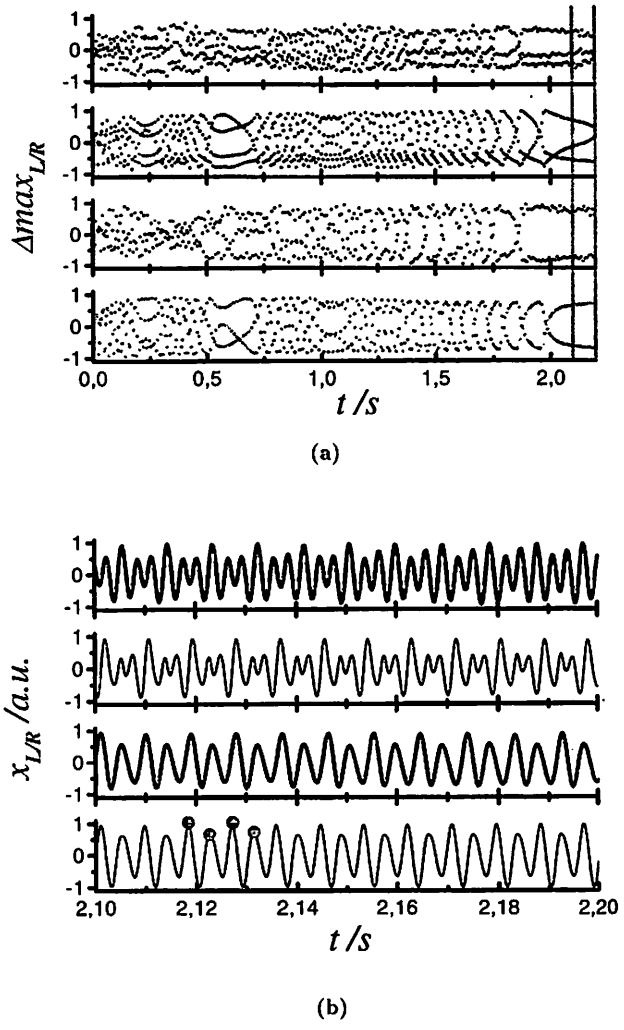


Figure 3. (a) Differentiated amplitude contours of the non-stationary left (blue) and right (red) asymmetric vocal fold oscillations. Prior to maximum detection the observed vocal fold oscillations have been filtered using a bandpass FIR filter (passband 50-750Hz) in order to remove higher order harmonics due to vocal fold contact and low frequency contamination arising from movements of the endoscope relative to the larnyx. On the ordinates the differences of subsequent oscillation maxima are plotted (HGG: dark dots, simulations: light dots). (b) Several periods of the simulated time series (fine lines) and the filtered HGG (thick lines) showing 3:2 phase-locking. This episode corresponds to the time interval which is highlighted by the green markers in differentiated amplitude scatter plots of this figure. The subsequent maxima (green dots) are extracted by using a peak picking algorithm with quadratic interpolation. Since the amplitudes of the experimental and the simulated data are of the same order of magnitude the time series have been divided by their maximum values, i.e. about 1 mm.

References

H. Bloss, C. Backert, and A. Roguse. CAMSYS high speed camera system. Fraunhofer Gesellschaft IIS, Erlangen, Germany, 1993.

L. Dolansky and P. Tjernlund. On certain irregularities of voiced-speech waveforms. *IEEE Transactions.* AU-16:51-56. 1968.

U. Eysholdt, M. Tigges, T. Wittenberg, and U. Proschel. Direct evaluation of high-speed recordings of vocal fold vibrations. *Folia Phoniatr. Logop.*, 48:163-170, 1996.

M.S. Fee, B. Shraiman, B. Pesaran, and P.P. Mitra. The role of nonlinear dynamics of the syriny in vocalizations of a songbird. *Nature,* 395:67-71, 1998.

L. Glass and M. Mackey. *From Clocks to Chaos.* Princeton University Press, 1988.

H. Herzel, D.A. Berry, I.R. Titze, and I. Steinecke. Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling. *Chaos,* 5:30-34, 1995.

H. Herzel. Possible mechanisms of vocal instabilities. In P.J. Davies and P.J. Fletcher, editors, *Vocal Fold Physiology: Controlling Complexity and Chaos,* pages 63-75. Singular Publishing Group, San Diego, 1996.

K. Ishizaka and N. Isshiki. Computer simulation of pathological vocal-cord vibration. *J. Acoust. Soc. Am.*, 60:1193-1198, 1976.

N. Isshiki, M. Tanabe, K. Ishizaka and D. Broad. Clinical significance of asymmetrical vocal cord tension. *Ann. Otol.*, 86:58-66, 1977.

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

P. Mergell, W.T. Fitch, and H. Herzel. Modeling the role of non-human vocal membranes in phonation. *J. Acoust. Soc. Am.*, 105:2020-2028, 1999.

S. Nowicki and R.R. Capranica. Bilateral syringeal interactions in vocal production of oscine bird sound. *Science,* 231:1297-1299, 1986.

P. Sirvio and K. Michelsson. Sound-spectrographic cry analysis of normal and abnormal new-born infants. *Folia phoniatr.*, 28:161-173, 1976.

M.E. Smith, G. S. Berke, B. R. Gerrat, and J. Kreiman. Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration. *J. Speech Hear. Res.*, 35:545-554, 1992.

I. Steinecke and H. Herzel. Bifurcations in an asymmetric vocal fold model. *J. Acoust. Soc. Am.*, 97:1571-1578, 1995.

I. Wilden, H. Herzel, G. Peters, and G. Tembrock. Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics,* 9:171-196, 1998.

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.

D.Wong, M. R. Ito, and N. B. Cox. Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases. *J. Acoust. Soc. Am.*, 89:383-394, 1991.

NCVS Status and Progress Report - 14
September 1999, 47-54

Quantifying Correlations in Pitch- and Amplitude Contours of Sustained Phonation

Hanspeter Herzel, Ph.D.
Institute for Theoretical Biology, Humboldt University, Berlin
Robert Reuter, M.S.
Institute of Electronics, Technical University, Berlin

Abstract

Nonlinear phenomena such as subharmonics, biphonation, or deterministic chaos induce characteristic patterns in pitch- and amplitude contours. We introduce a novel measure of correlations in contours termed MAC (Mean Average Correlation). For uncorrelated and weakly correlated noise the values of MAC are calculated, and they turn out to be rather small. Moreover, we analyze simulations of the circle map and find high values of MAC for subharmonic and chaotic data. We present examples of severely rough voices where correlations in contours are quantified by MAC. Finally, we study the connection between perceived voice quality and MAC with the aid of 120 perceptually evaluated voices.

Introduction

Many of the observed irregularities in voiced speech are related to the intrinsic nonlinearities in the vibrations of the vocal folds. Bifurcations of this nonlinear dynamical system to subharmonics, biphonation (two independent fundamental frequencies), or deterministic chaos often induce rough voice qualities [1-5]. Since the pioneering studies of Lieberman [6, 7] pitch- and amplitude contours are widely used tools in voice analysis. From series of periods and amplitudes various perturbation measures can be derived [8-11]. There are numerous attempts (see, e.g., [12-14] to relate perturbation measures to perceptual categories such as hoarseness, breathiness, and roughness. The perception of voice quality plays a central role in clinical practice and, therefore, a quantification via acoustic parameters would be helpful. However, there is only limited success in that direction [15]. There are several problems associated with the application of perturbation measures:

- The pitch detection itself is still problematic [16-18].
- The statistical properties of the derived perturbation measures have not been studied systematically.
- Trends, measurements noise, frequency jumps, or outliers may influence perturbation measures strongly.
- Roughness is a complex psychoacoustic category [19-22]. For example, the perceived roughness depends strongly on carrier and modulation frequencies. Consequently, the perception of subharmonics depends, for example, on the pitch [23-25].

In this paper we discuss some methodological aspects of perturbation analysis. In particular we argue that correlations in pitch- and amplitude contours are good indicators of nonlinear phenomena. We introduce a novel measure of the amount of correlations - the Mean Absolute Correlation MAC. This quantity is calculated for stochastic and chaotic data and will be used to distinguish roughness and breathiness.

Mean Absolute Correlation - MAC

According to the textbook [26] hoarseness is regarded as an overall measure of perturbations. Breathiness is associated with turbulent noise and a relatively low intensity of overtones [27]. Roughness is related to irregular vibration patterns of vocal folds leading to amplitude- and frequency modulations [12, 28]. Frequently, a four point scale (0 ... 3) is used to evaluate the degree of hoarseness (H), breathiness (B), and roughness (R).

It was suggested [3, 28-30] that correlations in pitch- and amplitude contours are appropriate indicators of roughness. This point of view can be justified by nonlinear dynamics theory: The characteristic bifurcations from a normal voice (a "limit cycle") are jumps to subharmonic

regimes, the appearance of another frequency, and the onset of deterministic chaos [3-5]. These nonlinear phenomena induce certain correlated patterns in pitch- and amplitude contours. Thus the amount of correlations may serve as a measure of roughness due to complex vibratory patterns of the vocal folds. Contrarily, turbulent noise induces primarily uncorrelated fluctuations of periods and amplitudes.

Consequently we test in this paper the hypotheses that the overall amount of perturbations (measured by jitter, shimmer, or standard deviations) quantifies hoarseness (roughness and breathiness) whereas correlations in contours distinguish roughness from breathiness.

In the following we introduce the corresponding statistical quantities. We denote the sequence to be analyzed by x_i (i = 1, 2, ..., L). These values x_i may represent periods or amplitudes directly or differences of them. In many cases it is desirable to study first- or second-order perturbation functions in order to remove trends [3, 10]. In this paper we study generally first-order differences, i.e. the x_i are differences of subsequent periods or amplitudes. The variance of the sequence can be estimated by

$$VAR = \frac{1}{L}\sum_{i=1}^{L}(x_i - \bar{x}_i)^2 \qquad \text{with} \qquad \bar{x}_i = \frac{1}{L}\sum_{i=1}^{L} x_i. \quad (1)$$

We estimate covariance coeffficients as follows:

$$C(k) = \frac{1}{L-k}\sum_{i=1}^{L-k} x_i \cdot x_{i+k} - \left(\frac{1}{L-k}\sum_{i=1}^{L-k} x_i\right)\cdot\left(\frac{1}{L-k}\sum_{i=1}^{L-k} x_{i+k}\right). \quad (2)$$

The mean absolute covariance is defined by

$$< |C(k)| >= \frac{1}{10}\sum_{k=2}^{11} |C(k)|. \quad (3)$$

We exclude the value C(l) from averaging since this value is affected by the trend elimination (see below). The mean is taken over a range of 10 values (k = 2, 3,..., 11) since most periodicities of interest are in that range [3, 28, 31]. Moreover, since contours contain typically a few hundred values (2 s with a pitch of 100 Hz give, e.g., L = 200) the correlation coefficients up to k = 11 can be estimated reliably.

Of course, the mean in Eq. (3) can be generalized easily to another k-range or one can introduce some k-dependent weights to improve the signal-to-noise ratio. In order to be specific, we choose in this paper an averaging given by Eq. (3).

Since covariances cannot be larger than the variance, the normalized quantity

$$MAC = \frac{< |C(k)| >}{VAR} \quad (4)$$

is in the range between zero and one. It will be shown in the remainder of this paper that MAC is indeed a convenient quantification of the strength of correlations in contours.

MAC for Random Noise

Any statistical quantity derived from contours is affected by the finite sample size of the available data. In this section we discuss the statistical properties of estimators of correlation functions [32, 33].

White Noise

For turbulent noise one can assume that subsequent pitch periods or amplitudes are uncorrelated. For simplicity we assume that the x_i are Gaussian white noise. In the limit $L \to \infty$ the correlation coefficients C(k) approach zero. For finite samples, however, there are systematic and statistical deviations from zero. It has been shown [33] that the bias E [C(k)] decays quickly proportional to 1/L. Here E [...] denote expectation values. The standard deviation, however, decays slowly as $\sqrt{1/L}$ and has to be considered in some detail. Straightforward calculations [34] show that the variance of correlation coefficients is about 1/L. These statistical fluctuations lead to the following expectation value of the Mean Average Correlation MAC [34]:

$$E[MAC] = \sqrt{\frac{2}{\pi \cdot L}} \quad (5)$$

For L = 100 we obtain, e.g., a value of about 0.08. The standard deviation of MAC for white noise is much smaller than its expectation value. Hence the value according to equation 5 can be considered as characteristic for white noise.

Correlated Noise

Here we discuss an autoregressive model of first order as a first approximation of correlations between subsequent values x_i

$$x_{i+1} = a \cdot x_i + R_i \qquad (a < 1). \quad (6)$$

Ri represent independent random numbers. The autocorrelation function is given by

$$\frac{C(k)}{VAR} = a^k. \quad (7)$$

Even for $L \to \infty$ we obtain a positive value of MAC:

$$MAC = \frac{1}{10}\sum_{k=2}^{11} a^k = \frac{a^2(1-a^{10})}{10(1-a)} \approx \frac{a^2}{10}. \quad (8)$$

Hence, weak correlations between subsequent periods or amplitudes give only small values of our correlation measure MAC. For example, a = 1/L leads to MAC ≈ 0.05 which

is in the order of the noise level calculated in the preceding section.

Effects of Trend Elimination

Trend elimination can be regarded as a high-pass filter. It is not surprising, therefore, that correlations are induced. Even for uncorrelated values y_i the first order differences

$$x_i = y_i - y_{i-1} \tag{9}$$

exhibit correlations. In the following we derive the corresponding correlation function of x_i. For simplicity we assume that the mean values of x_i and y_i vanish. Then we obtain

$$\begin{aligned} C(k) &= \frac{1}{L-k}\sum_{i=1}^{L-k} x_i \cdot x_{i+k} \\ &= \frac{1}{L-k}\sum_{i=1}^{L-k} (y_i y_{i+k} - y_{i-1}y_{i+k} - y_i y_{i+k-1} + y_{i-1}y_{i+k-1}) \end{aligned} \tag{10}$$

The term $y_i\, y_{i+k-1}$ induces correlations for k =1:

$$\frac{C(1)}{VAR} = -\frac{1}{2} \tag{11}$$

Second-order differences yield [34]

$$\frac{C(1)}{VAR} = -\frac{2}{3} \quad \text{and} \quad \frac{C(2)}{VAR} = +\frac{1}{6} \tag{12}$$

This implies that differentiated contours introduce alternations and may mimic period-doubling. Since we use in this paper first-order differences, we calculate MAC just for k-values above 1 in order to avoid spurious correlations.

MAC for Subharmonics and Chaotic Data

Period-Doubling and -Tripling

As described in the introduction, nonlinear phenomena induce typically correlations in pitch- and amplitude contours. For example, period-doubling (or "octave jumps") yield zig-zag sections in contours. The corresponding correlation functions alternate between -1 and +1. Consequently, we obtain $MAC_{P2} = L$ It can be shown that period 3 implies [34]

$$\frac{C(k)}{VAR} = +1 \quad \text{for} \quad k = 3, 6, 9, \ldots \tag{13}$$

$$\frac{C(k)}{VAR} = -\frac{1}{2} \quad \text{for} \quad k = 1, 2, 4, \ldots \tag{14}$$

This gives $MAC_{P3} = 0.65$.

These examples illustrate that period-doubling and period-tripling correspond to fairly high values of MAC compared to noisy contours. In order to study the variation of MAC for more complex data we apply our concept to simulated time series of the circle map.

Simulations of the Circle Map

It has been shown previously that subharmonics and chaotic vibrations of the vocal folds lead to rough voices [3, 35]. Moreover, analog simulations of a vocal fold model sound clearly rough [36]. Here we study for illustrative purposes a simple discrete model that generates easily subharmonic and chaotic contours. We analyze MAC for subharmonic and chaotic data from the well-known circle map [37, 38].

$$y_{i+1} = y_i + \Omega - \frac{K}{2\pi} sin(2\pi y_i) + R_i \quad (mod\ 1) \tag{15}$$

This is a representative model of two coupled oscillators. The quantity y_i refers to a phase variable. If there is no coupling (K = 0) there is just a constant phase shift Ω. For increasing coupling strength K entrainment regions associated with subharmonics and chaos appear [39-41].

Since voice instabilities are closely related to bifurcations of coupled oscillators, this model is particularly appropriate. In the above equation Ω represents the frequency ratio of the two oscillators and the trigonometric term models a nonlinear coupling. In order to model the omnipresent noise we add independent random numbers (equidistributed in [-0.001, 0.001]). This small amount of noise affects subharmonics and chaos only minimally [42]. The mod I operation guarantees that the variable y_i is confined to the interval [0,1].

Around each rational value of Ω = p : q we find for 0 < K < 1 entrainment regions ("Arnold tongues"), i.e., cycles with a period q. For K > 1 these regions begin to overlap and deterministic chaos appears.

Figure 1 shows sequences $x_i = y_i - y_{i-1}$ for K = 1.1 and increasing Ω. For each value of Ω a random initial condition y_o was chosen and the first hundred iterates have been

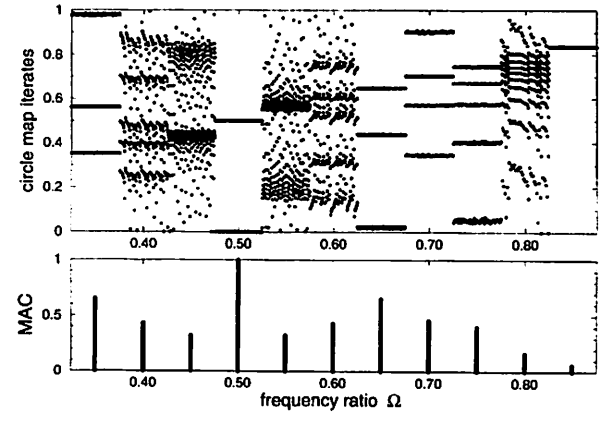


Figure 1. Contours from the circle map with superimposed weak noise (compare equation 15). The frequency ratio Q was taken at the values 0.35, 0.40, ...,0.85. For each value 400 data points are plotted. The lengths of the vertical bars in the lower graph display the value of the corresponding MAC.

discarded to remove transient. The following 400 values are plotted subsequently in Figure 1. This representation can be regarded, therefore, as a concatenation of several contours for slightly different parameters. It can be seen that there are various periodic regimes and chaotic bands in between. However, these chaotic oscillations are not totally random but exhibit some residual periodicities. This is due to weakly unstable cycles within the chaotic attractors.

As predicted, the measure MAC in the lower graph is about 1 for period 2 (at $\Omega = 0.5$) and 0.65 for period 3 (at $\Omega = 0.35$ or 0.65). These periodicities correspond to 1 : 2 and 1 : 3 or 2 : 3 entrainment. 1 : 1 entrainment at $\Omega = 0.85$ leads to almost constant values of the series x_i. Consequently, MAC is close to the value for random noise.

Higher periodicities (3 : 4 and 4 : 5 entrainment) and deterministic chaos lead to MAC-values clearly beyond the noise level of about 0.04. Thus, MAC reflects also correlations which persist in chaotic regimes. There are, however, also strongly chaotic systems with a fast correlation decay [43] where correlation functions cannot distinguish chaos from random noise.

MAC for Contours of Hoarse Voice Samples

Representative Pathological Voices

In this section we present contours, correlation functions, and MAC values for characteristic voice samples from the university hospital Charitè [35]. The examples are selected from 22 extremely hoarse voices with various pathologies. A fully automatic algorithm was applied to detect pitch periods and amplitudes [34]. It starts with a rough estimate of the pitch range via spectral analysis. The precise pitch detection is based on waveform matching of subsequent cycles. Moreover, a range of up to 2 seconds is chosen automatically from which pitch- and amplitude contours are analyzed. In this section we study only the middle 100 cycles of that range.

It can be seen from our figures 2 and 3 that various complex pattern are found reminiscent to simulations of the circle map in the preceding section. The correlation functions extract the dominant periodicities and the resulting MAC-values are clearly beyond the noise level for rough voices.

Figure 2 shows amplitude contours from three patients with papilloma. The first voice is predominantly breathy (H3,B3,R1) and, therefore, the amplitude fluctuations are seemingly random leading to a small value of MAC = 0.06. Contrarily, the other two cases exhibit strong oscillating components. The second voice (H3,B3,R2) is characterized by episodes of biphonation (two audible pitches). During these biphonic episodes the left and right folds are desynchronized [35]. The third voice (H3,B1,R3) exhibits strong low frequency modulations leading to the peak at k = 10 in the correlation function.

The contours in figure 3 are derived from patients with unilateral paralysis. In all three cases pronounced subharmonic components are found leading to oscillating correlation functions and, hence, to large MAC-values.

Excised Larynx Experiment

Figure 4 shows a highly irregular contour from an excised larynx experiment. A dog larynx was dissected and firmly mounted on a tube supplying heated and humified

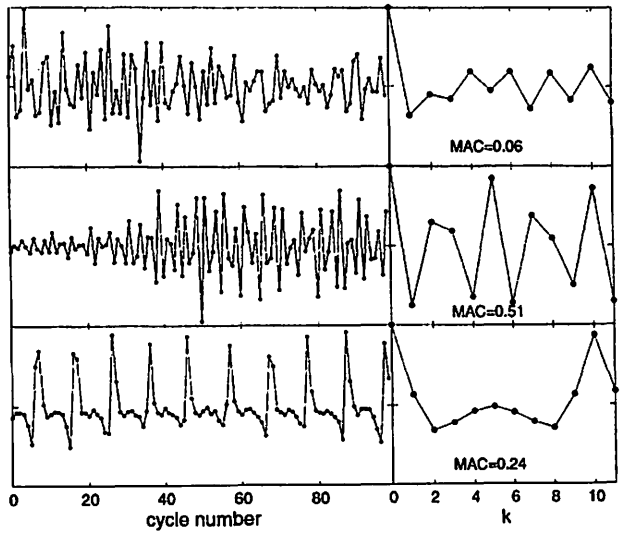


Figure 2. Differentiated amplitude contours from three patients with papilloma. The contours are normalized for visualization and, therefore, no vertical scales are given. The vertical range of the corresponding correlation functions (right) is between -1 and 1.

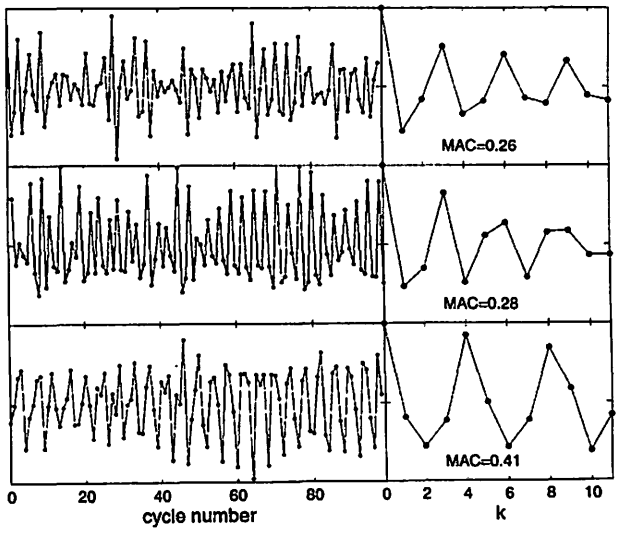


Figure 3. Differentiated amplitude contours from three patients with paralysis with strong subharmonic components.

air (see [44] for details). For large driving pressure a transition to irregular vibrations via period-doubling was observed (compare Figure 8 of [44]).

Beside the strong irregularity there is a subharmonic component leading to a period 3 in the correlation functions. Near the end of the segment a period 6 can be seen. The correlation functions in the lower graphs are derived from the first and second half of the contour. The function with the period 6 (lower right graph) leads to a particular high MAC-value.

These examples illustrate that rough voice signals are associated with complex patterns in pitch- and amplitude contours. The resulting correlations can be quantified with correlation functions and the value of MAC - a novel perturbation measure.

Perceptual Categories

As discussed in the introduction it would be desirable to quantify the more or less subjective characterization of voice qualities. The amount of correlations in contours seems to be good candidate to distinguish roughness and breathiness [3, 29]. Our quantity MAC was indeed small for a breathy voice (upper graph in figure 2) and large for rough voice signals.

In order to test dependencies between perceived voice quality and our discussed measures we use recordings of the vowel "a" of 120 patients from a recent multicenter study [45]. These voices have been evaluated by 30 phoniatricians using the 4 point HBR-scale [46]. In the following we relate the average HBR values to our statistical quantities VAR and MAC. The contours have been derived automatically as described above. The number of pitch periods per voice signal varies between L = 166 and L = 752 (mean 362, standard deviation 106).

First, we discuss hoarseness H as an overall measure of voice quality. As expected, any statistical quantity that measures the amount of amplitude perturbations is correlated with H ratings. For example, shimmer[%] and shimmer[dB] (see [8] for definitions) have correlation coefficients of 0.62. Since $\sqrt{VAR}$ of the amplitude contour is intimately related to shimmer[%] (correlation coefficient: 0.99), it is no surprise that the correlation between $\sqrt{VAR}$ and H-ratings is similarly high (0.64). Even stronger correlations can be found using semilogarithmic plots. Between log (VAR) and H-rating a correlation coefficient of 0.82 was obtained. The corresponding scatter plot is shown in Fig. 5a. Thus the hypothesis is supported that the C(O) = VAR can serve as a measure of hoarseness.

The second hypothesis was that MAC can differentiate between breathiness and roughness. Thus we calculated correlations between MAC from the amplitude contours and the difference of R- and B-ratings. For jitter[%], shimmer[%], shimmer[dB], and $\sqrt{VAR}$ the correlation coefficients to R-B were below 0.1. Thus conventional overall measures of perturbations are not at all able to discriminate between roughness and breathiness. The correlation coefficient between MAC and R-B was 0.32. This value is significantly above the noise level for 120 voices but reflects only weak correlations.

However, the ratings of the various experts exhibit a wide scatter. Consequently, the difference R-B is strongly fluctuating from listener to listener. For severely hoarse voices the ratings were more reliable. Thus we studied also the subset of 30 voices with an average H-rating above 2.25.

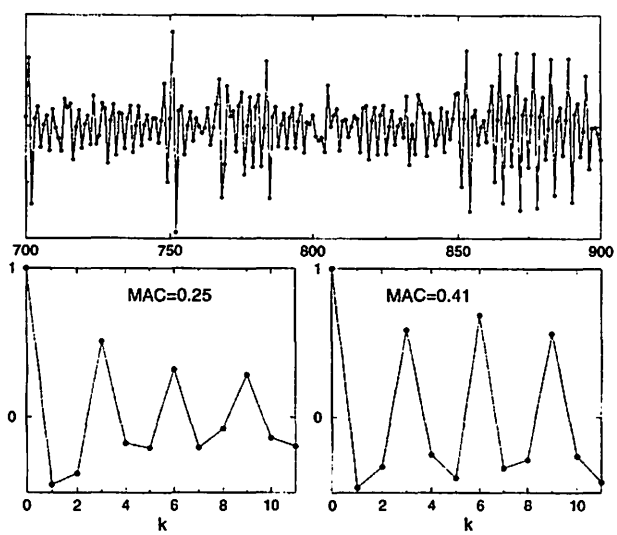


Figure 4. Differentiated pitch contour from an excised larynx experiment. The correlation functions below are calculated from the first (left) and second (right) 100 cycles of the contour above.

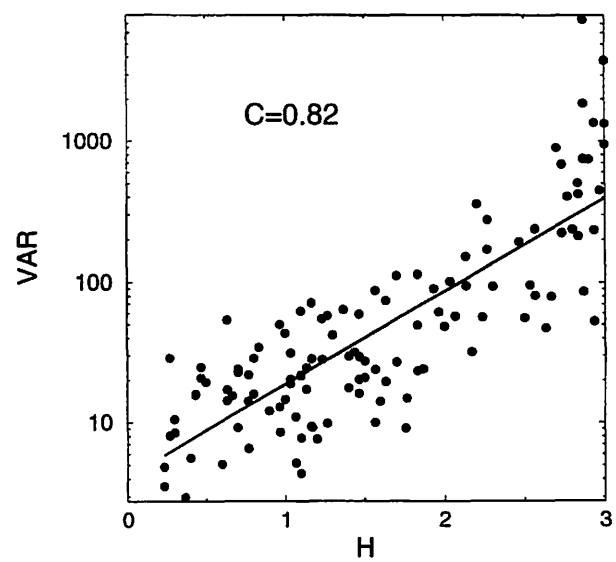


Figure 5. Scatter plot of the variances of 120 amplitude contours versus the mean rating of hoarseness H.

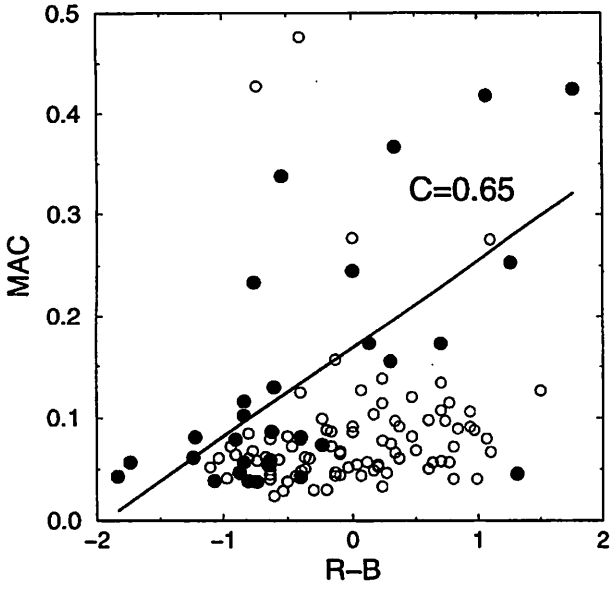


Figure 6. Scatter plot of MAC versus the differences of roughness and breathiness ratings R-B. Filled circles: 30 voices with hoarseness above 2.25; empty circles: other 90 voices.

For these voices we found a clear correlation between MAC and R-B (correlation coefficient: 0.65, see Fig. 6). However, larger studies are necessary to confirm this preliminary result.

Discussion

Our paper was devoted to the statistical analysis of pitch- and amplitude contours. The examples in the preceding sections were derived from acoustic signals. We emphasize that the same techniques are valuable for contours from EGG recordings or high-speed glottography. Recently, in a pioneering study of Mergell pitch-contours of left and right vocal fold vibrations have been analyzed separately [47].

As a first result we found that correlation functions and their absolute mean MAC are sensitive indicators of nonlinear phenomena such as period-doubling, biphonation, or deterministic chaos. In order to evaluate the significance of large MAC-values it was helpful to calculate the expectation value of MAC for random noise. It turned out that MAC remains relatively small for white and correlated noise. This distinguishes MAC from other measures based on correlation functions. Imaizurm suggested, for example, the difference of maximal and minimal correlation coefficients [29] as a measure of roughness. This quantity, however, is a difference of two fluctuating correlation coefficients and exhibits, therefore, large statistical fluctuations. In contrast, MAC is an average over 10 values and can be estimated more reliably. Moreover, its expectation value and variance can be calculated easily [34].

It was shown that a differentiation of contours can induce alternating correlation functions that could be misinterpreted as period-doubling. Even for random sequences first-order differences gives for k = 1 correlation coefficients of -1/2 and second-order differences yield even -2/3. Therefore, MAC is calculated in this paper for the range k = 2, 3,..., 11.

Finally, we looked for relations between perceptual categories and correlation functions. Whereas the variance C(0) quantifies hoarseness, the value of MAC can differentiate between roughness and breathiness. Correlation coefficients above 0.6 have been obtained that are quite high compared to earlier studies [12, 48]. However, the correlations to R-B were not very strict which can be related to the following aspects:

- Any rating has its own uncertainty due to different experiences and subjectivity.
- Roughness and breathiness are related to several properties of the acoustic signal and are hardly measurable by a single quantity.
- Psychoacoustic experiments show that perceived roughness depends on many details of the signal. A signal sounds rough if spectral peaks fall into the same critical band [22, 49, 50]. The distances of spectral peaks, however, depend on the pitch and the specific features of subharmonics and modulations. Contours describe only certain aspects of complex signals and cannot, therefore, correlate perfectly to roughness.

Recently more complicated roughness models from psychoacoustic theory have been tested [34, 51, 52]. They are based on the power-spectra of the voice signal and are, therefore, more closely related to the perception by listeners. A comparison of contour-based measures and psychoacoustically motivated techniques will be published elsewhere.

Acknowledgments

Partial funding for this research was provided by the Deutsche Forschungsgemeinschaft. We thank J. Wendler and M. Hess for providing voice samples and I. R. Titze and D. Berry at the National Center for Voice and Speech in Iowa-City many fruitful discussions. Suggestions of two anonymous referees are gratefully appreciated.

References

[1] W. Mende, H. Herzel, K. Wernike: Bifurcations and chaos in newborn cries. Physics Letters A 145 (1990) 418-424.

[2] I. R. Titze, R. Baken, H. Herzel: Evidence of Chaos in Vocal Fold Vibration. Vocal Fold Physiology: Frontiers in Basic Science (Ed. I. R. Titze). Singular Publishing Group, San Diego, 1993, pp. 143-188.

[3] H. Herzel, D. A. Berry, I. R. Titze, M. Saleh: Analysis of vocal disorders with methods from nonlinear dynamics. J. Speech and Hearing Res. 37 (1994) 1008-1019.

[4] I. Steinecke, H. Herzel: Bifurcations in an asymmetric vocal fold model. J. Acoust. Soc. Am. 97 (1995) 1874-1884.

[5] M. Tigges et al.: Observation and modelling of glottal biphonation. Acoustica 83 (1997) 707-714.

[6] P. Lieberman: Perturbations in Vocal Pitch. J. Acoust. Soc. Am. 33 (1961) 597-603.

[7] P. Lieberman: Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathological Larynges. J. Acoust. Soc. Am. 35 (1963) 344-353.

[8] R. J. Baken: Clinical Measurement of Speech and Voice. Allyn and Bacon, 1987.

[9] P. Milenkovich: Least mean square measures of voice perturbation. J. Speech and Hearing Res. 30 (1987) 529-538.

[10] N. B. Pinto, I. R. Titze: Unification of perturbation measures in speech signals. J. Acoust. Soc. Am. 87 (1990) 1278-1289.

[11] I. R. Titze: Principles of Voice Production. Prentice Hall, 1994.

[12] M. Hirano et al.: Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia. J. of Phonetics 14 (1986) 455-456.

[13] B. Hammarberg, B. Fritzell: Acoustic and perceptual analysis of vocal dysfunction. J. of Phonetics 14 (1986) 533-547.

[14] C. R. Rabinov et al.: Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. J. Speech and Hearing Res. 38 (1995) 26-32.

[15] D. Wong (Ed.): Proceed. Workshop on Acoustic Voice Analysis. Denver, 1995.

[16] W. Hess: Pitch Determination of Speech Signals. Springer-Verlag, Berlin, 1983.

[17] I. R. Titze: Comparison of Fo Extraction Methods for High-Precision Voice Perturbation Measurements. J. Speech Hear. Res. 36 (1993) 1120-1133.

[18] S. Bielamowicz et al.: Comparison of voice analysis systems for perturbation measurement. J. Speech and Hearing Res. 39 (1996) 126-134.

[19] H. v. Helmholtz: Die Lehre von den Tonempfindungen. Vieweg und Sohn, Braunschweig, 1862.

[20] R. W. Wendahl: Some parameters of auditory roughness. Folia phoniatr. 18 (1966) 26-32.

[21] E. Terhardt: On the Perception of Periodic Sound Fluctuations (Roughness). Acustica 30 (1974) 201-213.

[22] E. Zwicker, H. Fastl: Psychoacoustics. Springer, Berlin, 1990.

[23] R. F. Coleman: Effect of median frequency levels upon the roughness of jittered stimuli. J. Speech and Hearing Res. 12 (1969) 330-336.

[24] R. A. Newman, F. W. Emanuel: Pitch effects on vowel roughness and spectral noise for subjects in four musical voice classifications. J. Speech and Hearing Res. 34 (1991) 753-760.

[25] K. Omori, H. Kojima, R. Kakani, D. H. Slavit, S. M. Blaugrund: Acoustic characteristics of rough voice: subharmonics. J. of Voice 11 (1997) 40-47.

[26] J. Wendler, W. Seidner, G. Kittel, U. Eysholdt: Lehrbuch der Phoniatrie und Paedaudiologie. Thieme-Verlag, Stuttgart 1996.

[27] G. de Krom: Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. J. Speech and Hearing Res. 38 (1995) 794-811.

[28] Y. Koike: Application of some acoustic measures for the evaluation of laryngeal dysfunction. Studia Phonologica VII (1973) 17-23.

[29] S. Imaizumi: Acoustic measures of roughness in pathological voice. J. of Phonetics 14 (1986) 457-462.

[30] H. Kasuya, Y. Endo: Acoustic analysis, conversion, synthesis of the pathological voice. In: Vocal Fold Physiology: Voice Quality Control (Eds. O. Fujimura, M. Hirano), Singular Publ. Group, San Diego, 1995, pp. 305-319.

[31] H. Herzel. Possible Mechanisms of Vocal Instabilities. In: Vocal Fold Physiology: Controlling Complexity & Chaos (Eds. N. H. Fletcher, P. Davis) Singular Publ. Group, San Diego, 1996, pp. 63-75.

[32] R. A. Fisher: Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrica 10 (1915) 507-521.

[33] O. Weiss, H. Herzel: Correlations in protein sequences and property codes. J. Theor. Biol. 190 (1998) 341-353.

[34] R. Reuter: Untersuchung der Rauhigkeit der menschlichen Stimme auf der Grundlage der Nichtlinearen Dynamik und Psychoakustik. PhD Thesis, Techn. University Berlin, 1999.

[35] H. Herzel, J. Wendler: Evidence of chaos in phonatory samples. Proceedings EUROSPEECH, Genova, ESCA 1991, pp. 263-266.

[36] R. Reuter, R. Orglmeister, H. Herzel: Simulations of vocal fold vibrations with an analog circuit. Int. J. Bifurcation and Chaos 9 (1999) 1075-1088.

[37] H. G. Schuster: Deterministisches Chaos. VCH, Weinheim, 1994.

[38] R. W. Leven, B.-P. Koch, B. Pompe: Chaos in dissipativen Systemen. Akademie Verlag, Berlin, 1993.

[39] P. Berge, Y. Pomeau, C. Vidal: Order within chaos. Wiley, New York, 1986.

[40] L. Glass, M. Mackey: From clocks to chaos. Princeton University Press, 1988.

[41] D. Kaplan, L. Glass: Understanding nonlinear dynamics. Springer, Berlin, 1995.

[42] H. Herzel, B. Pompe: Effects of noise on a nonuniform chaotic map. Physics Lett. A 122 (1987) 121-125.

[43] H. Herzel, W. Ebeling: The decay of correlations in chaotic maps. Physics Lett. A 111 (1985) 1-4.

[44] D. A. Berry, H. Herzel, I. R. Titze, B. Story: Bifurcations in excised larynx experiments. J. of Voice 10 (1996) 129-138.

[45] M. Hess, M. Ulrich, M. Gross: Multicenter Studie zur subjektiven Beurteilung von Stimmstorungen. In: Aktuelle phoniatrisch -padaudiologische Aspekte (Ed. M. Gross) Renate Gross Verlag, Berlin, 1996.

[46] T. Nawka, L. C. Anders, J. Wendler: Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. Sprache Stimme Gehoer 18 (1994) 130-133.

[47] P. Mergell: Nonlinear dynamics of phonation - high-speed glottography and biomechanical modeling of vocal fold oscillations. PhD Thesis, Techn. University Berlin, 1998.

[48] N. Arends, D.-J. Povel, E. van Os. L. Speth: Predicting voice quality of deaf speakers on the basis of glottal characteristics. J. Speech and Hearing Res. 33 (1990) 116-122.

[49] J. G. Roederer: The Physics and Psychophysics of Music. Springer, New York, 1995. 20

[50] W. Aures: Ein Berechnungsverfahren der Rauhigkeit. Acustica 58 (1985) 268-281.

[511 D. Michaelis, T. Gramss, H. W. Strube: Glottal-to-noise excitation ratio - a new measure for describing pathological voices. Acustica 83 (1997) 700-706.

[52] D. Michaelis, M. Frohlich, H. W. Strube: Selection and combination of acoustic features for the description of pathological voices. J. Acoust. Soc. Am. 103 (1998) 1628-1639.

NCVS Status and Progress Report - 14
September 1999, 55-61

Perception of Pitch and Roughness in Vocal Signals with Subharmonics

Christine Bergan, M.M.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Pitch and roughness were rated as the extent of amplitude modulation (AM) and frequency modulation (FM) of a subharmonic ($F_o/2$) were varied. The objective was to determine the identification boundaries for pitch and roughness and to discover how both kinds of modulation affect these boundaries. Another objective was to judge the reliability between subjects when identifying pitch and roughness. Three procedures were used: ABX comparisons, Method of Adjustment, and Rating of Roughness. Results indicated that the crossover point to the lower pitch (associated with the subharmonic) occurred between 10% and 30% modulation, depending on modulation type and F_o. Subjects demonstrated highly variable perceptions of pitch and roughness with poor inter-subject reliability.

Introduction and Background

The ability to evaluate the perceptual qualities of the human voice is of practical interest to voice clinicians and singing teachers. If diagnosis and therapy are to improve, it is crucial that perceptual ratings of broad categories such as pitch, loudness, roughness, resonance, and breathiness can be broken down into finer subcategories. Ideally, the subcategories would reflect specific vocal fold oscillation properties. For example, roughness resulting from a subharmonic may be distinguishable from roughness resulting from aperiodicity. Aside from the existing inadequacy of current terminology for clinical judgment of the voice (Jensen, 1965), there is a lack of parametric studies of specific waveform characteristics and their effect on vocal qualities.

Recent research by Kreiman et al (1998) and Rabinov et al (1995) has demonstrated poor inter-subject reliability and validity when rating such qualities as roughness and breathiness. "Listeners agreed very poorly in the midrange of scales for breathiness and roughness, and mean ratings in the midrange of such scales did not represent the extent to which a voice possesses a quality, but served only to indicate that the listeners disagreed" (1998). De Bodt et al (1997) studied the effect of experience and professional background on perceptual rating of voice quality using the GRBAS scale. They presented 9 pathological voices to speech language pathologists and otolaryngologists with a 14-day test-retest interval. The test-retest reliability was moderate, with best agreement occurring for the G (grade) parameter and the worst for the S (strained) parameter. They concluded that professional background had a greater impact on perceptual rating than experience. Gerratt et al (1993) compared internal and external standards in voice quality judgments through the use of standard rating scales versus scales in which a set of anchors was presented prior to rating. Their results showed that the use of an anchored scale (when compared to an unanchored scale) significantly increases both inter and intra-subject reliability. Other research (Ebel, 1951) has also shown poor reliability of ratings.

A wealth of literature exists concerning the general perception of pitch and roughness. Pitch is affected by the intensity, duration, and spectrum of a voiced sound. When the intensity increases, the pitch of a low tone decreases, whereas the pitch of a high tone increases (Stevens, 1935). Beerands (1989) found that the pitch of short tones is less salient than the pitch of long tones. According to the ANSI standard of 1994, "Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends mainly on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus."

The sensation of two interfering simple tones (sinusoids) is characterized as roughness in speech and tonal dissonance in music. Maximal roughness and maximal dissonance are sometimes considered synonymous. The degree of interference among the harmonics of two sinusoids determines the amount of dissonance or consonance perceived for simple frequency ratios of the fundamentals. Perception of roughness may also result from either amplitude modulation or frequency modulation of the tone. Terhardt (1974) found that the roughness of a sinusoidally amplitude-modulated tone depends primarily on the relative amplitude modulation (relative fluctuation of the temporal envelope).

Roughness seems to be strongly correlated with the envelope fluctuations of a sound that has passed through a critical band filter. A parameter would be the ratio of fluctuation amplitude to a steady state mean value. "Then, the entire roughness is composed of the partial roughnesses which are contributed by adjacent critical bands." When subharmonics exist in the source waveform, due to a variety of possible pathologies or unusual vibration patterns, roughness is likely to be perceived because the fundamental and a subharmonic are in a critical band. Furthermore, the pitch may become uncertain.

Wendahl (1966a) showed that for a given jitter (usually thought of as a random frequency modulation), a signal with a low fundamental frequency (F_0) tended to be perceived as rougher than a signal with a high F_0. In a 50 Hz. Tone for example, even a sinusoidal (rather than random) modulation may be enough to create the perception of roughness and a lowered pitch. This will be borne out in our present study.

Purpose and Research Questions

The purpose of this study was to determine what perceptual judgments occur as a direct result of an $F_0/2$ subharmonic modulation, both in frequency (FM) and in amplitude (AM). We wished to determine an identification boundary between the F_0 related pitch and the $F_0/2$ related pitch and to investigate whether this perceptual boundary is discrete (categorical) or continuous. Subjects were presented with tokens of the a synthetic [a] vowel. The research questions to be answered were: (1) Where does the identification boundary occur as a function of modulation extent? (2) Is it the same for FM and AM? (3) How does the modulation affect the perception of roughness? (4) Does F_0 have an effect on the perception of pitch and roughness? (5) How variable is the subjects' ability to identify pitch and roughness?

Table 1.
Parameters for Simulation Model

Parameter	*Default*	*Range*	*Perceptual Effect of Changes*
Flow Amplitude	375 cm^3/sec	200-700	Loudness
Fundamental Frequency	100 Hz	50-1200	Pitch
Open Quotient – Q_o	.6	.1-1.0	Tightness/Breathiness
Skewing Quotient - Q_s	1.7	1-5	Timbre
DC Flow Added	0 cm^3/sec.	0-300	Breathiness
Area of Epilarynx Tube	.5 cm^2	0-1	Resonance/Ring
F_0 Modulation Frequency	0 Hz	0-12	Vibrato (Rate)
F_0 Modulation Extent	0 %	0-12	Vibrato (Amount)
Amplitude Modulation Frequency	0 Hz	0-12	Vibrato (Rate)
Amplitude Modulation Extent	0 %	0-12	Vibrato (Amount)
Nasal Coupling Area	0 cm^2	0-1	Nasality

Procedures

Ten subjects were recruited. All of them successfully passed a hearing test and were considered to have normal hearing. The ages of the subjects ranged from 22-57, with a mean age of 33 years. Amount of musical background or training ranged from none (1 subject) to moderate (5 subjects) to professional vocalists (4 subjects).

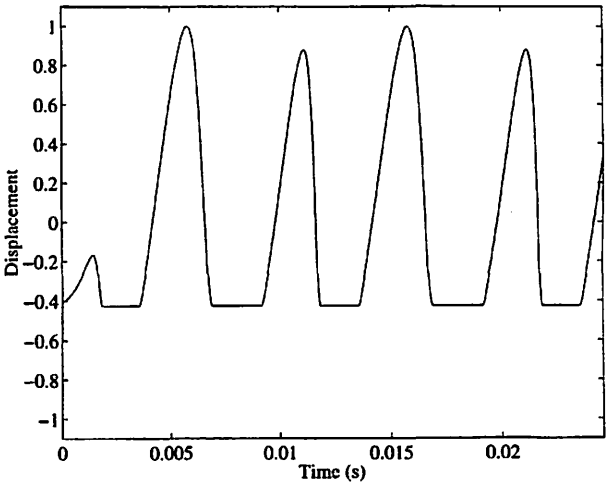


Figure 1. Glottal flow waveform with 20% amplitude modulation with F_0=200 Hz.

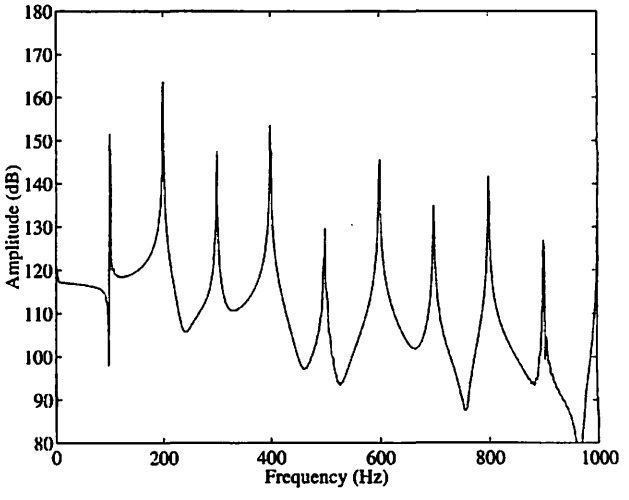


Figure 2. Magnitude spectrum of glottal flow with 20% frequency modulation with F_0=200 Hz.

Synthetic stimuli were generated with a computer model Speak, a computational glottal flow and vowel articulation model based on the linear source-filter theory (Titze, Mapes, and Story, 1994). In this model, both the glottal source and the filter parameters can be manipulated and controlled. The glottal parameters are: peak flow fundamental frequency, open quotient, skewing quotient, the size of the area of the epilaryngeal tube, and frequency and extent of AM and FM (imposed on the first two parameters). The filter defines the area function of the vocal tract. Table 1 represent all available inputs, with typical default values and ranges.

All default values were used, with the exception of fundamental frequency (which was chosen to be 100, 200, or 300 Hz.), F_o modulation frequency (which was always chosen to be half of the F_o, e.g, 50, 100, or 150 Hz.), and either the amount of amplitude modulation (%) or frequency modulation (%), which was varied from 2% to 98%. Figure 1 shows a simulated glottal waveform with 20% amplitude modulation. Note the alternation of successive amplitudes. Frequency modulation of the same extent (20%) is difficult to see on the time waveform, so we show a magnitude spectrum in Figure 2. Note here that the subharmonics are about 10-15 dB below the harmonics in two co-existing series.

For the Method of Adjustment task to match pitch, triangular waves (rather than pure tones) were generated by a function generator (Heath IG-1271, Benton Harbor, MI) to present the subject a complex, but modulated) spectrum. Ideally, the same SPEAK synthesizer would have been used, but a real-time version with simple dial control of F_o was not yet available. The intensity of the signal was kept at a comfortable dB SPL level for each subject averaging about 60 dB SPL. Subjects were allowed to adjust the frequency until they found what they believed to be a "match" to the presented token. Frequencies chosen by the subjects during the Method of Adjustment task were measured and recorded by the investigator (not in view of the subjects) using a frequency counter with a range of 5 Hz-80 MHz (B&K 1805, Chicago, IL).

Three different listening tasks were requested of the subjects. All tasks involved the presentation of tokens of 1-second duration of the vowel [a] presented at 100, 200, or 300 Hz. All three fundamental frequencies were presented either in a non-modulated form (1st procedure only) or were systematically modulated with their corresponding subharmonic ($F_o/2$) by varying the extent of AM or FM.

The first task involved presentation of the above tokens in an ABX forced alternative format in which the subjects first heard the *non-modulated* fundamental frequency F_o as A; then heard the non-modulated subharmonic

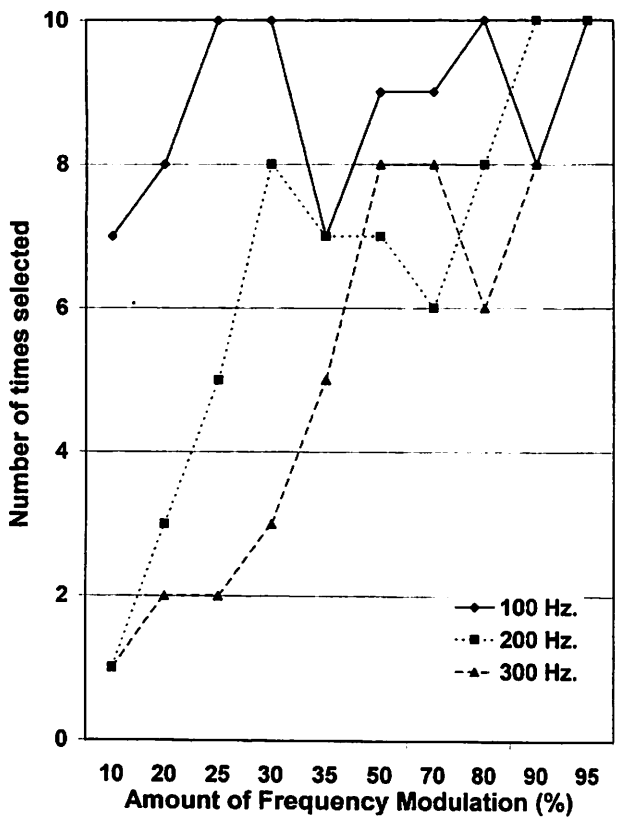


Figure 3. ABX comparisons for FM.

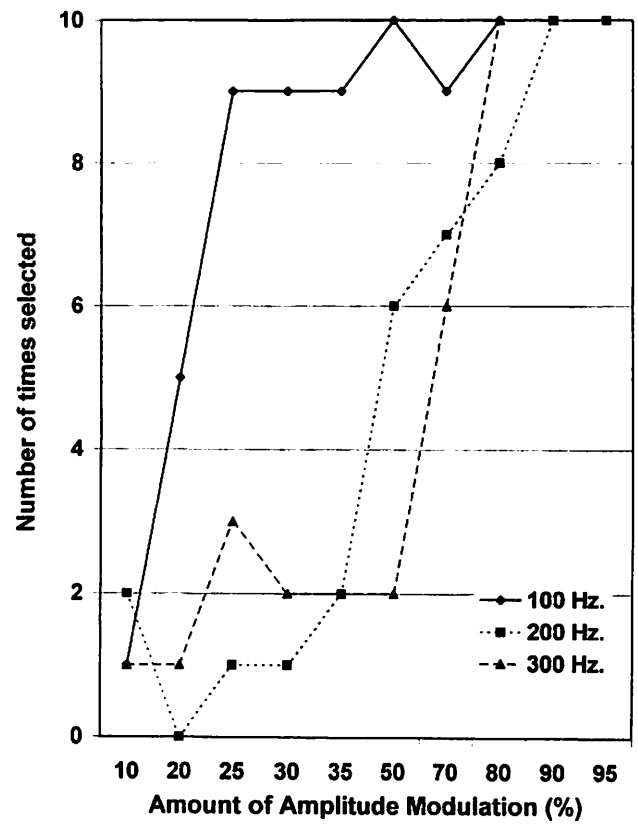


Figure 4. ABX comparisons for AM.

$F_o/2$ as *B*; and finally heard the token with the modulation *X*. They were asked to listen to *A* and *B* and determine if the pitch of *X* was more like *A* or *B*. There was a 2.5-second pause between *A* and *B* and between *B* and *X*, followed by a 5-second pause for response. All possible tokens were presented three times to increase the power and reliability of our results and to check for intra-subject reliability.

The second task involved the presentation of single tokens of the modulated tones, with the option to repeat the stimulus as many times as desired or to continue. The subject was allowed to adjust the frequency dial of the wave generator until they believed they had found a pitch to "match" the given token. When this match had been found, their choice was displayed on a frequency counter and was recorded by the investigator. All tokens were presented twice to check for intra-subject reliability.

The third task involved the presentation of the above stimuli again as single tokens. Subjects were instructed to rate each in terms of roughness on a scale of 1 (very smooth) to 10 (very rough). The subjects were given a set of anchors prior to this task to provide them with a frame of reference for the subsequent ratings. These anchors represented the entire range of all possible amplitude and frequency modulated tones.

Results and Discussion

Boundary identification was influenced by both F_o and type of modulation. For frequency modulated vowels (Figure 3), the identification boundary (5 out of 10 selections) appeared to occur with less than 10% modulation for 100 Hz, 25% for 200 Hz, and 35% for 300 Hz. For amplitude modulated tokens (Figure 4), the identification boundary appeared to occur at 20% for 100 Hz, and approximately 50% for both 200 and 300 Hz.

In the Method of Adjustment task (Figure 5 and 6), there was a high preference for choosing the subharmonic pitch as the true pitch, but intermediate pitches between the two octaves were systematically chosen. Thus, there was not a pitch dichotomy, but a gradual variation. This would suggest a continuous rather than discrete pitch perception of these modulated signals. The crossover to the subharmonic for FM was at 10% modulation for 100 Hz and at 20% for 300 Hz. For AM, the crossover occurred at 20% modulation for all fundamental frequencies.

From the above results, it appears that FM yields identification of the subharmonic *at a lower modulation extent* than does AM and that lower F_o produces *earlier* identification of the subharmonic as the true pitch.

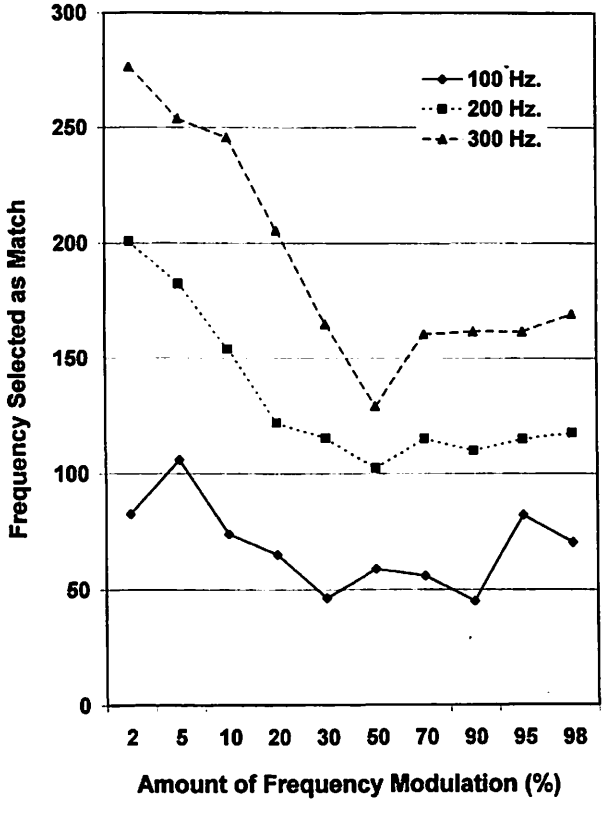


Figure 5. Method for adjustment comparisons for FM.

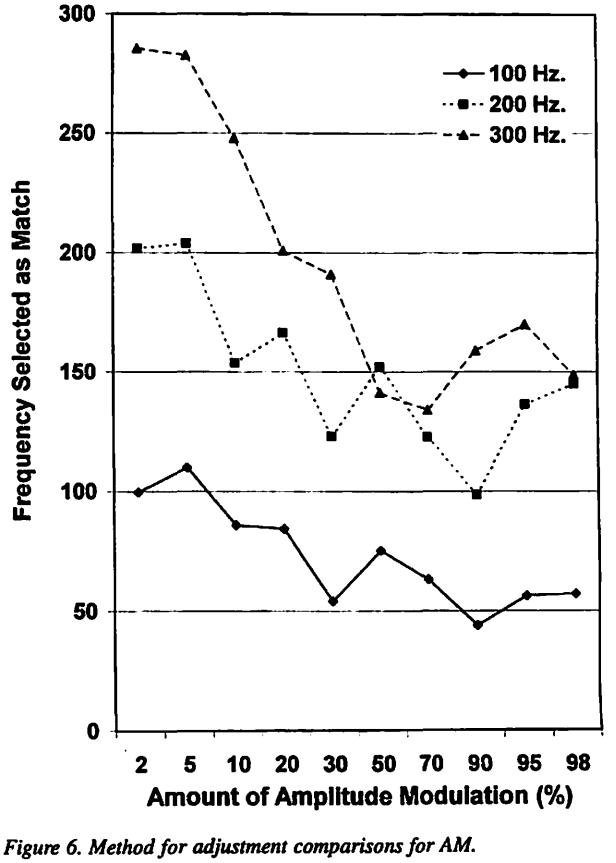


Figure 6. Method for adjustment comparisons for AM.

Perception of roughness was also influenced by F_o and type of modulation (Figures 7 and 8). FM tones received ratings of roughness of 5 or greater (out of 10%) for 10% modulation while AM tones didn't receive roughness ratings of 5 or greater until 20% modulation was reached. Thus, it appears that FM tones cause a somewhat higher rating of roughness (generally) than do AM tones, but statistical significance was not reached in our study. The lowest F_o (100 Hz.) was perceived as rougher than the other two F_os (200 and 300 Hz.) at *all* modulation amounts and for *both* AM and FM. This supports the findings by Wendahl (1966a).

With regard to the question of subject reliability, at least half of the subjects (those with little or no musical training) showed poor pitch matching abilities and their "matches" appeared to be random and inconsistent. This was demonstrated by their selection of neither the F_o, the $F_o/2$, nor any other harmonic component. However, it is interesting to note that even the musicians showed great inter-subject variability when choosing between the F_o and its subharmonic (Table 2; following page).

For perception of pitch in both AM and FM, the subject variance generally increased with increasing F_o. For FM, the variance was particularly great for 100 Hz at the 95% and 98% modulations. There was a slight overall decrease in the variance at the 90% level for *all three* F_os and for *both* AM and FM. It appears that there is poor inter-subject reliability in the perception of F_o in even the most "trained" ears and even less reliability in the "untrained" ears.

In perception of roughness, the lowest F_o (100 Hz) generally had greater variance in the ratings than the other two F_os (Table 3; following page). This was especially true for AM. Variance in roughness ratings for 200 and 300 Hz seemed to dramatically increase at about 10% and then tended to remain fairly high up to 98%. These findings support those made by Kreiman et al (1998) and Gerratt et al (1993). Inter-subject variability was high for all three procedures, for both AM and FM, and for both pitch identification and roughness rating.

Statistical analysis for the ABX task showed the comparison of selected F_o with the presented F_o yielded a chi-square value of 37.63 ($p \leq .001$). The comparison of selected F_o by type of modulation for AM yielded a chi-square value of 103.50 ($p \leq .001$). The comparison of selected F_o with the type of modulation for FM yielded a chi-square of 54.906 ($p \leq .001$). This would imply a very strong relationship between selected F_o and presented F_o, as well as selected F_o and type of modulation.

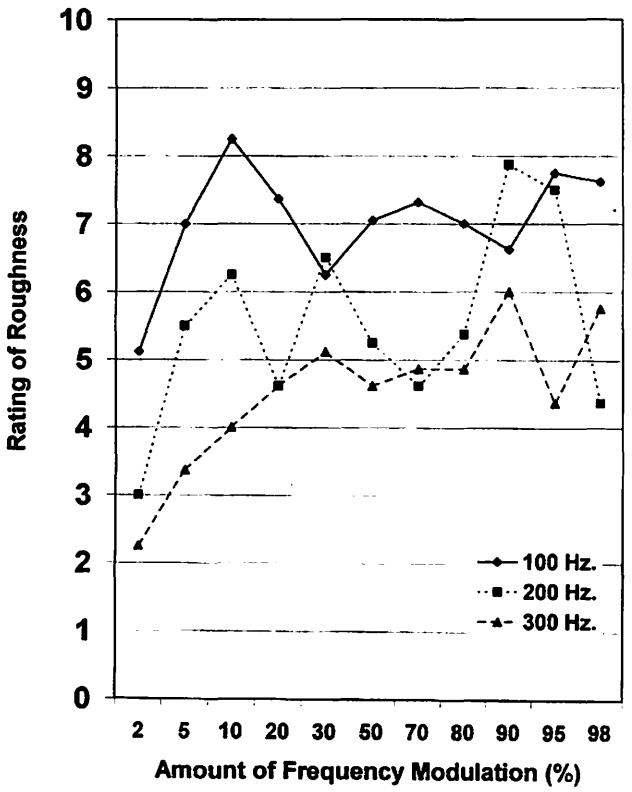


Figure 7. Roughness rating for FM.

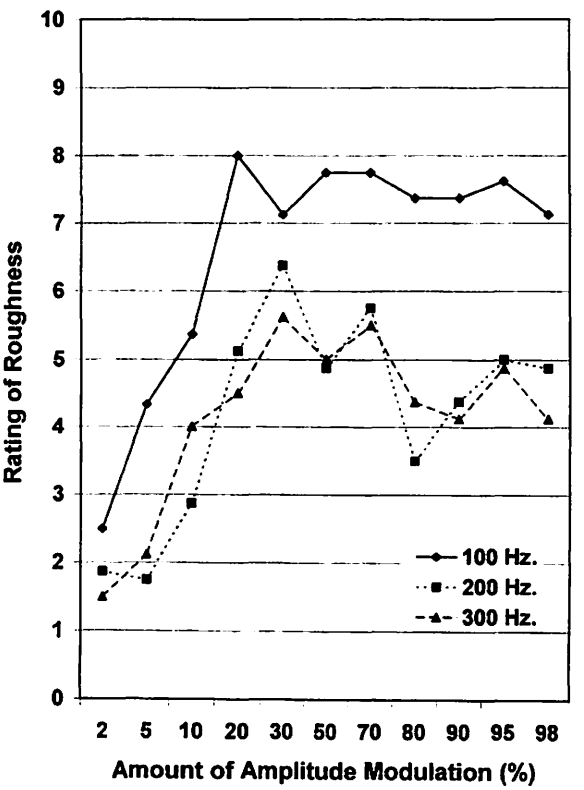


Figure 8. Roughness rating for AM.

Table 2.
Variabilities in the Method of Adjustment Task
(Note Overall Decrease in Variability at 90%)

Amount of Freqeuncy Modulation

		2%	5%	10%	20%	30%	50%	70%	90%	95%	98%
100 Hz.	mean	82.75	106	74	65.22	46.33	58.88	55.88	45	82.25	70.37
	stdev.	24.3	46.77	31.26	25.01	25.02	32.11	33.69	**12.83**	55.18	57.73
200 Hz.	mean	200.75	182.25	153.88	121.88	115.22	102.33	115	109.88	114.87	117.57
	stdev.	4.02	37.36	56.44	43.81	51.12	45.41	32.45	**34.32**	38.24	37.38
300 Hz.	mean	276.62	254	245.55	205.44	164.77	129	160.33	161.66	161.62	169.12
	stdev.	41.56	58.01	81.74	75.21	52.43	35.97	33.32	**28.83**	59.91	57.7

Amount of Amplitude Modulation

		2%	5%	10%	20%	30%	50%	70%	90%	95%	98%
100 Hz.	mean	99.6	109.9	85.9	84.33	54.11	75.33	63.22	43.88	56.37	57.12
	stdev.	1.64	40.73	31.78	23.62	8.66	54.25	24.17	**11.35**	17.96	20.71
200 Hz.	mean	201.8	204	153.7	166.44	123	152.22	122.55	98.55	136.25	145
	stdev.	31.97	12.07	64.12	55.24	43.66	58.25	44.36	**10.08**	49.09	49.66
300 Hz.	mean	285.6	282.7	247.8	200.88	191	141.22	133.88	158.88	169.87	148.75
	stdev.	29.41	33.01	65.61	89.08	94.47	32.62	23.71	**28.62**	55.27	20.67

Analysis of variance for comparison of selected F_o versus F_o for the Method of Adjustment Task showed an F-value of 12.38 ($p \leq .0001$) for FM tokens and an F-value of 14.35 ($p \leq .0001$) for AM tokens. However, caution must be taken in interpretation of the relationship between selected F_o and presented F_o since the selected F_o values were mean values and contained great variability. Comparison of selected F_o versus type of modulation gave an F-value of 2.21 ($p \leq .0001$) for FM and an F-value of 2.92 ($p \leq .0001$) for AM.

Statistical analysis of ratings of roughness yielded an F-value of 3.81 ($p \leq .001$) for comparison of the selected versus presented F_o in FM tokens, while an F-value of 6.45 ($p \leq .0001$) occurred for the comparison of selected versus presented F_o in AM tokens. Comparisons *between* type of modulation and selected rating, however, were not as significant, yielding an F-value of 1.29 ($p \leq .1889$) for FM tokens and an F-value of .67 ($p \leq .8548$) for AM tokens. This would imply a strong relationship between presented F_o and rating of roughness, but a weak relationship between type of modulation and rating of roughness.

Conclusions and Implications for Future Research

This study has brought into question the perception of pitch and roughness for vowel sounds containing subharmonics. Our findings suggest that there may not be a single "fundamental pitch" when listening to such sounds. A modulation of as little as 10% can cause the pitch to drop toward a lower octave associated with the subharmonic. But inter-subject variability is high. This may be explainable on the basis of the "best-fit" or (template-fitting) model proposed by Goldstein et al (1973). According to this model, subjects pick the best-fitting F_o that corresponds to a highly individualized stereotype.

Further research might involve the investigation of the perception of lower order subharmonics (F_o/n, where n=1, 2, 3, ...) or modulations that have no integer relation to F_o. One possible goal might be the development of a set of training tapes that contain a great variety of anchors for modulations typically observed in voice signals. These tapes could be used by clinicians, otolaryngologists, and vocal pedagogues who wish to find a more objective, accurate, and standardized means of perceiving vocal qualities.

Table 3.
Variabilities in the Perception of Roughness
(Numbers of Greatest Variance are Boldfaced)

Amount of Freqeuncy Modulation

		2%	5%	10%	20%	30%	50%	70%	80%	90%	95%	98%
100 Hz.	mean	5.12	7	**8.25**	7.37	6.25	7.5	7.32	7	6.62	7.75	7.62
	stdev.	1.88	1.41	1.49	1.5	2.96	2.5	2.55	2.93	1.59	**3.01**	2.72
200 Hz.	mean	3	5.5	6.25	4.62	6.5	5.25	4.62	5.37	**7.87**	7.5	4.37
	stdev.	1.19	1.77	1.75	1.84	1.19	1.75	2.19	2.06	1.45	1.3	**2.87**
300 Hz.	mean	2.25	3.37	4	4.62	5.12	4.62	4.87	4.87	6	4.37	5.75
	stdev.	0.7	1.68	2.5	2.61	2.41	2.77	2.47	2.9	**3.07**	2.61	2.86

Amount of Amplitude Modulation

		2%	5%	10%	20%	30%	50%	70%	80%	90%	95%	98%
100 Hz.	mean	2.5	4.33	5.37	8	7.12	7.75	7.75	7.37	7.37	7.62	7.12
	stdev.	2.13	2.32	3.33	1.6	1.95	2.65	2.54	**3.37**	2.66	3.02	2.47
200 Hz.	mean	1.87	1.75	2.87	5.12	**6.37**	4.87	5.75	3.5	4.37	5	4.87
	stdev.	0.83	0.7	1.45	**2.23**	2.19	1.8	2.12	1.69	1.06	2.13	1.88
300 Hz.	mean	1.5	2.12	4	4.5	**5.62**	5	5.5	4.37	4.12	4.87	4.12
	stdev.	0.75	1.12	2	2.44	**2.62**	2	1.3	2.26	2.1	2.53	2.59

Acknowledgments

We would like to express our sincere appreciation to Eric Hunter, Roger Chan, and Xuyang Zhang for their invaluable and most generous assistance in the preparation of this study. Funding for this study was from Grant #: P60 DC00976 from the National Institutes of Health.

References

ANSI (ASA-111-1994). *American National Standard Acoustical Terminology*. New York, NY: American National Standards Institute; 1994.

Beerends JG. The influence of duration of pitch in single and simultaneous complex tones. *J Acoust Soc Amer.* 1989;86:1835-1844.

DeBodt MS, Wuyts FL, Vande Heyning PH, Croux C. Test-retest study of the GRBAS scale: Influence of experience and professional background on perceptual rating of voice quality. *J Voice* 1997;11:74-80.

Ebel R. Estimation of the reliability of ratings. *Psychometrika* 1951;16:407-424.

Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgements. *J Speech Hear Res.* 1993;36:14-20.

Jensen PJ. Adequacy of terminology for clinical judgment of voice quality deviation. *Eye Ear Nose Throat Month.* 1965;44:77-82.

Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J Acoust Soc Amer.* 1998;104:1598-1608.

Rabinov CR, Kreiman J, Gerratt BR, Bielamowicz S. Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *J Speech Hear Res.* 1995;38:26-32.

Stevens SS. The relation of pitch to intensity. *J Acoust Soc Amer.* 1935;6:150-154.

Terhardt E. Pitch, consonance, and harmony. *J Acoust Soc Amer.* 1974;55:1061-1069.

Titze I, Mapes S, Story B. Acoustics of the tenor high voice. *J Acoust Soc Amer.* 1994;95:1133-1142.

Wendahl R. Some parameters of auditory roughness. *Folia Phoniatr.* 1966a;15:241-250.

NCVS Status and Progress Report - 14
September 1999, 63-72

Cellular Physiology of the Vocal Folds

Steven Gray, M.D.
Department of Surgery, Division of Otolaryngology–Head and Neck Surgery, The University of Utah School of Medicine

Introduction

Cells constitute the core biologic processes by which our vocal folds are alive. Although it is rare for any voice professional to be cognizant that vocal health or illness is often times dependent upon cellular well being or dysfunction, the laryngologist must be aware of this relationship. In essence, nearly every voice illness can either be related to a tissue change or disorder, or sub-optimal or inappropriate functional use of normal or abnormal tissue. As physicians we must decide if the tissue is diseased or if it is being used sub-optimally or both. If the tissue is diseased, then this implies it's cellular health is lacking and we should try to restore that. It is to this matter that this chapter will address. Although this chapter attempts to summarize the important information regarding cellular activity, vocal health, and voice disorders, it will become apparent that much knowledge is lacking.

The field of vocal health and vocal function has had a decidedly different past than many other organ systems. This is in part due to the large volume of research performed in this field by speech scientists, speech pathologists, and those with engineering backgrounds. As these investigators have pioneered research in this field, they have used the tools with which they are most comfortable and experienced. Consequently, we have a rich foundation in signal processing and acoustical analysis of voice and vocal problems. We are also developing a nice foundation in other areas of science such as computer modeling and integrated research such as that found in the neuromotor arena. However, voice research at a biologic cellular or molecular level has not been done with the same degree of interest. It is the author's hope that this chapter will help provide some of the building blocks for clinicians to understand vocal health and vocal pathology from a biologic and tissue basis and also to perhaps stimulate ideas and research in this area.

The vocal folds are composed of various tissue types: epithelium, lamina propria, striated muscle, nerves, vascular structures, and cartilage. Although each of these tissue types are important, this chapter will focus on the epithelium and lamina propria. Vocal pathologies that are related to muscle dysfunction are integrated into the neuromuscular system and often have a neurologic basis for their disorder. It is beyond the scope of this chapter to deal with those diseases. On the other hand, most vocal lesions are pathologies that are lamina propria in origin. The lamina propria is certainly an active and vibrant structure which is deserving of our attention for the next few pages.

The Epidermis

The vocal folds have a stratified squamous epithelium for their covering. Other parts of the larynx have a ciliated pseudocolumnar epithelium. The majority of the ciliated epithelium is found in the posterior glottis and there is also occasionally a small strip of ciliated epithelium right at the anterior comisure. However, the membranous vocal folds are covered with squamous epithelium. As opposed to normal skin, the epithelial cells appear to be metabolically functional until near the time of desquamation. Squamous cell epithelium is replenished by the top surface (old) cells being desquamated into the lumen and newer cells (being produced by division of the basal cells) moving superiorly or medially and taking their place. These metabolically active cells have caused some researchers to wonder about their role in water transport. Is this epidermis permeable to water and if so, how much? The answers to these questions have not yet been answered although research is beginning. (personal communication, Kimberly Fisher, Northwestern University)

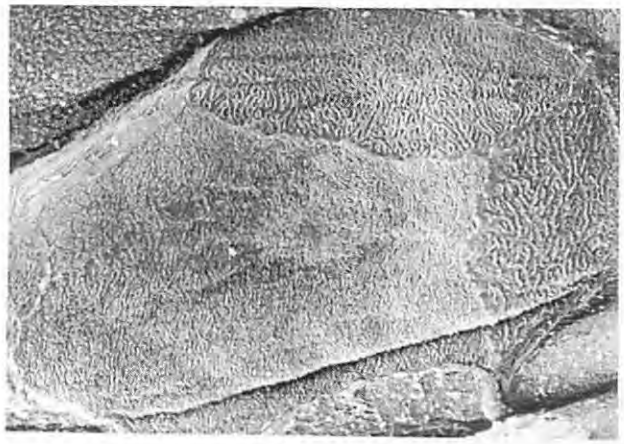


Figure 1A. This electron micrograph is of the cellular surface of the cell on the surface of the true vocal folds. You will note the microridge pattern that extends across the cell's surface. Near the surface, the cells become very flat and in this particular micrograph you can see that some cells have desquamated leaving a new, very flat cell underneath with a slightly different microridge pattern.

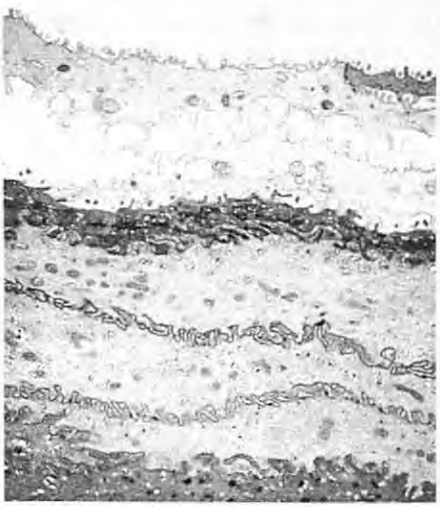


Figure 1B. This is a transmission electron micrograph of the surface of the epithelium of the vocal fold. At the top is a squamous cell which is just beginning to be desquamated. Note that its cellular attachments to the cell underneath it are no longer present. The cell is as if it were floating over the underlying cell. Also note the small microridge pattern present to this cell. These cells have many small intracellular organelles suggesting that this cell is metabolically active.

The luminal surface of these squamous cells is covered with a microridge pattern. See figure 1. The purpose of the micro-ridges is not completely known although some have postulated that it assists in mucous adherence, water absorption through increased surface area, or assists in traction during vocal fold vibration much like the treds of a tire prevents slippage.

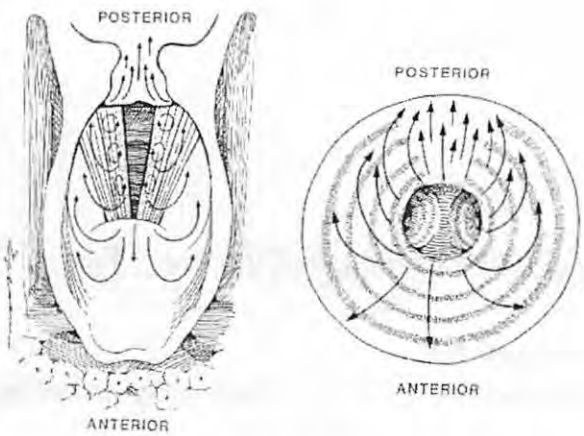


Figure 2. This schematic shows the direction of the ciliary flow found in human and canine experiments. The human experiments have not traced laryngeal mucociliary flow although the canine experiments have. Note that the mucociliary blanket flows posteriorly and circularly.

Probably one of the more interesting aspects of the epidermis is actually the mucociliary blanket which lies over the epidermis. The mucociliary blanket refers to a layer of mucous which covers the vocal folds. This mucous consists of two layers: a mucinous layer and a serous layer[29]. The mucinous layer is on top or is the more luminal layer and is filled with various mucin molecules. Part of the purpose of the mucinous layer is to prevent dehydration of the serous layer and of the underlying cilia and cells. The molecules in the mucin layer are designed to be protective to the underlying structures. The serous layer is the layer of the mucociliary blanket which is in more contact with the cilia and constitutes a much higher percentage of water. Therefore the serous layer is less viscous and is more watery, allowing the cilia to move through this layer with ease as compared with direct contact to the mucinous layer which is more thick and viscous. The mucociliary blanket is propelled up the trachea in a circular fashion with the cilia beating posteriorally and superiorly. The flow of the blanket is therefore posterior in a circular fashion and superior. Once the blanket reaches the posterior portion of the trachea, it goes directly upward and through the posterior glottis.

Studies have mapped the mucociliary flow by spraying or inhaling small micropledgets into a human or canine trachea and tracing their motion[29]. Pledgets which are sprayed immediately into the subglottis can be found to travel over the vocal folds, in essence being pushed by the mucociliary blanket over the non-ciliated squamous epithelium. About two to three millimeters lateral to the superior and leading edge of the vocal fold, the squamous epithelium again becomes a ciliated pseudocolumnar epithelium and at this point, the mucous blanket is pushed back posteriorly and superiorly until the mucous is swallowed. Based on unpublished research done in our laboratory on canine

animals, and the research of Fukuda, the mucociliary blanket flow is shown in figure 2[10]. It is this mucociliary flow which probably accounts for some laryngeal diseases having a propensity to occur in specific locations. For instance, it is common for many lung diseases which are infectious to have manifestations in the posterior glottis such as tuberculosis and chronic fungal infections. This location is predisposed to this disease due to the large volume of exposure through mucociliary blanket clearance.

The mucociliary blanket travels up the trachea at a somewhat amazing rate of 4 to 21 millimeters per minute in normal healthy subjects[41,44]. This fairly rapid moving, watery blanket keeps our vocal folds healthy and moist under normal conditions. However many activities may dry out the mucociliary blanket or may paralyze or partially slow the cilia thus impairing the function of the mucociliary blanket. For instance, it has been shown that one puff of a cigarette can slow the mucociliary blanket clearance to below normal levels[12,33]. The consequence of a mucociliary blanket which is moving slowly or which is too viscous has not been established in experimental models on voice production. However it is conceptually easy to understand that a very slow clearance of the mucociliary blanket will selectively predispose the mucous to more environmental toxins, more inflammatory agents, and dehydration. Furthermore, a viscous blanket or dehydrated blanket becomes harder to move for the cilia thus slowing it even more. Other environmental conditions such as cold or dryness can impair the mucociliary blanket clearance[28]. When cilia are destroyed or impaired recovery can take anywhere from a few hours to three weeks. In summary, the mucociliary blanket is an important first parameter of vocal health. Having a smooth, watery, fast flowing mucous blanket keeps our vocal folds moist and lubricated.

Basement Membrane Zone

The epidermis serves as a protective covering to the lamina propria. It also serves as a covering to help give shape and consistency to the lamina propria. The epidermis is secured to the lamina propria through the basement membrane zone. The basement membrane zone is a collection of protein and non-protein structures which together help the basal cells secure themselves to the rather amorphous mass of proteins present in the lamina propria. The epidermis is of course primarily cellular and the cells are attached to each other through desmosomes. Desmosomes are attachments between the cytoskeleton of cells to the cytoskeleton of the surrounding cells. These are strong attachments and can resist the pounding that our skin or vocal folds take. See figure 3. The lamina propria, on the other hand, is a very loose connection of fibrous and non-fibrous structures. This creates a dilemma which is, how does the epidermis secure itself to the lamina propria being that one

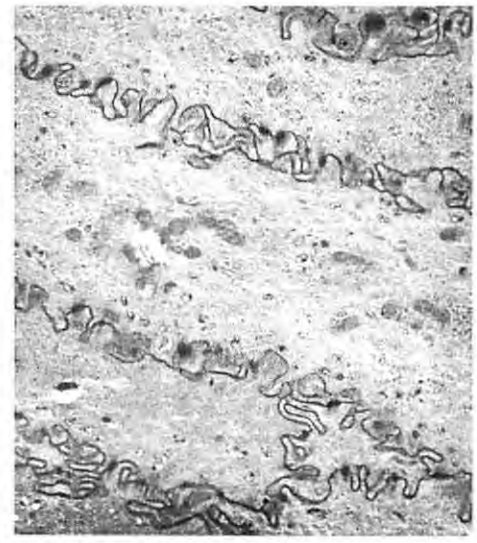


Figure 3A. This is a higher magnification of the squamous epithelium of the vocal folds. Cells are connected or held together to the opposite of another cell by tight areas of cellular junction called desmosomes. These desmosomes are represented by areas of black along the cellular junction. They are pointed out with arrows in this micrograph. Figure A is representative of normal vocal folds.

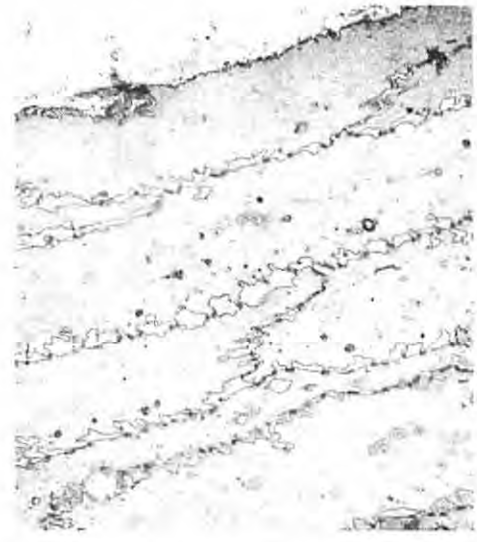


Figure 3B. Shows the epithelium from canine vocal folds when subjected to excessive vocal fold vibration. You will note that the cells are now held only to each other through these desmosomes. This indeed suggests that demosomes are some of the key anchoring elements of the vocal fold epithelium that helps the cells hold together during vibratory stress.

is primarily a cellular structure and the other a non-cellular structure. The body has resolved this dilemma with the basement membrane zone. See figure 4 (following page). Essentially the basal cells have anchoring filaments which secure the hemi-desmosome of the cell (half of a demosome) to the lamina densa and lamina lucida whose main substance is that of collagen type four. See figure 5. Similarly other structures, known as anchoring fibers (collagen type 7), loop

from the lamina densa into the lamina propria and back up into the lamina densa. Those who are familiar with the basement membrane zone will understand that this a simplified version but it will suffice to illustrate that securing the epidermis to the lamina propria during intense vibration is not an easy task and later we will see that this becomes critically important in some of our vocal pathologies.

It is important to point out that some of the proteins in the basement membrane zone, such as collagen type 7, have been shown to be genetically influenced[4]. For instance, the number of anchoring fibers (this is termed the "population density" of anchoring fibers) that one has in the basement membrane zone is genetically determined[13]. For example, the average person may have between 80 and 120 anchoring fibers per unit area of their basement membrane zone while someone who has a recessive form of the gene which doesn't create as many anchoring fibers may have only between 40 and 60 anchoring fibers per unit area. Those who are homozygous for the recessive gene will have few or no anchoring fibers[43]. This is an indication that there may be some genetic conditions that possibly predispose a voice to certain vocal conditions. For instance, if someone has more anchoring fibers than another, would their voice be less predisposed towards the occurrence of vocal nodules? The answer to that question is not known. Nor is it known if a lack of anchoring fibers is a key determinant in the formation of vocal nodules. Nevertheless, it does raise some interesting questions and makes us aware that genetics may play a role in vocal health.

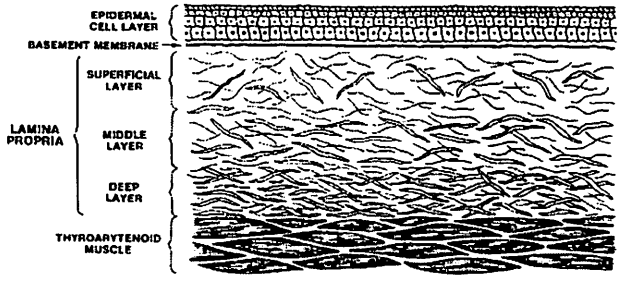


Figure 4. This schematic shows the epithelium in the three layers of the lamina propria with the basement membrane being a thin layer helping secure the epithelium and lamina propria together. Reprint Courtesy: Singular Publishing Group, Inc. (800)521-8545, 401 West "A" Street, #325 San Diego, CA 92101-7904. Gray SD, Hirano M, Sato K: Vocal Fold Physiology: Frontiers in Basic Science. *Titze IR (ed) "Molecular and Cellular Structure of Vocal Fold Tissue: 1-34, 1993.*

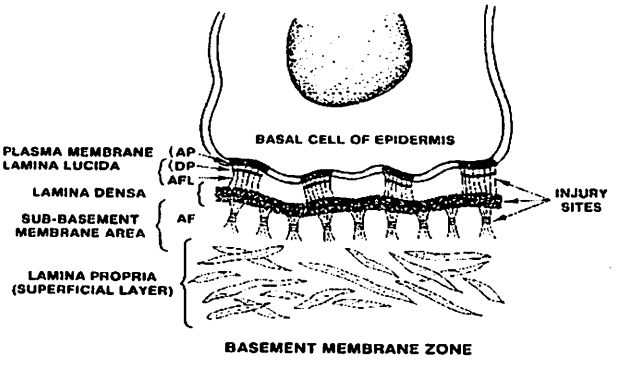


Figure 5. This schematic shows how the basement membrane zone is constructed. Essentially the basal cells hold themselves into the collagen type 4 and other laminar proteins through the use of anchoring filaments. Then anchoring fibers (collagen type 7) secure these laminar proteins to the lamina propria. Reprint Courtesy: Singular Publishing Group, Inc. (800)521-8545, 401 West "A" Street, #325 San Diego, CA 92101-7904. Gray SD, Hirano M, Sato K: Vocal Fold Physiology: Frontiers in Basic Science. *Titze IR (ed) "Molecular and Cellular Structure of Vocal Fold Tissue: 1-34, 1993.*

The Lamina Propria

The lamina propria of the vocal folds has traditionally been divided into three layers based on their histologic composition of elastin and collagen fibers: the superficial layer of the lamina propria (SLLP), the intermediate or middle layer of the lamina propria (MLLP), and the deep layer of the lamina propria (DLLP)[26,27]. The SLLP is characterized by fewer elastin fibers than the MLLP and DLLP. The MLLP is noticed by a rise in the elastin fiber concentration and the DLLP contains a modest decrease in elastin fibers coupled with an increase in collagen fibers (figure 6). In reality, such a contrast between layers is not readily noticeable in every individual specimen. However, as a general rule, these categories are descriptive and contain functional importance. The MLLP and DLLP together constitute what is termed the vocal ligament. This is an area of the lamina propria that bears longitudinal stress and thus has a more dense collagen fiber composition.

Hammond et. al found that the width of the layers based on the elastin fiber concentration varies with age but not so much by gender[19]. He described that the collagen concentration was fairly uniform throughout all layers. The collagen concentration varied more by gender and not by age[20]. The width of these layers also varies individually and since the collagen concentration is fairly level throughout all layers, the width is mostly dependent upon elastin fiber concentration.

Another useful division of the lamina propria is by biologic components. The lamina propria can be divided into cellular and non-cellular contents. In biology, the non-cellular molecules are also termed extracellular as opposed to intracellular molecules. In the field of extracellular biology, the term applied to the matrix of molecules found between the cells is the extracellular matrix or ECM. Therefore, in the lamina propria of the vocal folds, we may divide the tissue into the cells of the vocal fold and the extracellular matrix[31,30]. This is a useful division since the contents of the extracellular matrix influence the properties of tissue oscillation.

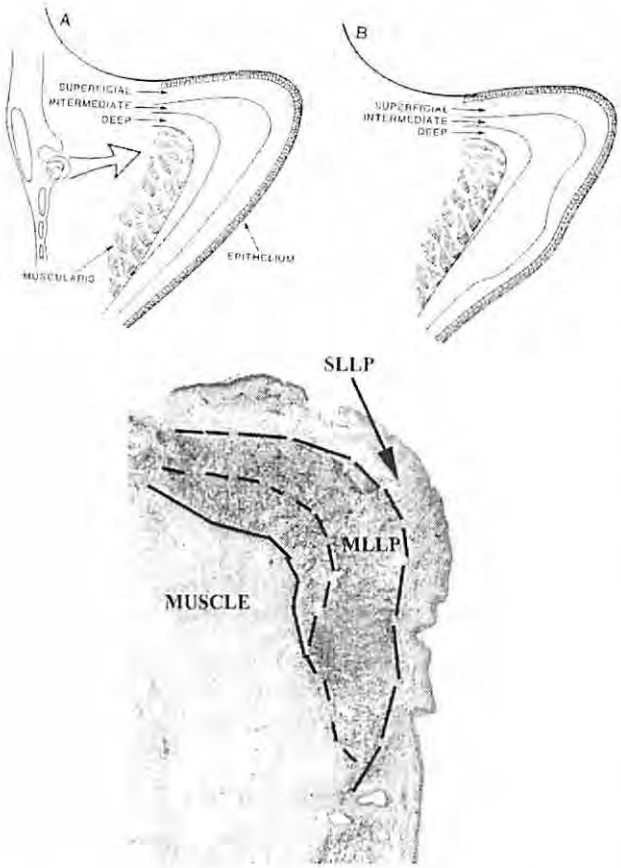


Figure 6 (a:top left). This schematic shows the three layers of the lamina propria. Reprint Courtesy: Annals Publishing Company. Gray SD, Dove H, Biclamowicz SA, Titze IR, Ludlow C: Experimental Approaches to Vocal Fold Alteration: Introduction to the Minithyrotomy. 108(1):2, 1999. (b: top right). This schematic shows a common variation in the shape of the middle layer of the vocal fold. Reprint Courtesy: Annals Publishing Company. Gray SD, Dove H, Biclamowicz SA, Titze IR, Ludlow C: Experimental Approaches to Vocal Fold Alteration: Introduction to the Minithyrotomy. 108(1):2, 1999. (c: bottom). This photograph of a histologic section of a vocal fold shows the three layers in a middle-aged male. Reprint Courtesy: Annals Publishing Company. Gray SD, Dove H, Biclamowicz SA, Titze IR, Ludlow C: Experimental Approaches to Vocal Fold Alteration: Introduction to the Minithyrotomy. 108(1):2, 1999.

Among the important cells of the lamina propria are the fibroblast, the myofibroblast and the macrophage. Other cells have been described but their role in phonation biology is not known. Catten et. al surveyed human vocal fold lamina propria for the population density of these cells[6]. He noted that about one-third of humans have a moderate concentration of macrophages present just below the basement membrane zone and the superficial layer of the lamina propria. Macrophages are cells which respond to and may cause inflammation. Their location, right below the epidermis, is suggestive that these cells are present to combat inflammatory agents which are crossing the epithelium. These agents may be bacteria, viruses, or environmental inhalants which are noxious to the tissue. Fibroblasts are cells which maintain the lamina propria. They are like custodians of the tissue. They perform the housekeeping roles of maintaining a healthy matrix. They replace old proteins and manufacture new proteins. They are present with a similar population density in all layers of the vocal fold. Myofibroblasts are fibroblasts which have differentiated into cells of repair[8]. They can be compared to a combination of the Red Cross and construction workers. These cells are only present when injury or damage has occurred and repair and construction is needed. Following tissue injury, these cells will show up to provide extracellular matrix repair and construction. Interestingly, these cells are found in the majority of normal human vocal folds. They are at the highest population density in the superficial layer with a decreasing population as tissue depth becomes deeper. This would likely indicate that in the normal human vocal fold some small amount of tissue injury is constantly present and that this tissue injury is greatest in the superficial layer.

The presence of myofibroblasts in almost all of our vocal folds suggest that all of us undergo some minimal or microscopic trauma frequently. It also suggests that vocal folds are extremely competent in repairing microscopic injury efficiently and without any significant compromise to the vocal fold tissue. It is when this injury reaches a more macroscopic state or that we don't allow the vocal folds to repair themselves that more serious forms of pathologies arise which are discussed later in this chapter. In order to maintain cellular health, we need to allow these cells to do their job. Clinically, most microscopic injury to the vocal folds seems to resolve itself within a day or two. Performers who experience vocal overuse one night will frequently mention that within two or three days their voice is back to a more normal performing standard. This is consistent with studies looking at basement membrane zone injury and repair. The basement membrane zone is extremely competent in repairing itself within a 36-48 hour time frame. This is mainly true for microscopic injury and if daily, constant, and vigorous injury is being experienced by the vocal fold tissue then the vocal folds may not be able to repair themselves adequately enough to prevent the onset of pathologies.

Extracellular Matrix Composition

The extracellular matrix has been divided by the class of molecules which make up the composition. In the vocal fold lamina propria this division is usually as follows: fibrous proteins, interstitial proteins, and other interstitial molecules such as carbohydrates and lipids[15,30]. The role of carbohydrates and lipids have not been studied and therefore no further comment will be made about these molecules. Both the fibrous proteins and the interstitial proteins have

Table 1.
Proteoglycans Found In the Human Vocal Fold

Proteoglycans	Function	Localization in Vocal Folds
Hyaluronic Acid	Creates and is the means for control of tissue viscosity, effects tissue flow resistance and tissue osmosis, provides space filling and space occupying molecules, probably helps determine lamina propria layer thickness[15,16]	Found throughout the ECM of the lamina propria, slightly more intense in the intermediate layer of the LP, found in macrophages and fibroblasts of the LP, evidence suggests gender specificity. Males>females
Decorin	Binds to collagen fibers resulting in delayed fibril formation and thinner fibril formation[10,20], may help reduce fibrosis and scar following injury	Found in the ECM of the lamina propria, may be more concentrated in the SLLP
Fibromodulin	Binds to collagen fibers resulting in delayed and thinner collagen fibers, both decorin and fibromodulin may effect ligament performance	Found in the ECM of the lamina propria, found mainly in the intermediate and deep layers of the LP, seems to be concentrated around the vocal ligament[22,37]
Versican	Has ability like hyaluronic acid to fill space, bind and organize water molecules	Found being manufactured by the fibroblasts and macrophages in the LP
Heparan Sulfate Proteoglycan	Binds to fibronectin, collagen IV, and laminin[23,24], may play a role in tissue morphogenesis	Found in the basement membrane zone of the vocal folds, found in the fibroblasts and macrophages of the LP

Aggregan was not found in vocal folds

Biglycan has not been searched for in the vocal folds

ECM – extracellular matrix
LP – lamina propria
SLLP – superficial layer of lamina propria

Data from Pawlak A, Hammond T, Gray SD: Immunocytochmical study of proteoglycans in vocal folds. Ann Otol Rhinol Laryngol. 105:6-11, 1996.

been the subject of more intense research in the recent years. We have already partially discussed two important fibrous proteins, that of collagens and elastins. Because of the ease of identifying these two proteins with histologic stains, these proteins have been studies for the last few decades. On the other hand, the interstitial proteins, those proteins found in between the fibrous proteins, have only been studied during this last decade. Pawlak described many of the interstitial proteins and are listed in table 1[39].

