NCVS Status and Progress Report
Volume 6/May 1994

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

Editorial and Distribution Information

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

Distribution of this report is not restricted.
However, production was limited to 800 copies.

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

Primary Sponsorship

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

Other Sponsorship

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

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

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

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

Memphis State University
Department of Audiology and Speech Pathology

The University of Illinois
Department of Speech and Hearing Science

NCVS Personnel

Administration

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

Area Coordinators
Research - Ingo Titze
Training - Erich Luschei
Continuing Education - Julie Ostrem
Dissemination - Barbara Bustillos

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

Investigators, Affiliates and Support Staff

Fariborz Alipour, Ph.D.
Julie Barkmeier, Ph.D.
David Berry, Ph.D.
Andrea Beckmann, B.S.
Patricia Benjamin, B.S.
Florence Blager, Ph.D.
Diane Bless, Ph.D.
James Brandenburg, M.D.
Barbara Bustillos
Stefanie Countryman, M.A.
Heather Dove, M.A.
Christopher Dromey, M.A.
Wendy Edwards, B.A.
Penny Farrell
Charles Ford, M.D.
Steven Gray, M.D.
Judith Grayhack, Ph.D.
Chwen-Geng Guo, M.S.
Marilyn Hetzel, Ph.D.
Michelle Hodge
Margaret Hoehn, M.D.
Henry Hoffman, M.D.
Yoshiyuki Horii, Ph.D.
Antonia Johnson, B.S.
Karen Jones, B.S.
Joel Kahane, Ph.D.
Judith King, Ph.D.
David Kuehn, Ph.D.
Jon Lemke, Ph.D.
Julie Lemke
Erich Luschei, Ph.D.
Sharon Maclay, M.A.
Brian McCabe, M.D.
Jim McKenna, B.A.
Martin Milder, B.S.
Paul Milenkovic, Ph.D.
Kenneth Moll, Ph.D.
Jerald Moon, Ph.D.
John Nichols, B.A.
Lorraine Olson Ramig, Ph.D.
Julie Ostrem., B.S.
Namrata Patil, M.D.
Annette Pawlas, M.A.
Mark Peters
Donald Robin, Ph.D.
Dawn Salyards, B.S.
Jackie Schalow, B.S.
Ronald Scherer, Ph.D.
Elaine Smith, Ph.D.
Marshall Smith, M.D.
Nancy Pearl Solomon, Ph.D.
Leatitia Thompson, Ph.D.
Ingo Titze, Ph.D.
Nancy Tye-Murray, Ph.D.
Vern Vail, B.S.
Katherine Verdolini, Ph.D.
William Winholtz, A.A.S.
Darrell Wong, Ph.D.
Raymond Wood, M.D.
Patricia Zebrowski, Ph.D.

Doctoral Students

Todd Brennan, M.S.
Cheng Hwa Chen, M.S.
Eileen Finnegan, M.A.
Susan Hensley, M.M., M.A.
Mark Leddy, M.S.
Phyllis Palmer, M.A.
Bruce Poburka, M.S.
Annie Ramos, M.S.
Eilee Savelkoul, M.S.
Alice Smith, M.A.
Julie Stierwalt, M.A.
Brad Story, B.S.
Kenneth Tom, M.A.

Postdoctoral Fellows

Kristin Larson, Ph.D.
Young Min, M.D.
Savita Prakash, M.D.

Visiting Scholars

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

Advisory Board

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

Contents

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

Part II. Tutorial reports and summaries of Training, Continuing Education and Dissemination of Information

Forward

The greatest reward that comes from assembling these status and progress reports is to see the productivity of our students and fellows. The training programs of the NCVS, and more generally of the Research and Training Centers (RTC's) that are sponsored by the National Institutes of Health, are bearing much fruit. In this issue we have original first authored contributions by Brad Story, Ina Steinecke, Stephen Austin, Carol Bauer, Young Min, Hyun-Sub Sim and Kristin Larson.

Dr. Young Min, Dr. Savita Prakash, and Dr. Kristin Larson are three of our current postdoctoral fellows. Dr. Min will be leaving us to do her residency in otolaryngology. We congratulate her, thank her, and wish her well. Dr. Larson is getting her laboratory organized in Denver. She will be with us for some time to come. In our two previous reports we also had first-authored publications by Dr. David Berry and Dr. Linnea Peterson. Dr. Prakash has already augmented her medical skills with exceptional knowledge of laboratory instrumentation. Dr. Berry continues to be productive in biomechanics of vocal fold movement and Dr. Peterson has gone on to residency at UCLA.

It is also appropriate to mention the visiting scientists that have blessed us with their contributions. Dr. Hanspeter Herzel, a physicist from Humboldt University in Berlin, has been with us for several years now, usually in the Spring. His contribution to the dynamics of nonlinear systems has been extremely valuable. We have also had the privilege of hosting his student Ina Steinecke, whose paper on asymmetrical vocal fold modeling appears in this issue. Other significant visitors have been Dr. Tzu-Yu Hsiao, Meijin Nakayama, and the distinguish vocal coach, Arthur Lessac.