Fibrous proteins and interstitial proteins have various roles within the extracellular matrix. Collagens provide strength and structure to the tissue. They are useful in bearing stress and resisting deformation when subjected to a force. Elastin fibers convey elasticity to the tissue. Elasticity is defined as the ability to be deformed and return to the original shape. The vocal folds must be deformed and retain the ability of returning to their original shape[40]. In-

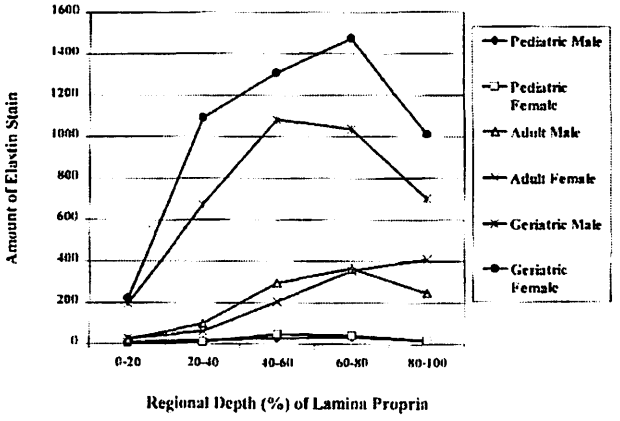


Figure 7. This figure displays the concentration of elastin fibers in human vocal fold lamina propria by depth in the vocal fold and by age. The x-axis displays depth in the vocal folds from superficial at 0% to deep in the lamina propria at 100%. 0% represents the basement membrane zone and 100% represents the vocalis muscle. The y-axis represents intensity or concentration of fibers. Displayed are various age groups and gender. Elastin fiber concentration was significantly different by age, but not by gender. Reprint Courtesy: Mosby Year Book, Inc. Otolaryngology Head & Neck Surgery Journal. Hammond TH, Gray SD, Butler J, et al: A Study of Age and Gender Related Elastin Distribution Changes in Human Vocal Folds. 119(4):317, 1998.

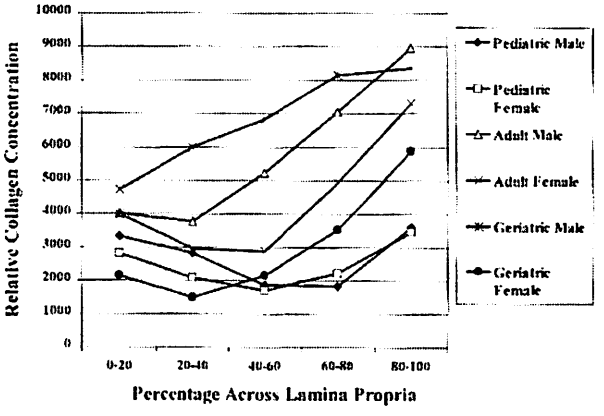


Figure 8. This shows the collagen fiber density in the human lamina propria vocal folds from superficial to deep. The x and y axis is similar to that described in figure 7. Displayed are various age groups by gender. Note that collagen fiber density was statistically different by gender but not by age. Reprint Courtesy: Annals Publishing Company. Hammond TH, Gray SD, Butler J: Age and Gender Related Collagen Distribution in Human Vocal Folds. In press, 1999.

terstitial proteins, on the other hand, effect the properties of viscosity. Viscosity is defined as how easy a substance flows. If a substance such as tissue flows easily, then the viscosity is low. Water for instance, would have a very low viscosity. Grease or something like motor oil would have a higher viscosity. Interstitial molecules not only effect but often control the viscosity of the vocal folds. Interstitial molecules also convey shock absorption properties to the tis-

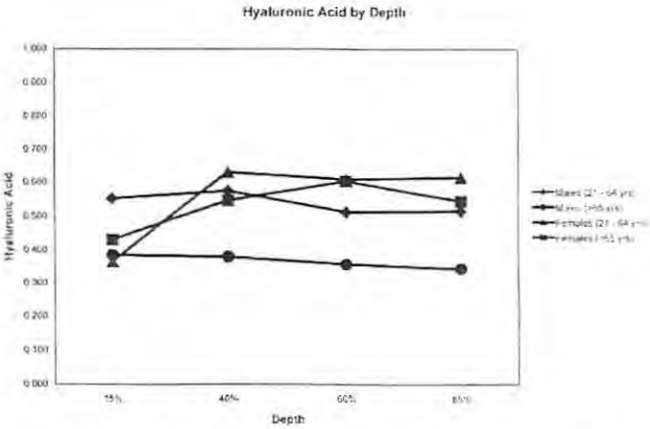


Figure 9. This figure shows concentration of hyaluronic acid in the human lamina propria vocal fold. Gender differences are present. X and Y axis is similar to that described in figure 7.

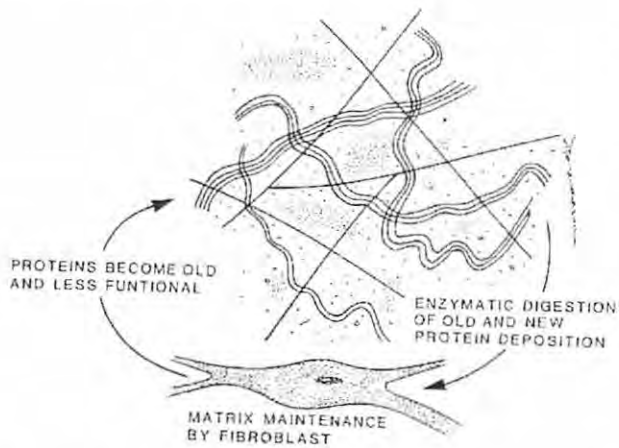


Figure 10. This schematic illustrates extracellular matrix turnover. Note that the fibroblast is the cell which is responsible for this. The fibroblast produces enzymes which degrade the extracellular matrix proteins and additionally, the fibroblast produces new proteins

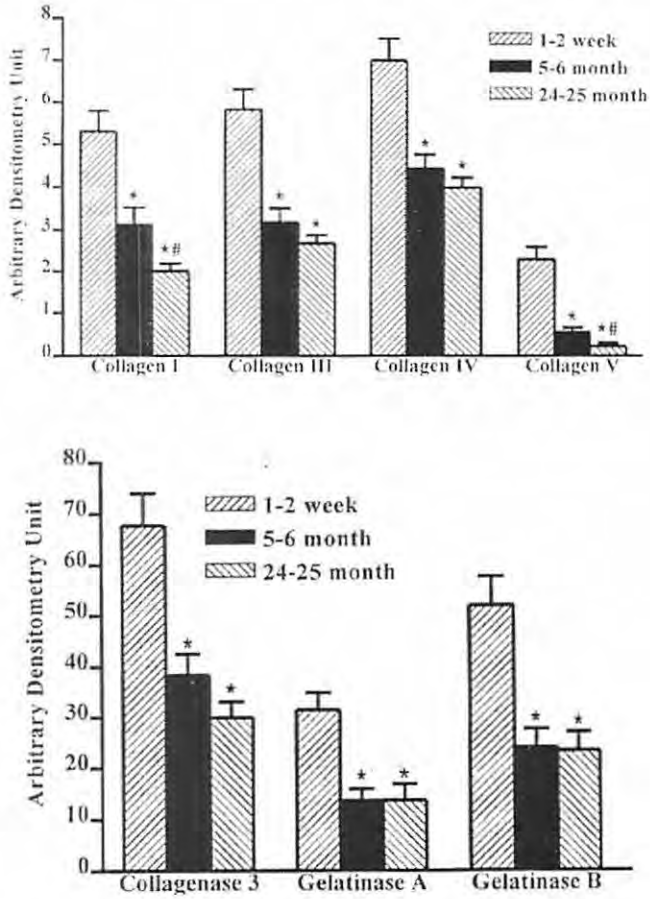


Figure 11 (A: top) This figure demonstrates collagen production in rat vocal fold lamina propria. Note that collagen production decreases significantly with age. (B: bottom). This figure illustrates collagenase production in rat vocal fold lamina propria. Note that collagenase production also decreases by age. Therefore, both collagen synthesis and collagen degradation decreases with age. The combined effect of this is decreased collagen turnover in the lamina propria. Although similar studies have not been done in human vocal folds, they have been performed in other human extracellular matrix areas. These studies also suggest that collagen turnover decreases significantly with age.

sue. Particularly the molecule of hyaluronic acid not only effects tissue viscosity but confers a dampening or shock absorption property[2]. Hyaluronic acid is an important component of synovial fluid in our knees, hips and other joints[32]. The greater the hyaluronic acid tissue concentration is, the higher the viscosity is and the more shock absorption properties the tissue has. Hammond et. al and Butler et. al found that hyaluronic acid concentration was higher in males than in females[18,5].

Distribution of the interstitial and fibrous proteins is influenced by age and gender and is maintained by the fibroblast. Patterns of distribution across the layers are shown for elastin fibers in figure 7,[19] for collagen fibers in figure 8,[20] and hyaluronic acid in figure 9[5]. A slow perusal of these graphs demonstrates many interesting trends and brings to mind many questions. Of interest is that the population density of elastin fibers in the lamina propria seems to be more correlated with age rather than gender. However, the proteins of hyaluronic acid and collagen appear to be more influenced by gender and not by age. Another critical question pertains to what is happening at the cellular level that causes the fibroblasts to produce one type of matrix composition 3 mm below the surface of the vocal fold while at just 1 mm below the surface, the fibroblasts produce a rather different matrix composition. One of the answers to that question may be regarding the effect of force on cells. It is becoming established that physical forces may influence the gene expression of that cell[22,34]. Perhaps

the forces found in the superficial layer of the lamina propria may be of a magnitude to cause different gene expression than that found at a slightly deeper level.

Extracellular Matrix Regulation

As indicated above, the regulation of this matrix is performed by the fibroblasts. The extracellular matrix proteins are constantly being surveyed and assessed. Old or injured molecules are destroyed enzymatically or phagocytized and new molecules are manufactured. This process of destroying old proteins and manufacturing new proteins is termed extracellular matrix turnover or regulation. Turnover refers to this process of constantly changing new for old proteins. This extracellular matrix turnover slows with age. (figure 10)

Regulation of these matrix proteins is accomplished by enzymes produced by the fibroblasts which destroy old proteins. Examples of these enzymes are collagenase, which enzymatically destroys collagen type one and collagen type three, and gelatinases, which destroy the collagen basement membrane proteins such as collagen type four. Elastase is an enzyme which enzymatically destroys elastin molecules. To make biology a little more complex, the body produces tissue inhibitors of metalloproteinases (TIMPS). Metalloproteinases are the class of enzymes comprising collagenases, elastases, and gelatinases[3,36]. TIMPS are produced by the body to inhibit the enzymatic activity of these enzymes. In order to determine regulatory aspects of the extracellular matrix of the vocal folds, we must consider the proteins which are producing extracellular matrix proteins, the enzymes which are degrading these proteins, and the TIMPS which are inhibiting these enzymes. Studies looking at all three of these have not been performed in human vocal folds. However, they have been performed in animal models. In these models, production of collagens in the vocal folds decreases dramatically with age. Enzymes which degrade collagens also decreases significantly with age. See figure 11. TIMPS for these enzymes do not decrease with age and stay at a relatively constant level throughout all age groups. The physiologic effect is that the turnover of collagen slows considerably with age. This means that collagens can become older and more aged before they are destroyed or replaced.

The effect of age on proteins is that proteins undergo increased cross linking with age. Cross linking is a process in which biologic bonds are made between parts of the molecule, which bonds are normally not there. The effect of these bonds is that they interfere with some of the normal biologic functions of the molecule. For instance, following cross linking, elastin molecules are not as elastic[37]. Collagen molecules, after cross-linking, become more stiff[12]. The older the molecule becomes, the more likely it is to experience cross linking and consequently, degradation of its biologic property. The physiological effect of this slowing extracellular matrix turnover is that as we age, our vocal folds become less elastic and more stiff. Future medical treatments may allow us to speed up turnover thus getting new or fresh proteins which are more elastic and less stiff.

Pathologies

Pathologies of the vocal folds are related to tissue changes. This infers that the cells of the vocal folds are creating and maintaining a tissue state that is not normal. Nodules, polyps, and Reinke's edema are all examples of tissue changes resulting in an abnormal sounding voice. As a general rule, these pathologies are confined to a superficial and occasionally intermediate layer. Articles have been written about the pathogenesis of nodules and polyps and therefore, these topics will be covered briefly in this chapter.[35,7,9]

Nodules

Nodules appear to be the result of injury to the superficial layer of the lamina propria and the basement membrane zone due to extensive vibration resulting in destruction of the tissue[7,14]. Nodules display a disorganized basement membrane zone with proliferation of certain basement membrane zone components which are not laid down in an organized fashion. Additionally, collagen type four and fibronectin are increased when compared to normal vocal folds. These types of histological findings are consistent with chronic repetitive injury to this tissue layer which results in an aberrant healing response with excessive collagen type four and fibronectin deposition. Undoubtedly, there are other pathologic protein depositions which have not yet been described

Polypoid Changes and Reinke's Edema

Polypoid changes have been associated with vascular lakes, increased fibrin deposition, and less fibronectin deposition[9,7]. The exact mechanism of injury is not as clear as that in nodules. However, it is proposed that these changes may be the result of chronic or acute voice injury whereas it is usually unlikely that a single acute injury results in nodules. Additionally, environmental effects, such as smoking, may play a role in the pathogenesis of these types of pathologies. Dikkers et al has provided some nice descriptions comparing polypoid changes with those of Reinke's edema[9]. There is considerable overlap between these types of lesions and their histological findings. It is apparent that the fibroblasts are certainly doing something different in these types of lesions than those found in normal vocal folds,

but the cellular activities in pathological states has not been elucidated.

Summary

In conclusion, the mucociliary blanket helps to protect the vocal fold and assists in vocal fold vibration by lubricating our vocal folds. The epithelium is designed for vibration and protects and gives form to the folds. The lamina propria and extracellular matrix are an important part of the vocal folds that influences the viscosity and elasticity of this tissue. This matrix suffers some loss of these biomechanical properties with age due to the slowing protein turnover that is present. A study of the cells of the vocal fold indicate that there is some mild inflammation present in many normal vocal folds and that most normal vocal folds undergo frequent injury. Cellular health is important. Since we know so little about cellular function in the vocal folds, only broad statements may be made about prevention and treatment of voice disorders. With more advances in this area, more specific programs of treatment will be developed. There are many areas of this science that need further study.

Acknowledgements

This work was partially supported by NIH Grant 1 RO1 DC04068-01.

References

1. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. Differentiated cells and the maintenance of tissues. *In* Molecular biology of the cell, 3 ed. New York and London, Garland, 1994, pp 1179-87.

2 Balazs EA, Gibbs DA: The rheological 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, pp 1241-1253.

3. Birkedal-Hansen H, Moore WGI, Bodden MK, et al: Matrix metalloproteinases: A review. Crit Rev. Oral Biol. Med. 4:197-250, 1993.

4. Briggaman RA, Wheeler CE: Epidermolysis bullosa dystrophica recessive: A possible role of anchoring fibrils in the pathogenesis. J Invest Dermatol 65:203-11, 1975.

5. Butler JE, Gray SD: Use of systemic dexamethasone in the treatment of nerve crush injuries in the rat model. Unpublished data.

6. Catten M, Gray SD, Hammond TH, et al: An analysis of cellular location and concentration in vocal fold lamina propria. Otolaryngology HNS J 118(5):663, 1998.

7. Courey M, Shohet J, Scott M, et al: Immunohistochemical characterization of benign laryngeal lesions. Ann Otol Rhinol Laryngol. 105:525-531, 1996.

8. Darby I, Skalli O, Gabbiani G: Alpha-smooth muscle action is transiently expressed by myofibroblasts during experimental wound healing. Lab Invest 63:21-9, 1990.

9. Dikkers F, Nikkels P: Benign lesions of the vocal folds: Histopathology and phonotrauma. Ann Otol Rhinol Laryngol 104:698-703, 1995.

10. Fukuda H, Kawaida M, Tatehara T, et al: A new concept of lubricating mechanisms of the larynx. *In* Fujimura O, ed: Vocal physiology: Voice production, mechanisms and functions. New York, Raven Press, Ltd, 1988, pp 83-92.

11. Goetnick PF: Proteoglycans in development. *In* Kimmel CB, ed: Current topics in developmental biology (vol 25). New York, Academic Press, 1991, pp 111-131.

12. Goodman RM, Yergin BM, Landa JF, et al: Relationship of smoking history and pulmonary function tests to tracheal mucous velocity in nonsmokers, young smokers, ex-smokers and patients with chronic bronchitis. Am Rev Respir Dis 117(2):205-214, 1978.

13. Gray SD, Shirley SN Pignatari, Harding, P: Morphologic ultrastructure of anchoring fibers in normal vocal fold basement membrane zone. J of Voice 8:48-52, 1994.

14. Gray SD, Hammond EH, Hanson D: Benign pathologic response of the larynx. Ann Otol Rhinol Laryngol 104:8-13, 1995.

15. Gray SD, Titze IR, Chan R, et al: Vocal fold proteoglycans and their influence on biomechanics. Laryngoscope 109(6):845-854, 1999.

16. Hammond E, Griffin J, Odell WD: A chorionic gonadotropin secreting human pituitary cell. J Clin Endocrinol Metab. 71:747-754, 1991.

17. Hammond EH, Hansen JK, Spencer LS, et al: Immunofluorescence of endomyocardial biopsy specimens: methods and interpretation. J Heart Lung Transplant. 12:113-124, 1993.

18. Hammond TH, Zhou R, Hammond EH, et al: The intermediate layer: A morphologic study of the elastin and hyaluronic acid constituents of normal human vocal folds. J of Voice 11:59-66, 1997.

19. Hammond TH, Gray SD, Butler J, et al: A study of age and gender related elastin distribution changes in human vocal folds. Otolaryngology HNS J 119(4):314, 1998.

20. Hammond TH, Gray SD, Butler JE: Age and gender related collagen distribution in human vocal folds. Ann Otol Rhinol Laryngol 1999; in review.

21. Hardingham TE, Fosang AJ: Proteoglycans: Many forms and many functions. FASEBJ. 6:861-70, 1992.

22. Harris RC, Haralson MA, Badr KF: Continuous stretch-relaxation in culture alters rat mesangial cell morphology, growth characteristics, and metabolic activity. Lab Invest. 66(5):548-554, 1992.

23. Heinegard D, Franze A, Hedbom E, et al: Common structures of the core proteins of interstitial proteoglycans. Ciba Found Symp. 124:69-88, 1986.

24. Heremans A, Van der Schueren B, De Cock B, et al: Matrix associated heparan sulfate proteoglycan: core protein-specific monoclonal antibodies decorate the pericellular matrix of connective tissue cells and the stromal side of basement membranes. J Cell Biol. 109:3199-3211, 1989.

25. Heremans A, De Cock B, Cassiman JJ, et al: The core protein of the matrix associated heparan sulfate proteoglycan binds to fibronectin. J Biol Chem. 265:8716-8724, 1990.

26. Hirano M: Structure of the vocal fold in normal and disease states. Anatomical and physical study. ASHA Report. 11:11-30, 1981.

27. Hirano M, Kakita Y: Cover-body theory of vocal fold vibration. *In* Daniloff RG (ed): Speech Science. San Diego, College-Hill Press, 1981, pp 1-46.

28. Hirsch JA, Tokayer KL, Robinson MJ, et al: Effects of dry air and subsequent humidification on tracheal mucous velocity in dogs. J Appl Physiol 39:242-246, 1975.

29. Kobayashi K, Wanner A: Mucociliary clearance and ciliary activity. *In* Chung KF, Barnes PJ (eds): Pharmacology of the respiratory tract. New York, Marcel Dekker, Inc, 1993, pp 621-654.

30. Labat-Robert J, Bihari-Varga M, Robert L: Extracellular matrix. FEBS Lett. 1990;268:386-93.

31. Lander AD: Proteoglycans. Guidebook to the extracellular matrix and adhesion proteins. *In* Kreis T, Vale R, (eds.). New York, Oxford University Press, 1993, pp 12-16.

32. Laurent TC, Laurent UBG, Fraser JR: Functions of hyaluronan. Annals of Rheu Dis. 54:429-432, 1995.

33. Lourenco RV: Distribution and clearance of aerosols. Am Rev Respir Dis 101:460-463, 1970.

34. Margolis LB: Induction of cell processes by local force. J Cell Sci. 98:369-373, 1991.

35. Mossallam I, Kotby M, Ghaly A, et al: Histopathological aspects of benign vocal fold lesions associated with dysphonia. *In* Vocal Fold Histopathology: A Symposium, edn 1. San Diego, College Hill Press, 1986, pp 65-80

36. Murphy G, Docherty AJP: The matrix metalloproteinases and their inhibitors. Am J Respir Cell Mol Biol. 7:120-125, 1992.

37. Niewoehner DE, Kleinerman J, Liotta L: Elastic behavior of post-mortem human lungs: Effects of aging and mild emphysema. J Appl Physiol. 39:943, 1975.

38. Oldberg A: Fibromodulin. *In* Kreis T, Vale R, (eds). Guidebook to extracellular matrix and adhesion proteins. New York, Oxford University Press, 1993, pp 55-56.

39. Pawlak A, Hammond T, Hammond E, et al: Immunocytochemical study of proteoglycans in vocal folds. Ann Otol Rhinol Laryngol., 105:6-11, 1996.

40. Perlman AL, Titze IR, Cooper DS: Elasticity of canine vocal fold tissue. J Speech and Hear Res. 27:212-219, 1984.

41. Sackner MA, Landa J, Hirsch JA, et al: Pulmonary effects of oxygen breathing: A 6-hour study in normal men. Ann Intern Med 82:40-43, 1975.

42. Schneider SL, Kohn RR: Effects of age and daibetes mellitus on the solubility of collagen from human skin, tracheal cartilage and dura mater. Exp. Gerontol. 17:185, 1982.

43. Tidman MJ, Eady RAJ. Evaluation of anchoring fibrils and other components of the dermal-epidermal junction in dystrophic epidermolysis bullosa by a quantitative ultrastructural technique. J Invest Dermatol 84:374-7, 1985.

44. Wanner A: Clinical aspects of mucociliary transport. Am Rev Respir Dis 116:73-125, 1977.

Synopsis Page

The chapter discusses some of the tissue, cellular architecture and physiology relevant to phonation biology. The mucociliary blanket and its role in external vocal fold lubrication is presented. The epithelium, basement membrane zone and lamina propria all have specific roles in oscillating tissue. Three cell types, the fibroblast, myofibroblast and macrophage maintain important and unique roles. Protein turnover in the lamina propria is important and slowing matrix turnover may be a leading factor in creating some of the characteristics associated with vocal senescence. Lastly, aspects of cellular health and cellular pathology are discussed.

NCVS Status and Progress Report - 14
September 1999, 73-77

Dissection Plane of the Human Vocal Fold Lamina Propria and Elastin Fiber Concentration

Steven Gray, M.D.
Division of Otolaryngology – Head and Neck Surgery, University of Utah School of Medicine
Karen Chan, B.S.
Statistical Data Center, LDS Hospital, Salt Lake City
Bryan Turner, M.D.
University of Utah

Abstract

To determine if a natural plane of dissection occurs in the normal human vocal fold, semi-blunt instruments (Bouchayer laryngeal dissector) were used to dissect the lamina propria. The depth of the plane of dissection was correlated with the elastin fiber concentration to discover if the plane occurred at a predictable point in the elastin concentration as it increased between the superficial and middle layers. Eight human larynges were dissected using an operative microscope. The dissection plane consistently occurred between 23-50% depth into the lamina propria. No consistent correlation was found with the elastin fiber concentration. Depth of the plane of dissection has an interesting association with age: dissection planes occurred more superficial in older specimens.

Introduction

The three layered structure of the lamina propria of the human vocal fold has received much attention from laryngologists over the two decades. Clinically, surgeons have experienced that surgery confined to the superficial layer of the lamina propria (SLLP) generally heals better than surgery that violates the middle or deep layers of the lamina propria. Lesions that invade the deeper layers of the vocal folds, such as cysts, may prompt the surgeon to take a different phonosurgical approach, since these lesions may have a less favorable prognosis than lesions limited to the SLLP. This concern about confining surgery to the SLLP has prompted caution and counsel among teachers of laryngology to correctly finding the SLLP and not inadvertently dissecting into the middle layer when excising lesions or raising microflaps.

As has been well described, the superficial layer is histologically different from the deeper two layers based on their collagen and elastin content. The middle and deep layers are histologically similar and form the vocal ligament(1,2). The deep layer of the lamina propria is quite tightly bound to the thyroarytenoid muscle, and dissection between the two is difficult and generally requires sharp dissection. In contrast, dissection within relatively normal (not scar) SLLP seems to be much easier and can often be performed safely with blunt (non-sharp) instruments or dissectors.

We proposed the following study to determine the relative ease and safety of developing microflaps in the SLLP. If the SLLP is truly a pliable, loose tissue then using non-sharp instruments for the majority of the dissection should be possible and perhaps preferred. We also wanted to answer the following two questions:

1. When raising microflaps, is the dissection prone to occur in the SLLP, at the junction of the SLLP and middle layer, or occasionally deeper?
2. Does the dissection plane have a relationship to the elastin content (and vocal ligament) of the lamina propria?

We are curious about the relationship of a surgical dissection plane as it relates to the elastin fiber concentration since an increase in the elastin fiber concentration has been the traditional histologic defining characteristic of the end of the SLLP and the beginning of the middle layer of the lamina propria. Depth measurements of the lamina pro-

pria, such as 1.5 mm deep, do not suffice in determining if the dissection occurs in the SLLP or whether the middle layer has been violated since the thickness of the SLLP varies individually and with age(3,4). Hammond et al. showed that SLLP thickness varies significantly with age, with young human vocal folds showing quite thick SLLP and geriatric folds showing thin SLLP.

Methodology

Eight human cadaveric larynges were used for the dissection. These larynges were harvested within 18 hours of death and had no history of laryngeal disease or manipulation such as intubation. Larynges were selected from an adult age range of 18 to 76 years old. The larynges were mounted by inserted a fixed tube into the attached first few tracheal rings and then fixing the thyroid cartilage with some pins to a fixed structure. A silk suture was placed through the thyroid cartilage at the most anterior portion and fixed securely so that the thyroid cartilage was pulled forward and slightly inferior. This put the vocal folds on stretch and opened up the larynx for easier surgical access to the vocal folds.

A Zeiss operation microscope with a 250-mm lens was used for the surgical dissection. An incision was made with a #15 blade through the epithelium just superior and lateral (about 1 mm) to the free edge of the vocal fold. Then using a Bouchayer laryngeal dissector, a plane of dissection was established beneath the epithelium (See Figure 1). Occasionally a small cup forcep was used to provide some counter tension for the dissection. Belluci scissors were sparingly used when small fibrous bands spanned the dissection plane. When Belluci scissors were used, which did not occur often, it was generally in the immediate intrafold region of the dissection. The microflap was raised inferiorly until the dissection plane was beyond (below) the vocal ligament.

Then the flap was replaced to the original position and the vocal folds were excised and fixed in formalin. Histologic sections were made through the mid-section of the membranous vocal fold. Four-micron thick sections were stained to detect elastic fibers using Verhoeff elastic tissue stain (known as Elastin-Van Giesan or EVG).

An image analysis system previously described to determine elastin fiber concentration was used(3). An optical interference filter of 630 nm with a 10-nm bandwidth was used to restrict the wavelengths of light bombarding the specimen. This wavelength was chosen because it deemphasizes the pink in the stain and accentuates the black staining of the elastin fibers. The elastin black staining is then measured and quantified using a gray scale. The intensity and extent for a given region, in this case a high power field, can be determined.

High-powered fields were taken from the epithelium to the dissection plane, and then fields were taken from the dissection plane to the vocalis muscle. High-powered fields started at the free edge of the vocal fold and were taken (imaged) in a straight line from the epithelium to the vocalis muscle, perpendicular to the vocal ligament. Elastin concentrations for each vocal fold were plotted and the place of the dissection plane marked.

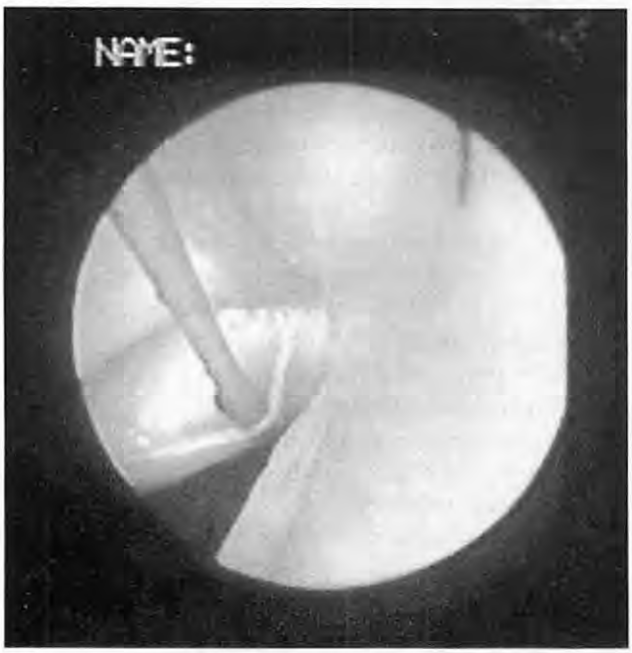


Figure 1a. This picture shows the methodology using the Bouchayer dissection instrument to bluntly raise a microflap.

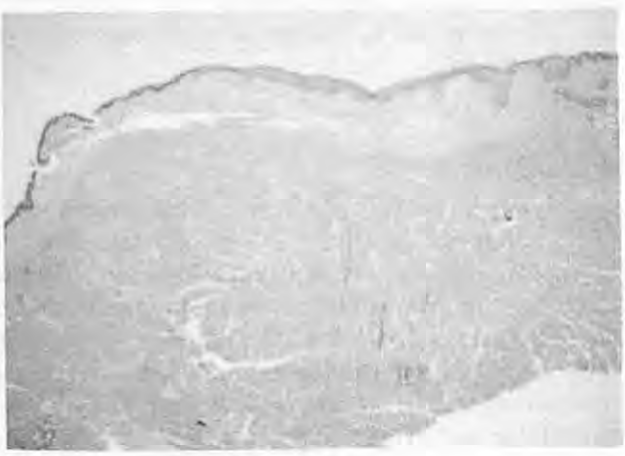


Figure 1b. This figure demonstrates the histological result and level of dissection associated with this microflap.

Data Analysis

The number of images (high power fields) used to cross the lamina propria varied because the thickness of the lamina propria varied between larynges. Each image was assigned a relative percent value based on its position in the overall depth of the lamina propria. For instance, in a sample that required ten fields, the first image spanned 0%-10%; thus, the first image was assigned the midpoint of this range, 5%. Similarly, the second image spanned 10%-20%; therefore, the second image's relative depth was 15%.

Elastin amounts by relative depth across the lamina propria were plotted for the eight larynges. A sequence of consecutive measurements across the lamina propria was defined as a pass. For a given location in a larynx (i.e., left or right anterior, left or right posterior), two or three passes across the lamina propria were performed. For each pass an attempt as to the relative location of the level of dissection (LOD) was made. All measurements for a given location in a larynx were averaged in order to estimate the relative depth of the LOD for that given location in a given larynx. If passes were performed on more than one location for a given larynx, and the location-specific LOD estimates were within 20% of each other, then the location-specific LOD estimates for a given larynx were averaged across location. The estimated LOD per individual larynx by age were plotted, and Pearson's correlation coefficient (r) was reported.

Association of elastin amounts and relative depth of the lamina propria were described by reporting Spearman's correlation coefficients (r_s) per individual larynx. Spearman's correlation coefficients were reported instead of Pearson's correlation coefficients because the scales for measuring intensity and relative depth of dissection were extremely different. For example, the scale for elastin fiber intensity for a given larynx might span 0 to 12-million, while the relative depth ranged from 0%-100%. Descriptive statistics for the relative depth of the LOD were also reported.

Results

The results are summarized in Table 1. Spearman's correlation coefficients (r_s), relative LOD depth (%), along with more detailed description of the elastin curves are summarized by specimen.

The total range of the dissection depth went from 50% to 23%, and the specimen with the deepest depth of dissection (50%) was the youngest at age 18. The specimens showed a mean dissection depth of 34% (n=5) into the vocal folds with a standard deviation of 11% and 95% confidence interval from 20% to 47% depth. Larynx #2755 had surgical artifact from an inadvertent Belluci scissor cut that changed the level of dissection by greater than 20%. The average age of the specimens (n = 7) was 43 years with a standard deviation of 18 years. The age of the specimens ranged from 18 to 76 years. The gender and age for one of the specimens was not listed from the coroner's office and, despite attempts for information, was not given. There were not enough specimens to ascertain gender difference.

Depth of the occurring dissection plane and age appeared to be negatively associated, correlation = -0.93 (n = 4, p = 0.066). The mean level of the dissection became more superficial with age. The result of this analysis is displayed in Figure 2.

Examination of the different passes from the eight specimens revealed two distinct curve patterns as they relate to elastin fiber concentration.

1. A rise and plateau pattern.
2. A rise and descend pattern.

Table 1.
Summary of Results by Larynx

Larynx No.	Curve Pattern	Gender	Age	Relative LOD Depth (%)	LOD Occurrence	r_s	p-value
2760	Rise/Descend	**	**	23	At Initial Rise	0.500	0.048
2084	**	Male	18	50	**	-0.245	0.467
2231	Rise/Plateau	Female	33	39	At Peak	0.685	0.042
2774	Rise/Descend	Male	37	33	At Peak	0.175	0.374
2755	Rise/Descend	Female	41	5*	**	0.661	0.038
2632	Rise/Plateau	Male	48	31	At Initial Rise	0.613	<0.001
2639	Rise/Descend	Male	49	31	At Peak	-0.279	0.177
2757	Rise/Plateau	Female	76	27	At Initial Rise	0.892	<0.001

*Sample omitted from descriptive statistic calculations because dissection depth was forced to be unnaturally thin.

**Unknown or unable to determine.

***Depth of dissection plane into vocal fold. (ex:.50-50% depth between epithelium and vocalis muscle)

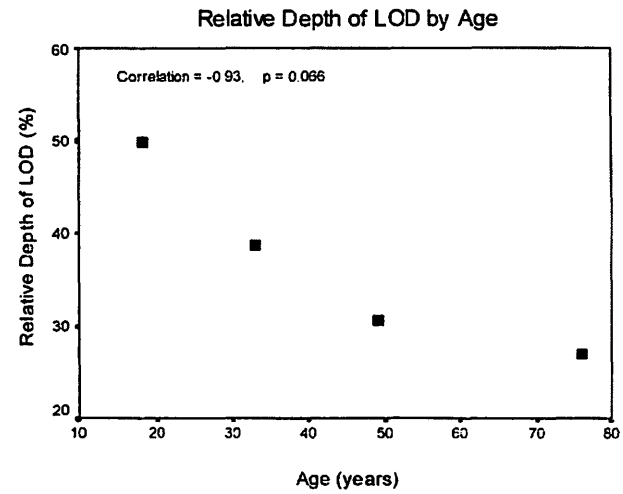


Figure 2. This figure shows the level of dissection as it correlates with age in males. Note the level of dissection is thicker in younger patients and becomes more superficial with age.

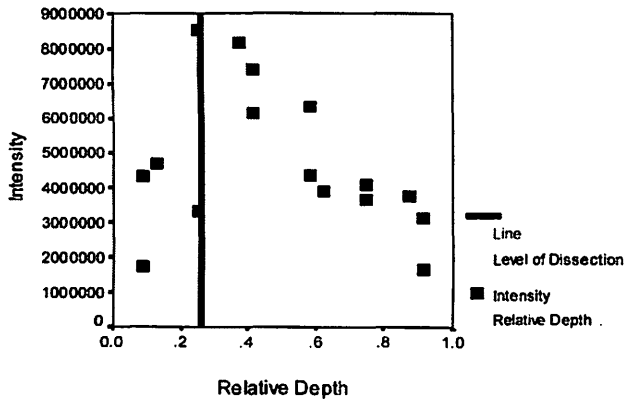


Figure 3. An example of one of the specimens correlating intensity of elastin fiber concentration and the level of dissection. In this example, the level of dissection occurred at the initial rise of the elastin fiber concentration. This project did not demonstrate a correlation between elastin fiber concentration and the level of the dissection plane. However, the dissection plane did occur between the rise of the elastin fiber concentration or at the peak of the elastin fiber concentration.

Three of the eight larynges displayed a rise and plateau pattern, while four of the eight specimens indicated the rise and descend pattern. One of the specimens had no discernable pattern. An example of the elastin curve pattern in relation to the level of dissection is shown in Figure 3.

Consistent association between elastin fiber concentration and relative depth into the lamina propria was not found. Correlation coefficients ranged from –0.279 to 0.892 for the eight larynges (see Table 1).

The plane of dissection did not consistently occur in the same location on the elastin curves. For either curve pattern, the dissection plane either occurred *just* as the elastin fiber concentration began to increase, or the plane occurred at the *maximum* peak of the elastin concentration. For example, in Figure 3, the dissection plane occurred when the elastin fiber concentration began to increase. Table 1 presents in tabular form the dissection plane location in relation to the elastin curve for all larynges used in this study.

Discussion

This study indicates that it is difficult to dissect deeply into the vocal ligament when blunt dissection is predominantly used in creating a microflap. The results would also indicate that as the superficial layer becomes the intermediate layer, an area exists that easily dissects and separates more than the surrounding tissue. Consistently the separation occurred with 95% probability between 20-47% depth. Interestingly, dissection did not routinely occur more superficially. Based on elastin fiber concentration only, one would expect the plane of dissection to occur anywhere in the first 40-50%. However, the dissection occurred either as the elastin fiber concentration was increasing or occurred at the peak of the elastin fiber concentration. This suggests that perhaps some other proteins are influencing the areas of lamina propria weakness and areas of easy separation.

The superficial layer does have elastin and collagen material. The elastin material in the superficial layer does not appear to be predominantly in fiber form and consequently does not stain with the EVG stain. But the SLLP does contain considerable elastin in the forms of elaunin and oxytalan. The SLLP also has other proteins as part of the normal extracellular matrix. These are fibronectin, decorin and hyaluronic acid. Fibronectin is a globular glycoprotein that has some adhesive qualities. Decorin and hyaluronic acid are proteoglycans that exert influence over collagen fibril formation and have viscoelastic properties. Together, these fibrous and interstitial proteins form a matrix with certain properties(4,5).

Similarly, the intermediate layer is composed of hyaluronic acid, a proteoglycan called fibromodulin, and an increased number of elastin fibers. Other proteins, not yet identified, are likely present in these layers and further research will contribute to our understanding of these layers. The composition difference between the two layers appears to be a) increase in elastin fibers, b) fibromodulin and c) perhaps differences in hyaluronic acid. The increase in elastin fibers has traditionally been the area of focus for the intermediate layer, but the dissection planes did not correlate tightly with the concentrations of elastin. The small chain proteoglycans, especially fibromodulin and decorin, bind to and interact with collagen fibers and constitute an important part of tendons and ligaments. It is certainly possible that the plane of dissection was influenced by the presence of these proteins rather than the elastin. Fibromodulin has been identified to be vocal fold lamina propria layer specific in some individuals(6). This proteoglycan has only been studied in a few vocal folds and consequently general statements applying to population groups cannot be made. Clearly, fibromodulin does appear to be a major constitutional difference between the two lamina propria layers.

The association between depth of the plane of dissection and age is intriguing. Hammond et al showed that the elastin fiber concentration increases with age, and consequently, the superficial layer may become thinner either because of infiltration from middle layer protein components or atrophy of superficial layer components(7). This study seems to fit with his findings, that the transition of the superficial and middle layer moves superficially as we age. The reason for this migration is not known. The clinical importance of these findings is that the microflaps which surgeons develop will likely be thinner in older patients, as the entire superficial layer seems to become thinner with age.

In summary, it appears that within the layered structure of the lamina propria, there is an area that is more likely to separate surgically than the surrounding tissue. Using blunt dissection, that dissection plane occurs with 95% probability between 20-47% depth. No association was found between the plane of dissection and the concentration of elastin fibers. It is likely that other molecular components of the extracellular matrix influence the ease of tissue separation and consequently the plane of dissection. However, the blunt dissection plane did move superficially with increasing age. The plane of dissection never went deeper than 50% into the fold. Researchers may also receive some confidence that when using blunt dissection in human larynges, components of the SLLP will be in the superficial flap while components of the vocal ligament will be deep to the plane of dissection.

Acknowledgement

This work was supported by NIH Grant #P60 DC00976 to the National Center for Voice and Speech from the National Institute on Deafness and Communicative Disorders (NIDCD).

References

1 Hirano M: **Structure of the vocal fold in normal and disease states. Anatomical and Physical Study**. *ASHA Report*. 1981;11:11-30.

2 Hirano M, Kakita Y: **Cover-body theory of vocal fold vibration**. In: Daniloff RG, ed. *Speech Science*. San Diego, CA: College-Hill Press; 1985:1-46.

3 Hammond TH, Gray SD, Butler JE, Zhou R, Hammond E: **A study of age and gender related elastin distribution changes in human vocal folds**. *Otolaryngology HNS J*. In press.

4 Gray SD, Titze IR, Alipour F, Hammond TH: **Biomechanical and Histologic Observations of Vocal Fold Fibrous Proteins**. *Annals of Otology, Rhinology and Laryngology*, in press.

5 Gray SD, Titze IR, Chan R, Hammond TH: **Vocal Fold Proteoglycans and Their Influence on Biomechanics**. *The Laryngoscope*, in press.

6 Pawlak A, Hammond TH, Hammond EH, Gray SD: **Immunocytochemical study of proteoglycans in vocal folds**. *Annals of Otology, Rhinology, and Laryngology*, 1996;105:6-11.

7 Hammond TH, Zhou R, Hammond EH, Pawlak A, Gray SD: **The intermediate layer: A morphologic study of the elastin and hyaluronic acid constituents of normal human vocal folds**. J of Voice 1997;11:59-66.

NCVS Status and Progress Report - 14
September 1999, 79-87

Splicing Patterns of Fibronectin (FN) Precursor mRNA and Expression of Genes Regulating FN in Rat Vocal Fold Extracellular Matrix

Hu Ding M.D.
Steven Gray, M.D.
University of Utah, School of Medicine
Department of Surgery, Division of Otolaryngology – Head and Neck Surgery

Abstract

Vocal tissue may become stiffer with age, suggesting a possible contribution of extracellular matrix aberration to vocal senescence. Using reverse-transcriptase polymerase chain reaction (RT-PCR), patterns of alternative splicing of the EIIIA, EIIIB and V regions of the fibronectin (FN) precursor mRNA (pre-mRNA) were determined in rat vocal folds, skin and lungs. mRNA levels for total FN, EIIIA, EIIIB and V regions, and FN-degradation related proteinases were measured for possible senescent changes in vocal folds of rats of 1.5±0.5 weeks (neonatal), 6.0±0.5 months (adult), and 24±0.5 months (elderly). Alternative splicing of FN pre-mRNA generates EIIIA+/-, EIIIB+/- and V+/- FN isoforms in the vocal folds, and EIIIA+/-, EIIIB- and V+/- isoforms in skin and lungs. Relative abundance in expression of isoforms derived from the same splicable regions of FN pre-mRNA changes with age and tissue, and in vocal folds of the elderly rats, significant overexpression of the EIIIB+ isoform was observed, with mRNA levels for total FN and EIIIA, also elevated. In summary, the results demonstrate differential patterns of alternative splicing of FN pre-mRNA, indicating that the EIIIB+ coded FN isoform may be important in voice senescence. There is an enhanced accumulation of FN molecules in the aged vocal folds, which may contribute to the increased stiffness of the elderly vocal folds.

Introduction

Human voice changes with age (43,19,29). One contributory factor is that with age human vocal folds become less elastic, more viscous and stiffer (9). These age-associated alterations in physical properties of vocal folds have been partially attributed to senescent changes in alignment and distribution of connective tissue fibers in the vocal folds (38,23).

Human vocal fold is a laminar structure (20,25). Configuration and structure of the layers appear to be the major determinants of their physical properties, and in turn, the performance of the vocal folds. At molecular levels, it is recognized that the composition and levels of extracellular matrix (ECM) proteins are critical in maintaining the integrity, elasticity, viscosity and stiffness of vocal folds (9). Thus, any abnormalities or aberration in vocal fold ECM proteins may theoretically cause alterations in physical properties of this organ and lead to voice changes.

Fibronectins (FNs) are high molecular weight, multi-domain/multi-functional adhesive glycoproteins present in plasma, extracellular matrices and basement membranes, and on the surface of a variety of cell types. FNs bind to a number of biological macromolecules including heparin, collagen, fibrin, and cell surface receptors (integrins). The biological functions of FNs known to date include cytoskeletal organization, cell morphology including oncogenic transformation, cell differentiation and migration including oncogenic metastasis, embryonic development, wound repair, and blood clotting (for review, see 37,22,36). As one of the most prominent structural glycoproteins in ECM, FNs play important roles in organization and integrity of extracellular matrices through their interactions with the other matrix molecules, in particular, collagens (37,22,36).

FN proteins are composed of three types of repeating homology units designated as types I, II, and III (22,36). Whereas FN is encoded by a single gene (36), a number of FN isoforms of variable molecular weight exist (34). A partial explanation for generation of FN protein species is that the single primary messenger RNA transcript (pre-

mRNA) of FN may undergo alternative splicing in three regions, namely EIIA, EIIB, and V, respectively (36,24,40). The functional significance of alternative splicing of FN pre-mRNA is not fully understood; however, it is believed that the modulation of FN functions in cell-cell and cell-substrate interactions in a variety of physiological situations is achieved by the selective expression of different isoforms (24,40).

FNs are abundantly present in the vocal folds as detected by immunostaining technique. Deposition of FN proteins in the vocal folds varies differentially in benign vocal lesions (18,12). That FN may participate in voice senescence has not been studied yet, but it is strongly indicated by the observation that, in tissues other than vocal folds, alternative splicing of FN pre-mRNA may change with age and developmental stages (26,7,33).

To understand the regulation of fibronectin synthesis in the vocal folds and elucidate mechanisms by which fibronectin may participate in voice senescence, we characterized the splicing patterns of the FN gene transcript in the vocal folds and evaluated senescent alterations in vocal fold mRNA levels for FN isoforms. Data obtained from the vocal folds was compared to the data from the skin and lungs.

Method and Materials

Animals

F344 Sprague Dawley rats were purchased from National Institute on Aging (Bethesda, MD). Three groups of rats, aged 1.5±0.5 weeks, 6.0±0.5 months, and 24±0.5 months, were studied. To obviate gender effects, only male rats were used. Prior to the study, animals were kept in the Animal Resource Center of University of Utah Health Science Center, with access of regular rat chow and drinking water ad libitum, for four days to allow the animals to be adapted to new environment. The animals were then anesthetized; through a middle-anterior incision on the larynges the vocal folds were exposed and removed under microdisecting microscope. Skin covering 1 cm of the tail tip and about 100 mg of lung tissue were also harvested from each rat. The tissues were immersed immediately in 200 ul to 1000 ul of ice-cold cell disruption solution, RNA STAT-60 (Tel-Test, Inc. Friendswood, TX), snap frozen with liquid nitrogen, and saved in –80 oC until being further processed for extraction of total RNA.

Extraction of Total RNA

Total RNA was prepared from the frozen tissues with guanidine-phenol extraction technique (11) using RNA STAT-60 as the cell disruption solution. The frozen tissues were thawed on ice and then finely homogenized, first with a PRO 200 Homogenizer (Intermountain Scientific Co., Salt Lake City, UT) followed with a glass Tenbroeck Tissue Grinders (VWR Scientific Products, Willard, OH). The following procedures for extracting total RNA from the tissue homogenate were performed according the manufacturer's instructions. The quantity of the total RNA was determined by absorbency at 260 nm, and the quality was confirmed by sharp bands of 18S and 28S rRNA after 2 ug total RNA is separated on a 1% agarose/formaldehyde gel and stained with ethidium bromide.

Reverse-Transcription (RT) of Total RNA to First Strand of Complementary DNA (cDNA)

To eliminate genomic DNA contamination, RNA samples were treated with RNase-free DNase before being reverse-transcribed to first strand cDNA. This was carried out by incubating 1 ug total RNA with 5 units of RNase-free DNase I (Amersham Pharmacia Biotech, Piscataway, NJ) in RT buffer (Boehringer Mannheim, Indianapolis, IN), at 37 °C for 20 minutes. The final concentration of $MgCl_2$ in the mixture was adjusted to 5 mM. After the incubation the samples were heated at 75 °C for 5 min to inactivate the DNase I, followed by cooling down to 4 °C.

The First Strand cDNA Synthesis Kit from Boehringer Mannheim was used for the RT reaction, but a modification of the manufacturer's protocol was adopted to achieve a significant increase in the yield of reverse transcripts. The total volume of the RT reaction was 20 ul and the final concentrations of the reagents in the reaction mixture were RT buffer (1x), $MgCl_2$ (5 mM), dNTP mixture (1 mM for each dNTP), random primers (total 0.08 A_{260} units), RNase inhibitor (2.5 units/ul), and AMV reverse transcriptase (≥2.5 units/ul). After the assembly of the RT mixture, the reactions were incubated in a thermal cycler (GeneAmp PCR System 2400 or 9600, Perkin Elmer, Norwalk, CA) at 25 °C for 10 min, 42 °C for 95 min, 99 °C for 5 min and 4 °C for 5 min, consecutively. The total volume of the reaction was then brought to 400 ul with diethyl pyrocarbonate (DEPC) treated H_2O, aliquoted and saved in –80 °C until use.

Amplification of cDNA With Polymerase Chain Reaction (PCR)

"Hot start PCR" with the application of an anti-Taq DNA polymerase antibody, TaqStart antibody (Clontech Lab. Inc. Palo Alto, CA) were employed. The total volume of the PCR reaction was 25 ul in 1x PCR buffer (10 mM Tris-HCl, 50 mM KCl, pH 8.3). The optimal concentrations of the reagents (especially $MgCl_2$) in the PCR reactions varied with different target genes, but for PCR amplifying most target genes a standard recipe was feasible. The final concentrations of the reagents in this standard recipe were dNTP (0.2 mM for each dNTP), $MgCl_2$ (1.5 mM), forward and reverse primers (0.5 uM for each), 1:1 mixture of Taq polymerase and TaqStart antibody (0.4 ul) and various amount of cDNA equivalent to 1.25 ng to 10 ng of total RNA. After the assembly of the PCR reactions, for most target genes the mixture was incubated in Perkin Elmer thermal cyclers with the following standard protocol: one cycle

of 94 °C for 1 min, followed by 35 cycles of 94 °C for 30 seconds (denaturing), 56 °C for 1 min (annealing) and 72 °C for 2 min (extension), and then by 1 cycle of 72 °C 5 min and cooled to 4 °C.

Experimental conditions for PCR were optimized including reagent concentrations, thermal cycler protocol and amount of cDNA added to the reaction. Table 1 lists the optimal concentrations of $MgCl_2$ and annealing temperature (AT) for the primers that have been used in this work. To help determine the optimal volumes of the cDNA, linearity of the PCR reactions, defined as the linear relationship between the amount of the starting cDNA that was added to the PCR reactions and the amount of the corresponding PCR products, was determined, for each gene and

Table 1.
Information of Primers Used in the Study

Gene Name	Primer Sequences	Size (bp)	Mg[3]	AT[4]	Ref.
Total FN	F[1]:5'-GTTGGCACTGACGAAGAGCC-3' R2: 5'-AAGCCAGAGTCAGATAACCG-3'	272	1.5	55	23
FN (EIIIA)	F: 5'-GTCAGTCCAGATCAAACAGA-3' R: 5'-GTGCTGTCTGGAGAAAGGTT-3'	558/287	1.5	55	43
FN(EIIIB)	F: 5'-TGACATCAGAAGAATCAAAACCAGTT-3' R: 5'-TTACACTGTCAAAGATGACAAGGAAA-3'	640/367	1.5	55	44
FN (V)	F: 5'-ATGAAATGATGTACTCAGAACTCT-3' R: 5'-ATTACTGGCTACATTATCAAGTATGAGAA-3'	619/544/259	1.5	55	44
GAPDH	F: 5'-ACCCCCAATGTATCCGTTGT-3' R: 5'-TACTCCTTGGAGGCCATGTA-3'	299	1.5	56	45
Matrilysin	F: 5'-GGGGACTGCAGACATCATAA-3' R: 5'-ACTTCTGGATGCCTGCAATG-3'	322	1.0	56	45
Stromelysin-1	F: 5'-ATCCGAGGTCATGAAGAGCT-3' R: 5'-TATGTGGGTCACTTTCCCTG-3'	317	1.0	56	45
Stromelysin-2	F: 5'-GTCCAAGCAGGTTACCCAAA-3' R: 5'-TTCAGTGTGTGTGTCACCGT-3'	307	2.0	56	45
Stromelysin-3	F: 5'-TATGACGGTGAGAAGCCAGT-3' R: 5'-TCGAGGAAACTTTCCAGGA-3'	320	1.5	56	45
TIMP-1	F: 5'-GACCTGGTCATAAGGGCTAAA-3' R: 5'-GCCCGTGATGAGAAACTCTTCACT-3'	216	1.5	56	46
TIMP-2	F: 5'-TGCAGCTGCTCCCCGGTGCAC-3' R: 5'-TTATGGGTCCTCGATGTCGAG-3'	590	1.5	56	47
TIMP-3	F: 5'-GTGGTGGGAAAGAAGCTGGTGAA-3' R: 5'-CACTAATTTCATTGTCATCAT-3'	570	1.5	56	48
TIMP-4	F: 5'-GCTCAGTCGCGGATCTGCAGTGTC-3' R: 5'-CTAGGGCTGGACGTGTCAACGTAT-3'	680	1.5	59	49

F: Forward primer
R: Reverse primer
Mg:$MgCl_2$ (mM)
AT: annealing temperature (°C)

each tissue, by adding different amounts of the starting cDNA to the PCR reactions and amplifying all reactions with the same PCR conditions. The densities of the PCR products were plotted against the volumes of the starting cDNA. The amount of starting cDNA was defined as the one that showed a linear trajectory before it reached the maximum plateau.

To quantitate the PCR products, the PCR products were electrophoresed through a 2.0% agarose gel with 0.5 ug/ml ethidium bromide. The density of the PCR products-formed bands were then determined with a Gel Documentation System 760 (GelExpert, Nucleotech Corporation, San Mateo, CA). cDNA for GAPDH was amplified from the same sample and electrophoresed and analyzed in the same ways. The ratios for target gene/GAPDH were determined and defined as the standardized densitometry values of the target genes as presented in the result.

The specificity of the PCR products was carefully verified. To clarify that a PCR product was due to contamination, a negative control reaction, where cDNA was omitted, was always amplified simultaneously with the reaction for the target gene. The fidelity of the PCR amplification was confirmed by directly sequencing the nucleotide sequence of the PCR product and comparing its homology to the native sequence of the target gene, and by visualizing their migration relative to the molecular weight standards (50 bp DNA ladder, GibcoGRL Life Technologies, Gaithersberg, MD)

Primers

The primers, previously designed according to the published gene sequences, were synthesized and crude purified commercially. Table 1 listed the sequences of the primers and their optimal $MgCl_2$ concentrations and annealing temperature (AT) that we have employed.

Data Analysis

Statistical analyses were carried out with GraphPad PRISM (Version 2.0, Intuitive Software for Science, San Diego, CA). Results are expressed as the mean ± S.E. Statistical significance, set at $P<0.05$, was determined using one-way analysis of variance (one-way ANOVA). Each data point represents a minimum of six individual RT-PCR assays.

Results

Splicing Patterns of FN Pre-mRNA in Rat Vocal Folds

EIIIA and EIIIB are single type III domains encoded by single exons either included or excluded during splicing (41). Alternative splicing of the V region generates three possible variants in the rat and five in the human, all of which arise from exon subdivision of the coding sequences (42). FN synthesized in situ, or "cellular FN" typically retains all these domains. The primers for FN amplification were initially selected from exons flanking introns for total FN and from exons spanning each of the alternatively spliced FN exons to obviate the inadvertent amplification of genomic DNA (Table 1). Primers spanning the EIIIA region result in a 526 bp product if this exon is spliced in (EIIIA+) and a 256 bp product if it is spliced out (EIIIA-). Similarly, the 640 bp and 367 bp PCR products correspond to the EIIIB+ and EIIIB- isoforms, respectively. Three different products of 619, 344, and 259 bp are expected with PCR amplification using V region primers. By virtue of alternative splicing and exon subdivision, the 619 bp product contains the complete V region, the 544 bp product retains part of the V region, and the 259 bp fragment lacks V region completely. As illustrated in Table 2, all the above-mentioned splicing patterns are present in rat vocal folds, generating isoforms of EIIIA+/EIIIA-, EIIIB+/EIIIB- and V+(619 bp)/V+(544 bp)/V-. The predominant products from alternative splicing of EIIIA and EIIIB are the spliced-out forms, while the more abundant mRNA species from

Table 2.
Age Associated Differential Alternative Splicing of Fibronectin Precursor mRNA in Rat Vocal Folds, Skin and Lungs

Ratio (%)	Vocal Fold 1-2 w	Vocal Fold 5-6 m	Vocal Fold 24-25 m	Tail Skin 1-2 w	Tail Skin 5-6 m	Tail Skin 24-25 m	Lungs 1-2 w	Lungs 5-6 m	Lungs 24-25 m
EIIIA +/-	13.4+2.1	10.3+2.3	15.3+3.3	34.9+3.2	43.2+3.9	38.7+4.3	26.8+3.1	22.8+3.2	31.0+2.4
EIIIB+/-	24.9+2.1	23.5+1.9	65.8+7.2						
V(619)/-	211.1+19.1	229.4+23.5	234.7+26.1	215.1+22.2	160.0+12.2	10.3+3.9	78.3+9.6	34.3+4.2	36.8+4.5
V(544)/-	242.2+16.8	333.2+23.1	321.2+26.3	275.2+19.1	359.3+36.5	57.9+8.2	116.5+12.1	115.8+7.9	82.5+4.1

alternative splicing of the V region are the spliced-in forms. The patterns of alternative splicing of regions EIIIA and V in the lungs and the tail skin are identical to those in the vocal folds (Table 2). However, in the lungs and the tail skin splicing of EIIIB generates only the spliced-out form mRNA species, visualized as a 240 bp band on agarose gel (Table 2).

Age-Associated Differential Alternative Splicing of FN Pre-mRNA in Rat Vocal Folds

Ratios of the mRNA species that are generated from alternative splicing of the FN pre-mRNA regions are used as the index of the relative abundance of the isoforms (26,7,33). Quantitatively, no new mRNA species, nor missing existing mRNA species was detected in all the three studied organs and in the different age groups. Quantitatively, the ratios of EIIIA+/EIIIA- in all three organs do not show significant age-dependent changes (Table 2). The ratio of EIIIB+/EIIIB- in the vocal folds of the adult rats is not statistically different from the ratio of the neonatal animals, but it is significantly increased by almost two times in the elderly rats (Table 2). As compared to the neonatal rats, the V+(619 bp)/V- ratios in the adult animals show either unchanged (vocal folds) or slightly but significantly decreased (skin and lungs), while the V+(544 bp)/V- ratios are either significantly increased (vocal folds and skin) or remain unchanged (lungs). In the elderly rats, both V+(619 bp)/V- ratio and V+(544 bp)/V- ratio remain unaltered in the vocal folds but significantly decreased in both skin and lungs, in comparison to the adult rats (Table 2).

Senescent Alterations of mRNA Levels of Total FN and FN Isoforms in Rat Vocal Folds

To determine the mRNA levels for total FN, primers flanking introns for total FN were used and amplification with these primers results in an expected product of 272 bp (47). For measuring total mRNA levels of the isoforms generated from splicing of EIIIA, EIIIB or V region, the levels of the mRNA species that are products of the same region were summed up and interpreted as the total mRNA level of that region. Messenger RNA levels of total FN in the vocal folds of the adult rats is not different from the neonates, but significantly elevate, by 25.24%, in the elderly rats ($P<0.05$) (Figure 1). Age-associated fluctuation in EIIIA mRNA levels is similar to total FN, i.e., there is no difference between the neonatal and the adult levels, but the levels in the elderly rats significantly increase (Figure 1). The EIIIB mRNA levels display age-related changes different from EIIIA region and total FN; it decreases in the adult rats by 31% versus the neonatal rats and significantly increases back in the elderly rats to a level that is 24% higher than the neonatal level ($P<0.05$). The mRNA levels for all the three V region-derived isoforms in the vocal folds do not change with age significantly (Figure 1).

Gene Expression of Fibronectin-Degrading Enzymes and Their Tissue Inhibitors in Vocal Folds of Rats With Different Age

Stromelysins 1-3 and matrilysin are the major matrix metalloproteinases for degradation of fibronectin core proteins (31,45). These FN-degrading proteinases are abundantly expressed in rat vocal folds (Figure 2) and their expressions display age-related fluctuations (Figure 3B). The peak levels of gene expression of FN-degrading enzymes are detected in the vocal folds of the neonatal rats; in adult rats the levels decline to 53% to 68% of those in neonates ($P<0.05$) but do not show further significant in the 24-25 month old rats (Figure 2). Messenger RNA levels for TIMPs 1-4 remain relatively unchanged during the life, i.e., there is no difference in gene expression for TIMPs 1-4 among the three groups of rats, with the only exception that TIMP-2 mRNA levels in the elderly rats is decreased to 58% of the neonatal level (Figure 3; following page).

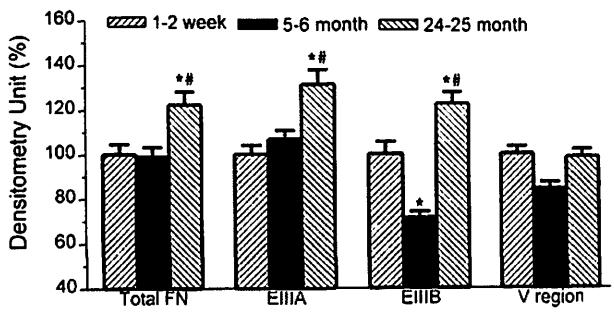


*Figure 1. Senescent alterations of mRNA levels of the FN and FN isoforms in rat vocal folds. *P<0.05 vs. the levels of rats of 1-2 weeks; #P,0.05 vs. the levels in rats of 5-6 months.*

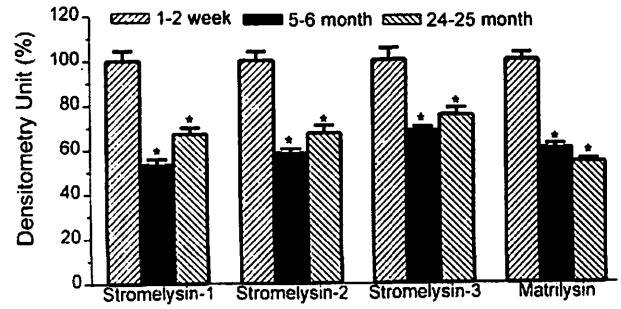


*Figure 2. Expression of genes for fibronectin-degrading proteinases in the vocal folds of rats with different ages. *P<0.05 vs. the levels of rats of 1-2 weeks.*

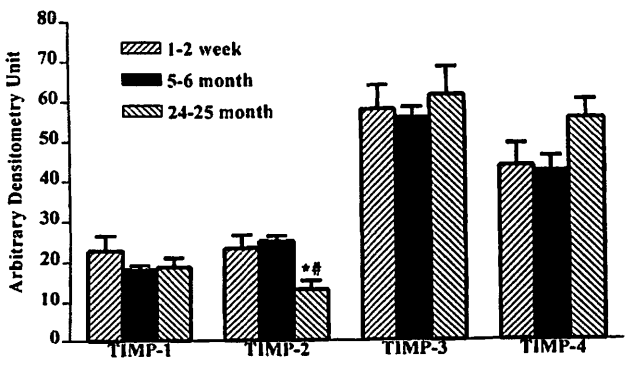


*Figure 3. Expression of genes for tissue inhibitors of metalloproteinases (TIMPs) in the vocal folds of rats with different ages. *P<0.05 vs. the levels of rats of 1-2 weeks; #P<0.05 vs. the level in rats of 5-6 months.*

Discussion

Alternative Splicing of FN Gene Transcript in the Vocal Folds, Tail Skin and Lungs

The present study is the first report on the alternative splicing patterns of FN pre-mRNA in rat vocal folds and skin and reveals that alternative splicing of the FN pre-mRNA in the vocal folds results in isofoms that are EIIIA+/EIIIA-, EIIIB+/EIIIB-, and V+/V-, while in the tail skin, we found EIIIA+/EIIIA-, EIIIB-, and V+/V- (Table 2). We also confirmed that FN isoforms of EIIIA+/EIIIA-, EIIIB-, and V+/V- are the products of the alternative splicing of FN gene transcript in rat lungs, as reported previously (26,33). Alternative splicing of EIIIA, EIIIB and V regions of the FN pre-mRNA generate collectively12 different FN isoforms in the rats (37,24,40), the splicing patterns, however, differ in individual tissues. For example, in rat skeletal muscle are found EIIIA-, EIIIB- and V-; brain, heart and kidney are EIIIA+/EIIIA-, EIIIB+/EIIIB-, and V+/V-; while spleen, liver, and lung are completely EIIIB- (or EIIIB+/EIIIB- ratios <0.01), but EIIIA+/EIIIA- and V+/V- (26,33). The functional significance of the tissue specificity in the alternative splicing of FN pre-mRNA is not fully elucidated. Available information indicate that each of the protein peptides resulted from alternative splicing of these special regions has unique physiological functions or pathogenic involvement (21,30,5,16). Therefore, tissue-specific patterns in the alternative splicing of the FN gene transcript enable individual tissues to produce different FN isoforms that may function in tissue-dependent ways. It is not clear as to what are the mechanisms that are responsible for the tissue specificity in alternative splicing of FN pre-mRNA, although a number of factors or molecules are known to participate in the regulation of the alternative splicing process (for review, see 24). It is likely that to-be-discovered tissue specific factors are involved in tissue specific manipulation of FN pre-mRNA splicing.

Age-Associated Alterations in Expression of FN Isoforms in the Vocal Folds

In this study, neither missing of existing isofoms nor appearing of new isoforms was observed in the vocal folds as well as the other tissues examined, the integrity of the alternative splicing machinery appears to be preserved during senescence process. There are, however, senescent alterations in relative expression of FN isoforms derived from the splicing of the same pre-mRNA region, but the changes are diverse in both tissues and in pre-mRNA regions. The ratios of EIIIA+/EIIIA- in the vocal folds, skin and lungs show no significant age-dependent changes. In vocal folds the ratio of EIIIB+/EIIIB- is not different between the neonates and the adults, but is over doubled in the elderly rats, indicating significantly increased expression of the EIIIB+ isoform (Table 2). The changes in the ratios of V+/V- vary diversely among the tissues (see the Results) but is not significant in the vocal folds. Previous studies (35,10,44) suggest that the structure, function, and molecular size of fibronectin are altered as a consequence of aging and there is statistically significant changes in splicing with aging in several rat tissues (26,7,33), indicating the possibility that FNs participate in tissue or organ aging by variations in isoform shifting. It is important to recognize that, as evidenced by this study as well as previous (26,7,33), age-related (or development-related) FN isoform shifting shows marked and diverse tissue variation. This implies that some FN isofoms may play more important roles in senescence of individual tissues than the others. Our data highlights the importance of the EIIIB+ isoform in vocal aging since this isoform is the only one that is exclusively increased in the aged vocal folds (Table 2). EIIIB exon coded peptide has an amino acid sequence that is more conserved in evolution than any other type III repeat (36), which strongly suggests a biological function for this FN isoform. However, so far there are no clues as to a specific role for EIIIB+ isoform. It is interesting that the alternative splicing of EIIIB is found under codevelopmental control, i.e., the EIIIB+ isoform is minimally expressed in tissues other than fetal or tumor or wounded tissues (8,32,15). It remains to be explored as to why the EIIIB+ isoform is upregulated in aged vocal folds. We speculate that this might be due to the physical or chemical stimulations that apply to vocal folds and TGF-b1 is likely one of the tissue mediator (4).

Age-Related Changes in Total FN and Total FN Isoforms Derived From Different Region Splicing in the Vocal Folds

Histochemical studies demonstrated the presence of unique deposition of fibronectin proteins, regardless of the identity of the isoforms, in different benign laryngeal lesions (18,12). To elucidate possible participation of FNs in senescent voice alterations, it is important to determine if

mRNA levels of total FN and total isoforms derived form each individual variable FN pre-mRNA regions change with age in the vocal folds. The data indicate that in the vocal folds, expression of total FN, EIIIA and EIIIB regions significantly elevate in the elderly rats as compared to the neonatal and the adult rats, but the expression of the V region is unchanged in this age group. Upregulation of total FN and regions of EIIIA and EIIIB in the aged vocal folds is unique and tissue specific because in a separate study, we found that there is age-dependent decrease in gene expression for collagenous and elastin proteins in rat vocal folds (13).

It is noteworthy that composition and levels of extracellular matrix (ECM) components are the reflection of the balance between the biosynthesis and degradation of ECM proteins. ECM turnover, either physiological or pathological, is in most cases a highly organized process that involves the selective action of a group of zinc- and calcium-dependent proteases, namely matrix metalloproteinases (MMPs) or matrixin protease family (31,45). In addition, in vivo activities of MMPs are controlled at several levels including their interactions with specific naturally occurring inhibitors, e.g., the tissue inhibitors of metalloproteinases (TIMPs) (31). TIMPs are cell-secreted nonspecific inhibitors that act as negative regulators of MMPs; four TIMPs have been cloned and well characterized and designated as TIMP-1, 2, 3, and 4, respectively. An imbalance between MMPs and TIMPs results in either metalloproteinase activation or suppression, and in turn, determines the rate of matrix accumulation and degradation (add reference). The core protein portion of fibronectins is degraded mainly by stromelysins (stromelysin 1, 2, and 3) and matrilysin (31,45). The data indicate that gene expression of fibronectin-degrading enzymes in the vocal folds appears to adopt a pattern different from the one of TIMPs. In comparison to neonatal rats, vocal expression of the fibronectin-degrading proteinases significantly diminishes in adult rats. After adulthood, however, mRNA levels for the proteinases do not show further decrease in the elderly rats (Figure 2). On the other hand, gene expression of TIMPs in the vocal folds remain at stable levels at different ages with only one exception where mRNA level of TIMP-2 in the elderly rats significantly decreased to 58% of the neonatal levels (Figure 3). Combination of enhanced expression of fibronectin and relatively stabilized expression of FN-degrading proteinases and TIMPs in aged vocal folds suggest a possible slowdown of FN degradation and a possible accumulation of FN proteins in aged vocal tissue.

Clinically it is well observed that the stiffness of vocal fold lamina propria increase with age, which may partially be responsible for the voice alterations in the elderly (43). Histology data indicate that senescent alterations in ultrastructures of vocal fold collagen and elastin fibers such as fiber alignment and arrangement, and in distribution and deposit of the fibers in the different layers of the vocal folds may be among the possible mechanisms (38,23). One such abnormal structural change is that, like their counterparts in other organs, the collagen and elastin fibers in the vocal folds may become increasingly cross-linked with age, rendering the fibers the resistance to digestive enzymes as well as to deformation and stiffness (27,28,3,1). The mechanisms underlying the enhanced cross-linking of the collagen and elastin fibers in aged tissues are multiple and still not fully understood (2). It is interesting that fibronectin has multi-binding domains through which FN can bind to receptors and other macromolecules including collagens. It is speculated that the enhanced accumulation of fibronectin molecules in the aged vocal folds may cause increased cross-binding between FNs and collagen molecules in the vocal folds, and in turn, the stereo conformation of the collagen fibers.

In summary, the present study reveals that alternative splicing of EIIIA, EIIIB and V regions of the FN pre-mRNA in the vocal folds generates EIIIA+/-, EIIIB+/-, and V+/- FN isoforms, while in the tail skin, isoforms of EIIIA+/-, EIIIB- and V+/-. There is age-associated and tissue specific alterations in the relative abundance of the isoforms generated from splicing of the same variable FN pre-mRNA region, and in the vocal folds, these changes pinpoint to an enhanced expression of EIIIB+ isoform in the elderly rats. In the aged vocal folds, mRNA levels for total FN and total EIIIA and EIIIB derived FN isofoms significantly rise, but levels of mRNAs for FN-degrading proteinase and inhibitors remain unchanged. These changes suggest a possible accumulation of FN molecules in the aged vocal folds.

References

1. Alnaqeeb MA, NS Al Zaid, G Goldspink. Connective tissue changes and physical properties of developing and aging skeletal muscle. *J Anatomy.* 139:677-689, 1984.

2. Bailey AJ, RG Paul, L Knott. Mechanisms of maturation and aging of collagen. Mechanisms *Aging Devel.* 106(1-2):1-56, 1998.

3. Betsch DG, E Baer. Structure and mechanical properties of rat tail tendon. *Biorheology* 17:83-94, 1980.

4. Borsi L, A Castellani, AM Risso, et al. Transforming growth factor-b regulates the splicing pattern of fibronectin messenger RNA presursor. *FEBS Lett.* 261:175-178, 1990.

5. Burton-Wurster N, HR Leipold, G Lust. Dibutyryl cyclic AMP decreases expressi8on of ED-A fibronectin by canine chondrocytes. *Biochem Biophys Res Cmmun.* 154:1088-1093, 1988.

6. Cai X, CS Foster, JJ Liu, et al. Alternatively spliced fibronectin molecules in the wounded cornea: Analysis by PCR. *Invest Ophthalomol Vis Sci.* 34:3585-3592, 1993.

7. Caputi M, FE Baralle, CA Melo. Alalysis of the linkage between fibronetin alternative apliced sites during ageing in rat tissues. *Biocimica et Biophysica Acta.* 1263:53-59, 1995.

8. Carnemolla B, E Balza, A Siri, et al. A tumor-associated fibronectin isoform generated by alternative splicing of messenger RNA precursors. *J Cell Biol.* 108:1139-1148, 1989.

9. Chan R. PhD dissertation (University of Iowa) 1998.

10. Chandrasekhar S, E Norton, AJT Millis, et al. Functional changes in celular fibronectin from late passage fibroblasts in vitro. *Cell Biol Int Rep.* 7:11-21, 1983.

11. Chomczynski P, N Sacchi. Single-step method of RNA isolation by acid Guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* 162:156-159, 1987.

12. Courey MS, JA Shohet, MA Scott, et al. Immunohistochemical characterization of benign laryngeal lesions. *Ann Otol Rhinol Laryngol* 105:525-531, 1996.

13. Ding H, SD Gray. Senescent expression of genes coding collagens, collagen-degrading metalloproteinases and tissue inhibitors of metalloproteinases in rat vocal folds: Comparison with skin and lungs. In review.

14. Espaillar A, SJ Lee, V Arrunategui-Correa, et al. Expression of fibronectin isoforms in rat cornea after an epithelial scrape wound. *Eye Res.* 13:325-330, 1994.

15. French-Constant C, L Van De Water, HR Dvorak, et al. Reappearance of an embryonic pattern of fibronectin splicing during would healing in the adult rat. *J Cell Biol.* 109:903-914, 1989.

16. Glukhova MA, MG Frid, BV Shekhonin, et al. Expression of extra domain A fibronectin sequence in vascular smooth muscle cells is phenotype dependent. *J Cell Biol.* 109:357-366, 1989.

17. Wells, GMA, C Graham, JA Cossings, et al. Quantitation of matrix metalloproteinases in cultured rat astrocytes using the polymerase chain reaction with a multi-competitor cDNA standard. *Glia* 18:332-340, 1996.

18. Gray SD, E Hammond, D Hanson. Benign pathologic responses of the larynx. *Ann Otol Rhinol Laryngol.* 104:13-18, 1995.

19. Hagen P, GD Lyons, DW Nuss. Dysphonia in the elderly: Diagnosis and managements of age-related voice changes. *South Med J.* 89(2):204-207, 1996.

20. Hirano H. Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniat.* 26:89-94, 1974.

21. Humphries MJ, A Komoriya, SK Akiyama, et al. Identification of two distinct regions of the type III connecting segment of human plasma fibronectin that promote cell type-specific adhesion. *J Biol Chem.*262:6886-6889, 1987.

22. Hynes R. Molecular biology of fibronectin. *Ann Rev Cell Biol.* 1:67-90, 1985.

23. Ishii K, WG Zhai, M Akita, et al. Ultrastructure of the lamina propria of the human vocal fold. *Acta Oto-Laryngol.* 116(5):778-782, 1996.

24. Kornblihtt AR, CG Pesce, CR Alonso, et al. The fibronectin gene as a model for splicing and transcription studies. *FASEB J.* 10:248-257, 1996.

25. Kurita S. Layer structure of the human vocal fold: morphological investigation. *Otologia (Fuoka).* 26:973-997, 1980.

26. Magnuson VL, M Young, DG Schattenber, et al. The alternative splicing of fibronectin pre-mRNA is altered during aging and in response to growth factors. *J Biol Chem.* 266:14654-14662, 1991.

27. Miyahara T, A Murai, T Tanaka, et al. Age-related differences in human skin collagen: solubility in solvent, susceptibility to pepsin digestion, and the spectrum of the solubilized polymeric collagen molecules. *J Geromtol.* 37:651-655, 1982.

28. Mohan S, E Radha. Age-related changes in rat muscle collagen. *Gerontololgy.* 26:61-67, 1980.

29. Morrison MD, P Gore-Hickman. Voice disorders in the elderly. *J Otolarynol. 15*(4):231-234, 1986.

30. Mould AP, A Komoriya, KM Yamada, et al. The CS5 peptide is a second site in the IIICS region of fibronectin recognized by the integrin alpha 4 beta 1. Inhibition of alpha 4 beta 1 function by RGD peptide homologues. *J Biol Chem.* 266:3579-3585, 1991.

31. Murphy G, AJP Docherty. The matrix metalloproteinases and their inhibitors. *Am J Respir Cell Mol Biol.* 7:120-125, 1992.

32. Oyama F, S Hirohashi, Y Shimosato, et al. Oncodevelopmental regulation of the alternative splicing of fibronectin pre-messenger RNA in human lung tissues. *Cancer Res.* 50:1075-1078, 1990.

33. Pagani F, L Zagato, C Vergani, et al. Tissue-specific splicing pattern of fibronectin messenger RNA precursor during development and aging in rat. *J Cell Biol* 113(5):1223-1229, 1991.

34. Paul JI, RO Hynes. Multiple fibronectin subunits and their post-translational modifications. *J Biol Chem.* 259:13477-13487, 1984.

35. Porter MB, OM Pereira-Smith, JR Smith. Novel monoclonal antibodies identify antigenic determinants unique to cellular senescence. *J Cell Physiol.* 142:425-433, 1990.

36. Romberger DJ. Fibronectin. *Int J Biochem Cell Biol.* 29(7):939-943, 1997.

37. Ruoslahiti E. Fibronectin and its receptors. *Ann Rev Biochem.* 57:375-413, 1988.

38. Sato K, M Hirano. Age-related changes of elastic fibers in the superficialy layer of the lamina propria of vocal folds. *Ann Otol Rhinol Laryngol.* 106(1):44-48, 1997.

39. Sava G, I Capozzi, A Bergamo, et al. Down-regulation of tumour gelatinase/inhibitor balance and preservation of tumour endotrhelium by an anti-metastatic ruthenium complex. *Int J Cancer.* 68:60-66, 1996.

40. Schwarzbauer JE. Alternative splicing of fibronetin-three variants, three functions, *Bioessays.* 13:527-533, 1991.

41. Schwarzbauer JE, RS Patel, D Fonda, et al. Multiple sites of alternative splicing of the rat fibronectin gene transcript. *EMBO J.* 6:2573-2580, 1987.

42. Schwarzbauer JE, JW Tamkun, IR Lemischika, et al. Three different fibronectin mRNAs arise by alternative splicing within the coding region. *Cell.* 35:421-431, 1983.

43. Sinard RJ, D Hall. The aging voice: How to differntiate disease from normal changes. *Geriatrics.* 53(7):76-79, 1998.

44. Sorrentno JA, AJT Millis. Structural comparisons of fibronectins isolated from early and late passage cells. *Mech Ageing Dev.* 28:83-97, 1984.

45. Stetler-Stevenson WG. Dynamics of matrix turnover during pathologic remodeling of the extracellular matrix. *Am J Phthol.* 148(5):1345-1350, 1996.

46. Stetler-Stevenson WG, PD Brown, M Onisto, et al. Tissue inhibitor of metalloproteinases-2 (TIMP-2) mRNA expression in tumor cell lines and human tumor tissues. *J Biol Chem.* 265(23):13933-13938, 1990.

47. Vitale AT, M Pedroza-Seres, V Arrunategui-Correa, et al. Differential expression of alternative apliced fibronectin in normal and wounded rat corneal stroma versus epithelium. *Invest Ophthalmol Vis Sci.* 35:2664-3672, 1994.

48. Wu IM, Moses MA. Cloning and expression of the cDNA encoding rat tissue inhibitor of metalloproteinases 3 (TIMP-3). *Gene.* 168:243-246, 1996.

49. Wu IM, MA Moses. Molecular cloning and expression analysis of the cDNA encoding rat tissue inhibitor of metalloproteinases-4. *Matrix Biology.* 16:339-342, 1998.

NCVS Status and Progress Report - 14
September 1999, 89-99

Advances and Refinements in Phonosurgery

Charles Ford, M.D.
Division of Otolaryngology–Head and Neck Surgery, The University of Wisconsin

Abstract

Scientific discovery, technological advances, and improved outcomes assessment have resulted in advances and refinements in phonosurgery. Three areas of substantial evolution are phonomicrosurgery, laryngeal framework surgery, and the use of implantable materials in vocal folds. Discovery of the importance of the superficial layers of the lamina propria has led to increased use of more limited medial microflap approaches and less frequent use of the classic lateral cordotomy flap approach. Alternative approaches to managing vocal fold scarring defects have addressed the separation of body and cover and provided suitable lamina propria replacement. Approaches to sulcus vocalis have been refined to address type II (linear vergeture) and type III (focal invasive pit) sulcus, where there is loss of lamina propria, while still recognizing the common nonpathological type I (physiological) sulcus. Technological advancements such as photodynamic therapy, tuned dye lasers, and laryngeal microdebridement have augmented the armamentarium for mechanical removal of laryngeal papillomata. Careful infusion-assisted microexcision and adjunctive medical management has been refined and made more effective. Laryngeal framework surgery has embraced the development of silastic, hydroxylapatite, expanded polytetrafluoroethylene, and titanium shims. Anatomical studies have helped to improve operative precision and safety, and have led to inventive variations in arytenoid repositioning that improve closure of the posterior subunit. Vocal fold augmentation by injection has been facilitated by innovative use of the rigid telescope and intraoperative videostroboscopy. Anatomical studies have focused on the infrafold region and rheological studies have attempted to match viscoelastic properties of injectable substances to those of vocal fold tissues. Alloplastic materials like Teflon have been largely supplanted by newer bioimplantables such as fat, collagen, and fascia.

Introduction

The primary goal of phonosurgery is to restore and improve voice. Advances in phonosurgery have been built on the established principles of exposure and technique that formed the foundation for current phonosurgical techniques. While some current techniques are the result of empiricism and anecdotal adventures, recent advances are more likely the result of careful basic and applied science. Disciplined clinical research has been enhanced by voice assessment tools that allow clinicians to study the acoustic and aerodynamic parameters of voice production and to comprehensively study the glottic waveform. Outcomes analysis has also given clinicians a better appreciation of the impact of voice disorders and the efficacy of surgical interventions.

Phonomicrosurgery

Phonomicrosurgery is the application of the microscope and micro-instrumentation in endoscopic laryngeal surgery. Recent advances might be considered refinements of the important, visionary contributions of microlaryngoscopy pioneers. The concept of phonosurgery was introduced by Albrecht, whose use of a gynecologic colposcope for microlaryngoscopy was met with skepticism by the otolaryngological community. Von Leden (1) recalls this event and his personal interactions with Lewy, Kleinsasser, Jako, Lynch, and others who catalyzed the evolution of laryngoscopes and optics. These seminal advances, which allowed the surgeon magnified vision while performing bimanual surgery through a laryngoscope, have facilitated the continued development of phonomicrosurgical tools and techniques. Hirano (2) recently emphasized the parallel advances in basic anatomy and physiology of phonation, pathophysiology of vocal fold diseases, and vocal function evaluation. Continued refinement of techniques is

dependent on accurate, objective voice assessment to precisely diagnose disorders and measure the results of surgical intervention.

Many of the technical advances in phonomicrosurgery involve improved visualization of the pathology and operative field. The keys to any direct laryngoscopic examination have not changed over the years. These principles were elegantly summarized by Zeitels and Vaughan (3): (1) position the patient supine with the neck flexed on the chest and extended at the occipito-atloid joint, (2) apply external counter pressure (usually with tape) to direct the laryngeal framework posteriorly, (3) conform the laryngoscope to the triangular shape of the glottis, (4) provide independent laryngeal suspension so the surgeon has both hands free, and (5) induce internal laryngeal distension to move the supraglottal tissues away from the true vocal folds. Various laryngoscopes of different size and shape have been designed to achieve improved exposure. Kantor and Berci (4) introduced a novel approach to avoid the optical disadvantage imposed by looking through the laryngoscope. In their system, a rigid telescope with an attached camera is passed through a separate channel within the laryngoscope. The surgeon operates while viewing a magnified, high-resolution image that fills the entire field. Yanagisawa et al (5) compared this approach to conventional microlaryngoscopy, noting its many advantages but also demonstrating the slight distortions introduced by the earliest version of this instrument. An alternative use of rigid endoscopy allows the surgeon to use 0, 30, and 70-degree telescopes through any laryngoscope to enhance visualization in difficult exposure cases (6), but this requires an assistant if the surgeon is to not be limited to a single-handed approach. Rigid and flexible endoscopes can also be used in the clinic setting for simple microlaryngeal phonosurgical procedures, injections, and biopsies (7,8).

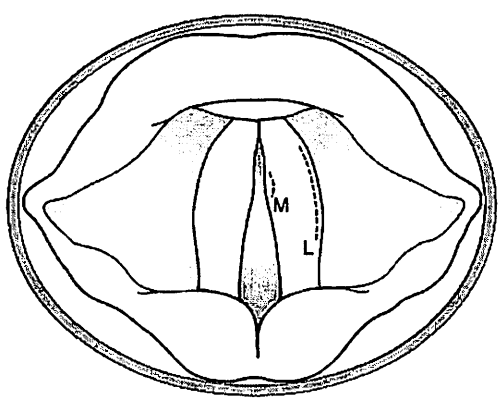


Figure 1. This diagram shows the fundamental difference between the lateral (L) and medial (M) microflap. The medial incision is limited by the approximate dimensions of the lesion.

Incisions, Microflaps, and Planes of Dissection

Many fundamentals of surgery apply to phonomicrosurgery for benign disease. Adequate exposure, careful tissue handling, hemostasis, preservation of normal tissues, and primary healing are necessary to preserve vocal fold function. As we have learned more about vocal fold structure and function, efforts to preserve the cellular and amorphous components of the vocal fold lamina propria have influenced phonosurgical refinements. The classical lateral microflap cordotomy approach (**Figure 1**) for removing benign vocal fold lesions with microinstrumentation involves making a knife incision on the superior-lateral surface of the vocal fold, and often extending this incision with microscissors or a spreading technique. This relatively long incision facilitates blunt dissection in the superficial lamina propria plane and allows easy identification of the vocal ligament. Although initially described as the exclusive microsurgical approach, it is now preferred only in cases where the lesion is large or diffuse, or where identification of the vocal ligament might be problematic (9). For example, in the scarred vocal fold with a poorly defined lesion, it is easier to first identify the vocal ligament laterally and then proceed to dissect around the lesion.

Sataloff (10) argues that the "mini-microflap" approach minimizes damage to the complex basement membrane attachments to the superficial lamina propria, and reports improved results with this approach. The essential difference with this approach is in the placement of the incision near the free edge of the vocal fold (**Figure 1**). Unlike the lateral incision, it is relatively short with anterior-to-posterior extension only as long as necessary to expose the lesion. Such an approach minimizes tissue damage and the risk of postoperative scarring.

Table I.
Phonomicrosurgical Approach to Benign Vocal Fold Pathology

BENIGN PATHOLOGY	MEDIAL MF	LATERAL FLAP	EXCISION ABLATION
Small, medial, focal	+	-	-
Thin mucosa, superficial	+	-	-
Large, unclear margins	-	+	-
Scarred, attached vocal ligament vocal ligament	-	+	-
Focal papilloma	+/-	-	+
Clinical leukoplakia	-	-	+
Mucosal bridge	-	-	+

Courey (11) distinguishes the medial microflap approach from the lateral microflap approach. The medial microflap is ideally suited for situations where: (1) the lesion involves only the medial surface of the vocal fold, (2) the overlying cover is very thin, (3) there is redundant cover, or (4) it appears easy to separate the lesion from the vocal ligament. With the medial microflap approach, it is possible to use a more focal incision, separate the pathological tissue, and spare normal tissue (**Table I**). The key concept here is to spare not only the vocal ligament, but also avoid damaging uninvolved superficial lamina propria. The superficial lamina propria regenerates poorly, and stiffness will result from re-epithelialization over deficient tissue.

Concern about the integrity of the superficial lamina propria is particularly relevant to current methods for microsurgical management of Reinke's edema. It is uncommon for the overlying epithelium to exhibit substantial keratotic changes or atypia, so surgery for Reinke's edema has focused on removal of the pathological tissue with suction while sparing nonredundant epithelial covering tissues. Zeitels (12) recently stressed the importance of preserving as much lamina propria as possible to prevent the epithelial basement membrane from adhering to the vocal ligament. Such adherence decreases oscillation and results in a stiffer vocal fold due to iatrogenic scarring. This phenomenon may be responsible for some prolonged postoperative dysphonias in patients with Reinke's edema. He also demonstrated that cold instrumentation is better suited than CO_2 lasers for resection of small superficial vocal fold lesions (13). In earlier work, Lumpkin (14) also noted that, when compared to laser surgery, cold lateral microflap surgery led to quicker healing and rehabilitation in patients with Reinke's edema. With increased understanding of the cellular and acellular components of the superficial lamina propria, it appears the explanation for this lies in the extent of lamina propria destruction. Gray (15) stressed the importance of the superficial lamina propria in preserving optimal viscoelasticity for vocal fold oscillation. Surgical damage to the lamina propria causes the normal beneficial proteins, elastins, and hyaluronic acids to be replaced with collagen fibers and fibronectin. This increases tissue viscosity, which means that greater energy is required to overcome frictional forces to permit oscillation. In phonosurgery, it is therefore obviously expedient to preserve the superficial lamina propria. When lamina propria is lost, correction should consider replacement with substances that restore freedom of motion and decrease tissue viscosity.

Scars and Sulcus Vocalis: Concepts and Surgery

There are currently several effective alternatives for management of scarred vocal folds. The traditional options have consisted of injections of fat or collagen, incisions to free up the mucosa or break up linear contractures, and excision or ablation of the scarred tissue. Alternative approaches involve inserting a cellular tissue layer between the epithelium and the underlying vocal ligament. Collagen injections appear to soften scar tissue and facilitate augmentation of scar bands and focal defects. Fat implantation is one way of introducing a cellular material to separate layers and restore a functional body and cover mechanism for oscillation. We are currently achieving success utilizing topical Mitomycin-C in treating laryngeal scars. This antibiotic with antineoplastic activity inhibits fibroblastic growth and appears to prevent scar tissue proliferation in cases where we have ablated subglottic scar with radial CO_2 laser incisions. Long-term results, safety, and overall efficacy of this approach await definitive clinical trials.

Sulcus vocalis is a characteristic condition in which normal lamina propria is lost and replaced with scar tissue. Hirano (16) called attention to the histopathology of sulcus problems and later Sato and Hirano (17) looked at the ultrastructure of these lesions. They noted numerous disturbances of the basement membrane and lamina propria, especially at the bottom of the sulcus. These disturbances included fragmentation of elastic fibers, and alteration of the quantity and quality of collagenous and elastic fibers.

Because sulcus vocalis is a descriptive term often applied to a wide variety of conditions, there has been some confusion over the true definition of the disorder. We have seen histological evidence of sulci on routine whole mounted

Table II.
Sulcus Vocalis: Nomenclature and Classification

SULCUS VOCALIS	TYPE I	TYPE II	TYPE III
Other terms	Pseudosulcus	Sulcus vergeture	Ruptured cyst
Dysphonia	Variable to normal	Moderate	Severe
Videostroboscopy	Variable to normal	Focal stiffness	Stiff, no wave
Supeficial lamina propria	Intact	Involved/lost	Involved/lost
Vocal ligament	Normal	Normal or attached	Invaded/lost
Vocalis muscle	Possibly atrophic	Normal	Involved +/-
******			
Lateral flap, incisions	No	Yes + augment	No
Excision	No	No	Yes + augment

larynges studied for cancerous lesions (18) and have observed sulcus-like depressions in the vocal folds of normal volunteer subjects, senile larynges, and patients with vocal fold paralysis. Ford et al (19) distinguished true pathological sulci from the numerous physiological events that cause subtle furrows along the medial edge of vocal folds (**Table II-previous page**). Physiological sulci (*type I*) have a normal lamina propria and minimal disturbance of mucosal wave. Pathological sulci have loss of the superficial lamina propria with a linear depression or vergeture (*type II*) or a penetrating pit extending into the vocal ligament or thyroarytenoid muscle (*type III*) (**Figure 2**). The disturbance of the lamina propria in the pathological types of sulcus vocalis results in stiffness, decreased mucosal wave, and marked hoarseness. Type II sulcus vocalis responds to lateral cordotomy with undermining to free up the epithelium, as described by Bouchayer (20). The response is often dramatic when the linear scar is broken up with multiple superior-to-inferior vertical incisions, as described by Pontes (21). When undermining, it is important to maintain the integrity of the flap to a level at least 3 mm inferior to the sulcus. The cuts should be of differing lengths for maximal effect (19). Type III sulcus vocalis is often associated with inflammation, cysts, mucosal bridges, and extensive penetrating scar tissue that must be excised for optimal rehabilitation. Complete dissection often results in a loss of tissue, thereby making subsequent measures such as implantation, injection, or thyroplasty necessary. Important adjunctive treatment includes voice therapy and vigorous control of laryngopharyngeal reflux. With all such cases of phonosurgery for scars and sulci, rehabilitation is often slow and patients should be informed that it might be 6 to 12 months before maximal improvement is noted.

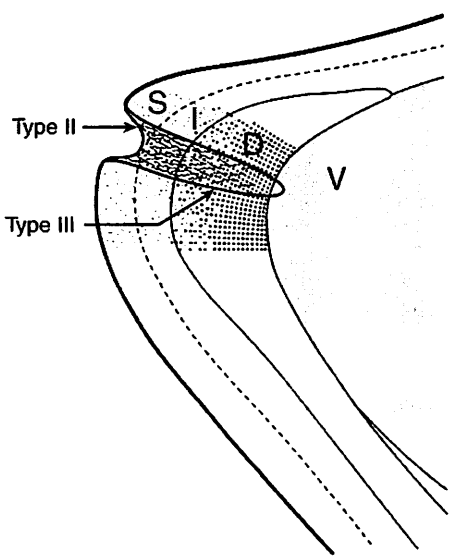


Figure 2. The 3 types of sulcus are diagramatically displayed to demonstrate the depth of involvement of the lamina propria. In type I, the lamina propria is not usually disturbed, so it is not depicted. In type II, there is loss of superficial lamina propria along the length of the vocal fold; these are typically linear defects. In type III, there is a focal pit often extending through the deepest layer of the lamina propria.

Management of Laryngeal Papillomatosis

Recurrent respiratory papillomatosis (RRP) remains a major therapeutic problem although there have been medical and phonosurgical advances addressing this disease. While RRP is a benign superficial laryngeal lesion that causes hoarseness, it is significant because it is the most common benign neoplasm of the larynx and has the potential to spread throughout the respiratory tract. Human papilloma virus has been identified as the cause, and types 11 and 16 appear associated with the most aggressive acting lesions (22). Unfortunately, antiviral medical treatments aimed at controlling the papilloma virus or associated viruses have not proven sufficiently effective to supplant surgery as the primary mode of therapy. There are numerous ongoing studies that hold promise of improving RRP management. These include the antiviral agents acyclovir and cidofovir (23), dietary indoles derived from cruciferous vegetables (indole-3-carbinol and diindolelylmethane) that alter estrogen metabolism (24), and immunomodulating agents such as interferon and ribavirin.

The current standard of treatment for laryngeal papillomatosis remains CO_2 laser ablation. Although clearly superior to earlier techniques of monocular avulsion and some microsurgical techniques, these procedures are far from ideal. Among the laser-induced injuries to the larynx that have been reported are anterior glottic webbing, destruction of the lamina propria with scarring, interarytenoid fixation, and glottic stenosis (25). Technical errors can lead to these types of direct tissue damage; there is also the risk of airway fires and other problems arising from a breach of laser safety protocols. So while there are continuing advances in laser technology, there has also been increasing interest in reducing the reliance on lasers for RRP treatment.

Shikowitz et al (26) have continued their work with photodynamic therapy using the photosensitizing agent dihematoporphyrin-ether (DHE), which is selectively retained in hyperplastic and neoplastic tissues like RRP. When activated by light at a specific wavelength (630 nm), DHE produces cytotoxic agents that destroy the tissues. In clinical trials, they established an optimal light dose (50 J) and drug dose of 4.25 mg/kg DHE. They found a 50% or greater reduction in the recurrence rate during the first year. Interestingly, the treatment, when compared to controls, appeared to have no effect on the persistence of latent papilloma virus DNA, based on Southern blot and PCR studies. Another laser approach takes advantage of the vascular nature of RRP lesions using a tuned flash pump dye (FPD) laser in the yellow spectrum (around 577 nm), which is strongly absorbed by hemoglobin (27). This laser has proven effective against cutaneous verrucous lesions that have a similar

vascular core. The major theoretical advantage of this approach is that the FPD laser coagulates rather than vaporizes tissue. This not only decreases damage and scarring of adjacent normal tissues, but it also eliminates the plume of smoke that is potentially hazardous to the surgical team.

Because of the potential destruction and hazards associated with tissue vaporization, we have found it useful to limit the use of the CO_2 laser in the treatment of large RRP lesions. Instead, we use very aggressive suction to control and remove pathologic tissues and eschar. The large-bore suction allows the lesions to be manipulated for more precise cutting or vaporization by the laser(**Figure 3**). For small focal lesions it is often preferable to substitute cold microsurgical techniques facilitated by saline infusion, and reserve the CO_2 laser for more limited cutting and dissection. Recently, we have found the microdebrider with the Skimmer angled blades (Xomed, Jacksonville, FL) useful for management of bulky papillomata (**Figure 3**), especially in areas that are difficult to expose adequately for laser ablation. This instrument selectively sucks the amorphous lesion into the cutting blades, allowing for surprisingly delicate tissue handling when the variable speed control foot switch mode is used. For focal lesions, Zeitles and Sataloff (28) described a useful technique of phonomicrosurgical resection designed to remove the entire lesion, spare the lamina propria, and promote healing with limited scarring. The keys to this approach are subepithelial infusion of saline with 1:10,000 epinephrine, and sufficient magnification to allow precise elevation of a microflap that circumscribes the lesion but preserves the uninvolved lamina propria.

Laryngeal Framework Surgery

Laryngeal framework surgery consists of procedures designed to alter vocal fold shape and tension to affect glottic closure and vocal pitch. Medialization thyroplasty is the most widely used laryngeal framework surgical procedure. Isshiki introduced the basic techniques of medialization thyroplasty 25 years ago (29) and he has continued to build on the concept of adjusting vocal fold position and tension to achieve optimal vocal function. Using an in vivo canine laryngeal model and excised larynges, he controlled airflow and subglottal pressure to demonstrate that a normal voice can be achieved throughout a range of glottal sizes (30,31). Displays of these data indicated that excessively tight closure can exacerbate vocal breathiness and roughness. This is an important consideration as evolving techniques of medialization affect the position, tension, and viscosity of the vocal fold. A relatively common reason for poor thyroplasty outcome is excessive anterior medialization, which results in increased homolateral vocal fold stiffness and impaired oscillation of the contralateral fold.

Placement of the cartilaginous window, management of the inner perichondrium, and choice of displacement shim are the key variables in medialization thyroplasty surgery (**Table III**).Adherence to key anatomical factors is important to the success and safety of new medialization techniques. Ford (32) cited common mishaps in the placement of cartilaginous windows for thyroplasty: (1) adhering to a formula for an implant different than the one being used, (2) failing to adjust the size of the implant to the size of the larynx, and (3) placing the window too far anterior or superior. Magnetic resonance imaging studies showed the failure of effective localized vocal fold displacement when these considerations were not employed in window placement and implant design (33). Analysis of coronal and cross-sectional histology and cadaver dissections demonstrated the importance of preserving the thyroid cartilage inner perichondrium—especially anteriorly—and placing the displacement shim posteriorly (32). Such precautions prevent

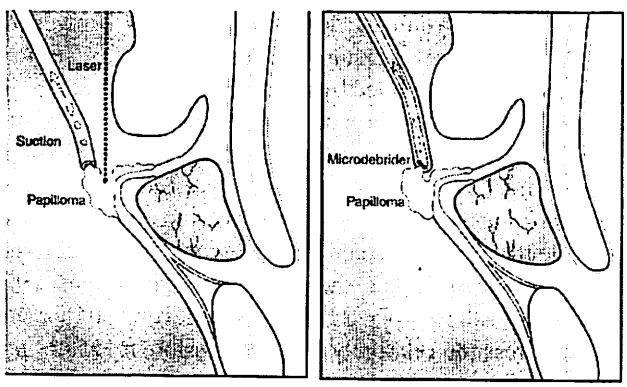


Figure 3. Adequate suctioning can limit thermal trauma when removing papillomata. In this drawing, suction is used aggressively as an adjunct to laser ablation (left) and can be used creatively to manipulate and remove diseased tissue with the microdebrider (right).

Table III.
Variables In Medialization Thyroplasty

PLACEMENT OF THYROID CARTILAGE WINDOW
1. Placement based on Isshiki model
2. Vary anterior-posterior limits based on shim design (avoid anterior third)
3. Vary superior-inferior limits based on shim design (avoid ventricle)

MANAGEMENT OF INNER PERICHONDRIUM
1. Preserve
2. Anterior-based flap
3. Incise or excise

IMPLANTS
1. Silastic (preformed or carved)
2. Hydroxylapatite (VoCom system)
3. Expanded polytetrafluoroethylene (Gore-Tex)
4. Titanium and other experimental materials

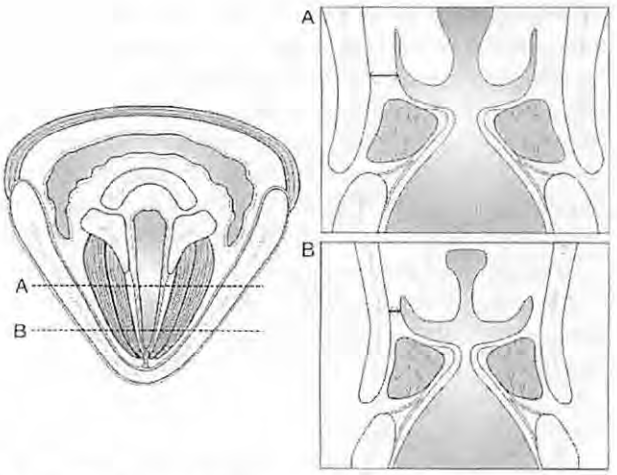


Figure 4. This schematic demonstrates the proximity of the ventricular saccule to the inner thyroid perichondrium. At the level of the posterior vocal fold (A) there is sufficient soft tissue interposed between inner perichondrium and ventricular mucosa to allow safe dissection. At the anterior glottis (B) it is easy to injure or penetrate the mucosa, leading to granuloma and extrusion.

causing an inadvertent penetrating injury of the laryngeal ventricular mucosa, where the ventricle and saccule often approximate the thyroid cartilage inner perichondrium within 1-2 mm in the anterior larynx (**Figure 4**).

Recently, there has been a proliferation of different materials and designs for medialization thyroplasty displacement shims. Montgomery (34) tried to simplify medialization thyroplasty by developing a preformed silastic shim. The Montgomery® thyroplasty implant system (Boston Medical Products, Westborough, MA) requires an incision of the inner perichondrium and extended posterior placement to medialize the vocal process of the arytenoid cartilage (35). Cummings and Flint (36) developed a different preformed displacement shim made of hydroxylapatite and designed to displace the membranous vocal fold. The VoCom® system (Smith and Nephew Richards, Memphis, TN) was based on laboratory studies using animal models that demonstrated the relative tissue tolerance of hydroxylapatite compared to silicone and Teflon (37). In addition to being a preformed prosthesis, this implant has the theoretical advantage of eventual integration into the thyroid cartilage for long-term stability. A recent report by Friedrich (38) demonstrated successful use of a modifiable preformed titanium vocal fold medialization implant (developed in collaboration with Heinz Kurz Company, Dusslingen, Germany) that comes in two sizes and spring-locks into the cartilage for stability. McCulloch and Hoffman (39) proposed the use of a commonly available material, expanded polytetrafluoroethylene (Gore-Tex), which can be cut in strips, passed through a thyroplasty window, and layered into the paraglottic space to achieve vocal fold medialization. The relative biocompatibility of this material was emphasized by Giovanni (40), who reported short-term favorable results.

Manipulation and placement of the arytenoid cartilage is another way to address glottic insufficiency. The arytenoid adduction procedure described by Isshiki (41) has found widespread acceptance for correcting large glottic chinks, posterior glottic incompetence, and malposition of the arytenoid. Among the various techniques described, that of Netterville has been most lucidly illustrated (42). Unlike the Isshiki procedure, Netterville preserves the cricoarytenoid joint and achieves exposure by rotating the thyroid cartilage and removing a posterior cartilaginous window. In an attempt to determine the biomechanical effects of arytenoid adduction on the larynx, Noordzij, Perrault, and Woo (43) used an ex vivo canine model. They attempted to show how suture tension affected glottal configuration (vocal fold length and position) and tension. They demonstrated that applying tension to a suture attached to the muscular process in the vector of the lateral cricoarytenoid and thyroarytenoid muscles medialized and lowered the vocal fold. They concluded that while arytenoid adduction medializes the membranous vocal fold (the *anterior subunit*), it did not alter stiffness. Patients with vocal fold paralysis often require both medialization and stiffening of the membranous vocal fold to restore effective oscillation. This suggests that while arytenoid adduction might provide excellent relief of aspiration and correct a *posterior subunit* insufficiency, additional measures might be required for optimal rehabilitation.

Modifications of medialization thyroplasty have been used to correct posterior glottic insufficiency, and the Montgomery prosthesis (34) was designed to specifically medially displace the vocal process of the arytenoid. Limitations of such an approach have stimulated interest in combined procedures and have led to further variations on medialization thyroplasty. In collaboration with Isshiki, Kojima et al demonstrated the limitations of a large prosthesis designed to displace the vocal process of the arytenoid (44). Using fresh human cadaver larynges, they showed that the vocal process was distant from the thyroid cartilage window and was surrounded by abundant muscle tissue. Inserting a large silicone prosthesis only achieved successful medialization of the vocal process and the membranous vocal fold. Computed tomography revealed that the cartilaginous portion of the vocal fold was hardly medialized despite the use of sufficiently large shims. In contrast, they achieved excellent placement of the posterior larynx with a smaller prosthesis, placed more posteriorly, and designed to compress the relatively easily accessible muscular process.

Clinical observations of the occasional limitations of single modality framework surgery has stimulated more careful analysis of the pathology in the selection of treatment alternatives. One of the major advances in framework surgery has been the increasing application of combinations of phonosurgical procedures to address various aspects of the pathology. The simplest example of combining procedures is the use of bilateral medialization thyroplasty. Postma, Blalock, and Koufman (45) reported on the use of bilateral medialization laryngoplasty in the management of symptomatic bowed vocal folds. In addition, they reported on the value of lipoinjection to enhance the voice results in some of these patients. The work is consistent with the earlier report of Ford, Bless, and Prehn (46) that used medialization thyroplasty as an adjunct to other medialization and augmentation procedures. Netterville, who popularized the concept of combining medialization thyroplasty with arytenoid adduction, reported on the Vanderbilt experience with silastic medialization and arytenoid adduction (47). Such an approach addresses the anatomical position of the vocal process as well as the placement and tension of the membranous vocal fold. An entirely different approach to the cricoarytenoid joint has recently been shown by Zeitels. In an attempt to more precisely place the arytenoid, he exposed the joint and directly pexed the arytenoid to the cricoid (48). Using this arytenopexy approach, he aimed to simulate not only the contraction of the lateral cricoarytenoid muscle but also the synchronous contraction of the interarytenoid, lateral thyroarytenoid, and posterior cricoarytenoid muscles. In more recent modifications of this technique, he proposed disarticulation and subluxation of the cricothyroid joint to improve vocal fold tension (49). Using the combination of arytenopexy, medialization thyroplasty, and cricothyroid subluxation, optimal placement and function is achieved, along with restoration of normal phonation times, two-octave dynamic ranges, and minimal acoustic perturbation.

The evolution of medialization thyroplasty continues to be driven by comprehensive objective assessments of voice. Consistent with most published results of thyroplasty, Lu et al (50) demonstrated significant improvement in glottic-gap size during phonation, maximum phonation time, transglottic airflow, jitter, harmonics-to-noise ratio, loudness, breathiness, and hoarseness assessments. There is a tendency for results to stabilize after 1 month, so there is little change noted after 6 months with incremental assessments[50]. In an attempt to correlate such results with the physical signs of injury and recovery, Gorham and colleagues (51) examined 50 thyroplasty patients incrementally and found initial evidence of postoperative irritation (erythema in 68%, edema in 76%, and hematoma in 28%) that reduced substantially by 1 month postoperatively.

A precautionary note was sounded by Janas et al (52), who studied pulmonary function in thyroplasty patients. They found that thyroplasty surgery worsened the FEF50%:FIF50% ratio in 12 of 15 patients and created new extrathoracic obstruction in 4 of 15 patients. Fortunately, these findings were only associated with symptoms in highly active younger patients. They suggested that the possibility of decreased laryngeal airflow should be discussed with younger, active patients during preoperative counseling.

In considering the risk of complications, the surgeon should be aware of the learning curve associated with medialization procedures. This was apparent in a recent survey of 1039 surgeons who performed the procedure (53). Forty-two percent of respondents reported one or more major complications, and this was significantly more common in surgeons who had performed less than 10 procedures in their career. The most common complications were failure to achieve voice improvement, airway compromise, and implant migration. Koufman and Postma (54) reviewed their experience with revision laryngoplasty and found the most common problems to be undercorrection of the posterior glottis, implant placement too high, implant extrusion, overcorrection anteriorly, and undercorrection of the glottic gap.

Injections and Implants

Placement of materials directly into the vocal fold remains a convenient and useful method of correcting glottic insufficiency. In addition to the traditional method of injecting through a laryngoscope without magnification, new approaches optimize contemporary optics to facilitate accuracy and reduce perioperative morbidity. Rosen (55) performed a variety of augmentation injection procedures using a rigid zero-degree telescope (4 mm diameter, 30 cm long) coupled with a video camera and passed through a slotted anterior commissure laryngoscope. The procedures used intravenous sedation and topical anesthesia techniques, and the aid of an operative assistant. The vocal folds were easily accessed for either medial or lateral injections. This approach seems particularly useful for medial, superficial lamina propria injection sites where precise placement is essential for correcting scarring and focal defects. Ford et al (8) described the use of rigid endoscopy in the office or voice lab to facilitate vocal fold injections. Such an approach provides superior optics, facilitates incremental injection with videostroboscopic monitoring, and avoids the need for heavy sedation or general anesthesia. Students, residents, and patients are afforded an excellent view of the procedure in real time or on videotape recordings. A variation of this approach was described by Arad-Cohen and Blitzer (56), who performed vocal fold augmentation injections using a flexible fiberoptic laryngoscope in an office setting. This approach might be better tolerated by patients with a heightened gag reflex. The flexible scope can also be used to

monitor transcutaneous placement of injectates by observing incremental tissue displacement when the needle is placed through the anterior cervical skin and larynx. Gray et al (15) extended this concept in developing the "minithyrotomy" approach. They described cutting a small hole in the anterior thyroid cartilage, just off the mid-line, and placing the injecting needle or instrument through this hole into the lamina propria of the vocal fold. Placement was guided by observing tissue movement with motion of the instrument. The authors felt that this affords a better opportunity to precisely place material in the lamina propria and critical infrafold region. Implant placement is further assured by avoiding mucosal puncture sites, so there is no chance of extrusion as often occurs with conventional injection techniques. The infrafold area—roughly 3 mm inferior to the leading edge of the vocal fold—is the site of origin of the mucosal wave. This site appears to be important for oscillation. Interestingly, it is this area of expanded intermediate layer of the lamina propria that is rich in elastin fibers, hyaluronic acid, and fibromodulin. These molecules are largely responsible for the elasticity and viscosity essential for oscillation (15).

Materials

The challenge of treating glottic insufficiency due to vocal fold paralysis, atrophy, or scarring continues to generate interest in implantable substances. For years it was believed that Teflon paste was inert and suitable for vocal fold augmentation. Similar alloplastic materials such as silicone and recently Bioplastique have been used in Asia and Europe. Vocal fold tissue reaction to Teflon has limited long-term results and possible local and distant complications (57-59). A new approach has been the development of bioimplants for vocal fold augmentation.

Initial studies using bovine collagen (60,61) led to the use of a variety of injectable autologous materials. Autologous fat has been used clinically as an injection and short-term results have been encouraging (62,63). Shaw et al (64) recently reported favorable 1-year results of autologous fat injections in 22 patients using acoustic analysis, videostroboscopy, perceptual ratings, and subjective patient ratings. To enhance graft survival, they employed a rigorous protocol that harvested fat by surgical excision, rinsed extensively to separate fat from blood, free fatty acids and cellular debris, and finally suspended fat in human insulin. Sataloff and coworkers (65) proposed using small fat grafts as implants in the scarred vocal fold. In this technique, a small incision is made on the superior surface of the vocal fold, and a suitable pocket is surgically created to facilitate placement of the graft. The major concern about all forms of fat—and other bioimplants—is persistence of the implant. In animal models, fat has shown variable resorption but some persistence at 6 months (66). Bower, Valentino, and Hoffman (67) documented persistence of fat at 5 months by histologic study in a laryngectomy patient. Using the feline model, Saccogna et al (68) found evidence of viable fat to 8 months post-implantation, but no substantial graft survival in cats sacrificed at 12 months. There is no consensus on the proper technique for fat implantation or injection and it remains unclear which situations are most favorable for long-term implant survival. As with other bioimplants, persistence of correction and function is perhaps more important than graft survival.

Collagen and fascia are other bioimplants that have proven useful in vocal fold augmentation. After reporting on the use of bovine cross-linked collagen in a broad population of patients with glottic insufficiency (61), we focused use of this material on specific populations (69) where other materials did not seem as useful. Burke (Pacific Voice Foundation, San Francisco, CA 1998) recently reported successful use of bovine collagen in treating over 300 patients with Parkinson's disease. However, concerns about the possibility of allergic responses have limited the clinical use of bovine collagen. We addressed this concern by applying the same technique with autologous collagen preparations. In animal models (70) and clinical studies (71), autologous collagen injections reacted in a manner similar to the bovine material but without the risk of allergic reactions. Another autologous material familiar to otolaryngologists is autologous fascia. Rihkanen (72) reported on the use of injection of autologous fascia for vocal fold augmentation. More recently, Tsunoda and colleagues (73) described the use of autologous fascia transplantation. This technique is similar to the technique described for fat implantation (65) and it appears applicable to management of scarred larynges.

As new materials and techniques continue to emerge, it is important to recognize the place of Teflon injection in the current armamentarium of the phonosurgeon. Teflon is still useful in situations where long-term voice outcome is not the primary concern. Examples include patients with terminal cancer and older patients disabled by aspiration who are not suitable surgical candidates. It remains important to salvage poor results in patients previously treated with Teflon. In the past, salvage approaches were limited to excision or CO_2 laser ablation (74) and in some cases, subsequent augmentation with collagen (57). One of the most exciting applications of the free-electron laser is the ablation of Teflon. This experimental laser, when tuned to a wavelength of 8.5mm, was found to ablate injected Teflon efficiently without appreciable charring or damage to the surrounding soft tissues (75). Netterville and his coworkers (76,77) introduced a very valuable approach to removing Teflon granulomas through a lateral laryngotomy. They create a thyroid cartilage window for exposure of the paraglottic space, through which the granuloma is completely removed. Successful reconstruction of the vocal fold

soft tissues is then achieved by inserting an inferiorly based sternohyoid muscle flap. We have found this approach useful for a variety of applications, including revision of thyroplasties and reconstruction when there is a break in the endolaryngeal mucosa.

With the variety of approaches and materials available for injection and implantation, it is important to maintain sight of principles necessary for optimal results. To achieve the most effective glottic closure, one must achieve medialization of the involved vocal fold in the same plane as the opposite side. Failure to do so can lead to occult glottic insufficiency where the airflow and breathy phonation appear inconsistent with the indirect laryngoscopic findings. This is important in primary and revision procedures. Proper voicing is dependent on approximation of the membranous vocal folds. This is not always accomplished by rotation of the vocal process and it is often necessary to fill out the membranous fold directly. However, it is counterproductive to overfill the anterior third of the vocal fold since premature contact of the anterior fold can impede oscillation of the contralateral vocal fold and accentuate posterior glottic insufficiency. In applying mechanical principles and seeking more ideal substances to implant, it is important to consider the biological properties of the materials used. Bioimplants are often more forgiving over time, but allowance must be made for resorption and assimilation. Alloplastic and inert materials induce permanent tissue changes and extreme care should be used in patient selection and technique.

Summary

Phonosurgery has undergone enormous changes in recent years. This is a result of many factors, including new materials and procedures, improved understanding of vocal function, and better tools of analysis. Basic principles remain focused on preservation of normal tissues and the use of materials and methods that best imitate normal structures. There is an increasing effort to match viscoelastic properties and minimize iatrogenic scarring of vocal folds. Through continued use of the tools for voice assessment, we will be able to further refine techniques. Increasing understanding of the structure and function of the vocal mechanism remains essential for the development of improved techniques, instruments, and materials.

Some of the major points in this review are:

- Preservation of the vocal fold lamina propria is the key to restoring normal structure and function.
- Medial microflaps are best for limited medial lesions with thin or redundant cover.
- There are 3 types of sulcus vocalis. Type I is physiological, with usually minimal disruption of mucosal wave and vocal function. Type II (vergeture) has a linear loss of the superficial lamina propria with disturbed mucosal wave, oscillation, and voice. Type III is a focal pit usually involving all layers of the lamina propria, may be associated with cysts, and severely alters vocal fold oscillation and voice.
- Management of laryngeal papillomata should introduce the least amount of physical or thermal damage to the surrounding normal tissues. This can be done by careful microdissection, various focal lasers, aggressive suctioning, and gentle microdebridement.
- Medialization thyroplasty is possible with shims of silastic, hydroxylapatite, or other materials but results are dependent on careful assessment of the pathology and understanding of the anatomy.
- Thyroid cartilage window placement should vary with the size of the larynx and shape of the planned implant, with an effort to avoid too-anterior or too-superior placement.
- Manipulation of the arytenoid by arytenoid adduction or arytenopexy is the best way to correct posterior subunit gaps, vocal fold tension, and inappropriate vocal fold level.
- Bioimplantable substances such as fat, collagen, and fascia are more forgiving, better tolerated, and match vocal fold viscoelastic properties better than alloplastic injectables. They are the most suitable implants in patients who expect long life and where optimal voice is a primary goal.
- Lateral laryngotomy with pedicled flap reconstruction provides the best current solution for total removal of extensive Teflon granulomas with reconstruction of glottic and paraglottic space tissues.

While this review has focused on the advancements and refinements of phonomicrosurgery, laryngeal framework surgery, and implants, other surgical advances in voice restoration have occurred. Notable are the advances in conservation and reconstructive surgery associated with laryngeal cancer rehabilitation, and refinements in the management of neuromuscular diseases of the larynx, including variations in the use of botulinum toxin for spasmodic dysphonia. The field of neurolaryngology is generating increasing interest and holds the keys to functional laryngeal transplantation, management of bilateral vocal fold paralysis and other neurological deficits.

As with all phonosurgery, careful, accurate assessment of the pathology is essential. It is often more important to determine the components of pathology that can be corrected and develop a comprehensive treatment program, than to seek a single ideal procedure or material.

Acknowledgment

This work was supported in part by the National Center for Voice and Speech Grant # P60DC00976 from the National Institutes on Deafness and Other Communication Disorders.

References

1. Von Leden H. Microlaryngoscopy: a historical vignette. *J Voice* 1988;1:341-346.

2. Hirano M. Phonosurgery: past, present, and future. *Transactions ABEA* 1995; 25-30.

3. Zeitels SM, Vaughan CW. "External counterpressure" and "internal distension" for optimal laryngoscopic exposure of the anterior glottal commissure. *Ann Otol Rhinol Laryngol* 1994;103:669-676.

4. Kantor E, Berci G, Partlow E , Paz-Partlow M. A completely new approach in microlaryngeal surgery. *Laryngoscope* 1991;101:676-679.

5. Yanagisawa E, Yanagisawa K, Horowitz JB, Mambrino LJ. Comparison of new telescopic video microlaryngoscopic techniques. *Ann Otol Rhinol Laryngol* 1992;101:51-60.

6. Yeh AR, Huang HM, Chen YL. Telescopic video microlaryngeal surgery. *Ann Otol Rhinol Laryngol* 1999;108:165-168.

7. Bastian RW, Delsupehe KG. Indirect larynx and pharynx surgery: a replacement for direct laryngoscopy. *Laryngoscope* 1996;106:1280-1286.

8. Ford CN, Roy N, Sandage M ,Bless DM. Rigid endoscopy for monitoring indirect vocal fold injection. *Laryngoscope* 1998;108:1584-1586.

9. Merati AL, Andrews RJ, Courey MS, Garrett CG, Ossoff RH. Phonomicrosurgical management of intracordal cysts. *Oper Tech Otolaryngol Head Neck Surg* 1999;9:230-237.

10. Sataloff RT, Spiegel JR, Heuer RJ, et al. Laryngeal mini-microflap: a new technique and reassessment of the microflap saga. *J Voice* 1995;9:198-204.

11. Courey MS, Garrett CG, Ossoff RH. Medical microflap for excision of benign vocal fold lesions. *Laryngoscope* 1997;107:340-344.

12. Zeitels SM, Bunting GW, Hillman RE, Vaughn T. Reinke's edema: phonatory mechanisms and management strategies. *Ann Otol Rhinol Laryngol* 1997;106:533-543.

13. Zeitels SM. Laser versus cold instruments for microlaryngoscopic surgery. *Laryngoscope* 1996;106:545-552.

14. Lumpkin SM, Bennett S, Bishop SG. Postsurgical followup study of patients with severe polypoid corditis. *Laryngoscope* 1990;100:399-402.

15. Gray SD, Bielamowicz SA, Dove H, Titze IR, Ludlow C. Experimental approaches to vocal fold alteration: introduction to the minithyrotomy. *Ann Otol Rhinol Laryngol* 1999;108:1-9.

16. Hirano M, Yoshida T, Tanaka S, Hibi S. Sulcus vocalis: functional aspects. *Ann Otol Rhinol Laryngol* 1990;99:679-683.

17. Sato K, Hirano M. Electron microscopic investigation of sulcus vocalis. *Ann Otol Rhinol Laryngol* 1998;107:56-60.

18. Nakayama M, Ford CN, Brandenburg JH, Bless DM. Sulcus vocalis in laryngeal cancer: a histopathologic study. *Laryngoscope* 1994;104:16-24.

19. Ford CN, Inagi K, Bless DM, Khidr A, Gilchrist KW. Sulcus vocalis: a rational analytical approach to diagnosis and management. *Ann Otol Rhinol Laryngol* 1996;105:189-200.

20. Bouchayer M, Cornut G, Loire R, Witzig E, Roch JB, Bastian RW. Epidermoid cysts, sulci, and mucosal bridges of the true vocal cord: a report of 157 cases. *Laryngoscope* 1985;95:1087-1094.

21. Pontes P, Behlau M. Treatment of sulcus vocalis: auditory perceptual and acoustic analysis of the slicing mucosa surgical technique. *J Voice* 1993;7:365-376.

22. Pou AM, Rimell FL, Jordan JA, et al. Adult respiratory papillomatosis: human papillomavirus type and viral coinfections as predictors of prognosis. *Ann Otol Rhinol Laryngol* 1995;104:758-762.

23. Snoeck R, Wellens W, Desloovere C, et al Treatment of severe laryngeal papilomatosis with intralesional injections of cidofovir [(S)-1-(3-hydroxy-2-phosphonylmethoxypropyl) cysosine]. *J Med Virol*, 1998;54:219-225.

24. Rosen CA, Woodson GE, Thompson JW, Hengesteg AP, Bradlow HL. Preliminary results of the use of indole-3-carbinol for recurrent respiratoy papillomatosis. *Otolaryngol Head Neck Surg* 1998;118:810-815.

25. Nuss R. Management of pediatric laryngeal papillomatosis. *Oper Tech Otolaryngol Head Neck Surg* 1999;9:210-213.

26. Shikowitz MJ, Abramson AL, Freeman K, Steinberg BM, Nouri M. Efficacy of dhe photodynamic therapy for respiratory papillomatosis: immediate and long-term results. *Laryngoscope* 1998;108:962-967.

27. Bower CM, Flock S, Waner M, Schaeffer R. Flash pump dye laser treatment of laryngeal papillomas. *Ann Otol Rhinol Laryngol* 1998;107:1001-1005.

28. Zeitels SM, Sataloff RT. Phonomicrosurgical resection of glottal papillomatosis. *J Voice* 1999;13:123-127.

29. Isshiki N, Morita H, Okamura H, Hiramoto M. Thyroplasty as a new phonosurgical technique. *Acta Otolaryngol (Stockh)*, 1974;78:451-457.

30. Isshiki N. Vocal mechanics as the basis for phonosurgery. *Laryngoscope* 1998;108:1761-1766.

31. Isshiki N. Mechanical and dynamic aspects of voice production as related to voice therapy and phonosurgery. *J Voice* 1998;12:125-137.

32. Ford CN. Intraoperative decision making in medialization laryngoplasty. *Oper Tech Otolaryngol Head Neck Surg* 1999;10:17-21.

33. Ford CN, Unger JM, Zundel RS, Bless DM. Magnetic resonance imaging (MRI) assessment of vocal fold medialization surgery. *Laryngoscope* 1995;105:498-504.

34. Montgomery WW, Montgomery SK, Warren MA. Thyroplasty simplified. *Oper Tech Otolaryngol Head Neck Surg* 1993;4:223-231.

35. Montgomery WW, Montgomery SK. Montgomery thyroplasty implant system. *Ann Otol Rhinol Laryngol* 1997;106(suppl 170):1-16.

36. Cummings CW, Purcell LL, Flint PW. Hydroxlapatite laryngeal implants for medialization: preliminary report. *Ann Otol Rhinol Laryngol* 1993;102:843-851.

37. Flint PW, Corio RL, Cummings CW. Comparison of soft tissue response in rabbits following laryngeal implantation with hydroxylapatite, silicone rubber, and Teflon. *Ann Otol Rhinol Laryngol* 1997;106:339-407.

38. Friedrich G. Titanium vocal fold medializing implant: Introducing a novel implant system for external vocal fold medialization. *Ann Otol Rhinol Laryngol* 1999;108:79-86.

39. McCulloch TM, Hoffman HT. Medialization laryngoplasty with expanded polytetrafluoroethylene – surgical technique and preliminary results. *Ann Otol Rhinol Laryngol* 1998;107:427-432.

40. Giovanni A, Vallicioni JM, Gras R, Zanaret M. Clinicall experience with Gore-Tex for vocal fold medialization. *Laryngoscope* 1999;109:284-288.

41. Isshiki N, Tanabe M, Sawada M. Arytenoid adduction for unilateral vocal cord paralysis. *Arch Otolaryngol* 1978;104:555-558.

42. Miller FR, Bryant GL, Netterville JL. Arytenoid adduction in vocal fold paralysis. *Oper Tech Otolaryngol Head Neck Surg* 1999;10:36-41.

43. Noordzij JP, Perrault DF, Woo P. Biomechanics of arytenoid adduction surgery in an ex vivo canine model. *Ann Otol Rhinol Laryngol* 1998;107:454-461.

44. Kojima H, Hirano S, Shoji K, Isshiki N. Anatomic study for posterior medialization thyroplasty. *Ann Otol Rhinol Laryngol* 1999;108:373-377.

45. Postma GN, Blalock PD, Koufman JA. Bilateral medialization laryngoplasty. *Laryngoscope* 1998;108:1429-1434.

46. Ford CN, Bless DM, Prehn RB. Thyroplasty as primary and adjunctive treatment of glottic insufficiency. *J Voice* 1992;6:277-285.

47. Netterville JL, Stone RE, Luken ES, Civantos FJ, Ossoff RH. Silastic medialization and arytenoid adduction: the Vanderbilt experience. A review of 116 phonosurgical procedures. *Ann Otol Rhinol Laryngol* 1993;102:413-424.

48. Zeitels SM, Hochman I, Hillman RE. Adduction arytenopexy: a new procedure for paralytic dysphonia and the implications for medialization laryngoplasty. *Ann Otol Rhinol Laryngol* 1998;107:1-24.

49. Zeitels SM. Adduction arytenopexy with medialization laryngoplasty and cricothyroid subluxation: a new approach to paralytic dysphonia. *Oper Tech Otolaryngol Head Neck Surg*, 1999;10:9-16.

50. Lu FL, Casiano RR, Lundy DS, Xue JW. Longitudinal evaluation of vocal function after thyroplasty type I in the treatment of unilateral vocal paralysis. *Laryngoscope* 1996;106:573-577.

51. Gorham MM, Avidano MA, Crary MA, Cotter CS, Cassisi NJ. Laryngeal recovery following type I thyroplasty. *Arch Otolaryngol Head Neck Surg* 1998;124:739-742.

52. Janas JD, Waugh P, Swenson ER, Hillel A. Effect of thyroplasty on laryngeal airflow. *Ann Otol Rhinol Laryngol* 1999;108:286-292.

53. Rosen CA. Complications of phonosurgery: results of a national survey. *Laryngoscope* 1998;108:1697-1703.

54. Koufman JA, Postma GN. Revision laryngoplasty. *Oper Tech Otolaryngol Head Neck Surg* 1999;10:61-65.

55. Rosen CA. Phonosurgical vocal fold injection: indications and techniques. *Oper Tech Otolaryngol Head Neck Surg* 1999;9:203-209.

56. Arad-Cohen A, Blitzer A. Office-based direct fiberoptic laryngoscopic surgery. *Op Tech Otolaryngol Head Neck Surg* 1999;9:238-242.

57. Nakayama M, Ford CN, Bless DM. Teflon vocal fold augmentation: Failures and management in 28 cases. *Otolaryngol Head Neck Surg* 1993;109:493-498.

58. Rubin HJ. Misadventure with injectable polytef (Teflon). *Arch Otol* 1975;101:114-116.

59. Ellis JC, McCaffrey TV, DeSanto LW, Reiman HV. Migration of Teflon after vocal cord injection. *Otolaryngol Head Neck Surg* 1987;96:63-66.

60. Ford CN, Martin DW, Warner TF. Injectable collagen in laryngeal rehabilitation. *Laryngoscope* 1984;94:513-518.

61. Ford CN, Bless DM, Loftus JM. Role of injectable collagen in the treatment of glottic insufficiency: a study of 119 patients. *Ann Otol Rhinol Laryngol* 1992;101:237-247.

62. Laccourreye O, Crevierbuchman L, Lepimpecbarthes F, Garcia D, Riquet M, Brasnu D. Recovery of function after intracordal autologous fat injection for unilateral recurrent laryngeal nerve paralysis. *J Laryngol Otol* 1998;112:1082-1084.

63. Brandenburg JH, Kirkham W, Koschkee D. Vocal cord augmentation with autogenous fat. *Laryngoscope* 1992;102:495-500.

64. Shaw GY, Szewczyk MA, Searle J, Woodroof J. Autologous fat injection into the vocal folds: technical considerations and long-term follow-up. *Larngoscope* 1997;107:177-186.

65. Sataloff RT, Spiegel JR, Hawkshaw M, Rosen DC, Heuer RJ. Autologous fat implantation for vocal fold scar: a preliminary report. *J Voice* 1997;11:238-246.

66. Milkus JL, Koufman JA, Kilpatrick SE. Fate of liposuctioned and purified autologous fat injections in the canine vocal fold. *Laryngoscope* 1995;105:17-22.

67. Bauer CA, Valentino J, Hoffman HT. Long-term result of vocal cord augmentation with autogenous fat. *Ann Otol Rhinol Laryngol* 1995;104:871-874.

68. Saccogna PW, Werning JW, Setrakian S, Strauss M. Lipoinjection in the paralyzed feline vocal fold: study of graft survival. *Otolaryngol Head Neck Surg* 1997;117:465-470.

69. Ford CN, Bless DM. Selected problems treated by vocal fold injection of collagen. *Am J Otolaryngol* 1993;14:257-261.

70. Staskowski PA, Ford CN, Inagi K. The histologic fate of autologous collagen injected in the canine vocal fold. *Otolaryngol Head Neck Surg* 1998;118:187-190.

71. Ford CN, Staskowski PA, Bless DM. Autologous collagen vocal fold injections: A preliminary clinical study. *Laryngoscope* 1995;105:944-948.

72. Rihkanen H. Vocal fold augmentation by injection of autologous fascia. *Laryngoscope* 1998;108:51-54.

73. Tsunoda K, Takanosawa M, Niimi S: Autologous transplantation of fascia into the vocal fold: a new phonosurgical technique for glottal incompetence. *Laryngoscope* 1999;109:504-508.

74. Dedo HH. Injection and removal of Teflon for unilateral vocal cord paralysis. *Ann Otol Rhinol Laryngol* 1992;101:81-86.

75. Lano CF, Reinisch L, Kuo T, et al. Ablation of teflon granulomas in the canine larynx with the free-electron laser. *Ann Otol Rhinol Laryngol* 1999;108:17-23.

76. Netterville JL, Rainey CL, Coleman JR, Reinisch L, Chang S, Ossoff RH. Lateral laryngotomy for the removal of teflon granuloma. *Ann Otol Rhinol Laryngol* 1998;107:735-744.

77. Coleman JR, Miller FR, Netterville JL. Teflon granuloma excision via a lateral laryngotomy. *Oper Tech Otolaryngol Head Neck Surg* 1999;10:29-35.

NCVS Status and Progress Report - 14
September 1999, 101-113

3-D Vocal Tract Imaging and Formant Structure: Varying Vocal Register, Pitch and Loudness

Kenneth Tom, Ph.D.
Department of Speech Communication, California State University Fullerton
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Eric Hoffman, M.D.
Division of Physiologic Imaging, Department of Radiology, The University of Iowa College of Medicine
Brad Story, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts

Abstract

Although advances in techniques for image acquisition and analysis have facilitated the direct measurement of three-dimensional vocal tract air space shapes associated with specific speech phonemes, little information is available with regard to changes in 3-D vocal tract shape as a function of vocal register, pitch and loudness. In this study, 3-D images of the vocal tract during falsetto and chest register phonations at various pitch and loudness conditions were obtained using electron beam computed tomography (EBCT). Eight image sets of the vocal tract were acquired for a single adult male subject: six during sung phonations in falsetto register and two during spoken phonations in chest register. Each volume set consisted of contiguous 3 mm axial slices encompassing the arch of the hard palate superiorly and the first tracheal ring inferiorly. Image processing included (1) segmentation, the differentiation of air space areas from surrounding tissue in the axial images; (2) reconstruction of the vocal tract shape in three dimensions using shape-based interpolation and (3) measurement of cross-sectional areas along the length of the vocal tract. Differences in vocal tract configuration and formant characteristics derived from the eight measured vocal tract shapes are reported.

Introduction

In the adult male voice, the highest fundamental frequencies of the speaking or singing voice are usually produced in falsetto register. Loud, high pitched phonation in falsetto register can be thought of as one end of the physiologic range of capabilities of the vocal production mechanism. Physiologically, however, falsetto phonation can be produced at various degrees of vocal intensity throughout the upper third to two-thirds of the fundamental frequency range in the adult male voice. In most voices there is some degree of overlap in fundamental frequencies that can be produced in either chest or falsetto register (Colton and Hollien, 1972; Hollien, 1974, 1977; Titze, 1988, 1994; Welch *et al.*, 1988).

Vocalization in falsetto register has long been a part of speech and song in many of the world's cultures. Linguistically, phonation in falsetto or breathy voice can function as a phonemic contrast or an element of prosody (Laver, 1980). High pitched falsetto vocalizations have been observed as a response to psychologically and/or emotionally traumatic situations (Scherer, 1995). It is used almost universally when imitating women's or children's voices (Large, 1972) and is frequently heard in character or cartoon voices on the theatrical stage and in film media. Falsetto singing appears in diverse cultural contexts, including the Beijing Opera and the *onnagata* tradition of Kabuki theater in Japan, where male actors have traditionally performed the roles of female characters (Malm, 1967; Alley, 1982; MacKerras, 1983); traditional chanting of Pygmy tribes in central Africa; Alpine yodeling in central Europe; and many genres of Western popular music. In Western classical music, it had become a tradition by the 17th century, especially

in the sacred music of England, to have choral voice parts for alto sung by countertenors, i.e. adult tenors or baritones using falsetto voice (Giles, 1982). More recently, there has been a revival of interest in the art of the countertenor vocal soloist in Europe and North America, especially in the performance of early music and Baroque opera. Countertenor soloists are adult males with tenor or baritone speaking voices, who generally use a falsetto-based vocal technique for singing (Giles, 1982; Welch *et al.*, 1988).

Falsetto phonation has also been used by speech pathologists in voice therapy and by teachers of classical singing voice to address vocal hyperfunction. Voice therapists have used falsetto as a therapeutic tool to treat functional voice disorders, including vocal nodules, ventricular phonation and muscle tension dysphonia (Boone, 1994; Colton and Casper 1990, Perlman 1992), engaging aspects of laryngeal behavior in falsetto that appear to be less taxing vocally to improve patients' vocal use patterns. Singing teachers may use falsetto phonation with male students to develop a light chest register ("head voice") or to decrease undesirable laryngeal effort in high pitch production (Frisell, 1964).

Although phonation in falsetto register is an integral part of the physiologic range of human vocal capability and its usage not uncommon, there are few comprehensive studies regarding physiologic control in falsetto speech or singing. Most studies examining physiologic control of vocal intensity, for example, have avoided high pitched or falsetto phonation (Holmberg *et al.*, 1988; Stathopoulos and Sapienza, 1993; Holmberg *et al.*, 1995). This is to a large degree due to practical and theoretical limitations of currently available measurement and analysis techniques used to obtain estimates of salient production variables, e.g. time-varying glottal behavior and vocal tract resonances. To circumvent the limitations of current approaches for estimating such physiologic control variables, an analysis by synthesis approach incorporating physiologically based modeling of falsetto phonation has been used (Titze, Mapes, and Story, 1994; Tom, 1996).

The objective of the current study was to acquire three-dimensional images of the vocal tract during phonation at various pitch and loudness conditions in falsetto and chest register for a single subject and to obtain the corresponding vocal tract length and cross-sectional area functions. These vocal tract measurements facilitated the specification of formant frequencies and bandwidths associated with each phonatory condition and also made it possible to utilize physiologically based computer simulation to estimate control variables in falsetto phonation. The formant structure of the high-pitched falsetto phonations, difficult to obtain from the wide-bandwidth oral airflow signal using acoustic LPC methods (Markel and Gray, 1976), was readily observed in the frequency response of the measured vocal tracts to a simulated impulse excitation.

Three-dimensional imaging of the vocal tract can be accomplished by obtaining a series of contiguous image slices through the portion of the body encompassing the vocal tract, segmenting the airway shape from its bordering tissues and reconstructing it in three dimensions. Images can be acquired with either electron beam computed tomography (EBCT) or magnetic resonance imaging (MRI). Each technique has its advantages and disadvantages. In terms of reducing the risk of any adverse side effects, MRI has the clear advantage. No hazardous effects have been observed from short term exposures to the magnetic fields currently used in MRI scanning systems. For imaging airways, however, MRI techniques have a number of disadvantages. Image resolution and accuracy are limited. Air to tissue boundaries can be distorted due to MRI imaging artifacts, effectively blurring the edges of the vocal tract slightly (Baer *et al.*, 1991). Tissues that are low in hydrogen content, such as bony structures and teeth, are captured poorly and appear to be the same gray scale density as air.

Scanning times for MRI image acquisition are much longer than for EBCT. Using current MRI technology, the time required to scan an entire vocal tract is approximately 4-5 minutes (Story, Titze and Hoffman, 1996), depending on the desired resolution and scanning parameters being used. The addition of pauses required for breathing when imaging the vocal tract during actual phonation increases total image acquisition time to 10 or more minutes per vocal tract shape. Under such circumstances subject fatigue and movement artifact become an important consideration. Scan time for imaging a vocal tract using EBCT techniques is much shorter, about 40-45 seconds. With pauses included for breathing between phoneme reiterations, total acquisition time is approximately 60-90 seconds. Because the present study included phonations at high effort conditions that could not be sustained over the 10-15 minutes of total image acquisition time required by MRI techniques, the use of MRI was not feasible.

The majority of studies on volumetric vocal tract imaging during speech have used MRI techniques. Baer *et al.* (1991) first demonstrated 3-D reconstruction of static vocal tract shapes during speech using MRI images. Vocal tract shapes for the four point vowels were acquired in sagittal, coronal and axial planes from two adult male subjects, with subsequent 3-D reconstruction and measurement of cross-sectional areas. Also using MRI techniques, Moore (1992) acquired images in sagittal and coronal planes for three vowels and two continuants for five adult male subjects and measured associated vocal tract volumes. He noted that vocal tract configurations were more stable when subjects actually phonated during image acquisition than when they simulated a phonatory posture without phonating. Both of these studies described the difficulties presented by the long acquisition times required (up to 45 minutes) by then current techniques. Sulter *et al.* (1992) used MRI to image

vocal tract shapes of an adult male subject, a trained classical singer, during productions of the vowels /i/, /ɛ/ and /ɑ/ measuring vocal tract length and cavity volumes. In another single subject study, Greenwood *et al.* (1992) used MRI to image five vowel shapes. Their corresponding area functions were derived from image slices configured similarly to cross-sections proposed in Mermelstein's (1973) vocal tract model. Dang *et al.* (1994) collected volumetric images of nasal tract passages, sinus cavities and vocal tract volumes using MRI. 3-D reconstructions of these structures were used to study the acoustics of the nasal volumes and to model nasal consonant production. Yang and Kasuya (1994) reconstructed volumetric MR images of the oral and pharyngeal spaces associated with five vowels produced by their three subjects, an adult male and female, and an 11 year old male. Tongue shapes and vocal tract configurations associated with fricative consonant production from four adult subjects were imaged, reconstructed in three dimensions and measured by Narayanan *et al.* (1995) and Narayanan (1995). The most comprehensive set of volumetric images of the vocal tract for speech sounds produced by a single speaker using MRI were produced by Story (1995). He acquired images for twenty-two complete vocal tract shapes, trachea, piriform sinuses and the nasal tract. A subset of these images, 18 vocal tract shapes associated with the phonation of 12 vowels, 3 nasals and 3 plosives, and their area functions were reported by Story *et al.* (1996). Using MRI technology and scanning parameters that were an improvement over those used in earlier studies, these image sets still required several minutes for image acquisition. Given pauses for inhalations required between reiterations of a target phoneme, image acquisition for each vocal tract volume set required 10 minutes or more of repeated execution of a phonation task, all while trying to avoid movement of vocal tract structures. Movement artifact due to subject fatigue and articulatory instability clearly becomes a concern when a phoneme target must be reiterated 30 times or more for each vocal tract shape.

Kiritani, Tateno, Iinuma & Sawashima (1977) explored the use of x-ray computed tomographic (CT) techniques to image cross-sectional areas of the vocal tract. Although they noted the potential for acquiring high resolution images of the vocal tract using CT, its implementation at that time was limited by the lack of adequate reconstruction and measurement software. X-ray techniques were also used by Perrier *et al.* (1992) and Beautemps *et al.* (1995) in their efforts to calculate cross-sectional areas of the vocal tract from measures of midsagittal width (width to area transformations).

For imaging airways, electron beam computed tomography (EBCT) techniques yield images of higher resolution than MRI images. The air-tissue boundary is captured with greater accuracy and bony structures and teeth are clearly imaged. Using currently available EBCT scanners, a sixty slice volumetric study of contiguous 3 mm slices encompassing the entire vocal tract can be scanned relatively quickly (about 40 seconds). This comparatively brief image acquisition time greatly reduces the potential for subject fatigue and associated movement artifact, which can blur resultant images. The chief disadvantage associated with EBCT is its use of ionizing radiation, which therefore limits the number of scans considered safe (International Commission on Radiological Protection, 1977; National Council on Radiation Protection and Measurements, 1987).

Image Acquisition and Analysis

Image Acquisition Protocol

Volumetric images of the vocal tract were scanned from a single male subject for sustained phonation of the vowel /ɑ/ under eight phonatory conditions, varying voice register, pitch and loudness levels. The conditions, summarized in Table I, included three sung falsetto register pitch levels (low pitch, 262 Hz; medium pitch, 349 Hz; high pitch, 466 Hz) and chest register speech phonation at a self-selected comfortable pitch level. Each of these four pitch levels was scanned at two intensity levels, moderately low intensity (*mezzo piano, mp*) and very loud intensity (*fortissimo, ff*). A scout scan preceded the acquisition of the eight image sets.

For the purposes of this study, a subject who could readily and consistently produce falsetto as well as chest register phonations throughout his fundamental frequency and intensity ranges was required. To increase the reliability of phonations produced during data collection, a trained singer was recruited. Singers trained in classical vocal technique develop vocal production habits which allow them to reliably produce tones of similar frequency, intensity and timbre. The subject for this study was a 45 year old adult male who has had extensive singing training in the Western classical tradition, including 12 years of vocal study as a baritone and 6 years of study as a countertenor. An active

Table I.
Phonatory Conditions for Vocal Tract Imaging

Vowel for all phonations = /ɑ/

Pitch/Condition	Register	Pitch (F_0)	Loudness
B-flat$_4$ ff	Falsetto	B-flat$_4$ (466 Hz)	very loud (ff)
B-flat$_4$mp	Falsetto	B-flat$_4$ (466 Hz)	moderately soft (mp)
F_4 ff	Falsetto	F_4 (349 Hz)	very loud (ff)
F_4 mp	Falsetto	F_4 (349 Hz)	moderately soft (mp)
C_4 ff	Falsetto	C_4 (262 Hz)	very loud (ff)
C_4 mp	Falsetto	C_4 (262 Hz)	moderately soft (mp)
Speech Loud	Chest	D_3 (147 Hz)	very loud
Speech Comf.	Chest	B-flat$_2$ (117 Hz)	comfortable loudness

performer, he has sung as a countertenor for the past 11 years performing early music using a falsetto-based singing technique to vocalize in the alto range. The subject's medical and recent health history were unremarkable and there was no history of speech or hearing disorders. The subject is a native speaker of General North American English and his speaking fundamental frequency was within normal limits. The subject's fundamental frequency range spanned from 69 Hz ($C\#_2$) to 392 Hz (G_4) in chest register and from 147 Hz (D_3) to 587 Hz (D_5) in falsetto register.

For each phonatory condition to be scanned, the subject was positioned comfortably in a supine position on the imaging table. His lower neck was supported and stabilized with a rolled towel. Head positioning was aligned before each set of scans such that the Frankfort plane was perpendicular to the imaging table and the anatomic midline centered. Each phonatory condition required approximately 40 seconds of actual scanner activation time. Total acquisition time depended on how long the subject could repeatedly prolong the target phoneme /ɑ/ at a particular phonatory condition without significant movement of articulatory structures. Including brief pauses for rest and breathing between vowel reiterations, during which scanning was interrupted, the total time required to image the vocal tract for each condition ranged from approximately 60-90 seconds.

The following system was devised to time the scanning interruptions when the subject needed to rest and inhale between vowel repetitions. Before each condition, the subject produced several trial utterances to gauge how many seconds he could produce steady prolongations without significant movement of vocal tract structures. When the utterance length was determined, the radiology technician timed pauses such that he stopped before the end of a vowel reiteration and reinitiated imaging as soon as the subject began the next reiteration, as monitored over the intercom.

EBCT Scanning Parameters

EBCT images were acquired with an Imatron C-150 scanner (Boyd and Lipton, 1983). Each volume set consisted of sixty contiguous, parallel, axial slices. Slice thickness was 3 mm. These scanned images encompassed the hard palate superiorly, the first tracheal ring inferiorly, the lips anteriorly, the posterior pharyngeal wall posteriorly, and the buccal walls to the left and right of vocal tract air space. Slice scan aperture was 100 msec. The field of view (FOV) for each slice was 21 cm and the image matrix was 512 x 512 pixels. The resolution in the plane of imaging (axial) was 0.410 mm, which is near the theoretical limit of the scanner's resolution. Scanning parameters are summarized on Table II.

The accuracy of the image acquisition and analysis procedures using the Imatron C-150 scanner has been assessed with a tubular phantom of known dimensions (Story, 1995). The phantom consisted of three connected sections of air-filled tubing placed in a closed water-filled plastic enclosure. Known and measured cross-sectional areas of the phantom differed by 1.8%-2.0%.

Table II.
Parameters for EBCT Image Acquisition Using the Imatron C-150 Scanner

slice thickness = 3 mm
slice scan aperture = 100 msec
FOV = 21 cm (field of view)
image matrix = 512 x 512 pixels
resolution = 0.410 mm/pixel

Image Analysis

Image analysis was accomplished in three stages, i.e. image segmentation, three-dimensional airway reconstruction, and airway measurement. These procedures were performed using UNIX based image display and measurement software called VIDA™ (volumetric image display and analysis), which was developed by Hoffman and colleagues (Hoffman *et al.*, 1992). Further information regarding all VIDA modules can be accessed via the Internet at http://everest.radiology.uiowa.edu. These image analysis techniques, as applied to vocal tract airway analysis, have been described in detail by Story (1995) and Story *et al.* (1996).

Vocal Tract Airway Segmentation

After the EBCT image sets were acquired, the image data were transferred via magnetic tape from the Imatron C-150 scanner to a UNIX-based computer workstation for analysis. The vocal tract was segmented, i.e. differentiated from surrounding tissue, using a seeded region growing technique whereby all airway voxels (three-dimensional pixels) were assigned a unique gray scale value (Hoffman *et al.*, 1983; Udupa, 1991). Prior to segmentation, a five percent reduction in gray scale values was made on the entire image set, assuring that the new gray scale assignment of the airway voxels would be the brightest value in the image. The threshold value for the border between tissue and air was determined by evaluating the brightness profile at an area of tissue to air transition and selecting the gray scale value that was 50% between the brightest area of tissue and the darkest area of the airway. Once this threshold was determined, actual segmentation could begin.

The process of segmentation consisted of the following sequence of steps. (1) A mouth termination plane was determined using the procedure defined by Mermelstein (1973) and "painted" on the midsagittal plane, as illustrated in Figure 1. When this plane was extended to all slices, it served as a delimiter for the anterior portion of the mouth,

Figure 1. Defining mouth termination plane by "painting" a bar to contain the oral air space (brightest pixels represent the airway).

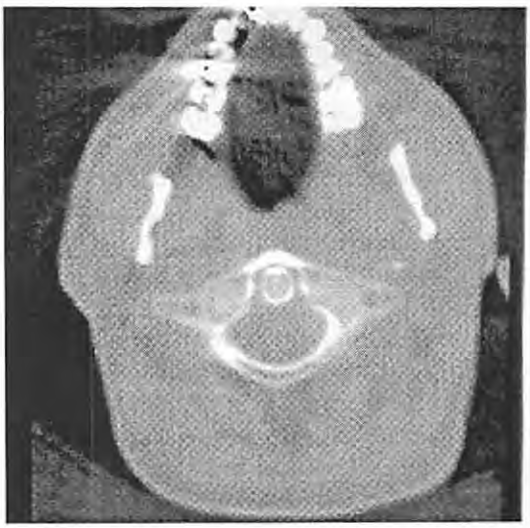


Figure 2. Segmentation: streaking artifacts on an axial slice of the oral cavity due to dental work.

effectively separating it from the oral airspace. (2) Because the corners of the lips were usually posteriorly oriented with respect to the painted plane, there were areas of oral air space which were not closed off from the air space outside of the vocal tract. This situation confused the automatic algorithm for assigning the vocal tract air space its unique color value, because the algorithm was designed to define enclosed regions. To address this, such "leaks" were contained by manually drawing a line using the same color value as the painted plane to close the gap on axial slices in which this occurred. This procedure was somewhat subjective and consequently introduced some amount of error in calculating the vocal tract air space in the anterior mouth region. (3) In addition to closing such airway leaks into the outer air space, manual editing was also required to address streak artifacts created by imaging of metallic dental fillings. Figure 2 illustrates the streaking artifact that occurred on a number of axial slices in the oral cavity. Knowledge of the subject's dental history aided in the procedure of outlining the air-tissue border manually, where streak artifact obscured the threshold value for automatic detection. This corrective editing may have contributed additional error in defining the vocal tract air space within the oral cavity, when such manual editing was necessary. Once these editing steps were completed, automatic seeded region growing based detection of vocal tract air space on each axial slice served to define the vocal tract region, whose voxel values were converted to one unique gray scale value.

Volume Reconstruction

Reconstruction of the vocal tract in three dimensions was accomplished using a process called shape based interpolation (Raya and Udupa, 1990; Udupa, 1991) on the segmented image set, yielding a stack of slices with the same voxel dimension (0.410 mm) along all three axes. The reconstructed 3-D image data from the shape based interpolated vocal tract was the basis for subsequent cross-sectional area measurements. The edges of the interpolated airway shape were also used to perform 3-D surface renderings of the vocal tract. Graphically represented as a three-dimensional object with the use of shading, surface renderings can be displayed at any number of angles or magnification levels. The display itself cannot be measured directly, but does allow the user to assess the quality of the segmentation procedure and to observe 3-D views of the vocal tract's outer shape.

Vocal Tract Measurement

To measure cross-sectional areas from the shape based interpolated data set, an algorithm originally designed to study the upper airway during sleep (Hoffman and Gefter, 1990; Hoffman *et al.*, 1992) was used. This algorithm computes the center line of a tube-shaped structure as well as cross-sectional areas from oblique sections perpendicular to the calculated centerline. Once the end points of the structure being analyzed (in this case the mouth and the glottis) are defined by the user, the iterative bisection algorithm computes a series of centroids along the airway, from which oblique cross sections can be calculated perpendicular to the local airway long axis. Tube length (in this case, vocal tract length), was quantified using methods described by Story *et al.* (1996).

The wave-reflection model of the vocal tract (Liljencrants, 1985; Kelly and Lochbaum, 1962) used in this study to compute acoustic wave propagation required the use of an even number of equally long vocal tract sections. The sampling frequency (44,100 Hz) used in this computer model also required that each vocal tract segment be

0.396825 cm thick. The total length of the final discretized vocal tract was consequently an even integer multiple of 0.396825 cm, e.g. a vocal tract modeled as 44 equivalent sections was 17.46 cm in length. The process of discretization consisted of (1) making a choice from the discretized vocal tract lengths available to best fit the measured data, (2) normalizing measured data to this length, (3) fitting the data to a cubic spline curve and (4) sampling the cubic spline curve at equally spaced intervals of 0.396825 cm.

The piriform sinuses could not be effectively measured using the iterative bisection method. However, since the axis of wave propagation in the piriform sinuses was roughly perpendicular to the axial plane, a pixel counting method was used to measure cross-sectional areas. Area functions were obtained by sampling at 2.05 mm intervals (every five slices) from the interpolated image data set. As was the case with cross-sectional areas of the main vocal tract, cross-sectional areas of the piriform sinuses were also discretized to fit an even number of cross- sectional cylinders for acoustic modeling. Because the combined acoustic effect of both piriform sinuses could be adequately modeled as a single tube, the area functions were combined into a single side branch to the main vocal tract airway. Area functions and frequency responses (from an impulse excitation) for each of the eight conditions were calculated, based on measurements from the 3-D vocal tract shapes.

Results and Discussion

Surface Renderings and Vocal Tract Area Functions

In each surface rendered airway, the air space of the vocal tract is represented graphically as an apparently solid gray scale object, with the articulatory speech organs removed. Sagittal, anterior and posterior views of the surface rendered vocal tract (with the piriform sinuses intact) for the falsetto register, high pitch (B-flat$_4$, 466 Hz), moderately soft condition are presented in Figures 3(a), 3(b), and 3(c). In the sagittal view, as shown in Figure 3(a), the rounded fist-like shape in the upper portion of the gray scale object is the air space within the oral cavity. The curved region rising gradually from the left and then falling towards right before it ends, depicts the contour of the air space beneath the velum and the arch of hard palate; the inferior surface of the fist-like shape, the contour of the air space superior to the tongue. The flat termination plane of the oral airway, per Mermelstein's procedure (1973), can be seen in the sagittal and anterior views, Figures 3(a) and 3(b) respectively. Descending from the fist-like shape inferiorly, the main vertical tube shape represents the air space in the oropharynx, laryngopharyx, glottis and upper trachea, respectively. The airway is smallest in the area of the glottis, which serves as the 0 cm reference point at the glottal end of the vocal tract for the purpose of measurement. In Figures 3(b) and 3(c), the two slender, vertically oriented, funnel-shaped objects lateral to the lower portion of the main airway are the piriform sinuses. In the posterior view in Figure 3(c), the two short vertical projections on the left and right sides of the oral cavity are the buccal cavities between the teeth/gum ridges and the mucosa of the buccal walls. The posterior view also illustrates in detail the transition of the posterior pharyngeal air space into the piriform sinuses.

Surface renderings (sagittal views) and graphs of corresponding area functions for the vocal tracts associated with each of the eight phonatory conditions are presented in Figures 4, 5, 6, and 7. Changes in cross-sectional area

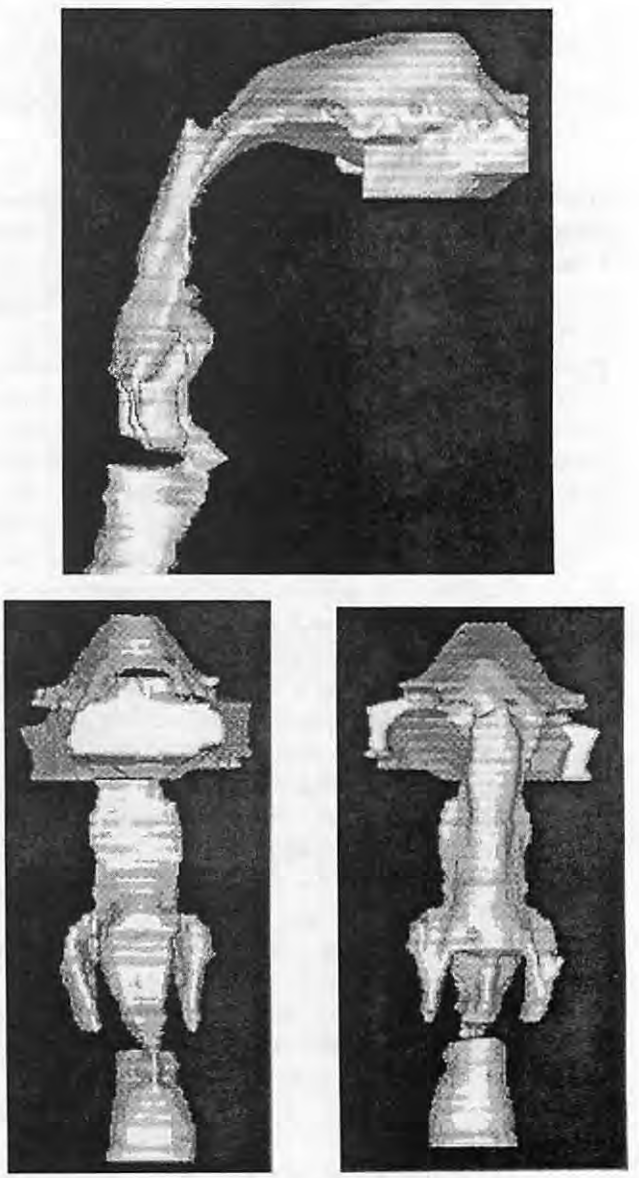


Figure 3. Surface rendering of shape based interpolated reconstructions of the 3-D vocal tracts, including the piriform sinuses, for the falsetto register, high pitch, moderately soft condition; (a-top) sagittal view, (b-bottom left) anterior view, and (c-bottom right) posterior view.

along the length of the vocal tract are illustrated in the graph to the right of each surface rendered vocal tract for each phonatory condition. Beginning at the origin, measured units on the x-axis represent distance in centimeters above the glottis. On the y-axis, measured units represent cross-sectional area in square centimeters.

A characteristic common to the vocal tract shapes across all eight phonatory conditions was a widening of the vocal tract airway above the glottis that starts at about 2 cm past the glottis, expands to its widest point at about 4 cm past the glottis and begins to narrow again at about 5 cm past the glottis. For the most part, this is a consequence of the changes in cross sectional area that occur as the piriform sinuses converge with the main vocal tract tube. This finding concurs with Story *et al.* (1996), who found that the location of this widening was consistent across all vowels. They suggested that this location's uniformity served to point out the consistency of the image analysis procedures in terms of defining the glottal termination. In the current study, the extent of this supraglottal widening varied with phonatory conditions. The absolute cross-sectional area at its widest point was greater, in some cases, significantly so, than those found by Story *et al.* (1996) or in previous research, to which they compared their findings (Baer *et al.*, 1991; Fant, 1960;

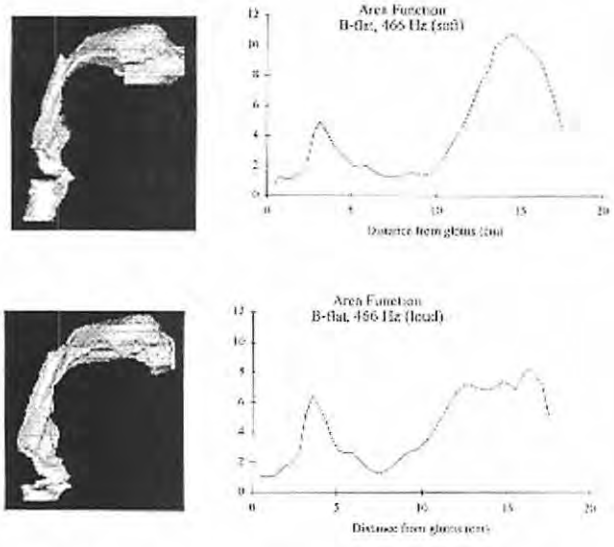


Figure 4. Surface rendered airways and area functions for falsetto register, high pitch conditions.

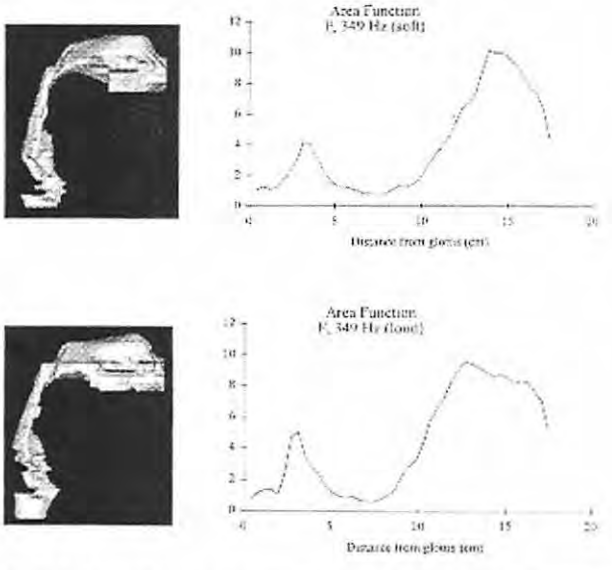


Figure 5. Surface rendered airways and area functions for falsetto register, medium pitch conditions.

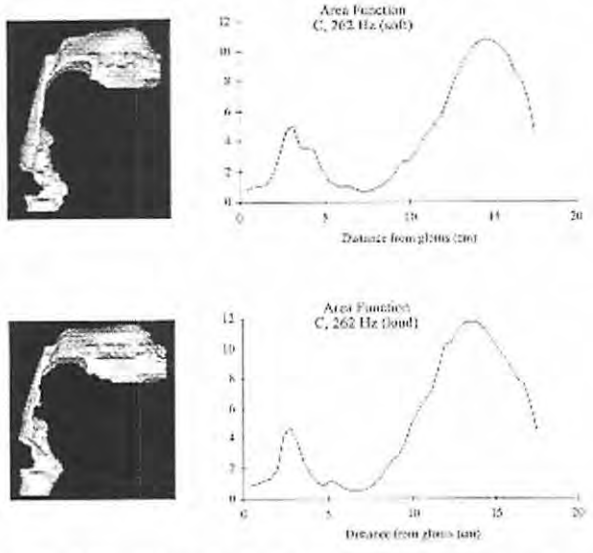


Figure 6. Surface rendered airways and area functions for falsetto register, low pitch conditions.

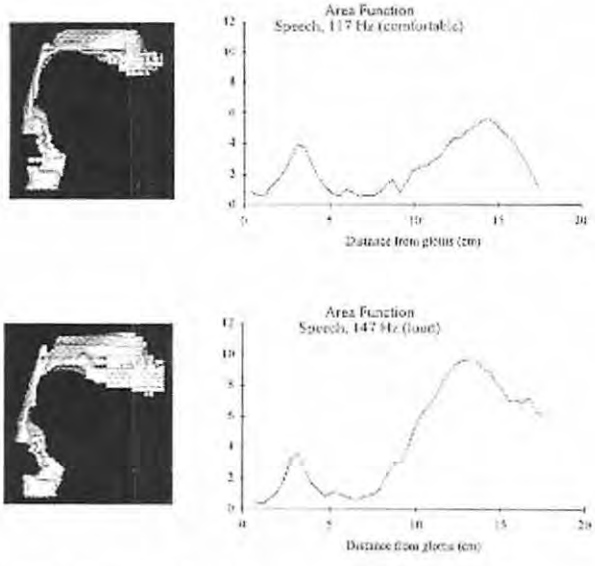


Figure 7. Surface rendered airways and area functions for chest register, speech pitch conditions.

Yang and Kasuya, 1994). The widest cross-sectional area of the supraglottal widening for the /ɑ/ vowel in these studies were approximately 1.1 cm^2 (Story *et al.*, 1996), 2.2 cm^2 and 2.9 cm^2 (Baer *et al.*, 1991), 4.1 cm^2 (Fant, 1960) and 1.9 cm^2 (Yang and Kasuya, 1994). In the current study this measure ranged from approximately 3.8 cm^2 for the loud speech condition in chest register to 6.2 cm^2 for the falsetto register, high pitched, very loud condition. For the speech condition and low pitch falsetto condition, the supraglottal widening reduced slightly in area as intensity increased. For the medium and high pitch falsetto conditions, the supraglottal widening increased in dimension.

Comparing Changes in Area Functions Due to Variations in Register, Pitch and Loudness

Changes in 3-D vocal tract configuration as a function of changes in vocal register, pitch and loudness can be assessed by comparing differences in area functions along the length of the vocal tract. The area functions associated with the vocal tract shape for the /ɑ/ prolongation in the comfortable speech condition are compared to those of the other phonatory conditions in Figure 8. In Figure 8(a), a comparison of comfortable (bold line) and very loud speech (narrow line) in chest register, the vocal tract gesture accompanying the change in vocal intensity was an overall increase in oral cavity area that begins approximately 7 cm from the glottis, just anterior to the vowel constriction (defining /ɑ/ vowel quality) in the oropharynx. At its widest point in the oral cavity, the cross sectional area almost doubled for the loud speech condition (from about 6 cm^2 to 10 cm^2) and the mouth opening increased significantly (from about 1 cm^2 to 6 cm^2). This occurred with a concurrent slight reduction in dimension of the supraglottal widening that occurred about 3.5 cm past the glottis.

Although the oral cavity dimensions are significantly greater in the low pitch falsetto condition as compared to comfortable speech (Figure 8 (b)), the within pitch/within register pattern of change in cross sectional areas as

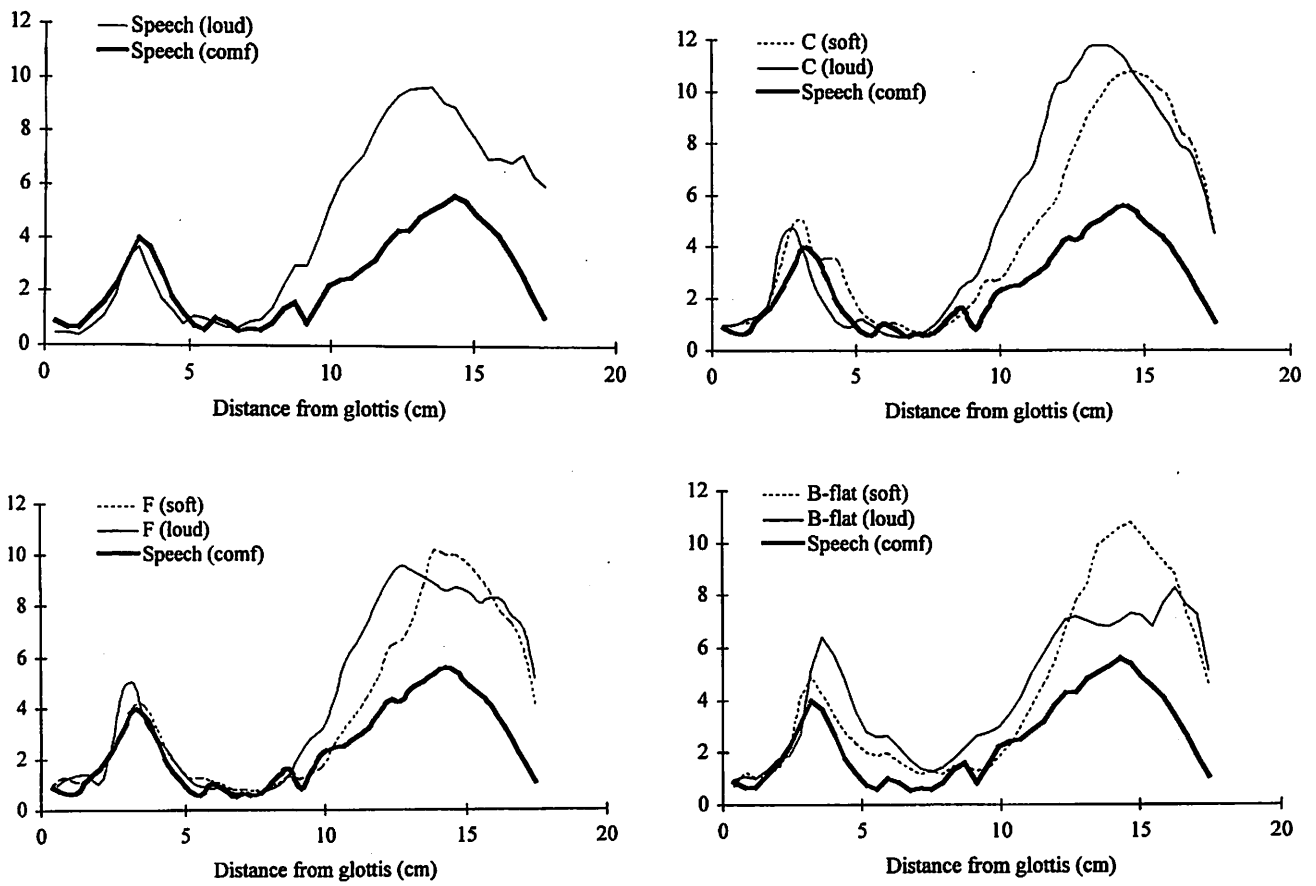


Figure 8. Comparisons of the area functions for chest register comfortable speech (bold line) and (a-top left) chest register loud speech (narrow line), (b-top right) falsetto register, low pitch (C_4 262 Hz), moderately soft (dashed line) and very loud (narrow line), (c-bottom left) falsetto register, medium pitch (F_4 349 Hz), moderately soft (dashed line) and very loud (narrow line), and (d-bottom right) falsetto register, high pitch (B-$flat_4$ 466 Hz), moderately soft (dashed line) and very loud (narrow line).

a function of loudness for the low pitch falsetto condition is similar to that for the speech conditions. The oral cavity increased in size as the supraglottal widening decreased in size. For falsetto phonations sung at the medium and high pitch levels, Figure 8(c) and 8(d), respectively, the oral cavity is significantly larger than for comfortable speech. The changes in vocal tract shape associated with soft (dashed line) versus loud intensity levels (narrow line) appear somewhat counterintuitive. The subject reduced rather than increased the volume of the oral cavity in the very loud condition. This reduction in oral cavity volume occurred with a simultaneous increase in the volume of the supraglottal widening in the lower pharynx. This anterior/posterior shift in relative volumes or cross-sectional areas may be a strategy to balance simultaneously the need for maintaining approximate vowel quality while preserving vocal timbre. At both the low and medium pitch conditions in falsetto, Figures 8(b) and 8(c), the volume of the oral cavity was only somewhat larger for the loud intensity condition. In contrast to the falsetto conditions, the oral cavity size and mouth opening are increased significantly in the loud speech condition.

The greatest contrasts in vocal tract configuration due to register, pitch and loudness can seen by comparing the area functions for speech phonation in chest register at a comfortable intensity level and those for high-pitched sung phonation in falsetto register at a very loud intensity level in Figure 8 (d). For speech phonation in chest register at a comfortable intensity level (bold line), the cross-sectional area peaked at 3.7 cm^2 at the widest point of the supraglottal widening, decreased to 1 cm^2 or less for the vowel constriction, increased to 5.8 cm^2 at the widest point in the oral cavity, then gradually reduced to about 1 cm^2 at the mouth termination. For high-pitched sung phonation in falsetto register at a very loud intensity level (narrow line), the cross-sectional area peaked at 6.4 cm^2 at the widest point of the supraglottal widening, decreased to less than 1.5 cm^2 for the vowel constriction, increased to 8.3 cm^2 at the widest point in the oral cavity, then closed down to an area of approximately 5 cm^2 at the mouth termination.

Quantitative Area Functions

As discussed in Section I.F., the "raw" area functions measured from the volumetric image data for the eight phonatory conditions were discretized for use in speech simulation. The process of discretization involved choosing the discretized vocal tract length (even multiples of vocal tract sections 0.396825 cm in length) that best fit the measured vocal tract lengths, normalizing measured data to this length, fitting the data to a cubic spline curve and sampling the cubic spline curve at equally spaced intervals of 0.396825 cm. Numerical area functions for the eight phonatory conditions based on the discretized vocal tracts are listed in Table III. The discretized vocal tract length that best fit all eight phonatory conditions was 17.46 cm.

Table III.
Vocal Tract Area Functions in Square Centimeters, at Equal Intervals of 0.396825 cm, for 8 Phonatory Conditions Varying Vocal Register, Pitch and Loudness

Section 1 is the Glottal End of the Vocal Tract

Section Number	B-flat$_4$ ff	B-flat$_4$ mp	F_4 ff	F_4 mp	C_4 ff	C_4 mp	Speech loud	Speech comf.
1	0.96	0.71	0.82	1.01	0.97	0.78	0.45	0.88
2	0.79	1.25	1.18	1.25	0.96	1.00	0.47	0.69
3	0.99	1.05	1.36	1.10	1.16	1.05	0.38	0.68
4	1.36	1.19	1.39	1.18	1.29	1.16	0.71	1.18
5	1.78	1.44	1.07	1.70	1.88	1.88	1.10	1.59
6	1.88	2.12	2.33	2.18	4.12	3.24	1.83	2.23
7	2.70	4.13	4.76	3.01	4.77	4.79	3.21	3.12
8	5.19	4.93	5.03	4.14	3.85	5.03	3.67	3.97
9	6.42	4.23	3.39	4.06	2.41	3.60	2.63	3.67
10	5.73	3.39	2.82	3.17	1.59	3.52	1.71	2.73
11	4.81	2.83	2.31	2.29	1.07	3.44	1.31	1.76
12	3.59	2.36	1.64	1.70	0.88	2.29	0.79	1.23
13	2.80	2.00	1.20	1.27	1.18	1.53	1.08	0.74
14	2.61	1.88	0.94	1.27	0.95	1.13	1.02	0.58
15	2.66	1.99	0.86	1.12	0.68	0.97	0.84	1.01
16	2.17	1.62	0.94	1.03	0.54	1.05	0.69	0.84
17	1.73	1.39	0.67	0.82	0.50	0.84	0.69	0.56
18	1.37	1.19	0.60	0.79	0.57	0.68	0.89	0.64
19	1.27	1.29	0.59	0.75	0.81	0.69	0.97	0.60
20	1.43	1.19	0.75	0.80	1.18	0.83	1.40	0.86
21	1.78	1.45	0.97	1.05	1.80	1.11	2.22	1.35
22	2.18	1.57	1.37	1.31	2.46	1.43	2.99	1.61
23	2.64	1.29	2.23	1.23	2.80	1.90	2.99	0.85
24	2.79	1.38	2.82	1.43	3.64	2.66	3.98	1.53
25	2.99	1.85	3.19	1.73	4.94	2.69	5.20	2.24
26	3.49	2.31	4.32	2.57	5.83	3.23	6.15	2.44
27	4.13	2.99	5.76	3.20	6.57	4.05	6.67	2.53
28	5.02	3.83	6.54	3.79	7.11	4.69	7.15	2.86
29	5.77	4.60	7.28	4.41	8.58	5.22	8.16	3.17
30	6.51	5.45	8.28	5.28	10.17	5.81	8.84	3.81
31	7.07	6.71	9.04	6.34	10.43	7.09	9.35	4.28
32	7.23	7.81	9.54	6.67	11.12	8.20	9.62	4.31
33	7.05	8.44	9.39	7.42	11.76	9.20	9.67	4.81
34	6.89	9.94	9.13	8.79	11.81	9.82	9.69	5.07
35	6.85	10.33	8.89	10.14	11.76	10.43	9.09	5.33
36	7.04	10.65	8.58	9.98	11.35	10.73	8.93	5.62
37	7.36	10.83	8.73	9.99	10.70	10.76	8.25	5.42
38	7.24	10.37	8.55	9.66	10.10	10.60	7.69	4.91
39	6.86	9.82	8.13	9.15	9.40	10.20	6.99	4.54
40	7.72	9.34	8.31	8.53	8.79	9.76	7.04	4.13
41	8.31	8.84	8.22	7.70	7.95	8.53	6.88	3.40
42	7.66	7.40	7.56	7.29	7.63	8.05	7.18	2.70
43	7.32	6.25	7.07	6.27	6.32	6.70	6.39	1.86
44	5.17	4.57	5.16	4.15	4.49	4.49	5.99	1.09
VT length	17.46 cm	17.46 cm	17.46 cm	17.46 cm	17.46 cm	17.46 cm	17.46 cm	17.46 cm

B-flat$_4$ = 466 Hz
F_4 = 349 Hz
C_4 = 262 Hz

ff = very loud
mp = moderately soft

Speech loud = speech, very loud
Speech comf. = speech, comfortable loudness

Table IV.
Vocal Tract Lengths Associated with Variations in Vocal Register, Pitch and Loudness Levels, and Length Differences Between Measured and Normalized Vocal Tract Lengths

Condition	Measured Length	Difference from Normalized Length
B-flat$_4$ ff	18.16 cm	+0.70 cm
B-flat$_4$ mp	17.02 cm	-0.44 cm
F_4 ff	17.48 cm	+0.02 cm
F_4mp	16.99 cm	-0.47 cm
C_4 ff	16.81 cm	-0.65 cm
C_4 mp	17.29 cm	-0.17 cm
Speech loud	16.55 cm	-0.91 cm
Speech comf	17.39 cm	-0.07 cm

ff = very loud
mp = moderately soft

Table V.
Piriform Sinus Area Functions in Square Centimeters, at Equal Intervals of 0.396825 cm, Expressed as a Single Branch for Acoustic Modeling Purposes

Section 1 Represents the Area Function of the Superior-Most Portion of the Piriform Sinuses

Section Number	B-flat$_4$ ff	B-flat$_4$ mp	F_4 ff	F_4 mp	C_4 ff	C_4 mp	Speech loud	Speech comf.
1	2.39	1.34	2.16	1.82	2.05	2.17	2.58	2.26
2	2.13	1.13	2.07	1.60	2.24	1.94	2.41	2.21
3	1.88	1.02	2.19	1.50	2.28	1.36	2.60	1.95
4	1.26	0.83	1.83	1.04	1.74	0.11	1.29	0.83
PS branch length	1.59 cm	1.59 cm	1.59 cm	1.59 cm	1.59 cm	1.59 cm	1.59 cm	1.59 cm

The measured vocal tract lengths for the eight phonatory conditions are presented in Table IV. Differences between measured lengths and the discretized length used in acoustic modeling (17.46 cm) are also noted on Table IV. In the medium and high pitch falsetto phonations, it is interesting to note that the change from soft to loud intensity was consistently associated with an increase in measured vocal tract length, while the opposite pattern occurred at speech and low falsetto pitch conditions. Increases in overall vocal tract length tend to decrease all formant frequencies uniformly and are associated with the perception of

Table VI.
Piriform Sinus Lengths Associated with Variations in Vocal Register, Pitch and Loudness Levels

Condition	L. Piriform Sinus	R. Piriform Sinus
B-flat$_4$ ff	1.76 cm	1.52 cm
B-flat$_4$ mp	1.68 cm	1.68 cm
F_4ff	1.76 cm	1.76 cm
F_4mp	1.60 cm	1.60 cm
C_4 ff	1.88 cm	1.88 cm
C_4 mp	1.15 cm	1.11 cm
Speech Loud	1.48 cm	1.48 cm
Speech Comf.	1.19 cm	1.19 cm

ff = very loud
mp = moderately soft

darker vowel coloring (Titze, 1994). This gesture can counteract the tendency for the larynx to rise and consequently shorten the vocal tract with increases in vocal effort.

Numerical area functions for the piriform sinuses used in acoustic modeling are listed on Table V. The measured lengths of the piriform sinuses are presented in Table VI. Some slight left-right asymmetries in piriform sinus length occurred when the superior-inferior alignment of the piriform sinuses was not completely perpendicular to the transverse imaging plane of the EBCT scanner. The main trend, with regard to changes in the 3-D shape of the piriform sinuses from soft to loud intensity within each pitch condition, was an increase in both length and cross-sectional areas.

Formant Structure

The relative locations of the first and second formants are most often associated with vowel discrimination; those of the third, fourth and higher formants being associated more with the perception of vocal timbre. The first four formants for each of the eight vocal tract shapes are summarized on Table VII. Formants were obtained with a wave-reflection vocal tract model mentioned in Section I.F. by using cross sectional area functions from the 3-D image data as input, and calculating its response to an impulse excitation. Because the configurations were a static composite of the slightly varying vocal tract shapes that occurred during imaging of vowel reiterations for a particular phonatory condition, these formants are, in a sense, an "average" of the formants for the numerous /ɑ/ prolongations.

For the comfortable speech condition, the first, second and third formants (F1, F2, F3) were 543 Hz, 993 Hz and 2585 Hz respectively. Normative average values for adult male F1, F2 and F3 for the spoken vowel /ɑ/ are

Table VII.
Formant Frequencies Associated with Variations in Vocal Register, Pitch and Loudness Levels

Pitch & Loudness	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
FALSETTO REGISTER				
B-flat$_4$ ff	601	1230	2751	3553
B-flat$_4$ mp	612	1139	2735	3611
F$_4$ ff	592	1030	2858	3728
F$_4$ mp	599	1102	2764	3643
C$_4$ ff	604	939	2729	3833
C$_4$ mp	582	1062	2738	3850
CHEST REGISTER				
Speech Loud (D$_3$)	682	1058	2740	3851
Speech Comf. (B-flat$_2$)	543	993	2585	3747

B-flat$_4$ = 466 Hz
F$_4$ = 349 Hz
C$_4$ = 262 Hz
D$_3$ = 147 Hz
B-flat$_2$ = 117 Hz

ff = very loud
mp = moderately soft
Speech Comf. = Speech, comfortable loudness
Speech Loud = Speech, very loud

730 Hz, 1090 Hz, and 2440 Hz, respectively (Peterson and Barney, 1952). The subject's values for both F1 and F2 are lower; almost 200 Hz lower for F1 and about 100 Hz lower for F2. F3 for the subject was more than 100 Hz above the norm. The F1 and F2 values produced by the subject are consistent with a more rounded /ɑ/ allophone, approaching /ɔ/. For the high pitched, very loud sung falsetto tone, F1, F2, and F3 were 601 Hz, 1230 Hz, and 2751 Hz, respectively. The subject's F1 for this condition was more than 100 Hz below normative values and the F2 and F3 were both higher. The F1 and F2 are consistent with a phonetically more neutral vowel quality, most likely due to the more open oral cavity.

In speech phonations in chest register at both comfortable and loud intensity levels, the vowel formants (F1, F2) produced by the subject were lower in frequency than normative formant values for adult male speakers (Peterson and Barney, 1952) for the vowel /ɑ/. The subject did, however, follow the reported tendency for F1 to be raised (10% or more) when increasing vocal intensity (Sundberg *et al.*, 1993): the F1 increased by 26% (from 583 Hz to 642 Hz). The F1 frequency was raised 4% in the loud condition for low pitch (C$_4$, 262 Hz) falsetto phonation. Contrary to this trend, the subject stabilized or slightly lowered F1 in sung falsetto phonations with increased vocal intensity at the medium (F$_4$, 349 Hz) and high pitch (B-flat$_4$, 466 Hz) conditions. The lowered F1 may have been part of the subject's vocal technique for effecting a perceptually darker (phonetically more rounded and/or centralized) vowel quality and warmer vocal timbre as pitch increased. The tendency for most speakers is to shorten the vocal tract with increased pitch by raising the larynx. This gesture increases F1 values and creates a brighter or strident vocal timbre. The lowered F1 also had the effect of creating a high energy bandwidth with a shallower spectral slope (locally) in the area between the fundamental and the second formant.

Conclusion

Volumetric imaging of the vocal tract using EBCT was used to document three-dimensional changes in vocal tract configuration during phonation, which occurred as a function of vocal register (chest or falsetto), pitch (low, medium and high in falsetto register, and speech) and loudness (soft versus loud) for a single male subject. The high resolution of the images acquired in this study were in part due to two factors which helped to minimize blurring due to movement artifact during image acquisition: (1) Imaging time was kept to a minimum with the use of EBCT techniques, and (2) extraneous movements during phonations were minimized by the participation of a subject whose vocal training lent itself to producing stable and consistent repetitions.

The 3-D image data were analyzed to obtain cross-sectional areas along the length of the vocal tract (glottis to lips) for eight phonatory conditions. These extracted area functions were used to identify formant frequencies for falsetto phonations, which are often not accurately extracted from the wide-bandwidth oral airflow signal using current LPC based inverse filtering techniques. Surface renderings of the three-dimensional vocal tract shapes provide qualitative information regarding changes in 3-D shape and relative spatial configuration of speech articulators across these phonatory conditions. Quantifications of vocal tract length, piriform sinus length, and cross sectional area functions from these image sets can be used in speech simulation applications. An analysis by synthesis approach to developing estimates of glottal adduction in falsetto phonation using the cross-sectional area functions, as extracted from the eight phonatory conditions in the current paper, will be presented in the future.

Although the results of the entire study are from one subject, a trained singer, the data may not be atypical of male speakers with a similar, but untrained falsetto phonation pattern, trained countertenors or other singers, who vocalize in the falsetto register. Steps towards a comprehensive description of falsetto phonation from a subject who has developed an optimal technique for vocalizing in the falsetto register across a range of frequencies and intensity levels may give us insight into a falsetto type that is not only reliable but associated with less risk for injury to the tissues of the vocal folds than other falsetto phonation types.

Table AI.
Radiation Exposure Summary

For 3mm slice thickness, skin dosage is 1.08 CGy

Organ dosages = skin dose multiplied by conversion factor

Bone marrow	SD x .08	(1.08 x 0.08 = 0.0864)
Lens (eye)	SD x .009	(1.08 x 0.009 = 0.00972)
Thyroid	SD x .07	(1.08 x 0.07 = 0.0756)

Human effective dose equivalent, $H_e = \sum W_i H_i$

W_i = weighting factor for organ i based on stochastic risk estimates

H_i = radiation dose to organ i

Organ		Weighting factor
Red Marrow		0.15
Thyroid		0.03
Bone (surfaces)		0.03
Remaining Tissues	5@	0.06

$H_e = \sum W_i H_i$ = (0.15 x 0.0864) + (0.03 x 0.0756) + (0.03 x .0864) + (0.06 x 0.00972)

for bone marrow + thyroid + bone surface + lens/eye

= 0.01296 + 0.002268 + 0.002592 + 0.0005832

= 0.0184032 CGy* per set of 3mm scans

including 8 scanning conditions and one scout scan total = 9 scans

9 sets of scans = 9 x 0.0184032 = 0.1656288 CGy or about **0.17 rem**

0.17 rem is well below the exposure limitations for the general public (0.3 rem/year). For comparison, a conventional chest CT = 3-4 rem, an upper GI exam (film and fluoro) = 2.1 rem, a cine CT x-ray exam = 0.14 rem.

* calculations verified by Dr. William Stanford, Department of Radiology, University of Iowa College of Medicine & University of Iowa Hospitals & Clinics (1992)

Acknowledgments

The authors would like to express their gratitude to Dr. William Stanford, Dr. Brad Thompson, and the members of the Division of Physiologic Imaging, Department of Radiology at the University of Iowa College of Medicine for their contributions to this study. This research 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.

Appendix

In order to keep levels of exposure to the ionizing radiation associated with EBCT imaging techniques at safe levels for this study, the total calculated human effective dose equivalent was kept well below the exposure limitations recommended by the National Council on Radiation Protection (NCRP, 1987) and the International Commission on Radiation Protection (ICRP, 1977). For the nine scans acquired in the current study (eight volume sets and a scout scan), the total human effective dose equivalent, calculated using NCRP and ICRP standards, was 0.17 rem. The exposure limitation for the general public set forth by the NCRP and ICRP is 0.3 rem/year. The calculations are summarized in Table AI.

References

Alley, R. (1982). *Peking Opera* (New World Press, Beijing).

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," J. Acoust. Soc. Am. **90**, 799-828.

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

Boone, D., and McFarlane, S. (1994). *The voice and voice therapy, 5th ed.* (Prentice-Hall, Englewood Cliffs, NJ).

Boyd, D. P., and Lipton, J. J. (1983). "Cardiac computed tomography," IEEE Proc. **71**, 298-307.

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

Colton, R. H., and Hollien, H. (1972). "Phonational range in the modal and falsetto registers," J. Speech Hear. Res. **15**, 708-713.

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

Fant, G. (1960). *The Acoustic Theory of Speech Production* (Moulton, The Hague).,

Frisell, A. (1964). *The tenor voice* (Bruce Humphries, Somerville, MA).

Giles, P. (1982). *The countertenor* (Frederick Muller, London).

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

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

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

Hoffman, E. A., Sinak, L. J., Robb, R. A., and Ritman, E. L. (**1983**). "Non-invasive quantitative imaging of shape and volume of lungs," Am. Physiol. Soc. 1414-1421.

Hollien, H. (**1974**). "On vocal registers," J. Phonetics, **2**, 125-143.

Hollien, H. (**1977**). "The registers and ranges of the voice," in *Approaches to Vocal Rehabilitation* edited by M. Cooper & H. C. Cooper (Charles C. Thomas, Springfield, IL), pp. 76-121.

Holmberg, E., Hillman, R., and Perkell, J. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am. **84** , 511-529.

Holmberg, E., Hillman, R., Perkell, J., Guiod, P.C., and Goldman, S.L. (**1995**). "Comparisons among aerodynamic, electroglottographic, and acoustical spectral measures of female voice," J. Speech Hear. Res. **38**, 1212-1223.

International Commission on Radiological Protection. (**1977**). *Recommendations of the international commission on radiological protection, ICRP Publication 26* (Pergamon Press, Oxford).

Kelly, J. L. and Lochbaum, C. C. (**1962**). "Speech synthesis," Proc. 4th Intern. Congr. Acoust., paper 642, 1-4.

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

Large, J. (**1972**). "The male operatic head register versus falsetto," Folia Phoniatrica **24**, 19-29.