We have learned much from all of them.

Ingo Titze, Director
May, 1994

Part I

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 6
May, 1994, 1-11

Voice Simulation with a Body Cover Model of the Vocal Folds

Brad Story, B.S.
Department of Speech Pathology and Audiology, The University of Iowa
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The Univeristy of Iowa

Abstract

The classic two-mass model of the vocal folds (Ishizaka and Flanagan, 1972) has been extended to a three mass model in order to more realistically represent the body-cover vocal fold structure (Hirano, 1974). The model consists of two "cover" masses coupled laterally to a "body" mass by nonlinear springs and viscous damping elements. The body mass, which represents muscle tissue, is further coupled laterally to a rigid wall (assumed to represent the thyroid cartilage) by a nonlinear spring and a damping element. The two cover springs are intended to represent the elastic properties of the epithelium and the lamina propria while the body spring simulates the tension produced by contraction of the thyroarytenoid muscle (TA). Additionally, the two cover masses are coupled to each other through a linear spring which can represent vertical mucosal wave propagation. Effects on the vocal fold movement due to changes in subglottal pressure, spring tensions, and vocal tract loading are shown.

Introduction

Review of Previous Models

Several models that simulate the self-sustained oscillation of the vocal folds have been proposed throughout the years. Flanagan and Landraf (1968) modeled vocal fold vibration with a single mass-spring oscillator driven by airflow from the lungs. The model produced reasonable self-sustained oscillations only with an inertive vocal tract load. Because of it's single degree of freedom in the tissue, it could not produce the vertical phase difference needed for flow-induced oscillation. The two mass model of Ishizaka and Flanagan (1972), subsequently referred to as IF72 in this paper, was able to sustain oscillation with or without a vocal tract and provided the degrees of freedom necessary to produce the vertical phase difference. This model has been widely used as a simple, low-dimensional model of the vocal folds. Titze (1973 and 1974) represented both the vertical and longitudinal modes of vocal fold vibration with a sixteen mass model. This model, subsequently called T73, consisted of eight coupled longitudinal sections, each with two masses in the coronal plane. In contrast to the IF72 model, the upper mass was coupled to only the lower mass and not to a rigid lateral boundary. The two masses (in each longitudinal section) were also allowed to have a vertical degree of freedom which simulated two-dimensional trajectories of the vocal fold tissue.

Koizumi et. al. (1987) has described several variations of the simple two mass model that incorporate elements of both the IF72 and the T73 models. Their modifications were intended to produce a more natural sounding artificial voice to be used primarily as a glottal source for speech synthesis.

Other more complex models that simulate the layered structure of the vocal folds provide amore precise physiological representation of human tissue. The continuum mechanics model of Titze and Talkin (1979) and a more recent finite element implementation (Alipour-Haghighi and Titze, 1983) are examples. Because these models have a very large number of degrees of freedom (DOFs), they are capable of producing a rather complex vibratory pattern composed of many different modes. However, recent modal analyses using the finite element model have shown that the vocal fold vibration is largely dominated by the first 2 to 3 modes of vibration (Berry et. al. 1993). The contributions of the higher modes have been

found to be much less significant than the lower ones. This finding implies that for some purposes a simpler model with a small number of degrees of freedom may capture enough of the vibratory characteristics to serve as a useful research tool.

Purposes of Models

Simulation of vocal fold vibration typically has one of two purposes. Either the model is used as part of a speech synthesis system, or it is used to study specific aspects of the mechanics of the vibration. Hirano (1974) states that, if the purpose of a vocal fold model is for physiological or clinical purposes, then it must have parameters that can be varied to simulate different conditions produced by various laryngeal adjustments or by pathological variations. The Flanagan-Landgraf (1968) and the IF72 models were initially created to serve as a voice source for a speech synthesis system, but the IF72 model was later used to study various pathologies of the vocal folds (Ishizaka and Isshiki, 1976). However, there is no direct physiological correlation between the spring stiffnesses and the effects of muscle contractions in either of these models. The T73 model was specifically designed to study the mechanics of vocal fold vibration with the hope that it would lead to a better understanding of the workings of the voice source. The Koizumi et. al. (1987) models were created to produce more natural sounding synthesized speech than the IF72 model, but all models have been used as a tool to study vocal pathologies (Smith et. al. 1992). In addition Wong et. al. (1991) have combined the two mass approach of the IF72 model with the longitudinal discretization of the T73 model to create a ten mass model which was used specifically to study vocal pathologies.

The continuum models obviously have as their purpose the detailed study of tissue movement throughout the vocal fold during vibration. Thus, studies of vocal pathologies as well as normal phonation are natural areas of pursuit with these models. However, the complexity that results from using many degrees of freedom makes these models computationally intensive and less well suited for use in a speech synthesis system.