Laver, J. (**1980**). *The phonetic description of voice quality* (Cambridge University Press, New York).

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

MacKerras, C. (**1983**). *Chinese theater, from its origins to the present day* (University of Hawaii Press, Honolulu).

Malm, W. P. (**1967**). *Music cultures of the Pacific, the Near East and Asia* (Prentice-Hall, Englewood Cliffs, NJ).

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

Mermelstein, P. (**1973**). "Articulatory model for the study of speech production," J. Acoust. Soc. Am. **53**, 1070-1082.

Moore, C.A. (**1992**) "The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images," J. Speech Hear. Res. **35**, 1009-1023.

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

Narayanan, S. S., Alwan, A. A., and Haker, K. (**1995**). "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am. **98**, 1325-1347.

National Council on Radiation Protection and Measurements. (**1987**). *Recommendations on limits for exposure to ionizing radiation, NCRP Report No. 91*. (National Council on Radiation Protection and Measurements, Bethseda, MD)

Perlman, A. (**1992**). Personal communication.

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 Hear. Res. **35**, 53-67.

Peterson, G. E., and Barney. H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175-184.

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

Scherer, K. R. (**1995**). "Expression of emotion in voice and music," J. Voice **9**, 235-248.

Stathopoulos, E. T., and Sapienza, C. M. (**1993**). "Respiratory and laryngeal function of women and men during vocal intensity variation," J. Speech Hear. Res. **36**, 64-75.

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

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging,"J.Acoust. Soc. Am. **100**, 537-554.

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

Titze, I. R. (**1988**). "A framework for the study of vocal registers," J. Voice **2**, 183-194.

Titze, I. R. (**1994**). Principles of voice production (Englewood Cliffs, NJ).

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

Tom, K. (**1996**). "Intensity control in male falsetto phonation: An analysis by synthesis approach," Ph.D. dissertation, University of Iowa.

Udupa, J. K. (**1991**). "Computer aspects of 3-D imaging in medicine: A tutorial," in *3D Imaging in Medicine,* edited by J. K. Udupa and G. T. Herman (CRC, Boca Raton).

Welch, G. F., Sergeant, D. C., and MacCurtain, F. (**1988**). "Some physical characteristics of the male falsetto voice," J. Voice **2**, 151-163.

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

NCVS Status and Progress Report - 14
September 1999, 115-121

An Investigation of a Modal-Falsetto Register Transition Hypothesis Using Helox Gas

Martin Spencer, B.A
Ingo Titze, Ph.D., Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

This study concerned the effect of the first subglottal formant (F_1') upon the modal-falsetto register transition in males and females. Phonations using air and a helium-oxygen mixture (helox) were used in a comparative study to tease apart possible acoustic and myoelastic contributions to involuntary register transitions. Recordings of first subglottal formant and its accompanying bandwidths, and the lower and upper shift point marking the outer boundaries of each abrupt register transition, were obtained via a neck-mounted accelerometer, and analyzed using spectrograms and power spectra on a K-5500 Sona-Graph. The four subjects had their hearing masked bilaterally with speech level noise in order to increase the likelihood of involuntary register transition via minimized auditory feedback. In three of the four test subjects registration was surmised to be primarily a laryngeal event; evidenced by the similar frequency dependency of voice breaks in both air and helox. It may be hypothesized that subglottal resonance influenced register transition in the fourth subject, as voice breaks rose with helox-induced phonation; however, this result did not reach statistical significance.

Introduction and Background

Over a century and a half ago, Müller (1837) described experiments during which certain lengths of sub- and supraglottal tubing coupled to excised larynges were found to cause the larynges to sputter, and if control conditions were sustained, to produce falsetto voice[1]. Acoustic variations in subglottal pressure (P_s) are known to be on the order of 40-60% of the mean DC subglottal pressure[2]. Vilkman and colleagues have more recently observed dramatic effects of acoustic-mechanical vocal source-tract interaction on vocal fold vibration, and state that without doubt there is an interaction of subglottal resonances upon vocal fold vibration[3].

Austin and Titze[4] have also investigated subglottal resonance effects. They reported that, in half of the excised larynges mounted to a pseudotrachea, the amplitude of vocal fold vibration increased as a function of in-phase acoustic subglottal pressure, and decreased as a function of out-of-phase acoustic subglottal pressure.

An earlier supposition by Hollien[5], later somewhat modified to apply to speech-only phonation[1], was that a voice register is a purely laryngeal event. Comparisons of phonation in helox and air by McGlone and Brown[6] supported laryngeal dominance in register transition over acoustic causes because shift points between chest and falsetto registers were found at similar fundamental frequencies, regardless of the nature of the expired medium. Švec et al. have recently concluded that chest-falsetto register transitions appear to be primarily influenced by the vibratory properties of the vocal folds evidenced by varying register transition frequencies in excised larynges phonating with fixed length air supply tubes[7]. Švec et al also argue that variability in the frequency range of register transitions de-emphasizes critical myoelastic balances as a triggering effect for registration. The different balances are possible, but not necessary for transitions to occur.

Analysis of vocal fold behaviors in terms of nonlinear dynamics has suggested that voice breaks may be caused by different attractors coexisting in the same parameter regions, and therefore even extremely small changes in parameters, such as muscle tension or subglottal pressures, may lead to desynchronization of principal vibratory modes of the vocal folds, including bifurcations and chaos[8,9,10].

It may be stated without too much controversy that the modal register in males and a modal-falsetto mixture in females are the typical phonatory postures used in speech[11,12].

The greater timbral richness of the modal register is largely due to increased thyroarytenoid contraction, resulting in a lax vocal fold cover and involvement of the vocal fold body in vibration. Vilkman and colleagues specify that, for modal register to occur, the vocal folds must assume a critical collision mass of sufficient vertical and longitudinal dimension[3]. This critical mass in vibration is formed by using ample subglottal pressure, medial fold compression and adjusting the biomechanical properties of the mucosa with vocal fold length. Falsetto register occurs at relatively higher pitches and is a natural consequence of an elongated cover containing a stiffened vocal ligament, and a lax muscle. The transition between modal and falsetto registers may be smooth, as in a trained classical singer, or abrupt, as in a yodeler or pubertal male speaker.

Titze[12] described two hypotheses about involuntary register transitions:

1. The maximum active thyroarytenoid stress hypothesis states that, when a maximal thyroarytenoid muscle contraction is reached, no further increase in F_o can be obtained until the TA muscle releases its stress and the vocal ligament becomes the main tension-bearing tissue. Unless the muscle and ligament tensions are perfectly balanced in transition, F_o and voice quality changes will be abrupt.
2. The subglottal resonance hypothesis states that involuntary register transitions may be the result of uneven assistance by tracheal pressures in vocal fold vibration. The modal-falsetto register shift may be triggered somewhere between maximal acoustic reinforcement of vocal fold vibration at ($F_o = 3/5\ F_1'$), and maximally impeded vibration (at $F_o = F_1'$). Acoustic subglottal pressure variations are superimposed onto a larger DC component, which is positive throughout the complete glottal cycle[13].

Measurements of subglottal formants and their accompanying bandwidths have led to somewhat different results (Table 1). Van den Berg[14] used canine and human cadavers to estimate F_1'=300 Hz and B_1'=120 Hz. Ishizaka and colleagues[15], using laryngectomy patients, found F_1' = 640 Hz and B_1'=155. Cranen and Boves[16] arrived at average values of F_1'=475 Hz and B_1'=275 Hz via frequency analysis of subglottal pressure signals obtained during normal speech production.

Cranen and Boves[17] analyzed subglottal pressure signals recorded during speech production via FFT and LPC and obtained an F_1' estimate of 510 Hz with a standard deviation of 26 Hz, and B_1' = 104 Hz. A modeling study was then used to investigate the influence of normal glottal leakage and vertical phase closure on discrepancies between their own F_1' measurements and data obtained by Ishizaka *et al.* Relatively small glottal leaks were found to cause the formants in the pressure signals to deviate appreciably.

Table 1.
Subglottal Formant Frequencies and Bandwidths

Investigators	Analysis Method	F_1' (Hz)	B_1' (Hz)
van den Berg[14]	canine/human cadavers	300	120
Ishizaka et al[15]	laryngectomy	640	155
Cranen and Boves[16]	speech production	475	275
Cranen and Boves[17]	speech production	510	104
The Current Study	pitch sweeps	538	218

Tracheal puncture is the most direct technique for measuring subglottal pressures, but also the most invasive for the subject[2]. Subglottal pressure may also may be measured by a pressure transducer built into a catheter, which is inserted through the glottis. Although still an invasive and uncomfortable procedure, it has been used successfully[16,17,18]. Transcutaneous measurement, using an accelerometer in the neck, is an indirect reflection of subglottal resonant activity. Although it involves a filtering effect through the neck tissues, it has the benefit of being non-invasive. Henke used this method to determine a subglottal resonance frequency of 530 Hz[19].

Materials and Methods

Helox gas was used as an experimental sound-propagating medium. Helox is a harmless gaseous mixture consisting of 20% oxygen (the same content as air) and 80% helium. Because of its lesser density than air, helox raises the speed of sound c and thereby the subglottal formant frequencies, since $F_n' \approx (2n - 1)(c/4L)$, in a resonator that is open on one end and closed on the other. Thus, the use of this gas in contrast with air was conceived to be a simple way of isolating the contributing influences of each of the two hypotheses for involuntary register transitions. Sundberg[20] has concluded that the differences in density between helox and normal air will have a negligible effect on the aerodynamics of vocal fold vibration, that is, the source characteristics.

Pitch sweeps (glissandi) were used to identify the formants and register shift points. It was felt that these gliding phonations have several advantages over steady vowels or musical scales for analysis. When produced at a moderate loudness, pitch sweeps may clearly reveal the presence of voice breaks between chest and falsetto registers, thus avoiding the perceptual subtleties of register identification via static fundamental frequencies[6]. Also the removal of a musical framework (specified notes) may help to trigger the lower and upper shift points of voice breaks in an unconstrained manner. Trendelenburg was an early pioneer in the investigation of a myoelastic genesis of register transitions who used "siren-like" gliding tones across the vocal compass[1].

Subjects

Two male subjects were M1, who had a brief history of choral singing, and M2, a professional baritone singer. Two female subjects were F1 and F2, both speech language pathology graduate students with choral experience, classified as second sopranos. All subjects were primarily chosen for their ability to sustain pitch sweeps over a moderate length of time at a high volume level, not necessarily for their singing experience. Also, each subject had informally demonstrated consistent involuntary pitch breaks in probe testing for experimental suitability. The subjects had vocal histories within normal limits and were examined by endoscopy for laryngeal health by a laryngologist. Subject M2 had a non-pathologic posterior glottal chink.

Instrumentation

Testing was performed with the subjects seated in an Industrial Acoustics Company Inc. (IAC) sound booth. Speech-weighted noise was presented to the subjects binaurally via a Grasson-Stadler GSI GI clinical audiometer and E.A.R. Tone 3A insert headphones. Masking was used to decrease the subjects' sensitivity to pitch control, thereby potentially increasing the likelihood of spontaneous and involuntary voice breaks. Voice breaks produced without masking may have been unintentionally regulated by a perceptual framework, such as musical intervals .

Skin acceleration recordings were made with a Vibro-meter accelerometer (Model 501 FB) affixed with double-sided adhesive tape to the epidermis covering the trachea midway between the supra-sternal notch and the cricoid cartilage. The accelerometer signal was amplified by a Vibro-meter Corp. (Model P16) power supply and routed through a Symetrix SX202 preamp into a Panasonic SV-3700 DAT deck.

Helox was initially administered to the subjects through a Hudson RCI rebreathing mask for a period of five minutes prior to the testing procedure. Subjects were encouraged to exhale forcefully to permit the helox to sufficiently infuse the residual capacity of the lungs. The rebreathing mask was switched with a standard nonrebreathing mask immediately prior to the testing procedure, thereby allowing the subjects relatively unimpeded sound radiation and self-control over oxygen intake requirements.

Tasks

An initial series of ten chest register pitch sweeps, each consisting of a continuous upward and downward slide, was recorded for the analysis of subglottal formants. The subjects were instructed to raise and lower pitch by the researcher (with hand signals) because their sense of pitch was masked. The subjects were not instructed to sing, and the resulting speech-like phonations were not characterized by a supported musical timbre, as in a sung *glissando*. In this part of the procedure, register transitions were avoided by keeping the sweeps in a lower F_o range. Continuously moving harmonic amplitudes could be recorded without quantal spectral changes induced by register transition

A second series of ten rising and falling pitch sweeps was expanded to encompass the transition into falsetto or middle register. Subjects were instructed to broaden their sensation of pitch range, or were conducted by the researcher to reach the higher pitches. The samples were later spectrographically analyzed to determine the lower and upper shift points of each instance of abrupt transition, and also the duration of the voice breaks.

The two series of pitch sweeps were then re-recorded using helox-induced phonation after an adequate period of helox respiratory exchange had occurred. Expired gases were not compositionally analyzed; however, a respiratory therapist had verified that the gases would be fully saturated with water in this procedure.

Formant Frequency and Bandwidth Measurements

The recorded accelerometer signals were analyzed at 125 ms intervals with a Kay Elemetrics DPS Sona-Graph (Model 5500). A combination of power spectrum, waveform, and intensity contour was displayed. The power spectrum utilized a Hamming window analysis with a 1024 point transform size, flat shaping, and no averaging. With this setup, it was possible to monitor the movements of individual harmonics as they passed through the F_1' region in both the ascending and descending segments of the pitch sweeps.

Subglottal formants and bandwidths were measured from the power spectrum sweep. Voice break data were located through colored spectrographic analysis using a 1kHz frequency range and 1s time axis on the K-5500. When an individual harmonic reached a local maximum (its magnitude rose maximally above the magnitude of both of its nearest neighbor) the harmonic frequency was chosen as F_1' and the bandwidth was estimated according to the following equation:

$$B_1' = (3dB)\frac{(f_3 - f_1)}{(A_2 - A_1)/2 + (A_2 - A_3)/2} \quad , \quad (1)$$

in which the resonance curve was linearized around the harmonic frequency f_2 as shown in Figure 1 (following page). The harmonic amplitudes A_1, A_2, and A_3 are also shown in the figure. For each subject, the mean F_1', B_1' and the standard deviations were calculated from all cases for which a local maximum could be detected.

Durations of the voice breaks, with lower shift point (LSP) and upper shift point (USP) frequencies were also extracted spectrographically using 1024 pt Hamming window analysis over a 1s duration.

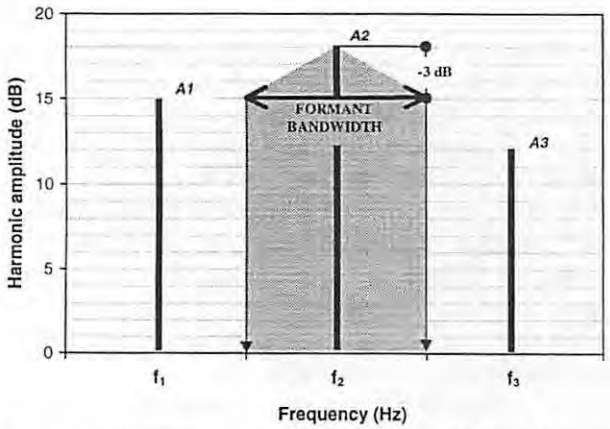


Figure 1. Subglottal formant calculation, where f_2 is the harmonic centered in the middle of the formant peak, and f_1 and f_3 are the adjacent harmonics. Bandwidth is determined by the averaged amplitude differences 3 dB down from the formant peak (A_2) relative to A_1 and A_3.

Results

The subjects reported that total bilateral masking levels of between 88-93 dB SPL were sufficient to eliminate auditory feedback. It was hoped that this reduction in feedback would reduce anticipation and awareness of register shifts, thereby contributing to increased involuntary pitch control. Of course, tissue conduction (bone and soft tissue) was not masked.

Mean durations of voice breaks were measured from initial establishment of the upper shift point (USP) to the eventual fade of the lower shift point (LSP) in ascending portions of the pitch sweeps, and visa-versa while descending. In all perceptually abrupt voice breaks, spectrography revealed the coexistence of both chest and falsetto registers for an average duration of 250 ms; a brief period of biphonation. Sequential power spectrum analysis (Figures 2a and 2b) showed the budding emergence of a secondary fundamental frequency which eventually reached a point of equal amplitude with the original F_o. Then the secondary F_o became more dominant as the primary F_o dropped away.

Falsetto phonation was consistently characterized by greater odd-numbered harmonic amplitudes, relative to even-numbered harmonics, when recorded with the accelerometer placed over the trachea. This suggests that the flow waveform had less of a flat portion in the closed phase[9].

Shift Points

Events in Figure 3 and Table 2 are listed in chronological sequence of occurrence; in the ascending half of a pitch sweep the LSP preceded the USP, and in the descending half the USP preceded the LSP. Abrupt voice breaks with a characteristic "crack" containing a leap interval occurred in 94% of the total obtained samples. The center point (CP) frequencies were also calculated.

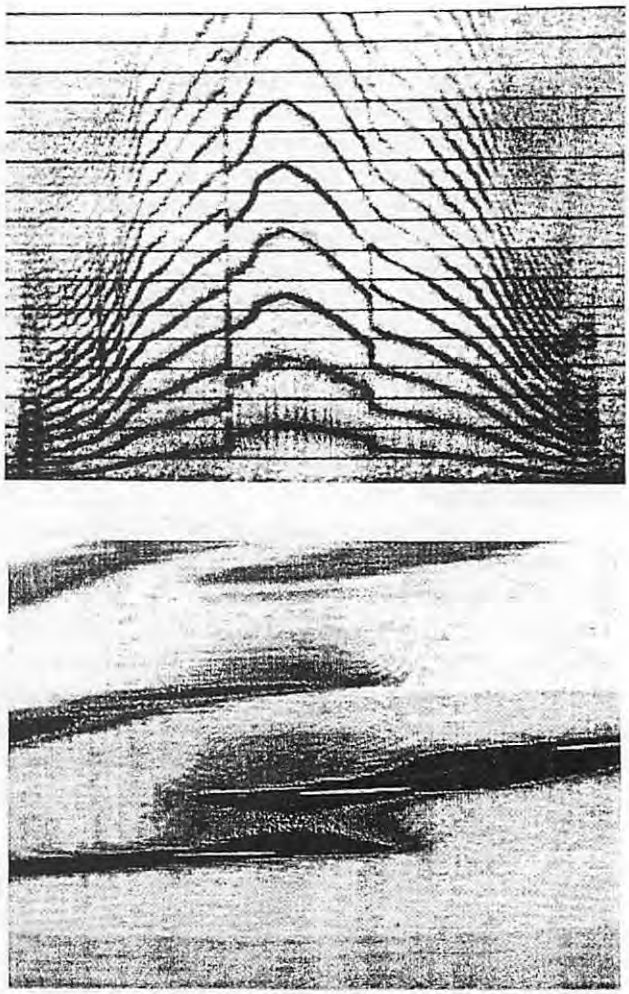


Figure 2. a-top) Narrow band spectrogram of a typical pitch sweep encompassing voice breaks in both ascending and descending portions (time axis = 8 seconds, 0-4000 Hz frequency range). b-bottom) A higher resolution image at the voice break occuring in the ascending portion more clearly showing an instance of biphonation (time axis = 500 ms, 0-1000 Hz frequency range).

Figure 3 shows that the LSP/USP measures for M2 and the two female subjects were remarkably similar in the two gaseous environment. M1 was the only subject to display an apparent rise in shift points induced by helox phonation; however, student t-tests did not identify any significant differences in within-subject air/helox comparisons at a 0.01 level of statistical probability.

M2's results reveal remarkably similar shift points in air and helox in both ascending and descending portions of the pitch sweeps. The approximate musical pitches marking the LSPs were E_4-F_4, and the USPs ranged from $F\#_4$-A_4. M2's leap intervals across the voice breaks were the largest in the group of subjects, averaging three semitones. M1's leap intervals varied from two to three semitones. The female subjects' LSPs varied between F_4-G_4, and their USPs

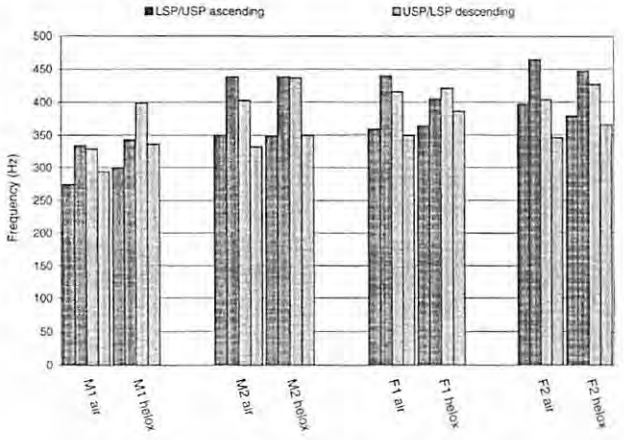


Figure 3. Mean upper and lower shift points (USP/LSP) occurring in ascending and descending portions of pitch sweeps in air or helox environments. Data measured via spectrographic analysis.

Table 2.
Observed Mean Lower and Upper Shift Points for Voice Breaks in Ascending and Descending Portions of Air and Helox Pitch Sweeps
Calculated Mean Voice Break Center Points in Italics. Standard Deviations in Brackets. All Figures in Hertz.

			Ascending Pitch			Descending Pitch		
Subject	Gas	Gender	LSP	Mean CP	USP	USP	Mean CP	LSP
M1	air	M	274 (23)	304	334 (37)	329 (38)	312	294 (37)
	helox		300 (19)	322	343 (26)	399 (93)	366	336 (81)
M2	air	M	350 (28)	379	438 (21)	403 (23)	367	332 (23)
	helox		349 (20)	394	438 (25)	437 (6)	394	350 (5)
F1	air	F	359 (8)	399	440 (19)	416 (18)	382	349 (21)
	helox		364 (57)	384	405 (61)	421 (19)	403	386 (17)
F2	air	F	397 (14)	431	464 (13)	404 (58)	375	346 (41)
	helox		379 (25)	413	447 (34)	427 (26)	396	365 (20)

ranged from G_4-$A\#_4$. F2 demonstrated a less than 1% difference between mean voice break center frequencies in helox and air. F1's mean voice break center frequencies actually showed a 4% drop, yet her F_1 increased by 77% (to be discussed next).

There are no profound male-female differences. For example, F1's air results (382 Hz and 399 Hz) are comparable to those of male M2 (367 Hz and 379 Hz), although he is a baritone and she is a second soprano.

Subglottal Formants

Table 3 lists estimates of F_1' for three of the four subjects, with the accompanying calculated bandwidths (B_1'). Subglottal formant measurements were not obtained for subject F2 due to an inability to locate F_1' through power spectrum analysis. In the other three subjects the approximate location of F_1' was fairly easily discerned; especially when observing a rapid series of spectral plots in which

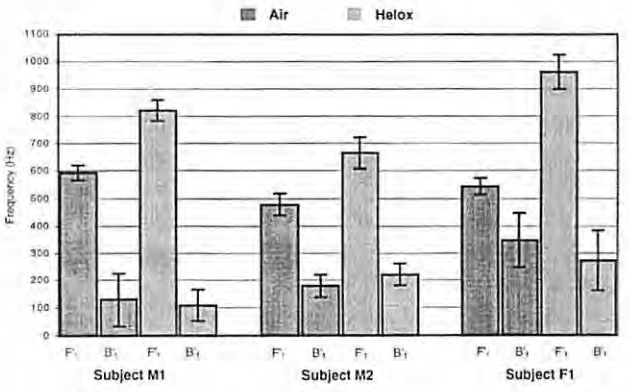


Figure 4. Means and standard deviation for first subglottal formant (F_1) and bandwidth (B_1) obtained with power spectrum analysis in air and helox induced pitch sweeps.

Table 3.
Observed First Subglottal Formant Frequency and Bandwidth Measures, Calculated air to Helox F_1' Ratios (R_F), and Fundamental Frequency to First Subglottal Formant Ratios
Standard Deviations in Brackets. Note That No Measurements of F_1' were Obtained for Subject F2.

Subject	Gas	Gender	B_1'	F_1'	F_1' ratios	F_0/F_1'
M1	air	M	593 (39)	129 (42)		0.5
	helox		822 (57)	108 (41)		0.4
					1.4	
M2	air	M	478 (28)	178 (96)		0.8
	helox		665 (24)	220 (59)		0.6
					1.4	
F1	air	F	544 (30)	347 (101)		0.7
	helox		962 (63)	272 (112)		0.4
					1.8	
F2	air	F	-	-		-
	helox		-	-		-

successive harmonics could be seen to rise and fall over F_1'. Clear accelerometer signals were obtained from subject F2, but when analyzed, gave no clues for possible F_1' location; there were simply no instances in which a center harmonic frequency could be seen to rise relative to both immediately adjacent harmonics.

Figure 4 is a plot of F_1' means and standard deviations. There are typical variances within and across subjects.

In air, the B_1' standard deviations ranged from 29-54% of their respective means, and in helox from 27-41%. These figures are comparable to the 24-58% variances previously reported in the literature for air-induced phonation[15,16,17].

Returning to Table 3, we make the note of the F_o/F_1' ratios. Titze[10] hypothesized that register changes should occur for values of F_o/F_1' between 0.6 and 1.0. Data obtained from M2 fall within this range, and therefore support Titze's claims. F_o/F_1' ratios calculated for the other two subjects, however, fall below the 0.6 value, giving less support to the hypothesis about subglottal interference.

Discussion

It is a finding of the present study that F_1' was not located at a single point, but rather shifted somewhat during the pitch glides (11-13% of the calculated F_1' value using air, and 7-14% using helox). The observed F_1' migration may have been caused by laryngeal elevation across the pitch sweep, affecting tracheal length, changing bronchial resonance due to varying lung volume, and perhaps also a frequency-dependent open quotient. A larger open quotient would also affect formant bandwidth, which may have deterred identification of F_1' in one subject[21,3,20].

Švec et al have observed a consistent hysteresis effect, in which upward voice breaks occurred at higher pitches and tensions than downward breaks[7]. In the current experiment, evidence of a hysteresis effect was consistently observed across all four subjects, but intriguingly less so in phonations using helox (Figure 3).

The formant frequency ratios (R_F) listed in Table 3 should reflect the difference in the speed of sound propagation in the two gases. The speed of sound in expired helox has been calculated to be approximately 1.8 times that of normal air[22]. Formant ratios should therefore follow this ratio. The lower values (1.4) for two subjects may be attributable to incomplete exchange of residual lung volumes of air, as well as a faulty mask seal and insufficient flow volume from the helox reservoir.

Subject M1 was the only subject whose results could be interpreted to be in some agreement with the subglottal resonance hypothesis, since both his F_1' and voice breaks were higher in helox-induced phonation. The average 12% increase in voice break center points, and the 39% increase in F_1', provided some evidence that a register transition may have been influenced by acoustic factors. However, t-tests did not validate a significant difference between the helox and air measures at a 0.01 level of probability.

Švec and Pešák [23] have noted that, at the moment of transition, while one register is being damped and the other is appearing, both will sound simultaneously for a brief period of time. Berry and colleagues[24] noted that in nonlinear systems different regimes of vibration may overlap, thereby forming "regions of coexistence"[7]. The existence of periods of biphonation, approximately 250ms in duration, was noted in this study to be consistent across all abrupt voice breaks.

Conclusions

Interactions between the first subglottal formant and register shift points have been investigated by making a comparison of pitch sweeps using helox and room air. Although subglottal formant frequencies were 40-80% higher with helox, the register shift points were similar for the two gases. In one of the four subjects, however, the subglottal resonance influenced the register transition with helox-induced phonation, though not to a statistically significant degree. The possibility exists that our subject pool did not include vocalists who have a typically "chesty" quality. Thus, a larger pool of subjects should be investigated. In general, however, the results suggest that registration is primarily a laryngeal event. Acoustic and myoelastic influences on register transition may exist along a continuum of blended interactions. It has been noted that there is a brief time interval of approximately 250 ms in which biphonation results from the simultaneous existence of modal and falsetto registers.

Acknowledgments

The authors thank Roger Chan, Ph.D., Eric Hunter, M.S., Emily Larson, M.A., Daniel McCabe, D.M.A., Brad Story, Ph.D., and Niro Tayama, M.D. for their assistance. Research was funded by Grant #: P60 DC00976 from the National Institute on Deafness and Other Communication Disorders.

References

1. Keidar A. *Vocal Register Change: An Investigation of Perceptual and Acoustic Isomorphism.* Ph.D. dissertation, Iowa City, Ia: University of Iowa; 1986.

2. Koike Y. Sub- and supraglottal pressure variation during phonation. In: Stevens KN, Hirano M, eds. *Vocal Fold Physiology.* University of Tokyo Press; 1981.

3. Vilkman E, Alku P, Laukkanen A-M. Vocal-fold collision mass as a differentiator between registers in the low-pitch range. *J Voice.* 1995;9:66-73.

4. Austin SF, Titze IR. The effect of subglottal resonance upon vocal fold vibration. *J Voice.* 1997;11:391-402.

5. Hollien H. On vocal registers. *J Phonetics.* 1974;2:125-143.

6. McGlone RE, Brown WS. Identification of the "shift" between vocal registers. *J Acoust Soc Am.* 1969;46:1033-1036.

7. Švec J, Schutte HK, Miller D. On pitch jumps between check and falsetto registers in voice: Data from living and excised human larynges. *J Acoust Soc Am.* In press.

8. Herzel H, Berry D, Titze IR, Saleh M. Analysis vocal disorders with methods from nonlinear dynamics. *J Speech Hear Res.* 1994;37:1001-1007.

9. Titze IR, Baken RJ, Herzel HP. Evidence of chaos in vocal fold vibration. In: Titze IR, ed. *Vocal Fold Physiology: Frontiers in Basic Science.* San Diego, Calif: Singular Publishing Group; 1992:143-182.

10. Titze IR. *Vocal Fold Physiology: Frontiers in Basic Science.* San Diego, Calif: Singular Publishing Company; 1993.

11. Sundberg J. *The Science of the Singing Voice.* DeKalb, IL: Northern Illinois University Press; 1987.

12. Titze IR. *Principles of Voice Production.* Englewood Cliffs, NJ: Prentice-Hall Inc.; 1994.

13. Titze IR. A framework for the study of vocal registers. *J Voice.* 1988;2:183-194.

14. van den Berg J. An electrical analogue of the trachea, lungs, and tissues. *Acta Physiol Pharmacol Neerlandica.* 1960;9:361-385.

15. Ishizaka K, Matsudaira M, Kaneko T. Input acoustic impedence mesuarements of the subglottal system. *J Acoust Soc Am.* 1976;60:190-197.

16. Cranen B, Boves L. Pressure measurements during speech production using semiconductor miniature pressure transducers: Inpact on models for speech production. *J Acoust Soc Am.* 1985;77:1543-51.

17. Cranen B, Boves L. On subglottal formant analysis. *J Acoust Soc Am.* 1987;81:734-746.

18. Miller DG, Schutte HK. Characteristic patterns of sub and supraglottal pressure variations within the glottal cycle. In: *Transcripts of the Twelfth Symposium: Care of the Professional Voice*, 1984; New York:The Voice Foundation.

19. Henke W. Signals from external accelerometer during phonation: Attributes and their internal physical correlates. *MIT Quarterly Progress Report.* 1974;114:224-231.

20. Sundberg J. Formants and fundamental frequency control in singing. An experimental study of coupling between vocal tract and vocal source. *Acoustica.* 1981;49:48-54.

21. Fujimura O, Lindqvist-Gauffin J. Sweep-tone measurements of vocal tract characteristics. *J Acoust Soc Am.* 1971;66:541,49 (2).

22. Stover WR. Technique for correcting helium speech distortion. *J Acoust Soc Am.* 1967;41:70-74.

23. Švec J, Pešák J. Voice breaks from modal to falsetto register. *Folia Phoniatr Logop.* 1994;46:97-103.

24. Berry D, Herzel H, Titze I, Story B. Bifurcations in excised larynx experiments. *J. Voice.* 1996;10:129-138.

NCVS Status and Progress Report - 14
September 1999, 123-130

Perceptual Voice Quality Changes Following Phonatory-Respiratory Effort Treatment ($LSVT_{(CM)}$) vs. Respiratory Effort Treatment for Individuals with Parkinson Disease

Christina Baumgartner, M.S.
Greeley, Colorado
Shimon Sapir, Ph.D., CCC-SLP
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Lorraine Ramig, Ph.D., CCC-SLP
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Department of Speech Language Hearing Sciences, The University of Colorado at Boulder

Abstract

Perceptual ratings of hoarseness and breathiness were used to assess the efficacy of two intensive methods for treating dysarthrophonia in individuals with idiopathic Parkinson disease. One method emphasized phonatory-respiratory effort (the Lee Silverman Voice Treatment, $LSVT_{(CM)}$) and the other emphasized respiratory effort alone (RET). Perceptual ratings were performed by two expert listeners based on random order presentation of the patients' pre- and post-treatment recordings of the "Rainbow Passage". The listeners were blinded to the patients and their treatment group. Statistically significant pre- to post-treatment improvement in hoarseness and breathiness was observed in the $LSVT_{(CM)}$ group but not in the RET group. The present findings are consistent with acoustic and physiologic findings reported previously, providing further evidence for the efficacy of the $LSVT_{(CM)}$.

Introduction

Parkinson disease (PD) is a progressive neurological disease caused by dopamine deficiency in the substantia nigra.[1] Approximately 1.5 million individuals in the USA suffer from PD, and at least 75% of them have voice and speech abnormalities related to their disease.[2-4] Some of these abnormalities, e.g., breathy phonation, hoarseness, reduced loudness, imprecise articulation, and reduced prosody are likely to affect speech intelligibility, which in turn may adversely affect the patient's communication and his or her social, economic, and psychological well being.[5-7]

Traditional speech therapy methods for dysarthric individuals with PD, typically administered once or twice a week and emphasizing articulation, rate and prosody intervention, have been largely ineffective.[8-9] In contrast, intensive voice therapy methods, administered almost daily and emphasizing simple phonatory effort tasks, have been found to produce favorable, long-term results in dysarthric individuals with PD.[10-12]

In 1987 Ramig and Mead 13 developed an intensive treatment program to improve vocal fold adduction and overall voice and speech production in individuals with Parkinson disease. The program, known as the Lee Silverman Voice Treatment, or $LSVT_{(CM)}$, is unique in that it focuses on a simple set of tasks designed to maximize phonatory and respiratory functions. This is done by instructing and constantly stimulating individuals to produce loud voice with maximum effort during sustained phonation and in various speech tasks. These individuals are also constantly reminded to monitor the loudness of their voice and the effort it takes to produce it.[14-15]

The loud and effortful phonatory tasks with the $LSVT_{(CM)}$ are aimed at improving respiratory drive, vocal fold adduction, laryngeal muscle activity and synergy, la-

ryngeal and supralaryngeal articulatory movements, and vocal tract configuration. These physiologic changes should improve voice quality and intensity, articulatory precision, prosodic inflection, resonance, and speech intelligibility. Such changes accompanying high loud phonation are expected based on similar effects seen in non-disordered speakers.[7, 16-17]

The implementation of high-effort, intensive phonatory-respiratory therapy is based on evidence from clinical practices in neurology and physical therapy[18-20] suggesting that when individuals with PD are pushed to higher effort levels, they learn to compensate for, or overcome, some of the deficits that underlie their motor impairment. This increase in effort level, especially when practiced intensively and daily, appears to help individuals with PD rescale or upscale the magnitude of their motor output, as seen in improved letterstroke in writing and stride length in walking[21-23]. In line with theories of motor learning[24-27], Ramig and her colleagues 28 have argued that intensive high-effort treatment of vocal functions, especially when coupled with proprioceptive feedback and auditory-vocal self-monitoring, should help individuals with PD rescale the magnitude of their speech motor output and habituate this level in conversation. Emphasis on self-monitoring is an important part of the treatment since motor deficits in individuals with PD appear to be related to factors such as impaired sensorimotor processing, inability to appropriately scale and regulate movement parameters, reduced ability to automatically execute learned motor plans, impairment in effort-demanding processes, and other abnormalities involving high level executive functions[23, 29-35].

Several acoustic, aerodynamic, stroboscopic and electroglottographic studies have demonstrated significant improvement in glottic closure, vocal fold vibratory movements, sound pressure level (SPL), and voice fundamental frequency (Fo) range and modulations following $LSVT_{(CM)}$.[28, 36-37] In a study where LSVT(CM) was compared with an alternative treatment method which emphasizes high respiratory effort (RET), the former method proved superior to the latter in improving SPL and phonatory function.[36] For example, whereas $LSVT_{(CM)}$ significantly increased vocal fold adduction and SPL, the RET produced inconsistent results, with some patients showing slight or moderate increases in SPL and vocal fold adduction, and others showing marked decreases in these variables post-treatment.

From a clinical standpoint, it is important to study perceptual changes that occur in voice and speech following treatment. Preliminary perceptual studies have already documented improvement in voice loudness, pitch inflection, speech intelligibility, and functional communication following $LSVT_{(CM)}$.[14, 28, 36] To our knowledge, perceptual changes in voice quality following $LSVT_{(CM)}$ have not been experimentally studied. Such information is important since abnormal voice quality can impact significantly on speech intelligibility and acceptability.[7]

The purpose of the present study was to assess the effects of $LSVT_{(CM)}$ and RET on the perception of voice quality in individuals with idiopathic PD. Two percepts of voice quality were chosen for the study — hoarseness and breathiness. These percepts characterize the voices in many individuals with PD. They are likely to be related to inadequate vocal fold adduction; suboptimal laryngeal muscle activation or synergy; muscle atrophy or fatigue; asymmetrical vocal fold tension or movements; stiffness or rigidity, or a combination of these.[37-42]

Acoustically, hoarseness and breathiness are characterized by excessive aperiodic energy superimposed on periodic (harmonic) energy.[43] In general, breathiness reflects transglottal air turbulence due to incomplete glottic closure, and hoarseness reflects irregular and asymmetrical vocal fold vibration.[43-45] Although hoarseness and breathiness are, to some degree, perceptually distinct, there is a considerable overlap between them in terms of their acoustic and physiologic characteristics.[46-48]

Studies in normal adult speakers have shown that as one increases loudness from normal to high levels, there are significant increases in SPL, subglottal air pressure, transglottal airflow, vibratory movement of the vocal folds, and glottal closure, and significant decreases in jitter and shimmer (acoustic indices of abnormal voice quality) as well as in breathiness or hoarseness.[16, 45, 49-54] There is also evidence that increasing voice intensity by vocal fold medial adduction and compression (through the contraction of the lateral cricoarytenoid, interarytenoid, and thyroarytenoid muscles) is more efficient than by increasing transglottal airflow or vocal fold tension.[55-58] Given these facts, and given the differential effects of $LSVT_{(CM)}$ and RET treatments on acoustic and physiologic measures mentioned above, one would expect to observe significant improvement in voice quality following $LSVT_{(CM)}$ and to a lesser degree following RET.

While anticipated that the $LSVT_{(CM)}$ will produce more favorable effect on voice quality, it is possible that as patients attempt to increase loudness, they may induce excessive tension in the vocal folds as well as excessive transglottal airflow. These physiologic changes may result in an increase in breathiness or hoarseness.[42, 59] Thus, one might argue that the RET may be a more reasonable approach to improve voice quality since it increases respiratory drive and is less likely to induce excessive vocal tension and abuse.

Another reason for comparing the two treatment methods was to assess the potential influences of extraneous variables such as the Hawthorne or placebo effects on treatment outcome. We reasoned that if the two treatment methods yielded different results, such differences are less likely to be related to extraneous effects and more likely to be attributed to treatment-specific mechanisms.

Method

Subjects

Initially, forty five individuals with idiopathic PD were included in the study and their voice was perceptually rated as described below. However, many of these individuals had a pre-treatment voice that was only mildly breathy or hoarse. To prevent possible ceiling effects, we decided to limit our study to only patients who had at least moderate amount of both breathiness and hoarseness pre-treatment. "Moderate amount" was defined as an average score of 25% on a perceptual rating scale (see below). Twenty individuals met this inclusion criterion. Of these individuals, thirteen (11M, 2F) were treated with the $LSVT_{(CM)}$ method and seven (5M,2F) were treated with the RET method. As can be seen in table A, on the average, the two treatment groups did not differ significantly ($p > 0.05$, $df = 1,18$; one-way ANOVA for unequal sample sizes) from each other in age (66.7 vs. 67.4, $F = 0.0349$), duration of PD since diagnosis (8.4 vs. 6.9, $F = 0.3747$), Hoehn & Yahr severity rating 60 (3.1 vs. 2.6, $F = 2.9321$), score on the motor examination section (section III) of the Unified Parkinson's Disease Rating Scale (UPDRS) 61 (34.5 vs. 28.2, $F = 0.9787$), Beck Depression Inventory (BDI) (9.5 vs. 10.7, $F = 0.2301$), and Montgomery Asberg Depression Rating Scale (MADRS) (8.2 vs. 7.3, $F = 0.0831$). These patients had been randomly assigned to their respective treatment group ($LSVT_{(CM)}$ or RET) after stratification on the variables just mentioned (excluding depression).

Treatment

Details of treatment have been described previously.[15, 36] Both forms of treatment were intensive with a duration of four one-hour sessions per week for four weeks. Both emphasized high effort levels and encouraged subjects to perform at maximum effort level throughout every session. Both types of therapies included repeated exercises for the first half of each session and speech tasks for the second half of each session.

The RET program targeted increased respiratory muscle activity to increase respiratory volumes and subglottal air pressure and loudness.[36] Treatment tasks included maximum inspiration and expiration,[62-63] maximum prolongation of /s/ and /f/ ,[64] and sustained intraoral air pressure using the Iowa Oral Performance Instrument [IOPI].[65] Subjects were encouraged to maximize their respiratory effort and were given frequent encouragement to "breathe" just prior to each of the sustained phonation, and during pauses while reading or performing conversational speaking tasks. Visual feedback of rib cage and abdomen excursions was provided to the individuals via NIMS Respigraph system PN SY03.[36] The RET did not address phonation or increasing phonatory effort, vocal fold adduction or voice pitch modulations.

The $LSVT_{(CM)}$ targeted increasing vocal effort to improve loudness. The main goal of the $LSVT_{(CM)}$ is to maximize phonatory efficiency by improving vocal fold adduction and overall laryngeal muscle activation and control.[15, 36] Special care is taken to increase vocal fold adduction without causing vocal hyper-adduction and strain. Upper extremity pushing and lifting tasks[66-67] during phonation were implemented to increase vocal fold adduction. Maximum prolongation of "ah" and maximum fundamental frequency range drills were completed. Subjects were encouraged to maximize phonatory effort and were given frequent encouragement to "think loud" during sustained phonation tasks, reading and conversational speaking tasks.[15,36] Attention was given to the respiratory system in the form of general reminders for subjects to take deep breaths "to be loud". The respiratory system was indirectly stimulated during all "think loud" speech tasks.[15, 36]

Table A.
Biographical and Medical Data —
Mean and Standard Deviation

	N, sex =	age	duration	H & Y	UPDRS	BDI	MADRS
$LSVT_{(CM)}$	11M, 2F	66.7 (7.8)	8.4 (5.4)	3.1 (0.6)	34.5 (9.9)	9.5 (4.4)	8.2 (4.3)
RET	5M, 2F	64.8 (9.6)	7.8 (4.8)	2.6 (0.8)	30.9 (14.5)	12.0 (6.0)	6.9 (5.6)

$LSVT_{(CM)}$ = Lee Silverman Voice Treatment
RET = Respiratory Effort Treatment
Duration = duration of PD since diagnosis
H & Y = Hoehn & Yahr stage of PD
UPDRS =Unified Parkinson's Disease Rating Scale (motor section III)
BDI = Beck Depression Inventory
MADRS = Montgomery Asberg Depression Rating Scale

Voice Recording Procedures

The subjects were recorded within 3 days before and after therapy. Pre-treatment and post-treatment voice recordings were made in an IAC sound-treated booth. As part of a larger protocol, subjects were instructed to read the "Rainbow Passage" [68] aloud at a comfortable rate and loudness. All subjects were seated with an AKG 410 microphone placed 8 cm from the lips. Pre-amplification was through an ATI-1000 amplifier. The data were recorded onto a Sony Digital PC-108M (DAT) eight-channel recorder.

Stimulus Tapes

Master tapes were created from the pre- and post-treatment voice recordings. The order of the recordings was randomized. Each voice sample of the "Rainbow Passage" was recorded two times consecutively when it was dubbed onto the master tapes. In order to offset the pos-

sible influence of loudness on the perceptual rating all the recorded samples were normalized, i.e., presented to the raters at the same SPL (70 dB), and at a distance of 50 centimeters.

Raters

Perceptual rating was done by two expert listeners, both speech pathologists, certified by the American Speech-Language-Hearing Association. These raters had 4.5 and 6.5 years of clinical and research experience post-masters degree, including extensive training and use of perceptual analysis of dysphonia and dysarthria. The raters had normal hearing and had no previous experience or interaction with the individuals in the present study.

Training

To familiarize themselves with the rating procedure, the raters were instructed on the operation of a computerized visual analogue rating scale. They were allowed 3 practice trials based on 3 random samples taken from the total sample. A visual analogue scaling procedure, which is based on an undifferentiated line, was used.[69] This scale has greater measurement sensitivity and produce more reliable results than an equal-appearing interval scale.[69]

Rating Procedures

Each rater listened to the master tapes in different orders in one session and rated the degree of breathiness and hoarseness they heard in each sample (see below). The raters listened individually free-field in an IAC sound-treated booth. They heard each voice sample consecutively and were allowed to listen to each sample as often as necessary. The computer monitor displayed a line for each percept. A new screen appeared for each "Rainbow Passage" sample. The listeners were instructed to click the mouse on the scale line at the point at which they perceived the extent of breathiness or hoarseness to be present. For each rating, the computer automatically calculated a percentage based on where the scale line was marked. Ratings were recorded on a spread sheet and subjected to statistical analyses.

Reliability Measures

Inter-rater reliability, measured with Cronbach Coefficient Alpha, was 0.84 and 0.77 for breathiness and hoarseness, respectively. Intra-rater reliability for breathiness and hoarseness was 0.83 and 0.69 for one rater and 0.86 and 0.73 for the other rater, respectively.

Statistical Analyses

Differences between means were tested for significance with a one-way ANOVA for unequal sample sizes (df = 1,18 for between group comparisons; df = 1,24 for $LSVT_{(CM)}$ pre- to post-treatment comparison; df = 1,12 for RET pre- to post-treatment comparison). Differences in frequency of occurrence were tested with chi square (X^2) analysis (two tails, with correction for continuity). The strength of a relationship between two variables was measured with a Pearson product correlation coefficient (r). Probability (p) values greater than 0.05 were considered non-significant.

Results

The results are summarized in Table B. This table shows the mean and standard deviation (in parentheses) of voice ratings and the percent change from pre- to post-treatment ratings for the two dependent variables (hoarseness and breathiness) in the $LSVT_{(CM)}$ and RET groups. Percent change was calculated for each subject by subtracting the post-treatment measure from the pre-treatment measure and dividing this difference by the largest of the two measures. The minus sign before the % change in hoarseness and breathiness indicates improvement. Figures 1 and 2 show the group means for hoarseness and breathiness pre- and post-treatment. Figure 3 and 4 show percent pre- to post-treatment change for each of the individuals in each of the treatment groups and for the two percepts. The seven individuals on the right side of Figures 3 and 4 were treated with the RET method, and the thirteen individuals on the left side were treated with the $LSVT_{(CM)}$ method.

As shown in Table B and Figures 1 and 2, the $LSVT_{(CM)}$ group made significant improvement in both hoarseness ($F = 12.3947$, $p = 0.005$) and breathiness ($F = 5.8882$, $p = 0.025$). Mean hoarseness rating decreased from 59.3 to 29.5, and mean breathiness rating decreased from 57.9 to 29.8. Mean percent pre- to post-treatment change in hoarseness and breathiness ratings was -54% and -59%, respectively. No statistically significant pre- to post-treatment changes were observed in the RET for hoarseness ($F = 2.0689$, $p > 0.05$) or breathiness ($F = 0.1920$, $p > 0.05$). In this group, mean hoarseness decreased from 52.4 to 39.8,

Table B.
Mean, Standard Deviation (in parentheses), and Percent Pre- to Post-Treatment Change in Perceptual Rating of *Hoarseness* and *Breathiness*

	HOARSENESS			BREATHINESS		
	pre	post	% change	pre	post	% change
$LSVT_{(CM)}$	59.3 (18.3)	29.5 (24.5)	-54% (30%)	57.9 (23.6)	29.8 (34.4)	-59% (38%)
RET	52.4 (15.5)	39.8 (17.2	-22% (38%)	43.2 (16.8)	38.9 (20.4)	-12% (30%)

$LSVT_{(CM)}$ = Lee Silverman Voice Treatment
RET = Respiratory Effort Treatment

and mean breathiness decreased from 43.2 to 38.9. Mean percent pre- to post-treatment change in hoarseness and breathiness ratings was -22% and -12%, respectively.

As can be seen in Figures 3 and 4, the majority of individuals in the $LSVT_{(CM)}$ showed marked improvement in both hoarseness and breathiness whereas only a few individuals in the RET made significant improvement. For example, eight (62%) of the thirteen individuals in the $LSVT_{(CM)}$ reduced hoarseness by at least 60% and eight (62%) reduced breathiness by at least 75% percent. Only one individual (14%) in the RET group reduced hoarseness by more than 60% and no individual reduced breathiness by more than 50%. These between-group differences are statistically significant for hoarseness (X^2 = 6.236 , p < 0.02) and breathiness (X^2 = 9.973, p < 0.01). One individual in the RET group (right-most data in Figures 3 and 4) increased hoarseness and breathiness markedly.

Percent change in hoarseness and in breathiness rating pre- to post treatment correlated highly and significantly in both the $LSVT_{(CM)}$ group (r = 0.9574, p < 0.001) and the RET group (r = 0.8125, p < 0.05). Correlation of hoarseness and breathiness ratings of the pre-treatment voice yielded r values of 0.654 and 0.355 for the $LSVT_{(CM)}$ and RET groups, respectively. Correlation of hoarseness and breathiness ratings of the post-treatment voice yielded r values of 0.9516 and 0.6863 for the $LSVT_{(CM)}$ and RET groups, respectively. Thus, the correlation between hoarseness and breathiness increased considerably from the pre- to post-treatment voice.

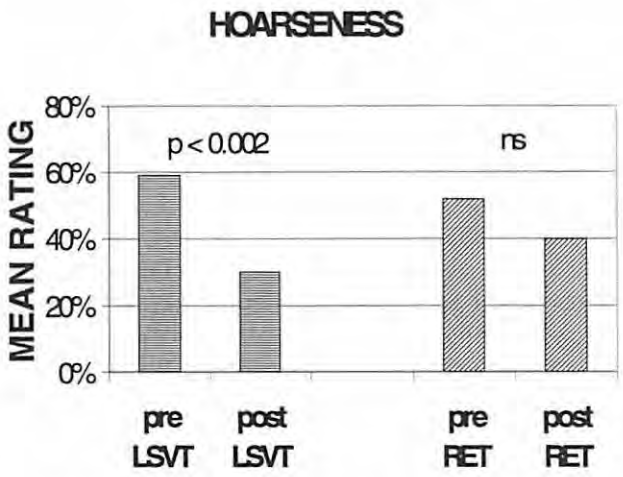


Figure 1. Mean, standard deviation (in parentheses), and percent change in hoarseness pre- and post-treatment in individuals with PD treated with $LSVT_{(CM)}$ (Lee Silverman Voice Treatment) versus RET (Respiratory Effort Treatment).

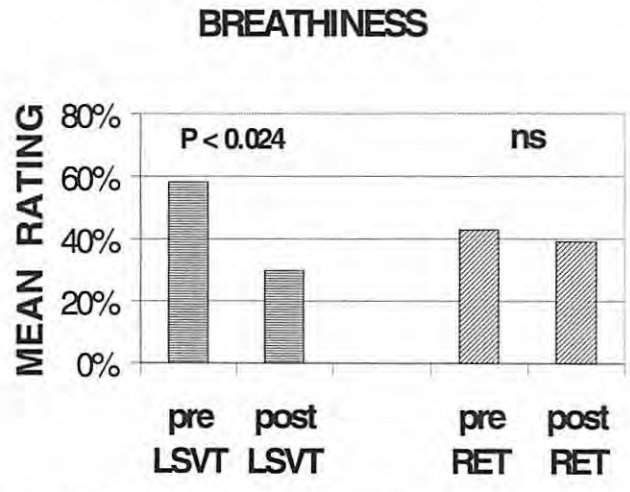


Figure 2. Mean, standard deviation (in parentheses), and percent change in breathiness pre- and post-treatment in individuals with PD treated with $LSVT_{(CM)}$ (Lee Silverman Voice Treatment) versus RET (Respiratory Effort Treatment).

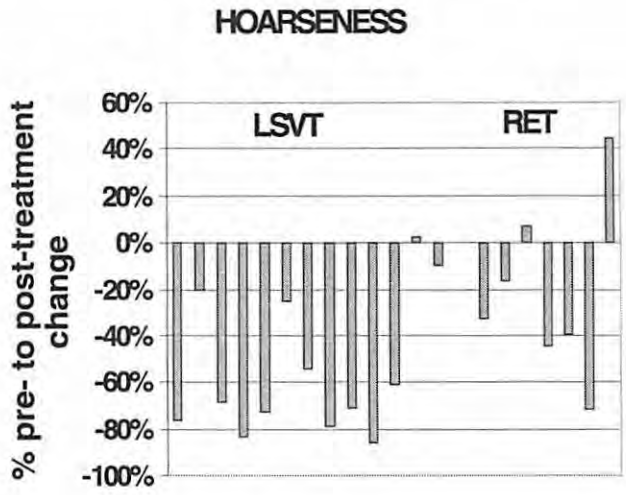


Figure 3. Pre- to Post-treatment percent change in hoarseness in individuals with PD treated with $LSVT_{(CM)}$ (Lee Silverman Voice Treatment) versus RET (Respiratory Effort Treatment).

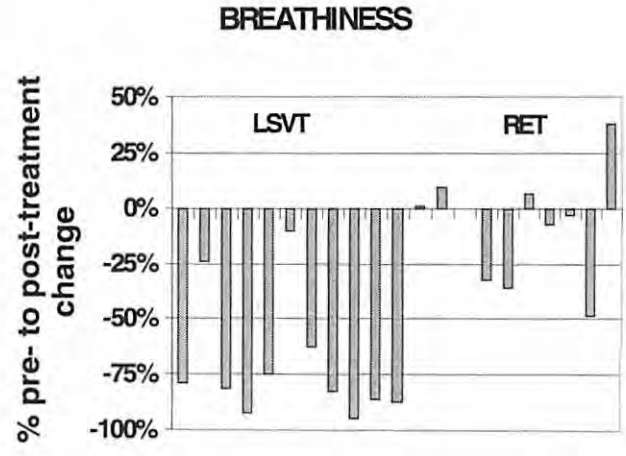


Figure 4. Pre- and Post-treatment ratings of breathiness in individuals with PD treated with $LSVT_{(CM)}$ (Lee Silverman Voice Treatment) versus RET (Respiratory Effort Treatment).

Discussion

In this perceptual study, individuals with PD treated with $LSVT_{(CM)}$ showed a significant reduction in hoarseness and breathiness post-treatment. This findings is consistent with acoustic and physiologic data reported earlier [14, 37, 70-72] attesting to the efficacy of the $LSVT_{(CM)}$ in the treatment of phonatory abnormalities associated with PD.

The lack of significant improvement in the RET group is more difficult to interpret. One possible explanation is that the sample size was too small and the variance across individuals too large to show significant effects. Another explanation is that the RET is less likely to produce favorable results because it does not improve vocal fold adduction and overall phonatory function to the extent that the $LSVT_{(CM)}$ does. The present findings and previous acoustic and physiologic data[16,71] support the latter explanation.

In viewing Figures 3 and 4 it is clear that the effects of either treatment were not uniform across individuals, although the trend was for the $LSVT_{(CM)}$ group to improve more and more consistently. We suspect that this non-uniformity is related in part to the degree and site of neuromuscular impairment and to the patients' level of motivation and cognitive ability to learn and use the vocal techniques taught. The non-uniformity may also reflect different magnitudes of glottal incompetence before and after therapy.

One of the reasons for comparing the two treatment methods was to assess the possible contribution of extraneous factors such as Hawthorne or placebo effects on treatment outcome. The fact that the two treatment methods did not produce the same results suggest that these results were most likely treatment-specific rather than related to extraneous factors. In the $LSVT_{(CM)}$ group, the main factors for improvement were likely to be increased vocal fold adduction and respiratory drive. In the RET group, the main factors for improvement were increasing respiratory drive, and perhaps increasing vocal fold adduction in some of the patients, probably due to overall increase in effort. Moreover, previous reports indicate that individuals with PD treated with RET show an increase in pause time, a decrease in utterance duration, and an increase in maximum duration of sustained phonation.[28,36] These changes suggest that the patients in the RET group indeed "learned" the respiratory tasks.

Another reason for comparing the effects of the two treatment methods was the concern that the $LSVT_{(CM)}$ may induce vocal hyperfunction or abuse, which could potentially deteriorate voice quality. The results of the present study and the results from previous studies[28,71] collectively suggest that the $LSVT_{(CM)}$ improves vocal function and quality and does not promote "hyperfunctional" phonation nor worsens voice quality.

In the present study, the percent pre- to post-treatment improvement in hoarseness and breathiness was highly correlated for both treatment groups. One explanation for this high correlation is that treatment had a similar effect on both percepts, i.e., when one percept improved the other percept improved to the same extent. Another explanation is that the raters perceived each voice sample as equally hoarse and breathy. The latter explanation is less tenable given that the correlation between hoarseness and breathiness ratings of the pre-treatment voice was relatively low, indicating that the two percepts were relatively distinct. Interestingly, the correlation between the two percepts was considerably higher in the post-treatment than the pre-treatment voice. This suggests that as voice quality improved with treatment, the two percepts were less distinguishable.

In this study we presented the speech samples to the raters at the same SPL level across patients. We did this to prevent influences of loud or soft speech on the perception of voice quality, as might be expected from previous studies with normal subjects.[7,16-17] Whether or not the normalization of SPL was necessary is an empirical question to be pursued in future studies.

It has been argued that perceptual measures of vocal quality are difficult to interpret, due to issues of validity and reliability.[73] Yet perceptual measures of voice and speech are necessary to assess the clinical significance of a particular treatment. This paradox is difficult to resolve. At present, the best way to assess the efficacy of a specific treatment is to provide converging evidence from both perceptual and more objective measures for the utility of the treatment under investigation. Thus, the present findings should be evaluated within the context of previous studies related to the efficacy of the $LSVT_{(CM)}$.[5,14,28,37,70-72] These studies, which include acoustic, physiologic and speech intelligibility measures, along with the present findings, provide consistent evidence for significant improvement in vocal function and speech production in dysarthric individuals treated with $LSVT_{(CM)}$.

Acknowledgements

The research presented here was supported by NIH grant No. R01DC01150. Our deepest gratitude is extended to the subjects who participated in this research study. The authors would like to thank Dr. Christopher Dromey, Dr. Kristin Baker, Dr. Susan Hensley and Ms. Annette Pawlas for their assistance during various parts of this research.

References

1. Hornykiewicz O, Kish S. Biochemical pathophysiology of Parkinson's disease. In Yahr M, Bergman K, eds. *Adv Neurol*. New York: Raven Press; 1986;45:19-34.

2. Logemann J, Fisher H, Boshes B, Blonsky E. Frequency and concurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *J Speech Hear Disord.* 1978;42: 47-57.

3. Sapir S, Pawlas A, Ramig L, Countryman S, O'Brien C, Hoehn M, Thompson L. Speech abnormalities in Parkinson Disease: Relation to medication, severity of motor impairment, duration of disease, medication, depression, gender, and age. *Neurology.* (in review).

4. Streifler M, Hofman S. Disorders of verbal expression in Parkinsonism. In Hassler R, Christ J, eds. *Advances in Neurology.* New York: Raven Press;1984.

5. Maclay S. Speech intelligibility gains in Parkinson's disease patients post voice treatment: Perceptual and acoustic correlates. Unpublished master's thesis, University of Colorado, Boulder; 1992.

6. Oxtoby M. *Parkinson's Disease Patients and their Social Needs.* London: Parkinson's Disease Society;1982.

7. Ramig L. The role of phonation in speech intelligibility: A review and preliminary data from patients with Parkinson's disease. In Kent R ed. *Intelligibility in Speech Disorders: Theory, Measurement and Management.* Amsterdam: John Benjamin;1992:119-155.

8. Hoberman S. Speech techniques in aphasia and Parkinsonism. *J Michigan State Med Societ.* 1958;57:1720-3.

9. LeDorze G, Doinne L, Ryalls J, Julien M, Oullet L. The effects of speech and language therapy for a case of dysarthria associated with Parkinson's disease. *Europ J Disord Communic.* 1992; 27: 213-24.

10. Johnson J, Pring T. Speech therapy and Parkinson's disease: a review and further data. *Brit J Disord Commun.* 1990;25:183-94.

11. Robertson S, Thompson F. Speech therapy in Parkinson's disease: A study of the efficacy and long-term effect of intensive treatment. *Brit J Disord Commun.* 1984;19: 213-24.

12. Scott S, Caird F. Speech therapy for Parkinson's disease. *J Neurol Neuros Psychiat.* 1983;46:140-4.

13. Ramig L, Mead C, Scherer R., Horii Y, Larson K, Kohler D. Voice therapy and Parkinson's disease: A longitudinal study of efficacy. Paper presented at the Clinical Dysarthria Conference, San Diego, CA. 1988.

14. Ramig L, Bonitati C, Lemke,J, Horii Y. Voice treatment for patients with Parkinson disease: Development of an approach and preliminary efficacy data. *J Med Speech-Lang Path.* 1994;2:191-209.

15. Ramig L, Pawlas A, Countryman,S. The Lee Silverman Voice Treatment ($LSVT_{(CM)}$): A Practical Guide to Treating the Voice and Speech Disorders in Parkinson Disease. Iowa City, IA: *National Center for Voice and Speech.* 1995b.

16. Dromey C, Ramig L Intentional changes in sound pressure level and rate: their impact on measures of respiration, phonation, and articulation. *J Speech Hear Res.* 1998;41:1003-18.

17. Schulman R. Articulatory dynamics of loud and normal speech. *J Acoust Soc Am.* 1989;85:295-312.

18. England A, Schwab R. The management of Parkinson's disease. *AMA Arch Int Med.* 1959;104:439-68.

19. Hallet M, Khosbin S. A psychological mechanism of bradykinesia. Brain. 1959;103:301-14.

20. McDowell F, Lee J, Sweet R. Extrapyramidal disease. In Baker A, Joynt R eds. *Clinical Neurology.* Philadelphia: Harper and Row; 1986:24-26.

21. Brooks V. The neural basis of motor control. New York: Oxford University Press;1986.

22. Muller F, Stelmach G. Scaling problems in Parkinson's disease. In: Requin J, Stelmach G, eds. *Tutorials in Motor Neuroscience.* Netherlands: Kluwer Academic Publishers;1991:161-74.

23. Stelmach G. Basal ganglia impairment and force control. In: Requin J, Stelmach G eds. *Tutorials in Motor Neuroscience.* Netherlands: Kluwer Academic Publishers;1991:137-48.

24. Adams J. A closed-loop theory of motor learning. *J Motit Behav.* 1971; 3:111-49.

25. Adams J. Use of the model's knowledge of the results to increase observer's performance. *J Human Mov Studies.* 1986;12:89-98.

26. Schmidt R. A schema theory of discrete motor skill learning. Psychol Rev. 1975;82: 225-60.

27. Schmidt R. *Motor control and learning.* Champaign, IL: Human Kinetic Publishers;1988.

28. Ramig L, Countryman,S, O'Brien C, Hoehn M, Thompson L. Intensive speech treatment for patients with Parkinson's disease: Short- and long-term comparison of two techniques. *Am Acad Neurol.* 1996;47:1496-1504.

29. Elias J, Treland J. Executive function in Parkinson's disease and subcortical disorders. *Semin Clin Neuropsychiat.* 1999;4:34-40.

30. Karayanidis F. Parkinson's disease: a conceptuatlization of neuropsychological deficits within an information-processing framework. *Biol Psych.* 1989; 29:149-79.

31. Marsden C. Function of the basal ganglia as revealed by cognitive and motor disorders in Parkinson's disease. *Can J Neurol Sci.* 1984;11:129-35.

32. Pillon B, Dubois B, Cusimano G, Bonnet A, Lhermitte F, Agid Y. Does cognitive impairment in Parkinson's disease result from non-dopaminergic lesions? *J Neurol Neurosurg Psychiat.* 1989;52: 201-6.

33. Schneider J, Diamond S, Markham C. Parkinson's disease: sensory and motor problems in arms and hands. *Neurology.* 1987;37: 951-6.

34. Taylor A, Saint-Cyr, J, Lang A. Frontal lobe dysfunction in Parkinson's disease. The cortical focus of neostriatal outflow. *Brain.* 1986;109:845-83.

35. Weingartner H, Burns S, Diebel R, LeWitt P. Cognitive impairments in Parkinson's disease: distinguishing between effort-demanding and automatic cognitive processes. *Psychiat Res.* 1986;11: 223-35

36. Ramig L, Countryman S, Thompson L, Horii Y. A comparison of two forms of intensive speech treatment for Parkinson disease. *J Speech Hear Res.* 1995; 38:1232-51.

37. Smith M, Ramig L, Dromey C, Perez K, Samandari R.. Intensive voice treatment in Parkinson's disease: Laryngostroboscopic findings. *J Voice.* 1995; 9:453-9.

38. Baker K, Ramig L, Luschei E, Smith M. Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging. *Neurology*. 1998; 51:1592-8.

39. Darley F, Aronson A, Brown J. Differential diagnostic patterns of dysarthria. *J Speech Hear Res*. 1969a;12: 246-69.

40. Darley F, Aronson A, Brown J. Clusters of deviant speech dimensions in the dysarthrias. *J Speech Hear Res*. 1969b;12:462-9.

41. Hanson D, Gerratt B, Ward P. Cinegraphic observations of laryngeal function in Parkinson's disease. *Laryngoscope*. 1984;94:348-53

42. Omori K, Slavit D, Matos C, Kojima H, Kacker A, Blaugrund S Vocal fold atrophy: quantitative glottic measurement and vocal function. *Ann Otol Rhinol Laryngol*. 1997;106: 544-51.

43. Toner M., Emanuel F, Parker D. Relationship of spectral noise levels to psychophysical scaling of vowel roughness. *J Hear Res*. 1990;33: 238-44.

44. Fukazawa T. el-Assuooty A, Honjo I. A new index for evaluation of the turbulent noise in pathological voice. *J Acoust Soc Am*. 1988;83:1189-93.

45. Gelfer M. A multidimensional scaling study of voice quality in females. *Phonetica*. 1993;50:15-27.

46. Dejonckere P, Remacle M, Fresnel-Elbaz E, Woisard V, Crevier-Buchman L, Millet B.Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. *Rev Laryngol Otol Rhinol*. 1996;Bord, 117:219-24

47. Leinonen L, Hiltunen T, Laakso M, Rihkanen H, Poppius H. Categorization of voice disorders with six perceptual dimensions. *Folia Phoniatr Logop*. 1997;49:9-20.

48. Wolfe V, Fitch J, Martin D. Acoustic measures of dysphonic severity across and within voice types. *Folia Phoniatr Logop* dica. 1997;49:292-9

49. Gelfer M. Fundamental frequency, intensity, and vowel selection: effects on measures of phonatory stability. *J Speech Hear Res*. 1995;38:1189-98.

50. Holmes L, Leeper H, Nicholson I. Laryngeal airway resistance of older men and women as a function of vocal sound pressure level. *J Speech Hear Res*. 1994;37:789-99.

51. Koike Y. Application of some acoustic measures for the evaluation of layrngeal dysfunction. Stud *Phonolog*. 1973;7:17-23.

52. Orlikoff R. Vocal stability and vocal tract configuration: An acoustic and electroglottographic investigation. *J Voice*. 1995;9:173-181.

53. Orlikoff R, Kahane J. Influence of mean sound pressure level on jitter and shimmer measures. *J Voice*. 1991;5:113-9.

54. Sodersten M, Hertegard S, Hammarberg B. Glottal closure, transglottal airflow, and vice quality in healthy middle-age women. *J Voice*. 1995;9:182-97.

55. Berke G, Hanson D, Gerratt B, Trapp T, Macagba C,Natividad M. The effect of air flow and medial adductory compression on vocal efficiency and glottal vibration. *Otol Head Neck Surg*. 1990;102:212-8.

56. Nasri S, Sercarz J, Azizzadeh B, Krieman J, Berke G. Measurement of adductory force of individual laryngeal muscles in an vivo canine model. *Laryngoscope*. 1994;104:1213-8.

57. Slavit D, McCaffrey T. Regulation of phonatory efficiency by vocal fold tension and glottic width in the excised canine larynx. *Ann Otol Rhino Laryngol*. 1991;100:668-77.

58. Tang J, Stathopoulos E. Vocal efficiency as a function of vocal intensity: a study of children, women, and men. *J Acoust Soc Am*. 1995;97:1885-92.

59. Hillman R, Holmberg E, Perkell J, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: an experimental framework and initial results. *J Speech Hear Res*. 1989;32:373-92

60. Hoehn M, Yahr M. Parkinsonism: onset, progression and mortality. *Neurology*. 1967;17:427-442.

61. Fahn S, Elton R, Committee, M. O. T. U. D. *Recent developments in Parkinson's disease*. (Vol. 2). New York: Macmillan Press;1987.

62. Hardy J. Cerebral Palsy. Englewood Cliffs: Prentice-Hall;1983.

63. Netsell R, Rosenbeck J. Treating the dysarthrias. In Netsell R ed. *A neurobiologic view of speech production and the dysarthrias*. San Diego: College-Hill Press;1986.

64. Leith D, Bradley M. Ventilatory muscle strength and endurance training. *J Appl Phys*. 1976;41:508-16.

65. Robin D, Goel A, Somodi L, Luschei E. Tongue strength and endurance: Relation to highly skilled movements. *J Speech Hear Res*. 1992;35:1239-45.

66. Aronson A. Clinical Voice Disorders. New York: Thieme-Stratton; 1990.

67. Froeschels E, Kastein S, Weiss D. A method of therapy for paralytic conditions of the mechanisms of phonation, respiration and glutintaion. *J Speech Hear Disord*.1955;20:3645-70.

68. Fairbanks G. *Voice and Articulation Drillbook*. New York: Harper and Brothers;1960.

69. Kreiman J, Gerratt B, Kempster G, Erman A, Berke G. Perceptual evaluation of voice quality: Review, tutorial and a framework for future research. *J Speech Hear Res*. 1993;36:21-40.

70. Brosovic G. Voice therapy and Parkinson disease: Measures of vocal fold adduction. Unpublished undergraduate honor's thesis University of Colorado Boulder;1994.

71. Dromey C, Ramig L, Johnson A. Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. *J Speech Hear Res*. 1995;38:751-63.

72. Ramig L, Dromey C. Aerodynamic mechanisms underlying treatment-related changes in SPL in patients with Parkinson disease. *J Speech Hear Res*. 1996;39:798-807.

73. Krieman J, Gerratt B. Validity of rating scale measures of voice quality. *J Acoust Soc Am*. 1998;104:1598-1608.

NCVS Status and Progress Report - 14
September 1999, 131-140

Intensive Voice Treatment (LSVT®) for Individuals With Parkinson's Disease: A Two-Year Follow-Up

Lorraine Ramig, Ph.D., CCC-SLP
Department of Speech Language Hearing Sciences, The University of Colorado-Boulder
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Shimon Sapir, Ph.D., CCC-SLP
Stefanie Countryman, M.A., CCC-SLP
Annette Pawlas, M.A., CCC-SLP
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Christopher O'Brien, M.D.
Colorado Neurological Institute, Englewood, Colorado
Margaret Hoehn, M.D.
Denver, Colorado
Leatitia Thompson, Ph.D.
Department of Neuropsychology, The University of Colorado Health Science Center

Abstract

The purpose of this study was to assess long-term (24 months) effects of two intensive treatment methods for speech and voice deficits in individuals with idiopathic Parkinson disease (PD). One method emphasized high phonatory-respiratory effort (the Lee Silverman Voice Treatment, LSVT®) and the other emphasized high respiratory effort alone (RET). Acoustic analyses of voice intensity and voice fundamental frequency inflection were used to assess changes in vocal function following treatment. The LSVT® was significantly more effective than the RET in improving vocal function and maintaining this improvement at two years follow-up. The present findings, along with others, provide additional evidence for the efficacy of the LSVT® in the treatment of voice and speech disorders in individuals with PD and other neurologic disorders.

Introduction

Approximately 1.5 million individuals in the USA suffer from Parkinson disease (PD). Of these individuals, at least 75% have voice and speech abnormalities related to their disease.[1-2] Some of these abnormalities, e.g., breathy phonation, hoarseness, reduced loudness, imprecise articulation, and reduced prosody are likely to affect speech intelligibility and oral communication, which in turn may adversely affect social, economic, and psychological well being.[3-5]

The physiological and neuropathological mechanisms underlying voice and speech deficits in individuals with PD are yet to be determined (see Sapir S, Pawlas A, Ramig L, Countryman S, O'Brien C, Hoehn M, Thompson L. Speech and voice abnormalities in Parkinson Disease: Relation to severity of motor impairment, duration of disease, medication, depression, gender, and age. Neurology; submitted). Voice abnormalities in individuals with PD have been attributed to inadequate vocal fold adduction; reduced laryngeal muscle activation or synergy; muscle atrophy or fatigue; asymmetrical vocal fold tension or movements; stiffness or rigidity of the vocal folds and/or respiratory muscles; or a combination of these.[6-12] Voice and speech abnormalities in individuals with PD have also been attributed to neurocognitive, neuroaffective, psychomotor, and other higher level cerebral dysfunction[24] (also see Sapir S, Pawlas A, Ramig L, Countryman S, O'Brien C, Hoehn M, Thompson L. Speech and voice abnormalities in Parkinson Disease: Relation to severity of motor impairment, duration of

disease, medication, depression, gender, and age. Neurology; submitted).

Traditional speech therapy methods for dysarthric individuals with PD, typically administered once or twice a week and emphasizing articulation, rate and prosody intervention, have been largely ineffective.[13-18] In contrast, intensive voice therapy methods, administered almost daily and emphasizing simple phonatory effort tasks, have been found to produce favorable, long-term results in dysarthric individuals with PD.[19-21]

In 1987 Ramig and Mead [22] developed an intensive treatment program to improve vocal fold adduction and overall voice and speech production in individuals with Parkinson disease. The program, known as the Lee Silverman Voice Treatment, or LSVT®, is unique in that it focuses on a simple set of tasks designed to maximize phonatory and respiratory functions. This is done by instructing and constantly stimulating individuals to produce loud voice with maximum effort during sustained phonation and in various speech tasks. These individuals are also constantly reminded to monitor the loudness of their voice and the effort it takes to produce it.[23-24]

The loud and effortful phonatory tasks with the LSVT® are aimed at improving respiratory drive, vocal fold adduction, laryngeal muscle activity and synergy, laryngeal and supralaryngeal articulatory movements, and vocal tract configuration. These physiologic changes should improve voice quality and loudness, articulatory precision, prosodic inflection, resonance, and speech intelligibility. Such changes accompanying high loud phonation are expected based on similar effects seen in non-disordered speakers. [5,25-26]

The implementation of high-effort, intensive phonatory-respiratory therapy is based on evidence from clinical practices in neurology and physical therapy [27-29] suggesting that when individuals with PD are pushed to higher effort levels, they learn to compensate for, or overcome, some of the deficits that underlie their motor impairment. This increase in effort level, especially when practiced intensively and daily, appears to help individuals with PD rescale or upscale the magnitude of their motor output, as seen in improved letterstroke in writing and stride length in walking.[31-33] In line with theories of motor learning, [33-36] Ramig and her colleagues [37] have argued that intensive high-effort treatment of vocal functions, especially when coupled with proprioceptive feedback and auditory-vocal self-monitoring, should help individuals with PD rescale the magnitude of their speech motor output and habituate this level in conversation. Emphasis on self-monitoring is an important part of the treatment since motor deficits in individuals with PD appear to be related to factors such as impaired sensorimotor processing, inability to appropriately scale and regulate movement parameters, reduced ability to automatically execute learned motor plans, impairment in effort-demanding processes, and other abnormalities involving high level executive functions. [32,38-44] Several acoustic, aerodynamic, stroboscopic, electroglottographic, and perceptual studies have demonstrated significant improvement in glottic closure, vocal fold vibratory movements, sound pressure level (SPL), voice fundamental frequency (Fo) range and modulations, voice quality and speech intelligibility following LSVT® [11,37,45-46] (see also Baumgartner C, Sapir S, Ramig L. Perceptual voice quality changes following phonatory-respiratory effort treatment (LSVT®) vs. respiratory effort treatment for individuals with Parkinson Disease. J Voice submitted.

The LSVT® has been previously compared with an alternative treatment method which emphasizes high respiratory effort (RET).[46] The comparison with the RET group was carried out both to evaluate the role of increased respiratory drive alone in the improvement of loudness in individuals with PD and to rule out extraneous factors such as the Hawthorne or placebo effects in interpreting treatment outcome. The greater improvement in vocal function with the LSVT® compared with the RET previously reported[46-47] (see also Baumgartner C, Sapir S, Ramig L. Perceptual voice quality changes following phonatory-respiratory effort treatment (LSVT®) vs. respiratory effort treatment for individuals with Parkinson Disease. J Voice ;submitted) is in line with evidence from physiological studies in normal adult individuals. These studies have shown that as one increases loudness from normal to high levels, there are significant increases in SPL, subglottal air pressure, transglottal airflow, vibratory movement of the vocal folds, vocal fold adduction and glottal closure.[25,48-54] Similarly, studies have shown that increasing voice intensity by vocal fold medial adduction and compression (through the contraction of the lateral cricoarytenoid, interarytenoid, and thyroartytenoid muscles) is more efficient than by increasing transglottal airflow or vocal fold tension.[55-58] Given these facts, and given the differential effects of LSVT® and RET treatments on acoustic and physiologic measures mentioned above, one would expect to observe greater improvement in vocal function following LSVT® than following RET.

The positive impact of the LSVT® program on voice and speech has recently been demonstrated not only in individuals with PD but also in individuals with multiple sclerosis (see Sapir S, Pawlas A, Ramig L, Seeley E, Fox C, Corboy J. Effects of intensive phonatory-respiratory treatment (LSVT®) on voice in individuals with Multiple Sclerosis. Neurology ;submitted). The LSVT® program has also been shown to improve swallowing in individuals with neurologically-based dysphagia.[45-46]

The above studies, documenting the efficacy of the LSVT® program, have been based on data obtained immediately following therapy, or 6 or 12 months following therapy. Given the slowly progressive nature of PD and some of the neurocognitive deficits associated with PD (e.g., impaired memory and self-monitoring skills) there is a need

to study the effects of LSVT® over a protracted time. The purpose of this study was to assess the impact of LSVT® and RET on vocal functions two years after treatment.

Method

Subjects

Thirty three individuals with idiopathic PD were the subjects of this study. Twenty one of them (17 males, 4 females) were in the LSVT® program and 12 (7 males, 5 females) were in the RET Program. Mean and standard deviation values of treatment group characteristics of age, time since PD was first diagnosed, score on the Unified Parkinson's Disease Rating Scale (UPDRS),[59] stage of disease, [60] and clinical speech and voice severity ratings before voice treatment and at 24 months follow-up are reported in Tables A and B. There were no statistically significant differences between the two groups and between the four times (pre-treatment, immediately post-treatment, 12 months follow-up, and 24 months follow-up) on any of these variables. The two groups also were not different in terms of the changes in medication they received during the 2 year period of the study. All subjects were considered "optimally medicated" by their neurologist, a movement disorders specialist, throughout the study.

After stratification on these variables, the subjects were randomly assigned to one of the two treatment groups. Patient attrition and lack of subject compliance with the experimental protocol in the two treatment groups resulted in unequal group sizes and smaller number of women.

Treatment

Details of treatment have been described previously.[24,46] Both forms of treatment were intensive, with a duration of four one-hour sessions per week for four weeks. Both emphasized high effort levels and encouraged subjects to perform at maximum effort level throughout every session. Both types of therapies included repeated exercises for the first half of each session and speech tasks for the second half of each session.

Table A.
Mean (Standard Deviation) Values of Pretreatment Group Characteristics

	Treatment group	
	LSVT (n = 21)	RET (n = 12)
Age (yr)	61.33 (11.40)	63.25 (7.14)
Time since diagnosis (yr)	7.19 (5.38)	5 (4.59)
Speech severity rating*	1.24 (1.18)	1.67 (1.87)
Voice severity rating*	2.52 (1.12)	2.25 (1.06)

* Severity ratings of speech and voice deficits are on a scale of 1 to 5: 1 = mild; 2 = mild/moderate; 3 = moderate; 4 = moderate/severe; and 5 = severe.

The RET program targeted increased respiratory muscle activity to increase respiratory volumes and subglottal air pressure and loudness.[46] Treatment tasks included maximum inspiration and expiration,[61-62] maximum prolongation of /s/ and /f/ ,[63] and sustained intraoral air pressure using the Iowa Oral Performance Instrument [IOPI]. [64] Subjects were encouraged to maximize their respiratory effort and were given frequent encouragement to "breathe" just prior to each of the sustained productions, and during pauses while reading or performing conversational speaking tasks. Visual feedback of rib cage and abdomen excursions was provided to the individuals via NIMS Respigraph system PN SY03.[46] The RET did not address phonation or increasing phonatory effort, vocal fold adduction or voice pitch modulations.

The LSVT® targeted increasing vocal effort to improve loudness. The main goal of the LSVT® is to maximize phonatory efficiency by improving vocal fold adduction and overall laryngeal muscle activation and control.[24,46] Special care is taken to increase vocal fold adduction without causing vocal hyper-adduction and strain. Upper extremity pushing and lifting tasks [13,65] during phonation were implemented to increase vocal fold adduction. Maximum prolongation of "ah" and maximum fundamental frequency range drills were completed. Subjects were encouraged to maximize phonatory effort and were given frequent encouragement to "think loud" during sustained phonation tasks, reading and conversational speaking tasks.[24,46] Attention was given to the respiratory system in the form of general reminders for subjects to take deep breaths "to be loud". The respiratory system was indirectly stimulated during all "think loud" speech tasks.[24,46]

The treatment intensity, high effort, clinician feedback, daily homework, daily quantification of treatment variables and carryover were all presented and stimulated equally in both treatment groups. Two clinicians delivered the treatment to all the subjects; both clinicians administered both forms of treatment and were randomly assigned to individuals patients. The clinicians worked together to ensure consistency and equivalent high effort and motivation across both forms of treatment. No other additional treatment was administered after the initial 16 sessions.

Table B.
Mean Scores (Standard Deviation) on the UPDRS, Stage of Disease, BDI and Cognitive Tests for the Treatment Groups Before and 24 Months After Treatment

	Treatment group			
	LSVT		RET	
	Before Treatment	24 months after treatment	Before treatment	24 months after treatment
UPDRS (n = 16 and 7)	27.66 (12.03)	29.19 (15.13)	12.86 (12.38)	19.21 (18.27)
Stage (n = 17 and 10)	2.56 (0.63)	2.65 (0.68)	2.20 (0.89)	2.40 (0.97)
BDI (n = 13 and 7)	11.77 (5.05)	9.54 (4.59)	9.0 (4.20)	8.86 (5.70)
Cognition (n = 14 and 9)	44.34 (4.12)	45.65 (5.43)	49.65 (3.09)	50.85 (4.54)

Procedures and Analysis

Pretreatment experimental data were collected within the week before speech treatment was initiated. Post-treatment data were collected within the week after treatment and were collected at the same time after medication. Additional post-treatment speech data collection sessions were completed at 6, 12, and 24 months after the initial therapy program. The results of the 6 and 12 months follow-up have been reported elsewhere [38] and will not be included in this study. The results of the 24 month follow-up will be compared here relative to the pre- and post-treatment data.

Neurological Assessment

Routine neurologic assessment, including standardized testing (UPDRS and Hoen & Yahr staging), were carried out on all subjects within one month of the initiation of speech treatment, as well as at 12 and 24 months after speech treatment.

Neuropsychological Assessment

A battery of neuropsychological tests were administered at 1 month before or within the first week of speech treatment to determine the status of cognitive functioning in these subjects. The measures were focused on attention and concentration, learning and memory, vocabulary, auditory verbal comprehension, and visual spatial skills.[66-67] The results of this testing produced 13 scores, which were converted to T scores that correct for demographic variables of age, education, and sex.[68] After completion of the cognitive evaluation, subjects were interviewed by the clinical neuropsychologist. At this time, subjects also completed the Beck Depression Inventory (BDI),[69] a self-report depression scale. The BDI scores were compared before treatment and 24 months after treatment. The cognitive testing was completed before treatment and 24 months after treatment to avoid learning effects. Results for both BDI and cognition are presented in Table B.

Otolaryngological Assessment

An otolaryngologic history and video-laryngoscopic examination were obtained on all subjects before the commencement of speech treatment. Subjects were excluded from the study if on examination there was evidence of laryngeal pathology (e.g., severe gastric reflux and benign mucosal lesions) not related to PD that would contraindicate speech and voice therapy. Additional details of pretreatment [70] and pretreatment to post treatment laryngologic characteristics are reported elsewhere.[11] Results of analysis of variance revealed no significant difference between the treatment groups for glottal incompetence at baseline.

Speech Assessment

For clinical purposes, standard speech and voice assessments (e.g., motor speech examination) were completed at the time of the first pretreatment speech data collection session. None of the subjects exhibited oral motor or speech and voice characteristics uncommon to PD. The severity of speech disorder ratings presented in Table A were determined by clinical observations.

Experimental speech data collection. Subjects were seated in a medical examining chair located in an IAC sound-treated booth; transducers were positioned and calibrated using standard procedures. A Bruel and Kjaer sound level meter (model 2230) was positioned 30 cm from the lips, and a head-mounted microphone (AKG C410) was positioned 8 cm form the lips. A calibration tone of known intensity was generated at the subjects' lips at the beginning and end of each recording session. The signals were recorded onto a Sony PC-108M eight-channel digital audio tape (DAT) recorder. Data were collected while subjects performed the following tasks: Maximum duration of sustained vowel "AH" phonation' reading of the phonetically

Table C.
Pre-, Post-Treatment, & Follow (FU) Means & Standard Deviations of SPL & STSD measures of Sustained "AH" Phonation, Reading the "Rainbow Passage" Aloud, & Conversational Speech (Monologue)
LSVT=Lee Silverman Voice Treatment
RET=Respiratory Effort Treatment

			pre	post	FU	PRE to POST	PRE to FU
SPL "AH"							
	LSVT (n=21)	Mean	68.26	82.36	76.5	F = 149.88	F = 39.32
		SD =	4.45	3.92	4.1	p = 0.000	p = 0.000
	RET (n=12)	Mean	69.19	68.69	70.1	F = 0.5082	F = 0.172
		SD =	5.31	4.79	7.01	p > 0.20	p > 0.20
SPL Rainbow							
	LSVT (n=21)	Mean	66.18	75.31	69.7	F=49.68	F = 14.23
		SD =	3.79	4.22	3.19	p = 0.000	p = 0.001
	RET (n=11)	Mean	65.79	68.03	66.4	F = 3.2537	F = 0.1471
		SD =	2.6	3.36	5.54	p > 0.05	p > 0.20
SPL monol							
	LSVT (n=12)	Mean	64.7	69.36	67.0	F = 31.30	F = 9.88
		SD =	2.56	3.39	1.87	p = 0.000	p = 0.009
	RET (n=6)	Mean	64.72	65.76	65.7	F = 0.0729	F = 0.013
		SD =	2.76	2.72	4.32	p > 0.20	p > 0.20
STSD Rainbow							
	LSVT (n=20)	Mean	1.9	2.48	2.29	F = 35.65	F = 17.78
		SD =	0.53	0.71	0.65	p = 0.000	p = 0.000
	RET (n=12)	Mean	1.87	2.17	2.03	F = 25.44	F = 3.278
		SD =	0.46	0.36	0.35	p = 0.000	p=0.098
STSD monol							
	LSVT (n=11)	Mean	1.74	2.09	2.39	F = 7.832	F = 5.280
		SD =	0.32	0.56	1.03	p = 0.019	p = 0.044
	RET (n=9)	Mean	2.25	2.14	2.13	F = 0.285	F = 0.285
		SD =	0.8	0.73	0.56	p = 0.608	p = 0.608

balanced "Rainbow Passage",[71] and 25 to 30 seconds of conversational speech (a monologue). Not all subjects were able to complete the conversational monologue task because of limited utterances and duration or task confusion.

Data Analysis

Vocal intensity measures for sustained phonation, reading, and monologue were analyzed using a custom-built software program. The sound level meter signal was preamplified and then digitized at 5K samples per second into a VAX 4000/200 system computer through a 16-bit resolution DSC-200 A/D converter. The software program displayed the sound level meter signal in decibels (dB).

To obtain measures of mean fundamental frequency (Fo) and F0 variability during reading and monologue, the microphone signal was digitized at 5K samples per sound into the same VAX system described above. The files were downloaded onto a 486 computer and then analyzed using Cspeech software. [72] The program calculated mean Fo and standard deviation (SD) in Hertz. Using standard procedures, the Hertz SD was then converted to express frequency variability in semitones (STSD). The STSD takes into account the nonlinear increases in Fo as function of mean Fo increases.

Results

The means and standard deviation (in parentheses) of the SPL and STSD data are summarized in Table C. Differences between means were analyzed statistically using one way analysis of variance (ANOVA) (group) with repeated measures on time (immediately pre-treatment, immediately post-treatment, and 24 months follow-up). Orthogonal contrasts were used to evaluate group differences in variables over time. The F and p values of these analyses are also provided in Table C. Figures 1-5 provide graphic displays of the means of SPL and STSD as a function of treatment group, speech tasks, and time of speech recordings (pre- vs. immediately post-treatment vs. follow-up 24 months post-treatment).

Twenty percent of the data were re-analyzed to determine measurement reliability. Repeated measures of SPL and STSD data yielded correlation coefficients greater than 0.97. Intra-patient reliability for vocal intensity measures have been assessed in previous studies and have been shown to yield correlation coefficients between 0.75 and 0.95, with most correlation coefficients in the upper range.[37,46]

As seen in Table C and Figures 1-3, the LSVT® resulted in a significant improvement in mean SPL and STSD for the three speech tasks from pre- to immediately post-treatment and from pre-treatment to 24 months follow-

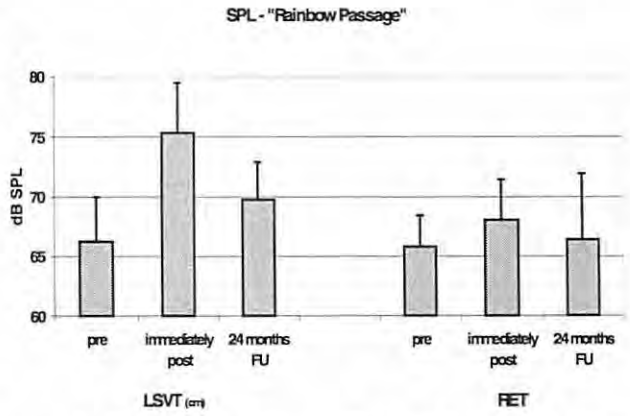


Figure 2. Mean SPL of reading the "Rainbow Passage" immediately pre-treatment (PRE), immediately post-treatment (POST), and 24 months after treatment (FU) in the LSVT and RET groups. Differences from PRE to POST and from PRE to FU are significant in the LSVT group (p = 0.000 and p = 0.001, respectively) but not in the RET group.

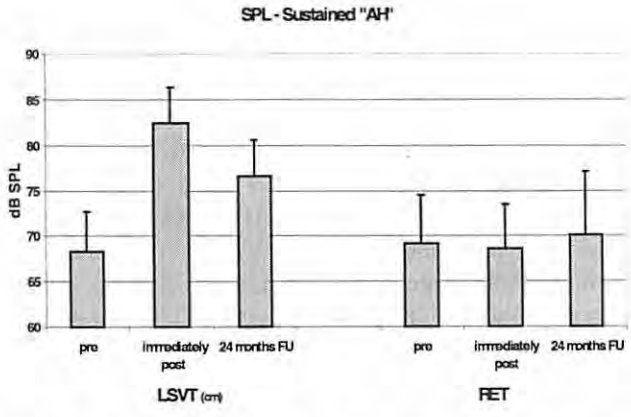


Figure 1. Mean SPL of sustained "AH" immediately pre-treatment (PRE), immediately post-treatment (POST), and 24 months after treatment (FU) in the LSVT and RET groups. Differences from PRE to POST and from PRE to FU are significant in the LSVT group (p = 0.000) but not in the RET group.

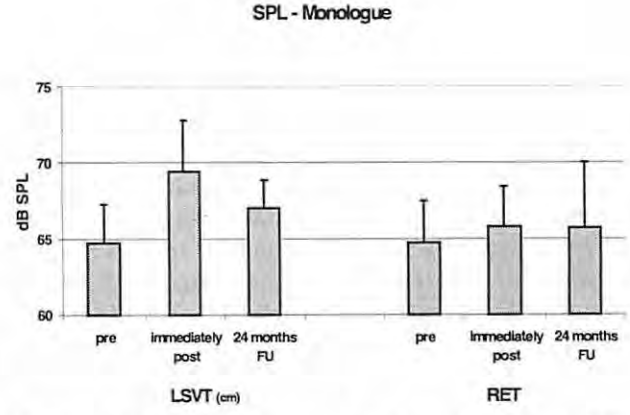


Figure 3. Mean SPL of the monologue immediately pre-treatment (PRE), immediately post-treatment (POST), and 24 months after treatment (FU) in the LSVT and RET groups. Differences from PRE to POST and from PRE to FU are significant in the LSVT group (p = 0.000 and p = 0.009, respectively) but not in the RET group.

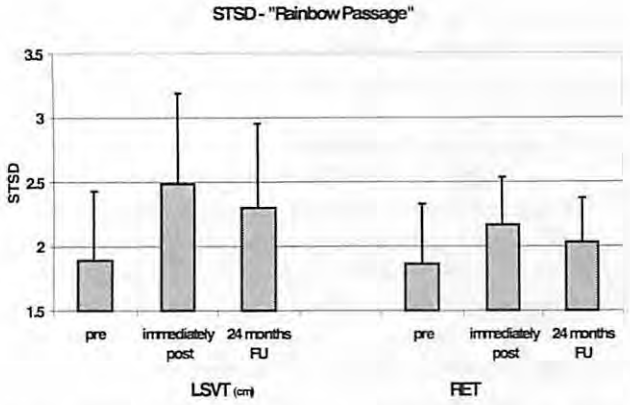


Figure 4. Mean STSD of reading the "Rainbow Passage" immediately pre-treatment (PRE), immediately post-treatment (POST), and 24 months after treatment (FU) in the LSVT and RET groups. Differences from PRE to POST and from PRE to FU are significant in the LSVT group (p = 0.000). The difference from PRE to POST is also statistically significant in the RET group (p = 0.000). The difference from PRE to FU in the RET group is not significant.

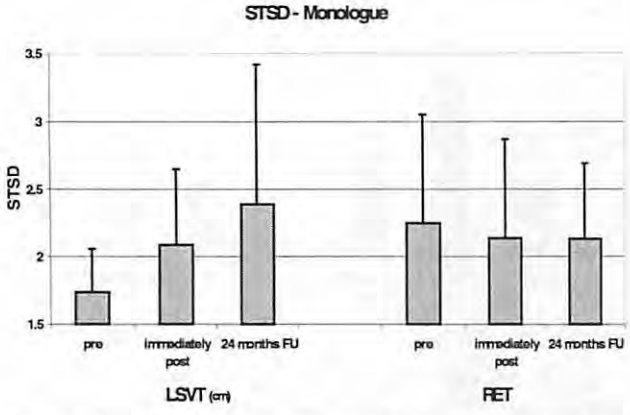


Figure 5. Mean STSD of the monologue immediately pre-treatment (PRE), immediately post-treatment (POST), and 24 months after treatment (FU) in the LSVT and RET groups. Differences from PRE to POST and from PRE to FU are significant in the LSVT group (p = 0.019 and p = 0.044, respectively) but not in the RET group.

up. Specifically, compared to pre-treatment, mean SPL values for post-treatment were significantly higher for sustained "AH" (by 14.1 dB, $p < 0.000$), "Rainbow Passage" (by 9.13 dB, $p < 0.000$), and monologue (by 4.66 dB, $p = 0.000$). Compared to pre-treatment, mean SPL values for the 24 months follow-up were statistically higher for sustained "AH" (by 8.24 dB, $p < 0.000$), "Rainbow Passage" (by 3.52dB, $p < 0.001$), and monologue (by 2.3dB, $p = 0.009$). Compared to pre-treatment, mean STSD values for post-treatment were significantly higher for the "Rainbow Passage" (by 0.58 STSD, $p = 0.000$), and monologue (by 0.35 STSD, $p = 0.019$). Compared to pre-treatment, mean STSD values for the 24 months follow-up were statistically higher for the "Rainbow Passage" (by 0.39 STSD, $p = 0.000$), and monologue (by 0.65 STSD, $p = 0.044$).

The RET failed to show significant improvement in SPL or STSD for any but one of the speech tasks from pre-treatment to post-treatment and from pre-treatment to 24 months follow-up. The exception is a significant improvement from pre-treatment to immediately post-treatment on the STSD measures during reading the "Rainbow Passage" (by 0.30 STSD, $p = 0.000$).

Discussion

It is logical to expect that, due to the degenerative nature of PD and the neurocognitive deficits that are associated with the disease (e.g., poor memory and self-monitoring skills), individuals with PD will not be able to retain the skills they have learned during therapy for a protracted time.[13-14,17-18] Indeed, our previous experience with more traditional treatment methods of voice and speech abnormalities in PD has taught us that individuals with PD will not maintain what they learn even a month or two after therapy. It is therefore remarkable that the vocal improvement with the LSVT® in this study was maintained for at least two years after therapy, without any additional therapeutic intervention. The fact that the RET did not produce any favorable effects on voice suggests that the improvement with the LSVT® was neither due to extraneous factors such as the Hawthorne or placebo effects nor simply related to increased respiratory drive.

We would like to offer three possible reasons why LSVT® but not RET produce these long term effects. The first one is that the patients learned to increase vocal fold adduction and improve laryngeal muscle activation and synergy, thus rendering the phonatory system more efficient. This interpretation is in line with previous physiologic studies of patients treated with LSVT®, demonstrating improved glottic closure and greater vibratory motions of the vocal folds after treatment.[11,47] It is not clear whether the increase in STSD with the LSVT® reflects simply an increase in vocal fold tension and subglottal pressure associated with increasing loudness, or whether it also reflects intentional activation of laryngeal muscles to improve intonation. We suspect that both explanations are correct since, perceptually, individuals treated with LSVT® often improve both loudness and prosody.[37]

The second explanation is that the LSVT®, by emphasizing loud phonation, high vocal effort, and self-monitoring of both loudness and effort, have helped the patients overcome some of the higher level deficits associated with PD, especially deficits in proprioceptive processing, scaling motor output parameters, motor learning, programming and memory, and servo-regulation of movement.[31,38,73-77] Indeed, physical therapy treatment techniques used to rehabilitate individuals with PD often emphasize intensive motor relearning, maximizing motor output and effort, increasing drive and goal-directed activity, and enhancing sensory

awareness to promote internal cueing, self monitoring, and upscaling of motor output.[16,75,78-84] These techniques help patients maximize motor performance and maintain that performance over a long period of time. Since the RET involved similar intensive treatment, one wonders why it did not produce favorable results. One explanation is that the target of treatment was respiration rather than phonation and that the lack of emphasis on the phonatory system did not allow patients treated with this method to maximize phonatory output. Another explanation is that what helped the LSVT® group was hearing their own voice as they practiced loud phonation, an auditory feedback that was not available to those treated with the RET.

The third explanation for the long term effect with the LSVT® is that the emphasis on loud phonation and high effort levels stimulated centers in the brain that are associated with drive and goal directed activities. These neuropsychological activities are highly related to the limbic system, which is also involved with the regulation of emotive vocalization and intensity of vocalization.[85-87] Regarding the latter, Jurgens & von Cramon[90] have argued that the limbic system, and the neocortical and subcortical systems associated with it, do not participate in motor coordination, nor in the execution of phonatory gestures; rather, they seem to function as a drive-controlling mechanisms that determines, by its activity, the readiness to phonate as well as the intensity of phonation. Thus, the LSVT®, by emphasizing loud and effortful phonation, may have constantly stimulated these systems in the brain, which may be impaired in individuals with PD, and which may have become more functional with LSVT®. Recent findings from a PET study provide preliminary evidence for the impact of LSVT® on cortical and subcortical areas.[88]

Most likely, it is the combination of these explanations that yielded the significant improvement in the LSVT® group. Moreover, the effects of the LSVT® were probably related to the simplicity of the task ("loud") and its upscaling effect on the entire speech motor system[37] (see also Sapir S, Pawlas A, Ramig L, Seeley E, Fox C, Corboy J. Effects of intensive phonatory-respiratory treatment (LSVT®) on voice in individuals with Multiple Sclerosis. Neurology; submitted) and the positive motivational feedback subjects in the LSVT® group received when they spoke intelligibly.

Interestingly, the LSVT® often improves not only phonation but also articulation.[47] Moreover, even in non-disordered speakers, increased loudness results in improved articulation.[25-26] The reason for this improvement in not clear. However, we suspect that when individuals are asked to produce loud phonation, or when they voluntarily increase loud phonation, they invariably activate those "primitive" centers in the brain that increase overall output of the vocal mechanism. The limbic system is strongly implicated in this loudness control mechanism, because of its role in regulating the intensity of vocalization. Importantly, studies in primates have shown that when certain centers in the brain, such as the periaquaductal gray and the anterior cingulate, are stimulated artificially to elicit vocalization, not only the respiratory and phonatory muscles are activated, but also the "articulatory muscles", i.e., the muscles of the tongue, jaw, and lips.[89-91]

To conclude, the present findings suggest that intensive treatment, emphasizing high effort loud phonation and self-monitoring of this phonation, can result in marked and long-terms improvement in vocal function. Recent studies suggest that the effects of LSVT® extend beyond loud phonation, and include improved voice quality, prosody, articulation, and speech intelligibility[3,37,47] (see also Sapir S, Pawlas A, Ramig L, Seeley E, Fox C, Corboy J. Effects of intensive phonatory-respiratory treatment (LSVT®) on voice in individuals with Multiple Sclerosis. Neurology; submitted *and* Baumgartner C, Sapir S, Ramig L. Perceptual voice quality changes following phonatory-respiratory effort treatment (LSVT®) vs. respiratory effort treatment for individuals with Parkinson Disease. J Voice ;submitted). We suggest that in addition to increased motor drive these positive and long-lasting effects of LSVT® are related, at least partially, to improvement in neurocognitive, psychomotor, and other high level executive functions, which are probably mediated via the limbic system and its neocortical and subcortical connections. Our explanations for the effects of LSVT® are tentative, and obviously in need of empirical substantiation beyond the evidence discussed here.

The results of the present study should increase awareness among physicians and other clinicians about the efficacy of the LSVT® and similar approaches in the treatment of speech and voice disorders in individuals with PD and other neurologic disease. Improved oral communication can make a significant positive impact on quality of life.

Acknowledgements

The research presented here was supported by NIH grant No. R01DC01150. Our deepest gratitude is extended to the subjects who participated in this research study. The authors would like to thank Dr. Christopher Dromey, Dr. Kristin Baker, and Dr. Susan Hensley for their assistance during various parts of this research.

References

1. Logemann J, Fisher H, Boshes B, Blonsky E. Frequency and concurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. J Speech Hear Disord 1978;42: 47-57.

2. Streifler M, Hofman, S. Disorders of verbal expression in Parkinsonism. In Hassler R, Christ J eds. Advances in Neurology. New York: Raven Press, 1984.

3. Maclay S. Speech intelligibility gains in Parkinson's disease patients post voice treatment: Perceptual and acoustic correlates. Unpublished master's thesis, 1992; University of Colorado, Boulder.

4. Oxtoby, M. Parkinson's Disease patients and their social needs. London: Parkinson's Disease Society; 1982.

5. Ramig L. The role of phonation in speech intelligibility: A review and preliminary data from patients with Parkinson's disease. In Kent R ed. Intelligibility in speech disorders: theory, measurement and management. Amsterdam: John Benjamin,1992; 119-155.

6. Baker K, Ramig L, Luschei E, Smith M. Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging. Neurology 1998;51:1592-8.

7. Darley F, Aronson A, & Brown J. Differential diagnostic patterns of dysarthria. J Speech Hear Disord, (1969a). 12, 246-69.

8. Darley, F, Aronson, A, & Brown, J. (1969b). Clusters of deviant speech dimensions in the dysarthrias. J Speech Hear Disord 1969;12:462-9.

9. Hanson D, Gerratt B, Ward P. Cinegraphic observations of laryngeal function in Parkinson's disease. Laryngoscope 1984; 94:348-53

10. Omori K, Slavit D, Matos C, Kojima H, Kacker A, Blaugrund S. Vocal fold atrophy: quantitative glottic measurement and vocal function. Ann Otol Rhinol Laryngol 1997;106:544-51.

11. Smith M, Ramig L, Dromey C, Perez K, Samandari R. Intensive voice treatment in Parkinson's disease: Laryngostroboscopic findings. J Voice 1995; 9:453-9.

12. Solomon N, Hixon T. Speech breathing in Parkinson's disease. J Speech Hear Disord 1993;36:294-10.

13. Aronson A. Clinical Voice Disorders. New York: Thieme-Stratton, 1990.

14. Greene M. The voice and its disorders. London: Pitman Medical, 1980.

15. Hoberman S. Speech techniques in aphasia and Parkinsonism. J Michigan State Med Societ 1958;57:1720-3.

16. LeDorze G, Doinne L, Ryalls J, Julien M, Oullet L. The effects of speech and language therapy for a case of dysarthiria associated with Parkinson's disease. Europ J Disord Communic 1992;27:213-24.

17. Sarno M. Speech impairment in Parkinson's disease. Arch Phys Med Rehabil 1968; 49:269-75.

18. Weiner W, Singer C. Parkinson's disease and nonpharmacological treatment programs. J Am Geriatr Soc 1989; 37:359-63.

19. Johnson J, Pring, T. Speech therapy and Parkinson's disease: a review and further data. Brit J Disord Commun 1990; 25:183-94.

20. Robertson S, Thompson F. Speech therapy in Parkinson's disease: A study of the efficacy and long-term effect of intensive treatment. Brit J Disord Commun 1984; 19:213-24.

21. Scott S, Caird F. Speech therapy for Parkinson's disease. J Neurol Neurosurg Psychiat 1983;46:140-4.

22. Ramig L, Mead C, Scherer R, Horii Y, Larson K, Kohler D. Voice therapy and Parkinson's disease: A longitudinal study of efficacy. Paper presented at the Clinical Dysarthria Conference; 1988; San Diego, CA.

23. Ramig L, Bonitati C, Lemke J, Horii Y. Voice treatment for patients with Parkinson disease: Development of an approach and preliminary efficacy data. J Med Speech-Lang Path 1994; 2:191-209.

24. Ramig L, Pawlas A, Countryman S. The Lee Silverman Voice Treatment ($LSVT_{®}$): A Practical Guide to Treating the Voice and Speech Disorders in Parkinson Disease. Iowa City, IA: National Center for Voice and Speech, 1995.

25. Dromey C, Ramig L. Intentional changes in sound pressure level and rate: their impact on measures of respiration, phonation, and articulation. J Speech Hear Res 1998;41:1003-18

26. Schulman R. Articulatory dynamics of loud and normal speech. J Acoust Soc Am 1989;5:295-312.

27. England A, Schwab R. The management of Parkinson's disease. AMA Arch Int Med 1959;104:439-68.

28. Hallet M, Khosbin S. A psychological mechanism of bradykinesia. Brain 1980;103:301-14.

29. McDowell F, Lee J, Sweet R. Extrapyramidal disease. In Baker A, Joynt R eds. Clinical Neurology. Philadelphia: Harper and Row 1986:24-26.

30. Brooks V. The neural basis of motor control. New York: Oxford University Press, 1986.

31. Muller F, Stelmach G. Scaling problems in Parkinson's disease. In: Requin J, Stelmach G, eds. Tutorials in motor neuroscience. Netherlands: Kluwer Academic Publishers 1991:161-74.

32. Stelmach G. Basal ganglia impairment and force control. In: : Requin J, Stelmach G, eds. Tutorials in motor neuroscience. Netherlands: Kluwer Academic Publishers 1991:137-48.

33. Adams J. A closed-loop theory of motor learning. J Motor Behav 1971; 3:111-49.

34. Adams J. Use of the model's knowledge of the results to increase observer's performance. J Human Mov Studies 1986;12:89-98.

35. Schmidt R. A schema theory of discrete motor skill learning. Psychol Rev 1975; 82:225-60.

36. Schmidt R. Motor control and learning. Champaign, IL: Human Kinetic Publishers, 1988.

37. Ramig L, Countryman S, O'Brien C, Hoehn M, Thompson L. Intensive speech treatment for patients with Parkinson's disease: Short- and long-term comparison of two techniques. Am Acad Neurol 1996;47:1496-1504.

38. Elias J, Treland J. Executive function in Parkinson's disease and subcortical disorders.Semin Clin Neuropsychiat. 1999; 4:34-40.

39. Karayanidis F. Parkinson's disease: a conceptuatlization of neuropsychological deficits within an information-processing framework. Biol Psych 1989; 29:149-79.

40. Marsden C. Function of the basal ganglia as revealed by cognitive and motor disorders in Parkinson's disease. Can J Neurol Sci 1984; 11(1 suppl 1):129-35.

41. Pillon B, Dubois B, Cusimano G, Bonnet A, Lhermitte F, Agid, Y. Does cognitive impairment in Parkinson's disease result from non-dopaminergic lesions? J Neurol Neuros Psychiat 1989;52:201-6.

42. Schneider J, Diamond S, Markham C. Parkinson's disease: sensory and motor problems in arms and hands. Neurology 1987;37:951-6.

43. Taylor A, Saint-Cyr J, Lang A. Frontal lobe dysfunction in Parkinson's disease. The cortical focus of neostriatal outflow. Brain 1986;109:845-83.

44. Weingartner H, Burns S, Diebel R, LeWitt P. Cognitive impairments in Parkinson's disease: distinguishing between effort-demanding and automatic cognitive processes. Psychiat Res 1984;11:223-35.

45. de Angelis E, Mourao L, Ferraz H, Behlau M, Pontes P, Andrade L. Effect of voice rehabilitation on oral communication of Parkinson's disease patients. Acta Neurol Scand 1997; 96(4):199-205.

46. Ramig L, Countryman S, Thompson L, Horii Y. A comparison of two forms of intensive speech treatment for Parkinson disease. J Speech Hear Disord 1995; 38:1232-51.

47. Dromey C, Ramig L, Johnson A. Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. J Speech Hear Res 1995;38:751-63.

48. Gelfer M. A multidimensional scaling study of voice quality in females. Phonetica 1993; 50:15-27.

49. Gelfer M. Fundamental frequency, intensity, and vowel selection: effects on measures of phonatory stability. J Speech Hear Res 1995;38:1189-98.

50. Holmes L, Leeper H, Nicholson I. Laryngeal airway resistance of older men and women as a function of vocal sound pressure level. J Speech Hear Res 1994;37:789-99.

51. Koike Y. Application of some acoustic measures for the evaluation of layrngeal dysfunction. Studia Phonologica 1973; 7:17-23.

52. Orlikoff R. Vocal stability and vocal tract configuration: An acoustic and electroglottographic investigation. J Voice 1995; 9:173-181.

53. Orlikoff R. Kahane J. Influence of mean sound pressure level on jitter and shimmer measures. J Voice 1991;5:113-9.

54. Sodersten M, Hertegard S, Hammarberg B. Glottal closure, transglottal airflow, and vice quality in healthy middle-age women. J Voice 1995; 9: 182-97.

55. Berke G, Hanson D, Gerratt B, Trapp T, Macagba C, Natividad M. The effect of air flow and medial adductory compression on vocal efficiency and glottal vibration. Otol Head Neck Surg 1990; 102:212-8.

56. Nasri S, Sercarz J, Azizzadeh B, Krieman J, Berke G. Measurement of adductory force of individual laryngeal muscles in an vivo canine model. Laryngoscope 1994; 104:1213-8.

57. Slavit D, McCaffrey T. Regulation of phonatory efficiency by vocal fold tension and glottic width in the excised canine larynx. Ann Otol Rhino Laryngol 1991; 100:668-77.

58. Tang J, Stathopoulos E. Vocal efficiency as a function of vocal intensity: a study of children, women, and men. J Acoust Soc Am 1995; 97:1885-92.

59. Fahn S, Elton R, Committee, M. O. T. U. D. Recent developments in Parkinson's disease. (Vol. 2). New York: Macmillan Press, 1987.

60. Hoehn M, Yahr M. Parkinsonism: onset, progression and mortality. Neurology 1967; 17:427-442.

61. Hardy J. Cerebral Palsy. Englewood Cliffs: Prentice-Hall, 1983.

62. Netsell R, Rosenbek J. Treating the dysarthrias. In Netsell R ed. A neurobiologic view of speech production and the dysarthrias. San Diego: College-Hill Press, 1986.

63. Leith D, Bradley M. Ventilatory muscle strength and endurance training. J Appl Physiol 1976;41:508-16.

64. Robin D, Goel A , Somodi L, Luschei E. Tongue strength and endurance: Relation to highly skilled movements. J Speech Hear Res 1992; 35:1239-45.

65. Froeschels E, Kastein S, Weiss D. A method of therapy for paralytic conditions of the mechanisms of phonation, respiration and glutintaion. J Speech Hear Disord 1955; 20:3645-70.

66. Goodglass H, Kaplan E. The assessment of aphasia and related disorders. Philadelphia:Lou and Fabiger, 1972.

67. Wechsler D. WAIS-R Manual. New York: The psychological Corporation, 1981.

68. Heaton R, Grant J, Mathews C. Comprehensive norms for an expanded Halsted-Reitan battery. Odessa, FL.: Psychological Assessment Resources, 1981.

69. Beck A, Ward C, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiat. 1961;4:561-571.

70. Perez K, Ramig L, Smith M, Dromey C. The Parkinson larynx: tremor, and videostroboscopic findings. J voice 1996 (in press).

71. Fairbanks G. Voice and Articulation Drillbook. New York: Harper and Brothers, 1960.

72. Milenkovic P. Least mean square measures of voice perturbation. J Speech Hear Res 1987;30:529-538.

73. Bohannon R. Physical rehabilitation in neurologic diseases. Curr Opin Neurol 1993;6(5):765-772.

74. Homberg V. Motor training in the therapy of Parkinson's Disease. Neurology 1993;43:S45-S46.

75. Morris M, Iansek R, Matyas T, Summers J. The pathogenesis of gait hypokinesia in Parkinson's disease. Brain 1994; 117:1168-1181.

76. Morris M, Iansek R, Matyas T, Summers, J. Ability to modulate walking cadence remains intact in Parkinson's disease. J Neurol Neuros Psychiat 1994;57(12):1532-1534.

77. Richards M, Cote L, Stern Y. The relationship between visuospatial ability and perceptual motor function in Parkinson's disease. J Neurol Neuros Psychiat 1993; 56:400-6.

78. Katsikitis M, Pilowsky I. A controlled study of facial mobility treatment in Parkinson's disease. J Psychosom Res 1996; 40(4):387-96.

79. Muller V, Mohr B, Rosin R, Pulvermuller F, Muller F, Birbaumer, N. Short-term effects of behavioral treatment on movement initiation and postural control in Parkinson's disease: a controlled clinical study. Movement Disord 1997; 12(3):306-314.

80. Pacchetti C, Aglieri R, Mancini F, Martignoni E, Nappi G. Active music therapy and Parkinson's disease: methods. Function Neurol 1998;13:57-67.

81. Palmer S, Mortimer A, Webster D, Bistevins R, Dickinson G. Exercise therapy for Parkinson's disease. Arch Physic Med Rehab 1986; 67(10):741-5.

82. Soliveri P, Brown R, Jahanshahi M, Marsden C. Effect of practice on performance of a skilled motor task in patients with Parkinson's disease. J Neurol Neuros Psychiat 1992;55:454-60.

83. Sunvisson H, Lokk J, Ericson K, Winblad B, Ekman S. Changes in motor performance in persons with Parkinson's disease after exercise in a mountain area. J Neurosci Nursing 1997; 29(4):255-60.

84. Viliani T, Pasquetti P, Magnolfi S, Lunardelli M, Giorgi, C, Serra P, Taiti P. Effects of physical training on straightening-up processes in patients with Parkinson's disease. Disabil Rehabil 1999; 21:68-73.

85. Eccles J. The emotional brain. Bulletin et Memoires de L'Academie Royale de medecine de Belgique. 1980; 135:697-711.

86. Jurgens U, von Cramon D. On the role of the anterior cingulate cortex in phonation: a case report. Brain and Language 1982;15:234-48.

87. Sapir S, Aronson A. Aphonia after closed head injury: aetiologic considerations. Brit J Disord Commun 1985; 20:289-96.

88. Liotti M, Vogel D, New P, Ramig L, Mayber H, Cook C, Fox P. A PET study of functional organization of premotor regions in Parkinson Disease following intensive speech and voice treatment ($LSVT_{®}$).Presented at the American Academy of Neurology (AAN); April 4, 1999; Toronto.

89. Larson C. The midbrain periaqueductal gray: a brainstem structure involved in vocalization. J Speech Hear Res 1985; 28(2):241-9

90. Sapir S, Campbell C, Larson C. Effect of geniohyoid, cricothyroid, and sternothyroid muscle stimulation on voice fundamental frequency of electrically elicited phonation in Rhesus Macaque. Laryngoscope 1983; 91:457-468.

91. West R, Larson C Neurons of the anterior mesial cortex related to faciovocal activity in the awake monkey. J Neurophysiol 1995; 74,:1856-69.

NCVS Status and Progress Report - 14
September 1999, 141-147

Effects of Intensive Phonatory-Respiratory Treatment (LSVT®) on Voice in Individuals with Multiple Sclerosis

Shimon Sapir, Ph.D.
Annette Pawlas, M.A.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Lorraine Ramig, Ph.D.
Department of Speech Language Hearing Sciences, University of Colorado-Boulder
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Erika Seeley, B.Sc.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Cynthia Fox, M.A.
Department of Speech and Hearing Sciences, The University of Arizona – Tucson
The National Center for Neurogenic Communication Disorders
John Corboy, M.D.
Health Sciences Center, The University of Colorado - Denver

Abstract

Many individuals with multiple sclerosis (MS) exhibit a variety of voice and speech problems, with vocal weakness and fatigue being common complaints. The purpose of this study was to assess the impact of an intensive phonatory-respiratory treatment program (Lee Silverman Voice Therapy, LSVT®) on vocal function in two women (ages 47 and 48) with a long history (12 and 15 years) of MS. These women complained of vocal weakness and fatigue associated with their illness. These voice problems were chronically present in spite of the fact that the symptoms of MS were in remission in these women during the study. Statistically significant improvement from pre- to post-treatment and from pre- to 6 month post-treatment follow up were observed in vocal sound pressure level (SPL) ($p < 0.005$) for different speech tasks and for duration of sustained vowel phonation ($p < 0.005$). Statistically significant improvement ($p < 0.001$) was also observed in the perceptual rating of voice loudness after treatment. The present findings provide further support for the efficacy of LSVT® in the treatment of various neurologically-based speech disorders.

Introduction

Approximately 40-45% of individuals with multiple sclerosis (MS) suffer from dysarthria.[1-2] These individuals may exhibit different speech abnormalities, depending on the severity of the neurologic involvement, location of the lesions in the central nervous system (CNS), and the progressive or fluctuating nature of the disease. In most of these individuals, the dysarthria is of the spastic-ataxic type.[3] Some of the prominent features of the dysarthria include impaired vocal loudness, breathy or harsh voice, vocal instability, and imprecise articulation.[3-5] Complaints of vocal weakness and fatigue, often accompanied by dysphagia, are common in dysarthric individuals with MS.[2,6-7]

Although oral communication is a significant aspect of quality of life, employment, and social functioning, very few studies have assessed the effects of speech treatment for individuals with MS.[7-8] Farmakides & Boone [7] reported that 85% of 68 patients who received speech therapy improved their speech. The therapy was designed according to the specific deficits each of the patients exhibited. Detailed information about treatment was not reported. Evaluation of treatment was done perceptually by three judges using a 4-point scale. Farmakides & Boone argued

that the improvement in their patients was related to a reduction in disuse atrophy, which presumably had prevented them from making full use of residual muscle function. Hartelius et al [8] treated seven dysarthric individuals with MS. Five of these patients improved with therapy, as measured acoustically and perceptually. The other two patients did not improve significantly. Hartelius et al [8] commented that the lack of improvement in these two individuals may have been due to the intensity of the treatment they had received, which was much greater than that provided for the five individuals who improved. Hartelius et al [8] concluded that "..It is conceivable that the MS population is not the most suitable for intensive therapy (i.e., 4-5 sessions a week), due to the fatigue that is frequently a part of the symptom spectrum." (p. 137).

Petajan, Gappmaier, White, Spencer, Mino, & Hicks[9] have discussed the interaction between fatigue, disuse atrophy, and depression in individuals with MS. They argued that these three factors can be minimized by exercising the system that is prone to fatigue. They demonstrated that individuals with MS who had received intensive aerobic training (3 X 40-minute sessions per week) improved significantly in fitness and quality of life, whereas patients who did not receive such training did not improve. Given the findings by Petajan et al,[9] this study was designed to evaluate whether an intensive treatment program known as the Lee Silverman Voice treatment (LSVT®) might improve vocal function in individuals with MS. The LSVT® uses high effort vocal exercises administered intensively (4 X 50-minute sessions per week for four weeks) to retrain voice in individuals with neurological disorders. The LSVT®) has been shown to be highly effective in improving vocal function in individuals with Parkinson's disease (PD), who also often complain of voice weakness and fatigue.[2,10]

Methods and Materials

Two women with MS were treated with LSVT® as described below. Subject #1 was 47 years old, with a progressive type of MS, initially diagnosed 12 years earlier. This individual was on medication, which included Baclofen, Ditropan, Hiprex, Micronase, and Prozac. Subject #2 was 48 years old, also with progressive MS, which had been diagnosed 15 years earlier. This individual was also on medication, which included Baclofen, Desiprimine, Qualfenisin, Minipress, and Noraprin. There were no medication changes during the course of the study.

In both individuals, the MS symptoms were in remission during the study, with the exception of a voice problem described by a speech pathologist as "reduced in vocal loudness" and by the subjects as "weak and easily fatigable." These individuals reported communicative problems such as not being well understood, especially in noisy situations, and often being asked to repeat a sentence because of poor intelligibility. They also felt that talking was effortful and that they were less inclined to engage in conversation. Because of these problems, they sought voice therapy to improve oral communication. They were recruited for this study from a local MS support group.

The LSVT® [11] is an intensive treatment program designed to improve vocal fold adduction and overall voice and speech production in individuals with neurologic disease who have dysphonia or dysarthria. The program is unique in that it focuses on a simple set of tasks designed to maximize phonatory and respiratory functions. This is done by instructing and constantly stimulating individuals to produce loud voice with increased effort during sustained phonation and in various speech tasks. These individuals are also constantly reminded to monitor the accompanying vocal effort it takes to produce the louder voice. [12-13]

The goals of increased vocal loudness and effort during phonatory tasks with the LSVT® are aimed at improving respiratory drive, vocal fold adduction, laryngeal muscle activity and synergy, laryngeal and supralaryngeal articulatory movements, and vocal tract configuration. These physiologic changes have been documented to improve voice quality and intensity, articulatory precision, prosodic inflection, resonance, and speech intelligibility.[12,14-17] (also Baumgartner C, Sapir S, Ramig L. Perceptual voice quality changes following phonatory-respiratory effort treatment (LSVT®) vs. respiratory effort treatment for individuals with Parkinson disease. J Voice submitted). Such changes accompanying high effort phonation are expected based on similar effects seen in non-disordered speakers.[18]

The implementation of high-effort, intensive phonatory-respiratory therapy is based on evidence from clinical practices in neurology and physical therapy,[19-21] suggesting that when individuals with neurologic disease are pushed to higher effort levels, they learn to compensate for, or overcome, some of the neuromuscular deficits that underlie their physical impairment. In line with theories of motor learning,[22-23] Ramig and her colleagues [10] have argued that intensive high-effort treatment of vocal function, especially when coupled with proprioceptive feedback and auditory-vocal self-monitoring, should help individuals with neurologic disease rescale the magnitude of their speech motor output and habituate this level in conversation. Details of the LSVT® have been described previously. [13]

The LSVT® targets increasing vocal effort to improve vocal loudness to a level that is within normal limits. The main goal of the LSVT® is to maximize phonatory efficiency by improving vocal fold adduction and overall laryngeal muscle activation and control.[13,24] Special care is taken to increase vocal fold adduction without causing vocal hyper-adduction and strain. Upper extremity pushing and lifting tasks [25-26] during phonation are implemented to increase vocal fold adduction. Maximum prolongation of "ah" and maximum fundamental frequency range exercises

are practiced. Subjects are encouraged to maximize phonatory effort and are given frequent encouragement to "think loud" during sustained phonation tasks, reading and conversational speaking tasks.[24] Attention is given to the respiratory system in the form of general reminders for subjects to take deep breaths "to be loud". The respiratory system is indirectly stimulated during all "think loud" speech tasks. [24]

The LSVT$_{®}$ is administered in four one-hour sessions per week for four weeks. It emphasizes high effort levels and encourages subjects to perform at maximum effort level throughout every session. It includes phonation exercises for the first half of each session and speech tasks for the second half of each session.

Voice Recording Procedures

Each individual was audio-recorded seven times: three times immediately prior to treatment to establish a baseline performance or level of function (pre-1, pre-2, pre-3), two times immediately after the completion of the treatment (post-1 and post-2), and two times at a 6-month follow-up (fu-1 and fu-2). These recordings were obtained in a sound treated room as each of the subjects was seated in a chair and performed these tasks: 1) sustaining vowel "ah" phonation for as long as possible for six repetitions; 2) reading the "Rainbow passage," [27] a phonetically balanced text; 3) describing a picture ("The Cookie Theft Picture"; [28]); and 4) speaking free on a self-chosen topic ("monologue"). The audio recordings were made with an 8-channel Sony PC-108M digital audio recorder. These recordings were transduced with a head mounted microphone (AKG-C410), which was positioned 8 cm from the lips, and with a Bruel and Kjaer sound level meter (Model 2230), which was positioned 30 cm from the lips. These distances were constantly monitored throughout recording sessions. A calibration tone of known intensity was generated at the individual's lips at the beginning and end of each recording session as well as following any adjustments of input levels onto the digital audio tape (DAT) recorder. In addition, the experimenter hand-recorded the peak vocal SPL measures which were continuously displayed at 1-s intervals from the digital output of the SLM during all speaking and voice tasks. The same experimenter collected all handwritten vocal SPL data.

Acoustic Analysis

Sound pressure level (SPL) was calculated for sustained phonation, reading, picture description, and monologue using the continuously hand-recorded peak SPL that was displayed at 1-s intervals from the digital output of the sound level meter. Mean vocal SPL measures derived from hand-recorded second-to-second peak vocal SPL have been reported to be comparable to mean vocal SPL measures derived from a custom-built software program.[24,29]

Duration of Sustained Phonation

Duration of sustained phonation was measured from the onset to the offset of the "ah" waveform, presented on the computer screen after being digitized from the audiotape.

Perceptual Analysis

In order to evaluate the impact of pre to post-treatment changes on listener perception, two studies were carried out. In one study vocal loudness associated with the readings of the "Rainbow passage" and the description of the "Cookie Theft" before and after treatment was assessed perceptually by 11 naïve raters who had no previous experience in analyzing or rating voices of individuals with neurologic disease. These raters were drawn from a larger pool of listeners, and were included in this study because of their high intra- and inter-rater reliability (Cronbach alpha +=0.90-1.0). All raters were volunteers and all had normal audiometric testing. Their ages ranged from 20 to 51 (mean = 34).

Each rater was seated in a sound treated booth. He or she was presented with audio-recorded samples of the pre- and post-treatment utterances and asked to rate the loudness of each sample on a line scale. This line scale was marked "too soft" on one end, and "normal" on the other end, with no marks in between these ends. The rater responded via a computer mouse by clicking on the scale line. This line was 100mm long, with 0 corresponding to normal loudness and 100 to "too soft". The rater was instructed to make a mark anywhere along the line to indicate the level of voice loudness he or she heard.

All raters were presented with the recording samples through a speaker. The distance of the speaker from the rater was the same for all raters. Also, all recordings were presented to the listeners at the same SPL level as it was originally recorded. This was done by adjusting the volume dial on the tape recorder so that the SPL at the listener's ear (measured with the sound meter level) was the same as the SPL of the original recording. The position of the volume dial was marked. The listeners were instructed not to move the chair, adjust the volume, or lean closer to the speaker during the experiment. The position of the chair and the volume dial were checked before and after each listener.

Each rater sat alone in the booth and was responsible for changing the tapes in the DAT machine and making the rating on the visual analog scale line which was presented on the computer screen. The order of the audio-recorded samples was randomized within and across raters, to counterbalance fatigue, inattention, and other extraneous effects.

In another perceptual study, 4 speech-language pathologists were asked to listen to pairs of voice recordings and to indicate which of the two in each pair was louder.

They recorded their responses on a response sheet. The pairs represented pre- and post-treatment samples of patient #1 or patient #2 voices during reading of the "Rainbow passage" or during picture description. The order of the samples within the pairs was randomized so the listeners had no a priori information as to whether the specific sample was recorded before or after treatment. Each of the 4 listeners was seated alone in an IAC booth and was presented with the stimuli in a similar manner as described above for the previous perceptual study. The pairs were repeated in random order to assess intra- and inter-listener reliability. Intra- and inter-listener reliability measures indicated greater than 90% agreement.

Finally, the patients' speech-language summary reports before and after LSVT® were reviewed to obtained information about possible qualitative changes experienced by the patients after treatment.

Results

Tables A through D summarize the results of the acoustic and perceptual analysis of the voice pre-, post- and 6 month follow-up (FU) for patients # 1 and #2. A one way analysis of variance (ANOVA) was used to compare pre- versus post-treatment and pre- versus FU. The F and p values of these analyses are provided in the Tables. The SPL (30 cm) and duration data are provided in terms of means and standard deviation (in parentheses). The perceptual rating data are provided in terms of means and standard error of measurements (SEM). The perceptual comparison data were not subjected to statistical anlyses. The numbers simply indicate how many of the 4 listeners judged the voice to be louder, pre or post treatment.

Patient # 1

As can be seen in Tables A and B, Patient #1 showed significant improvement pre- to post-, and pre- to FU in SPL, duration, and perceptual measures, indicating that for this individual, LSVT® resulted in marked improvement in vocal function. Specifically, the voice significantly increased in SPL for sustained "AH" phonation (average pre- to post increase 13.1 dB, $p < 0.005$; pre- to FU increase: 13.4 dB, $p < 0.005$), for the "Rainbow Passage" (pre-post: 13.1 dB, $p < 0.005$; pre-FU: 9.6 dB, $p < 0.005$), for the picture description (pre-post: 9.8 dB, $p < 0.01$; pre-FU: 7.6 dB, $p < 0.025$), and for the monologue (pre-post: 11.7 dB, $p < 0.001$; pre-FU: 6.1 dB, $p < 0.001$). The changes in SPL from pre- to post-treatment and from pre- to follow up for Patient #1 are illustrated graphically in Figure 1.

Duration of sustained "AH" phonation increased significantly from pre- to post (13.1 sec increase, $p < 0.001$) and from pre- to FU (9.5 sec increase, $p < 0.001$). Percep-

Table B.
Perceptual Data for Pating #1

		Pre	Post	SE M	F =	p <
Percept rating of loudness (0 = normal, 100 = too soft)	"Rainbow"	35.0	2.64	2.059	41.74	0.001
	Picture	46.5	1.68	2.41	65.06	0.001
Perceptual comparison of pre vs. post "louder" (N=4)	"Rainbow"	0	4			
	Picture description	0	4			

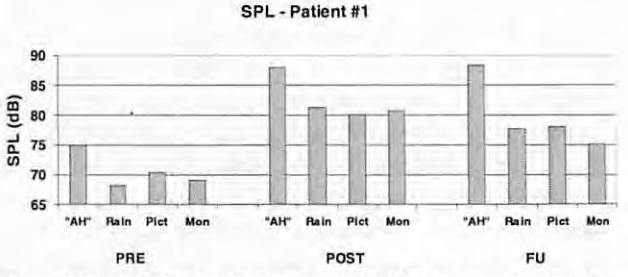


Figure 1. Patient # 1's SPL during sustained "AH", reading the "Rainbow Passage", describing a picture, and a monologue.

Table A.
SPL and Duration Data for Patient #1

Measurements	Task	Pre 1	Pre 2	Pre 3	Post 1	Post 2	FU 1	FU 2	pre vs post F =	pre vs post p <	pre vs FU F=	pre vs FU p <
SPL (dB) (30 cm)	sustained "AH"	75.01 (2.41)	76.05 (3.47)	73.79 (0.95)	89.54 (2.68)	86.65 (2.25)	87.28 (2.40)	89.48 (2.61)	92.357	0.005	130.41	0.005
SPL (dB) (30 cm)	"Rainbow"	67.91 (2.31)	66.96 (3.34)	69.77 (3.2)	81.79 (1.96)	80.81 (2.44)	77.86 (2.3)	77.78 (2.01)	135.02	0.005	81.24	0.005
SPL (dB) (30 cm)	Picture description	68.15 (2.82)	71.03 (3.05)	72.05 (2.78)	80.69 (2.92)	79.73 (3.56)	78.73 (2.59)	77.23 (2.8)	40.01	0.01	22.17	0.025
SPL (dB) (30 cm)	Monologue	68.74 (3.19)	69.13 (3.33)	69.33 (2.62)	79.72 (3.11)	81.95 (2.93)	75.65 (4.05)	75.09 (3.4)	186.98	0.001	223.75	0.001
Duration (sec)	sustained "AH"	12.3 (1.77)	11.76 (2.82)	12.96 (1.76)	24.88 (0.89)	26.17 (2.71)	21.71 (4.74)	22.03 (3.5)	401.79	0.001	421.85	0.001

tual rating of the voice loudness improved significantly from pre- (35.0) to post-treatment (2.65) for the "Rainbow passage" and from pre- (46.5) to post-treatment (1.63) for the picture description. The magnitude of improvement was significant ($p < 0.001$) for both the "Rainbow passage" and the picture description. Finally, 4 listeners (100%) judged the voice to be louder post-treatment.

Review of the speech-language summary reports suggested significant improvement in functional communication and in quality of life. Prior to therapy, this patient experienced difficulties communicating with others, especially in noisy situations. She often had to repeat what she had said and speak with high vocal effort for people to understand her. Speaking in front of a group was also difficult for her. As a singer, she had difficulties singing the typical range she used to have pre-morbidly. After the LSVT® program, she reported increased confidence in her voice and herself, a stronger voice, much less vocal effort, and clearer speech, with people no longer asking her to repeat what she said. She also reported that her singing improved, although it was still limited compared to what it was pre-morbidly.

Patient # 2

As can be seen in Tables C and D, with a few exceptions, Patient #2 also showed significant improvement in pre- to post-, and pre- to FU in SPL, duration, and perceptual measures, indicating that for this individual, LSVT® resulted in marked improvement in vocal function. Specifically, the voice significantly increased in SPL for sustained "AH" phonation (average pre- to post increase 22.9 dB, $p < 0.005$; pre- to FU increase: 21.3 dB, $p < 0.005$), for the "Rainbow Passage" (pre-post: 9.6 dB, $p < 0.05$; pre-FU: 7.4 dB, $p < 0.05$), and for the monologue (pre-post: 8.8 dB, $p < 0.05$; pre-FU: 3.5 dB, $p < 0.05$). SPL did not increase significantly for the picture description (pre-post: 4.4 dB, $p > 0.05$; pre-FU: 2.1 dB, $p > 0.05$). The changes in SPL from pre- to post-treatment and from pre- to follow up for Patient #2 are illustrated graphically in Figure 2.

Duration of sustained "AH" phonation increased significantly from pre- to post-treatment (6.6 sec increase, $p < 0.005$) but not from pre- to FU (6.6 sec increase, $p > 0.05$). Perceptual rating of the vocal loudness improved significantly from pre- (15.6) to post-treatment (2.59) for the "Rainbow passage" and from pre- (14.8) to post-treatment (5.59) for the picture description. The magnitude of improvement was significant for both the "Rainbow passage" ($p < 0.002$) and the picture description ($p < 0.021$). Finally, 4 listeners (100%) judged the voice to be louder post-treatment for the "Rainbow passage", and 3 listeners (75%) for the picture description.

Table D.
Perceptual Data for Pating #2

		Pre	Post	SEM	pre vs post F =	p <
Percept rating of loudness 0 = normal, 100 = too soft	"Rainbow"	15.6	2.59	2.059	10.78	0.002
	Picture description	14.8	5.59	2.41	5.71	0.021
Perceptual comparison of pre vs. post "louder" (N=4)	"Rainbow"	0	4			
	Picture description	0	3			

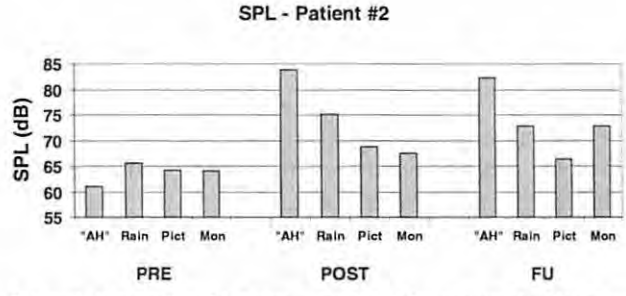


Figure 2. Patient #2's SPL during sustained "AH", reading the "Rainbow Passage", describing a picture, and a monologue.

Table C.
SPL and Duration Data for Patient #2

		Pre 1	Pre 2	Pre 3	Post 1	Post 2	FU 1	FU 2	F =	p <	F=	p <
SPL (dB) (30 cm)	sustained "AH"	59.81 (3.36)	62.34 (2.59)	60.81 (2.77)	83.90 (2.88)		82.97 (1.47)	81.58 (1.78)	242.52	0.005	387.22	0.005
SPL (dB) (30 cm)	"Rainbow"	63.40 (2.14)	66.06 (2.59)	57.13 (2.46)	75.11 (1.36)		72.87 (3.86)	71.09 (2.33)	18.66	0.05	16.71	0.05
SPL (dB) (30 cm)	Picture description	63.13 (2.2)	63.22 (2.26)	66.69 (2.96)	68.67 (2.24)		66.37 (2.08)	68.52 (3.56)	3.4	ns	3.28	ns
SPL (dB) (30 cm)	Monologue	62.43 (1.94)	64.67 (2.41)	64.76 (2.21)	72.80 (1.92)		67.46 (1.96)	67.25 (2.15)	33.69	0.05	11.88	0.05
Duration (sec)	sustained "AH"	21.61 (1.64)	20.87 (1.44)	21.0 (1.69)	27.76 (2.22)		23.31 (1.87)	32.31 (0.73)	209.29	0.005	3.9	ns

A review of the speech-language summary reports of this individual indicated significant improvement in her ability to communicate and in other aspects of her life. After the LSVT® program she wrote: "[my voice] definitely feels stronger", "I speak more clearly and louder", "with not a lot of effort", and "don't have people asking me to repeat [what I said]".

Discussion

In this study, the two individuals with MS who were treated with the LSVT® program showed significant improvement in vocal SPL duration of sustained vowel phonation and perceptual ratings of vocal loudness. The improvement in vocal function was of significant magnitude, and this effect was maintained 6 months following treatment. Moreover, the subjects' experiences, as documented in the speech-language summary reports, clearly indicate that they gained a great deal from the treatment program in terms of their ability to communicate more effectively, with much less vocal fatigue, and with significantly more confidence. These findings suggest that intensive phonatory-respiratory treatment can produce clinically significant and long-term improvement in vocal function in individuals with MS.

The physiologic mechanisms underlying these changes were not studied here, but previous studies have shown that similar improvement with the LSVT® in individuals with PD is associated with improved vocal fold adduction and vocal fold vibratory movements.[10,17,24] These physiologic events probably underlined the changes observed in the present study, suggesting that the LSVT® improved vocal fold vibratory efficiency. It is also likely that increased respiratory drive contributed to the improvement observed here.

The present findings contradict the recommendation by Hartelius et al [8] that individuals with MS should not be treated with an intensive program due to the patient's vocal weakness and fatigue. In fact, our findings are consistent with Petajan and his colleagues' [9] suggestion that intensive training can actually improve function in those parts of the body that are prone to fatigue. Our findings are also in line with the suggestion of Farmakides and Boone [7] that vocal training probably minimizes disuse atrophy and increases muscle strength and coordination, in spite of the patient's pre-treatment tendency to experience vocal fatigue.

The present findings are encouraging. The specific neuropathologic mechanisms underlying the vocal symptoms in the patients of this study are yet to be established. Given that the LSVT® has been proven effective in individuals with other neurologic disease such as PD [15,17,24] (also Baumgartner C, Sapir S, Ramig L. Perceptual voice quality changes following phonatory-respiratory effort treatment (LSVT®) vs. respiratory effort treatment for individuals with Parkinson disease. J Voice submitted) it is likely that the LSVT® will be of clinical benefit to other individuals with MS, regardless of the specific neuropathologic mechanisms underlying their vocal fatigue and weakness. Importantly, LSVT® has a significant impact not only on voice loudness but also on other aspects of speech, such as voice quality, articulation and prosody.[10]

(Also Baumgartner C, Sapir S, Ramig L. Perceptual voice quality changes following phonatory-respiratory effort treatment (LSVT®) vs. respiratory effort treatment for individuals with Parkinson disease. J Voice submitted). Moreover, there is evidence to suggest that LSVT® can significantly improve swallowing function in dysarthric individuals with dysphagia.[30-31] Unfortunately, only a small percentage of individuals with chronic neurologic disease receive treatment to improve voice and speech. For example, Hartelius and Svenson [2] reported that of the 460 individuals with PD or MS they studied, 44-70% experienced voice and speech impairment since the onset of their disease, yet only 3% of the PD individuals and 2% of the MS patients received therapy for their communicative problems. There is a need for additional studies to demonstrate the positive effects of LSVT® and other treatment methods on voice and speech in individuals with neurologic impairment. Such studies will not only improve treatment, but also increase awareness and encourage physicians and other clinicians to refer dysarthric individuals for treatment.

Acknowledgement

This work was supported in part by NIH grant #R01DC01150.

References

1. Darley F, Brown J, Goldstein N. Dysarthria in multiple sclerosis. J Speech Hear Res 1972; 15: 229-245.

2. Hartelius L, Svensson P. Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: A survey finding. Folia Phoniatrica et Logopedica 1994; 46:9-17.

3. Darley F, Aronson A, Brown J. Motor speech disorders. Philadelphia, Saunders, 1975.

4. FitzGerald F, Murdoch B, Chenery H. Multiple sclerosis: Associated speech and language disorders. Australian J Human Commun Disord 1987;15:15-33.

5. Hartelius L, Buder E, Strand E. Long-term phonatory instability in individuals with multiple sclerosis. J Speech Hear Res 1997;40:1056-72.

6. Daly D, Code C, Andersen H. Disturbances of swallowing and esophageal motility in patients with multiple sclerosis. Neurology 1962 ;12:250-256.

7. Farmakides M, Boone D. Speech problems of patients with multiple sclerosis. J Speech Hear Disord 1960; 25:385-90.

8. Hartelius L, Wising C, Nord L. Speech modification in dysarthria associated with multiple sclerosis: An intervention based on vocal efficiency, contrastive stress, and verbal repair. J Med Speech-Lang Path 1997; 5:113-40.

9. Petajan J, Gappmaier E, Whit A, Spencer M, Mino L, Hicks R. Impact of aerobic training on fitness and quality of life in multiple sclerosis. Ann Neurol 1996; 39:432-41.

10. Ramig L, Countryman S, O'Brien C, Hoehn M, Thompson L. Intensive speech treatment for patients with Parkinson's disease: Short- and long-term comparison of two techniques. Am Acad Neurol 1996; 47:1496-1504.

11. Ramig L, Mead C, Scherer R, Horii Y, Larson K, Kohler D. Voice therapy and Parkinson's disease: A longitudinal study of efficacy. Paper presented at the Clinical Dysarthria Conference, 1988, San Diego, CA.

12. Ramig L, Bonitati C, Lemke J, & Horii Y. Voice treatment for patients with Parkinson disease: Development of an approach and preliminary efficacy data. J Med Speech-Lang Path 1994; 2:191-209.

13. Ramig L, Pawlas A, Countryman S. The Lee Silverman Voice Treatment (LSVT): A Practical Guide to Treating the Voice and Speech Disorders in Parkinson Disease. Iowa City, IA: National Center for Voice and Speech, 1995.

14. Brosovic G. Voice therapy and Parkinson disease: Measures of vocal fold adduction. Unpublished undergraduate honor's thesis. 1994, University of Colorado, Boulder.

15. Dromey C, Ramig L, &Johnson A. Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. J Speech Hear Res_1995; 38:751-63.

16. Ramig L, Dromey C. Aerodynamic mechanisms underlying treatment-related changes in SPL in patients with Parkinson disease. J Speech Hear Res 1996; 39:798-807.

17. Smith M, Ramig L, Dromey C, Perez K, Samandari R. Intensive voice treatment in Parkinson's disease: Laryngostroboscopic findings. J Voice 1995; 9:453-9.

18. Schulman R. Articulatory dynamics of loud and normal speech. J Acoust Soc Am 1989 ;85:295-312.

19. England A, Schwab R. The management of Parkinson's disease. AMA Arch Int Med 1959;104:439-68.

20. Hallet M, Khosbin S. A psychological mechanism of bradykinesia. Brain 1980;103:301-14.

21. McDowell F, Lee J, Sweet R. Extrapyramidal disease. In Baker A, Joynt R eds. Clinical Neurology. Philadelphia: Harper and Row, 1986; 24-26.

22. Adams J. A closed-loop theory of motor learning. J Motor Behav 1971; 3:111-50.

23. Schmidt R. Motor control and learning. Champaign, IL: Human Kinetic Publishers, 1988.

24. Ramig L, Countryman S, Thompson L, Horii Y. A comparison of two forms of intensive speech treatment for Parkinson disease. J Speech Hear Res 1995; 38:1232-51.

25. Aronson A. Clinical Voice Disorders. New York: Thieme-Stratton, 1990.

26. Froeschels E, Kastein S, Weiss D. A method of therapy for paralytic conditions of the mechanisms of phonation, respiration and glutintaion. J Speech Hear Disord 1955; 20:3645-70.

27. Fairbanks G. Voice and Articulation Drillbook. New York: Harper and Brothers, 1960.

28. Goodglass H, Kaplan E. The assessment of aphasia and related disorders. Boston Diagnostic Aphasia Examination (2nd edition). Philadelphia:Lea and Febiger, 1983.

29. Fox C, Ramig L. Vocal sound pressure level and self-perception of speech in men and women with idiopathic Parkinson disease. Am J Speech-Lang Path 1997; 6:85-94.

30. de Angelis E, Mourao L, Ferraz H, Behlau M, Pontes P, Andrade L Effect of voice rehabilitation on oral communication of Parkinson's disease patients. Acta Neurol Scand 1997; 96(4):199-205.

31. Merson R, Rolnick M Speech-language pathology and dysphagia in multiple sclerosis. Phys Med Rehabil Clin N Am 1999; 9:631-41

NCVS Status and Progress Report - 14
September 1999, 149-161

Speech and Voice Abnormalities in Parkinson Disease: Relation to Severity of Motor Impairment, Duration of Disease, Medication, Depression, Gender and Age

Shimon Sapir, Ph.D., CCC-SLP
Annette Pawlas, M.A., CCC-SLP
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Lorraine Ramig, Ph.D., CCC-SLP
The University of Colorado-Boulder
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Stefanie Countryman, M.A., CCC-SLP
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Christopher O'Brien, M.D.
Colorado Neurological Institute, Englewood, Colorado
Margaret Hoehn, M.D.
Denver, Colorado
Leatitia A. Thompson, Ph.D.
Department of Neuropsychology, The University of Colorado Health Science Center

Abstract

The neurologic mechanisms underlying speech and voice disorders in individuals with idiopathic Parkinson disease (IPD) are poorly understood. The purpose of this study was to address this issue by assessing the relationships between speech (articulation, fluency, prosody) and voice (e.g., hoarse, harsh, breathy), rated perceptually, severity of motor impairment (measured by UPDRS, section III), disease duration, depression (measured by the MADRS), age and gender in 42 individuals with IPD. Abnormal voice was most prevalent (86%), followed by articulation (50%), fluency (33%) and prosody (33%). Sixty three percent of individuals had abnormal voice in the absence of speech problems. Individuals with low UPDRS scores, short duration of IPD, and low MADRS scores were more likely to have voice abnormalities alone. Individuals with high UPDRS scores, long duration of IPD, and high MADRS scores were significantly more likely to have both speech and voice abnormalities. Speech and voice abnormalities did not correlate with age or gender. The prevalence of abnormal voice and articulation for these medicated individuals was nearly identical to that reported previously for non-medicated individuals with PD. These findings suggest that voice is affected by Parkinson disease before speech abnormalities which develop later in the disease process. Although both speech and voice abnormalities in PD have been traditionally linked to rigidity and dopamine deficiency, we interpret our findings as well as others to suggest alternative neuropathologic mechanisms for these abnormalities.

Introduction

Parkinson disease (PD) affects over 1.5 millions of Americans.[1] The great majority of these individuals will develop speech abnormalities during the course of their disease.[2-5] These abnormalities will likely have adverse effects on the ability of these individuals to communicate effectively, and on their economic, social, and psychological well being.[6-9] While the etiology, physical pathologies and neurophysiology of PD have been investigated extensively,[10-11] comprehensive studies of speech and voice characteristics of PD and their development during the course of the disease have been scarce. Such information is needed to help elucidate the neuropathologic mechanisms underlying these

speech abnormalities, and to develop more effective methods to diagnose, treat and prevent communicative disorders in individuals with PD.[12,8]

Epidemiological studies are often used to identify symptom clusters that are unique to a particular neurologic disease and that can reflect the severity or stages of the disease. A similar approach has been taken to characterize speech and voice abnormalities that are associated with a particular neurologic disease.[3,13-15] Logemann et al. [3] attempted to characterize the speech abnormalities acquired in association with PD and to stage the course of these abnormalities as the disease progresses. To this end, they studied 200 individuals with PD. These individuals were not medicated during the time of the study. Some of the newly diagnosed individuals had never been on medication for their PD symptoms, whereas others were off their medication for at least two weeks prior to the study. Most individuals had idiopathic PD but some had postencephalitic PD. These individuals represented all five stages of the disease, as determined by the Columbia system.[16] The presence of specific voice, articulation, and other speech abnormalities was determined by perceptual phonetic analyses of the individuals' speech during the reading of a paragraph and during conversation. Two expert listeners, both speech pathologists, performed the phonetic analyses. Logemann et al. [3] reported that 89% of the individuals had abnormal voice (e.g., breathiness, roughness, hoarseness, tremulousness) and 45% had abnormal articulation. Other speech abnormalities were present but were less frequent. Of the individuals with abnormal voice, 45% had only voice abnormality, whereas the others had an additional articulatory disorder. The articulatory abnormalities tended to cluster for subgroups of individuals. For example, some individuals had articulation errors associated with back-tongue movement (e.g., "k" and "g"), other individuals had additional problems with tongue-blade articulation (e.g., "s", "z", "sh", "dg", and "ch"), and still other individuals had additional problems with lip articulation (e.g., "p", "b", "f', "v"). Logemann et al. [3] interpreted their data to suggest that either the subgroups represented specific patterns of neurologic lesions, or that these clusters of symptoms represented a progression in dysfunction, beginning with phonation, and gradually extending to include the articulation and other aspects of speech (e.g., resonance and prosody).

While Logemann, et al. [3] identified high incidence and early onset of voice disorders in individuals with PD, they did not relate their findings to factors such as severity of motor impairment, duration of illness, medication, depression, gender and age. Such observations may provide important insights regarding the neurophysiology, neuropathology, and progression of the speech abnormalities associated with PD. This information is necessary to develop more effective methods of treating speech and voice problems in individuals with PD.

It is commonly assumed that speech and voice abnormalities in individuals with IPD are related to dopamine deficiency and rigidity.[17-22] Thus, given the progressive nature of the disease, one would expect to see strong correlations between the prevalence and number of speech and voice abnormalities and the severity of motor impairment and duration of illness in individuals with PD. If speech and voice abnormalities in PD are related to factors other than rigidity and dopamine, the correlation between these abnormalities, severity of motor impairment, and duration of PD may not be strong.

It has been repeatedly reported [23-26] that medication has little or no ameliorating effect on speech and voice defects in individuals with PD. However, it is not clear to what extent medication for PD may actually have an adverse effect on speech and voice, given that such medication can produce iatrogenic complications in some individuals, for example, dyskinesia, tremor, and affective disorders.[15,27-30] Potentially, these motor and affective disorders can affect speech and voice (Darley, Aronson, & Brown, 1975; Sapir & Aronson, 1990). Thus, to clarify the relationship between medication for PD and speech and voice, it may be useful to compare individuals with IPD who are medicated with those who are not medicated for PD. If the two groups show similar symptomatology and prevalence of symptoms, this would suggest that the symptoms are related to the disease rather than to medication. If on the other hand the medicated group has significantly higher prevalence of speech and voice abnormalities, this may suggest that some of these abnormalities may be related to medication.

Between 40% and 50% of individuals with PD suffer from depression or a related mood disorder. [31] Mood disorders in individuals with neurologic disorders may be premorbid, reactive,[32] or endogenous, related to disturbances in the basal ganglia, thalamus, limbic system, and frontal lobe.[33-36] Mood disorders are also known to affect voice, speech and language functions,[30,37-40] as well as to be affected by communication disorders.[30,41] Thus, it is not clear to what extent speech and voice abnormalities in PD are related to depression, and what the direction of causality might be.

Gender may be another variable that affects speech and voice in individuals with PD. There are anatomical and physiologic differences between men and women involving the phonatory and vocal tract systems.[42-45] Also, progesterone and estrogen can affect symptoms of PD [46-47] and these hormones are known to affect the voice.[48-50] Thus, different voice symptomatology is expected in men and women with PD. Indeed, a recent study by Hertrich & Ackermann, [51] provides acoustic and electroglottographic evidence for gender-specific vocal dysfunction in PD. Other gender differences have been observed in individuals with PD. For example, Lyons, et al [28] concluded from their study that as PD progresses, gender differences emerge, with men exhibiting more severe Parkinsonian motor features and women experiencing more levodopa-induced dyskinesia. Given these findings, one would expect to see gender-re-

lated differences not only in voice, but also in other aspects of speech.

Speech and voice abnormalities in individuals with PD may be related to the normal aging process, or to an interaction between this process and PD and medication. Several studies have demonstrated significant differences in respiratory and phonatory functions in young and old normal adults [43,52-56] as well as differences between individuals with PD and individuals without neurologic disease. [53,57] The purpose of the present study was to provide information on the relationships between speech and voice abnormalities in individuals with PD and their relationship with severity of motor impairment, duration of PD, medication, depression, gender, and age.

Methods

Subjects

Forty-two individuals (32 males and 10 females) with IPD participated in this study. Their biographical and medical data are summarized in table A. These data include gender, age, duration of disease, Hoehn & Yahr stage of Parkinson disease, score on the motor section of the Unified Parkinson's Disease Rating Scale (UPDRS),[58-59] score on the Montgomery Asberg Depression Rating Scale (MADRS), and specific anti-Parkinson and anti-depression medication.

These individuals sought treatment for their speech and voice problems, and participated in the present study prior to treatment. They had received no prior voice/speech treatment at the time of this study. Their average age was 64.0 years (SD =10.7, range 32 - 83 years). The diagnosis of IPD was determined by a neurologist specializing in movement disorders. The average Hoehn and Yahr stage of disease [60] was 2.6 (SD = 0.7, range: 1 - 4). The average duration of the disease, measured since it was first diagnosed, was 6.8 years (SD = 5.2, range: 1 - 20 years). The average score on the motor section of the UPDRS was 25.9 (SD = 13.8, range: 1 - 48). All individuals were receiving anti-Parkinson medications, and some were also on medication for depression. UPDRS motor examinations were completed while individuals were "on" their anti-Parkinson medications and reflect bilateral ratings. All individuals on medication were considered neuropharmacologically stable and optimally medicated by their neurologist. Speech data were collected while the individuals were 'on' their medication. These individuals were recorded approximately one hour after receiving their medication.

A clinical neuropsychologist rated each individual for depression using the Montgomery Asberg Depression Rating Scale (MADRS). [61] Results revealed a mean score of 6.7 (SD= 4.4, range: 1-19). Scores of 6 or greater are indicative of depression. By this criterion, twenty-three individuals (55%) in this study had some degree of depression.

Data Collection of the Speech and Voice Samples

Individuals were seated in an IAC sound-treated booth with a headset microphone (AKG 410) positioned 8 cm in front of their lips. They were asked to read aloud a standard phonetically balanced passage entitled the "Rainbow Passage." [62] This passage takes about 60 seconds to

Table A.
Biographical, medical, and voice/speech data for forty-two individuals with IPD, listed by gender and age and arranged in columns by gender, age, duration of PD, Hoehn & Yahr Scale score, Unified Parkinson's Disease Rating Scale score, medication, and abnormal articulation, fluency, prosody, and voice quality. The right most column indicates the total number of speech and voice abnormalities for each patient.

	gend	age	dur	H&Y	UPDRS	MADRS	meds	artic	fluen	pros	voico	total
1	F	32	1	2	36	9	a,j					0
2	F	49	9	2	4	14	a,d,g				1	1
3	F	52	13	3	17	7	a,f,n	1			1	2
4	F	63	3	3	42	19	a,c	1		1	1	3
5	F	64	2	3	29	9	a,c,o					0
6	F	68	1	2	1	0	a,b,c					0
7	F	68	4	2	12	7	a,c		1		1	2
8	F	73	10	4	48	12	a,c,f	1	1	1	1	4
9	F	74	1	1.5	12	11	b,c				1	1
10	F	83	7	3	33	11	a	1	1	1	1	4
11	M	49	2	2	13	3	a,b,c				1	1
12	M	50	18	2	19.5	8	a,c,d					0
13	M	51	6	2	3	1	a,c	1			1	2
14	M	51	6	1.5	3	6	a,c	1		1	1	3
15	M	52	14	3.5	32	6	a,d,e,o	1	1	1	1	4
16	M	52	18	3	34	8	a,c,d,e	1	1		1	3
17	M	54	3	1	3	5	b,c,h					0
18	M	54	15	3.5	45.5	5	b,l,k		1		1	2
19	M	56	7	2	2	3	a,c,h					0
20	M	61	4	3	22.5	13	a,c,d	1	1	1	1	4
21	M	61	1	2	16	4	a,c				1	1
22	M	62	3	2	26	7	a,c				1	1
23	M	62	4	3.5	35	5	k,m	1		1	1	3
24	M	64	7	2	26	1	a				1	1
25	M	64	10	3	43	13	a	1	1	1	1	4
26	M	65	8	3	19	3	a,e,h			1	1	2
27	M	66	12	3	41	11	a,c,l	1	1	1	1	4
28	M	67	18	3.5	32	3	b,c,j	1	1		1	3
29	M	67	8	3	16	8	a				1	1
30	M	68	6	3	40.5	7	a,k				1	1
31	M	69	6	2	36	2	a,k	1		1	1	3
32	M	70	4	3	32	1	a,c	1	1		1	3
33	M	71	4	3	43	3	a,m	1			1	2
34	M	71	11	3.5	43	6	j,h	1	1		1	3
35	M	72	5	3	30	3	a,c,h			1	1	2
36	M	73	10	3	27	5	a,h	1	1		1	3
37	M	75	7	2	32	13	a,c		1		1	2
38	M	75	20	3.5	47	15	b,m	1		1	1	3
39	M	76	1	2	20	0	a				1	1
40	M	76	5	3	16	3	b,c,l	1			1	2
41	M	79	1	3	33	6	a,c	1		1	1	3
42	M	81	2	2	24	5	a,c,m				1	1
ave	64.0		7.1	2.6	25.9	6.7	sum=	21	14	14	36	2.0
SD=	10.7		5.3	0.7	13.8	4.5	freq.=	50%	33%	33%	86%	1.3

Parkinson medication:
a = sinemet, b = sinemet-cr, c = eldyprel, d = symmetrel, f = pergolide, g = cogentine, h = artane, I = amantadine, j = permax, k = deprenyl, l = parsidol. Anti-depressants: n= amitriptyline, 0 = tofranil, p = elavil. Higher socres on the UPDRS indicate greater severity of motor impairment.

read and is at approximately a 6th grade reading level. After preamplification through an ATI-1000 amplifier, the microphone signal was recorded onto a Sony Digital PC-108M (DAT) eight-channel tape recorder. A Bruel and Kjaer Type 2230 sound level meter was placed in the booth 50 cm from the individual's mouth. The signal from the sound level meter also was recorded onto the DAT. All data were collected by the same researcher.

Development of the "Master" Audio Tapes

A computer random number generator determined the ordering of the 42 reading samples which were subsequently dubbed onto two "master" digital audio tapes (DAT). Eighteen percent (18%) of the samples were repeated to assess intra-judge reliability.

Since intensity levels may confound perceptual judgments of voice quality and speech intelligibility, [63] intensity levels were normalized across all reading samples during the dubbing procedure. These samples were presented to the listeners at 70 dB SPL at a distance of 50 cm.

Speech and Voice Characteristics

Four major categories were evaluated perceptually: voice quality, articulation, fluency, and prosody. The listeners had to determine which of these categories was abnormal for each sample they heard, based on the following definitions. Voice quality was defined as abnormal if it had one or a combination of these qualities: hoarse, harsh, rough, breathy, glottal fry, "mucus"-like/crackle/"wet", tremulous, pressed, strained/strangled, or unsteady. Articulation was defined as abnormal if it was characterized by any or a combination of the following: imprecise consonant articulation (consonants are distorted, slurred, reduced in sharpness or crispness), omissions of phonemes (consonants, vowels, etc.) from words, effortful production of phonemes, and substitutions of phonemes. Fluency was considered abnormal if it was characterized by any or a combination of these: prolongations of phonemes, repetition of phonemes at the beginning of syllables or words, repetition of words, repetition of parts of words, and pallilalia. Prosody was defined as abnormal if it was characterized by any or a combination of the following: bizarre or unnatural stress patterning, intonation, or rate-rhythm.

Listeners

Two expert listeners, both certified speech/language pathologists, participated in the perceptual judgment of the speech and voice of each individual. Each listener had extensive training and experience in the perceptual analysis of abnormal speech in individuals with dysphonia and dysarthria. These listeners had at least 5 years of clinical experience, were females with normal hearing, and unfamiliar with the subjects.

Training

Prior to completing the listening procedure, the listeners participated in a three-hour session to familiarize themselves with the experimental procedures. They reviewed the listening procedure, the rating form, the computerized scanning form (i.e. answer sheet), the definitions and examples of the speech and voice characteristics to be rated, and the operation of the equipment. On a different day they were administered a practice listening session using two individual samples that were not part of the data pool.

Listening Procedure for Rating Speech and Voice Characteristics

The listeners individually rated the samples while seated in an IAC sound-treated booth. They were informed that they were rating the speech and voice of individuals with Parkinson disease. To reduce the potential for learning effects on the ratings, the order of the master tapes was randomized across listeners, thereby allowing each listener to hear the samples in a different order. To limit fatigue, the listeners were instructed to rate a maximum of 2 hours per session and up to 6 hours per week.

The tapes were played on a Technics Digital Audio Tape Deck (SV-DA10) through a Technics Stereo Integrated Amplifier (SU-V303). On the master tapes, each individual was identified by a number, age and sex. To ensure that the master tapes were played at a constant intensity level, the listeners were instructed to adjust the intensity to a comfortable level at the beginning of each listening session and to keep this intensity level fixed throughout each session. The listeners were instructed to play each reading sample as often as necessary in order to accurately rate the speech and voice characteristics.

In the present study we were interested in the prevalence of speech and voice abnormalities rather than the severity of these abnormalities in individuals with PD. Therefore, for each speech sample presented to the listener, the listener had to indicate whether each of the four major categories (voice, articulation, fluency, and prosody) was abnormal, using "true" or "false" replies. The present/absent rating paradigm is consistent with previous descriptive speech and voice studies in Parkinson disease.[3] The ratings were recorded on a computerized scanning form for data analysis.

Reliability Measures

Intra-judge reliability was calculated as percent agreement for the 18% of the samples which were repeated randomly on the tape. Percent intra-judge agreement for voice, articulation, fluency, and prosody was 100%, 86%, 100%, and 86% for one listener, and 88%, 100%, 88%, and

88% for the other listener, respectively. To consider any of the four major speech categories in the individual's speech as abnormal, both listeners had to indicate that such abnormality was present. Thus, the prevalence data to be reported below are based on 100% agreement between the two listeners. Intra-subject reliability for reading of the "Rainbow Passage" on two different occasions (within a week) has been reported in a previous study using the same individuals as the subjects here.[64] In that study, repeated reading produced highly consistent acoustic measures, indicating no significant differences in the speech and voice of these subjects during the two readings.

Comparison of Medicated and Unmedicated Individuals With PD

In order to assess the effects of medication on speech and voice in individuals with PD, the prevalence data from the 42 individuals in this study was compared with the prevalence data from the 200 umedicated individuals reported by Logemann et al. [3] Given that the definitions and inclusions of particular speech abnormalities were not always identical across the two studies, we chose to include only two major categories that were comparable across the studies. These were abnormal voice (termed "laryngeal disorders" by Logemann et al. [3]) and articulation.

Table B.
Mean (Standard Deviation), and Prevalence Data of Voice/Speech and Non-Voice/Speech Variables with Chi-Squire and ANOVA Analyses
p Values are at the bottom of each section
p > 0.05 is not significant (ns)

MOTOR IMPAIRMENT	n=	age	duration	H & Y	UPDRS	MADRS	articul.	fluency	prosody	voice	sum speech
updrs 1 - 19.5	15	59.4 (9.5)	6.1 (4.7)	2.1 (0.6)	10.4 (6.9)	5.5 (3 8)	27%	7%	13%	73%	1.2 (0.9)
updrs 32 - 48	19	65.6 (11.7)	9.0 (5.9)	3.1 (0.6)	38.3 (5.6)	8.2 (4 8)	79%	58%	53%	95%	2.8 (1.1)
F or X^2		2.741	2.348	***	***	2.989	7.294	7.52	4.078	1.593	21.961
p <		ns	ns	***	***	ns	0.01	0.01	0.05	ns	0.0001

DURATION	n=	age	duration	H & Y	UPDRS	MADRS	articul.	fluency	prosody	voice	sum speech
1 - 3 yrs	12	63.6 (13.9)	1.8 (0.9)	2.1 (0.6)	21.3 (12 8)	6 5 (5.2)	17%	0%	17%	67%	1.0 (1.0)
10 - 20 yrs	11	61.5 (9.6)	14.5 (3.6)	3.2 (0.5)	36.5 (10 7)	8.5 (3.5)	82%	73%	45%	91%	2.9 (1.2)
F or X^2		0.1787	***	***	9.5695	1.1673	7.326	10.368	1.093	0.814	16.324
p <		ns	***	***	0.01	ns	0.01	0.01	ns	ns	0.0001

DEPRESSION	n=	age	duration	H & Y	UPDRS	MADRS	articul.	fluency	prosody	voice	sum speech
madrs 0-3	13	65.7 (8.7)	5.7 (4.3)	2.5 (0.68)	21.0 (13.7)	2.0 (1.2)	46%	15%	23%	85%	1.76 (1.0)
madrs 11-19	10	68.3 (9.6)	8.3 (5.4)	2.8 (0.8)	32.5 (15.10)	13.2 (2.4)	70%	60%	70%	100%	3.0 (1.2)
F or X^2		0.462	1.665	***	3.826	***	0.5175	3.1829	3.335	0.304	7.5825
p <		ns	ns	***	ns	***	ns	ns	ns	ns	0.025

AGE	n=	age	duration	H & Y	UPDRS	MADRS	articul.	fluency	prosody	voice	sum speech
age 32 -56	12	50.2 (6.1)	9.3 (6.1)	2.3 (0.8)	17.5 (15.6)	6.3 (3.4)	42%	25%	17%	67%	1.5 (1.4)
age 70 - 83	14	74.9 (3.98)	6.3 (5.2)	2.8 (0.7)	31.4 (11.1)	6.7 (4.8)	64%	43%	36%	100%	2.5 (1.0)
F or X^2		***	1.8865	***	6.8309	0.0786	0.5757	0.2923	0.42	3.2517	2.1895
p <		***	ns	***	0.025	ns	ns	ns	ns	ns	ns

GENDER	n=	age	duration	H & Y	UPDRS	MADRS	articul.	fluency	prosody	voice	sum speech
Females	10	62.6 (14.7)	5.1 (4.4)	2.6 (0.8)	23.4 (16.4)	9.9 (5.0)	40%	30%	30%	70%	1.7 (1.6)
Males	32	64.5 (9.4)	7 7 (5.5)	2.6 (0 7)	26.7 (13 1)	5.7 (3 9)	53%	34%	34%	91%	2.1 (1 2)
F or X^2		0.2359	1.8553	***	0.4331	7.886	0.1312	0.0164	0.0164	1.2304	0.2948
p <		ns	ns	***	ns	0.01	ns	ns	ns	ns	ns

Statistical Analyses

The specific methods to assess differences between variables are discussed in the Results section. Prevalence data (expressed in terms of % occurrence) were tested for significant differences using a two-tail Chi-square test (X^2) with correction for continuity. Differences between means were tested for significance with a one way analysis of variance (sums of squares, unequal sample sizes). The strength of a relationship between two variables was tested with a Pearson product correlation coefficient (r). Statistical results with a p value greater than 0.05 were considered non-significant.

Results

The raw data of the perceptual voice/speech analysis are shown in Table A, right columns. "1" indicates the presence of a voice or speech abnormality as determined by both listeners. The right-most column shows the sum of speech and voice abnormalities for each individual. Table B (previous page) provides statistical analyses of the perceptual speech and voice data the non-speech data (gender, age, duration of PD, Hoehn & Yahr stage, UPDRS, and MADRS scores).

General Prevalence

Of the 42 individuals studied, 86% had an abnormal voice quality, 50% had abnormal articulation, 33% had abnormal fluency, and 33% abnormal prosody. Abnormal voice was significantly more prevalent than abnormal articulation (X^2= 14.45, $p < 0.01$), fluency (X^2= 26.435, $p < 0.001$), and prosody (X^2=26.435, $p < 0.001$). Abnormal articulation was not significantly different in prevalence than abnormal fluency and prosody (each X^2= 2.937, $P > 0.05$). Sixteen individuals had normal articulation, fluency, and prosody. Of these individuals, ten (63%) had an abnormal voice quality.

Prevalence of abnormal voice and articulation in medicated (this study) and unmedicated patients (Logemann et al. [3])

The two studies yielded nearly identical results for abnormal voice (89% vs. 86%, X^2=0.7578, $p > 0.05$) and articulation (45% vs. 50%, X^2=0.1771, $p > 0.05$).

Speech and voice abnormalities as a function of severity of motor impairment, duration of PD, depression, age and gender

The prevalence and average number of speech abnormalities as a function of motor impairment, duration of PD, depression, age, and gender are summarized in Table B and illustrated in Figures 1-6. Table B is divided into five sections. In each section, one variable is held as independent variable, and all other variables as dependent variables. The data in each of the independent variables are divided into two categories. The results of the statistical analyses are at the bottom of each section.

Severity of Motor Impairment

(Figures 1 and 6). Thirty four individuals were included in this comparison, those who had low UPDRS scores (< 20; n=15) were in one group, and those who had high UPDRS scores (> 30; n= 19) in another group. In both groups, abnormal voice quality was most prevalent, and this prevalence did not differ significantly as a function of UPDRS level (X^2 = 1.5928, $p > 0.05$). Individuals with high UPDRS scores were significantly more likely to have abnormal articulation (X^2 = 7.294, $p < 0.01$), fluency (X^2 = 7.52, $p < 0.01$) and prosody (X^2 = 4.0783, $p < 0.05$), and more likely to have high number of speech and voice abnormalities ($F = 21.961$ $p < 0.0001$) than individuals with low UPDRS scores. Age, Duration, and MADRS as dependent variables did not differ significantly as a function of low vs. high UPDRS scores.

Duration of PD

(Figures 2 and 6). Thirty three individuals were included in this comparison. Those who were diagnosed with PD recently (1-3 years; n=12) were in one group, and those who had PD for many years (10-20 yrs; n= 11) in another group. In both groups, abnormal voice quality was most prevalent, and this prevalence did not differ significantly across the two groups (X^2 = 0.81362, $p > 0.05$). In-

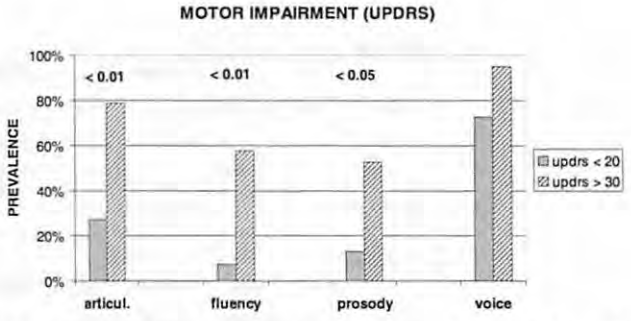


Figure 1. Prevalence of voice quality, articulation, fluency, and prosody abnormalities as a function of severity of motor impairment (measured by the UPDRS, Section III).

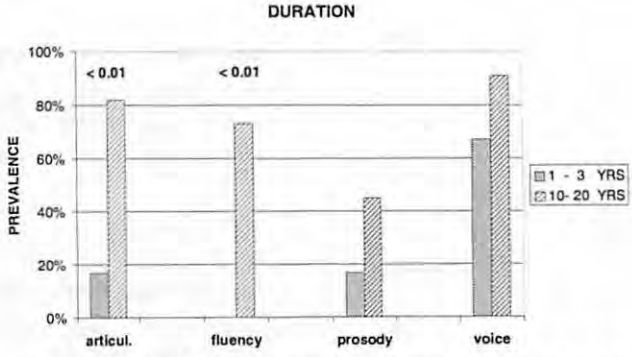


Figure 2. Prevalence of voice quality, articulation, fluency, and prosody abnormalities as a function of duration of PD since first diagnosed.

dividuals with long duration of PD were more likely to have abnormal articulation (X^2 = 7.326, p < 0.01) and fluency (X^2 = 10.368, p < 0.01), and a higher number of speech and voice abnormalities (F = 16.324, p < 0.0001) compared to individuals with relatively recent onset of PD. The prevalence of abnormal prosody did not differ significantly across the two groups (X^2 = 1.093, p > 0.05). As dependent variables, UPDRS differed significantly across the two groups (higher UPDRS scores in individuals with longer duration of PD) (F = 9.5695, p < 0.01) but Age and MADRS did not.

Depression

(Figures 3 and 6). Twenty three individuals were included in this comparison. Those whose MADRS scores were low (0-3; n= 13) were in one group, and those whose MADRS scores were high (11-19 yrs; n= 10) were in another group. In both groups, abnormal voice quality was most prevalent, and this prevalence did not differ significantly across the two groups (X^2 = 0.304, p > 0.05). There was no significant relationship between the two groups in the prevalence of abnormal articulation (X^2 = 0.5175, p > 0.05), fluency (X^2 = 3.1829, p > 0.05) and prosody (X^2 = 3.335, p > 0.05). However, individuals with high MADRS scores were more likely to have high number of speech and voice abnormalities (F = 7.5825, p < 0.025). Age, Duration, and UPDRS as dependent variables did not differ significantly between the low and the high MADRS groups.

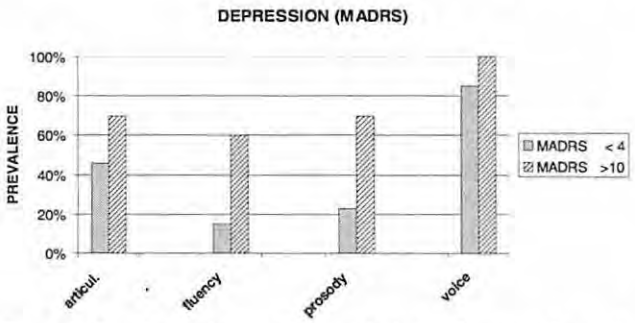


Figure 3. Prevalence of voice quality, articulation, fluency, and prosody abnormalities as a function of depression (measured by the MADRS).

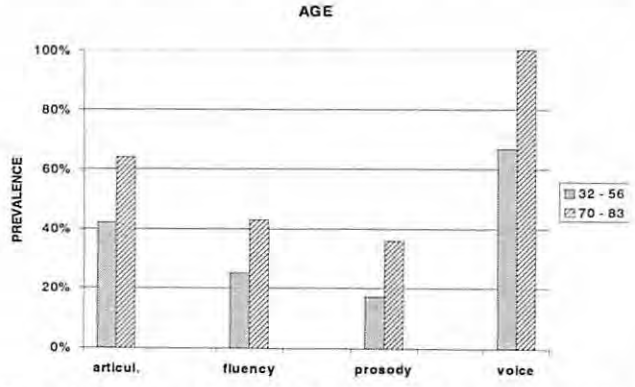


Figure 4. Prevalence of voice quality, articulation, fluency, and prosody abnormalities as a function of age.

Age

(Figures 4 and 6). Twenty six individuals were included in this comparison. Those who were within the age range of 32-56 years were in one group ("younger group"; n= 12), and those within the age range of 70-83 years in another group ("older group"; n= 14). In both groups, abnormal voice quality was most prevalent, and this prevalence did not differ significantly across the two groups (X^2 = 3.2517, p > 0.05). There was no significant difference between the two groups in the prevalence of abnormal articulation (X^2 = 0.5757, p > 0.05), fluency (X^2 = 0.2923, p > 0.05) and prosody (X^2 = 0.420, p > 0.05), nor in the number of speech and voice abnormalities (F = 2.1895, p > 0.05). As dependent variables, UPDRS scores were significantly higher in the older group (F = 6.8309, p < 0.025). Duration and MADRS scores did not differ significantly across the groups.

Gender

All the 42 individuals in this study (10 women, 32 men) were included for comparison. Abnormal voice quality was most prevalent speech disorder in both men and women, with no significant difference in prevalence across gender (X^2 = *1.2304, p > 0.05).* Men and women did not

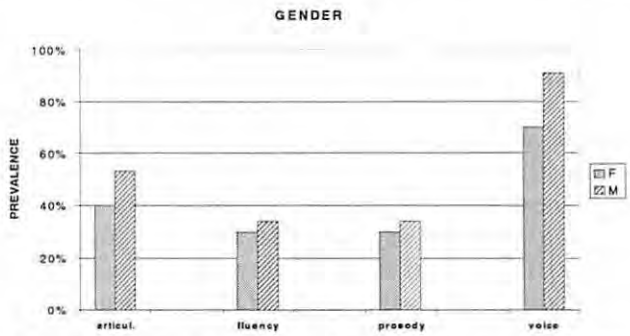


Figure 5. Prevalence of voice quality, articulation, fluency, and prosody abnormalities as a function of gender.

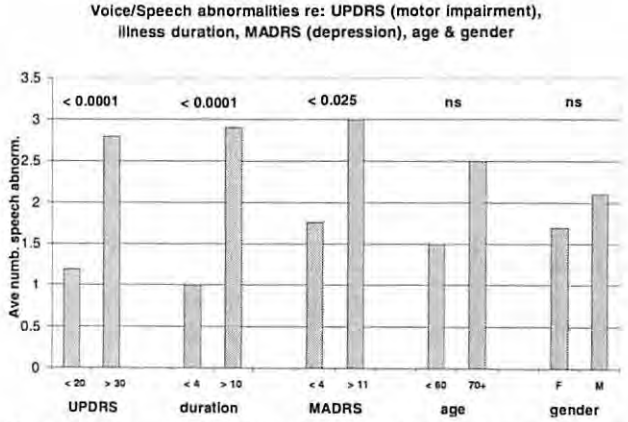


Figure 6. Average number of speech and voice abnormalities as a function of severity of motor impairment (measured by the UPDRS, section III), duration of PD since first diagnosed, depression (measured by the MADRS), age, and gender.

Table C.
Results of Pearson Product Correlation Analyses Between the Number of Speech and Voice Abnormalities and UPDRS, MADRS, and Duration

df = 40. ** p < 0.01 * p < 0.05

Speech and voice vs. UPDRS	r = 0.5422**
Speech and voice vs. Duration	r = 0.3689*
Speech and voice vs. MADRS	r = 0.3081*
UPDRS vs. Duration	r = 0.3356*
UPDRS vs. MADRS	r = 0.3441*
Duration vs. MADRS	r = 0.2180

differ in the prevalence of articulation (X^2 = 0.1312, p > 0.05), fluency (X^2 = 0.0164, p > 0.05), and prosody (X^2 = 0.0164, p > 0.05), nor in the number of speech and voice abnormalities (F = 0.2948, p > 0.05). As dependent variables, MADRS scores were significantly higher in women than in men (F = 7.886. p < 0.01). Age, Duration, and UPDRS scores did not differ significantly across gender.

Correlation Analyses

To assess the strength of the relationship between speech and voice abnormalities and UPDRS, Duration, and MADRS a Pearson product correlation analysis was applied. The results of these analyses are shown Table C. We first correlated the number of speech and voice abnormalities with UPDRS, MADRS, and Duration. The highest correlation was between the number of speech and voice abnormalities and UPDRS scores (r = 0.54, p < 0.01), followed by number of speech and voice abnormalities and Duration (r = 0.37, p < 0.05), and number of speech and voice abnormalities and MADRS (r = 0. 31, p < 0.05). We also correlated UPDRS, Duration, and MADRS with each other. As can be seen in Table C, UPDRS scores correlated with Duration (r = 0.3356, p < 0.05) and MADRS (r = 0.3441, p < 0.05). The correlation between Duration and MADRS scores was low and not statistically significant (r = 0.2180, p > 0.05).

Discussion

The present findings, along with others [3,12,65] suggest that abnormal voice is the most common speech disorder, and probably the first to develop in individuals with PD.

Why the voice is the first and most prominent disorder is unclear. One explanation might be that the neural mechanisms underlying vocalization are different than those underlying articulation, fluency and prosody. More specifically, vocalization is a phylogenically older system, and it appears to involve limbic (e.g., anterior cingulate cortex), brainstem (e.g., periaquaductal gray), and other subcortical systems.[66-68] Articulation, fluency, and prosody are very much related to linguistic, cognitive, and other higher level functions. These functions are subserved largely by neocortical mechanisms [69-71] although they may be influenced by subcortical mechanisms, such as those associated with the thalamic and parathalamic nuclei.[72-73] Thus, it is possible that the pattern of speech and voice deterioration in PD may reflect a progressive involvement of different subcortical and cortical systems. How these systems are related to each other and to the more classical structures associated with PD, such as the substantia nigra and basal ganglia, is not clear.

In this study the prevalence and number of speech and voice abnormalities were significantly correlated with severity of motor impairment and duration of PD. One possible interpretation of these findings is that as the disease progresses, motor impairment and speech and voice abnormalities develop concurrently, presumably because both are related to dopamine deficiency and rigidity. However, this traditional explanation of the present results is problematic for several reasons. Given that the motor section of the UPDRS loads heavily, though not exclusively, on rigidity, and given that dopamine treatment significantly improves rigidity,[59,74] one would expect to see, at the very least, a high correlation between the number of speech abnormalities and the UPDRS scores. But this was not the case in this study. The correlation between these two variables was only moderate (r = 0.54).

There are reasons to suspect that speech abnormalities in individuals with PD are not related to rigidity and dopamine. For example, individuals with PD may have large fluctuations in their rigidity during their "on" or "off" days, yet their speech and voice remain unchanged.[23,75] Also, common dopamine treatment, as well as other treatment methods for PD such as pallidotomy, stereotaxic brain stimulation, and fetal cell grafting, may produce marked reduction in rigidity, but with little or no effect on speech and voice. [25,76-78] Furthermore, one would expect that the prevalence of voice and articulation abnormalities in the present study would be different than that in Logemann et al's study, [3] since the two groups of patients differed in terms of medication (one group was treated with dopamine medication, the other was not). However, the two groups had nearly identical prevalence data, suggesting (assuming that the two groups were comparable in other important variables) that dopamine treatment probably did not affect the prevalence of speech abnormalities in the two groups. Thus, the overall evidence suggests that there is a dissociation between speech and voice abnormalities and rigidity and dopamine deficiency in individuals with PD.

But if speech and voice abnormalities in PD are not related to rigidity and dopamine, what might be their underlying neuropathology? At this point we can only offer a tentative hypothesis, based primarily on inferential evidence.

There is evidence to suggest that speech and voice abnormalities characteristic of PD, such as hypophonia, hypoprosodia, and hypokinetic articulation are also common in individuals with lesions to frontolimbic, parathalamic, and other subcortical areas other than the basal ganglia and substantia nigra.[67,73,79-87] Such lesions are not typically associated with rigidity, but rather with disturbances of affect, drive, sensorium, arousal, self-monitoring, goal-directed activity, and other neurocognitive and psychomotor functions. Furthermore, these disturbances appear to involve neural systems that are largely non-dopaminergic.[88-102] Given these facts, it is tempting to hypothesize that speech and voice abnormalities in PD may be related, at least partially, to the aforementioned disturbances and to non-dopaminergic mechanisms. Alternatively, given that some movement disorders in PD, such as abnormal posture, balance, and freezing, do not respond to common dopamine medication, and they appear to be mediated via special types of dopaminergic mechanisms, [103-104] it may be that speech and voice abnormalities in PD are mediated through these types of special dopaminergic mechanisms. If our hypothesis proves correct, this would mean that the treatment of speech and voice abnormalities in PD may require medication and behavioral and physiological approaches that are geared toward the unique neurochemical and neurobehavioral mechanisms hypothesized to underlie these speech and voice abnormalities.

Indeed, there is evidence that certain behavioral and physical therapy techniques for PD can produce significant, long-term improvement in speech and voice and in certain movement disorders such those involving posture, balance, locomotion, and gait. Many of these techniques have been developed on the premise that movement disorders in PD are related to sensory, neurocognitive, and psychomotor deficits.[105-112] These techniques emphasize intensive motor relearning, maximizing motor output and effort, increasing drive and goal-directed activity, and enhancing sensory awareness to promote internal cueing, self monitoring, and upscaling of motor output.[108,113-129]

In the present study, the number of speech and voice abnormalities correlated significantly, though weakly, with MADRS scores. There are a number of possibilities for why speech and voice abnormalities were related to depression. One possibility is that depression was a psychological reaction to the speech and voice abnormalities and/or motor impairment. However, given that the women in this study were more likely to be depressed than men, and given that the prevalence and number of speech and voice abnormalities and UPDRS scores did not differ significantly across gender, the explanation that depression was a psychological reaction to the speech or motor abnormalities is doubtful. The possibility that depression, motor impairment, and speech and voice abnormalities were epiphenomena, all related to different neuropathologic mechanisms of PD, is an alternative explanation. Still another explanation is that depression exacerbated motor impairment and/or speech and voice abnormalities. However, the low correlation between MADRS and UPDRS and the low correlation between MADRS and speech and voice abnormalities is not supportive of this explanation. Finally, it is possible that both dopamine and antidepressant medication reduced the symptoms of depression to the level where they were imperceptible to the observer, which in turn may have lowered the correlations between MADRS and speech and voice and motor abnormalities.

In the present study we found no significant relationship between speech and voice abnormalities and age or gender. These finding are surprising, given the evidence suggesting that these parameters should influence speech and voice. The lack of significant correlations may be related to the relatively small sample size, or to the way in which speech and voice abnormalities were assessed here. For example, it is possible that age and gender each had a significant effect on the severity of speech and voice abnormalities, but since we did not assess severity, we could not detect this effect. Also, it is possible that, like depression, age and gender produced subtle speech and voice abnormalities that were not detectable by perceptual analysis.

The speech and voice abnormalities in this study showed low or moderate correlations with UPDRS, duration of PD and MADRS. UPDRS, duration, and MADRS also showed low or moderate inter-correlations. The lack of high correlations suggests that these variables were not strongly related to each other. Alternatively, the low correlations may reflect the heterogeneity of symptoms and etiologies commonly observed in individuals with PD. [130]

To conclude, the present findings are preliminary, and should be interpreted as such. Improved methodology, such as using a larger sample, random subject selection, and sophisticated assessment of speech and voice functions (e.g., acoustic, physiologic, and narrow phonetic analyses) may provide stronger and more refined information on the relationships between speech abnormalities and other aspects of PD. Nevertheless, the majority of the findings in this study are consistent with other observations in clinical and research publications. Moreover, they point to new directions in the study of speech and voice abnormalities in PD. Understanding the unique neural mechanisms that might underlie these abnormalities is likely to improve their medical and behavioral treatment.

Acknowledgment

This work was supported in part by: NIH grants #RO1DC01150, OE-NIDRR #H133G40108 and the Hearst Foundation. We thank Dr. Christopher Dromey, Susan Hensley, Kathe Perez, and Christina Taskoff, as well as the individuals with IPD who participated in this study, for their contribution to the study

References

1. Lieberman A, Gopinathan G, Neophytides A, Goldstein M. Parkinson's disease handbook. New York: American Parkinson Disease Association, 1992.

2. Atarashi J, Uchida E. . A clinical study of Parkinsonism. Recent Advances in Research in the Nervous System 1959; 3:871-882.

3. Logemann J, Fisher H, Boshes B, Blonsky E. Frequency and concurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson individuals. J Speech Hear Disord 1978; 42:47-57.

4. Martilla R, Rinne V.K. Changing epidemiology of Parkinson's disease: Predicted effects of levadopa treatment. Acata Neurol Scand 1979; 59:80-87.

5. Selby G. Parkinson's disease. In: Vinken P, Bruyn G eds. Handbook of clinical neurology, Vol 6. Amsterdam: North Holland Publishing Company, 1968;173-211.

6. Habermann B. Day-to-day demands of Parkinson's disease. West J. Nursing Research 1996;18:397-413.

7. Marr J. The experience of living with Parkinson's disease. J Neurosci Nursing 1991 23:325-329.

8. Mutch R, Strucwick A, Roy S, Downie, A. Parkinson's disease: Disability, review, and management. BMJ 1986;293:675-677.

9. Oxtoby M. Parkinson's disease individuals and their social needs. London: Parkinson's Disease Society, 1982.

10. Jellinger K. The pathology of parkinsonism. In: Marsden C, Fahn S eds. Movement disorders 2. London: Butterworths, 1990; 124-165.

11. Koller W, Paulson G eds. Therapy of Parkinson's disease. New York: Marcel Dekker, 1995.

12. Hartelius L, Svensson P. Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey. Folia Phoniatr Logop 1994; 46(1):9-17.

13. Darley F, Aronson A, Brown J. Clusters of deviant speech dimensions in the dysarthrias. J Speech Hear Res 1969; 12:462-496.

14. Darley F., Aronson E, Brown J. Differential diagnostic patterns of dysarthria. J Speech Hear Res 1969; 12:246-269.

15. Darley F, Aronson A, Brown J. Motor speech disorders. Philadelphia: WB Saunders, 1975.

16. Yahr M, Duvoisin R, Hoehn M, Schear M, Barrett R. L-Dopa (L-3,4 dihydroxyphenylanine)—its clinical effects in parkinsonism. Trans American Neurologic Association 1968; 93:56-63

17. Gath I, Yair E. Analysis of vocal tract parameters in parkinsonian speech. J Acoust Soc Am 1988; 84(5):1628-34.

18. Hanson D, Gerratt B, Ward P. (1984). Cinegraphic observations of laryngeal function in Parkinson's disease. Laryngoscope 1988; 94(3):348-53.

19. Hartelius L, Wising C, Nord L. Speech modification in dysarthria associated with multiple sclerosis: An intervention based on vocal efficiency, contrastive stress, and verbal repair strategies. J Med Speech-Lang Path 1997; 5:113-140.

20. Hunker C, Abbs J, Barlow S. The relationship between parkinsonian rigidity and hypokinesia in the orofacial system: a quantitative analysis. Neurology 1982; 32(7):749-754.

21. Ludlow C, Bassich C. Relationships between perceptual ratings and acoustic measures of hypokinetic speech. In: McNeil M, Rosebek J, Aronson A eds. The dysarthrias: Physiology, acoustics, perception, management. San Diego: College-Hill Press, 1984;163-196.

22. Rosenfield D. Pharmacologic approaches to speech motor disorders. In Vogel D, Cannito M eds. Treating disordered speech motor control. Austin: Pro-ed, 1991; 111-152.

23. Larson K, Ramig L, Scherer R. Acoustic and glottographic voice analysis during drug related fluctuation in Parkinson disease. J Med Speech-Language Path 1994; 2:227-239.