The Model

Body-Cover Concept

The body-cover concept (Hirano, 1974) is generally used to describe the vocal fold structure (Figure 1). It suggests that the vocal fold can be divided into two tissue layers with different mechanical properties. The body layer consists of muscle fibers and some tightly connected collagen fibers of the vocal ligament. The cover layer consists of pliable, non-contractile tissue (the epithelium, the superficial layer and the intermediate layers of the lamina propria) that acts as a flexible sheath around the body layer. The cover typically is loosely connected to the body during vibration. Based on his findings, Hirano (1974) suggests that the vocal fold should be treated as a double structured vibrator whose stiffness parameters should be based on the relative activations of the thyroarytenoid and cricothyroid muscles.

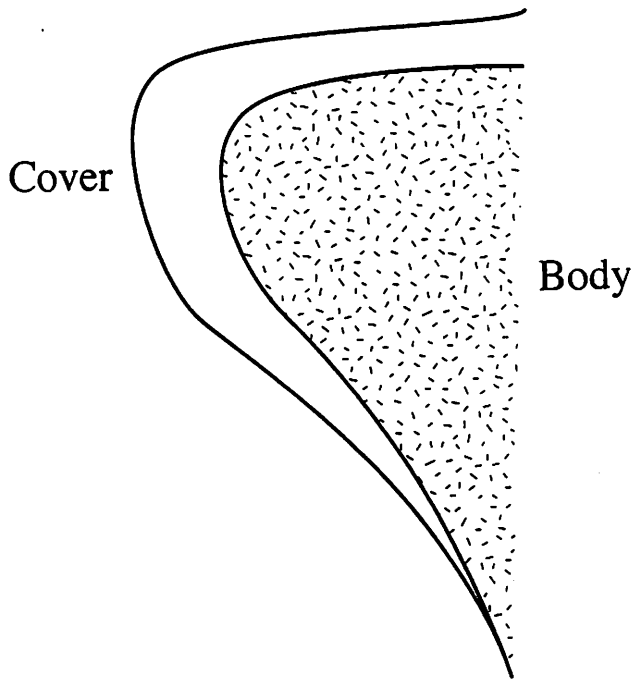


Figure 1. Body-cover structure of the vocal folds.

A three-mass model of the vocal folds is proposed here as a lumped element approximation of the body-cover structure. The model is essentially the classic IF72 two mass model with a third mass added to simulate the effect of the body component. It will be used to simulate a section of the vocal fold 0.3 cm in thickness (vertical dimension in coronal view), 0.23 cm in depth, and 1.0 cm in length. The cover portion of this 0.3 cm thick vocal fold is divided into two equally thick elements. A single larger mass is used to simulate the body layer. This discretization of the body-cover structure into a lumped element system is shown in Figure 2 (next page).

Both the upper and lower cover masses are coupled to the body mass through nonlinear springs and damping elements. The two cover masses are also coupled to each other through a coupling spring. This coupling spring accounts for the shear forces in the cover. The body mass is coupled to a rigid boundary, i.e., the thyroid cartilage. For the purposes of this paper, the right and left vocal folds are considered to be symmetric; that is, the same movement is considered to occur on the left as on the right. The asymmetric case will be the subject of a future study.

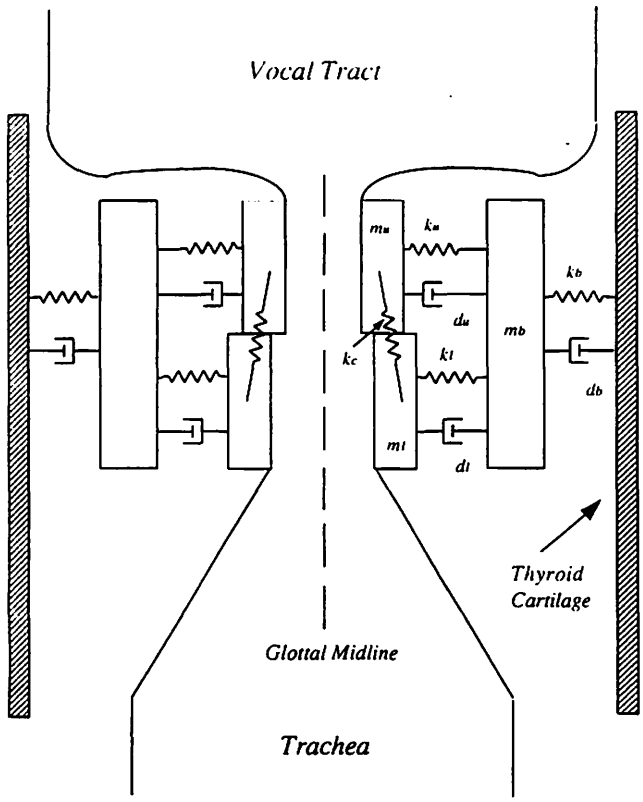


Figure 2. Lumped-element representation of the body-cover structure of the vocal folds.