24. Leanderson R, Meyerson B, Persson A. Lip muscle function in parkinsonian dysarthria. Acta Otolaryngol 1972; 74:354-357.

25. Baker K, Ramig, L, Johnson A, Freed C. Preliminary speech and voice analysis following fetal dopamine transplants in 5 individuals with Parkinson disease. J Speech Lang Hear Res 1997; 40(3):615-626.

26. Stewart C, Winfield L, Hunt A, Bressman S, Fahn S, Blitzer A, Brin M. Speech dysfunction in early Parkinson's disease. Movement Disorders 1995; 10:562-5.

27. Drasher D, Findley L. Dopaminergic induced changes in cognitive and motor processing in Parkinson's disease: an electrophysiological investigation. J Neurol Neuros Psychiat 1991; 54:603-9.

28. Lyons K, Hubble J, Troster A, Pahwa R, Koller W. Gender differences in Parkinson's disease. Clin Neuropharm 1998; 21:118-121.

29. Miyawaki E, Lyons K, Pahwa R, Troster A, Hubble J, Smith D, Busenbark K, McGuire D, Michalek D, Koller W. Parkinson's disease. Clin Neuropharm 1997; 20:523-530.

30. Sapir S, Aronson A. The relationship between psychopathology and speech and language disorders in neurologic patients. J Speech Hear Disord 1990; 55(3): 503-9.

31. Zesiewicz T, Gold M, Chari G, Hauser R Current Issues in Depression in Parkinson's Disease. Am J Geriatric Psychiat 1999; 7:110-8.

32. Winokur G, Black D, Nasrallah A. Depression secondary to other psychiatric disorders and medical illness. Am J Psychiat 1988; 23:1-12.

33. Fetoni V, Soliveri P, Monza D, Testa D, Girotti F. Affective symptoms in multiple system atrophy and Parkinson's disease: response to levodopa therapy. J Neurol Neuros Psychiat 1999; 66:541-4.

34. Konig P. Psychopathological alterations in cases of symmetrical basal ganglia sclerosis. Bio Psych 1989;25:459-68.

35. Swerdlow N, Koob G. Dopamine, schizophrenia, mania, and depression: Toward a unified hypothesis of cortico-striato-pallido-thalamic function. Behav Brain Sci 1987; 10:197-245.

36. Watson S, Khachaturian H, Lewis M, Akil H. Chemical neuroanatomy as a basis for biological psychiatry. In Berger P, Brodie H eds. American handbook of psychiatry (vol 8). New York: Basic Books, 1986.

37. Darby J, Simmons N, Berger P. Speech and voice parameters of depression: A pilot study. J Commun Disord 1984;17:75-85.

38. Hoffman G, Gonze J, Mendlewicz J. Speech pause time as a method for the evaluation of psychomotor retardation in depressive illness. Brit J Psychiat 1985;146:535-8.

39. Nilsonne A. Speech characteristics as indicators of depressive illness. Acta Psychiat Scand 1988;77:253-63.

40. Speedie L, O'Donnell W, Rabins P, Pearlson G, Poggi M, Gonzalez Rothi L. Language performance deficits in elderly depressed patients. Aphasiology 1990;4:197-205.

41. Murry T, Cannito M, Woodson G. Spasmodic dysphonia. Emotional status and botulinum toxin treatment. Arch Otol Head Neck Surg 1994;120:310-6

42. Carrie S, Livesey J, Carding P, Birchall J, Welch A. A comparison of combined glottography in men and women. Clin Otolaryngol 1993;18:505-507.

43. Holmes L, Leeper H, Nicholson I. Laryngeal airway resistance of older men and women as a functional of vocal sound pressure level. J Speech Hear Res 1994; 37:789-799.

44. Story B, Titze I, Hoffman E. Vocal tract area functions for an adult female speaker based on volumetric imaging, J Acoust Soc of Am; 104:471-87.

45. Titze I. . Physiologic and acoustic differences between male and female voices. J Acoust Soc America 1989; 85:1699-707.

46. Saunders-Pullman R, Gordon-Elliott J, Parides M, Fahn S, Saunders H, Bressman, S. The effect of estrogen replacement on early Parkinson's disease. Neurology 1999; 52:1417-21.

47. Strijks E, Kremer J, Horstink M. Effects of female sex steroids on Parkinson's disease in postmenopausal women. Clin Neuropharmacol 1999; 22(2):93-7.

48. Boulet M., Oddens B. Female voice changes around and after the menopause—an initial investigation. Maturitas 1996; 23(1):15-21.

49. Lindholm P, Vilkman E, Raudaskoski T, Suvanto-Luukkonen E, Kauppila A. The effect of postmenopause and postmenopausal HRT on measured voice values and vocal symptoms. Maturitas 1997; 28(1):47-53.

50. Pattie M, Murdoch B, Theodoros D, Forbes K. Voice changes in women treated for endometriosis and related conditions: the need for comprehensive vocal assessment. J Voice 1998; 12(3):366-71

51. Hertrich I, Ackermann H. Gender-specific vocal dysfunction in Parkinson's disease: electroglottographic and acoustic analyses. Ann Otol Rhinol Laryngol 1995; 104:197-202.

52. Hoit J. Hixon T. Age and speech breathing. J Speech Hear Res 1987; 30:351-366.

53. Baker K, Ramig L, Luschei E, Smith M. Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging. Neurology 1998; 51:1592-1598.

54. Muller P. The aging voice. Semin Speech Lang 1997; 18:159-68.

55. Sato K. Hirano M. Age-realted changes in the macula flava of the human vocal fold. Ann Otol Rhino Laryngol 1995; 104:839-44.

56. Sato K, Hirano M. Age-related changes of elastic fibers in the superficial layer of the lamina propria of the vocal folds. Ann Otol Rhinol Laryngol 1997;196:44-8.

57. Solomon N, Hixon T. Speech breathing in Parkinson's disease. J Speech Hear Res 1993; 36:294-10.

58. Fahn S, Elton R, members of the UPDRS Development Committee. Unified Parkinson's Disease Rating Scale. In: Fahn S, Marsden C, Goldstein M, Caine D eds. Recent developments in Parkinson's disease, New York: Macmillan Press, 1987;2, 153-63.

59. Stebbins G, Goetz, C. Factor structure of the Unified Parkinson's Disease Rating Scale: Motor examination section. Move Disord 1998;13:633-6.

60. Hoehn M, Yahr M. Parkinsonism: Onset, progression and mortality. Neurology 1967;19:427-442.

61. Montgomery S, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiat 1979; 134:382-389.

62. Fairbanks G. Voice and articulation drill book. New York: Harper, 1960.

63. Rostolland D. Acoustic features and shouted voice. Acoustica 1982; 50:118-125.

64. Ramig L, Countryman S, Thompson L, Horii Y. Comparison of two forms of intensive speech treatment for Parkinson disease. J Speech Hear Res 1995; 38:1232-1251.

65. Aronson A. Clinical voice disorders. New York: Thieme, 1990.

66. Behbehani M. Functional characteristics of the midbrain periaqueductal gray. Prog Neurobiol 1995;46:575-605.

67. Devinsky O, Morrell M, Vogt B. Contribution of anterior cingulate cortex to behaviour. Brain 1995;118:279-306.

68. Larson C. Brain mechanisms involved in the control of vocalization. J Voice 1988; 4:301-311.

69. Brown J. On the neural organization of language: thalamic and cortical relationships. Brain Lang 1975; 2:18-30

70. Dingwall W. The evolution of human communication systems. In Whitaker H, Whitaker H eds. Studies in Linguistics; vol 4, New York: Academic Press, 1979; 1-81.

71. Lamandella J. The limbic system in human communication. In: Whitaker H, Whitaker H eds)\. Studies in Neurolingusitics, vol 3, New York: Academic Press, 1977; 157-222.

72. Graff-Radford N, Damasio H, Yamada T, Eslinger P, Damasio A. Nonhaemorrahagic thalamic infarction: clinical, nuropsychological and electrophysiological findings in four anatomical groups defined by computerized tomography. Brain 1985; 108:485-516.

73. Meissner I, Sapir S, Kokmen E, Stein S. The paramedian diencephalic syndrome: a dynamic phenomenon. Stroke 1987; 18(2):380-385.

74. Schrag A, Schelosky L, Scholz U, Poewe W. Reduction of Parkinsonian signs in patients with Parkinson's disease by dopaminergic versus anticholinergic single-dose challenges. Mov Disord 1999; 14:252-5

75. Daniels N, Oates J, Phyland D, Feiglin A, Hughes A. Vocal characteristics and response to levodopa in Parkinson's disease. Move Disord 1996; 11 (suppl. 1):117.

76. Kupsch A, Earl C. Neurosurgical interventions in the treatment of idiopathic Parkinson disease: neurostimulation and neural implantation. J Mol Med 1999; 77(1):178-84.

77. Nasser J, Confort C, Ferraz A, Bouza A. Preliminary results in surgery of Parkinson's disease. Arq Neuropsiquiatr 1998 ;56 (3B):533-539.

78. Volkmann J, Sturm V, Weiss P, Kappler J, Voges J, Koulousakis A, Lehrke R, Hefter H, Freund, H.J. Bilateral high-frequency stimulation of the internal globus pallidus in advanced Parkinson's disease. Ann Neurol 1998; 44(6):953-961.

79. Chesson A. Aphasia following a right thalamic hemorrhage. Brain and Language 1983;19:306-16.

80. Cummings J, Benson D, Houlihan J, Gosenfeld L. Mutism: Loss of neocortical and limbic vocalization. J of Nervous and Mental Disease 1983;171:255-259.

81. Fisher C. Abulia minor versus agitated behavior. Clin Neuruos 1983; 31:9-31.

82. Gorelick P, Hier D, Benevento L, Levitt S, Tan, W. Aphasia after left thalamic infarction. Arch Neurol 1984;41:1296-8.

83. Jurgens U, von Cramon D. On the role of the anterior cingulate cortex in phonation: A case report. Brain and Language 1982; 15:234-248.

84. Lazzarino L, Nicolai A Aphonia as the only speech disturbance from bilateral paramedian thalamic infarction. Clin Neurol Neuros 1988; 90:265-7

85. Ozeren A, Sarica Y, Efe R. Thalamic aphasia syndrome. Acta Neurol Belg 1994 ;94(3):205-8.

86. Petrovici J. Speech disturbances following stereotaxic surgery in ventrolateral thalamus. Neurosurg Rev 1980; 3(3):189-95

87. Sapir S, Aronson, A. Aphonia after closed head injury: aetiologic considerations. Br J Disord Commun 1985; 20(3):289-96.

88. Brown P, Marsden C. What do the basal ganglia do? Lancet 1998; 351(9118):1801-1804.

89. Brown R, Marsden C, Quinn N, Wyke M. Alterations in cognitive performance and affect-arousal state during fluctuations in motor function in Parkinson's disease. J Neurol Neurosurg Psychiatry 1984; 47(5):454-465.

90. Cooper J, Sagar H, Jordan N, Harvey N, Sullivan E. Cognitive impairment in early, untreated Parkinson's disease and its relationship to motor disability. Brain 1991; 114 (Pt5): 2095-2122.

91. Dubois B, Pillon B. Cognitive deficits in Parkinson's disease. J Neurol 1997; 244(1):2-8.

92. Farina E, Cappa S, Polimeni M, Magni E, Canesi M, Zecchinelli A, Scarlato G, Mariani C. Frontal dysfunction in early Parkinson's disease. Acta Neurol Scand 1994; 90(1):34-38.

93. Karayanidis F. Parkinson's disease: a conceptualization of neuropsychological deficits within an information-processing framework. Biol Psychol 1989; 29(2):149-179.

94. Marsden C. Function of the basal ganglia as revealed by cognitive and motor disorders in Parkinson's disease. Can J Neurol Sci 1984; 11(1 Suppl):129-35.

95. Piccirilli M, D'Alessandro P, Finali G, Piccinin G, Agostini L. Frontal lobe dysfunction in Parkinson's disease: prognostic value for dementia? Eur Neurol 1989; 29(2):71-6.

96. Pillon B, Dubois B, Cusimano G, Bonnet A, Lhermitte F, Agid Y. Does cognitive impairment in Parkinson's disease result from non-dopaminergic lesions? J Neurol Neurosurg Psychiat 1989; 52(2):201-6.

97. Schneider J, Diamond S, Markham C. Parkinson's disease: sensory and motor problems in arms and hands. Neurology 1987; 37(6):951-6.

98. Smith Y, Bevan M, Shin, E, Bola, J. Microcircuitry of the direct and indirect pathways of the basal ganglia. Neuroscience 1998 86(2):353-87.

99. Stuss D, Guberman A, Nelson R, Larochelle S The neuropsychology of paramedian thalamic infarction. Brain and Cognition 1988; 8:348-78

100. Taylor A, Saint-Cyr J, Lang , A. Frontal lobe dysfunction in Parkinson's disease. The cortical focus of neostriatal outflow. Brain 1986; 109:845-83.

101. Van Spaendonck K, Berger H, Horstink M, Buytenhuijs E, Cools A. Executive functions and disease characteristics in Parkinson's disease. Neuropsychologia 1996; 34(7):617-26.

102. Weingartner H, Burns S, Diebel R, LeWitt P. Cognitive impairments in Parkinson's disease: distinguishing between effort-demanding and automatic cognitive processes. Psychiatry Res 1984; 11(3):223-35.

103. Bloem B, Beckley D, van Dijk J, Zwinderman A, Remler M, Roos R. Influence of dopaminergic medication on automatic postural responses and balance impairment in Parkinson's disease. Mov Disord 1996); 11(5):509-521.

104. Rogers M. Disorders of posture, balance, and gait in Parkinson's disease. Clin Geriatr Med 1996; 12(4):825-45.

105. Bohannon R. Physical rehabilitation in neurologic diseases. Curr Opin Neurol 1993; 6(5):765-772.

106. Elias J, Treland J. Executive Function in Parkinson's Disease and Subcortical Disorders. Semin Clin Neuropsychiatry 1999; 4(1):34-40.

107. Georgiou N, Iansek R, Bradshaw J, Phillips J, Mattingley J, Bradshaw J. An evaluation of the role of internal cues in the pathogenesis of parkinsonian hypokinesia. Brain 1993; 116 (Pt 6):1575-1587.

108. Homberg V. Motor training in the therapy of Parkinson's disease. Neurology 1993; 43(12 Suppl 6):S45-S46.

109. Morris M, Iansek R, Matyas,T, Summers J. The pathogenesis of gait hypokinesia in Parkinson's disease. Brain 1994; 117:1168-1181.

110. Morris M, Iansek R, Matyas T, Summers J. Ability to modulate walking cadence remains intact in Parkinson's disease. J Neurol Neurosurg Psychiat 1994; 57(12):1532-1534.

111. Muller F, Stelmach G. Scaling problems in Parkinson's disease. In: Requin J, Stelmach GE, eds. Tutorials in motor neuroscience. Netherlands: Kluwer Academic Publishers 1991: 161-174.

112. Richards M, Cote L, Stern Y. The relationship between visuospatial ability and perceptual motor function in Parkinson's disease. J Neuro Neuros Psychiat 1993; 56:400-6.

113. Dam M, Tonin P, Casson S, Bracco F, Piron L, Pizzolato G, Battistin L. Effects of conventional and sensory-enhanced physiotherapy on disability of Parkinson's disease patients. Adv Neurol 1996; 69:551-555.

114. de Angelis E, Mourao L, Ferraz H, Behlau M, Pontes P, Andrade L. Effect of voice rehabilitation on oral communication of Parkinson's disease patients. Acta Neurol Scand 1997; 96(4):199-205.

115. Gauthier L, Dalziel S, Gauthier S. The benefits of group occupational therapy for patients with Parkinson's disease. Am J Occup Ther 1987; 41(6):360-5.

116. Katsikitis M, Pilowsky I. A controlled study of facial mobility treatment in Parkinson's disease. J Psychosom Res 1996; 40(4):387-96.

117. Le Dorze G, Dionne L, Ryalls J, Julien M, Ouellet L. The effects of speech and language therapy for a case of dysarthria associated with Parkinson's disease. Eur J Disord Commun 1992; 27(4):313-24

118. McIntosh G, Brown S, Rice R, Thaut M. Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. J Neuro Neuros Psychiat 1997; 62(1):22-26.

119. Muller V, Mohr B, Rosin R, Pulvermuller F, Muller F, Birbaumer N. Short-term effects of behavioral treatment on movement initiation and postural control in Parkinson's disease: a controlled clinical study. Mov Disord 1997; 12(3):306-314.

120. Pacchetti C, Aglieri R, Mancini F, Martignoni E, Nappi G. Active music therapy and Parkinson's disease: methods. Funct Neurol 1998; 13(1):57-67.

121. Palmer S, Mortimer J, Webster D, Bistevins R, Dickinson G. Exercise therapy for Parkinson's disease. Arch Phys Med Rehabil 1986; 67(10):741-5.

122. Ramig L, Bonitati C, Lemke J, Horii Y. Voice treatment for individuals with Parkinson disease: Development of an approach and preliminary efficacy data. J Med Speech-Lang Path 1994; 2:191-209.

123. Ramig L, Countryman S, O'Brien C, Hoehn M, Thompson L. Intensive speech treatment for individuals with Parkinson's disease: Short- and long-term comparison of two techniques. American Academy of Neurology 1996; 47:1496-1504.

124. Schenkman M, Donovan J, Tsubota J, Kluss M, Stebbins P, Butler R. Management of individuals with Parkinson's disease: rationale and case studies. Phys Ther 1989; 69(11):944-55.

125. Scott S, Caird F. The response of the apparent receptive speech disorder of Parkinson's disease to speech therapy. J Neuro Neuros Psychiat 1984; 47(3):302-4.

126. Scott S, Caird F. Speech therapy for Parkinson's disease. J Neuro Neuros Psychiat 1993; 46(2):140-4.

127. Soliveri P, Brown R, Jahanshahi M, Marsden C. Effect of practice on performance of a skilled motor task in patients with Parkinson's disease. J Neuro Neuros Psychiat 1992; 55(6):454-60.

128. Sunvisson H, Lokk J, Ericson K, Winblad B, Ekman S. Changes in motor performance in persons with Parkinson's disease after exercise in a mountain area. J Neurosci Nurs 1997; 29(4):255-60.

129. Viliani T, Pasquetti P, Magnolfi S, Lunardelli M, Giorgi C, Serra P, Taiti P. Effects of physical training on straightening-up processes in patients with Parkinson's disease. Disabil Rehabil 1999; 21(2):68-73.

130. Graham J, Sagar H. A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: identification of three distinct subtypes. Mov Disord 1999; 14:10-20.

NCVS Status and Progress Report - 14
September 1999, 163-171

Influence of Order of Stimulus Presentation on Speech Motor Learning: A Principled Approach to Treatment for Apraxia of Speech

Taresa Knock, B.A.
Kirrie Ballard, Ph.D.
Department of Speech Pathology & Audiology, The University of Iowa
Donald Robin, Ph.D.
Department of Speech Pathology & Audiology, The University of Iowa
Department of Communicative Disorders, San Diego StateUniversity
Richard Schmidt, Ph.D.
Failure Inc., Los Angeles CA

Abstract

The present study was designed to examine the efficacy of applying a principled approach to the treatment of acquired apraxia of speech (AOS). Specifically, we sought to determine if application of principles of motor learning would improve speech production skills in two subjects with severe AOS. In particular, we examined one main principle, random practice and compared it to blocked practice. Of importance is the fact that most speech treatments utilise blocked practice, but the literature on motor learning clearly shows that blocked practice facilitates acquisition of target behaviours, but not retention and transfer, the later two considered true indices of learning. Results showed that random practice facilitated retention and transfer whereas blocked practice did not. The present study provides preliminary evidence that these principles may have a similar effect on learning of skilled speech motor acts as they do on learning of limb movements. This study is the first in a program of research planned in our laboratory to investigate principles of motor learning in the motor speech system. The results reported here are encouraging and provide justification and focus for further investigation.

Introduction

Apraxia of speech (AOS) has been defined as a "phonetic-motoric disorder of speech production caused by inefficiencies in the translation of a well-formed and filled phonologic frame to previously learned kinematic parameters assembled for carrying out the intended movement" (McNeil *et al.* 1997, p. 329). Numerous treatment approaches have been developed to remediate the motor speech disorder of acquired AOS (e.g. Wambaugh and Doyle 1994). These typically have presumed to alter the organisation of the neuromotor system but many have not made direct reference to a theoretical basis for the disorder. No treatments have been motivated by theories of how the motor system learns skilled actions, although in a general way some claim to enhance motor learning. While these approaches have claimed to be effective in improving speech production skills, many have not been tested with well-controlled experimental designs and the relative efficacy of different treatment approaches has not be compared. Furthermore, all have focused on acquisition of speech behaviours (i.e. performance during treatment) rather than retention and transfer to novel stimuli and responses (Wambaugh and Doyle 1994). There is now a wealth of data from studies of motor learning in limb systems, largely from the perspective of schema theory of motor control and learning (Schmidt 1975, Schmidt and Lee 1999), indicating that retention and transfer are better indicators of long-term learning than acquisition performance (e.g. Chamberlin and Lee 1993, Schmidt and Bjork 1992). This research has identified a set of principles, which enhance long-term learning of motor acts. The present study represents a preliminary investigation of the influence of principles of motor learning on the

relearning of speech skills in individuals with severe AOS. By "learning" we are specifically referring to *retention* of skills post-treatment. The present study represents an approach to treatment that is grounded firmly in a theory of how the motor system learns skilled actions.

Robin and colleagues (e.g. Robin 1992, 1999ab) have advocated the use of principles of motor learning in patients with motor speech disorders. Recent textbooks on motor speech disorders have also made strong claims about using these principles as a framework for treatment (Duffy, 1995, McNeil *et al.* 1997, Yorkston *et al.* 1998). This interest stems from the large body of research that has examined learning of novel limb movements in normal individuals. These studies have aimed to test the tenets of the schema theory of motor control and learning (Schmidt 1975, Schmidt and Lee 1999).

Schema theory presumes that learning results from developing and refining a representation of an action. This representation prescribes the relation between how we move and the demands of a given task. The two main concepts underlying schema theory are generalised motor programmes (GMP) and parameters. GMP's contain an abstract code about relative timing of events and the relative force with which events are to be produced. Parameters specify details about how the GMP will be expressed including absolute duration of movement, absolute force of muscle contractions, and the muscles or limb used to make the movement. These two concepts combine to reduce the demands for storage of programmes for actions and explain the relatively invariant features of actions. While the effects of principles of motor learning may be explained equally well using alternative theoretical approaches such as dynamic systems (e.g. Kelso 1995, Thelen & Smith 1994), we focus on schema theory in the present study to place our results in the context of this body of literature.

The concepts of schema theory can be used to explain the motor speech disorder of AOS. However, the issue of defining what constitutes a motor programme in speech production is far from resolved (e.g. Smith *et al.* 1995). If it is assumed that the process of motor programming includes GMPs and parameters that set specific aspects (e.g., absolute speed) of the program, then AOS may be thought of as a breakdown in the ability to activate and / or select a GMP, an inability to correctly set the parameters specific to a situation, or both (Clark & Robin, 1998).

Research on motor learning based on schema theory has led to the specification of several principles pertaining to the structure of practice and response feedback that promote retention and transfer of treated behaviours (e.g. Schmidt 1975, Schmidt and Bjork 1992, Schmidt and Lee 1999). It is beyond the scope of the present study to review all of these principles (see Schmidt and Lee 1999 for an extensive review). Here, we focus on those that are specifically examined in our study. It has been demonstrated that when several novel limb actions representing different GMP's are practised in a blocked order acquisition is more rapid than when actions are practised in random order. However, the random condition results in greater retention of skilled actions post-treatment and greater transfer of treatment effects to novel actions (e.g. Del Rey *et al.* 1994, Lee and Magill 1983, Shea and Morgan 1979, Shea and Wright 1991, Wulf and Schmidt 1994). Furthermore, retention and transfer are enhanced when feedback is provided with low frequency (i.e. on 30-60% of trials) (Lee *et al.* 1990, Salmoni *et al.* 1984, Schmidt 1991, Schmidt and Bjork 1992, Vander Linden *et al.* 1993, Weeks *et al.* 1993). Feedback can be provided in two forms — knowledge of results (KR) and knowledge of performance (KP). KR indicates the accuracy of a response while KP provides specific information regarding qualitative features of a response (e.g., bio-feedback). Providing only KR-feedback, particularly in the later stages of training, promotes retention of skills (see Schmidt and Lee 1999). Finally, feedback has a greater effect on learning when it is given about 3-4 sec after the response and when a 3-4 sec delay intervenes between the feedback and the subsequent stimulus presentation (Swinnen *et al.* 1990).

Schmidt and Lee (1999) suggested that these principles facilitate retention and transfer of treatment effects as they serve to increase the difficulty of the learning environment and more closely approximate a natural context. For example, random practice introduces difficulty as it forces retrieval and organisation of a different response on every trial (Lee and Magill 1985, Lee and Weeks 1987, Schmidt and Bjork 1992) whereas the blocked condition results in practice of movement execution only. The former more closely simulates a natural context and so should also result in greater stimulus transfer. Presenting KR-feedback only, with lower frequency and delays between responses and feedback, guides the individual's understanding of acceptable versus unacceptable responses, facilitates development of self-evaluation skills, and permits these processes to occur uninterrupted (see Schmidt and Bjork 1992).

Past researchers have attempted to use some principles of motor learning in treatments for AOS. For example, many treatment programs for AOS advocate a high number of practice trials (e.g. Phonetic Placement Therapy: Van Riper and Irwin 1958, Eight-Step Task Continuum: Rosenbek *et al.* 1973, Prompts for Restructuring Oral Muscular Phonetic Targets (PROMPT): Chumpelik 1984, Square *et al.* 1985, Minimal Pairs Treatments: e.g. Wambaugh *et al.* 1998). However, many studies have advocated conditions that are opposed to the above findings. For example, Yorkston *et al.* 1998 stated that feedback should be provided with high frequency and immediately following an

individual's response. While this may facilitate acquisition of treatment effects, it appears to be detrimental to retention of treated behaviours.

The present study utilised single subject experimental designs to examine the influence of random versus blocked order of practice on the efficacy of Phonetic Placement Therapy (Van Riper and Irwin 1958) in two individuals with severe AOS. Both treatment conditions were presented with low frequency delayed KR-feedback. Our specific hypotheses stated that, compared to blocked stimulus presentation, random stimulus presentation would result in a) slower acquisition of trained speech motor skills, b) greater retention of trained skills at one and four weeks post-treatment, and c) increased transfer of treatment effects to production of the treated behaviours with a novel stimulus and to related novel responses. Considering transfer to novel responses, we predicted that transfer would occur to speech targets that represented similar GMP's. While it is unclear what constitutes a motor programme in speech production, we hypothesised that, if GMP's correspond roughly to a phoneme, then transfer may occur to speech targets that utilise the same programmes but in a different sequence. This form of transfer may not be forthcoming in AOS, however, as these individuals are known to have particular difficulty in sequencing speech sounds. In addition, transfer was not expected to occur to targets that included untrained GMP's. This study is the first in a program of research planned in our laboratory to investigate principles of motor learning in the motor speech system.

Method

Subjects

Two Caucasian males participated in this study. Both were diagnosed with severe AOS with aphasia secondary to a single left-hemisphere cerebrovascular accident (CVA) within the distribution of the left middle cerebral artery. A CT scan performed on Subject 1 six months post-onset indicated the lesion was restricted to the left temporo-parietal area. Participant 1 was 38 years of age and 36 months post-onset at the beginning of treatment. For Subject 2, a CT scan two years post-onset revealed an area of reduced attenuation in the left fronto-temporo-parietal area. A right-sided hemiparesis was present. Participant 2 was 65 years of age and 96 months post-onset at the beginning of treatment. Both participants were right-handed native English speakers with no history of speech, language, or reading disorders or progressive neurologic disease. Subject 2 had a history of depression and alcohol abuse. Both participants passed a pure tone audiometric screening at 40 dBSPL at 500, 1000, and 2000 Hz bilaterally and had 20/40 vision (corrected or uncorrected).

Speech, Language, and Neuropsychological Testing

Formal testing was conducted within two weeks prior to the experiment (see Table 1). Classification of aphasia type was difficult due to the participants' speech production impairments and was, in part, guided by known lesion information. Subject 1 was tentatively classified with conduction aphasia based on results of the Western Aphasia Battery (WAB; Kertesz 1982) and presence of a posterior lesion. We were unable to classify Subject 2's aphasia type with confidence. In both cases, auditory comprehension was impaired but superior to expressive abilities with scores on the WAB comprehension subtests ruling out a Wernicke's aphasia (see Table 1). For both participants, the subtests of the WAB measuring nonlinguistic cognitive skills revealed mildly to moderately impaired performance (see Table 1). However, these skills were superior to speech and language skills.

Table 1.
Results of Formal Speech, Language, and Neuropsychological Testing

Test	Subject 1	Subject 2
Western Aphasia Battery[1]		
Fluency	0	1
Comprehension	7.2	7.4
Repetition	0	0.7
Naming	0	1.1
Aphasia Quotient	14.4	26.4
Reading	5.4	2.6
Writing	4.5	2.3
Praxis	9.2	9.0
Construction	7.8	5.6
Ravens Colored Progressive Matrices (/37)	27	22
Cognitive Quotient	41.3	40.1
Northwestern University Sentence Comprehension test for Aphasia[2]		
Canonical sentences (/20)	60%	75%
Noncanonical sentences (/20)	75%	55%
Assessment of Agility and Control of the Oral Mechanism or Speech Production[3]		
Diadochokinetic rate[4]		
/p/	0.7 / sec	3.0 / sec
/t/	2.3 / sec	3.6 / sec
/k/	unable to elicit	unable to elicit
/ptk/	unable to elicit	unable to elicit
Shape cancellation test		
Percent correct	100%	100%
Time to completion	2 min 14 sec	2 min 55 sec

[1] Kertesz (1982)
[2] Thompson (n.d.)
[3] Hall *et al.* (n.d.)
[4] All results indicate performance below the normal level according to normative data from Kent *et al.* (1987)

To define sentence comprehension in more detail, the Northwestern University Sentence Comprehension Test for Aphasia (Thompson n.d.) was administered. This is a sentence-to-picture matching task that tests comprehension of semantically reversible canonical sentences (i.e. simple actives as in "The thief chased the artist" and subject relatives as in "I saw the thief who chased the artist") and noncanonical sentences (i.e. passives as in "The artist was chased by the thief" and object relatives as in "I saw the artist who the thief chased"). Both participants demonstrated impairment on both sentence types (see Table 1). Subject 2 demonstrated an "agrammatic-type" comprehension deficit with above-chance comprehension of canonical sentences and at-chance comprehension of noncanonical sentences.

To evaluate verbal and nonverbal oral motor skills, the Assessment of Agility and Control of the Oral Mechanism for Speech Production (Hall *et al.* n.d.) was administered (see Table 1). Both subjects demonstrated oral apraxia. Speech production was characteristic of severe AOS (e.g. Kent and Rosenbek 1983) with the few utterances produced being notable for delayed initiation of movements, reduced rate, distortions of phonemes, difficulty sequencing movements, and articulatory groping. To confirm the diagnosis of AOS, the articulator visuomotor tracking task of Robin and colleagues (Hageman *et al.* 1994, Robin *et al.* in prep) was administered. This is a nonverbal task that assesses motor control of the speech articulators. Robin and colleagues have demonstrated that this task is sensitive to even mild AOS (Hageman *et al.* 1994, Robin *et al.* 1997, Robin *et al.* in prep) and performance is correlated with perceptual measures of articulatory precision and speech intelligibility. Both participants demonstrated the characteristic profile of AOS whereby tracking of predictable visuomotor targets is impaired while tracking of unpredictable targets is similar to normal. This result confirmed the diagnosis of AOS.

A shape cancellation test was administered to rule out visuo-spatial deficits. Both participants performed at the normal level (see Table 1).

Materials

Different stimulus sets were developed for each participant, based on their profile of deficits and stimulability for speech behaviours. These are described below.

Subject 1

For Subject 1, a set of cards with printed syllables and words was used to elicit the targeted speech behaviours during baseline testing and experimental probes. Stimuli included consonant-vowel (CV) syllables (/pa/, /ba/, /ta/, /fa/, /va/, /sa/, /pi/, /bi/, /ti/, /fi/, /vi/, and /si/), VC syllables (/ap/, /ab/, /at/, /af/, /av/, and /as/), and CVC words (initiated by /p/, /b/, /t/, /f/, /v/, and /s/). CVC words were controlled for frequency (plosives: mean = 56.50, SD = 4.38, fricatives: mean = 54.00, SD = 8.89, Carroll *et al.* 1971), length in syllables (one), and initial consonant (10 each of /p/, /b/, /t/, /f/, /s/, and /v/).

Subject 1 participated in two phases of treatment. In phase one,production of the CV syllables /pa/, /ba/, /ta/, /fa/, /va/, and /sa/ was treated. In phase 2, production of the VC syllables /ap/, /ab/, /at/, /af/, /av/, and /as/ was treated. During pre-practice (see Treatment procedure), the target syllables were represented both orthographically and with two diagrams showing position of the articulators for the consonant and for the vowel. During practice (see Treatment procedure), the stimuli were identical to those used in the probes – orthography only.

Subject 2

Subject 2 was unable to read subsequent to CVA and so black-and-white line drawings were used to elicit speech behaviours during baseline testing and experimental probes. The stimuli consisted of 12 CVC words initiated by /p/, /b/, /t/, /f/, /v/, and /s/. Six of these words (pat, bat, tab, face, vase, and safe) were targeted for treatment. The remaining six (cap, beak, tack, fish, save, and shave) were used to demonstrate experimental control. Words were controlled for length in syllables (one) and pictureability (elicited the target word from five normal controls).

Subject 2 participated in one phase of treatment. During pre-practice, the stimuli depicted black-and-white pictures with three diagrams demonstrating position of the articulators for the two consonants and the vowel. As this subject demonstrated excessive struggle behaviour and frustration when trying to recall the articulatory postures for each target word during practice, all stimulus cards for the practice were modified to include the stimulus picture and the diagrams of articulatory postures. These modified practice stimuli were also used for retention testing but the diagrams were not presented on the stimuli used during the baseline and experimental probes.

Experimental Design

A single-subject alternating treatments design (ATD) plus multiple baseline across subjects and behaviours designs were utilised (McReynolds and Kearns 1983). In an ATD, each subject undergoes all treatment conditions, serving as his / her own control. During the treatment phase/ s, all treatment conditions are applied within a single session. Traditionally, a single dependent variable is subjected to both treatment conditions to determine which treatment effects greatest change. This is not possible in studies of learning and so one attempts to select independent sets of behaviours of equivalent difficulty and each behaviour is

linked to one treatment. In this case, syllables or words with plosive consonants formed one set and fricative consonants formed a second set. Previous studies of AOS have reported no transfer from treatment for plosive production to production of fricatives and vice versa (Rubow *et al.* 1982). Therefore, we assumed that effects of each treatment could be isolated to the set of behaviours to which that treatment condition was applied. However, plosives and fricatives may not be equivalent in difficulty. To address this problem, the pairing of treatment condition and behaviour set was counterbalanced across subjects in the first phase of treatment and, for Subject 2, within subject across the two phases of treatment. In this way, a differential treatment effect would be demonstrated if learning occurred with a specific treatment condition regardless of which behaviour set it was linked with.

In a multiple baseline across subjects design, participants are exposed to progressively longer baseline phases to ensure that behaviour change is a direct result of treatment. In a multiple baseline design across behaviours, one behaviour is entered into treatment while others are held in baseline. Thus, experimental control is demonstrated when performance on a behaviour only changes with treatment application. This design permits maintenance of experimental control when transfer of treatment effects to select behaviours occurs.

Baseline Testing Procedures

Subject 1 received three baseline probes and Subject 2 received five. Baseline sessions for Subject 1 consisted of eliciting 10 each of the 12 CV syllables, six VC syllables, and six types of CVC words, for a total of 240 items per baseline. Baseline sessions for Subject 2 involved eliciting the six CVC words targeted for treatment 10 times each and the remaining words five times each, for a total of 90 items. All items were presented in random order with no modelling or feedback. Treatment commenced once baseline probes were completed and stable performance was demonstrated.

Treatment Procedure

Each treatment phase consisted of 12 sessions. Treatment sessions were 90-120 minutes long, two to three times per week. Each treatment session included two sections – one presenting the random practice condition and one the blocked condition. The order of sections was randomised within subject across sessions. For Subject 1, plosive CV syllables (/pa/, /ba/, /ta/) were trained in random order and fricative syllables (/fa/, /va/, /sa/) in blocked order in phase 1 of treatment. Plosive VC syllables (/ap/, /ab/, /at/) were trained in blocked order and fricative syllables (/af/, /av, /as/) in random order in phase 2. For Subject 2, CVC words initiated by plosives (pat, bat, tab) were trained in blocked order and fricative CVC words (face, vase, safe) in random order.

Consistent with previous studies examining the principles of motor learning, each treatment section included both a pre-practice and a practice component. The pre-practice component applied Phonetic Placement Therapy to prepare the subject for the practice stimuli. In this approach production of speech behaviours is trained using orthographic or picture stimuli, diagrams and verbal descriptions of articulatory features, and modelling. The pre-practice protocol involved (a) using the above methods to describe speech sounds to the subject, (b) modelling of the targets by the examiner five times, and (c) the subject attempting production with feedback from the examiner to shape more accurate production. Once the subject produced five consecutive attempts correctly, without a model, the practice trial began. Practice consisted of eliciting the three target behaviours for that section, either under random or blocked order stimulus presentation, a total of 50 times each (total of 150 responses per practice) without examiner models. Three different feedback schedules were developed so that the same items did not receive feedback in each session. For the blocked condition, the order in which the syllables or words were presented was randomised across sessions. The participants were offered breaks between the two sections of treatment.

Participants' responses to all stimuli were transcribed and scored online and recorded on audiotape for reliability purposes. Responses were scored as correct if all sounds in the syllable or word were perceived as accurate by the examiner.

Experimental Probe Procedure

An experimental probe identical to baseline testing was administered after every third treatment session, but not on the same day. This probe measured transfer of treatment effects to production of treated behaviours in a novel stimulus context and to novel responses. Responses were transcribed and scored on line and recorded on audiotape for reliability purposes.

Retention Testing

Retention testing consisted of two procedures, each performed at one and four weeks post-treatment. First, the procedure used for the practice components of the treatment sessions was used but with no feedback provided. Thus, retention of the treated behaviours was tested in the same order context as in training. Second, the experimental probe was administered to measure retention of transfer effects. This probe was administered to Subject 1 only. Subject 2 did not demonstrate transfer effects during treatment and so retention of such effects was not possible.

Reliability

Reliability was calculated on about 25% of all tests and practice sessions. Inter-rater point-to-point reliability for scoring of baseline and experimental probes was 97.7% (SD = 4.4) and intra-rater reliability was 96.8% (SD = 5.1). Inter-rater reliability for scoring of practice trials and retention testing was 95.2% (SD = 5.1) for plosives and 92.0% (SD = 7.3) for fricatives and intra-rater reliability was 93.7% (SD = 5.9) for plosives and 90.9% (SD = 8.5) for fricatives. Inter-rater reliability on the independent variable was 94.3% (SD = 6.5).

Results

Data representing acquisition, retention, and transfer for Subjects 1 and 2 are presented in figures 1 and 2, respectively. The figures represent (a) percent correct productions of trained CV (phase 1) and VC (phase 2) syllables for person 1 and trained CVC words for Subject 2, (b) retention of treated behaviours over time, and (c) transfer of treatment effects to a novel stimulus and to production of novel responses.

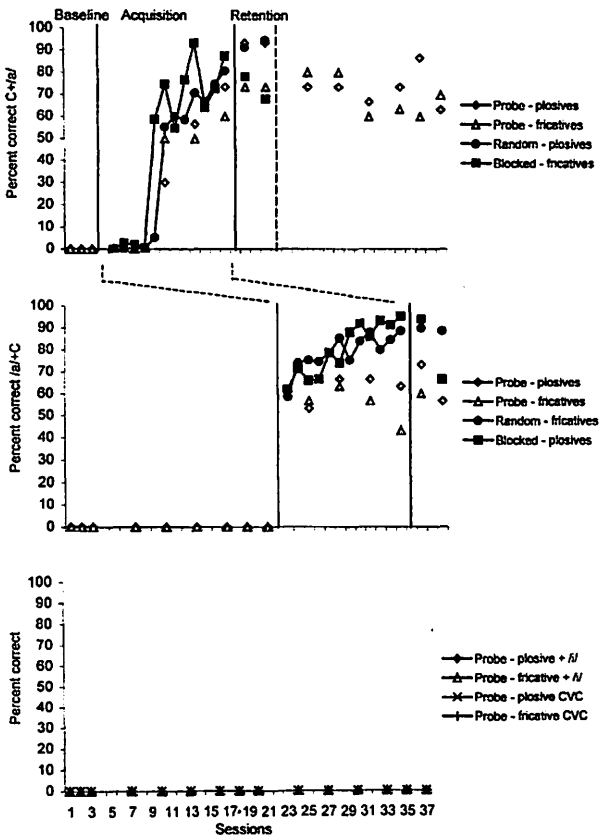


Figure 1. Performance of Subject 1 during baseline, the two phases of treatment, and retention testing. Black circles and triangles represent performance during the random and blocked practice trials and during retention testing with stimuli used during practice. Remaining symbols represent performance on other variables during baseline testing and experimental probes during treatment and retention phases. C denotes consonant and V vowel.

Both participants demonstrated stable performance across baseline testing and, therefore, were entered into treatment at the scheduled times. The level of performance for both participants was 0% correct across all baselines. Performance on the practice trials during the treatment phase represents the acquisition of trained speech behaviours. The subjects did not show more rapid acquisition of the speech behaviours presented in blocked order. Both sets of behaviours were acquired at about the same rate. At the completion of phase 1 of treatment for Subject 1, acquisition performance was at 81% correct in the random condition and 87% in the blocked condition (see figure 1). The same trend was evident in phase 2. Subject 2 demonstrated a similar trend during acquisition, with 84% correct production of randomly trained behaviours and 68% for those trained in blocked order in the final practice session (see figure 2). For both subjects, performance on behaviours presented in blocked order during the first treatment phase was considerably more unstable, or variable, across sessions than behaviours presented in random order (see figures 1 and 2).

Retention testing utilising the stimuli of the practice sessions showed greater retention of the behaviours trained in random order, particularly at four weeks post-treatment, for both subjects. Subject 1 performed at 91% and 95% accuracy on the randomly trained targets in the one and four week retention tests, respectively, while perfor-

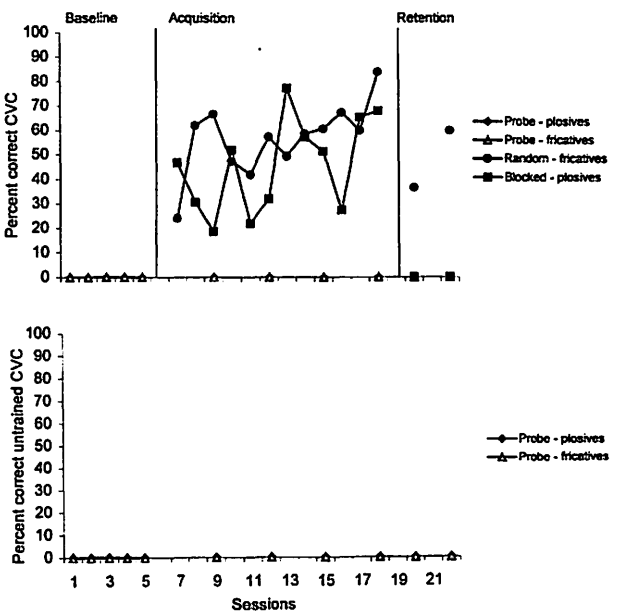


Figure 2. Performance of Subject 2 during baseline, the one phase of treatment, and retention testing. Black circles and triangles represent performance during the random and blocked practice trials and during retention testing with stimuli used during practice. Remaining symbols represent performance on other variables during baseline testing and experimental probes during treatment and retention phases. C denotes consonant and V vowel.

mance on the blocked behaviours was at 78% and 68%, respectively. Note that, in phase 1, the discrepancy between the two behaviour sets increased over time (see figure 1). In phase 2, this difference was not evident at one week post-treatment but emerged at the second retention test. Subject 2 exhibited similar results (see figure 2). The targets trained in random order were performed with 37% and 60% accuracy in the one and four week retention tests, respectively, while the behaviours trained in blocked order were not retained at all.

Experimental probes administered during the treatment phase tested transfer of trained behaviours to a novel stimulus and to novel responses. During treatment phase 1, Subject 1 showed transfer of treatment effects to the novel stimulus condition for both behaviour sets. However, retention of this transfer effect was greater for those behaviours trained in random order (see figure 1). No transfer of treatment effects to production of novel responses was evident (see figure 1), but the effects of treatment were over-generalised with trained behaviours being substituted for the novel items. Subject 2 did not demonstrate transfer to the novel stimulus condition or to novel responses. Therefore, the post-treatment experimental probes were not administered.

Discussion

This experiment was conducted to investigate the effects of random versus blocked stimulus presentation on learning motor speech skills in severe AOS. Subjects received both treatment conditions in parallel throughout treatment so that the effects of these two conditions were compared within subject. Based on studies of motor learning in normal individuals (e.g. Del Rey *et al.* 1994, Lee and Magill 1983, Shea and Morgan 1979, Shea and Wright 1991), it was predicted that random stimulus presentation would result in slower acquisition of trained behaviours and greater retention of trained behaviours at one and four weeks post-treatment. Furthermore, greater transfer of treatment effects was expected to a novel stimulus condition and to novel responses related to behaviours trained in the random condition.

In the present study, practising motor speech skills in random order did not result in slower acquisition of targeted behaviours, as previously reported. Both behaviour sets were acquired at a similar rate. However, performance on the blocked behaviours tended to be more variable across practice sessions. This finding is not consistent with studies of limb motor learning in normal individuals where blocked practice facilitates acquisition and where performance during the acquisition phase for both conditions does not demonstrate such variability (see Schmidt and Lee 1999 for a review). In our two subjects with AOS, it seemed that the variable performance across sessions for the blocked condition reflected a greater difficulty in accessing or executing the accurate GMPs in this condition compared to the random condition. This hypothesis is supported by the observation that, frequently, performance in the blocked condition was facilitated for the behaviour practised first. That is, Subject 1 demonstrated fewer instances of consonant distortions (i.e. inaccurate articulator placement or onset of voicing) and Subject 2 fewer substitutions of stereotypical utterances on the first behaviour. Repeatedly practising this first behaviour may have pushed the subjects into a stable set and inhibited access to programmes for subsequent speech targets in these individuals. In normal subjects, this "pathological" inhibition is not observed. While this hypothesis is speculative, it provides a direction for future study.

Both subjects demonstrated the predicted result of greater retention of behaviours trained in random order than blocked order, particularly at four weeks post-treatment. In the case of Subject 2, this effect was dramatic with no retention of the blocked behaviours. These findings support previous studies that have examined the influence of a random order of practice on learning of novel limb actions in normal individuals. Schmidt & Bjork (1992) argued that this effect is observed because a random practice order promotes recall and execution of actions, which in turn results in more successful recall of actions post-treatment. On the other hand, the blocked condition promotes a "stable set" in which only execution of the action is practised, such that recall is not enhanced. We contend that retention performance is a more rigorous and meaningful indicator of a treatment's efficacy than acquisition. These data demonstrate that conditions during practice influenced retention of skills in two individuals with severe AOS.

On closer examination of the data for acquisition and retention of trained behaviours, it was noted that accurately producing the voicing distinction between /p/ and /b/ and between /f/ and /v/ presented great difficulty for both subjects. This observation supports the initial diagnosis of a disorder at the phonetic-motoric level rather than the linguistic-phonological level. For both subjects, ability to use voicing to differentiate these sounds improved as treatment progressed. However, for Subject 1 in particular, ability to produce this voicing distinction was retained for the randomly trained behaviours and progressively lost for the behaviours trained in blocked order. This trend appeared evident for Subject 2 also, but was less clear given that he did not retain any behaviours trained in blocked order.

Results for transfer of treatment effects were not consistent across subjects. Subject 1 demonstrated transfer of production of trained behaviours to the novel stimulus condition of the experimental probe and, in phase 1 of treatment, retention of this transfer effect was greater for behaviours in the random condition. This result supports previous research and again supports the use of this principle. However, no transfer was observed to novel responses. While this may appear a negative result, it is possible that

the items selected to examine transfer were not appropriate. As noted above, there is no consensus on what is represented in a GMP for speech targets. Therefore, it is a challenge to define transfer items that are related to the trained behaviours and so should co-vary. As Lee (1988) stated, transfer is greatest when the processing involved in producing the treated behaviour is the same as that required for production of the transfer behaviour. At this point in time, theory may not be explicit enough to enable us to predict with confidence what behaviours will co-vary in treatment.

Subject 2 did not demonstrate transfer of treatment effects to either the novel stimulus condition or to novel responses. There is at least one possible explanation for this result. As noted above, this individual had difficulty recalling all target behaviours in the practice sessions and the practice stimuli were modified to include diagrams of articulatory postures. The cards used to elicit the target behaviours during the experimental probe were not modified. Therefore, the probe may have represented a more difficult transition for this subject compared to Subject 1, as not only did the stimulus context alter but the exact form of the eliciting stimulus. While Subject 2 also did not demonstrate transfer to novel CVC words, this is not entirely unexpected. Most of the words selected to test transfer included novel consonants and vowels. The GMPs underlying at least some of these phonemes were not trained and were unlikely to be theoretically related to the trained programmes. The finding that Subject 2 demonstrated improvements under random practice conditions and retained these effects four weeks post-treatment supports the efficacy of this approach. However, a lack of transfer stresses the need in many patient populations for training to be carried out in functional contexts.

One final issue of concern relates to experimental design. For Subject 1, the second treatment phase may have been influenced by the previous administration of the first phase. The linking of treatment condition with behaviour set was counterbalanced in an attempt to offset these potential effects. It is likely that order effects were present, as performance in the second phase was high from the first treatment session and the tendency for less stable performance in the blocked condition was not observed. However, the finding that the positive effect of random practice emerged in both phases, regardless of the behaviour set with which it was linked, is a strong indicator that the facilitation of learning in this condition is a robust effect.

Conclusions

These data represent the first well-controlled investigation of the influence of principles of motor learning in the motor speech system and in individuals with motor speech disorders. While various textbooks on these disorders have advocated the application of these principles, no empirical data had been collected to support or refute their use. The present study provides preliminary evidence that these principles may have a similar effect on learning (i.e. retention) of skilled speech motor acts as they do on learning of limb movements. We acknowledge that these results are preliminary and require replication and extension in a larger groups of subjects. The results are encouraging and provide clear justification and focus for further investigation.

Acknowledgements

The authors acknowledge Timothy Lee for helpful discussion during the writing of this paper. This research was supported, in part, by the National Institutes of Health NIDCD grant P60 DC00976.

References

Carroll, J., Davies, P. and Richman, B. (1971) *The American Heritage Word Frequency Book* (American Heritage Publishing Company Inc., New York).

Chamberlin, C. and Lee, T. (1993) Arranging practice conditions and designing instruction. In R. N. Singer, M. Murphy and L. K. Tennant (Eds) *Handbook of Research on Sport Psychology* (MacMillan Publishing Company, New York), pp. 213-241.

Chumpelik, D. (1984) The PROMPT system of therapy: Theoretical framework and applications for developmental apraxia of speech. *Seminars in Speech and Language*, **5**, 139-156.

Clark, H. and Robin, D. A. (1998) Generalised motor programme and parameterisation accuracy in apraxia of speech and conduction aphasia. *Aphasiology*, **12**, 699-713.

Del Rey, P. Liu, X. and Simpson, K. J. (1994) Does retroactive inhibition influence contextual interference effects? *Research Quarterly for Exercise and Sport*, **65**, 120-126.

Duffy, J. R. (1995) *Motor Speech Disorders: Substrates, Differential Diagnosis and Management.* (Mosby, St. Louis).

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.

Hall, P. K., Jordan, L. S. and Robin, D. A. (n.d.) *Assessment of Agility and Control of the Oral Mechanism for Speech Production.* University of Iowa, Iowa City, IA.

Kelso, J. A. S. (1995) *Dynamic Patterns: The Self-Organization of Brain and Behavior* (MIT Press/Bradford, Cambridge, MA).

Kent, R.D., Kent J. F. and Rosenbek, J. C. (1987) Maximum performance tests of speech production. *Journal of Speech and Hearing Disorders*, **52**, 367.

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 and Stratton, New York).

LEE, T. D. (1988) Transfer-appropriate processing: A framework for conceptualizing practice effects in motor learning. In O.G. Meijer and K. Roth (Eds) *Complex Movement Behaviour: 'The' Motor-Action Controversy* (Elsevier Science Publisher, B.V. North-Holland), pp. 201-215.

LEE, T. D. and MAGILL, R. A. (1983) The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 730-746.

LEE, T. D. and MAGILL, R. A. (1985) Can forgetting facilitate skill acquisition? In D. Goodman, R. B. Wilberg and I. M. Franks (Eds) *Differing Perspectives on Motor Learning, Memory, and Control* (Elsevier Science Publisher, B.V. North-Holland), pp. 3-22.

LEE, T. D., MAGILL, R. A. and WEEKS, D. J. (1985) Influence of practice schedule on testing schema theory predictions in adults. *Journal of Motor Behavior*, **17**, 283-299.

LEE, T. D. and WEEKS, D. J. (1987) The beneficial influence of forgetting on short-term retention of movement information. *Human Movement Science*, **6**, 233-245.

LEE, T. D., WHITE, M. A. and CARNAHAN, H. (1990) On the role of knowledge of results in motor learning: Exploring the guidance hypothesis. *Journal of Motor Behavior*, **22**, 191-208.

MCNEIL, M. R., ROBIN, D. A. and SCHMIDT, R. A. (1997) Apraxia of speech: Definition, differentation, and treatment. In M. R. McNeil (Ed.) *Clinical Management of Sensorimotor Speech Disorders* (Thieme, New York), pp. 311-344.

MCREYNOLDS, L. V. and KEARNS, K. P. (1983) *Single-Subject Experimental Designs in Communicative Disorders* (University Park Press, Baltimore, MD).

ROBIN, D. A. (1992) Developmental apraxia of speech: Just another motor problem. *American Journal of Speech-Language Pathology: A Journal of Clinical Practice*, **1**, 19-22.

ROBIN, D. A. (1999a). Apraxia of speech in adults: A model driven approach to understanding the disorder, diagnosis and treatment. Paper presented to San Diego State University Colloquium, San Diego, CA.

ROBIN, D. A. (1999b). Acquired apraxia of speech: From Theory to Clinic. Workshop presented to the Kansas Speech, Language, and Hearing Association, Kansas City, KS.

ROBIN, D. A., HAGEMAN, C., MOON, J. B., CLARK, H. C., WOODWORTH, G. and FOLKINS, J. W. (in prep.) *Visuomotor tracking abilities of speakers with apraxia of speech or conduction aphasia*. University of Iowa, Iowa City, IA.

ROBIN, D. A., SOLOMON, N. P., MOON, J. B. and FOLKINS, J. W. (1997) Nonspeech assessment of the speech production mechanism. In M. R. McNeil (Ed.) *Clinical Management of Sensorimotor Speech Disorders* (Thieme, New York), pp. 49-62.

ROSENBEK, J. C., LEMME, M. L., AHERN, M. B., HARRIS, E. H. and WERTZ, R. T. (1973) A treatment for apraxia of speech in adults. *Journal of Speech and Hearing Disorders*, **38**, 462-472.

RUBOW, R.T., ROSENBEK, J.C., COLLINS, M.J., and LONGSTRETH, D. (1982) Vibrotactile stimulation for intersystemic reorganization in the treatment of apraxia of speech. *Archives of Physical Medicine and Rehabilitation*, **63**, 150-153.

SALMONI, A. W., SCHMIDT, R. A. and WALTER, C. B. (1984) Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, **95**, 355-386.

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

SCHMIDT, R. A. (1991). Frequent augmented feedback can degrade learning: evidence and interpretations. In R. Requin and G. E. Stelmach (Eds) *Tutorials in Neuroscience* (Kluwer Academic Publishers, Dordrecht).

SCHMIDT, R. A. and BJORK, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, **3**, 207-217.

SCHMIDT, R. A. and LEE, T. D. (1999) *Motor Control and Learning: A Behavioral Emphasis* 3rd Ed (Human Kinetics, Champaign, IL).

SHEA, C. H. and MORGAN, R. L. (1979) Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 179-187.

SHEA, J. B. and WRIGHT, D. L. (1991) When forgetting benefits motor retention. *Research Quarterly for Exercise and Sport*, **62**, 293-301.

SMITH, A., GOFFMAN, L., ZELAZNIK, H. N., YING, G. and MCGILLEM, C. (1995) Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research*, **104**, 493-501.

SQUARE, P. A., CHUMPELIK, D. and ADAMS, S. (1985) Efficacy of the PROMPT system of therapy for the treatment of acquired apraxia of speech. In R. H. Brookshire (Ed.) *Clinical Aphasiology Conference Proceedings* (BRK, Minneapolis), pp. 319-320.

SWINNEN, S., SCHMIDT, R. A., NICHOLSON, D. E. and SHAPIRO, D. C. (1990) Information feedback for skill acquisition: Instantaneous knowledge of results degrades learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 706-716.

THELEN, E. and SMITH, L. B. (1994) *A Dynamic Systems Approach to the Development of Cognition and Action* (MIT Press/Bradford, Cambridge, MA).

THOMPSON, C. K. (n.d.) Northwestern University Sentence Comprehension Test for Aphasia. Northwestern University, Evanston, IL.

VANDER LINDEN, D. W., CAURAUGH, J. H. and GREENE, T. A. (1993) The effect of frequency of kinetic feedback on learning an isometric force production task in nondisabled subjects. *Physical Therapy*, **73**, 79-87.

VAN RIPER, C. and IRWIN, J. (1958) *Voice and Articulation* (Prentice Hall, Englewood Cliffs).

WAMBAUGH, J. L. and DOYLE, P. J. (1994). Treatment for acquired apraxia of speech: A review of efficacy reports. *Clinical Aphasiology*, **22**, 231-243.

WAMBAUGH, J. L., KALINYAK-FLISZAR, M. M., WEST, J. E. and DOYLE, P. J. (1998) Effects of treatment for sound errors in apraxia of speech and aphasia. *Journal of Speech, Language, and Hearing Research*, **41**, 725-743.

WEEKS, D. J., ZELAZNIK, H. and BEYAK, B. (1993) An empirical note on reduced frequency of knowledge of results. *Journal of Human Movement Studies*, **25**, 348-358.

WULF, G. and SCHMIDT, R. A. (1994) Feedback-induced variability and the learning of generalized motor programs. *Journal of Motor Behavior*, **26**, 348-361.

YORKSTON, K. BEUKELMAN, D. and BELL, D. (1998) Neuromotor Speech Disorders: Nature, Assessment, and Management (Paul H. Brookes, Baltimore, MD).

NCVS Status and Progress Report - 14
September 1999, 173-190

Toward a New Understanding of Apraxia of Speech: Theory, Analysis, and Treatment

Kirrie Ballard, Ph.D.
Jay Granier, M.A.
Donald Robin, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Researchers have interpreted the behaviours of individuals with acquired apraxia of speech (AOS) as impairment of linguistic phonological processing, motor control, or both. Acoustic, kinematic, and perceptual studies of speech in more recent years have led to significant advances in our understanding of the disorder and wide acceptance that it affects phonetic-motoric planning of speech. However, newly developed methods for studying nonspeech motor control are providing new insights, indicating that the motor control impairment of AOS extends beyond speech and is manifest in nonspeech movements of the oral structures. We present the most recent developments in theory and methods to examine and define the nature of AOS. Theories of the disorder are then related to existing treatment approaches and the efficacy of these approaches is examined. Directions for development of new treatments are posited. It is proposed that treatment programmes driven by a principled account of how the motor system learns to produce skilled actions will provide the most efficient and effective framework for treating motor-based speech disorders. In turn, well-controlled and theoretically motivated studies of treatment efficacy promise to stimulate further development of theoretical accounts and contribute to our understanding of AOS.

The debate over the precise characterisation of acquired apraxia of speech (AOS) has been active for many years. Researchers have interpreted the behaviours of individuals with AOS as an impairment of linguistic phonological processing, motor control, or both. While the notion of AOS as a phonetic-motoric disorder is now generally accepted, it frequently co-occurs with aphasia and differentiating between the respective phonetic-motoric and linguistic impairments has proven difficult. This paper presents the most recent developments in theory and methods to examine and define the nature of AOS. Theories of the disorder are related to existing treatment approaches. The efficacy of these approaches is examined and directions for development of new treatments are identified.

The term apraxia was coined by Steinthal (1871, cited by Roy 1978) and elaborated upon by Liepmann (1900, 1905, 1913). These early researchers regarded apraxia as a disorder of purposeful (voluntary) movement not attributable to the loss of strength, co-ordination, or mental faculty and restricted to certain body parts and functional activities. Darley *et al.* (1975) explored the syndrome of AOS, describing it as an impairment of volitional speech production in the face of preserved linguistic and motor execution abilities. Specifically, they identified the clinical features of (a) effortful groping for articulatory postures, (b) consonant phonemes more affected than vowels, (c) inconsistent, or variable, errors across productions, (d) errors that increase complexity of articulation rather than simplify, (e) errors that approximate the target within one to two features, (f) errors that represent perseveration, anticipation, and transposition of phonemes, (g) schwa insertion in consonant clusters, and (h) awareness of errors. Kent and Rosenbek (1983) also included such features as (a) slowed speech rate with prolongation of segments and transitions between segments (b) impaired co-ordination of voicing with movement of other articulators, and (c) difficulty with initiation of utter-

ances. Although many of these characteristics have not been rigorously tested, they have survived as diagnostic indicators of AOS and are widely utilised in clinical settings. Furthermore, there has been a focus on the clinical characteristics with less attention given to theorising about the underlying nature of the disorder.

Theoretical Accounts of Apraxia of Speech

A handful of theories have been offered to explain the impairment in AOS (e.g. Clark and Robin 1998, Dogil *et al.* 1994, Kelso and Tuller, 1981, Kent and Adams 1989, Kent and McNeil 1987, Mlcoch and Noll 1980, Rogers and Storkel in press, Van der Merwe 1997, Whiteside and Varley 1998). To place these theories in a broader perspective, it is beneficial to have an understanding of models of normal language production. Most language production models that have been proposed (e.g. Bock 1982, Garrett 1975, Levelt 1989) posit several stages from formulation to articulation of a message. While aspects of these models are strictly serial in nature, subunits of a message are processed in parallel (e.g. Bock 1982, Bock and Irwin 1980). Initially, a thought is formulated and then this thought is converted into abstract semantic units. From semantic information, a syntactic frame can be generated that represents the syntactic structure to be used in expressing the thought. Following these semantic and syntactic stages, a phonological representation of the message must be developed followed by a phonetic representation. According to Levelt (1989), at the phonological level the morphological and metrical composition of words is spelled out, followed by the segmental composition (i.e. consonants and vowels). Finally, phonetic syllable programmes are derived from the string of segments and these programmes specify articulatory gestures and vocal tract configuration.

The prevailing theoretical approach to AOS claims that the processes that build the phonological representation of a message are intact but the phonetic-motoric level of production is disrupted (e.g. McNeil *et al.* 1997, Shriberg *et al.* 1997a, Van der Merwe 1997). Shriberg *et al.* (1997a) summarised these approaches by referring to AOS as a deficit in sequencing the spatiotemporal aspects of movement at a prearticulatory level. McNeil *et al.* (1997) provided a more refined definition of AOS based on a growing body of experimental data. They stated that it is a phonetic-motoric disorder that affects the translation of an intact phonological representation of a message into the learned kinematic parameters for an intended movement.

Van der Merwe (1997) developed a model for considering diagnosis and management of motor speech disorders, including AOS and dysarthrias. This model includes linguistic-symbolic, motor planning, motor programming, and execution levels and relates these to neural substrates. It has numerous components in common with previous models of production (e.g. Levelt 1989) and theories dealing specifically with motor programming (e.g. Schmidt, 1975), but is unique in being specifically applied to accounting for speech and language pathologies. Initially, there is an intention to communicate, driven by the fronto-limbic system. Following this, semantic, syntactic, and morphological planning occurs and finally selection and combination of phonemes to form a phonological representation. These processes are driven by the traditional language areas – temporo-parietal cortex and Broca's area. It is from this point on that the model expands into more detail. Following generation of a phonological representation, the motor planning phase is implemented. Extensive connections between Broca's area, Wernicke's area, prefrontal cortex, premotor cortex (lateral area 6), the supplementary motor area (medial area 6), and somatosensory cortex (areas 5 and 7) underlie a complex association between stored representations and motor and sensory information for planning the speech string. Core motor plans, or the invariant spatial and temporal specifications of phonemes, are retrieved from a sensorimotor memory store and the consecutive movements involved in fulfilling these specifications are planned at the level of the articulator and inter-articulator synchronisation. At this point, variance in the realisation of temporal and spatial aspects is directly related to factors such as phonetic context and potential for coarticulation, linguistic influences on segmental duration, frequency and familiarity of the motor goal, interarticulatory synchronisation, overall speech rate, and awareness of initial conditions of the articulators such as physical perturbations. While response feedback can have no role in motor planning, it is possible that the process can be monitored centrally. That is, the adapted plan is compared to an internal stored representation that specifies permitted variance. When the limits to adaptation of spatial and temporal features of the core motor plan to the above variables are overstepped, the listener will perceive distortions and substitutions of sounds. The cerebellum and / or cortical sensorimotor areas may subserve this capacity. Finally, the articulator-specific subroutines that comprise a motor plan are temporally organised and fed forward to the motor programming level.

In Van der Merwe's model (1997), the motor programming phase involves specification of muscle-specific spatiotemporal and force parameters such as muscle tone, resistance, and absolute force, direction, range, and rate of movement. These programmes can be modified in response to sensory feedback as the movement unfolds. Neural substrates for programming include the supplementary motor area, basal ganglia, lateral cerebellum, frontolimbic system, and the primary motor cortex. The final phase in the model is execution, where the actual articulation is controlled and performed via the final common pathway. The primary motor cortex, lower motor neurones, peripheral nerves and motor

units are the neural structures directly involved in executing movement. However, other structures such as the supplementary motor area, cerebellum, basal ganglia, and thalamus may also be involved in processing ongoing feedback during motor execution.

Van der Merwe proposed that the acquired motor speech disorders of AOS and the dysarthrias can be related directly to impairments in motor planning, programming, or execution. The behavioural profile AOS may be explained by disturbances at the motor planning level – retrieval of core motor plans for phonemes, sequential organisation of movements for a single phoneme or a series of phonemes, adaptation to phonetic context, interarticulatory co-ordination, central monitoring, and relaying the motor plans to the motor programming level. It seems that motor speech disorders such as spastic, hypokinetic, hyperkinetic, and ataxic dysarthria may arise from damage affecting programming alone or both programming and execution. Flaccid dysarthria is thought to arise from damage affecting the motor execution level only. This model is possibly the most detailed and comprehensive attempt to explain impairments in the speech production process, relating subcomponents to underlying neural structures, diagnosis of motor speech disorders, and principled development of treatment strategies for such disorders. It provides a series of testable hypotheses for examining the nature of AOS and, thus, for developing treatment goals.

Kelso and Tuller (1981) proposed the coalitional theory of AOS which is consistent with, though less extensive than, the model of Van der Merwe (1997). They viewed AOS as a breakdown in the interaction between an individual and the environment that results in failure to meet behavioural goals. Kelso and Tuller argue that, for skilled actions to be co-ordinated, the neuromuscular system must be organised into functional units also known as coordinative structures (e.g. Easton 1972, Fowler 1977, Kelso *et al.* 1979, Turvey 1977). In the case of speech, these functional units govern the spatiotemporal relations between articulators during speech production. When a group of muscles is recruited as a functional unit, the relative timing of neuromuscular events between muscles within the group remains constant with changes in the absolute timing and magnitude of the activity (Turvey *et al.* 1978). Numerous studies have now demonstrated that this finely tuned spatiotemporal co-ordination between articulators is disrupted in AOS (e.g. Itoh *et al.* 1979ab, Freeman *et al.* 1978, Kent and McNeil 1987, Kent and Rosenbek 1983, Ziegler and von Cramon 1986, but see Seddoh *et al.* 1996). Collectively, these studies have supported the interpretation that AOS is a disorder affecting the phonetic-motoric level of speech production.

Kent and Adams (1989) also refer to a breakdown in co-ordination of articulator movements in AOS. They argue that, when the integrity of the motor system is compromised by AOS, the correlation between articulator movements diminishes and variability in production of target movement patterns increases. They comment that a neurologically-based temporal co-ordination disorder can account for many of the behavioural characteristics of AOS, although compensatory techniques or processes that code the contextual articulatory requirements for speech segments may play a role.

Two alternative theoretical approaches to AOS include theories of attentional resource allocation or resource capacity limitations (Clark and Robin 1998, Kent and McNeil 1987, Rogers and Storkel in press, Whiteside and Varley 1998) and linguistic-based accounts (Dogil *et al.* 1994, Dogil and Mayer unpub). To explain the frequently noted dysprosody of AOS, Kent and McNeil (1987) invoked the notion of a resource allocation problem within a motor control-based approach to speech production. They claimed that the phonetic representation codes information on syllable and segment structure separately so that these two classes of information may be differentially affected. This vulnerability to error at the phonetic-motoric programming level forces the speaker to allot more resources to the task of ordering segmental and syllabic information. They deduce that this increased resource demand results in lengthening of syllables and intersyllabic pauses and so gives rise to the secondary characteristics of dysprosody. Levelt (1989) argued that it is at this level of developing a phonetic plan that processes become available for prearticulatory editing, which is a more controlled process requiring allocation of resources.

Whiteside and Varley (1998) proposed a "cognitive-based" account of AOS which posits two routes for phonetic encoding. The direct route accesses stored "verbo-motor patterns" which specify the relative timing and force of the components of coordinative structures. It appears that a verbo-motor pattern is similar to the concept of a motor programme (e.g. Schmidt 1975) or motor plan (Van der Merwe 1997). The direct route is used for encoding frequently used phoneme sequences or syllables and utilises minimal computational resources. The indirect route is used for encoding very low frequency or novel syllables and words and involves computing the phonetic representation anew on a phoneme by phoneme basis. This route is demanding of computational resources. Whiteside and Varley claimed that individuals with AOS have lost access to verbo-motor patterns, or motor programmes, via the direct route and must compute phonetic representations phoneme by phoneme. This process predicts the reduced coarticulation seen in AOS (Mayer 1995, McNeil *et al.* 1994, Zeigler and von Cramon 1985). Furthermore, the authors proposed that the indirect route is not used efficiently in compensating for the loss of the direct route of encoding. This poor compensation is thought to result in articulatory groping (e.g. Darley *et al.* 1975), increased segmental and intersegmental dura-

tions (Bauman 1978, Collins *et al.* 1983, Freeman *et al.* 1978, Kent and Rosenbek 1983, Mercaitis 1983, Ryalls 1981, 1987, Strand 1987, Strand and McNeil 1996), and interarticulatory discoordination (Freeman *et al.* 1978, Itoh *et al.* 1979ab, Kent and McNeil 1987, Kent and Rosenbek 1983, Zeigler and von Cramon 1986). This inefficiency in coping apparently may be due to at least two factors. The indirect route may not be effective in isolation or individuals with AOS may have a coincident deficit in allocating processing resources that has a detrimental effect on their ability to utilise the more resource demanding indirect phonetic encoding route.

Rogers and Storkel (in press) proposed that AOS represents a strict resource capacity limitation. They claimed that the phonological output buffer, holding the output of speech programming processes, is limited to one syllable. In a parameter re-mapping task, normal controls and subjects with either aphasia or AOS plus aphasia were presented with a pair of printed words that differed in the feature-based similarity of the initial phoneme – shared voicing and manner, shared place and manner, or no shared features. Subjects were required to repeat the word pair as quickly and as accurately as possible. Only correct productions were analysed for the latency between the offset of the initial consonant of the first word and the onset of the initial consonant in the second word (CC interval) and the latency from offset of the final consonant of the first word and onset of the initial consonant in the second word (IP interval). For both dependent variables, there was no difference between conditions or groups for normal controls and subjects with aphasia alone. In the AOS plus aphasia group, the CC interval was longer across all conditions compared to the other groups. This slowed production of word pairs was interpreted as a phonological similarity effect. That is, when two consecutive words are phonologically similar, production of the second word is slowed (see Rogers and Storkel 1998). Rogers and Storkel (1998) proposed that this effect arises during reprogramming of the phonological buffer for the second word. The study of Rogers and Storkel (1998) involved programming one word at a time while the task used by Rogers and Storkel (in press) encouraged subjects to program two words at a time. Thus, in the latter study, slowed production was taken as evidence that individuals with AOS cannot program two words (or syllables) into the phonological buffer at a time. Some support for this explanation is provided in a study by Rochon *et al.* (1990). These authors note that maintaining verbal material in working memory requires a phonological store, or buffer, and sub-vocal articulatory rehearsal (e.g. Baddeley and Hitch 1974). They present evidence supporting a short-term memory impairment in AOS that results from reduced ability to perform articulatory rehearsal rather than impaired access to phonological representations. However, using a different task to that of Rogers and Storkel (in press), Rochon *et al.* based their conclusions on the finding that the phonological similarity effect was absent in their subjects with AOS. The study of Rochon *et al.* provides support for a phonetic-motoric, rather than a phonological, disruption in AOS.

The representational account of Dogil *et al.* (1994) and Dogil and Mayer (unpub.) takes the opposing stance that the disorder of AOS can be explained as a purely linguistic, or phonological, impairment. These authors examined AOS cross-linguistically and posited that it reflects defective implementation of phonological representations at the phonology-phonetics interface. The deficit manifests as phonological (e.g. phoneme substitutions) and phonetic (e.g. phoneme distortions) errors. Normal phonological representations are thought to differ in degree of "specification", depending on their role in the phonological structure of the language. Permanent under specification maintains that phoneme features that do not appear at any stage of derivation or representation are not specified for segments. These features represent sound properties that are phonologically irrelevant in a system (Trubetzkoy 1939) and are associated with laryngeal consonants (e.g. /h/, which are placeless), schwa-like vowels (which are targetless), and plain coronals (which are [-lateral]). Under specification of features seems to persist from phonology into phonetics (Cohn 1990, 1993, Keating 1988, Pierrehumbert and Beckman 1988, Stemberger 1993, Vollmer 1997, Zsiga 1997) with permanently under specified segments being strongly influenced by coarticulation at the phonetic-motoric level. On the other hand, features that are contrastive in the language are fully specified and less coloured by coarticulation. For segments that demonstrate very low contextual variability, temporal relations are crucial for making phonological contrasts and so "over specification" is required (Boyce *et al.* 1991, Keating 1990). Such segments include multiply articulated stops, implosives, and clicks (as in the Xhosa language). It should be noted here that some phonologists argue for eliminating the concept of under specification in phonological theory (e.g. Mohanan 1991, Broe 1993).

Dogil and colleagues suggested that AOS represents over specification of phonetic representations, rather than under specification. The reduced ability to construct under specified representations results in greatly limited coarticulation, particular difficulty with under specified speech sounds (e.g. laryngeals and schwa-like vowels), and intact representation of highly specified sounds (e.g. clicks). The authors presented data on coarticulation and laryngeal segment production in German individuals with AOS and production of clicks in individuals with AOS who speak Xhosa. Consistent with prediction, spectrographic analysis revealed considerably reduced or absent coarticulation, which they interpreted as full specification of features for each sound. Furthermore, production of typically under

specified laryngeals in German speakers was impaired while production of typically over specified clicks in Xhosa speakers was intact. These data were interpreted as support for a linguistically based theory of AOS whereby phonological and phonetic units are fully specified for each distinct articulatory gesture. The authors claimed that this has the cascading effect of severely disrupting coarticulation, which leads to disruptions in speed of articulation and precise timing of articulatory gestures. This representational account of AOS is relatively new and further experimentation is required to determine its explanatory power. However, it cannot explain the observation by Robin and colleagues (e.g. Clark and Robin 1998, Hageman *et al.* 1994, Robin *et al.* in prep) that individuals with pure AOS demonstrate nonspeech motoric impairments when tested with motor planning tasks that are more sensitive than the standard clinical battery (see below).

Experimental Analyses With Speech Tasks

While theoretical characterisations of the disorder of AOS appear quite clearly to identify specific behavioural manifestations of the disorder, clinical descriptions have lacked diagnostic power failing to clearly differentiate between certain aphasic syndromes and AOS (Buckingham 1979, Duffy 1995, McNeil *et al.* 1997). Profiles of speech behaviours observed in individuals with AOS and patients with fluent aphasias, such as conduction aphasia (CA), show considerable overlap (Kent and McNeil 1987). For example, CA has been described as an impairment in phonological encoding that affects generating and maintaining phonological codes (Friedrich *et al.* 1984), in phonemic encoding (Brown 1975), a stage in motor encoding (Yamadori and Ikumura 1975), or in pre-articulatory programming (Kohn 1984). Blumstein (1981) observed that descriptions of the speech error patterns of individuals with posterior aphasia are remarkably similar to those in AOS: (a) more frequent consonant errors than vowel errors, (b) more frequent substitutions than distortions or omissions of phonemes, (c) more errors on initiation of words than in word-final position, (d) consonant cluster reductions, and (e) substitutions that closely approximate the target phoneme.

One reason for the confusion between disorders may be the vague definitions provided for frequently used terminology. For example, while the descriptions of AOS, nonfluent aphasia, and CA include the phrase "effortful speech production", no operational definitions are provided that might distinguish the disorder groups. That is, there are no well-defined, objective criteria by which to judge the effortfulness of speech in subjects with AOS and differentiate it from that seen in the aphasias. Duffy (1995) suggested that, in AOS, effortfulness may reflect visible and audible groping of the articulators while McNeil *et al.* (1997) suggested it reflects inconsistent articulatory errors with successive attempts at the target. In CA, effortfulness may reflect word-finding blocks, perhaps due to inefficient access to a phonological buffer, with persistent attempts at the word level to correct erroneous productions (Buckingham 1979, McNeil *et al.* 1997). However, objectively and successfully teasing apart these underlying causes of effortfulness may not be possible with the tests available in most clinical settings.

McNeil *et al.* (1995) set out to describe and differentiate the speech errors of subjects with AOS, ataxic dysarthria, and CA. They postulated that consistency of error location, variability in error type, and successive approximations toward a speech target are factors that may be used to differentiate these subject populations. Subjects produced a set of stimulus words, repeating each word three times following the clinician's model. Notably, the findings contradicted accepted clinical descriptions of the disorders. Subjects with AOS were found to have high consistency of error location within a word and low error type variation within a word. Subjects with CA demonstrated the opposite pattern: low consistency of error location and high error type variation. However, the subjects with ataxic dysarthria performed as expected and were similar to the AOS group. Other findings differentiating CA from AOS were that subjects with CA produced fewer attempts across successive trials, produced most aborted attempts at the word level, and tended to achieve targets after successive attempts. The individuals with ataxic dysarthria produced very few attempts and starters and, therefore, were not included in this part of the analysis. The authors claimed that, as individuals with AOS produced most attempts at the single sound level and those with CA produced most attempts at the word level, AOS represents a phonetic-motoric impairment and CA a linguistic impairment.

Others researchers have sought support for the hypothesis that AOS represents errors at the phonetic level while CA represents errors at the phonological level. As Code (1998) points out, phonetic level errors will have implications for the phonological level particularly in perception by the listener. Code (1998) reviewed a series of studies that presented detailed analyses of the phonetic and phonological errors of these subject groups. Essentially, individuals with AOS tend to simplify articulation by producing targets in motorically easier, or unmarked, forms. One specific example of this is the voicing contrast. Voiced consonants represent the marked form of a sound while voiceless consonants represent the unmarked form. Individuals with AOS tend to have difficulty producing voiced consonants (DeRenzi et al. 1966, Fry 1958, MacNeilage 1982). Code and Ball (1982) presented evidence from a single case of AOS that supports the hypothesis that these voicing errors represent phonetic level impairment rather than a phoneme substitution process. They noted that their subject was unable to generate voicing in voiced fricatives but was sen-

sitive to, and able to manipulate, other phonetic level features that signal a voiced fricative – duration of the preceding vowel and duration of the friction on the consonant. The individual's errors clearly did not represent impairment at the phonological level.

Examining disturbance of temporal characteristics of speech has proven fruitful in distinguishing motoric from linguistic speech disorders. Seddoh *et al.* (1996) argued that temporal abnormalities may arise from impairment to either the linguistic or motoric systems but that the nature of the abnormalities will differ. That is, specific time-based information serves linguistic functions. For example, in English syllable timing rules exist which serve to vary the duration of unstressed syllables as a function of the sequential relationship with stressed syllables (Bolinger 1976, Cooper *et al.* 1977, Klatt 1976, Lehiste 1972). Seddoh *et al.* (1996) argued that it should be possible to differentiate temporal abnormalities due to motoric versus linguistic impairments.

One frequently cited temporal characteristic of apraxic speech is a reduction in overall speech rate. In detailed analyses of this phenomenon, AOS speakers have been compared to individuals with normal language or aphasia and found to demonstrate discoordination of voice onset time (Freeman *et al.* 1978, Itoh *et al.* 1979a, Kent and Rosenbek 1983), increased vowel durations (Collins *et al.* 1983, Freeman *et al.* 1978), and increased consonant durations (Kent and Rosenbek 1983). These effects have been observed for multisyllabic words and phrases and multisyllabic nonsense words and phrases (Bauman 1978, Collins *et al.* 1983, Kent and Rosenbek 1983, Mercaitis 1983, Ryalls 1981, 1987, Strand 1987, Strand and McNeil 1996). Furthermore, AOS speakers have protracted intersegment durations and transition durations (Kent and Rosenbek 1983, Mercaitis 1983). Other studies have demonstrated abnormal temporal features of speech on acoustic (Collins *et al.* 1983, Kent and Rosenbek 1983, McNeil *et al.* 1990a), kinematic (McNeil and Adams 1991), and perceptual (Odell *et al.* 1990, 1991) measures. These abnormalities all affect the realisation of phonemes at a segmental, intersegmental, or word level and may serve to decrease speech rate.

The study by Kent and McNeil (1987) was also designed to examine the nature of the observed slow speech rate of individuals with AOS. They found that the speech of individuals with AOS was slower than individuals with CA and normal subjects, with increased segment and intersegment durations. Individuals with AOS also were unable to effectively increase speech rate on command, suggesting motoric inflexibility. This is contrary to McNeil *et al.* (1997), who stated that AOS speakers are able to increase their rate of speech but at the expense of phonemic integrity. While subjects with CA also demonstrated pauses, these were not as pervasive and their speech attempts demonstrated periods of normal prosody. While they were able to increase speech rate more effectively than subjects with AOS, indicative of greater motoric flexibility, the number of phonemic and phonetic errors increased at the faster speaking rate. Kent and McNeil (1987) concluded that both disorders represent some degree of phonetic-motoric impairment: (a) CA primarily affects phonetic coding with secondary effects on working memory for phonetic coding and motor control processes, (b) AOS primarily affects motor control processes with a secondary effect on phonetic coding.

Robin *et al.* (1989) examined the rate of articulator movements during speech as well as inter-articulator temporal co-ordination in five individuals with AOS. They measured peak articulatory velocity in the lower lip as well as co-ordination of articulatory movements in the lower and upper lip in relation to perceptual errors in speech. They found that the subjects were able to generate high peak articulatory velocities during speech and did not demonstrate discoordination in the relative timing of upper and lower lip movements. The latter finding is in contrast to Itoh *et al.* (1979a) who studied co-ordination of lingual and velar movements, although Robin *et al.* commented that temporal co-ordination between the lips may be less sensitive to disruption (Konno *et al.* 1987). Furthermore, Robin *et al.* reported that peak articulatory velocity and co-ordination of relative timing of the lips were not systematically related to speech rate or presence of phonetic errors in speech in the individuals tested. Consistent with previous studies (e.g. Kent and McNeil 1987, Kent and Rosenbek 1983), these subjects had difficulty manipulating rate of speech for syllable and sentence level material and demonstrated increased duration of segment and intersegmental components. Robin *et al.* suggested that while speakers with AOS tend to have a slowed rate of speech with increases in durational elements, these characteristics do not arise from slowed velocity of articulatory movements.

Square-Storer and Apeldoorn (1991) presented an interesting study of the acoustic characteristics of three individuals with pure AOS and related these results to site of lesion. Subjects were diagnosed according to criteria developed at the Mayo Clinic (Darley *et al.* 1975). Notably, this system does not include dysprosody as a defining characteristic. One subject demonstrated a left basal ganglia lesion, another a left parietal and subcortical lesion, and the third bilateral parietal lesions. One- to four-syllable words were elicited two to three times by having subjects read, repeat, or name objects. Duration of segments, syllables, words, pauses, and additions and amplitude of the syllable nuclei were measured and related to the prosodic features of speech rate and stress patterns. Based on their results, Square-Storer and Apeldoorn suggested that the description of AOS as having unrelieved periods of dysprosody (Kent and Rosenbek 1983, Robin *et al.* 1989) may be rel-

evant only for more severe cases as their subject with milder AOS did not demonstrate this feature. Furthermore, the features of slowed speech rate and abnormal stress patterns were demonstrated in the two subjects with subcortical involvement and not in the subject with bilateral parietal damage. Given that the two former subjects demonstrated hemiplegia, the authors suggested that disruptions in rate and stress patterns may be signs of a concomitant unilateral upper motor neurone dysarthria. This study certainly highlights the problems of strictly diagnosing subjects *a priori* when clinical tools that adequately differentiate AOS from other motor speech disorder and some linguistic disorders are not yet available.

Another avenue researchers have employed, in attempts to differentiate AOS and CA, is degree of variability. The features traditionally used to define AOS serve to give individuals with AOS the appearance of being variable in their attempts to produce speech. While increased variability between and within subjects has been used to describe individuals with aphasia and speech motor disorders (e.g. Kent and McNeil 1987), increased levels of variability on temporal measures during speech have been taken to signify instability in motor control systems (DiSimoni 1974ab, Janssen and Wieneke 1987, Kent and Forner 1980, Sharkey and Folkins 1985, Smith 1992, 1994, Smith and Kenney 1994, Tingley and Allen 1975, Wieneke and Janssen 1987). That is, variability may reveal reduced control in reaching intended motor targets due to impairment. Also, age is a factor with variability decreasing through childhood and increasing again in normal older adults, likely due to maturation and deterioration of the motor control system, respectively (Ballard *et al.* in prep, DiSimoni 1974ab, Kent and Forner 1980, Smith 1992, 1994, Smith and Kenney 1994, Tingley and Allen 1975). With increased variability AOS speakers are bound to produce erroneous movement patterns, or patterns outside the acceptable range of normal, with greater frequency than normal subjects. Folkins (1985) argues that some of this increased variability may also reflect increased flexibility in dealing with an unstable motor control system to achieve perceptually acceptable tokens.

Several researchers have considered the variability of temporal characteristics of speech in relation to durational measures (Kent and McNeil 1987, McNeil *et al.* 1989, Robin *et al.* 1989, Seddoh *et al.* 1996). Kent and McNeil (1987) considered the temporal characteristics of sentence production in normal subjects (n = 3) and individuals with AOS (n = 3) or CA (n = 2). Subjects produced two sentences at a comfortable rate and a faster rate. Segment durations were significantly longer in the speech of the AOS and CA subjects compared to the normal subjects. However, the AOS subjects demonstrated greatest variability in performance. At the fast speaking rate, the AOS and CA subjects performed similarly in terms of segment duration values and variability. Intersegment durations were longer and more variable in the AOS individuals than the CA and normal individuals, especially at the faster speaking rate. On measurements of voice-onset time, all subjects with AOS and one subject with CA demonstrated longer than normal voice onset times. The second subject with CA had voice onset times close to or briefer than normal. Similarly, both AOS and CA subjects demonstrated more variability in second formant transitions than the normal subjects. Thus, some temporal parameters of speech may be affected in both AOS and CA.

Seddoh *et al.* (1996) studied four individuals with CA, five with AOS, and normal age-matched controls. They measured accuracy and variability of stop gap duration, voice onset time, second formant transition duration for a vowel, steady state vowel duration, and consonant-vowel duration in perceptually accurate repetitions of a given phrase (i.e. "That's a pop / pea / Bob / bee a day"). They found that speakers with AOS and CA were not clearly differentiated by measures of mean duration of segmental and intersegmental components of speech but were differentiated by the amount of variability on these measures. Contrary to Kent and McNeil (1987), they found that both AOS and CA demonstrated extended segment and intersegment durations indicative of abnormal temporal control during speech. On stop gap duration, AOS had significantly longer mean duration than the CA and normal groups, who did not differ. This is taken to support a motoric deficit in AOS and absence of motor involvement in CA. On vowel duration, both AOS and CA subjects demonstrated longer mean durations than normal, although for the CA subjects this was across all target words and for AOS subjects it was only for words with the tense vowel /i/. The AOS subjects demonstrated greater token-to-token variability as individuals and also significantly greater variability as a group on stop gap duration, vowel duration, and consonant-vowel duration compared to age-matched normal controls. Compared to speakers with CA, AOS speakers were more variable on stop gap duration and consonant-vowel duration. However, relatively normal levels of variability on voice onset time and second formant transitions in individuals with AOS prompted the authors to conclude that some aspects of temporal control are preserved. Subjects judged to have a more severe speech impairment demonstrated the greatest token-to-token variability. In comparing the CA speakers and age-matched normal subjects, there was no difference in variability of stop gap duration but CA speakers showed significantly greater variability for consonant-vowel duration. CA speakers did not differ from AOS or age-matched controls on variability of vowel duration. From these data, the authors concluded that both AOS and CA subjects present with abnormal temporal characteristics but the greater variability of the AOS subjects indicates a motoric deficit in AOS and a phonological deficit in CA. However, the results lend sup-

port to Kent and McNeil's (1987) conclusion that CA may also involve a subtle sensorimotor deficit.

Seddoh *et al.* (1996) also reflected on the issue of motoric flexibility. All tokens analysed in their experimental task were perceptually accurate. However, differences in acoustic characteristics between AOS, CA, and normal subject groups were still detectable. The authors suggested that subjects with AOS demonstrate a greater flexibility in compensating for their motoric instability so that they still achieve perceptual accuracy in a constrained task. Although this interpretation is consonant with that of Folkins (1985), it is in contrast to Kent and McNeil (1987) who hypothesised that the inability of their subjects to manipulate speech rate reflected motoric inflexibility.

Further support for increased variability in AOS comes from McNeil *et al.* (1989) who studied variability of peak articulatory velocities of the lower lip in four subjects. They found that, in word targets, the subjects with AOS were more variable than normal but did not differ on mean velocity.

Taken together, the studies reviewed above best support theoretical explanations of AOS which place the locus of the deficit in motor planning processes affecting the translation of an intact phonological representation of a message into the phonetic-motoric representation prior to execution by the articulators (Kelso and Tuller 1981, Kent and Adams 1989, McNeil *et al.* 1997, Van der Merwe 1997, Whiteside and Varley 1998). This disruption would affect retrieval of motor plans and / or specification of the spatiotemporal parameters of movements. These disruptions would manifest as trial and error groping as the individual attempts to retrieve a motor plan or sequence motor plans, distortion of phonemes, segmental and intersegmental durational changes with reduced speech rate, and loss of interarticulator co-ordination. The framework of Van der Merwe is the most explicit and can account for all of these disruptions. Capacity limitation theories, such as Rogers and Storkel (in press) provide an alternative perspective but may only account for some of these characteristics. However, none of these theories address the now robust literature demonstrating that the motor control impairments of AOS extend to nonspeech movements.

Experimental Analyses With Nonspeech Tasks

While it is necessary to consider the impairment of AOS in the context of speech production tasks, also studying nonspeech behaviours has the potential to disambiguate which characteristics are a result of the underlying motor impairment and which are related to the interaction between the motor and linguistic systems. Hageman *et al.* (1994), Itoh *et al.* (1979a), McNeil and Kent (1990), and McNeil *et al.* (1990b) have reported that subjects with AOS demonstrate impaired movement of the articulators during both speech and nonspeech tasks. This line of research should reveal the motoric disturbances that give rise to the profile of intact and disrupted perceptual and acoustic characteristics that have been detailed (see McNeil *et al.* 1997 for an extensive review). The motoric disturbances underlying AOS can be revealed in nonspeech tasks provided that the measure is sensitive to such disturbances. It is likely that standard clinical measures of oral and speech motor programming often lack the sensitivity to detect these disturbances, particularly in cases of milder impairment.

A number of nonspeech motor control tasks have been developed which are thought to be sensitive to the demands on the motor system during speech. These include tasks requiring subjects to control static position or isometric force of articulators (Barlow and Abbs 1986, McNeil *et al.* 1990b) and pursuit tracking of a visual signal with the articulators (Clark and Robin 1998, Hageman *et al.* 1994, McClean *et al.* 1987, Moon *et al.* 1993, Robin *et al.* in prep). In these tasks, a target level of force or position or a target movement pattern is displayed visually, for example on an oscillographic screen. A transducer is placed on the articulator of interest and the transduced signal is overlaid on the visual display. The subject is instructed to match the two signals and so achieve a given level of force or a target position for a given period of time or, in the case of visuomotor tracking, follow a given movement pattern. These techniques have clear application to assessment of static and dynamic aspects of motor control of the articulators.

Barlow and Abbs (1986) examined the stability of force in the lip, jaw, and tongue in subjects with spastic dysarthria and related these nonspeech measurements to perceptual accuracy of speech. Control of muscle force was more variable, or less stable, than in normal controls and demonstrated a significant correlation with perceptual judgements of speech intelligibility. McNeil *et al.* (1990b) compared subjects with AOS, ataxic dysarthria, CA, and normal controls on isometric force and static position control. According to the traditional view of AOS, one would not predict impairments in muscle force. They reported significantly less stable control of both static position and force in AOS and ataxic dysarthria compared to control subjects. It is noted that the mechanism underlying the similarity in performance between AOS and dysarthria is not necessarily the same. Performance of individuals with CA fell between normal controls and the apraxic and dysarthric groups. Consistent with previous studies by McNeil and colleagues, McNeil *et al.* (1990b) suggested that CA involves some degree of orofacial sensorimotor impairment coexisting with the primary linguistic deficit.

The visuomotor tracking task has long been utilised by researchers studying normal motor control and motor skill learning in the limbs (see Poulton 1974) and impaired motor control of upper extremities in Parkinson's Disease

(e.g. Flowers 1978). It was first applied to the oromotor system by McClean *et al.* (1987) who examined normal subjects and individuals with acquired dysarthria. In visuomotor tracking with the articulators, subjects gaze at an oscilloscope screen and use their lower lip, jaw, or voice (i.e. F_0) to track the movement of a horizontal bar (i.e. the target signal) with a cursor (i.e. the tracker signal). Three predictable target signals (i.e. 0.3, 0.6, and 0.9 Hz) and an unpredictable target signal (i.e. a complex signal composed of ten equal amplitude frequencies from 0.1 to 1 Hz) have been used. The velocity and amplitude of movements required by the task are within the range used during speech and, to some degree, approximate movements that occur during speech. Furthermore, it may be argued that a specific " programme" or stored plan of movement is implemented during jaw closure regardless of the behavioural context, but the parameters of that motor programme, such as the absolute speed and amplitude of the movement, will differ with context (e.g. Schmidt and Lee 1999). That is, a single motor programme that drives jaw closure may be implemented during mastication and speech but the speed and amplitude of the movement (i.e. parameters of the motor programme) will differ for these two activities (but see Smith and Denny 1990).

Hageman *et al.* (1994) tested the jaw, lower lip, and voice tracking abilities of normal subjects and individuals with AOS. They reported that normal subjects' accuracy in tracking declined as the frequency of predictable signals increased and tracking was poorest for the unpredictable signal. On predictable signals, AOS subjects were consistently poorer than normal controls but demonstrated a similar decline in accuracy with increasing frequency. While the normal controls tracked the predictable targets with smooth articulator movements, the subjects with AOS produced "jerky" movements. Notably, subjects with AOS tracked unpredictable target patterns more accurately than predictable targets and at a level of accuracy similar to normal subjects. The authors argued that, for predictable targets, normal subjects were able to formulate and follow an internal model, or motor programme, of the target signal (Hageman *et al.* 1994). That is, subjects initially perform in a reactive, or feedback, mode having to closely monitor the target in order to extract a pattern and develop a motor programme to execute the task. Once the programme is developed, subjects move into a predictive, or feedforward, mode implementing the programme with only occasional sampling of the target to ensure accuracy and execute corrections if accuracy is deficient. Subjects with AOS, on the other hand, seemed unable to develop such an internal model or programme. They appeared to remain in a reactive mode relying solely on feedback to execute the task. This interpretation is supported by the "normal" tracking of the unpredictable signal by subjects with AOS. In unpredictable target tracking, it is not possible to develop a programme of movement and so both subject groups are forced to use feedback mechanisms to execute the task. Hageman *et al.* concluded that AOS reflects an impairment in developing or implementing motor programmes for articulator movements both in speech *and nonspeech* tasks.