The springs k_u and k_l are considered to represent the effective coupling stiffnesses between the body and cover, which vary primarily with the contraction of the cricothyroid (CT) muscle. Spring k_b represents the effective stiffness of the body, which varies with both the TA contraction and CT contraction. The values used for the masses and stiffnesses (discussed later) were obtained from considerations of the physical properties of body and cover tissues.

A transmission line vocal tract (Kelly and Lochbaum, 1962; Liljencrants, 1985) was attached to the three mass model both supraglottally and subglottally. This allowed for interaction of the vocal folds with the acoustic pressures generated above and below the vocal folds.

Equations of Motion

The equations of motion for the three masses are written in terms of the coupling forces and the external driving forces that are exerted on each mass,

$$F_u = m_u\ddot{x}_u = F_{ku} + F_{du} - F_{kc} + F_{eu} + F_{uCol} \quad (1a)$$

$$F_l = m_l\ddot{x}_l = F_{kl} + F_{dl} + F_{kc} + F_{el} + F_{lCol} \quad (1b)$$

$$F_b = m_b\ddot{x}_b = F_{kb} + F_{db} - [\, F_{ku} + F_{dl} + F_{kl} + F_{dl} \,] \quad (1c)$$

where the following forces are identified,

F_{du}, F_{dl}, and F_{db} - forces due to damping
F_{ku}, F_{kl}, and F_{kb} - lateral spring forces
F_{kc} - spring force due to the coupling of m_u and m_l
F_{uCol} and F_{lCol} - forces generated only during collision with the opposite vocal fold
F_{eu} and F_{el} - external forces generated by the glottal flow
F_u, F_l, and F_b - forces of the accelerating masses.

If x_u, x_l, and x_b are defined to be the displacements of each mass and x_{uo}, x_{lo}, and x_{bo} are the initial positions of each mass then the displacements from equilibrium will be $(x_u - x_{uo})$, $(x_l - x_{lo})$, and $(x_b - x_{bo})$. With these definitions, the equations for the lateral spring forces can be written. For the upper mass,

$$F_{ku} = -k_u[\{(x_u - x_{uo}) - (x_b - x_{bo})\} + \eta_u\{(x_u - x_{uo}) - (x_b - x_{bo})\}^3] \quad (2)$$

where η_u is the coefficient of nonlinearity and $[(x_u - x_{uo}) - (x_b - x_{bo})]$ gives the net compression or elongation of the spring. Similarly, for the lower and body masses the spring forces are,

$$F_{kl} = -k_l[\{(x_l - x_{lo}) - (x_b - x_{bo})\} + \eta_l\{(x_l - x_{lo}) - (x_b - x_{bo})\}^3] \quad (3)$$

and,

$$F_{kb} = -k_b[(x_b - x_{bo}) + \eta_b(x_b - x_{bo})^3] \quad (4)$$

All of the coefficients of nonlinearity (η's) were set to a value of 100, as was done in the IF72 model. The coupling force is determined by the relative displacements of the upper and lower masses and a linear spring (k_c). This force is given by,

$$F_{kc} = -k_c[(x_l - x_{lo}) - (x_u - x_{uo})] \quad (5)$$

When the left and right vocal folds are in collision an extra nonlinear spring is switched on to simulate the effect of impact. The extra spring forces have the form:

$$F_{uCol} = -h_{uCol}[(x_u - x_{uCol}) + \eta_u(x_u - x_{uCol})^3] \quad (6a)$$

$$F_{lCol} = -h_{lCol}[(x_l - x_{lCol}) + \eta_l(x_l - x_{lCol})^3] \quad (6b)$$

where the h's are the linear spring coefficients, with values of $3k_u$ and $3k_l$. The η's are nonlinear coefficients, set to 500 as in the IF72 model; the x_{uCol} and x_{lCol} are the displacements where collision occurs for medial motion.

The damping forces are given by

$$F_{dl} = -d_l(\dot{x}_l - \dot{x}_b) \quad (7a)$$

$$F_{du} = -d_u(\dot{x}_u - \dot{x}_b) \quad (7b)$$

$$F_{db} = -d_b \dot{x}_b \quad (7c)$$

where the *d*'s are the damping coefficients. The damping coefficients are computed with the following equations during an open glottis condition,

$$d_l = 2\zeta_l(m_l k_l)^{1/2} \quad (8a)$$

$$d_u = 2\zeta_u(m_u k_u)^{1/2} \quad (8b)$$

$$d_b = 2\zeta_b(m_b k_b)^{1/2} \quad (8c)$$

During the closed glottis condition the damping ratio (ξ) is increased in a stepwise fashion ($\xi \rightarrow \xi + 1$) for the upper and lower masses while the body mass damping is not changed. Leaving the body damping unchanged is justified by assuming that only the elements in collision suffer increased energy losses. Titze (1976) disputes the idea that the damping ratio should be increased stepwise during collision, but the phonation simulated with this model was not acceptable without the extra collision damping. There is indeed more energy lost during collision, but how it is dissipated is not known with certainty. In a continuum model, momentum in the x-direction is transferred to the y-direction during collision. This energy of deformation cannot be reclaimed. Clearly, then, there is energy lost during collision, but whether it is through simple increases in damping constants isnot known.

Pressure Equations

The equations of motion for the three mass system are coupled to the aerodynamic driving forces via the glottal area. The intraglottal pressure exerts a force on the upper and lower cover masses which is the driving force that produces oscillation. This pressure will depend on the open area of the glottis in the upper and lower mass regions. Titze and Alipour-Haghighi (forthcoming) have outlined a general method of calculating the intraglottal pressure which will be adapted here for our three mass model. The method is based on the following assumptions:

1) The flow detaches at the minimum glottal diameter.

2) Bernoulli type flow exists from the subglottal region to the minimum glottal diameter.

3) A constant diameter jet exists from the minimum diameter to the glottal exit. Pressure is considered to be constant in this region.

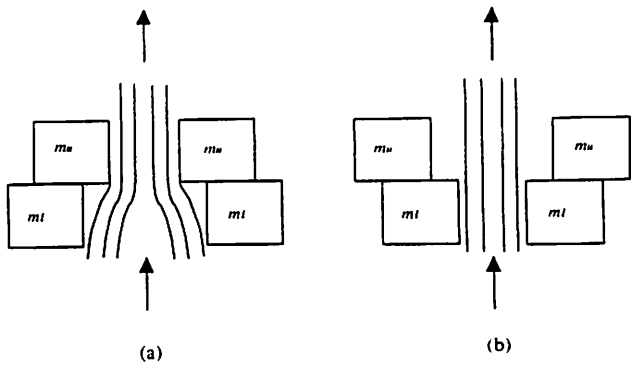


Figure 3. Assumed flow patterns in the glottis, a) Bernoulli region in lower section, jet region in upper, b) Jet region in both upper and lower sections.

4) Pressure recovery after glottal exit (expansion and reattachment) follows the equations derived by Ishizaka and Matsudaira (1972).

Figure 3 shows convergence and divergence of the glottis during a cycle of vibration. Note that for divergence a constant jet pressure is assumed throughout the glottis.

Titze and Alipour-Haghighi (forthcoming) present simplified equations for the pressure inthe glottis. In the region where Bernoulli flow is applicable, the pressure can be computed as,

$$P(a) = P_s - (P_s - P_i)\left(\frac{a_m}{a}\right)^2 \quad (9a)$$

where P_s and P_i are the subglottal and supraglottal pressures respectively, a_m is the minimum glottal area, a is the glottal area at some point upstream of the minimum, and $P(a)$ is the pressure at that upstream point. The pressure within the minimum area and at downstream points (in the glottis) is simply set to,

$$P = P_i \quad (9b)$$

The forces exerted on each cover mass by tne pressures in the glottis (P_u and P_l) are,

$$F_{eu} = P_u L_s T_u \quad (10a)$$

$$F_{el} = P_l L_s T_l \quad (10b)$$

where L_s is the length of the vocal fold that is effectively in vibration and T_u and T_l are the upper and lower vertical thicknesses.

Flow Equations

The three mass model is coupled to a vocal tract in which subglottal, pharyngeal, oral, and nasal sections are included. Since a wave reflection algorithm is being used we can derive the flow equation based on the incident and

reflected pressures at the glottis. The formulation of the flow equations are given in Titze (1984); only the results will be presented here. The flow through the glottis is,

$$u = \left(\frac{a_m c}{k_t}\right)\left\{\frac{-a_m}{A^*} \pm \left[\left(\frac{a_m}{A^*}\right)^2 + \left(\frac{4k_t}{c^2\rho}\right)\left(P_s^+ - P_i^-\right)\right]^{1/2}\right\} \quad (11)$$

where k_t is a transglottal pressure coefficient (Scherer and Titze, 1983), ρ is the density of air, a_m is the minimum glottal area, and c is the speed of sound. A* is defined as an effective vocal tract area for acoustic loading of the glottis,

$$\frac{1}{A^*} = \frac{1}{A_s} + \frac{1}{A_i} \quad (12)$$

where A_s and A_i are the areas of the first sections of the subglottal and supraglottal ducts, respectively. P_s^+ and P_i^- are the incident pressures above and below the glottis and are known from previous calculations. Once the flow has been computed, the reflected pressures, P_s^- and P_i^+, can be found using the relations,

$$P_s^- = P_s^+ - \left(\frac{\rho c}{A_s}\right)u \quad (13a)$$

$$P_i^+ = P_i^- + \left(\frac{\rho c}{A_i}\right)u \quad (13b)$$

The vocal tract shape used for the simulations in this paper is simply a uniform tube; nasal sections were not used.