Hageman *et al.* (1993) compared subjects with AOS, ataxic dysarthria, and normal controls. They reported that, contrary to individuals with AOS, those with ataxia demonstrated poorer correlations between target and tracker than normal subjects for both predictable and unpredictable signals. These results for ataxic dysarthria indicate that it is a motor execution disorder, rather than a motor programming disorder, so that tracking is affected for both target types.

A recent study by Robin *et al.* (in prep.) considered the relation between nonspeech VMT ability and perceptual measures of speech accuracy in individuals with AOS and CA. Subjects tracked both predictable and unpredictable targets with the lip and jaw. Consistent with previous work (Hageman *et al.* 1993, 1994), tracking performance for individuals with AOS was poorer than normal for predictable signals only and individuals with CA were not differentiated from normal subjects in either condition. For subjects with AOS only, performance on tracking predictable signals was highly correlated with measures of speech accuracy. Given that other behaviours such as peak articulatory velocity (Robin *et al.* 1989), temporal co-ordination of upper and lower lip movements during speech (Robin *et al.* 1989), and accuracy and variability of segment and intersegment durations (Seddoh *et al.* 1996) have not proven useful in predicting speech accuracy, these findings support the use of the tracking paradigm in assessment of AOS. Furthermore, these findings strongly support the notion that AOS is a disorder of motor planning that affects articulator movements in both speech and nonspeech tasks while CA is primarily a linguistic impairment.

Clark and Robin (1998) further explored the concept of a motor programming impairment in AOS by examining the formation of motor programmes and the setting of temporal and amplitude parameters in normal controls and individuals with AOS or CA. They predicted that all subjects with AOS would show reduced accuracy of the motor programme and that some may show reduced temporal parameterisation accuracy and / or reduced amplitude parameterisation accuracy. Furthermore, they predicted that individuals with CA would perform similarly to normal controls as they do not have a motor programming impairment. A variation on the visuomotor tracking task was employed with the jaw only. Each subject was required to gaze at a monitor, study a given movement pattern (i.e. a waveform) that appeared on the monitor, and then attempt to replicate the pattern by opening and closing the jaw after the pattern had disappeared from the screen. The screen remained blank while the subject attempted to replicate the

pattern, and two seconds later, feedback was provided by displaying the subject's production superimposed over the target waveform. Accuracy of the motor programme was determined by calculating the residual or relative difference between the subject's production and the target waveform, after correcting for absolute temporal and amplitude differences. Accuracy of temporal and amplitude parameters was determined by calculating absolute differences between the duration and amplitude of the subject's production and the target waveform.

Contrary to prediction, Clark and Robin (1998) found dissociation between motor programme accuracy and parameterisation accuracy among their AOS subjects. Either motor programme formation or parameterisation was impaired, but not both. The authors suggested that the motor programming impairment in AOS affects motor programme accuracy or parameterisation accuracy and that the dissociation between the two may reflect individual differences in resource allocation strategies. This dissociation was also observed within subject, where individuals with AOS were able to improve parameterisation accuracy but only at the expense of motor programme accuracy or vice versa. This effect may be due to an impaired ability to allocate attentional resources. Thus, the speaker with AOS may be incapable of producing accurate motor programmes or setting parameters correctly during complex motor activities and be forced to choose which of the two processes to execute more accurately. An alternative interpretation is that the process of developing, recalling, and / or implementing a reasonably accurate GMP or of manipulating parameter values may be more demanding of attentional resources in the individual with AOS. Therefore, in a finite system, fewer resources are available for one or the other and performance deteriorates along one dimension.

In summary, nonspeech physiological research into AOS has indicated that subjects with AOS have difficulty with developing or implementing models of movement patterns and / or setting the absolute timing and amplitude parameters of those movement patterns. It appears that this impairment also may have an impact on the availability and allocation of attentional resources for performing and adapting actions (Clark and Robin 1998, Rogers and Storkel 1999). The difficulty with motor control for nonspeech tasks appears to be one of the most consistent findings in the literature and is not wholly unexpected (see below). These data demand that models of AOS focusing predominantly or entirely on speech production need substantial revision. As is the case with the dysarthrias (e.g. Darley *et al.* 1975, Duffy 1995), where the basic nonspeech condition and movement control define the speech anomaly, we propose that AOS must be defined at multiple levels and the fundamental impairment needs to be considered at the level of nonspeech motor control.

The nonspeech motor control problems of AOS can be accounted for from a number of different theoretical postures. For instance, our work largely has been guided by theories of motor programming and, in particular, the concepts of generalised motor programmes and the parameters that set absolute aspects of programmes (e.g. Schmidt and Lee 1999). The nonspeech data also can be interpreted within recent models of dynamic systems (e.g. Kelso 1995, Thelen and Smith 1994). In the following discussion we present these two models of motor control and propose a broader view of AOS that encompasses both nonspeech and speech motor control difficulties.

A Broader View of Acquired Apraxia of Speech

While current theories of AOS account for the effects of the impaired motor system on speech with varying degrees of specificity, they do not explain the effects on nonspeech actions. We propose that, in order to incorporate findings from nonspeech tasks, we need to conceptualise AOS as a disorder of motor control that also may have an impact on the availability and allocation of attentional resources for performing and adapting actions (Clark and Robin 1998, Rogers and Storkel 1999). We rely on such theoretical stances as that presented by Folkins and Bliele (1990) and Saltzman (1986) which claim that the motor system is not necessarily organised around presumed units of language or speech. Rather, it is assumed that the motor system has its own cognitive architecture which is activated and monitored, in part, by the language system.

We propose, on a theoretical basis and from empirical data, that AOS is a motor control disorder that has its basis in nonspeech anomalies. When these motor control anomalies are such that they interfere with the production of speech, a phonetic-motoric disorder emerges that is called AOS. In our view, the issue that confronts scientists and clinicians is how to model the underlying motor anomalies in AOS. Once this has been accomplished, extension of such modelling to include speech can occur. Two opposing models of motor control have explanatory power in understanding both the speech and nonspeech profiles of AOS – schema (e.g., Schmidt 1975, Schmidt and Lee 1999) and dynamic systems (e.g. Kelso 1995, Thelen & Smith 1994) theory.

Schema theory (Schmidt 1975, Schmidt and Lee 1999) has been one of the most influential theoretical approaches to motor control and learning and was developed, through the study of normal individuals learning novel limb movements or performing skilled actions. Briefly, this theory assumes that learning results from developing and refining a representation of an action that prescribes the relation between how we move and the demands of a given task. Two primary concepts in schema theory are generalised motor programmes (GMP) and parameters. GMPs contain

an abstract code about the relative timing of actions and the relative force with which actions are to be produced. Parameters specify details about how a GMP will be expressed in terms of absolute duration of movement, absolute force of muscle contractions, and the muscles or limb used to make the movement. Through manipulation of parameters, numerous and even novel actions can be completed using a single GMP. These two concepts in combination serve to reduce storage demands for representations of action and explain the relatively invariant features across different productions of an action. As with most fields, there is terminological confusion when applying this work to the area of speech motor control. For example, in the model of Van der Mewre (1997) described above, the term "motor plan" seems to correspond rather directly to Schmidt's concept of GMPs and the term "motor programme" to Schmidt's concept of parameters.

AOS has traditionally been defined as a disorder of motor programming (e.g. Darley *et al.* 1975). Within the framework of schema theory, the speech errors are considered to be the result of disruption to the programmes that drive sound production. There is a historical precedent to using schema theory to explain the speech impairment in AOS (e.g. Kent and Rosenbek, 1983, Mlcoch and Noll, 1980). The issue of defining what constitutes a motor programme in speech production is far from resolved (e.g. Smith et al. 1995). However, if one assumes that there are GMPs that need to be parameterised for speech production at some level, then AOS can be thought of as a breakdown in activation and selection of GMPs, in specification of the parameters for those programmes, or both. Importantly, this model also can account for the nonspeech findings in AOS. Specifically, Clark and Robin (1998) reported that individuals with AOS were unable to execute a model jaw movement pattern accurately and appeared to trade-off accuracy of GMP execution for parameterisation accuracy, or vice versa. Likewise, we interpret the tracking data described above within this model. Normal subjects are able to develop a GMP rapidly in order to perform visuomotor tracking of predictable targets in a predictive mode. They also are able to manipulate the temporal parameter adeptly so as to track predictable signals of increasing frequencies with high accuracy. Individuals with AOS have difficulty in either developing the GMP or in utilising feedback to maintain accuracy. They also are unable to manipulate the temporal parameter to produce a given speed of movement in the context of the visuomotor tracking task. Tracking of unpredictable targets does not permit development of a GMP and performance on this condition does not distinguish individuals with intact motor control and those with AOS.

An alternative approach to motor control is based on dynamic systems (e.g. Kelso 1995, Thelen & Smith 1994). Rather than rely on concepts of motor programs and parameters, proponents of the dynamic systems approach argue that the "invariant" features in movement are represented as "stable states". In relation to motor skills, these states are action patterns that emerge through (a) the interaction of the parts of a system with each other and with the external environment, (b) inherent constraints on the system, and (c) the available supply of energy. In essence, each time one performs an act, it emerges as a new form. Some aspects of the act will be stable and predictable over repeated performances while others will reflect the variable, flexible, and adaptive aspects that take their form according to changes in the stimuli that were present at the time of a given action. As a given set of conditions reoccurs, the emergent behavioural patterns increase in stability and develop into a stable state and the order and complexity in the system may increase. The nature of dynamic systems permits great plasticity to adapt and reorganise in response to new conditions and to learn new acts. These phenomena are demonstrated in the developing human as new behaviours are learned and adaptations are made to accommodate the changing cognitive and neuromuscular systems (Thelen and Smith 1994). Similarly, as changes in cognitive and neuromuscular systems occur with ageing or with impairment, stability is disturbed and patterns of behaviour are disrupted or lost.

Speech sounds can be viewed as emergent properties with production of speech sounds in the adult representing stable states with the behavioural goal of perceptual adequacy. In this model, AOS can be viewed as a breakdown in some aspect of stable state formation. It is possible that, in the impaired individual, the system cannot self-organise so that stable behavioural patterns do not emerge. Alternatively, some states may be pathologically stable, or inflexible (Folkins 1985), so that the system remains in that state when it is no longer appropriate (e.g. perseveration). Furthermore, the system may achieve stable states but be overly sensitive to perturbation or feedback or be unable to adapt effectively to such perturbation or feedback. Kelso and Tuller's (1981) coalitional theory of AOS is based on early formulations of dynamic systems theory within the field of speech motor control. The dynamic systems approach accounts for the nonspeech data above. The data of Clark and Robin (1998) appear to implicate a system that can self-organise into a stable state but that does not adapt effectively to changes in the stimulus. The case of high GMP accuracy but low parameterisation accuracy may represent an inflexible, or excessively stable state. The converse – low GMP and high parameterisation accuracy – may represent an overly sensitive stable state. Data from the visuomotor tracking studies suggest that, when tracking predictable signals, individuals with AOS have difficulty achieving or remaining in a stable state. The unpredictable tracking condition, however, does not permit a behavioural pattern to emerge and, again, normal and impaired individuals perform similarly.

Although quite speculative at this point in time, these models provide a framework for research on clinical applications addressing description, diagnosis, and treatment of motor speech disorders. Specifically, we present three clinical issues that bare directly on the relation between speech and nonspeech motor control in AOS. First, it is known that there are individuals with AOS who perform within normal limits on routine oral motor examinations that include tasks such as alternating labial protrusion and retraction. Our contention (see Robin *et al.* 1997) is that most nonspeech motoric tasks used in clinical settings are too far removed from speech so that they permit detection of only fairly severe apraxic disturbances. Given that even patients with very mild AOS perform abnormally on the visuomotor tracking task described above, this task appears to be a sensitive measure of praxis for nonspeech oral motor tasks. Considering the converse situation, we predict that any individual who performs poorly on the relatively less sensitive routine oral motor exam will show some degree of AOS.

A second issue relates to development of a diagnostic marker for AOS. As is the case with most disorders of speech and language, a systematic approach to accurate diagnosis of AOS will most likely stem from a detailed and sensitive range of both nonspeech and speech tasks. However, data from our laboratory (e.g. Clark and Robin 1998, Hageman *et al.* 1994, Robin *et al.* in prep) are intriguing to consider in this light. Though preliminary in nature, our studies of visuomotor tracking ability have demonstrated that every individual with AOS that has been tested to date has an impaired ability to track predictable targets in the face of "normal" tracking of unpredictable signals. These data lead to the hypothesis that nonspeech visuomotor tracking performance is a diagnostic marker for AOS. If this hypothesis holds, then the presence of impaired tracking of predictable targets but normal tracking of unpredictable targets, regardless of other cognitive or linguistic (i.e. aphasia) problems, will support a diagnosis of apraxia. Note that Shriberg *et al.* (1997bc) have argued that prosodic abnormalities may be diagnostic markers for developmental apraxia of speech. However, like many models of AOS, their model fails to account for the nonspeech findings in the childhood apraxia literature and does not consider the fact that prosodic anomalies are robust throughout all motor speech disorders including the dysarthrias.

The final clinical issue that has driven our work with AOS is the development of a principled approach to treatment. Our findings over a range of experimental contexts, have led us to the area of motor control and learning (Schmidt and Bjork 1992, Schmidt and Lee 1999) as one that can provide a model driven approach to treatment of motor speech disorders. We review approaches to treatment and their efficacy below and present some preliminary data supporting the application of specific principles of motor learning (Schmidt and Lee 1999) developed and tested in other motor systems.

Treatment

Taking into account the findings from studies of both speech and nonspeech tasks above, well-motivated and principled approaches to treatment can be developed based on broader theories of motor control and learning. To date, surprisingly few studies exploring treatment approaches to AOS and efficacy of these approaches have been published (Wambaugh and Doyle 1994). The majority of these studies have proposed methods that presume to alter the organisation of the neuromotor system and, thus, bring about relearning of motor speech skills. Specifically, they focus on developing motor plans, programmes, by retraining the production of spatiotemporal aspects of sounds and sound sequences. They do not directly provide practice in independently recalling these newly developed, or relearned, motor plans – an approach which would likely facilitate retention and transfer to novel contexts (see below). They typically do not focus on varying the phonetic context in which sounds are practised with the explicit goal of affecting coarticulation or interarticulatory co-ordination. References to available theories of motor control and learning are rare in these studies and treatment goals typically are not overtly motivated by theoretical accounts (but see Wambaugh *et al.* 1996). In the context of Van der Merwe's framework (1997), AOS may represent an impairment in one or more of the following processes – (a) retrieval of motor plans for phonemes, (b) sequential organisation of movements for a single phoneme or a series of phonemes, (c) adaptation to phonetic context in the form of coarticulation, (d) interarticulatory co-ordination, (e) central monitoring by comparison with an internal stored representation of the plan, and (f) relaying the motor plans to the motor programming level for specification of parameters of movement. While some of these processes may not be amenable to intervention, it seems relevant to determine which are affected for a given individual and structure therapy accordingly.

In direct contrast to the speech motor learning literature, numerous studies have considered motor learning of limb movements in normal children and adults. Many of these studies are based upon the schema theory of motor control and learning (Schmidt 1975, Schmidt and Lee 1999). In reviewing this body of literature, it appears that development of flexible and skilful actions is influenced by several principles of motor learning that pertain to the structure of practice and the nature of feedback (see Schmidt and Lee 1999 for an extensive review). Briefly, learning is facilitated when (a) several skills to be learned are practised in random order rather than in blocks, (b) individuals produce a high number of trials within a practice session, (c) feed-

back is given on about 50% to 60% versus 100% of trials, (d) feedback is given simply on correctness of a response rather than specific aspects of how the response was performed, particularly in later stages of acquisition, and (e) a brief unfilled 3-4 sec delay is present between the stimulus and response and between the response and the next stimulus. Schmidt and Bjork (1992) observed that conditions facilitating acquisition of a skill during practice may actually interfere with, or minimise, retention and transfer. Obviously, retention can not occur without acquisition but this research indicates that retention is the more informative measure of a training programme's efficacy.

It is reasonable to expect that principles of motor learning will apply equally to the speech motor system. A few specific principles, such as performing a high number of trials, have been applied in many treatment programmes for AOS (e.g. Phonetic Placement: Van Riper and Irwin 1958, Eight-Step Task Continuum: Rosenbek *et al.* 1973, Melodic Intonation Therapy: Sparks *et al.* 1974, Prompts for Restructuring Oral Muscular Phonetic Targets (PROMPT): Chumpelik 1984, Square *et al.* 1985, Phonetic Derivation and Phonetic Approximation: see Square-Storer 1989, Phonetic Contrasts or Minimal Pairs Treatments: e.g. Wambaugh *et al.* 1998). However, these studies have not directly compared the permutations of these principles, such as high versus low number of trials. Also, little empirical evidence has been amassed to support or refute the application of other motor learning principles in treating the speech production impairments of AOS. It is also the case that many treatment programmes that purport to be "motor-based" actually have procedures that are in direct antithesis to the principles of motor learning. Finally, the focus of the majority of studies has been on performance during acquisition of a treated skill rather than retention and transfer of treatment effects (Wambaugh and Doyle 1994).

Wambaugh and Doyle (1994) reviewed 28 studies reporting on the efficacy of treatment for AOS. Of the 26 studies that treated speech production, as opposed to alternative communication modalities, the focus was on acquisition performance. Only eight studies reported on retention of treatment effects (Dworkin *et al.* 1988, Lane and Samples 1981, LaPointe 1984, Raymer and Thompson 1991, Simmons 1980, Square *et al.* 1985, Thompson and Young 1983, Warren 1977), with four of these claiming positive results. Six studies reported on transfer to untreated responses (Dworkin *et al.* 1988, LaPointe 1984, Raymer and Thompson 1991, Square *et al.* 1986, Square-Storer and Hayden 1989, Thompson and Young 1983), with one study claiming positive results (Square *et al.* 1986) and three only limited transfer. Three studies systematically tested transfer to untreated stimuli (Dworkin *et al.* 1988, Raymer and Thompson 1991, Southwood 1987), with one claiming positive results (Dworkin *et al.* 1988). Notably, four of the eleven studies reporting on retention and / or transfer did not utilise a controlled experimental design (Lane and Samples 1981, Square *et al.* 1985, 1986, Square-Storer and Hayden 1989) so that results from these studies must be interpreted with caution. The studies by Square and colleagues (Square *et al.* 1985, 1986, Square-Storer and Hayden 1989) applied the PROMPT method. However, a subsequent single case study reported by Freed *et al.* (1997) demonstrated positive retention effects for PROMPT using a well-controlled multiple baseline across behaviours design. This study was not designed to examine transfer across responses and stimuli. Of the remaining seven reports, the majority employed drill type activities with high numbers of trials. No studies systematically tested the effects of principles of motor learning on acquisition of treated skills and retention and transfer of treatment effects.

Since the review of Wambaugh and Doyle (1994), Wambaugh *et al.* (1998) have published a treatment efficacy study using a minimal pairs treatment approach. They employed multiple baseline designs across speakers and behaviours targeting one minimal pair phoneme contrast (e.g. /S/-/s/), at word level, in each phase of treatment. The three subjects acquired all or most of the targeted phonemes in trained words and demonstrated transfer to production of the target phonemes in untrained words. However, some loss of skill was observed at retention testing six weeks post-treatment and transfer to novel responses and novel phrase and sentence level stimuli was limited. This treatment protocol involved presentation of targeted behaviours in blocks with high frequency verbal feedback. It is possible that retention may have been facilitated through application of principles of motor learning such as random ordering of a set of stimuli and less frequent feedback.

Ballard *et al.* (1999) presented the first study to test the influence of principles of motor learning on treatment for AOS. They examined the effect of practising speech skills in random versus blocked order in the context of Phonetic Placement Therapy (Van Riper and Irwin 1958). In this preliminary study, multiple baseline designs across subjects and behaviours and an alternating treatments design were combined to study two subjects with severe AOS and concomitant oral apraxia and aphasia. For one subject, the targeted behaviours were consonant-vowel syllables and, for the other subject, consonant-vowel-consonant words. Both subjects demonstrated less erratic performance during acquisition on behaviours treated in random order compared to those treated in the blocked condition. Consistent with prediction, retention performance was superior for the behaviours trained in the random practice condition. Neither subject demonstrated transfer to untrained responses. However, the first subject transferred the treatment effects

to an untrained stimulus (i.e. a different probe task), with retention of transfer effects post-treatment being greatest for those behaviours treated in the random condition. While this study represents preliminary data, results are positive and warrant continued work in this area.

Summary and Conclusions

Significant advances have been made in recent years toward understanding the nature of the motor speech disorder of AOS. Acoustic, kinematic, and perceptual studies of speech have led to more refined definitions of the behavioural characteristics for use in differential diagnosis. New theoretical accounts have been proposed that are embedded within well-known models of speech and language production. Development and application of methods for studying nonspeech motor control are providing new insights, indicating that the motor control impairment of AOS is more pervasive that previously thought. We have provided a broader account of AOS to accommodate these data and claim that the disorder involves a fundamental impairment of praxis in the articulatory motor system that crosses both speech and nonspeech motor tasks.

Many theoretical and analytical developments in this field have not yet been incorporated into treatment approaches for AOS. Furthermore, few treatment studies have taken advantage of the extensive body of literature supporting the use of specific principles that maximise learning of motor skills in other systems. The time is at hand to re-examine AOS with the goal of modifying existing approaches to treatment, developing new approaches, and rigorously comparing the efficacy of these manipulations. Current theoretical accounts now provide a stronger framework for setting specific treatment goals and examining transfer to related responses and stimuli. Treatment programmes that are driven by a principled account of how the motor system learns should provide the most efficient and effective framework for treating motor-based speech disorders. In turn, well-controlled and theoretically motivated studies of treatment efficacy promise to stimulate development of these theoretical accounts and contribute to our understanding of this enigmatic disorder.

Acknowledgements

This research was supported in part by the National Institutes of Health NIDCD grant P60 DC00976.

References

BADDELEY, A. D. and HITCH, G. J. (1974) "Working memory". In G. Bower (Ed.) *Recent Advances in Learning and Motivation Vol. 7* (Academic Press, New York), pp. 47-90.

BALLARD, K. J., ROBIN, D. A., KNOCK, T. L. and SCHMIDT, R. A. (1999) *Influence of frequency of feedback and order of stimulus presentation on treatment for apraxia of speech.* Paper presented at the Clinical Aphasiology Conference, Key West, Florida.

BALLARD, K. J., ZIMBA, L. and ROBIN, D. A. (in prep) *Age-related changes in motor control: articulator visuomotor tracking.* University of Iowa, Iowa City, IA.

BARLOW, S. and ABBS, J. (1986) Fine force and position control of select orofacial structures in the upper motor neuron syndrome. *Experimental Neurology*, **94**, 699-713.

BAUMAN, J. A. (1978) Sound duration: A comparison between performances of subjects with central nervous system disorders and normal speakers. Unpublished PhD dissertation, University of Colorado.

BLUMSTEIN, S. (1981) Phonological aspects of aphasia. In M. T. Sarno (Ed.) *Acquired Aphasia* (Academic Press, New York), pp. 129-155.

BOCK, J. K. (1982) Towards a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, **89**, 1-47.

BOCK, J. K. and IRWIN, D. E. (1980) Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, 19, 467-484.

BOLINGER, D. L. (1976) Length, vowel, juncture. *Bilingual Review*, **3**, 43-61.

BOYCE, S. E., KRAKOW, R. A. and BELL-BERTI, R. (1991) Phonological underspecification and speech motor organisation. *Phonology*, **8**, 219-236.

BROE, M. (1993) Specification theory: The treatment of redundancy in generative phonology. Unpublished PhD dissertation, University of Edinburgh.

BROWN, J. W. (1975) The problem of repetition: A study of "conduction" aphasia and the "isolation" syndrome. *Cortex*, **11**, 37-52.

BUCKINGHAM, H. (1979) Explanation in apraxia with consequences for the concept of apraxia of speech. *Brain and Language*, **8**, 202-226.

CHUMPELIK, D. (1984) The PROMPT system of therapy: Theoretical framework and applications for developmental apraxia of speech. *Seminars in Speech and Language*, **5**, 139-156.

CLARK, H. and ROBIN, D. A. (1998) Generalised motor programme and parameterisation accuracy in apraxia of speech and conduction aphasia. *Aphasiology*, **12**, 699-713.

CODE, C. (1998) Models, theories, and heuristics in apraxia of speech. *Clinical Linguistics and Phonetics*, **12**, 47-65.

CODE, C. and BALL, M. J. (1982) Fricative production in Broca's aphasia: A spectrographic analysis. *Journal of Phonetics*, **10**, 325-331.

COHN, A. C. (1990) Phonetic and phonological rules of nasalization. Unpublished doctoral dissertation, University of California at Los Angeles.

COHN, A. C. (1993) Nasalization in English: Phonology and phonetics. *Phonology*, **10**, 43-81.

COLLINS, M. J., ROSENBEK, J. C. and WERTZ, R. T. (1983) Spectrographic analysis of vowel and word duration in apraxia of speech. *Journal of Speech and Hearing Research*, **26**, 224-230.

COOPER, W.E., LAPOINTE, S.G. and PACCIA, J. M. (1977) Syntactic blocking of phonological rules in speech production. *Journal of the Acoustical Society of America*, **61**, 1314-1320.

DARLEY, F. L., ARONSON, A. E. and BROWN, J. R. (1975) *Motor speech disorders*. (W. B. Saunders Company, Philadelphia).

DERENZI, E., PIECZURO, A. and VIGNOLO, L. A. (1966) Oral apraxia and aphasia. *Cortex*, **2**, 50-73.

DISIMONI, F. G. (1974a) Influence of vowel environment on the duration of consonants in the speech of three-, six- and nine-year old children. *Journal of the Acoustical Society of America*, **55**, 360-361.

DISIMONI, F. G. (1974b) Some preliminary observations on temporal compensation in the speech of children. *Journal of the Acoustical Society of America*, **56**, 697-699.

DOGIL, G. and MAYER, J. Selective phonological impairment: A case of apraxia of speech. Unpublished manuscript, University of Stuttgart.

DOGIL, G., MAYER, J. and VOLLMER, K. (1994) *A representational account for apraxia of speech*. Paper presented at the Fourth Symposium of the International Clinical Phonetics and Linguistics Association, New Orleans, Louisiana.

DUFFY, J. R. (1995) *Motor Speech Disorders: Substrates, Differential Diagnosis and Management*. (Mosby, St. Louis).

DWORKIN, J. P., ABKARIAN, G. G. and JOHNS, D. F. (1988) Apraxia of speech: The effectiveness of a treatment regimen. *Journal of Speech and Hearing Disorders*, **53**, 280-294.

EASTON, T. A. (1972) On the normal use of reflexes. *American Scientist*, **60**, 591-599.

FLOWERS, K. (1978) Some frequency response characteristics of Parkinsonism on pursuit tracking. *Brain*, **101**, 19-34.

FOLKINS, J. W. (1985) Issues in speech motor control and their relation to the speech of individuals with cleft palate. *Cleft Palate Journal*, **22**, 106-122.

FOLKINS, J. W. and BLIELE, (1990) Taxonomies in biology, phonetics, phonology, and speech motor control. *Journal of Speech and Hearing Disorders*, **55**, 596-611.

FOWLER, C. (1977) *Timing control in speech production*. Indiana University Linguistics Club, Bloomington.

FREED, D. B., MARSHALL, R. C. and FRAZIER, K.E. (1997) Long-term effectiveness of PROMPT treatment in a severely apractic-aphasic speaker. *Aphasiology*, **11**, 365-372.

FREEMAN, F., SANDS, E. and HARRIS, K. S. (1978) Temporal co-ordination of phonation and articulation in a case of verbal apraxia: A voice onset time study. *Brain and Language*, **6**, 106-111.

FRIEDRICH, F. J., GLENN, C. G. and MARIN, O. S. M. (1984) Interruption of phonological coding in conduction aphasia. *Brain and Language*, **22**, 266-291.

FRY, D. B. (1958). Phonemic substitutions in an aphasic patient. *Language and Speech*, **1**, 52-61.

GARRETT, M. F. (1975) The analysis of sentence production. In G. H. Bower (Ed.) *The Psychology of Learning and Motivation Vol. 9*. (Academic Press, New York), pp. 133-177.

HAGEMAN, C. F., ROBIN, D. A., MOON, J. B. and FOLKINS, J. W. (1993) Visuomotor tracking in neurogenic disorders. Paper presented at the American Speech-Language-Hearing Association Convention, San Antonio, Texas.

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.

ITOH, M., SASANUMA, S. and USHIJIMA, T. (1979a) Velar movements during speech in a patient with apraxia of speech. *Brain and Language*, **7**, 227-239.

ITOH, M., SASANUMA, S., TATSUMI, I. and KOBAYASHI, Y. (1979b) Voice onset time characteristics of apraxia of speech. *Annual Bulletin No. 13 Research Institute of Logopedics and Phoniatrics, University of Tokyo*, pp. 123-132.

JANSSEN, P. and WIENEKE, G. (1987) The effects of fluency inducing conditions on the variability of in the duration of laryngeal movements during stutterers' fluent speech. In H. F. M. Peters and W. Hulstijn (Eds) *Speech Motor Dynamics in Stuttering* (Springer-Verlag, Wien), pp. 337-344.

KEATING, P. A. (1988) Underspecification in phonetics. *Phonology*, **5**, 275-292.

KEATING, P. A. (1990) The window model of coarticulation: articulatory evidence. In J. Kingston and M. E. Beckman (Eds) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Cambridge University Press, Cambridge), pp. 451-470.

KELSO, J. A. S. (1995) *Dynamic Patterns: The Self-Organization of Brain and Behavior* (MIT Press/Bradford, Cambridge, MA).

KELSO, J. A. S., SOUTHARD, D. L. and GOODMAN, D. (1979) On the nature of human interlimb coordination. *Science*, **203**, 1029-1031.

KELSO, J. A. S. and TULLER, B. (1981) Toward a theory of apractic syndromes. *Brain and Language*, **12**, 224-245.

KENT, R. D. and ADAMS, S. G. (1989) The concept and measurement of coordination in speech disorders. In S. A. Wallace (Ed.) *Perspectives on the Coordination of Movement* (Elsevier Science Publishers B. V., North-Holland), pp. 415-451.

KENT, R. D. and FORNER, L. L. (1980) Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, **8**, 157-168.

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 Press, 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.

KLATT, D. H. (1976) Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208-1221.

KOHN, S. (1984) The nature of the phonological disorder in conduction aphasia. *Brain and Language*, **23**, 97-115.

KONNO, K., SUGISHITA, M. and HIROSE, H. (1987) Articulatory movements in apraxia of speech. *Annual Bulletin RILP*, **21**, 177-191.

LANE, V. W. and SAMPLES, J. M. (1981) Facilitating communication skills in adult apraxics: Application of blissymbols in a group setting. *Journal of Communication Disorders*, **14**, 157-167.

LAPOINTE, L. L. (1984) Sequential treatment of split lists: A case report. In J. Rosenbek, M. McNeil and A. Aronson (Eds) *Apraxia of Speech: Physiology, Acoustics, Linguistics, Management* (College-Hill, San Diego) pp. 277-286.

LEHISTE, I. (1972) The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, **51**, 2018-2024.

LEVELT, W. J. M. (1989) *Speaking: From Intention to Articulation* (MIT Press, Cambridge, MA).

LIEPMANN, H. (1900) Das Krankheitsbild der apraxia (motorischen asymboli) auf Grund eines Falles von einseitiger apraxie. *Monatschrift Psychiatrie Neurologie*, **9**, 15-40.

LIEPMANN, H. (1905) Die linke hemisphaere und das handeln. *Muchener medizinische Wochenschrift*, **52**, 2322-2326, 2375-2378.

LIEPMANN, H. (1913) Motor aphasia, anarthria, and apraxia. *Transactions of the 17th International Congress of Medicine* (Section XI, Part II), pp. 97-106.

MACNEILAGE, P. F. (1982) Speech production mechanisms in aphasia. In S. Grillner, B. Lindblom, J. Lubker and A. Persson (Eds) *Speech Motor Control* (Pergamon Press, Oxford), pp. 43-60.

MAYER, J. A. (1995) A representational account for apraxia of speech. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, **1**, 82-85.

MCCLEAN, M. D., BEUKELMAN, D. R. and YORKSTON, K. M. (1987) Speech-muscle visuomotor tracking in dysarthric and nonimpaired speakers. *Journal of Speech and Hearing Research*, **30**, 276-282.

MCNEIL, M. R. and ADAMS, S. (1991) A comparison of speech kinematics among apraxic, conduction aphasic, dysarthric, and normal geriatric speakers. *Clinical Aphasiology*, **19**, 279-294.

MCNEIL, M. R., CALIGUIRI, M. and ROSENBEK, J. C. (1989) A comparison of speech kinematics among apraxic, conduction aphasic, ataxic dysarthric and normal geriatric speakers. *Clinical Aphasiology*, **17**, 173-193.

MCNEIL, M. R., HASHI, M. and SOUTHWOOD, H. (1994) Acoustically derived perceptual evidence for coarticulatory deficits in conduction aphasia and apraxia of speech. *Clinical Aphasiology*, **22**, 203-218.

MCNEIL, M. R. and KENT, R. D. (1990) Motoric characteristics of adult aphasic and apraxic speakers. In G. E. Hammond (Ed.) *Cerebral Control of Speech and Limb Movements* (Elsevier Science Publishers, Amsterdam), pp. 349-386.

MCNEIL, M. R., LISS, J. M., TSENG, C. and KENT, R. D. (1990a) Effects of speech rate on the absolute and relative timing of apraxic and conduction aphasic sentence production. *Brain and Language*, **38**, 135-158.

MCNEIL, M. R., ODELL, K. H., MILLER, S. B. and HUNTER, L. (1995) Consistency, variability, and target approximation for successive speech repetitions among apraxic, conduction aphasic, and ataxic dysarthric speakers. *Clinical Aphasiology*, **23**, 39-55.

MCNEIL, M. R., ROBIN, D. A. and SCHMIDT, R. A. (1997) Apraxia of speech: Definition, differentation, and treatment. In M. R. McNeil (Ed.) *Clinical Management of Sensorimotor Speech Disorders* (Thieme, New York), pp. 311-344.

MCNEIL, M. R., WEISMER, G., ADAMS, S. and MULLIGAN, M. (1990b) Oral structure nonspeech motor control in normal, dysarthric, and apraxic speakers: Isometric force and static position control. *Journal of Speech and Hearing Research*, **33**, 255-268.

MERCAITIS, P. A. (1983) Some temporal characteristics of imitative speech in non-brain-injured, aphasic, and apraxic adults (Dissertation Abstracts International, Ann Arbor, MI).

MLCOCH, A. G. and NOLL, J. D. (1980) Speech production models as related to the concept of apraxia of speech. In N. J. Lass (Ed.) *Speech and Language: Advances in Basic Research and Practice Vol. 4.* (Academic Press, New York), pp. 201-238.

MOHANAN, K. P. (1991) On the bases of radical underspecification. *NLLT*, **9**, 285-325.

MOON, J. B., ZEBROWSKI, P., ROBIN, D. A. and FOLKINS, J. W. (1993) Visuomotor tracking ability of young adult speakers. *Journal of Speech and Hearing Research*, **36**, 672-682.

ODELL, K., MCNEIL, M. R., ROSENBEK, J. C. and HUNTER, L. (1990) Perceptual characteristics of consonant production by apraxic speakers. *Journal of Speech and Hearing Disorders*, **55**, 345-359.

ODELL, K., MCNEIL, M. R., ROSENBEK, J. C. and HUNTER, L. (1991) Perceptual characteristics of vowel and prosody production in apraxic, aphasic, and dysarthric speakers. *Journal of Speech and Hearing Research*, **34**, 67-80.

PIERREHUMBERT, J. and BECKMAN, M. (1988) Japanese tone structure. *LI Monograph Series No. 15* (MIT Press, Cambridge, MA).

POULTON, E. C. (1974) *Tracking skill and manual control.* (Academic Press, New York).

RAYMER, A. M. AND THOMPSON, C. K. (1991) Effects of verbal plus gestural treatment in a patient with aphasia and severe apraxia of speech. In T.E. Prescott (Ed.) *Clinical Aphasiology Vol. 20* (Pro-Ed, Austin, TX), pp. 285-298.

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., MOON, J. B., CLARK, H. C., WOODWORTH, G. and FOLKINS, J. W. (in prep.) *Visuomotor tracking abilities of speakers with apraxia of speech or conduction aphasia.* University of Iowa, Iowa City, IA.

ROBIN, D. A., SOLOMON, N. P., MOON, J. B. and FOLKINS, J. W. (1997) Nonspeech assessment of the speech production mechanism. In M. R. McNeil (Ed.) *Clinical Management of Sensorimotor Speech Disorders* (Thieme, New York), pp. 49-62.

ROCHON, E., CAPLAN, D. and WATERS, G. (1990) Short-term memory processes in patients with apraxia of speech: Implications for the nature and structure of the auditory verbal short-term memory system. *Journal of Neurolinguistics*, **5**, 237-264.

ROGERS, M. A. and STORKEL, H. L. (1998) Reprogramming phonologically similar utterances: The role of phonetic features in pre-motor encoding. *Journal of Speech, Language, and Hearing Research*, **42**, 258-274.

ROGERS, M. A. and STORKEL, H. L. (in press) Planning speech one syllable at a time: The reduced buffer capacity hypothesis in apraxia of speech. *Aphasiology*.

ROSENBEK, J. C., LEMME, M. L., AHERN, M. B., HARRIS, E. H. and WERTZ, R. T. (1973) A treatment for apraxia of speech in adults. *Journal of Speech and Hearing Disorders*, **38**, 462-472.

ROY, E. A. (1978) Apraxia: A new look at an old syndrome. *Journal of Human Movement Studies*, **4**, 191-210.

RYALLS, J. H. (1981) Motor aphasia: Acoustic correlates of phonetic disintegration of vowels. *Neuropsychologia*, **19**, 365-374.

RYALLS, J. H. (1987) Vowel production in aphasia: Towards an account of the consonant vowel dissociation. In J. H. Ryalls (Ed.) *Phonetic Approaches to Speech Production in Aphasia and Related Disorders* (Little Brown and Co., Boston, MA).

SALTZMAN, E. (1986) Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer and C. Fromm (Eds) *Generation and Modulation of Action Patterns* (Springer-Verlag, Berlin), pp. 129-144.

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

SCHMIDT, R. A. and BJORK, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, **3**, 207-217.

SCHMIDT, R. A. and LEE, T. D. (1999) *Motor Control and Learning: A Behavioral Emphasis* 3rd Ed (Human Kinetics, Champaign, IL).

SEDDOH, S., ROBIN, D. A., SIM, H-S., HAGEMAN, C., MOON, J. B. and FOLKINS, J. W. (1996) Speech timing in apraxia of speech versus conduction aphasia. *Journal of Speech and Hearing Research*, **39**, 590-603.

SHARKEY, S. G. and FOLKINS, J. W. (1985) Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control. *Journal of Speech and Hearing Research*, **28**, 8-15.

SHRIBERG, L. D., ARAM, D. M. and KWIATKOWSKI, J. (1997a) Developmental apraxia of speech: I. Descriptive and theoretical perspectives. *Journal of Speech, Language, and Hearing Research*, **40**, 273-285.

SHRIBERG, L. D., ARAM, D. M. and KWIATKOWSKI, J. (1997b) Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research*, **40**, 286-312.

SHRIBERG, L. D., ARAM, D. M. and KWIATKOWSKI, J. (1997c) Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, **40**, 313-337.

SIMMONS, N. N. (1980) Choice of stimulus modes in treating apraxia of speech: A case study. In R. H. Brookshire (Ed.) *Clinical Aphasiology Conference Proceedings* (BRK, Minneapolis), pp. 302-307.

SMITH, A., GOFFMAN, L., ZELAZNIK, H. N., YING, G. and MCGILLEM, C. (1995) Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research*, **104**, 493-501.

SMITH, A. and DENNY, M. (1990) High frequency oscillations as indicators of neural control mechanisms in human respiration, mastication, and speech. *Journal of Neurophysiology*, **63**, 745-758.

SMITH, B. L. (1992) Relationships between duration and temporal variability in children's speech. *Journal of the Acoustical Society of America*, **91**, 2165-2174.

SMITH, B. L. (1994) Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adult speech. *Journal of Phonetics*, **22**, 155-175.

SMITH, B. L. and KENNEY, M. K. (1994) Variability control in speech production tasks performed by adults and children. *Journal of the Acoustical Society of America*, **96**, 699-705.

SOUTHWOOD, H. (1987) The use of prolonged speech in the treatment of apraxia of speech. In R. H. Brookshire (Ed.) *Clinical Aphasiology Conference Proceedings* (BRK, Minneapolis), pp. 277-287.

SPARKS, R., HELM, N. A. and ALBERT, M. (1974) Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex*, **10**, 303-316.

SQUARE, P. A., CHUMPELIK, D. and ADAMS, S. (1985) Efficacy of the PROMPT system of therapy for the treatment of acquired apraxia of speech. In R. H. Brookshire (Ed.) *Clinical Aphasiology Conference Proceedings* (BRK, Minneapolis), pp. 319-320.

SQUARE, P. A., CHUMPELIK, D. A., MORNINGSTAR, D. and ADAMS, S. (1986) Efficacy of the PROMPT system of therapy for the treatment of acquired apraxia of speech: A follow-up investigation. In R. H. Brookshire (Ed.) *Clinical Aphasiology Conference Proceedings* (BRK, Minneapolis), pp. 221-226.

SQUARE-STORER, P. A. (1989) Traditional therapies for apraxia of speech – reviewed and rationalized. In P. A. Square-Storer (Ed.) *Acquired Apraxia of Speech in Aphasic Adults* (Lawrence Erlbaum, Hove and London), pp. 145-161.

SQUARE-STORER, P. A. and APELDOORN, S. (1991) An acoustic study of apraxia of speech in patients 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 Publishing Co., Baltimore), pp. 271-288.

SQUARE-STORER, P. and HAYDEN, D. C. (1989) PROMPT treatment. In P. Square-Storer (Ed.) *Acquired Apraxia of Speech in Aphasic Adults* (Lawrence Erlbaum, Hove and London), pp. 190-219.

STEMBERGER, J. (1993) Glottal transparency. *Phonology*, **10**, 107-138.

STRAND, E. A. (1987) Acoustic and response time measures in utterance production: A comparison of apraxic and normal speakers. Unpublished doctoral dissertation, University of Wisconsin-Madison.

STRAND, E. A. and MCNEIL, M. R. (1996) Effects of length and linguistic complexity on temporal acoustic measures in apraxia of speech. *Journal of Speech and Hearing Research*, **39**, 1018-1033.

THELEN, E. and SMITH, L. B. (1994) *A Dynamic Systems Approach to the Development of Cognition and Action* (MIT Press/Bradford, Cambridge, MA).

THOMPSON, C. K. and YOUNG, E C. (1983) *A phonological process approach to apraxia of speech: An experimental analysis of cluster reduction.* Paper presented at the American Speech and Hearing Association Convention, Cincinnati, Ohio.

TINGLEY, B. and ALLEN, G. (1975) Development of speech timing control in children. *Child Development*, **46**, 186-194.

TRUBETZKOY, N. (1939) *Principles of Phonology.* (University of California Press, Los Angeles, CA).

TURVEY, M. T. (1977) Preliminaries to a theory of action with reference to vision. In R. Shaw and J. Bransford (Eds) *Perceiving, Acting, and Knowing: Toward an Ecological Psychology* (Erlbaum, Hillsdale), pp. 211-265.

TURVEY, M. T., SHAW, R. E. and MACE, W. (1978) Issues in the theory of action. In J. Requin (Ed.) *Attention and Performance Vol. VII* (Erlbaum, Hillsdale, NJ).

Van der Merwe , A. (1997) A theoretical framework for the characterization of pathological speech sensorimotor control. In M. R. McNeil (Ed.) *Clinical Management of Sensorimotor Speech Disorders* (Thieme, New York), pp. 1-25.

VAN RIPER, C. and IRWIN, J. (1958) *Voice and Articulation* (Prentice Hall, Englewood Cliffs).

VOLLMER, K. (1997) Koartikulation und glottale Transparenz. Unpublished doctoral dissertation, University of Stuttgart. Phonetik-AIMS 3.5 (Working papers of the Chair of Experimental Phonetics, University of Stuttgart)

WAMBAUGH, J. L. and DOYLE, P. J. (1994). Treatment for acquired apraxia of speech: A review of efficacy reports. *Clinical Aphasiology*, **22**, 231-243.

WAMBAUGH, J. L., DOYLE, P.J., KALINYAK, M. M. and WEST, J. E. (1996) A minimal contrast treatment for apraxia of speech. *Clinical Aphasiology*, **24**, 97-108.

WAMBAUGH, J. L., KALINYAK-FLISZAR, M. M., WEST, J. E. and DOYLE, P. J. (1998). Effects of treatment for sound errors in apraxia of speech and aphasia. *Journal of Speech, Language, and Hearing Research*, **41**, 725-743.

WARREN, R. L. (1977) Rehearsal for naming in apraxia of speech. In R. H. Brookshire (Ed.) *Clinical Aphasiology Conference Proceedings* (BRK, Minneapolis), pp. 80-90.

WHITESIDE, S. P. and VARLEY, R. A. (1998) A reconceptualisation of apraxia of speech: A synthesis of evidence. *Cortex*, **34**, 221-231.

WIENEKE, G. and JANSSEN, P. (1987) Effect of speaking rate on speech timing variability. In H. F. M. Peters and W. H. Hulstijn (Eds) *Speech Motor Dynamics in Stuttering* (Springer-Verlag, Wien), pp. 345-352.

YAMADORI, A. and IKUMURA, G. (1975) Central (or conduction) aphasia in a Japanese patient. *Cortex*, **11**, 73-82.

ZIEGLER, W. D. and VON CRAMON, D. (1985) Anticipatory coarticulation in a patient with apraxia of speech. *Brain and Language*, **26**, 117-130.

ZIEGLER, W. D. and VON CRAMON, D. (1986) Disturbed coarticulation in apraxia of speech: Acoustic evidence. *Brain and Language*, **29**, 34-47.

ZSIGA, E. C. (1997) Features, gestures, and Igbo vowels: An approach to the phonology-phonetics interface. *Language*, **73**, 227-274.

NCVS Status and Progress Report - 14
September 1999, 191-201

Personality Traits and Psychological Factors in Voice Pathology: A Foundation for Future Research

Nelson Roy, Ph.D.
Department of Communication Disorders
Division of Otolaryngology-Head & Neck Surgery, The University of Utah
Diane Bless, Ph.D.
Department of Communicative Disorders
Division of Otolaryngology-Head & Neck Surgery, The University of Wisconsin-Madison

Abstract

It has been argued that personality, emotions, and psychological problems contribute to or are primary causes of voice disorders, and that voice disorders in turn create psychological problems and personality effects. This article (1) briefly reviews the literature surrounding the role of psychological and personality processes in individuals with functional dysphonia (FD), vocal nodules (VN), and spasmodic dysphonia (SD); (2) provides an overview of recent concepts in personality and trait structure; and (3) summarizes the fundamental tenets of a theoretical synthesis proposed by Roy and Bless (in press) to explain the dispositional bases of FD and VN. This theory links FD and VN to the signal sensitivities and behavioral response biases of neurotic introverts and neurotic extraverts, respectively. In a companion article (this issue), the merits of the Roy and Bless theory are evaluated.

The human voice has long been regarded as a "barometer of emotion" and a "mirror of personality" implying that the voice reflects individual differences in emotional state and personality disposition (Aronson, 1990; Diehl, 1960). The early association between voice and personality is revealed in the derivation of the word "personality." Moses (1954) explained that the term first originated from the Latin *persona*, which meant the mouthpiece of a mask worn by actors (per sona: the sound of the voice passes through). The term's meaning changed from the mask to the actor, the "person" in a theatrical production. The word eventually evolved to represent any person and ultimately "personality," but over the years the symbolic connection with the voice was lost. Almost 50 years has passed since Moses published his classic text describing the complex interplay of emotional dynamics, personality, and voice production; however, little progress has been made in our understanding of the relation between personality, voice, and voice disorders.

Despite few advances, voice clinicians continue to endorse the voice-psychology association by offering emotional or personality factors as possible causal explanations when the voice becomes disordered (Aronson, 1990; Colton & Casper, 1996; Stemple, 1993). The debate over the role of personality or psychological processes is most controversial in the case of such disorders as functional dysphonia (FD), vocal nodules (VN), and spasmodic dysphonia (SD). It is unclear whether personality and emotional maladjustment contribute to or are primary causes of these voice disorders, or alternatively, whether these voice disorders create psychological problems or personality effects (Cooper, 1973). Today many of the same questions that existed in Moses' time remain unanswered: Do personality or psychological differences exist among voice-disordered groups? If so, which personality factors might be causally significant?

This article briefly reviews the literature surrounding the role of psychological and personality processes in individuals with these voice disorders, and summarizes a theory that identifies personality as a contributing factor in the development of FD and VN. Finally, in a companion article, data are provided that evaluate the predictions and

fundamental tenets of the theory at a superfactor trait level of personality description.

Personality and Vulnerability to Voice Disorders and Illness

The notion that personality characteristics might influence vulnerability to illness and illness progression has already received much attention in the health psychology literature (Akistal, Hirshfeld, & Yerevanian, 1983; Smith & Williams, 1992; Stone & Costa, 1990; Suls & Rittenhouse, 1990). It has been argued that different personality traits might predispose one to certain diseases/disorders, influence symptomatology and course, and in turn be affected by the experience of illness (Contrada, Leventhal, O'Leary, 1990; Holroyd & Coyne, 1987). These models fall into two general categories. The first category includes a range of predisposition or vulnerability models, in which personality characteristics are proposed either to directly cause the development of the disorder, or to indirectly modify the course or expression of the disorder. The second category, referred to as disability (scar) hypothesis models, asserts that certain disorders affect personality; that is, the experience of the disorder/illness causes personality changes. These changes may represent either unresolved residual symptoms of the disorder or relatively stable post episode character adaptations (Akistal et al., 1983). These models, which relate personality and illness, provide a useful context to seat the companion investigation, which aims to explore the role of personality and psychological processes in select voice pathologies.

Limitations of Existing Research

The voice literature is replete with speculations linking FD, VN, and SD to psychological precursors and personality variables. However, there is a scarcity of objective research and the majority of writings on this topic are based primarily on anecdote and clinical impressions.

The existing research is generally disappointing for a variety of reasons and has been critiqued previously in reviews by Roy and Bless (in press) and Green (1988). In short, most studies mix voice disorder types and sexes into a single group, rendering interpretation of the results difficult, if not impossible. Furthermore, investigators often neglect to compare their findings with other voice-disorder groups. This hinders interpretations of commonality versus specificity (Deary, Scott, Wilson, White, MacKenzie, Wilson, 1997). Thus, it is unknown whether the results are unique to the subgroup of patients studied or applicable to voice-disordered patients in general. Finally, information is not provided regarding the severity of vocal handicap or duration of the vocal symptoms. Therefore it cannot be ascertained whether psychological or personality differences merely represent the outcome of the vocal disability.

These methodological inadequacies make it very difficult to generalize the results and evaluate the specific nature of the voice disorder-personality relationship. There is currently no clear evidence of whether personality or psychological processes should be considered causal, correlational, or consequential. With these caveats in mind, the following sections will summarize empirical research findings as they relate to FD, VN, and SD. Despite methodological differences and shortcomings, some interesting trends do surface.

Functional Dysphonia

Functional dysphonia refers to a voice disorder in the absence of identifiable neurological or structural pathology (Koufman & Blalock, 1982; Morrison, Nichol, & Rammage, 1986). Some controversy surrounds the term "functional dysphonia," and confusion exists because FD is often used as the general descriptive term for a host of medically unexplained voice disorders. Functional dysphonia is sometimes broadly synonymous with "hysterical," "psychogenic," "conversion,", "psychosomatic," "hyperkinetic," "hyperfunctional," "muscle misuse," or "muscle tension" dysphonia. As is obvious by the range and variety of labels, theorists' opinions differ concerning the relative contribution of psychological factors to the formation of functional voice disorders (Cooper, 1973).

Functional dysphonia, which may account for more than 10% of cases referred to multidisciplinary voice clinics (Schalen & Andersson, 1992), occurs predominantly in women, and like many other voice disorders, commonly follows upper respiratory infection symptoms (Aronson, Peterson & Litin, 1966; Friedl, Friedrich & Egger, 1990; Gerritsma, 1991; Kinzl, Biebl & Rauchegger, 1988; Milutinovic, 1991). The disorder is frequently transient and varies in its response to treatment (Bridger & Epstein, 1983; Fex, F., Fex, S., Shiromoto & Hirano, 1994; Koufman & Blalock, 1982; Roy & Leeper, 1993). Functional *dysphonia* and *aphonia* are sometimes regarded as disorders represented on a continuum of 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). Most studies investigating personality and/or psychological processes group both disorders under the designation "psychogenic voice disorder," reflecting the etiological supposition. Some authors caution 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).

Psychological Mechanisms in FD

Diverse psychopathological processes contributing to voice symptom formation in FD have been proposed. An extensive review of all possible mechanisms is beyond the scope of this paper; thus, only a brief overview is provided. The interested reader is referred to Roy & Bless (in press) for a more complete exploration of the putative psychological and personality processes involved in FD.

The dominant psychological explanation for medically unexplained voice loss is the concept of conversion disorder introduced by Freud (Aronson, 1990; Butcher, 1995; Greene & 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 (American Psychiatric Association (APA), DSM IV, 1994). The conversion symptom represents an unconscious simulation of illness that apparently prevents conscious awareness of emotional conflict or stress, thereby displacing the mental conflict and reducing anxiety. When the laryngeal system is involved, it is referred to as conversion dysphonia or aphonia.

Butcher and colleagues (Butcher, Elias, Raven, Yeatman & Littlejohns, 1987; Butcher, Elias, Raven, 1993; Butcher, 1995) argue against conversion disorder as the most common cause of dysphonia unaccounted for by pathological findings. In the place of conversion, Butcher (1995) offers two alternative psychological models to account for partial or complete functional voice loss. The first model is a slightly reformulated psychoanalytic explanation, while the second is based upon cognitive-behavioral principles. Despite somewhat different causal pathways, both models clearly emphasize the inhibitory effects of excess laryngeal musculoskeletal tension on voice production.

The subject of poorly regulated laryngeal muscle tension is also a theme in the writings of Morrison and colleagues (Morrison & Rammage, 1993; Rammage, Nichol & Morrison, 1987) and Aronson (1990) among others (Colton & Casper, 1996; Greene & Mathieson, 1989). Nichol, Morrison & Rammage (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. Rammage and associates (1987) also proposed that a relatively minor organic change, such as edema, infection, or reflux laryngitis, might trigger functional misuse, particularly if the individual is exceedingly anxious regarding his or her voice or health. The same authors speculated 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 personality and psychological factors as strongly influential in the development of FD, they have virtually ignored the possibility that such processes could be the negative consequence of coping with an incapacitating voice disorder (i.e., the disability hypothesis). Because voice problems can be associated with a number of adverse consequences, including laryngeal discomfort, fatigue, and impairment of social and/or occupational functioning (with a concomitant loss of self-esteem and social support), it is not unreasonable to postulate that chronic voice problems might lead to general personality changes, such as the development of heightened feelings of distress and dissatisfaction and social withdrawal. Depression, anxiety, and tension are frequent psychological concomitants of chronic illness (Dubovsky & Weissberg, 1982; Nemiah, 1961; Reiser, 1980). The notion that such sequelae could be considered outcomes of a severe voice disturbance, rather than causal agents, has received little research attention.

Assessment of Psychological Processes in FD

Empirical evidence to support the various psychological mechanisms offered to explain FD has seldom been provided. It should be recognized that many of these theories are untestable from a scientific perspective. Only a few studies exist that have used standardized instruments to assess the personality or psychological characteristics of patients with FD. A complete review of the relevant findings and interpretations is provided in Roy, McGrory, Tasko, Bless, Heisey & Ford (1997). As Roy and colleagues indicate, direct comparison of the observations is restricted because of significant methodological differences. Such differences might account for the disparate results regarding the frequency and degree of hysterical personality traits (Aronson et al., 1966; Gerritsma, 1991; Kinzl et al., 1988), conversion reaction (House & Andrews, 1987; Pfau, 1975), and psychopathological symptoms (Aronson et al., 1966; Gerritsma, 1991; House & Andrews, 1987; Kinzl et al., 1988; Pfau, 1975). Despite their methodological differences/inadequacies, these studies have identified a trend toward elevated levels of (1) anxiety, (2) somatic complaints, and (3) introversion in the FD population. Patients have been described as socially anxious, nonassertive, and with a tendency toward restraint (Friedl et al., 1990; Gerritsma ,1991). None of these researchers have attempted to explicitly integrate their findings into a coherent theory of personality as a contributing factor for FD.

In a recent study, Roy and colleagues (1997) described the personality/psychological characteristics of female subjects with 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 general medical outpatient control group, the FD subjects scored significantly higher on 7 of 10 clinical scales, suggesting an elevated degree of emotional maladjustment. The FD subjects were described as depressed, anxious, somatically preoccupied, and introverted. The results suggested that in spite of symptom improvement after voice therapy, these subjects continued to exhibit poor levels of adaptive functioning. With the exception of the Depression scale, all clinical scales were viewed as assessments of character, not mood (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989; Duckworth & Anderson, 1995; Graham, 1987; Graham, 1990; Marks, Seeman, & Haller, 1974; Newmark, 1979). Therefore, the authors interpreted the data to support a dispositional (trait-like) vulnerability for the development of functional symptoms, including laryngeal problems.

Vocal Nodules

Vocal nodules are benign callous-like lesions of the vocal folds often attributed to chronic, repetitive phonotrauma producing biomechanical tissue stresses and reactive histological changes. Vocal nodules are considered to be a common manifestation of vocal hyperfunction, that is, abuse and/or misuse of the vocal mechanism due to excessive and/or "imbalanced" muscular forces (Hillman, Holmberg, Perkell, Walsh & Vaughan, 1989, 1990), and may account for almost 4% of an otolaryngology caseload (Nagata, Kurita, Yasumoto, Maeda, Kawasaki, & Hirano, 1983). In adults, at least two-thirds of patients with VN 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 supposed cause(s) of the vocal nodule rather than the nodule itself. The short-term results of behavioral treatment programs or surgical excision are generally favorable (Bouchayer & Cornut, 1988; Lancer, Syder, Jones & Le Boutillier, 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 (Bridger & Epstein, 1983). One factor that may interfere with successful treatment is poor extraclinical compliance with therapy recommendations (Verdolini-Marston, Burke, Lessac, Glaze, & Caldwell, 1995).

For the most part, authors have attempted to distinguish VN from other mass lesions and other voice disorders, such as FD. However, some clinicians have classified VN as a functional disorder and have emphasized the role of psychological precursors and predisposing personality factors (Arnold, 1962; Aronson, 1990; Wilson, 1987). One common view is that people with vocal nodules are talkative and have aggressive tendencies (Arnold, 1962; Aronson, 1990; Green, 1989; Mosby, 1970; Nemec, 1961; Toohill, 1975; Wilson, 1971; Wilson and Lamb, 1974; Withers & Dawson, 1960). Elevated levels of anxiety, emotional reactivity, and maladjustment, as well as high levels of extraversion have been found among patients with VN (Mosby, 1970; Peter & Brandell, 1980; Toohill, 1975; Yano, Ichimura, Hoshino & Nozue, 1982).

In a recent study using the MMPI, Roy, McGrory & Bless (1995) identified elevated levels of psychological distress and somatic complaints in a group of adult female VN patients when compared to a medical outpatient control group. Goldman, Hargrave, Hillman, Holmberg and Gress (1996) confirmed these findings when they also identified elevated levels of anxiety, somatic complaints, and voice use among VN patients when compared to non-voice-disordered controls. No differences, however, were identified between the VN subjects and a voice-disordered control group—who were free of mucosal disease.

Recently, White, Deary and Wilson (1997), using the General Health Questionnaire (GHQ) and Eysenck Personality Questionnaire (EPQ), found no significant differences in personality traits when comparing dysphonic patients (both functional and organic) with outpatient otolaryngology controls. They did, however, identify elevated levels of psychological distress in both voice-disordered groups and concluded that it was impossible to identify those dysphonia patients with a major underlying psychological upset based solely on the 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 VN. Further research is necessary to better appreciate the relationship between personality and VN development and maintenance.

Spasmodic Dysphonia

Spasmodic dysphonia is a poorly understood voice disorder characterized by intermittent voice arrests and strained-strangled voice quality (Cannito, 1991). The effortful voice spasms fluctuate in severity and may remit for hours or even days at a time. Male to female ratio ranges from 1:1 to 1:4. The intermittency and laryngeal specificity of symptoms have historically led many to invoke a psychological explanation for this enigmatic disorder (Bloch, 1965; Brodnitz, 1962; Heaver, 1960). As reported in Cannito (1991), so universal was this opinion that in his review of the literature, Arnold (1959) wrote that since its original description, "all authors agreed that spastic dysphonia represented a psychoneurotic disorder of pneumophonic coordination (p.4)." In the psychoanalytic tradition, many clinicians viewed SD symptoms as a hysterical conversion reaction whereby intrapsychic conflict was unconsciously

converted into a voice disturbance (Brodnitz, 1962; Heaver, 1960). In a retrospective review of 130 cases of SD seen during his career, Brodnitz (1976) reported that 41% of patients could identify severe emotional trauma preceding the onset of SD, and a further 22% exhibited symptoms consistent with severe neuroses. Descriptions of this sort typically included symptoms of anxiety, depression, and somatic preoccupation.

Despite some controversy (Cooper, 1980), there is near consensus that SD is an action-induced focal laryngeal dystonia, whose onset of symptoms is often related to emotional upset or environmental stresses (Blitzer, Lovelace, Brin, Fahn & Fink, 1985; Finitzo & Freeman, 1989; Ludlow, Hallett, Sedory, Fujita, Naughton, 1990). The psychiatric literature on SD remains sparse and in only six studies has an attempt been made to assess psychological factors in SD.

Aronson, Brown, Litin & Pearson (1968) administered the MMPI and used psychiatric interviews. Traits most frequently observed in the so-called emotionally involved patients included, but were not limited to, suppressed anger, compulsiveness, and verbal repression. Elevations on scales of depression and social introversion were noted; however, the MMPI results failed to distinguish the subjects with SD from a general medical outpatient population.

Izdebski, Dedo, and Boles (1984) evaluated the case histories of 200 SD patients who were subsequently submitted to recurrent laryngeal nerve sectioning. Despite identifying a higher incidence of somatic complaints/illnesses among the SD group when compared to normal controls, the authors inferred from the case histories that there was "no specific clustering of factors or patterning of events that could be regarded as contributing to the causation of SD" (p.10). These authors concluded that their findings unequivocally failed to support a psychological basis for SD.

Cannito (1991) examined the emotional characteristics of 18 SD patients compared with normal controls, matched for age and sex. Statistically significant elevations were noted in the group with SD on psychometric measures of depression, anxiety, and somatic complaints. Cannito suggested that emotional/psychological factors associated with SD may affect contemporary research, and an improved understanding of the disorder should include knowledge pertaining to the contribution of emotional factors.

Recently, Kiese-Himmel and Zwirner (1996) administered standardized psychometric tests to18 subjects with adductor SD, and did not identify any significant differences in emotional instability, hypochondriasis, somatization or depression when compared to published test norms. However, the authors suggested that the personality structure of nearly half of the patients showed a tendency toward increased achievement orientation and trait anxiety. Many patients had experienced mild-to-moderate psychosocial stress within the two-year period preceding the onset of symptoms. These investigators concluded that SD is probably the result of a combination of unknown neurological and psychosocial factors.

Other invesigators have explored changes in measures of depression and anxiety following treatment for SD with botulinum toxin injections. Murry, Cannito and Woodson (1994) reported reduced levels of depression and anxiety at one week after injection and maintenance of these improvements during the ensuing two-month postinjection period. Liu and colleagues (1998) confirmed elevations on scales of anxiety, depression, and somatization symptoms in a group of patients with SD when compared to a matched healthy control group. One month following botulinum toxin injection however, significant improvements in affective adjustment and quality of life were observed. Collectively, these researchers concluded that elevated depression and anxiety might best be regarded as state-dependent characteristics, most suitably viewed as concomitant rather than causal.

It is apparent from the preceding discussion that the presence and character of psychological processes and personality factors in SD is unclear. Experimental verification is needed to elucidate the relationship between psychological sequelae and SD.

A Theory of the Dispositional Bases of Functional Dysphonia and Vocal Nodules

It is evident that the cause of common voice problems such as FD, VN and SD is poorly understood, and may involve the convergence of multiple factors, including organic, psychological, and social features. One 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 particular voice disorders.

Roy and Bless (in press) proposed a theory that identified personality traits as important factors in the development and maintenance of FD and VN. The theory is based upon Newman and colleagues' (Newman & Wallace, 1993a, b; Patterson & Newman, 1993; Wallace & Newman, 1991) coupling of a biological theory of personality (Eysenck, 1967; Eysenck, H. & Eysenck, M., 1985) with a neuropsychological model of the conceptual nervous system (Gray, 1975). Based on differences in personality, the theory predicts "*unique*" and "*contrasting*" signal sensitivities and behavioral response biases for individuals with FD and VN. Roy and Bless proposed that specific 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 is postulated to serve as a persistent diathesis, rendering an individual vulnerable for recurrence of symptoms. Given

the presumed neuropathological origins of SD, personality was not proposed to play a causal role. The following sections briefly outline the theory's foundations and its predictions.

A. Personality and its Hierarchical 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). Traits are typically defined as durable dispositions (response tendencies) that reflect individual differences (Tellegen, 1985). Genetic and environmental factors have been implicated in trait formation (Plomin, Loehlin, Defries, 1985), and certain personality traits vary with age and sex (Eysenck & Eysenk, 1975).

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 traits at the top (Goldberg, 1993; John, 1990) (Figure 1). 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; John, 1990). The so-called "Big Three" dimensions are commonly referred to as (1) Extraversion vs. Introversion (E), (2) Neuroticism vs. Stability (N), and (3) Constraint vs. Disinhibition (CON).[1]

The "Big Three" personality dimensions are typically derived using factor analytic techniques and thus are not necessarily tied to actual psychobiological processes. While many personality theories exist (see Peterson, 1988 for a review), some investigators such as Hans Eysenck (1967; Eysenck & Eysenck, 1985) and Jeffrey Gray (1982; 1987) have linked the superfactors to specific psychophysiological processes and neurobiological substrates. Extraversion (E) and Neuroticism (N) play a vital role in the Roy and Bless theory, which borrows from a synthesis of Eysenck's and Gray's biological theories of personality to account for the development of FD and VN (Newman & Wallace, 1993a, 1993b; Patterson & Newman, 1993; Wallace & Newman, 1991). The reader is referred to Roy and Bless (in press) for a complete explication of the theory and its assertions.

B. Eysenck's Personality System

H. J. Eysenck (1967) developed an integrated biopsychosocial theory of personality that is based primarily on two of the superfactor dimensions of personality, Extraversion (E) and Neuroticism (N). Extraversion involves the willingness to engage and confront the environment, including the social environment. Extraverts (high E) tend to be dominant, sociable, and active, whereas introverts (low E) tend to be quiet, unsociable, passive, and careful. Eysenck views E as reflecting stable differences in the activity level of the ascending reticular activating system, and thus cortical arousal.

Neuroticism, the second personality dimension, can be likened to emotionality and is related to anxious, depressed, tense, and emotional characteristics. High N individuals tend to be emotionally unstable, worried, or highly reactive to environmental stimuli. Eysenck (1967) identified the visceral brain as the neural substrate of N. This included the septum, hippocampus, cingulum, amygdala, and hypothalamus. Neuroticism magnifies response tendencies derived from E (Eysenck & Eysenck, 1975). Therefore, neurotic introverts tend to be more introverted, and neurotic extraverts tend to be more extraverted, when compared to their stable counterparts.[2]

C. Gray's Theory of Personality and Nervous System Function

Gray (1975, 1982, 1985) has proposed a general theory of personality linked to a conceptual nervous system model that consists of a set of three interacting components:

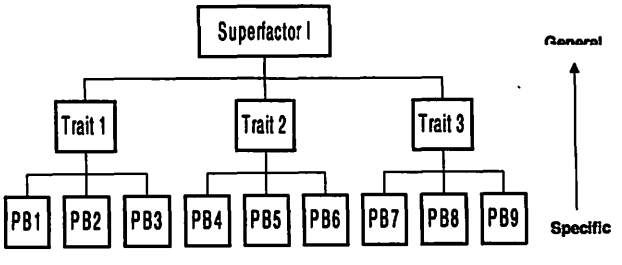


Figure 1. The hierarchical organization of personality. Different levels in the trait hierarchy represent different levels of breadth or abstraction in personality description. Several narrow, specific traits cluster to define a single broad dimension or superfactor. These lower level constituent traits are linked to many interrelated "psychological behaviors (PB)" including actions, thoughts and feelings.

[1] Constraint vs. disinhibition (CON) is centered on the basic issue of impulse control. High constraint individuals are cautious, restrained, refrain from risky adventures, and accept 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).

[2] Eysenck also proposed a third personality dimension known as Psychoticism (P) or "tough-mindedness." This broad personality factor shares many of the impulsive features of the Constraint dimension (reversed), but also includes a predilection for aggressive behavior. These three dimensions E, N, and P are measured by the 90-item Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975) to be reviewed in the companion article (this issue). The EPQ P scale is considered a measure of the constraint dimension (reversed).

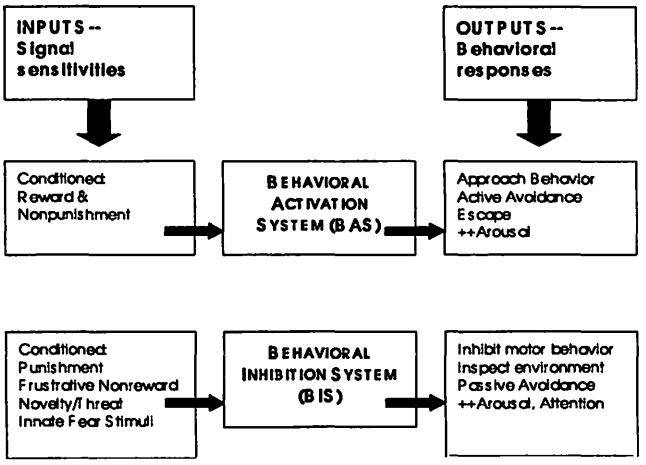


Figure 2. The signal inputs and behavioral outputs of Gray's (1975, 1982, 1985) behavioral activation and behavioral inhibition systems

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. The BAS is considered the "go" system, and promotes the initiation of goal-directed motor behavior, including approach, escape, and active avoidance (Figure 2).

The BIS, on the other hand, is responsible for organizing reactions to conditioned signals of punishment, signals of frustrative nonreward, and novel or threat stimuli. Frustrative 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 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 nonreward by not performing a given action (i.e., response suppression).

The third component of Gray's model, the NAS, serves to prepare or ready the organism to respond to BAS or BIS inputs that have motivational or emotional significance. All three components–the BAS, BIS, and NAS–have been linked to specific neural substrates.

Several particulars regarding Gray's theory should be recognized. First, 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. Second, an increase in the activity of either behavioral system results in augmented NAS activity via excitatory outputs from the BAS and BIS. Third, the NAS has excitatory connections affecting responses mediated by the behavioral systems, so that as NAS activity increases, the speed and strength of behavioral responses increase proportionately.

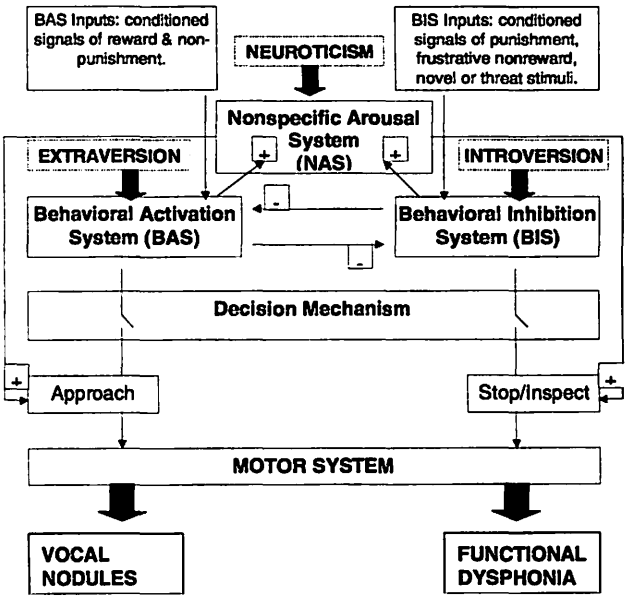


Figure 3. The Roy and Bless (in press) theory of the dispositional bases of vocal nodules and functional dysphonia, adapted from Newman & colleagues synthesis of Eysenck's and Gray's biological theories of personality. The three systems within Gray's conceptual nervous system model are mapped onto E and N. Functional dysphonia and vocal nodules are viewed as behavioral consequences of the signal sensitivities and response biases of BIS dominant neurotic introverts and BAS dominant neurotic extraverts respectively.

D. The Significance of Extraversion and Neuroticism in FD and VN

Newman and colleagues (Newman & Wallace, 1993a, 1993b; Patterson & Newman, 1993; Wallace & Newman, 1991) 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 an adaptation of this synthesis that provides the foundation for the Roy and Bless (in press) theory of the dispositional bases of FD and VN (Figure 3).

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, nonreward sensitive = stop/inspect, behavior. Neuroticism directly reflects the reactivity of the NAS: an individual is neurotic by reason of possessing a more reactive (i.e., labile) NAS than a stable individual. Neuroticism augments response tendencies associated with the two behavioral systems; therefore, as N increases, extraverts tend to act in a more extraverted manner, and intro-

verts tend to behave in a more introverted manner. In this model, the conjunction 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).[3] 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).

This model and its presumed signal sensitivities and response biases have implications for both types of voice disorders, FD and VN. These implications will be described in the section that follows.

E. Applying the Theory to Individuals with FD and VN

Roy and Bless (in press) contend that the signal sensitivities and response biases of BIS-dominated *neurotic introverts* contribute to the development of FD. They hypothesize that FD is related to anxiety, inhibitory laryngeal motor behavior, and elevated laryngeal tension states. According to the theory, elevated neuroticism and low extraversion (high N, low E) should characterize the personality of individuals with FD. Partial or complete voice loss in the absence of structural pathology therefore reflects the cumulative effects of heightened NAS and BIS, with resultant motor inhibition and elevated tension states.

Roy and Bless (in press) also allege that VN development is in part a result of 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 extended voice use and abuse, VN patients 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, Roy and Bless reason that VN patients should score high on indices of extraversion (dominance, sociability) and neuroticism (emotional reactivity), and low on measures of constraint (reflecting impulsivity). Neuroticism serves to potentiate the signal sensitivities and response biases of extraversion leading to impulsivity.

To summarize, Roy and Bless (in press) link the disorders of FD and VN to personality differences related to dissimilar signal sensitivities and response biases of neurotic introverts and neurotic extraverts, respectively. The hyperreactivity of the BIS is a prime constituent in the pathogenesis of FD, whereas hyperreactivity of the BAS is pathogenic in VN development. Both behavioral systems are amplified by the NAS.

In the companion article that follows, assessment of these broad personality dimensions is undertaken. Such an evaluation 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, 1983; Roy, Bless, Heisey & Ford, 1997). Improved understanding of its influence could help to explain voice therapy failure and refine treatment strategies in some cases (Gunther, Mayr-Graft, Miller, & Kinzl, 1996). 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.

[3] The term "neurotic" as it is employed in this context should not be mistaken as synonymous with the Freudian concept of the neurotic (i.e., an individual with a clinical neurosis). Rather, neurotic is used to describe individuals who score above the median on N-sensitive personality measures, such as the Eysenck Personality Quesitionnaire (Eysenck & Eysenck, 1975) or the Multidimensional Personality Questionnaire (Tellegen, 1982). It should be noted that individuals who are high scorers on N are not necessarily clinically disturbed or even dysfunctional.

Acknowledgements

This work was supported in part by the National Center for Voice and Speech through Grant P60 00976 for the National Institute on Deafness and Other Communication Disorders. Parts of this manuscript were presented at the 1997 American Speech Language Hearing Association Annual Convention, Boston, MA. The authors gratefully acknowledge Lisa Roteliuk, M.S. for her assistance in manuscript preparation. Please address correspondence to: Nelson Roy Ph.D., Department of Communication Disorders, 390 South 1530 East, Room 1219, Salt Lake City, Utah 84112-0252, (801) 585-0428, Fax: (801) 581-7955.

References

Akiskal, H. S., Hirschfeld, R. M. A., & Yerevanian, B. I. (1983). The relationship of personality to affective disorders. *Archives of General Psychiatry*, 40, 801-810.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.

Arnold, G. A. (1959). Changing interpretations of a persistent affliction. *Logos*, 2, 3-14.

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

Aronson, A. E. (1990). *Clinical Voice Disorders: An Interdisciplinary Approach* (3rd ed.). New York: Thieme.

Aronson, A. E., Brown, J. R., Litin, M. E., & Pearson, J. S. (1968). Spastic dysphonia I: Voice, neurologic and psychiatric aspects. *Journal of Speech and Hearing Disorders*, 33, 203-218.

Aronson, A. E., Peterson, H. W., & Litin, E. M. (1966). Psychiatric symptomatology in functional dysphonia and aphonia. *Journal of Speech and Hearing Disorders*, 31, 115-127.

Blitzer, A., Lovelace, R. E., Brin, M. F., Fahn, S., & Fink, M. E. (1985). Electromyographic findings in focal laryngeal dystonia (spastic dysphonia). *Annals of Otology, Rhinology, Laryngology*, 94, 591-594.

Bloch, P. (1965). Neuro-psychiatric aspects of spastic dysphonia. *Folia Phoniatrica*, 17, 301-364.

Boone, D. R., & McFarlane, S. (1988). *The Voice and Voice Therapy* (4th ed.). 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* (pp. 453-481). Springfield, IL: Charles C. Thomas.

Brodnitz, F. S. (1976). Spastic dysphonia. *Annals of Otorhinolaryngology*, 85, 210-214.

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.

Contrada, R. J., Leventhal, H., & O'Leary, A. (1990). Personality and health. In L.A. Pervin (Ed.), *Handbook of personality theory and research* (pp. 638-669). New York: Guilford Press.

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.

Cooper, M. (1980). Recovery from spastic dysphonia by direct voice rehabilitation. In B. J. Urban (Ed.), *The Proceedings of the 18th Congress of the International Association of Logopedics and Phoniatrics*, Vol. 1. Rockville, MD: American Speech-Language-Hearing Association.

Deary, I. J., Scott, S., Wilson, I. M., White, A., MacKenzie, K., Wilson, J. A. (1997). Personality and psychological distress in dysphonia. *British Journal of Health Psychology*, 2, 333-341.

Diehl, C. F. (1960). Voice and personality: An evaluation. In D. A. Barbara (Ed.), *Psychological and psychiatric aspects of speech and hearing* (pp. 171-203). Springfield, IL: Charles C. Thomas.

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.

Dubovsky, S. L., & Weissberg, M. P. (1982). *Reactions to Illness: In Clinical Psychiatry in Primary Care* (2nd ed.). 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.

Eysenck, H. J. (1967). *Biological Basis of Personality*. Springfield, IL: Thomas.

Eysenck, H. J. & Eysenck, M. W. (1985). *Personality and Individual Differences: A Natural Science Approach*. New York: Plenum Press.

Eysenck, H. J., & Eysenck, S. B. (1975). *Manual of the Eysenck Personality Questionnaire*. San Diego, CA: Educational and Industrial Testing Service.

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.

Finitzo, T., & Freeman, F. J. (1989). Spasmodic dysphonia, whether and where: Results of seven years of research. *Journal of Speech and Hearing Research*, 32, 541-55.

Friedl, W., Friedrich, G., & Egger, J. (1990). Personality and coping with stress in patients suffering from functional dysphonia. *Folia Phoniatrica*, 42, 144-149.

Friedl, W., Friedrich, G, Egger, J., & Fitzek, I. (1993). Psychogenic aspects of functional dysphonia. *Folia Phoniatrica*, 45, 10-13.

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.

Graham, J. R. (1987). *The MMPI: A Practical Guide* (2nd ed.). 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. (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. (1988). The inter-relationship between vocal and psychological characteristics: A literature review. *Australian Journal of Human Communication Disorders*, 16, 31-43.

Green, G. (1989). Psychobehavioral characteristics of children with vocal nodules: WPBIC ratings. *Journal of Speech and Hearing Disorders*, 54, 306-312.

Greene, M. C., & Mathieson, L. (1989). *The Voice and its Disorders* (5th ed.). London: Whurr Publishers.

Gunther, V., Mayr-Graft, A., Miller, C., & Kinzl, H. (1996). A comparative study of psychological aspects of recurring and non-recurring functional aphonias. *European Archives of Otorhinolaryngology*, 253, 240-244.

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 (pp. 250-263)*. Springfield, IL: 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.

Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1990). Phonatory function associated with hyperfunctionally related vocal fold lesions. *Journal of Voice*, 4, 52-63.

Holroyd, K. A., & Coyne, J. (1987). Personality and health in the 1980's: Psychosomatic medicine revisited? *Journal of Personality*, 55, 359-375.

House, A. O., & Andrews, H. B. (1987). The psychiatric and social characteristics of patients with functional dysphonia. *Journal of Psychosomatic Research*, 3, 483-490.

Izdebski, K., Dedo, H. H., & Boles, L. (1984). Spastic dysphonia: A patient profile of 200 cases. *Journal of Otolaryngology*, 5, 7-14.

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.

Kiese-Himmel, C., & Zwirner, P. (1996). Psychological factors in spasmodic dysphonia. *Laryngorhinootologie*, 75, 397-402.

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., & Le Boutillier, A. (1988). The outcome of different management patterns for vocal cord nodules. *Journal of Laryngology and Otology*, 102, 423-427.

Liu, C. Y., Yu, J. M., Wang, N. M., Chen, R. S., Chang, H. C., Li, H. Y., Tsai, C. H., Yang, Y. Y., & Lu, C. S. (1998). Emotional symptoms are secondary to the voice disorder in patients with spasmodic dysphonia. *General Hospital Psychiatry*, 20, 255-259.

Ludlow, C., Hallett, M., Sedory, S., Fujita, M., & Naughton, R. (1990). The pathophysiology of spasmodic dysphonia and its modification by botulinum toxin. In: Beradelli, A., Benecke, R., Manfredi, M., Marsden, C., (Eds.), *Motor Disturbances II* (pp.274-288). Orlando: Academic Press Inc.

Marks, P. A., Seeman, W., & Haller, D. L. (1974). *The Actuarial Use of the MMPI with Adolescents and Adults*. Baltimore: Williams & Wilkins Company.

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. A. (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.

Moses P. J. (1954). *The Voice of Neurosis*. New York: Grune & Stratton.

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.

Murry, T., & Woodson, G. (1992). Comparison of three methods for the management of vocal fold nodules. *Journal of Voice*, 6, 271-276.

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: Implications for disinhibitory psychopathology in children. *Clinical Psychology Review*, 13, 699-720.

Newman, J. P., & Wallace, J. F., (1993b). Cognition and psychopathy. *In Psychopathology and Cognition*, New York: Academic Press.

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 Surgery*, 108, 643-647.

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 study on the self-concept of children with vocal nodules*. Paper presented at the ASHA convention. Detroit, MI.

Peterson, C. (1988). *Personality*. Orlando, FL:Harcourt, Brace, Jovanovich.

Pfau, E. M. (1975). Psychologische unter-suchungsergegnisse sur atiologie der psychogenen dysphonien. *Folia Phoniatrica*, 25, 298-306.

Plomin, R., Loehlin, J. C., & Defries, J. C. (1985). Genetic and Environmental Components of "Environmental" Influences. *Developmental Psychology*, 21, 391-402.

Rammage, L. A., Nichol, H., & Morrison, M. D. (1987). The psychopathology of voice disorders. *Human Communications Canada*, 11, 21-25.

Reiser, D. E. (1980). Reactions to illness. In: D. E Reiser, & A. K. Schroder, (Eds.), *Patient Interviewing: The Human Dimension*. Baltimore: Williams & Wilkins.

Roy, N. & Bless, D. M. (in press). Toward a theory of the dispositional bases of functional dysphonia and vocal nodules: Exploring the role of personality and emotional adjustment. In R. D. Kent & M. J. Ball (Eds.), *The Handbook of Voice Quality Measurement*, San Diego: Singular Publishing Group.

Roy, N., Bless, D. M., Heisey, D., & Ford, C. F. (1997). Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short- and long-term treatment outcomes. *Journal of Voice*, 11, 321-331.

Roy, N., & Leeper, H. A. (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 convention, New Orleans, LA.

Roy, N., McGrory, J. J., Tasko, S. M., Bless, D. M., Heisey, D., & Ford, C. N. (1997). Psychological correlates of functional dysphonia: An evaluation using the Minnesota Multiphasic Personality Inventory. *Journal of Voice*, 11, 443-451.

Schalen, L., & Andersson, K. (1992). Differential diagnosis and treatment of psychogenic voice disorder. *Clinical Otolaryngology*, 17, 225-230.

Smith, T. W., & Williams, P. G. (1992). Personality and health: Advantages and disadvantages of the 5-factor model. *Journal of Personality*, 60, 395-423.

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.

Stone, S. V., & Costa, P. T., Jr. (1990). Disease-prone personality or distress-prone personality? The role of neuroticism in coronary heart disease. In H. S. Friedman (Ed.), *Personality and disease* (pp. 178-202). New York: Wiley.

Suls, J., & Rittenhouse, J. D. (1990). Models of linkages between personality and disease. In H. S. Friedman (Ed.), *Personality and Disease* (pp. 38-64). New York: Wiley.

Tellegen, A. (1982). Brief Manual for the Multidimensional Personality Questionnaire. Unpublished manuscript, University of Minnesota, Minneapolis.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A.H. Tuma & J.D. Maser (Eds.). *Anxiety and the Anxiety Disorders* (pp. 681-706). Hillsdale, NJ: Erlbaum.

Toohill, R. J. (1975). The psychosomatic aspects of children with vocal nodules. *Archives of Otolaryngology*, 101, 591-595.

Verdolini-Marston, K. L., Burke, M. K., Lessac, A., Glaze, L., & Caldwell, E. (1995). Preliminary study of two methods of treatment for laryngeal nodules. *Journal of Voice*, 9, 74-85.

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.

Watson, D., Clark, L. A., & Harkness, A. R. (1994). Structures of personality and their relevance to psychopathology. *Journal of Abnormal Psychology*, 103, 18-31.

White, A., Deary, I. J., Wilson, J. A. (1997). Psychiatric disturbance and personality traits in dysphonic patients. *European Journal of Disorders of Communication*. 32, 121-128.

Wilson, D. K. (1987). *Voice Problems of Children* (3rd.ed.). Baltimore, MD: Williams & Wilkins.

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 Rorschach Protocol Interpretation. *Acta Symbolica*, 43-55.

Withers, B. T., & Dawson, M. H. (1960). Psychological aspects...Treatment of vocal nodule cases. *Texas State Journal of Medicine*, 56,43-46.

Yano, J. L., 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 - 14
September 1999, 203-219

Personality and Voice Disorders: A Superfactor Trait Analysis

Nelson Roy Ph.D.
Department of Communication Disorders
Division of Otolaryngology-Head & Neck Surgery, The University of Utah
Diane Bless Ph.D.
Department of Communicative Disorders
Division of Otolaryngology-Head & Neck Surgery, The University of Wisconsin-Madison
Dennis Heisey Ph.D.
Department of Surgery, The University of Wisconsin-Madison

Abstract

To determine whether personality factors play causal, concomitant, or consequential roles in common voice disorders, a vocally normal control group and four groups with voice disorders—functional dysphonia (FD), vocal nodules (VN), spasmodic dysphonia, (SD) and unilateral vocal fold paralysis (UVFP)—were compared on measures of personality and psychological adjustment. Superfactor group comparisons revealed that the majority of FD and VN subjects were classified as introverts and extraverts, respectively. Comparisons involving the SD, UVFP, and control subjects did not identify consistent personality differences. The disability hypothesis, which suggests that personality features and emotional maladjustment are solely a negative consequence of vocal handicap, was not supported. Personality variables and their behavioral consequences may therefore contribute to FD and VN. Results are presented within the context of a dispositional theory offered by Roy and Bless (in press).

The role of personality in common voice disorders remains enigmatic. Considerable controversy surrounds whether personality factors and psychological adjustment should be considered causal, concomitant, or consequential to disorders such as functional dysphonia (FD), vocal nodules (VN) and spasmodic dysphonia (SD). Roy and Bless (in press) proposed a theory to delineate the dispositional bases of FD and VN. The authors speculated that personality composition, and the consequent cognitive processing and behavioral patterns, provides an important footing for the development of FD and VN. The "superfactor" personality trait dimensions of Extraversion (E) and Neuroticism (N) play a vital role in the theory, which represents a synthesis of Eysenck's (1967) biological theory of personality and Gray's (1975) neuropsychological model of the conceptual nervous system. The synthesis mapped the extraversion dimension onto the relative strengths of Gray's behavioral activation and inhibition systems and identified the neuroticism dimension with Gray's nonspecific arousal system. This theory linked the disorders of FD and VN to the signal sensitivities and response biases of neurotic introverts and neurotic extraverts, respectively.

The purpose of this research was to evaluate individual differences in personality and aspects of psychological functioning by comparing a variety of voice-disordered groups on measures of personality and emotional adjustment. Using a superfactor approach to personality description, the following research questions were addressed: (1) Do personality differences exist between voice-disordered groups at the superfactor trait level? (2) If so, are these differences consistent with hypotheses derived from the Roy and Bless (in press) theory describing the dispositional bases of FD and VN? (3) To what extent are these personality differences related to age? (4) Do group differences exist in the presence and degree of self-reported depression and anxiety? If so, are these differences related to the level of vocal handicap, and/or the duration of voice symptoms? (5) For clinical purposes, do scales of personality offer useful diagnostic information that would potentially distinguish the groups?

Method

Participants

In order to address our research questions, four voice-disordered groups and a non-voice-disordered medical control group were evaluated using measures of personality and psychological distress. Because the voice disorders of interest occur predominantly in female populations, only female subjects were recruited to participate. This restriction permitted comparisons between and within groups.

With the exception of 12 Canadian subjects within the FD group, all subject selection and recruitment was based at the University of Wisconsin-Madison and conducted by the Division of Otolaryngology–Head and Neck Surgery. Consecutive subjects with voice disorders were identified and recruited to participate. The non-voice-disordered medical outpatient control group was also recruited from the same institution by the attending otolaryngologist or nurse during a nonemergent visit.

A total of 292 potential subjects were invited to participate. From that total, 208 individuals (71.2%) agreed to complete all questionnaires. Despite receiving reminder notices, 32 subjects failed to return their completed questionnaires, resulting in 177 total respondents, or 61.5% of all subjects recruited. Eight subjects from the UVFP group were eventually excluded, based upon subsequent comprehensive chart review that revealed that these subjects did not meet the strict inclusion described below. A summary of response ratios for individual groups is described in Table 1. Response ratios ranged from 51% to 78%. The final number of participants in this research study was 169, or 57.8% of all subjects recruited.

Table 1.
A summary of subject recruitment and participation by group membership. Response ratios (%) for each group indicate the percentage of originally invited subjects who agreed to participate or completed all questionnaires (i.e., respondents).

Groups	Total # of Subjects Invited to Participate	Total # of Subjects Agreed to Participate (Response Ratio %)	Total # of Respondents (Response Ratio %)
Functional Dysphonia	N = 66 45 US; 21 Canadian	N = 55 (83.3%) 39 US; 16 Canadian	N = 45 (68.2%) 33 US; 12 Canadian
Vocal Nodules	N = 71	N = 45 (63.4%)	N = 37 (52.1%)
Spasmodic Dysphynia	N = 45	N = 36 (80%)	N = 35 (77.8%)
Unilateral Vocal Fold Paralysis	N = 45	N = 30 (66.7%)	N = 23 (51.1%) 8 later excluded for criteria reasons
Controls	N = 65	N = 42 (64.6%)	N = 37 (57%)

Groups

Group 1: Functional Dysphonia (FD) — This experimental group consisted of 45 women (mean age 49.0 (13.3); range 22 to 79 years) with disordered voices in the absence of neurological and/or structural pathology, who had not undergone previous laryngeal microsurgery (Koufman & Blalock, 1982). The diagnosis of FD was made if comprehensive laryngeal examination and medical investigation by both a laryngologist and a speech language pathologist specializing in voice disorders failed to identify laryngeal disease or laryngeal neuropathology sufficient to account for the dysphonia.

Group 2: Vocal Nodules (VN) — This group consisted of 37 women (mean age 33.7 (9.9); range 18 to 63 years) with disordered voices secondary to bilateral vocal fold nodules. Videolaryngostroboscopic examination confirmed the presence of bilateral nodules at the junction of the anterior one-third and posterior two-thirds of both vocal folds.

Group 3: Spasmodic Dysphonia (SD) — This group consisted of 35 women (mean age 55.1 (15.0); range 30 to 78 years) who had received the diagnosis of adductor SD following case history review and auditory-perceptual, videolaryngostroboscopic, aerodynamic, and in some cases electromyographic evaluations. Auditory-perceptual characteristics of adductor SD include sustained or intermittent strained-strangled voice, hard glottal initiation, choked vocal attack, and staccato and stutterlike blocks (Izdebski & Dedo, 1981; Ludlow, Naunton & Bassich, 1984). The presence of voice tremor was noted but was not considered an exclusionary criterion. In addition to uncontrolled breaks in phonation during connected speech, subjects typically demonstrated improved performance during whisper and falsetto, normal laughing, throat clearing, and coughing (Cannito, 1991). Subjects had no previous history of other neurological or speech disorders (Ludlow & Connor, 1987). Fiberoptic examination of the larynx during connected speech often revealed intermittent or relatively sustained involuntary hyperadduction of the true vocal folds and/or supraglottic structures (Finitzo & Freeman, 1989; Parnes, Lavarato, & Myers, 1979). All subjects had received periodic intracordal botulinum toxin injections. The final diagnosis of SD was corroborated by (1) an unsuccessful trial of voice therapy, and (2) an observable reduction in vocal symptoms following toxin injection (as evidenced by auditory-perceptual improvements and the patient's subjective report).

Group 4: Unilateral Vocal Fold Paralysis (UVFP) – This group consisted of 15 women (mean age 53.8 (12.4); range 33 to 69 years) with disordered voices *uniquely* associated with unilateral vocal fold paralysis. Videolaryngostroboscopy, aerodynamic evaluation, and in some cases laryngeal electromyographic assessment confirmed the diagnosis. This group constituted a fairly *etiologically homogeneous* voice-disordered group. Specifi-

cally, subjects with vocal fold paralysis due to motor neuron disease, intracranial lesions such as tumors, degenerative diseases of the central nervous system, brainstem encephalitis, cerebrovascular disease, or cancer were excluded. Patients with other cranial nerve involvement and dysfunction or polyneuropathies were also excluded to ensure a relatively pure voice-disordered group for comparison purposes. The etiology underlying the majority of paralysis cases was therefore either iatrogenic or idiopathic. These "idiopathic" laryngeal palsies are often attributed to neuritis secondary to viral infection, among other causes (Lewis, 1958).

Group 5: Non-voice-disordered otolaryngology controls (Controls) — The control group consisted of 37 consecutive female non-voice-disordered otolaryngology patients (mean age 45.3 (13.6); range 19 to 71 years). Selecting voice-disordered and control subjects from a comparable medical setting and level of medical care, specifically tertiary (specialty) hospital-based care, was preferred over selecting subjects from a dissimilar level of care, (e.g., primary care-general practice). This approach obviated possible confounds such as "help-seeking behavior," and "psychological distress," which could be related to assessment and management of medical problems at dissimilar levels of patient care (Deary, Scott, Wilson, White, MacKenzie & Wilson, 1997).

These non-voice-disordered subjects attended the otolaryngology clinic for physical complaints unrelated to voice production. They had no history of previous voice disorder, voice therapy, or medical or surgical management aimed at correcting disordered voice. Patients who received surgical or medical treatment for head and neck cancer were excluded. This head and neck cancer group was not considered analogous to the voice-disordered groups, in view of the disparate levels of physical involvement, perceived life threat, and emotional stresses confronting them.

Procedures

Description of the Self-Report Instruments

The next few sections summarize the self-report measures used to assess subjects' broad personality dimensions and aspects of emotional adjustment. Each subject completed the questionnaires and returned them in a postage-paid envelope provided by the researchers. These popular self-report instruments were selected because of their excellent test construction and psychometric properties.