Parameter Values

The numerical values used in this model should have some physiological relevance. This section demonstrates how the initial numbers were obtained and then fine tuned to produce acceptable phonation.

The masses were obtained by computing the approximate volume that is consumed by the particular element and then multiplying by the tissue density. For the cover masses, the depth is made up of the epithelial layer, the superficial layer, and the intermediate layer of the lamina propria. According to Hirano (1977) and Hirano et. al. (1981) the epithelial layer has an approximate depth of 0.005 cm while the superficial layer and intermediate layer of the lamina propria are each about about 0.03 cm deep. The depth of the cover is then,

$$\begin{aligned} D_{cover} &= 0.005 \text{ cm} + 0.03 \text{ cm} + 0.03 \text{ cm} \\ &= 0.065 \text{ cm} \end{aligned} \quad (14)$$

At this point we must assume an effective length of the vibrating vocal fold and define a vertical thickness for each mass. A length of 1 cm was chosen for the vocal fold length and the thickness of each mass will be 0.15 cm for a total thickness of 0.3 cm. Using the thickness, depth, and length we can compute the volume consumed by the portion of the cover that we are intending to simulate. The cover mass is then the product of the volume and the tissue density. The density of the mucosa is approximately 1.02 g/cm^3 (Perlman, 1985). Thus the mass of the cover is,

$$\begin{aligned} M_{cover} &= (0.065 \text{ cm})\,(0.3 \text{ cm})\,(1.0 \text{ cm})\,(1.02 \text{ g/cm3}) \\ &= 0.0199 \text{ g} \end{aligned} \quad (15)$$

In this model each cover element will be assumed to have the same mass so that,

$$m_l = m_u = \frac{M_{cover}}{2} \approx 0.01\ g \quad (16)$$

For the body mass we assume the same length and thickness as the cover. The depth of the body is determined by the combination of the depth of the deep layer of the lamina propria and the depth of the portion of the muscle that is effectively in vibration. Hirano et. al. (1981) reported that the depth of the deep layer is approximately 0.05 cm. However, the amount of the muscle tissue that is involved in vibration depends on the amplitude of vocal fold vibration and on the relative activation levels of the CT and TA muscles. Titze et. al. (1989a) define a ratio of the cross section of the TA muscle in vibration to the total cross-sectional area in vibration (including the cover and deep layer of the lamina propria). They suggest that this ratio may have a value of about 0.3 for a "soft" loudness condition and 0.6 for a "loud" loudness condition. Choosing a value of 0.5 for this ratio to simulate a condition between soft and loud, an effective muscle depth can be computed. Previously it was shown that the depth of the cover is ~0.065 cm and the depth of the deep layer (of the ligament) is ~.05 cm which combine to produce a total depth of 0.115 cm. If the ratio of TA cross-section to total cross-sectional area is 0.5 then the depth of the muscle will also need to be 0.115 cm, assuming that the body and cover have the same thickness. Now the depth of the body is the combination of the deep layer and the portion of vibrating muscle,

$$\begin{aligned} D_{body} &= 0.05 \text{ cm} + 0.115 \text{ cm} \\ &= 0.165 \text{ cm} \end{aligned}$$

The body mass is then the product of the volume and the tissue density. The density of the muscle has been measured to be 1.04 g/cm^3 (Perlman, 1985) and we will assume for simplicity that the density of the deep layer of the lamina propria (ligament) has the same density as the muscle. Sothe mass of the body element is,

$$m_b = (0.165\text{ cm})\,(0.3\text{ cm})\,(1.0\text{ cm})\,(1.04\text{ g/cm3}) \quad (17)$$
$$= 0.05148\text{ g} \qquad \text{(app. 0.05 g)}.$$

The values of the effective stiffness coefficients are determined mainly by the longitudinal stress in the tissue fibers. This stress can be converted to an equivalent coupling stiffness by equating the expressions for fundamental frequency of a vibrating string and a vibrating mass,

$$F_o = \left(\frac{1}{2L_s}\right)\left(\frac{\sigma}{\rho}\right)^{1/2} = \left(\frac{1}{2\pi}\right)\left(\frac{k}{m}\right)^{1/2} \quad (18)$$

where σ is the longitudinal stress in the tissue fibers and L_s is the vocal fold length. Now solving for k gives,

$$k = \frac{\pi^2 \sigma m}{\rho L_s^2} \quad (19)$$

The values for mass have already been computed, the density of the tissue is known, and the length has been chosen to be 1 cm. Thus, only a value for stress (σ) is needed to calculate k.

Alipour-Haghighi and Titze (1991) give passive stress-strain curves for both body and cover tissues taken from an excised canine larynx. This study shows that a strain of 10% in the cover will produce a stress of about 4.0 kPa while the same amount of strain in the body tissue produces about 3.5 kPa of stress. We now can use these values of stress in equation (15). With $\rho_{cover} = 1.02\text{ g/cm}^3$ and $\rho_{body} = 1.04\text{ g/cm}^3$ we get the following estimates of k for the body and cover,

$$k_{cover} = 5.0\text{ N/m}$$
$$k_{body} = 50.0\text{ N/m}.$$

But these are only passive stiffnesses, so if there is any TA contraction the stiffness of the body will be higher. Alipour-Haghighi et. al. (1989) obtained a stress-strain relationship for tetanically stimulated vocal fold muscle which showed that a stress of approximately 75 kPa was produced for a strain of 10%. Using this value for the stress (σ) in equation (19) gives a stiffness coefficient of ~850.0 N/m. This value represents the stiffness of the vocal fold muscle that could be expected during supra-maximal stimulation. However, during normal phonation the level of activation will be much lower than supra-maximal and hence the stiffness will also be much lower. It was found that reasonable phonation was achieved if the body stiffness was set to 100.0 N/m which is higher than the passive stiffness but much lower than the supra-maximal condition. It was also determined that the well documented vertical phase difference was best modeled if the stiffness of the upper cover element was slightly smaller than the lower. Thus the following stiffness values were used:

$$k_u = 3.5\text{ N/m}$$
$$k_l = 5.0\text{ N/m}$$
$$k_b = 100.0\text{ N/m}.$$

The spring constant, k_c, accounts for the shear forces between the masses. In a distributed system, this shear determines the mucosal wave velocity. It is this parameter over which the human appears to have the least control, its properties being primarily dependent upon vocal health. The mucosal wave velocity has been measured in the range from 1.0 to 2.0 m/s (Baer, 1975; Titze et al., 1993; Sloan et al., 1993). In this model, a value of 2.0 N/m was chosen for k_c, which produced a wave velocity of approximately 1.1 m/s.

Kaneko et. al. (1974) estimated the vocal fold damping ratio to be approximately 0.1 - 0.2 at a fundamental frequency of 30-40 Hz while Isshiki (1977) reported damping ratios of 0.2 - 0.4 at a fundamental frequency of 130 Hz. Experience with the three mass model has shown that phonation is more acceptable using damping ratios in line with Isshiki's data rather than those of Kaneko. The following damping ratios have been assigned to the three elements,

$$\xi_u = 0.4$$
$$\xi_l = 0.4$$
$$\xi_b = 0.2$$

The parameter values defined in this section constitute a system that simulates essentially normal phonation. This means that a steady vibration is achieved with the amplitudes of vibration, vertical phase difference, fundamental frequency, and mucosal wave velocity all in the range of the published information on these quantities.

Simulation Results

Four Cases of Hirano (1974)

It was desired to test this model on realistic laryngeal configurations. Hirano (1974) describes four typical laryngeal adjustments that are reflected in terms of the relative stiffnesses of the body and cover layer. Each of these four conditions has been simulated with the three mass model by varying the three spring constants. The model parameters for each case are shown in Table I. For each case, the laryngeal configuration and the characteristics of the vocal fold vibration which Hirano describes will be repeated and then followed with a comparison of the simulation results.

Figure 4a shows the case in which the stiffness of both the body and cover is low due to a small amount of contraction in the TA and CT muscles. Hirano (1974)

Table 1.
Parameter values for the four cases of Hirano (1974).*

Values	Case A	Case B	Case C	Case D
m_u	0.01	0.01	0.01	0.01
m_l	0.01	0.01	0.01	0.01
m_b	0.05	0.05	0.05	0.015
T_u	0.15	0.15	0.15	0.15
T_l	0.15	0.15	0.15	0.15
k_u	3.5	3.5	3.5	79.0
k_l	5.0	5.0	5.0	80.0
k_b	30.0	700.0	100.0	300.0
k_c	2.0	2.0	2.0	2.0
x_{uo}	0.01799	0.01799	0.01799	0.01799
x_{lo}	0.018	0.018	0.018	0.018
x_{bo}	0.30	0.30	0.30	0.30
P_l	0.80	0.80	0.80	0.80

* *m's* = mass (g), *k's* = stiffness (N/m), *T's* = thickness (cm), x_o's = prephonatory displacement (cm), P_l = kPa.

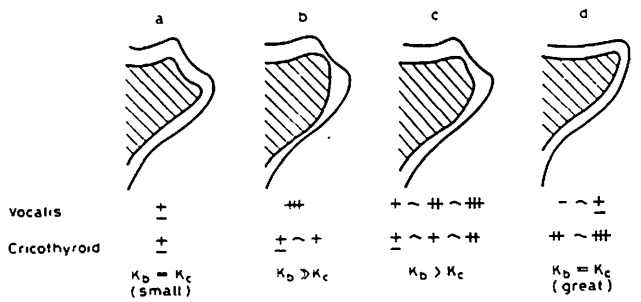


Figure 4. Four laryngeal configurations based on different relative stiffnesses of body and cover elements (from Hirano, 1974).