Eysenck Personality Questionnaire (EPQ; H.J. Eysenck & S.B. Eysenck, [1975a]. The EPQ is one of the most widely used inventories in personality research today. It is a 90-item self-report questionnaire and is scored on one validity scale and three personality scales. The examinee responds "yes" or "no," depending on whether the subject views the statement as descriptive of her behavior. It has been used to identify three broad personality dimensions: Extraversion (E), Neuroticism (N), and Psychoticism (P). Each of these traits is measured by means of 24 yes or no statements, selected on the basis of item and factor analysis. Additionally, the EPQ contains a Lie scale (L) designed to measure response style/impression management through a tendency to "fake good." The EPQ scales have been shown to possess good reliability (H.J. Eysenck & S.B. Eysenck, 1975a) and validity (H.J. Eysenck & S.B. Eysenck, 1971; S.B. Eysenck & H.J. Eysenck, 1970; Verma & Eysenck, 1973).

The State-Trait Anxiety Inventory - Trait Scale (STAI -Trait; Spielberger, Gorusch, Luschene, Vagg, & Jacobs, 1983). The STAI-Trait is a norm-referenced inventory consisting of 20 statements describing various manifestations of anxiety/distress. Subjects use a four-point scale to indicate how frequently the individual experiences the feeling. The "trait" version is designed to assess enduring or persistent anxiety. Thus, it is felt to reflect dispositional anxiety, not mood. These scales have been submitted to a substantial degree of development and testing, particularly for use in investigating psychological concepts and behavior under various stress conditions of different groups. Both subtests are popular tools for clinical and research purposes and demonstrate acceptable psychometric properties.

The Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; Beck & Steer, 1993). The BDI consists of 21 items, each with a 0-3 grading system. Each item consists of four or five self-evaluative statements indicating increasing severity of depression, for example:

0 I do not feel sad or blue
1 I feel blue or sad
2a I am blue or sad all the time and I can't snap out of it
2b I am so sad or unhappy that it is very painful
3 I am so sad or unhappy that I can't stand it.

For some items two alternative statements are provided at the same weight. A total depression score is calculated by summing the highest scores from each item.

This instrument has undergone extensive standardization (Beck et. al. 1961), which indicates adequate internal consistency (Beck and Beamersderfer, 1974; Schwab, Bialow & Holzer, 1967) and external validity (Bech, Gram, Dein, Jacobsen, Bitger & Bolwig, 1975; Crawford-Little & McPhail, 1973; Metcalfe & Goldman, 1965). Against the Hamilton Depression Rating Scale, high correlations have been found, ranging from 0.72 to 0.82 (Bech et al., 1975; Davies, Burrows, Poynton, 1975; Schwab et al., 1967). The validity of change scores on the BDI (sensitivity) has been shown in a number of studies (Metcalfe and Goldman, 1965; Bech et al., 1975). In summary, the BDI is a widely used instrument, which is sensitive to clinical change, has good discrimination, and is covalidated with clinical rating scales and behavioral observation scales.

Voice Handicap Profile (VHP; Adapted from ASHA SID 3 Prototype). This self-report measure was designed to assess the global impact of the voice disturbance on the patient's social and vocational functioning. Because

at the time of this research, no single instrument with known psychometric characteristics existed to achieve this goal, it was necessary to develop such a tool. Item selection was adapted from a prototype currently under investigation by the ASHA Special Interest Division (SID 3)-Voice and Voice Disorders. The profile consists of 10 statements and an overall severity rating scale (see appendix A). Subjects rate on a 4-point scale how frequently they experience the condition or state. The last item requires the subject to provide an overall severity rating of their voice disturbance, where 1 equals normal, and 7 equals severe impairment. The total score provides an index of the subject's current self-rating of the degree of their voice difficulties.

Results

Do personality differences exist between groups at the superfactor level of personality description? To what extent are these differences related to age?

To assess the relation between personality and diagnostic status, we compared the voice-disordered groups and the non-voice-disordered control group on each personality and psychological adjustment measure using a one-way analysis of variance (ANOVA). Fisher's protected least significant difference (LSD) procedure was used to compare the means. Because personality differences have been associated with age (Eysenck & Eysenck, 1975), the contribution of age to the voice disorder-personality relationship was evaluated, using age as a covariate in an analysis of covariance (ANCOVA). Pairwise LSD comparisons were then repeated using age-adjusted mean scores. To ensure robustness of the analysis, both the ANOVA and ANCOVA (age on the original scale) were repeated on rank-transformed data. This analysis is equivalent to the Kruskal-Wallis k-sample test, the F test generated by the parametric procedure applied to the ranks. Again, pairwise comparisons were performed on the rank-transformed data using Fisher's protected LSD procedure.

As a general rule, the results from the ranked analyses did not differ qualitatively from the unranked analyses. When the results from the two analyses differed substantially (i.e., rendered opposite interpretations), and/or if the differences were relevant to the theoretical formulations, results from both analyses are reported.

To illustrate group comparisons on each personality scale, column graphs depict group mean scores and standard errors in descending order. The lines drawn above the columns connect sets of homogeneous groups based on LSD comparisons of the original untransformed data. Groups covered/connected by a line were not significantly different and thus constituted a homogeneous subset. Groups not covered/connected by a line were significantly different ($p<.05$, two-tailed).[1]

The Eysenck Personality Questionnaire (EPQ)

To assess the relation between personality and diagnostic status at the superfactor level, groups were compared using the "Big Three" scales derived from the EPQ: E, N, and P. The results are also summarized for the L scale.

Extraversion (E). A significant omnibus F test revealed differences between groups on the E scale ($F(4, 164) = 6.99$, $p<.0001$). E reflects individual differences in sociability, dominance, energy, and enthusiasm. The significant omnibus F test was followed by all pairwise LSD tests, which showed that the VN group was clearly more extraverted (high E) than the FD ($p<.0001$), SD ($p<.0014$) and UVFP ($p<.0031$) groups. However, the VN group did not differ significantly from the control group on the E scale ($p=.18$) (ranked data $p=.094$). Additionally, the FD ($p<.0011$) and UVFP ($p<.049$) groups were less extraverted (low E) than the control group. Figure 1 graphically illustrates the differences between the groups on E, with the VN and FD groups scoring at opposite ends of the E continuum. Age was not a significant covariate in the ANCOVA ($F(1,163) = 0.79$, $p<.38$).

Neuroticism (N). The N scale is sensitive to individual differences in emotional stability and emotional reactivity. A significant omnibus test confirmed differences between groups ($F(4, 164) = 4.27$, $p<.0026$). The FD group was significantly more emotionally reactive (high N) than the SD ($p<.0027$), UVFP ($p<.0069$), and control groups ($p<.0006$). However, the difference between the VN and FD groups only approached statistical significance ($p=.071$). The VN group also scored higher than the controls on this scale, although this difference did not achieve statistical significance ($p=.108$). Figure 1 illustrates that the FD and VN groups scored highest on N when compared to the other groups. No other differences were detected between groups. The ANCOVA did not reveal a significant age effect ($F(1, 163) = 0.79$, $p<.38$).

Psychoticism (P). The P scale is sensitive to a combination of personality constructs, including impulsivity and aggressiveness (high P), and agreeableness and conscientiousness (low P).[2] A significant omnibus test confirmed differences between groups on the P scale ($F(4,164) = 3.65$, $p<.007$). As shown in Figure 1, the VN group scored significantly higher on P when compared to the FD ($p<.0101$), SD ($p<.0013$), and control groups ($p<.0012$). No statistical

[1] Before embarking on the analysis of personality differences between groups, a comparison of the Canadian and US FD subjects was undertaken to determine whether results could be pooled for this group. T-test comparisons of the mean scores for all scales revealed no significant differences between the Canadian and US FD subjects ($p<.05$, two-tailed). Therefore, results are reported for the pooled data.

[2] The Psychoticism scale assesses P, the third dimension in Eysenck's Big Three model of personality. Recall that P contains many of the same features of the Constraint dimension reversed, so that High P is essentially equivalent to low CON.

difference was detected between the VN and UVFP groups ($p<.095$). The ANCOVA did not reveal a significant age effect ($F(5, 163) = 1.90, p<.17$).

Lie Scale (L). The L scale assesses differences in self-disclosure style and attempts at impression management. The omnibus test confirmed differences between groups on this validity scale ($F(4, 164) = 3.56, p<.0082$). Analysis of group differences revealed that the VN group scored significantly lower than the FD ($p<.002$) and SD ($p<.003$) groups on this validity scale (Figure 1). The ANCOVA did not reveal a significant age effect ($F(5, 163) = 3.37, p<.068$) and the results were unaltered when scores were adjusted for age.

Stepwise Logistic Discriminant Analysis — EPQ Superfactors

Do scales of personality exist that distinguish the groups? Do these variables have diagnostic value?

Several stepwise logistic discriminant analyses were computed to identify personality scales that successfully distinguished the voice-disordered subjects from one another and from the non-voice-disordered control subjects. Personality variables entered the model in a stepwise fashion, in the order of their discriminatory ability. Variable entry stopped when the remaining variables contributed no further significant discriminatory information. All personality variables remaining in the model were significant at the 0.05 critical level. Each logistic discriminant analysis was first performed with age excluded as an entry variable, and then repeated with age allowed to enter.

The area under the Receiver Operating Characteristic curve (A_{ROC}) was also computed to determine the discriminatory potential of the personality variables identified by the logistic analysis. A_{ROC} represents the probability that a randomly chosen "diseased" subject (or a subject with a particular voice disorder diagnosis, i.e., "a case") is (correctly) rated or ranked with greater suspicion than a randomly chosen non-diseased subject (or a subject with another voice disorder type, i.e., a "non-case"). A_{ROC} will vary from 0.5 (no apparent accuracy) to 1.0 (perfect accuracy). The null hypothesis, that the A_{ROC} is 0.5, was tested for significance with the overall model likelihood ratio Chi-square statistic (Hanley & McNeil, 1982). Results are reported separately for each pairwise group comparison on the EPQ.

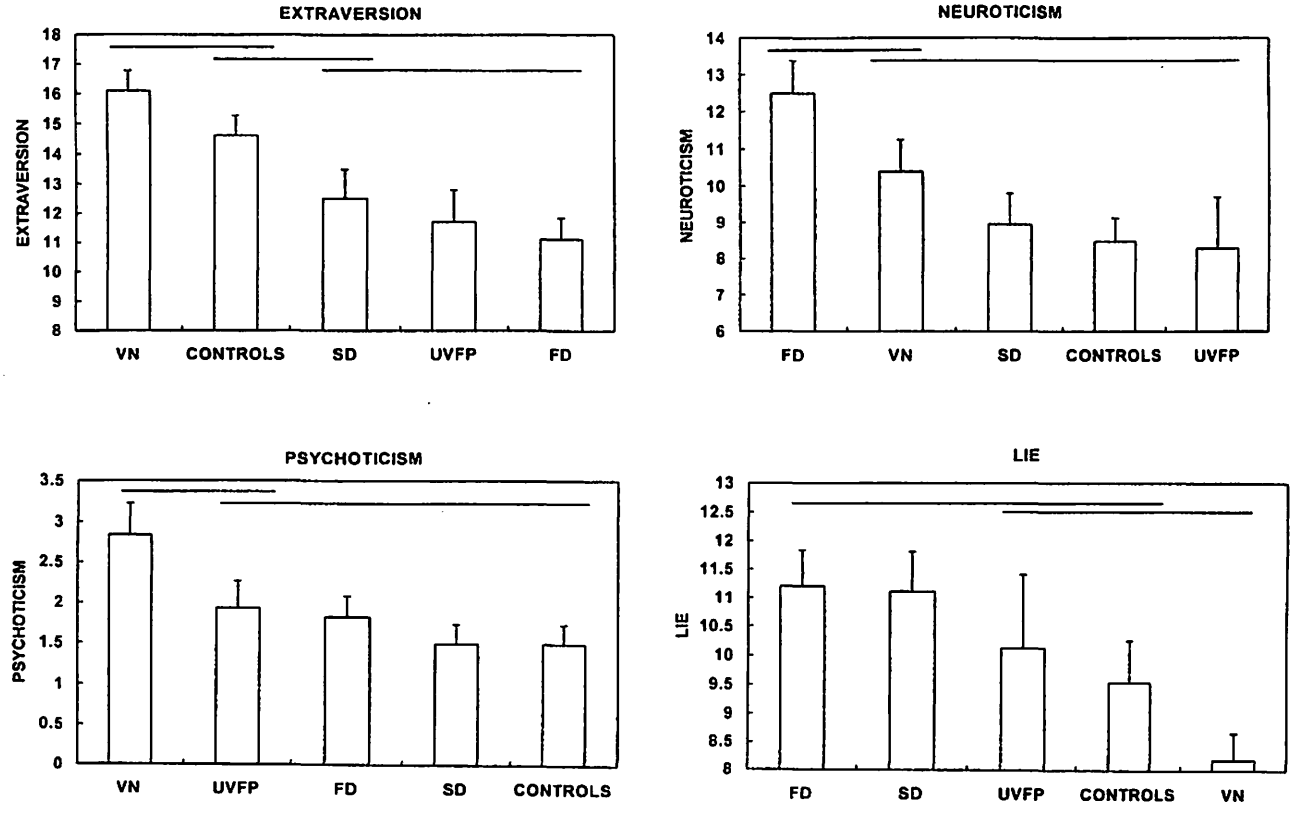


Figure 1. Rank ordered group means and standard errors for each scale of the Eysenck Personality Questionnaire (EPQ). Lines drawn above select columns (groups) indicate non-significant differences between those groups (i.e., homogeneous sets of groups). Groups not covered by a line are significantly different ($p<.05$).

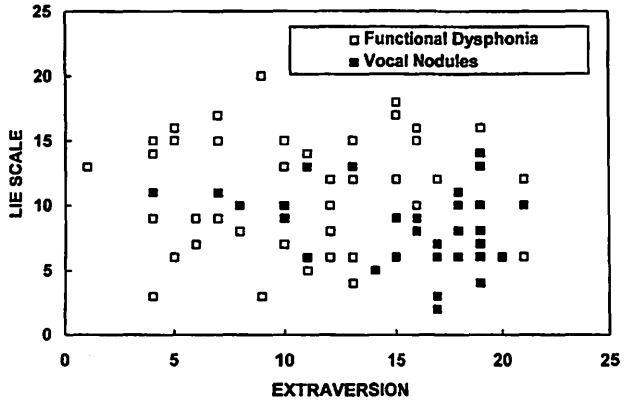


Figure 2. Scatterplot of scores for individual FD and VN subjects for the Eysenck Personality Questionnaire (EPQ) E and L scales.

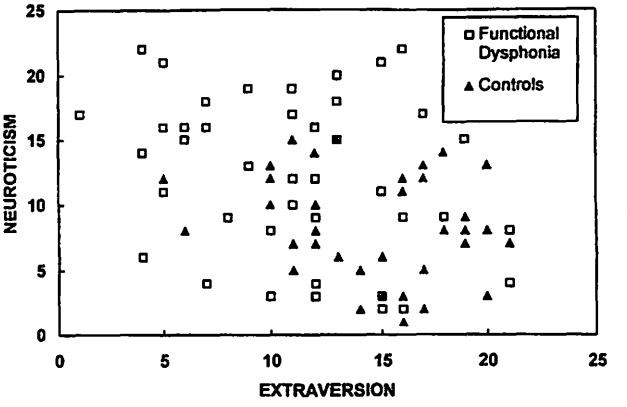


Figure 3. Scatterplot of scores for individual FD and control subjects on the EPQ scales of E and N.

To determine the diagnostic value of the discriminatory variables, sensitivity and specificity estimates were calculated. The diagnostic index was based on the predicted probability of a "case" from the logistic regression model. The cutoff used for case assignment was the midpoint between the average probabilities of the case group and control (non-case) group. This was computed as the overall average for the two groups, but with the subject weighted by the reciprocal of the group size. When a voice disorder group and the control group were compared, the predicted value was the probability that the subject was voice disordered (i.e., a "case"). However, when two different voice disorder types were compared (for example, FD versus VN), one diagnosis was designated the "case," and the other designated the "non-case." Estimates of sensitivity (the percentage of correctly identified cases) and specificity (the percentage of correctly identified non-cases) are reported for each discriminant analysis.

Functional Dysphonia vs. VN

The first stepwise discriminant analysis compared the VN group with the FD group. When age was excluded as a potential discriminatory variable, the E and L scales successfully distinguished the FD subjects from the VN subjects. The A_{ROC} was a respectable 0.826 ($p<.0001$). Figure 2 graphically illustrates the discriminatory value of these two scales. The reader should note that for each scatterplot, some individual observations may be hidden because of data overlap. The analysis was repeated with age admitted as a potential entry variable. Age entered significantly, along with E and L, increasing the A_{ROC} =.894 ($p<.0001$). The estimated sensitivity and specificity were 80% and 78%, respectively.

Functional Dysphonia vs. Controls

With age excluded as an entry variable, the stepwise logistic discriminant analysis comparing the FD group with the control group identified two personality scales, E and N, which provided significant discriminatory information. The area under the ROC curve was (A_{ROC} = 0.756, $p<.0001$). The results, illustrated in Figure 3, show that the majority of FD subjects were more neurotic and less extraverted when compared to the control subjects. The estimated sensitivity and specificity was 73% and 71%, respectively. When admitted as a variable, age did not provide any further discriminatory information.

Functional Dysphonia vs. SD and UVFP

The personality scale N successfully discriminated the FD subjects from the SD (A_{ROC} = .670, $p<.0058$) and UVFP subjects (A_{ROC} = .698, $p<.018$). For the FD vs. SD comparison, the estimated sensitivity and specificity were 63% and 62%, respectively. For the FD vs. UVFP comparison, the estimated sensitivity and specificity were 64% and 60%, respectively. Age did not enter as a significant discriminative variable.

Vocal Nodules vs. Controls

With age excluded as an entry variable, the P scale discriminated the VN subjects from the control subjects. Recall that P subsumes traits such as impulsivity and disagreeableness (aggressiveness). The area under the ROC curve was A_{ROC}=.725 ($p<.0012$). Age subsequently entered as a significant variable along with P, increasing the area under the ROC to A_{ROC} = 0.809 ($p<.0001$). The estimated sensitivity and specificity were 73% and 68%, respectively.

Vocal Nodules vs. UVFP

With age excluded as an entry variable, only one personality factor, E, significantly discriminated the VN subjects from the UVFP subjects (A_{ROC} =.791, $p<.0017$). However, when age entered as a significant discriminative variable, two personality variables, E and N, also entered significantly, resulting in a substantial increment in the area under the ROC curve (i.e., A_{ROC} = .948, $p<.0001$). The estimated sensitivity and specificity were 89% and 80%, respectively. The scatterplot shown in Figure 4 shows that the

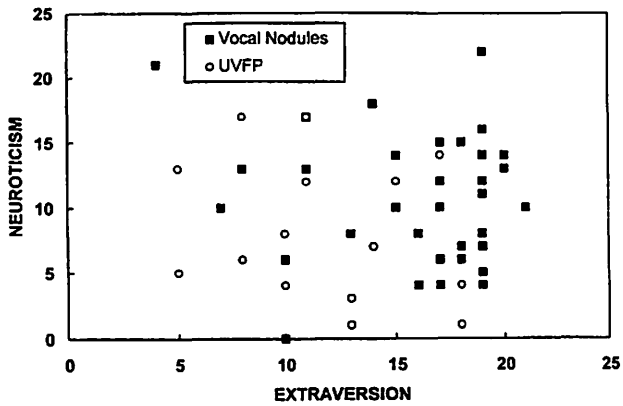


Figure 4. Scatterplot of scores for individual VN and UVFP subjects on the EPQ scales of E and N.

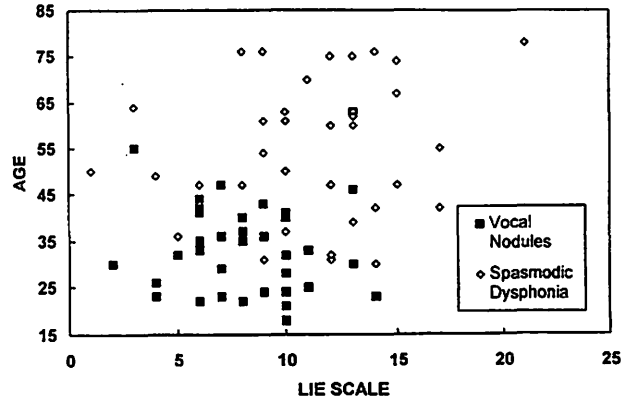


Figure 5. Scatterplot of scores for VN and SD subjects using the variable age and scores from the EPQ Lie Scale .

VN subjects scored higher on both N and E when compared to the UVFP subjects.

Vocal Nodules vs. SD

With age excluded from the analysis, two scales, E and L, distinguished the VN group from the SD group (A_{ROC}=.768, p<.0002). When age entered as a significant discriminant variable, only the L scale entered significantly, increasing the overall area under the ROC curve to a respectable A_{ROC} = .901(p<.0001). The VN subjects were generally younger and less concerned with impression management than the SD subjects, as shown in Figure 5. The estimated sensitivity and specificity were 89% and 74%, respectively.

SD vs. UVFP, Controls and (b) UVFP vs. Controls

No EPQ variables met the 0.05 significance level for entry into the model when comparing the SD group with either the control or the UVFP groups. However, the E scale weakly distinguished the UVFP subjects from the control subjects (A_{ROC} = .692, p<.024). Sensitivity and specificity were estimated to be 65% and 53%, respectively. When admitted, age did not enter as a significant variable.

As hypothesized by Roy and Bless (in press), are the majority of subjects with FD and VN classified as neurotic introverts and neurotic extraverts respectively?

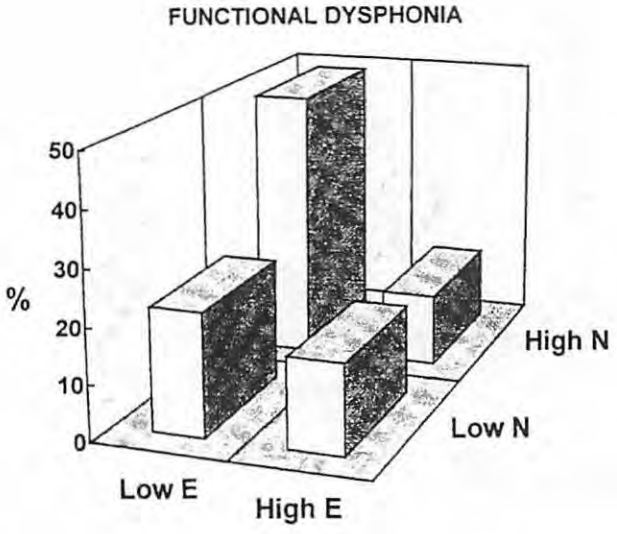
The Roy and Bless theory predicts that most FD subjects will score as high N-low E (neurotic introverts) and that most VN subjects will score as high N-high E (neurotic extraverts). Based on the presumed neurological origins of SD and UVFP, the theory does not predict any personality typologies for these voice disorder groups. To determine whether the subjects with FD and VN possessed these personality features, median scores from the entire subject pool were calculated separately for the E and N scales. The median score for the E scale was 14, and the median score for the N scale was 10. Individuals scoring greater than the median on E were classified as "high E" subjects, and those individuals scoring at or below the median were classified as "low E." Similarly, all subjects scoring above the median on N were classified as "high N," and those subjects scoring at or below the median on N were discretized as "low N." Dichotomizing each group in this way generated four possible classifications: high N-high E (neurotic extraverts), low N-high E (stable extraverts), high N-low E (neurotic introverts), and low N-low E (stable introverts). A Chi-square analysis confirmed differences between groups in proportions of each personality type (Chi-square (12, 169) = 31.96, p<.001). Chi-square was tested against equal probability in all cells.

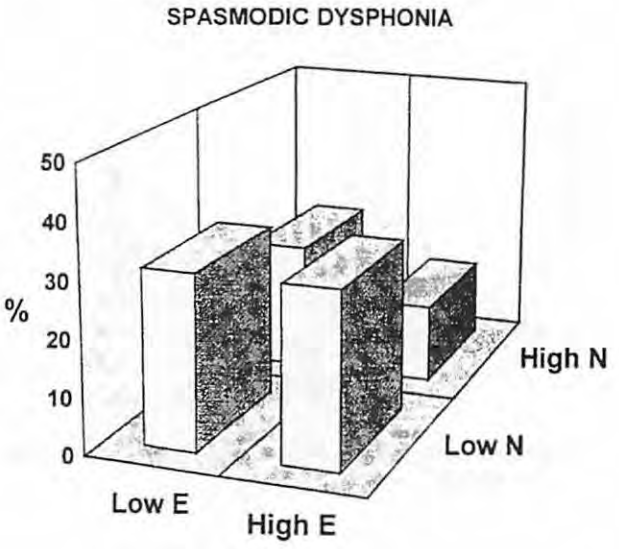
The results are displayed in Figures 6, 7, and 8. In Figure 6, a comparison of the distribution of FD and VN subjects revealed that the 49% of subjects with FD occupied the high N-low E quadrant of the cell matrix, whereas the majority of VN subjects occupied the low N-high E (43%) or high N-high E (32 %) quadrants. The pattern of the results illustrates the preponderance of low E in FD and high E in VN. Figure 7 confirms that the SD group was not dominated by a single personality dimension or combination of dimensions. However, as indicated in the logistic discriminant analysis, 47% of subjects with UVFP appeared to score below the median on N and E (i.e., stable introverts). The control subjects appeared equally divided between high E and low E (Figure 8).

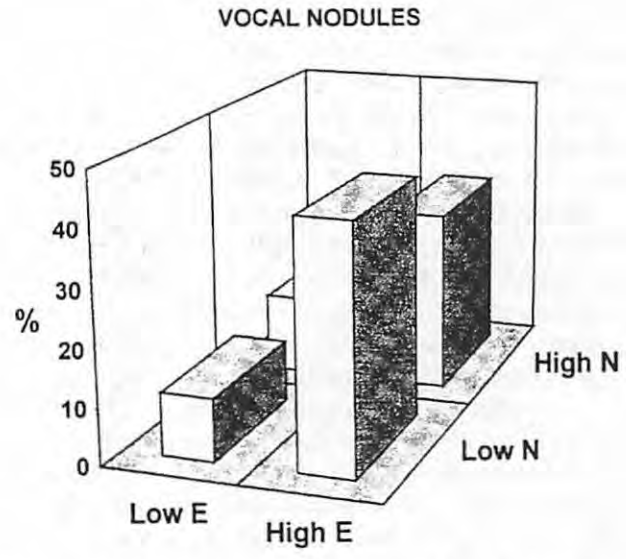
Do group differences exist in the degree of depression and anxiety? If so, are these differences related to the level of vocal handicap, and/or the duration of voice symptoms?

State-Trait Anxiety Inventory (STAI)

A significant F test revealed omnibus differences between groups on scores obtained from the STAI (F(4, 1565) = 4.63, p<.0015). The trait version of the STAI evalu-







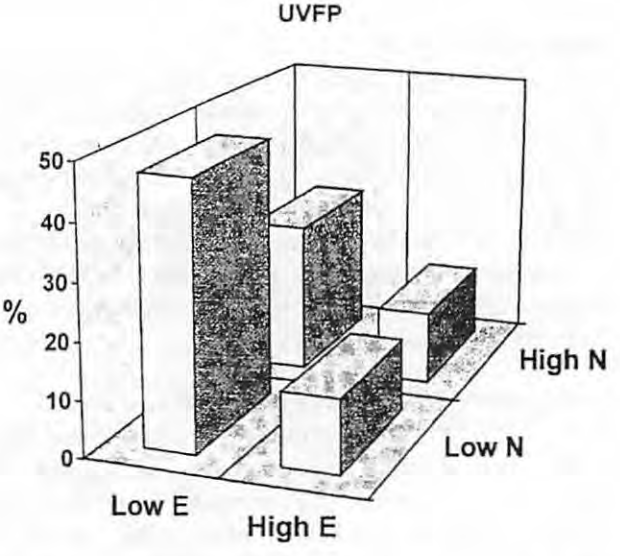


Figure 6. Distribution of subjects with FD and VN based on dichotomized scores from the EPQ superfactors Extraversion (E) and Neuroticism (N). The majority of FD subjects scored as high N-low E, whereas most VN subjects scored as low N-high E, or as high N-high E.

Figure 7. A 3-D column graph illustrating the distribution of subjects with SD and UVFP based on dichotomized scores from the EPQ superfactors of E and N. Scores for the subjects with SD distribute fairly evenly throughout the cell matrix, whereas the subjects with UVFP collect primarily in the low E cells.

ates differences in long-standing, relatively stable (trait) anxiety. The FD group reported more trait anxiety than the control ($p<.0001$), VN ($p<.0141$), SD ($p<.0021$), and UVFP ($p<.0029$) groups (Figure 9). No significant age effect was identified in the ANCOVA ($F(1, 156) = 0.10$, $p<.75$).

Beck Depression Inventory (BDI)

The omnibus test on the original data was not rejected ($F(4, 150), = 2.00$, $p<.098$); however, rank-transformed BDI data revealed omnibus group differences ($F(4, 150) = 2.61$, $p<.038$). This justified further exploration of pairwise contrasts. The FD group reported more symptoms of depression than the control ($p<.025$), VN ($p<.033$), and SD ($p<.013$) groups (Figure 10). When scores were adjusted for age, the difference between the FD and VN groups no longer remained significant ($p<.19$). No other differences between groups were identified.

The voice handicap profile (VHP) was designed to assess the global impact of the voice disturbance on the patient's social and vocational functioning. To assure internal consistency of the VHP, Cronbach's coefficient alpha was calculated separately for each voice-disordered group, using the first 10 items of the scale. Coefficient alpha calculations

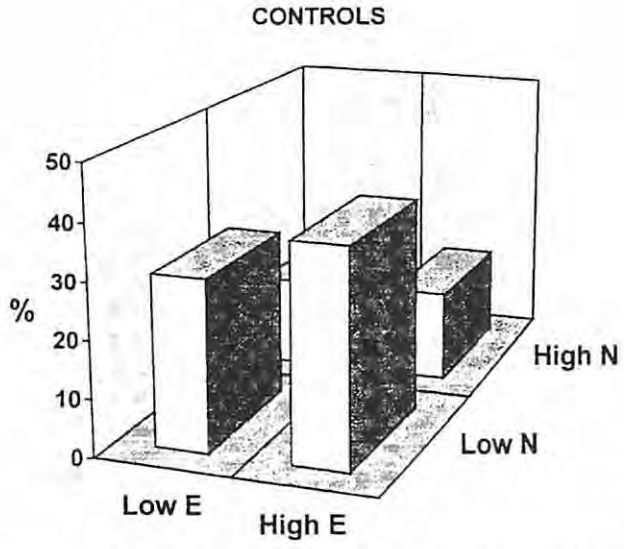


Figure 8. A 3-D column graph illustrating the distribution of control subjects based on dichotomized scores from the EPQ superfactors E and N. The majority of controls were classified as low N-low E or low N-high E.

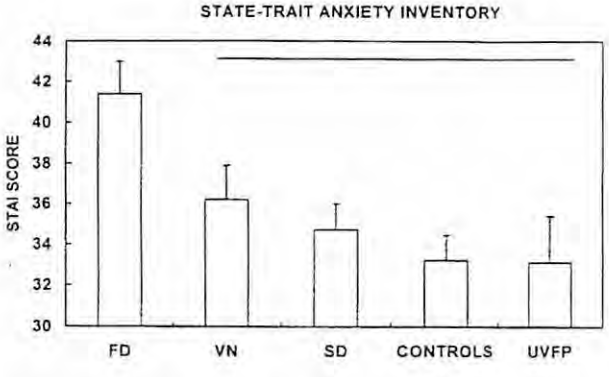


Figure 9. Rank-ordered group means and standard errors for scores obtained from the STAI (Trait scale).

ranged from 0.82 to 0.95, indicating acceptable internal consistency (FD 0.91; SD 0.87; VN 0.82; UVFP 0.95).

A total vocal handicap score for subjects was computed using their self-ratings from the voice handicap profile. The sum of the scores from the first 10 items was added to the severity rating scale to generate a total handicap score. A significant F test revealed omnibus differences between groups ($F(4, 151) = 34.78$, $p<.0001$). As expected, the non-voice-disordered control group scored significantly lower on the handicap scale when compared to all other voice-disordered groups ($p<.0001$). Pairwise LSD comparisons of group means also revealed that the UVFP group reported higher levels of overall handicap when compared to the VN group ($p<.026$). However, no other significant differences were identified among the other voice-disordered groups. Group means and standard errors for the voice disability profile are shown in Figure 11.

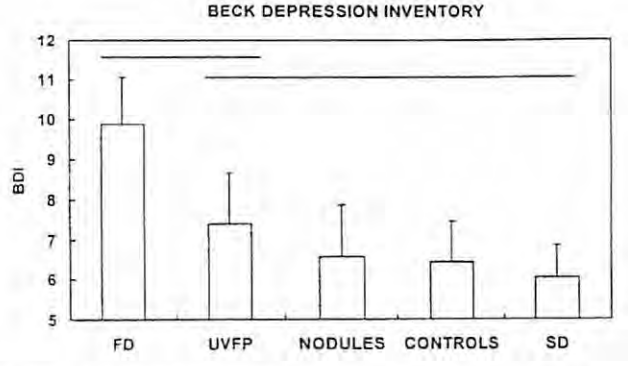


Figure 10. Rank-ordered group means and standard errors on the Beck Depression Inventory.

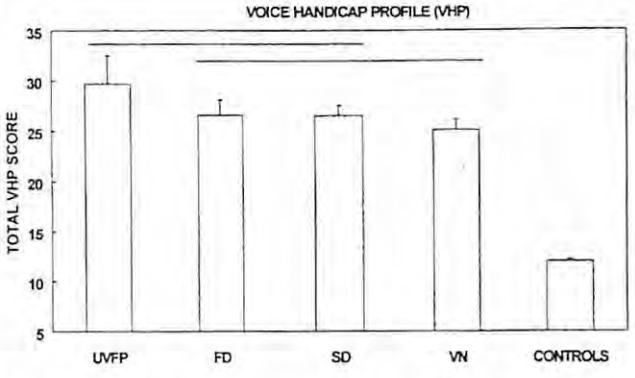


Figure 11. Rank-ordered group means and standard errors for total voice handicap profile scores.

To determine whether subjects' ratings of their level of vocal handicap were related to any of the scales of personality, anxiety, or depression, Spearman correlation coefficients were calculated for each voice disorder group. Although no interscale correlations achieved statistical significance, it is noteworthy that correlations between the total voice handicap score, the State-Trait Anxiety Inventory, and the Beck Depression Inventory scores did approach significance for the SD group only ($r = .30$, $p<.076$; $r = .30$, $p<.079$, respectively).

To determine whether the duration of vocal symptoms was related to self-ratings of vocal handicap, personality, and emotional adjustment, Spearman correlation coefficients were calculated for each group. Duration of voice symptoms was determined by medical chart review. For each subject, the time elapsed (in years) from the onset of the voice problem, to the date they participated in the research study, was calculated. The F test revealed omnibus differences between groups ($F(3, 123) = 5.711$, $p<.0011$). The SD group had longer-standing voice problems when compared to the other voice-disordered groups ($p<.05$) (see Figure 12). No significant correlations were detected across groups. To assess the influence of age on duration measures,

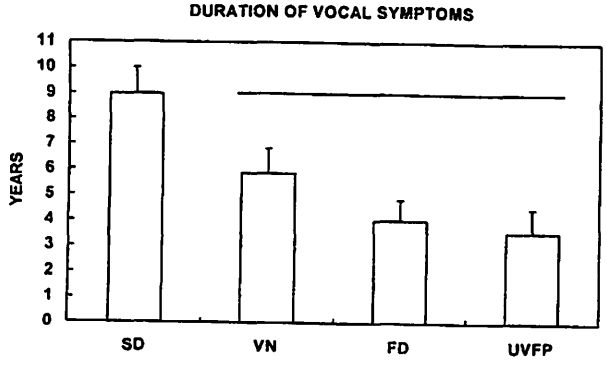


Figure 12. Rank-ordered group means and standard errors for duration of vocal symptoms.

the analysis was repeated with age entered as a covariate. A significant age effect was detected ($F(1, 122) = 8.207$, $p<.005$); however, Spearman correlations calculated with age partialed out did not change the results.

Discussion

It has been argued that personality, emotions, and psychological problems contribute to or are primary causes of voice disorders, and that voice disorders in turn create psychological problems and personality effects. This investigation compared a non-voice disordered otolaryngology control and four voice-disordered groups on self-report measures of personality and emotional adjustment. At the superfactor trait level, the FD and VN groups differed in significant ways from one another, from the other voice-disordered groups, and from the non-voice-disordered control group. Results largely support the contention that individuals with certain personality traits may be susceptible to developing FD or VN. In contrast, less support was found for the disability (scar) hypothesis, which argues that voice disorders lead to general personality changes. This raises the question as to how the results can be interpreted within the general theoretical framework presented in the companion article (this issue).

Personality Features of Individuals with FD

A fundamental principle of the Roy and Bless (in press) theory linking personality and FD is that subjects with FD should be neurotic (high N) introverts (low E). This combination of traits is said to produce heightened anxiety, inhibitory behavior of the laryngeal system, and elevated tension states in the context of behavioral inhibition system (BIS) inputs, such as uncertainty, threat, punishment, and frustrative nonreward. The discriminant analysis of EPQ superfactors, which compared the FD and control groups, provides the most compelling evidence supporting the FD-neurotic introvert hypothesis. Two personality factors, N and E, were identified as discriminating the FD subjects from the controls. Consistent with the hypotheses, FD subjects scored high on N and low on E (neurotic introverts).

The theory also asserts that most FD subjects should be classified as BIS-dominant introverts. This contention was also supported by the results. When E scores were dichotomized above and below the sample median, the majority of FD subjects (71%) were classified as low E, thereby furnishing support for the FD-introversion hypothesis. On the whole, these findings lend support for the low E-BIS dominance association in FD.

N is identified with Gray's nonspecific arousal system and is conceived as amplifying the stimulus sensitivities and response biases of the BIS. Eysenck (1967) suggested that neuroticism should be conceptualized as a propensity to react vigorously to all sorts of environmental stimuli. Evidence for the central role N plays in FD was furnished by discriminant analyses, which compared the FD group with all other groups in pairwise contrasts. In each contrast (with the notable exception of FD vs. VN), the superfactor N was identified as a powerful discriminatory variable. N alone successfully discriminated the FD groups from the SD and UVFP groups. Moreover, when E and N scores were dichotomized above and below the sample median, almost one-half of the FD subjects were classified as high N-low E, providing further support for the neurotic introvert hypothesis.

Another prediction of the theory was that the combination of high N and low E would be associated with elevated reports of anxiety. This prediction was also supported by the data. The FD groups scored significantly higher than all groups on the State-Trait Anxiety Inventory.

To summarize, the data revealed substantial differences in personality between the FD group, the other voice-disordered groups, and the non-voice-disordered control group at the superfactor level of analysis. Moreover, the majority of these differences conformed to the predictions of Roy and Bless (in press) regarding the FD group. The discussion is now turned to the subjects with vocal nodules.

Personality Features of Individuals with VN

The theory offered by Roy and Bless (in press) predicts that VN subjects, in accord with behavioral activation system (BAS) dominance, should demonstrate elevated scores on the Extraversion (E) scale. Analysis of the results at the superfactor level confirmed that the VN group did score higher on E than all of the voice-disordered groups. The importance of E was accented in both univariate and logistic discriminant analyses of the EPQ "Big Three" superfactors. Over three-quarters of the VN subjects (76%) were classified as high E when E scores were dichotomized into high and low E based on the sample median. The prominence of extraversion was also validated by the logistic discriminant analysis (with age excluded as a variable) whereby

high E was identified in every pairwise comparison involving the VN and voice-disordered groups.

Yano, Ichimura, Hoshino and Nozue (1982) also identified elevated extraversion among patients with vocal nodules and polyps when compared to two control groups (a vocally normal medical control group and vocal cord neoplasm group), and concluded that excessive voice use, which elevates the risk for developing vocal nodules, is a behavioral manifestation of an extraverted personality. Because Yano et al (1982) found elevated E scores even among industrial workers who had to communicate in noisy environments, they concluded that personality was as important as occupational voice use in nodule development. While the VN group in this study was clearly more extraverted than the other voice disordered groups, the difference between the VN and control groups on the EPQ E scale only approached significance.

Neuroticism plays a pivotal role in the theory of the dispositional bases of VN. N is believed to amplify the signal sensitivities and response biases of the BAS, causing extraverts to behave in a more extraverted and impulsive manner. The evidence for elevated N at the superfactor level is generally less robust. The difference between the VN group and the control group only approached statistical significance. However, the EPQ discriminant analysis comparing the VN and UVFP groups identified N and E as important discriminatory variables. In this analysis, the VN group scored higher on both E and N, which provided partial validation of the VN-N relationship.

The present theoretical formulation predicts that behavioral disinhibition (impulsivity) is the consequence of the conjunction of the extraversion and neuroticism traits. Furthermore, it is argued that this disinhibition is to some extent responsible for the development and maintenance of the vocal pathology. From this perspective, subjects with VN are unable to alter ongoing or dominant response sets (i.e., stop talking, vocal overuse-abuse) in the presence of salient social reward cues. At a superfactor level, features of disinhibition are captured in the EPQ Psychoticism scale (P). The theory holds that VN subjects should score low on constraint, or alternatively, high on P.

Examination of the results at the superfactor level supported the predictions of elevated P in the VN group. Subjects with VN were significantly less inhibited and scored highest on the EPQ P scale when compared to subjects in the other groups. This important finding was bolstered by the discriminant analysis of the EPQ superfactors, which identified elevated P as the personality dimension that best discriminated the VN group from the control group. Recall that P contains elements of impulsivity and disagreeableness (aggression). These results associate low Constraint with the VN group.

The role of impulsivity in the development and maintenance of vocal nodules has not received much attention in the voice literature. Impulsivity is admittedly a complicated construct, but most current descriptions of impulsivity include some form of *behavioral excess*, in the sense of doing something that potentially leads to trouble. The behavior is viewed as impulsive, because good judgment would suggest that it be inhibited (Fowles, 1987). Vocal nodules are thought to be related to forms of behavioral excess; for instance, excessive voice use and abuse. The theory espoused by Roy and Bless holds that when anticipating social rewards, VN subjects will fail to inhibit vocal overuse and abuse in spite of obvious signs of trouble, such as laryngeal discomfort and audible voice deterioration. Engaging in persistent vocal behavior in spite of its untoward effects is viewed as impulsive, because good judgement would suggest that it be inhibited. Green (1989) has identified similar impulsivity, hyperactivity, and distractibility in children with vocal nodules when compared with non-voice-disordered, age-matched peers. Thus, the association between impulsivity and vocal nodules seems to exist in both children and adults. This is an important finding worthy of further research attention.

In addition to disinhibition, the P scale of the EPQ is also assumed to capture features of elevated aggression (Eysenck & Eysenck, 1975b). It is not unreasonable to speculate that combining the traits of extraversion (sociability and dominance) and aggressiveness could contribute to abusive vocal patterns. For example, in order to be heard and noticed in social situations, VN subjects may vocalize excessively at elevated volumes and pitch. This pattern could be augmented by an aggressive verbal style, promoting further vocalization toward the extremes of loudness and pitch. Sustained vocalizing in this manner could result in vocal fold biomechanical stresses (i.e., increased collision forces) and mucosal changes, leading to vocal nodule development and maintenance (Hillman, Holmberg, Perkell, Walsh, Vaughan, 1989; Titze, 1994). This abusive pattern might be aggravated by an impulsive personality style whereby subjects with VN may not engage in stop/reflect behavior when the first signs of dysphonia or laryngeal discomfort appear.

Another interesting finding worthy of discussion is the discriminatory value of the L scale of the EPQ in distinguishing FD subjects from VN subjects. Eysenck and Eysenck (1975a) suggest that in addition to measuring an individual's tendency to respond in a socially desirable manner, "the L scale must measure some stable personality function; unfortunately little is known about the precise nature of this function (p.7)." Low scores on L suggest few attempts at impression management and a bias toward exaggerated self-disclosure. While neither the FD nor the VN groups scored differently than the controls on L, the VN group did score lowest of all groups and significantly lower when compared to both the FD and SD groups. Thus, the VN group seemed not to engage in impression management

and in fact tended to overdisclose. Likewise, Yano and coworkers (1982) reported a similar finding in that their VN/polyp group scored significantly lower on the L scale when compared to a non-voice-disordered control group. Unfortunately, the authors did not comment on the meaning of this low L scale result. The tendency to overdisclose is certainly an enticing finding worthy of further clarification.

To summarize, only partial support was found for the predictions of high N and high E leading to an impulsive personality style in the subjects with VN. While VN subjects as a group appeared extraverted and disinhibited based on measures of extraversion (high E) and constraint (high P), respectively, the evidence for elevated N at the superfactor level was generally weak.

On the whole, the subjects with FD and VN in this investigation fell on opposite ends of the extraversion-introversion continuum, with FD subjects scoring closest to the introverted pole and VN subjects scoring closest to the extraversion pole. Elevated scores on the neuroticism measure suggested amplified emotional reactivity in the FD group. The combination of high N and low E was interpreted to produce increased anxiety in the FD group, whereas high N and high E was interpreted as provoking impulsivity and aggressiveness in the VN group. Collectively, these results provide partial support for the central tenets of the Roy and Bless (in press) theory delineating the dispositional bases of both voice disorders.

Several caveats are in order, however. These results do not provide direct evidence of the signal sensitivities or response biases outlined in the Roy and Bless (in press) model. For instance, the proposition that BIS-dominant FD subjects are sensitive to threat or punishment cues and respond with passive avoidance was in no way tested in this research. Similarly, the belief that BAS-dominant VN subjects are sensitive to reward cues and engage in persistent approach behavior was not evaluated. Thus, the key mechanistic assertions of the theory await verification. It should be recognized, however, that Newman et al. have already completed substantial experimental research evaluating the signal sensitivities and response biases of these personality typologies (i.e., neurotic introverts and neurotic extraverts). A complete review of this literature is beyond the scope of this paper, but suffice it to say that many of the fundamental predictions of Newman's theoretical synthesis have received considerable empirical support (for a review see Wallace & Newman, 1991; Bachorowski & Newman, 1990; Newman & Nichols, 1986; Patterson, Kosson, & Newman, 1987; Wallace & Newman, 1990). Research needs to be undertaken to directly assess and confirm the existence of such biases in subjects with FD and VN.

Although group trends supported the hypotheses, it should also be noted that a number of subjects did not conform to the personality typologies described above. For example, several VN subjects were classified as stable introverts, which clearly runs counter to the predictions. It is evident that while personality seems to be an important consideration in the development of VN, it does not meet the requirements of *necessary or sufficient* causal status. If there is a causal relationship, it probably exists only in the presence of other physiological and vocal variables. It is reasonable to suggest that personality traits like extraversion and disinhibition may simply tip the balance toward the formation of nodules in the presence of a physiological or anatomical predisposition toward them. In individual cases, for example, the occupational voice use demands of singers, actors, aerobic instructors, and teachers may be the primary causal factor, and personality may play a minor role. In this study, the small number of VN subjects classified as introverts (n= 9) did not permit identification of occupational, medical, or demographic patterns/trends that might explain the origin of the lesions. Future research should continue to explore these exceptional cases to identify factors that might be causally significant.

If personality variables are related in some way to nodule formation, clarification of the physiological mechanism involved is required. Numerous medical factors influencing nodule development have been identified including allergies, respiratory tract abnormalities, endocrine imbalance, irritation due to fumes, and tobacco and alcohol consumption (Arnold, 1962; Baynes, 1966; Colton & Casper, 1996; Deal, McClain, & Sudderth, 1976; Herrington-Hall, Lee, Stemple, Niemi, & McHone, 1988; Kawase, Sawashima, Hirose, & Ushijima, 1982; Ohlsson, Jarvholm, Lofqvist, Naslund, & Stenborg, 1987; Senturia & Wilson, 1968 Silverman & Zimmer, 1975; Toohill, 1975). The complex interplay between personality, voice use patterns, and anatomical/physiological vulnerability and these medical factors deserves further study.

Voice Handicap and Personality, Depression and Anxiety

Because voice problems can be associated with a number of adverse consequences, including laryngeal discomfort, fatigue, and impairment of social and/or occupational functioning (with a concomitant loss of self-esteem and social support), it is not unreasonable to posit that chronic voice problems might lead to general personality changes, such as the development of heightened feelings of distress and dissatisfaction (high N) and social withdrawal (low E). If the disability hypothesis is correct, then personality measures, such as high N and low E in the case of FD, would simply be negative consequences of the voice disorder. However, close examination of the results provides weak support for this hypothesis. The personality scales N and E were unrelated to the self-reported level of vocal handicap. It appears then that the personality differences between groups were not related to the level of vocal handicap and

perhaps should not be viewed as a consequence of coping with voice problems.

By and large, the voice disordered groups did not differ substantially from one another on the voice handicap scale, yet significant differences were found between groups on the personality scales. For example, the FD, SD, and VN groups reported almost identical levels of vocal handicap, yet the VN subjects appeared extraverted, the FD subjects introverted, and the SD subjects ambiverted. If a voice disorder is responsible for producing social withdrawal (i.e. low E), the disability hypothesis cannot account for the high E tendencies in the VN group. Furthermore, the subjects with UVFP reported the highest voice handicap scores, yet they scored lowest on the N scale, indicating little emotional distress. Therefore, it appears that an individual can have a voice disorder and remain very outgoing and sociable (e.g. subjects with VN) and free of emotional distress (e.g. subjects with UVFP).

Additional evidence against the disability hypothesis exists in the data evaluating levels of depression and anxiety. While the FD group reported the highest levels of depression and anxiety, these affective problems were unrelated to their level of vocal handicap. Although no interscale correlations achieved statistical significance, it is noteworthy that for the *SD group only*, correlations between the total voice handicap score, the State-Trait Anxiety Inventory, and the Beck Depression Inventory scores did approach significance. Some researchers have taken the position that, when identified, such affective disturbances are the consequence of coping with voice difficulties (Murry, Cannito, & Woodson, 1994). This investigation, which employed a variety of voice disorder diagnoses, provides modest support for the outcome or disability hypothesis. Admittedly, however, the present methodology is only an *indirect* test of the disability hypothesis; future investigations should assess personality longitudinally to *directly* measure the effects of the voice disorder on personality and psychological adjustment.

Duration of Voice Symptoms and Personality, Depression and Anxiety

Another factor that may relate to the disability hypothesis is the length of time an individual must contend with the voice problem. Long-standing voice problems may negatively influence personality and psychological adjustment more than shorter duration problems. This contention was not supported by the results. None of the personality or adjustment scales significantly correlated with voice disorder duration.

Personality Differences among Individuals with SD and UVFP

While the dispositional theory offered by Roy and Bless (in press) generated several predictions regarding the salience of specific personality traits associated with FD and VN, it also predicted that personality *would not* be an important etiological factor in the disorders of SD and UVFP. In view of the presumed neuropathological origins of SD and UVFP, it was predicted that the personality profiles of these groups should not differ from each other or from the non-voice-disordered control group. Collectively, the results supported this prediction. Analysis of the EPQ superfactor results confirmed that the SD group did not differ from the controls or from the UVFP group on the E, N, or P scales. Almost identical results were obtained when comparing the UVFP group with the controls; however, the UVFP group did score lower on E than the controls ($p<.049$). Because of the small number of subjects in the UVFP group, these researchers are cautious when interpreting this modest relationship. While it is tempting to explain this finding as evidence to support the disability hypothesis, which proposes that low E (introversion) is an outcome of a communicative disability, more subjects are required to establish the validity of this effect. Further research may provide clarification.

The SD and UVFP groups did not differ from one another or from the controls on the anxiety or depression measures. These negative results concerning the presence of emotional factors in SD are at odds with the report of Cannito (1991) who identified elevated levels of anxiety, depression, and somatic complaints in a group of patients with SD. However, Cannito compared his experimental SD group to a *healthy* nondysphonic control group, while this investigator used a *medical* outpatient control group. As Cannito admitted, the affective differences that he observed perhaps could be explained on the basis of reaction to a medical disorder.

Furthermore, differences surrounding subject selection limits direct comparisons. Cannito included a variety of SD classifications, which were dominated by "mixed" and "abductor" SD subtypes (67%). Our investigation included subjects with the adductor subtype of SD only. This difference in group composition may partly explain the conflicting results. In addition, Cannito did not report the length of time subjects were symptomatic, or whether the SD subjects were receiving treatment at the time of assessment. Both of these factors may have influenced the results.

In a later study, Murry and colleagues (1994) reported elevated levels of depression and anxiety in a newly diagnosed adductor SD group prior to receiving their initial botulinum toxin injection. The authors employed the same data from Cannito's (1991) healthy control group for comparison purposes, thus rendering comparisons with our study difficult. The authors then dichotomized the SD group into depressed and non-depressed subtypes, further complicating matters. When the measures were readministered approximately 1 week and 2 months following botulinum toxin injection, only the "depressed" SD patients exhibited improvements in emotional adjustment. The authors concluded

that elevated depression and anxiety may be a result of acquiring SD. The absence of affective disturbance in our SD group may be related to (1) differences in control group selection, and (2) the fact that all SD subjects in this study were receiving botulinum toxin injections during the period of assessment, which may have moderated the "state" component of depression. Unlike FD and VN, we did not find any evidence to support personality as a potential risk factor in the development of SD, nor did we find evidence to support group elevations in anxiety or depression.

Further Comparisons with Existing Research

These results are consistent with the Roy, McGrory et al (1997) investigation that studied the personality-psychological correlates of a group of remitted subjects with FD. In spite of the fact that the FD patients had been successfully treated and were assessed several months post-treatment, they continued to report elevated anxiety, diffuse somatic symptoms, and introversion. The authors rejected the disability hypothesis as an explanatory construct, and offered that their findings were more consistent with a general trait of somatopsychic distress (i.e., elevated N). They argued that such a trait may represent dispositional vulnerability for the development, maintenance, and relapse of functional voice symptoms. The results from this investigation lend support to their dispositional vulnerability explanation.

These results are also compatible with the findings of Gerritsma (1991), who administered Wilde's Amsterdam Biographical Questionnaire to 82 patients with FD and aphonia.[3] The author identified elevated scores on the Neuroticism (N) and Neurotic Somatization (NS) scales, and low scores on the E scale. Interestingly, Gerritsma interpreted this finding by regarding aphonia/dysphonia "as a sign of extremely introverted reacting (p.17)." His interpretation, although not identical, shares some conceptual similarity with the theoretical formulation of nonspecific arousal system (NAS) potentiated BIS dominance in FD.

Clinical Implications

Although specific episodes of voice loss may be transient, documentation of a link between personality and voice pathology, combined with knowledge regarding the stability of personality through adulthood, suggests that in the absence of significant characterological change, personality may act as a persistent risk factor for voice pathology. A logical extension of this view is that the disorders of FD and VN should follow a pattern of remission and relapse given the persistent dispositional diathesis. It also prescribes that during the life course, voice problems in these two groups should not be singular isolated events. Rather, subjects should report a previous history of dysphonia, perhaps varying in severity. Although few studies of long-term treatment outcomes exist, the available research seems to substantiate these contentions, at least in the case of FD. In recent work (Roy, Bless, Heisey, & Ford, 1997), the results of behavioral voice therapy with a group of subjects with FD were evaluated. Despite initially responding favorably to behavioral treatment, 68% of subjects experienced some form of relapse during the 36-month follow-up period. Based upon patient report, these relapses were generally self-limiting, and less severe than the original dysphonia. In other work, Roy & Leeper (1993) also found that at least 80% of subjects with FD had reported/experienced previous episodes of voice loss, which is consistent with the notion that voice loss is not an isolated event. Gerritsma (1991) and Bridger & Epstein (1983) reported similar findings.

In the case of VN, few data exist evaluating long-term treatment outcomes. In clinical circles, clinicians often report frustration regarding lack of patient compliance with therapy advice, especially suggestions aimed at limiting the type and extent of voice use. Verdolini-Marston, Burke, Lessac, Glaze & Caldwell (1995) recently examined two forms of intervention for patients with VN. The authors concluded that the two treatments had similar effectiveness. For both therapies, the probability of a good outcome covaried directly with use of therapy techniques outside of the clinical setting. It seems reasonable to speculate that such extraclinical compliance might be influenced by patient's personality factors, such as extraversion, aggression, and impulsivity. One would assume that such factors might moderate a patient's response to therapy efforts. Thus it would be interesting to evaluate whether such personality variables are indeed related to treatment outcomes in patients with VN.

If voice therapies for vocal nodules are aimed at some form of behavioral inhibition (i.e., to reduce excess voice use and vocal abuse), an individual with a predilection toward behavioral activation and impulsivity may recognize the untoward effects of these vocal behaviors, but may be less able to inhibit such activity in the context of motivationally significant reward cues. Clinicians should recognize that voice therapy techniques that require inhibition of vocal behavior (e.g., modified voice rest, reduced loud talking) or even altering technical aspects of voice production, seem to be at odds with the personality and presumed behavioral inclinations of many patients with VN. This type of controlled processing is precisely what appears to be dysregulated in impulsivity. Therefore, the patient's inability to modify (self-regulate) such vocal behavior may

[3] The Amsterdam Biographical Questionnaire consists of 4 scales. The neuroticism (N) scale and (2) Neurotic Somatization (NS) scales measure neurotic instability as expressed through psychoneurotic and functional physical complaints, respectively; the Extraversion scale (E) measures social extraversion versus introversion; and the Test Defensively (T) scale measures self-defensive vs. self-critical test attitude (i.e., lie scale). This questionnaire was based on a precursor of the EPQ known as Maudsley Personality Inventory.

be unrelated to motivation and more directly related to enduring personality dispositions. This understanding would hopefully lead to methods that integrate more completely an individual's personality into the therapy plan. Recognizing the role of personality may also help clinicians to explain voice therapy failure.

From a clinical standpoint, it is clear that regardless of etiological interpretation the identification of emotional disorders in FD patients is critical for proper management. It is unlikely that lasting gains in behavioral voice therapy will be made with patients who remain clinically anxious and depressed (Butcher, Elias, Raven, Yateman, and Littlejohn, 1987). Such patients should be referred for psychological or psychiatric consultation regardless of whatever other behavioral or medical voice-related regimens may be indicated (Cannito, 1991).

The impressive discriminatory value of some of the personality scales also merits further discussion. Estimates of sensitivity and specificity sometimes exceeded 80% and thus provided excellent accuracy as potential diagnostic indices. For example, these scales may be potentially useful as a clinical tool to distinguish FD patients from SD and VN patients. Further research is needed to determine the inventory items that provided the most discriminatory power. The development of an abbreviated but reliable screening tool could raise the index of suspicion of one disorder type over another. At the very least, information regarding personality and psychological functioning could assist in the assessment and treatment process by identifying subjects in need of adjunctive psychological counseling and/or at risk for relapse.

Further Suggestions for Future Research

While the "Big Three" scales (E, N, CON/P) represent the highest order traits that reflect the most general level in the hierarchy of dispositions, relying solely on these composite superfactors can be misleading and fail to provide the necessary resolution to adequately describe personality. Because different levels of the trait hierarchy represent different levels of breadth or abstraction in personality description (Briggs, 1989; Costa & McCrae, 1995), decomposition of the superfactors into constituent traits affords a clearer analysis of both the type and range of the content subsumed within each of the broad factors. Analysis at a lower level of the hierarchy, which includes several component traits, can offer important information that is obscured at the highest level. Ideally then, future personality assessment should be conducted so as to survey different levels of the trait hierarchy (Hull, Lehn, & Tedlie, 1991).

Additional behavioral studies are needed with respect to the operation of both the BAS and BIS and their putative role in behavioral dysregulation in FD and VN. The current study was limited by its exclusive reliance on self-report measures of personality and psychopathology. Future studies should use multimethod assessments of personality and draw on information from multiple sources, such as family members, peers, and clinicians. The relations observed in the current study require replication with multimethod data in order to more effectively separate construct variance from method variance. Further research also is required to determine whether personality differences related to gender exist among these voice-disordered groups. Although males with FD and VN are a minority, it would be interesting to determine whether they share similar personality traits with their female counterparts.

For the past several decades, voice scientists and clinicians have essentially ignored the field of personality psychology. The results of this investigation suggest that the relation between personality and voice disorders merits serious attention for both practical and theoretical reasons. For instance, the relation between personality and long-term treatment outcomes in FD and VN needs to be investigated more fully. If personality represents an enduring factor in voice vulnerability, then the lingering question of whether personality influences can be moderated in any significant manner needs to be addressed. Identification of other predisposing anatomical or physiological factors in VN and FD may help define the interaction between personality and voice disorder vulnerability. Most voice therapy techniques focus on the overt disorder of phonation; until more is known of the etiologic factors/triggers, it may be unrealistic to expect great advances in long-term "cure" rates. The results of this investigation seem to suggest, as Moses (1954) did over 40 years ago, that exploring the characteristics of the "person" behind the voice may be as fruitful as studying the structure that produces it.

Acknowledgements

This work was supported in part by the National Center for Voice and Speech through Grant P60 00976 for the National Institute on Deafness and Other Communication Disorders. Parts of this manuscript were presented at the 1997 American Speech Language Hearing Association Annual Convention, Boston, MA. Please address correspondence to: Nelson Roy Ph.D., Department of Communication Disorders, 390 South 1530 East, Room 1219, Salt Lake City, Utah 84112-0252, (801) 585-0428, Fax: (801) 581-7955.

References

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

Bachorowski, J., & Newman, J. P. (1990). Impulsive motor behavior: The effects of personality and goal salience. *Journal of Personality and Social Psychology*, 58, 512-518.

Baynes, R. A. (1966). An incidence of chronic hoarseness among children. *Journal of Speech and Hearing Disorders*, 31, 172-176.

Bech, P., Gram, L. F., Dein, E., Jacobsen, O., Bitger, J., & Bolwig, T. G. (1975). Quantitative rating of depressive states. *Acta Psychiatrica Scandinavia*, 51, 161-170.

Beck, A. T., & Beamersderfer, A. (1974). Assessment of depression: The depression inventory. In P. Pichot, (Ed.), *Psychological Measurement*. Karger: Basel.

Beck, A. T., & Steer, R. A. (1993). *Beck Depression Inventory Manual*. San Antonio: The Psychological Corporation.

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

Bridger M. M., & Epstein, R. (1983). Functional voice disorders: A review of 109 patients. *Journal of Laryngology and Otology*, 97, 1145-1148.

Briggs, S. R. (1989). The optimal level of measurement for personality constructs. In D. M. Buss & N. Cantor (Eds.), *Personality Psychology: Recent Trends and Emerging Directions* (pp. 246 -260). New York: Springer-Verlag.

Butcher, P., Elias, A., Raven, R., Yeatman, J., & Littlejohn, 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.

Colton, R., & Casper, J. K. (1996). *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. Baltimore: Williams & Wilkins.

Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, 64, 21-50.

Crawford-Little, J., & McPhail, N. I. (1973). Measures of depressive mood at monthly intervals. *British Journal of Psychiatry*, 122, 447.

Davies, B., Burrows, G., & Poynton, C. A. (1975). Comparative study of four depression rating scales. *Australian and New Zealand Journal of Psychiatry*, 9, 21-24.

Deal, R. E., McClain, B., & Studderth, J. F. (1976). Identification, evaluation, therapy and follow-up for children with vocal nodules in a public school setting. *Journal of Speech and Hearing Disorders*, 41, 390-397.

Deary, I. J., Scott, S., Wilson, I. M., White, A., MacKenzie, K., & Wilson, J. A. (1997). Personality and psychological distress in dysphonia. *British Journal of Health Psychology*, 2, 333-341.

Eysenck, H. J. (1967). *Biological basis of personality*. Springfield, IL: Thomas.

Eysenck, H. J. & Eysenck, S. B. (1971). The orthogonality of psychoticism and neuroticism: A factorial study. *Perceptual and Motor Skills*, 33, 461-462.

Eysenck, H. J., & Eysenck, S. B. (1975a). *Manual of the Eysenck Personality Questionnaire*. San Diego, CA: Educational and Industrial Testing Service.

Eysenck, H. J., & Eysenck, S. B. (1975b). *Psychoticism as a dimension of personality*. London: Hodder & Stoughton.

Eysenck, S. B. & Eysenck, H. J. (1970). Crime and personality: An empirical study of the three factor theory. *British Journal of Criminology*, 9, 241-250.

Finitzo, T., & Freeman, F. (1989). Spasmodic dysphonia, whether and where: Results of seven years of research. *Journal of Speech and Hearing Research*, 32, 541-555.

Fowles, D. C. (1987). Application of a behavioral theory of motivation to the concepts of anxiety and impulsivity. *Journal of Research in Personality*, 21, 417-435.

Gerritsma, E. J. (1991). An investigation into some personality characteristics of patients with psychogenic aphonia and dysphonia. *Folia Phoniatrica*, 43, 13-20.

Gray, J. A. (1975). *Elements of a Two Process Theory of Learning*. London: Academic Press.

Green, G. (1989). Psychobehavioral characteristics of children with vocal nodules: WPBIC ratings. *Journal of Speech and Hearing Disorders*, 54, 306-312.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.

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.

Hull, J. G., Lehn, D. A., & Tedlie, J. C. (1991). A general approach to testing multifaceted personality constructs. *Journal of Personality and Social Psychology*, 61, 932-945.

Izdebski, K., & Dedo, H. H. (1981). Spastic dysphonia. In J. Darby (Ed.), *Speech evaluation in medicine (pp.105-127)*. New York: Grune & Stratton.

Kawase, N., Sawashima, M., Hirose, H., & Ushijima, T. (1982). A statistical study of vocal cord nodule, vocal cord polyp and polypoid vocal cord with special reference to the physical and social histories of patients. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (Tokyo)*, 16, 235-245.

Koufman, J. A., & Blalock, P. D. (1982). Classification and approach to patients with functional voice disorders. *Annals of Otology, Rhinology, Laryngology*, 91, 372-377.

Lewis, G. W. (1958). Zoster sine heptete. *British Medicine Journal*, 2, 418-421.

Ludlow, C. L., & Connor, N. P. (1987). Dynamic aspects of phonatory control in spasmodic dysphonia. *Journal of Speech and Hearing Research*, 30, 197-206.

Ludlow, C. L., Naunton, R. F., & Bassich, C. J. (1984). Procedures for selection of spastic dysphonia patients for recurrent laryngeal nerve section. *Otolaryngology, Head & Neck Surgery*, 92, 24-31.

Metcalfe, M., & Goldman, E. (1965). Validation of an inventory for measuring depression. *British Journal of Psychiatry*, 111, 239-245.

Moses P. J. (1954). *The Voice of Neurosis*. New York: Grune & Stratton.

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.

Newman, J. P., & Nichols, S. L. (1986). Effects of punishment on response latency in extraverts. *Journal of Personality and Social Psychology*, 50, 624-630.

Ohlsson, A. C., Jarvholm, B., Lofqvist, A., Naslund, P. E., & Stenborg, R. (1987). Vocal behavior in welders: A preliminary study. *Folia Phoniatrica*, 39, 98-103.

Parnes, S. M., Lavarato, A. B., & Myers, E. N. (1979). Study of spastic dysphonia using videofiberoptic laryngoscopy. Annals of Otology, Rhinology and Laryngology, 87, 322.

Patterson, C. M., Kosson, D. S. & Newman, J. P. (1987). Reaction to punishment, reflectivity, and passive avoidance learning in extraverts. *Journal of Personality and Social Psychology*, 52, 565-575.

Roy, N., & Bless, D. M. (in press). Toward a theory of the dispositional bases of functional dysphonia and vocal nodules: Exploring the role of personality and emotional adjustment. In R. D. Kent & M. J. Ball (Eds.), *The Handbook of Voice Quality Measurement*. San Diego: Singular Publishing Group.

Roy, N., Bless, D. M., Heisey, D., & Ford, C. F. (1997). Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short- and long-term treatment outcomes. *Journal of Voice*, 11, 321-331.

Roy, N., & Leeper, H. A. (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., Tasko, S. M., Bless, D. M., Heisey, D., & Ford, C. N. (1997). Psychological correlates of functional dysphonia: An evaluation using the Minnesota Multiphasic Personality Inventory. *Journal of Voice*, 11, 443-451.

Schwab, J. J., Bialow, M. R., & Holzer, C. E. (1967). A comparison of two rating scales for depression. *Journal of Clinical Psychology*, 23, 94-96.

Senturia, B. H., & Wilson, F. B. (1968). Otorhinolaryngologic findings in children with voice deviations. *Annals of Otology, Rhinology, and Laryngology*, 77, 1027-1041.

Silverman, E. M., & Zimmer, C. H. (1975). Incidence of chronic hoarseness among school children. *Journal of Speech and Hearing Disorders*, 40, 211-215.

Spielberger, G. D., Gorusch, R. L., Lushene, R., Vagg, P., & Jacobs, G. A. (1983). *Manual for the State-trait Anxiety Inventory (Form Y Self-Evaluation Questionnaire)*. Palo Alto, Calif: Consulting Psychologists Press.

Titze, I. R. (1994). *Priniciples of voice production*. Englewood Cliffs, NJ: Prentice-Hall.

Toohill, R. J. (1975). The psychosomatic aspects of children with vocal nodules. *Archives of Otolaryngology*, 101, 591-595.

Verdolini-Marston, K. L., Burke, M. K., Lessac, A., Glaze, L., & Caldwell, E. (1995). Preliminary study of two methods of treatment for laryngeal nodules. *Journal of Voice*, 9, 74-85.

Verma, R. M., & Eysenck, H. J. (1973). Severity and type of psychotic illness as a function of personality. *British Journal of Psychiatry*, 122, 573-585.

Wallace, J. F., & Newman, J. P. (1990). Differential effects of reward and punishment on response speed in anxious and impulsive individuals. *Personality and Individual Differences*, 11, 999-1009.

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.

Yano, J. L., 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.

Appendix A

Voice Handicap Profile

Read each statement and then circle the number that best indicates how you presently feel about you voice.

1=Almost never 2=Sometimes 3=Often 4=Almost Always

1. I have to alter daily activities because of my voice.
   1 2 3 4

2. My voice interferes with communication.
   1 2 3 4

3. My voice is distracting to others.
   1 2 3 4

4. It is difficult for others to hear me in noisy environments.
   1 2 3 4

5. My voice gets tired during the day.
   1 2 3 4

6. It takes a lot of energy to produce voice.
   1 2 3 4

7. I miss work because of my voice.
   1 2 3 4

8. I think about my voice problem.
   1 2 3 4

9. My voice sounds worse than other speakers.
   1 2 3 4

10. People make comments about my voice.
    1 2 3 4

11. On a scale of 1 to 7, where 1= normal, 4 = moderate impairment, and 7 = severe impairment, circle the number that best describes how bad your voice is.

    1 2 3 4 5 6 7
    Normal Moderate Severe

NCVS Status and Progress Report - 14
September 1999, 221-226

Preparation of the Speech-Language Pathologist Specializing in Voice: An Educational Survey

Miriam vanMersbergen, M.A., CCC-SLP
Loyala University Health System
Medical College of Wisconsin Physicians and Clinics
Julie Ostrem, B.S.
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology

Abstract

This study investigates the academic and clinical preparation a speech-language pathologist receives in the area of voice. Surveys were sent to the graduate programs in speech-language pathology accredited by the American Speech-language Hearing Association in 1994, and again in 1999. Results from the 1994 survey indicated that students received limited information and clinical experience in voice. Although most programs required a voice disorders course, fewer mandated coursework in normal voice production or clinical experience with voice patients. The follow-up survey showed that the educational environment in 1999 is similar to 1994 results.

Introduction

The field of speech-language pathology serves patients with a wide spectrum of disorders. A certified clinician may provide therapy to patients with conditions as diverse as traumatic brain injury, dysfluency, or voice disorders. The complex nature of communication and the accelerating pace of research generate a virtually limitless bank of information for the practitioner to absorb. This need for a wide and varied knowledge base presents a challenge to the educational institutions training speech-language pathologists in master's programs approximately two years in length.

The entire profession finds itself in debates discussing this dilemma. Members of the American Speech-Language-Hearing Association (ASHA) have expressed an obvious and immediate interest in these discussions. In 1988, ASHA established a framework for members to organize special interest divisions to exchange information on current topics – including academic preparation - with colleagues with common professional interests. The Speech Science and Orofacial Disorders (Special Interest Division #5) sought to gauge the academic preparation of those interested in speech science. The SID#5 study concluded that speech science instruction tended to be introductory in nature, with scanty coursework offered at the graduate level.[1]

Special Interest Division #3 (Voice and Voice Disorders) members have likewise been involved in debates concerning academic preparation of speech-language pathologists who wish to specialize in voice. Many recent discussions and input from members have focused on the pros and cons of specialty recognition in voice within the field of speech-language pathology.[2]

The adequacy of master's degree training has been in question by other specialty areas of speech-language pathology as well. In 1992, the National Center for Neurogenic Communication Disorders presented a two-part Telerounds broadcast. This presentation included a panel of clinicians and academicians who debated the standard of training programs preparing speech-language pathology students to practice in medical settings. While no conclusions were drawn, salient arguments included: the level of quality care the patient and hospitals expect from speech-language pathologists; the responsibility of educational institutions in preparing competent practitioners; and the cost of education compared to starting salaries of speech-language pathologists.[3]

Other investigations of the adequacy of speech-language pathology training programs have focused on the treatment of those with fluency disorders and of tracheotomy patients. In a study by Mallard, Gardner, and Downey[4], 155 school-based clinicians were asked to report their clinical preparation for treatment of various communication disorders. Not only did the majority of clinicians report lacking the confidence and desire to work with patients who stutter, the group identified a lack of clinical training as a primary reason for this difficulty. The study also found that these clinicians rated voice lower than stuttering on confidence/desire scales. Manley, Frank and, Melvin[5] reported that only half of medically-based speech-language pathologists surveyed felt comfortable treating patients with tracheostomies. The authors reported that the clinicians with the greatest confidence in servicing this population had both strong academic and clinical training.

Indeed, debates concerning the training of other voice professionals are currently in progress. In the training of teachers of singing, questions have arisen about the quantity, quality and appropriate timing of instruction.[6] The Voice and Speech Trainers Association, a professional organization for theater voice and dialect coaches, issued a statement regarding the optimal training for a theater voice coach. VASTA's recommendation includes both academic and practical internships.[7]

To further complicate matters in this debate, past surveys have suggested that few speech-language pathologists feel comfortable or even desire to treat patients with voice disorders. St. Louis' and Durrenberger's 1993 study found that speech-language pathologists found voice one of the least desirable disorders to treat.[8] The clinicians cited reasons for this low desirability to treat voice disorders as a lack of training and experience. The St. Louis study also confirms that practitioners serving those with voice and fluency disorders are the minority in the field of speech-language pathology.

One suggested solution is specialty training and recognition of that training. For those who are concerned about specialty training in voice and other areas, there is disagreement about the optimal placement within a speech-language pathologist's career. Some believe specialty training should be integrated into the master's program as supplemental clinic hours, expansion to three-year graduate programs, or the development of clinical doctorates. Required internships have also been suggested. Others have proposed additional post-graduate training, such as extended clinical fellowship years or specialty training and certification. Finally, others believe specialty training can be obtained through continuing education programs after attainment of licensure.

Each of the proposed solutions carries with it inherent drawbacks. Starting salaries for new speech-language pathology graduates make it difficult to justify extended master's programs. In 1997, the average starting salary for speech-language pathologists was $30,000 for an academic year and $38,000 for a calendar year.[9] It is feared that lengthy and expensive graduate programs coupled with low salaries will discourage high-quality candidates to the discipline. Additionally, there are concerns about the availability of qualified supervisors for internships, extended clinical fellowships or other apprenticeships. Continuing education opportunities may be sparse, expensive, or theoretical rather than practical in nature. Additionally, continuing education programs may not be adequately targeted for either the new clinician attempting to build a knowledge base for specialization recognition.

In 1996, ASHA reported that the majority of respondents to a questionnaire and membership priority ratings did not see a need for specialty recognition in voice.[10] One of the reasons cited was a need within the profession for "generalists." Third-party payers and health maintenance organizations may reinforce this point of view by preferring to pay generalist fee rates to specialist fees.

At the crux of the issue is the quality of care available to clients who require the services of competent and confident voice therapist. A study by Smith et al[11] documented that a majority of patients with voice disorders reported adverse effects on the professional, social and psychological facets of their lives. These negative effects argue a case for competent and confident professionals to treat voice disorders. How well is the educational system preparing the new speech-language pathologist to assist the client regain quality of life?

The current study addresses a simple objective: to determine the baseline understanding of normal voice production, voice disorders, and voice therapy techniques for the new master's graduate in speech-language pathology. The study was repeated following a five-year interval to assess educational trends in light of the current debates surrounding training. To date, there have been no previous organized studies addressing academic preparation in the area of voice.

Methods

Data Collection

The 215 graduate programs in speech-language pathology, registered with ASHA as of May 1994, and the 207 graduate programs registered with ASHA as of January 1999 were solicited with letters and surveys. The initial cover letter explained that the purpose was to gather knowledge about graduate training in the area of voice and voice disorders so that continuing educational materials could be developed for new practitioners. A letter included with the second survey explained a similar purpose, including the

intention of comparing results with those from 1994. The questionnaires from 1994 and 1999 were identical in content. A copy of the current, complimentary Guide to Vocology from the National Center for Voice and Speech was included in both surveys as an appreciation for participating in the survey.[12, 13]

Thirty-eight schools responded to the first survey within a three-month period of time. In order to increase the response rate, non-respondents were telephoned and the questionnaire was completed orally. Questionnaires from thirty-one schools were completed through the telephone survey. Thus, the total response was 69, or 32 percent.

The second set of surveys were mailed January 2, 1999, to the chairs of the 207 accredited programs of speech-language pathology in the United States. By April 27, 1999, 71 surveys had been completed and returned, for a response rate of 34 percent.

Data Analysis

Data were entered into Microsoft Excel for arithmetic and statistical processing. Data were then analyzed in three areas: 1) student interest, 2) academic training, and 3) clinical training.

Student interest questions required respondents to state the number of graduate students in their programs, the number of students interested in voice and voice disorders, and the number of students with backgrounds in vocal performance. The criteria for a student with a primary interest in voice included those students who opted for a masters' theses in voice, asked for extra clinical hours in voice disorders, or stated preferences to advisors or voice instructors. A background in voice performance was defined as current activity in the music, theater, or broadcasting field, or as participation in advanced training in music, theater, or broadcasting. Each school's response was tallied for each question. The number of graduates interested in voice was divided by the total number of graduate students. Likewise, the number of graduate students with a vocal performance background was divided by the total number of graduate students.

Academic training questions required respondents to choose the number of credit hours offered and required in voice *production* and the number of credit hours taken by students. Credit hour choices included four categories: no credit hours, 1-3 credit hours, 4-6 credit hours, and 7+ credit hours. Academic training questions also required respondents to select the number of credit offered and required in voice *disorders* and the number of credit hours taken by students. Again, credit hour choices included: no credit hours, 1-3 credit hours, 4-6 credit hours, and 7+ credit hours. Responses were tallied under each category, and the percentage of programs within each category was calculated.

Finally, clinical training questions addressed two areas: the ASHA clinical practica hours in voice disorders required by each program and the average ASHA clinical practica hours taken by students. Because ASHA no longer specifically requires any voice hours for school accreditation, each program was queried regarding internal requirements. Respondents were asked to choose the number of clinical hours required and taken. Clinical hour choices included: none, 5-10 clinical hours, 11-15 clinical hours, 16-25 clinical hours, and 25+ clinical hours. Responses were tallied under "hours required" and "hours taken" and number of programs (in percent) was calculated for each category.

Results

Student Interest

Table 1 reflects student interest in voice across the two time periods. In 1994, 14.7 percent of all graduate students met at least one requirement defining them as "having an interest in voice". In 1999, 10 percent of all graduate students met at least one of the student interest criteria. The 1994 survey reported 3.7 percent of all students within the programs possessed backgrounds in voice, as compared to the 1999 result of 2.5 percent.

Academic Training

Voice Production. Figure 1 (following page) depicts the comparison of academic coursework *offered* in voice production in 1994 and in 1999. About half (52 percent) of the programs surveyed offer 1-3 credit hours in voice production for both time periods. Interestingly, the percentage of programs offering *no* credit hours in voice production increased from 1994 to 1999 (9 to 13 percent). Only 5 programs of 71 (7 percent) in the 1999 survey stated that they offer a course solely devoted to voice production; in 1994, the total was 2 programs of 69 (3 percent). The other programs offered voice production information in combination with coursework in voice disorders, speech science, or anatomy and physiology classes.

Table 1.
Percentage of Students from the Programs Surveyed With a Primary Interest in Voice Disorders and With a Background in Voice Performance in 1999 and in 1994

	1999	1994
Student Interest in Voice Disorders	10%	14.7%
Students with Backgrounds in Voice Performance	2.5%	3.7%

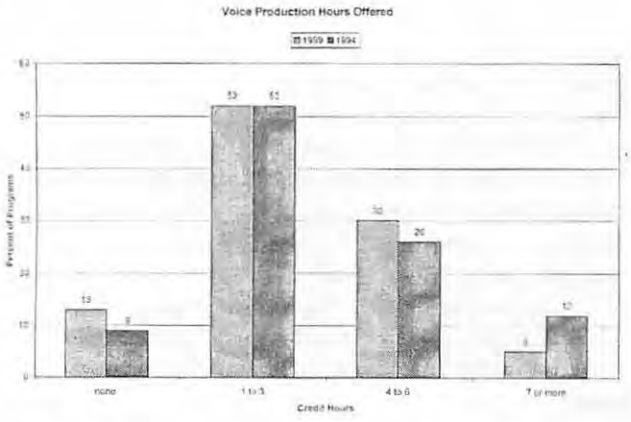


Figure 1. Academic coursework, in credit hours, offered *in voice production in 1999 and in 1994. Percentage of programs surveyed in 1999 and 1994 for each of the categories: none, 1-3 credit hours, 4-6 credit hours, and 7+ credit hours.*

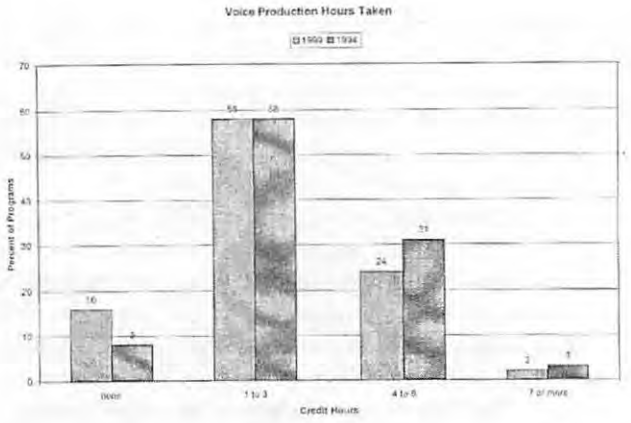


Figure 3. Academic coursework, in credit hours, taken *in voice production in 1999 and in 1994. Percentage of programs surveyed in 1999 and 1994 for each of the categories: none, 1-3 credit hours, 4-6 credit hours, and 7+ credit hours.*

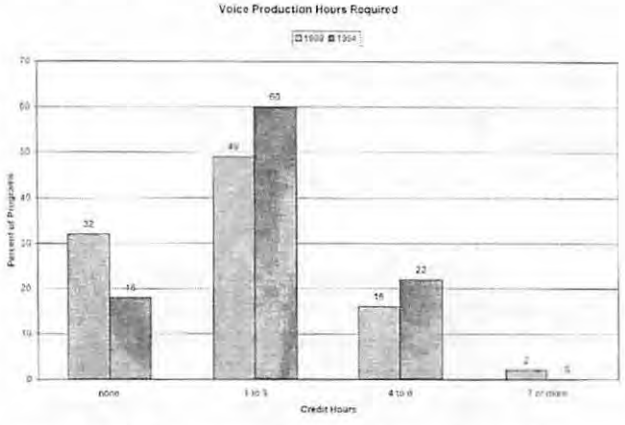


Figure 2. Academic coursework, in credit hours, required *in voice production in 1999 and in 1994. Percentage of programs surveyed in 1999 and 1994 for each of the categories: none, 1-3 credit hours, 4-6 credit hours, and 7+ credit hours.*

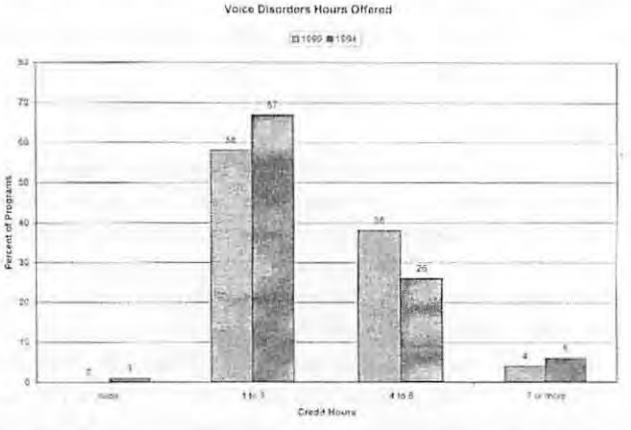


Figure 4. Academic coursework, in credit hours, offered *in voice disorders in 1999 and in 1994. Percentage of programs surveyed in 1999 and 1994 for each of the categories: none, 1-3 credit hours, 4-6 credit hours, and 7+ credit hours.*

Figure 2 represents the findings for academic coursework in voice *required* by graduate speech pathology programs in 1994 and 1999. In 1994, 60 percent of programs required 1-3 credit hours in voice production as compared to 49 percent in 1999. Nearly one-third of the programs currently does not require coursework in voice production.

Figure 3 represents the findings for academic coursework *taken* by students in voice production for both the 1999 and 1994 surveys. Results for 1994 and 1999 were identical, with 58 percent of students taking 1-3 credit hours in voice production. Of note, the percentage of students taking *no* credit hours in voice production rose between 1994 and 1999 (a difference of 9 percent).

Voice Disorders. Figure 4 represents the findings for academic coursework *offered* in voice disorders for both the 1999 and the 1994 surveys. In 1994, two-thirds (67 percent) of programs required 1-3 credit hours in voice production as compared to 58 percent of the programs with such requirements in 1999. Only rarely (1 percent in 1999 and 0 percent in 1994), does an academic program have no offerings in voice disorders.

Figure 5 represents the findings for academic coursework *required* in voice disorders for both the 1999 and the 1994 surveys. Findings were similar across the two time periods, although fewer programs in 1999 do not require voice disorders credits for graduation (10 percent versus 17 percent in 1994). About two-thirds (67 percent in 1999 and 64 percent in 1994) of the programs require 1-3 credit hours of voice disorders classwork.

Figure 6 represents findings for academic coursework *taken* by students during both time periods.

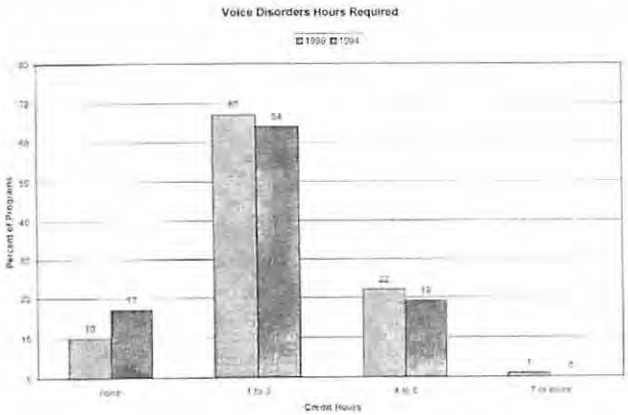


Figure 5. Academic coursework, in credit hours, required *in voice disorders in 1999 and in 1994. Percentage of programs surveyed in 1999 and 1994 for each of the categories: none, 1-3 credit hours, 4-6 credit hours, and 7+ credit hours.*

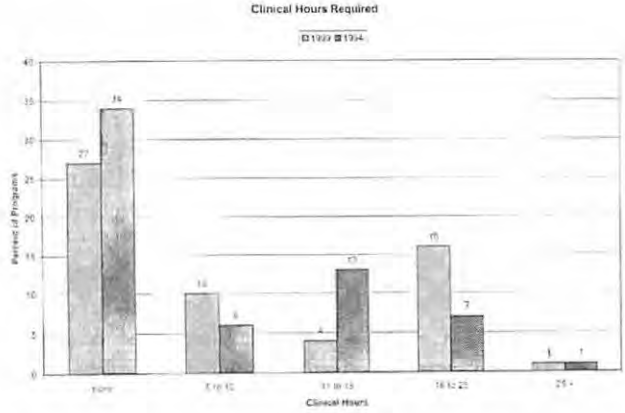


Figure 7. ASHA clinical hours, required *in voice diagnostic and treatment in voice disorders in 1999 and in 1994. Percentage of programs surveyed in 1999 and 1994 for each of the categories: none, 5-10 credit hours, 11-15 credit hours, and 25+ credit hours.*

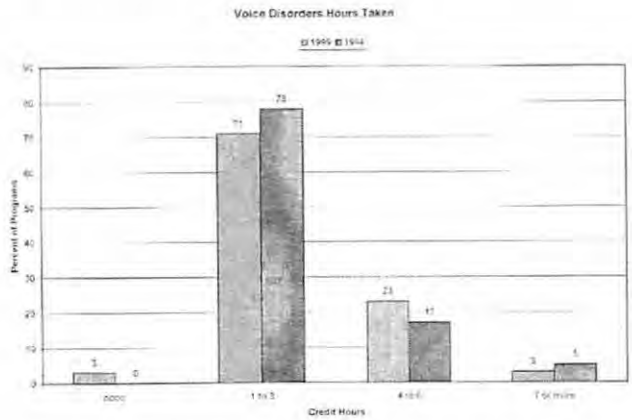


Figure 6. Academic coursework, in credit hours, taken *in voice disorders in 1999 and in 1994. Percentage of programs surveyed in 1999 and 1994 for each of the categories: none, 1-3 credit hours, 4-6 credit hours, and 7+ credit hours.*

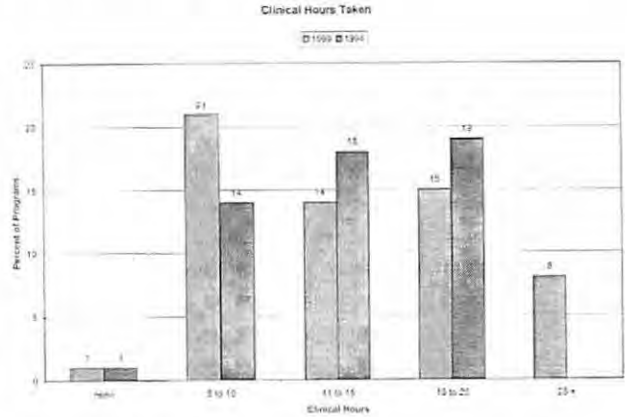


Figure 8. ASHA clinical hours, taken *in voice diagnostic and treatment in voice disorders in 1999 and in 1994. Percentage of programs surveyed in 1999 and 1994 for each of the categories: none, 5-10 credit hours, 11-15 credit hours, and 25+ credit hours.*

Across both time periods, the majority of students (71 percent in 1999 and 78 percent in 1994) take between 1-3 credits in voice disorders.

Clinical Training

Figure 7 represents the findings for ASHA clinical practica hours required in voice disorders for both the 1999 and 1994 surveys. There were few differences across the time periods. Interestingly, however, nearly a third of students (27 percent in 1999 and 34 percent in 1994) could graduate without *any* clinical voice experience. Very few programs (1 percent for each time period) require 25 clinical voice hours or more.

Figure 8 represents the findings for ASHA clinical practica hours taken by students. Almost all students receive some clinical exposure to voice patients. Interestingly, the trend is moving toward the extremes – in 1999, there was an increase in students with only 5-10 clinic hours (21 percent versus 14 percent) and an increase in students with 25 or more hours of clinical experience (8 percent versus 0 percent).

Discussion

From a statistical perspective, drawing conclusions from a simple survey leads to certain caveats. First, no inferences can be made about a general population based on the survey's inherent bias. Specifically, those who responded to the survey may have stronger programs in voice than non-respondents. Furthermore, any statistical comparisons between 1994 and 1999 would be difficult since no inference to a general population can be responsibly made. Second, survey answers to academic and clinical questions were based on numerical ranges rather than specific numbers to

bolster the ease of completing the survey, and thus, increasing the response rate. Finally, respondent error cannot be statistically managed and therefore must be factored into any conclusion or observation made from the study.

However, some general observations can be noted.

Academic offerings are nearly identical to coursework actually taken, despite requirements imposed by the program. This comparison is consistent across both time periods and true for both voice production and voice disorders coursework. This leads to the question: would graduate speech pathology students take more voice-oriented courses if they were offered? Does a lack of faculty members qualified to teach voice courses limit the amount of available opportunities in voice study?

Secondly, the students' higher tendency in the 1999 survey to select minimal (5-10 hours) or significant (25 hours or more) clinical hours with voice patients raises a provocative question. Can this be interpreted as evidence that students do seek as much specialty training - either in voice disorders or away from voice disorders - as they are able to get within their academic programs?

Finally, reviewing requirements (Figures 2, 5 and 7) established by academic programs, many new speech-language pathologist may graduate without coursework devoted to voice production or voice disorders and without experience with voice clients. Many others in the current academic climate receive only limited exposure to voice in a course focusing on disorders. Although courses in voice disorders provide the future clinician with acceptable technical skills, we believe these courses do not provide a broad enough understanding of the vocal mechanisms. This knowledge gap could prevent the clinician from critically evaluating new therapies. It also could limit the practitioner's ability to keep abreast of research findings through peer-reviewed journals and professional meetings.

A simple survey such as this may raise more questions than it answers, even those posed in the introductory discussion. Speech-language pathology has reached a critical stage in its evolution as a health care discipline. The acceleration of knowledge in many subspecialties – neurogenics, language, and other areas in addition to voice – has overflowed the limitations of a two-year master's program. Institutions of higher learning cannot be expected to graduate new clinicians well versed in all areas. Furthermore, with the ever growing need for proof of treatment efficacy, how can the profession as a whole insure that its practitioners indeed understand the disorders and adequately implement the treatments of voice –or any other disorder- with only minimal, basic, exposure? Until post-graduate programs such as specialty internships or expanded clinical fellowships can be established, innovative continuing education programs such as those offered by the Internet and other distance learning mechanisms provide the first step to improve the care of services to the voice client.

Acknoweldgements

The authors gratefully acknowledge financial support from Grant P60 DC00976 from the National Institute on Deafness and Other Communication Disorders, a division of the National Institutes of Health. They also express thanks to Penelope Hall, Ph.D. and Cynthia Fix, MA, CCC-SLP, for reviewing a preliminary version of the paper, and to Julie Stark, AA, for her assistance in collecting and organizing data.

References

1. SID#5 Speech Science and Orofacial Disorders Newsletter. Rockville, Maryland:ASHA, 1995.

2. SID#3 Voice and Voice Disorders Newsletter. Rockville, Maryland:ASHA, September 1994.

3. Are University Training Programs Adequately Preparing Students for Speech Language Pathology in Medical Settings? National Center for Neurogenic Communication Disorders Telerounds, University of Arizona Board of Regents, #4 (10-7-92) and #6 (12-9-92).

4. Mallard AR, Gardner LS, Downey CS. Clinical training in stuttering for school clinicians. J Fluency Disord 1998;13:253-259.

5. Manley SB, Frank EM, Melvin CF. Preparation of speech-language pathologists to provide services to patients with a tracheostomy tube: A survey. Amer J Speech-Lang Path 1999;8:2:171-180.

6. Hollien H. That golden voice - Talent or training? J Voice 1993;7:195-205.

7. VASTA Guidelines for the Preparation of Voice and Speech Teachers. Voice and Speech Trainers Association, 1995.

8. Salary Report, Speech-Language Pathologists & Audiologists, Rockville MD: American Speech-Language Hearing Association, 1997.

9. SID#3 Voice and Voice Disorders Newsletter, Rockville, Maryland:ASHA, June 1996.

10. St Louis KO, Durrenberger CH. What communication disorders do experienced clinicians prefer to manage? ASHA, December 1993.

11. Smith E, Verdolini K, Gray S, Nichols S, Lemke J, Barkmeier J, Dove H, Hoffman H. Effect of voice disorders on quality of life. J Med Speech-Lang Path 1996;4:223-244.

12. Verdolini, K, DeVore K, Ostrem, J. Guide to Vocology. Iowa City, IA: The National Center for Voice and Speech, 1993.

13. Verdolini, K, DeVore K, McCoy, S, Ostrem, J. Guide to Vocology. Iowa City, IA: The National Center for Voice and Speech, 1998.

CUMULATIVE AUTHOR INDEX

VOLUMES 1-14

NCVS Status and Progress Report - 14
September 1999, 227-230

NCVS Status and Progress Report
Cumulative Author Index, Volumes 1-14