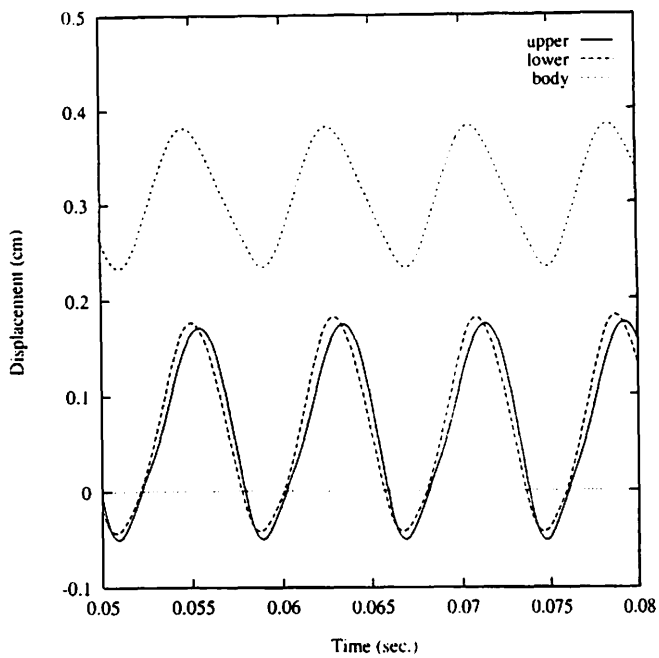
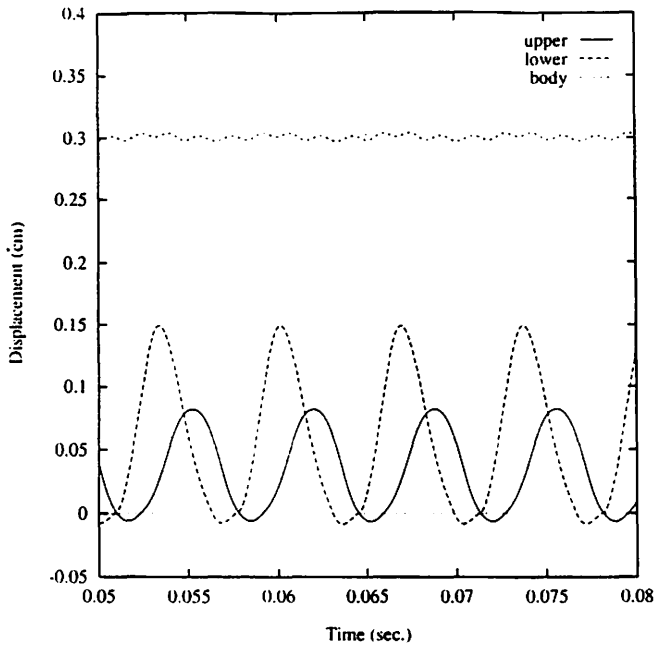


Figure 5. Displacements of the three masses for case A in Table 1.

suggests that both body and cover are quite lax and will be equally involved in the movement. This, he claims, is typical of soft phonation at low pitch. This case was simulated by setting the stiffness coefficients to be those given in Table I for case A. The simulation is shown in Figure 5. The displacements of each of the three masses during vibration are shown on the vertical axis while time is on the horizontal. Zero displacement is considered to be the glottal midline and any negative displacement implies collision with the opposite vocal fold. It can be seen in this figure that the cover and the body essentially move together. This case produced a fundamental frequency of 125 Hz, a lower to upper amplitude ratio of 1.03, and a vertical phase difference of 17 deg/mm, which suggests a high mucosal wave velocity of 2.7 m/s.

When the TA is contracted to a much greater degree than the CT, the stiffness of the body is much higher than that of the cover. In this case, which Hirano (1974) claims to berepresentative of loud heavy voices at medium pitch levels, the vibration presumably takes place mainly in the cover as sketched in Figure 4b. The simulation of this case is shown in Figure 6. It is observed that there is very little movement of the body mass while considerable movement takes place in the cover masses. F_o was 113 Hz with a vertical phase difference of 65 deg/mm. The amplitude ratio (lower to upper mass) was found to be 3.0 while the mucosal wave velocity was a low 0.2 m/s.

In case C the TA contraction is slightly more dominant than that of the CT. Vocal fold movement will involve both the body and cover but may involve the cover slightly more. Hirano (1974) suggests that this case is considered to be "normal" phonation. The simulation results are shown in Figure 7. It is observed that both the body and cover are involved in the motion with the cover having a somewhat greater amplitude. This case produced an F_o of 138 Hz, which is slightly higher than in the previous two cases. The vertical phase difference is 44 deg/mm and the mucosal wave velocity is 1.1 m/s. Titze, Jiang, and Hsiao (1993) have recently reported vertical phase differences of 27-61 deg/mm with most of the data clustering in the 40-55 deg/mm range. They also measured mucosal wave velocities of 0.5-2.2 m/s. The three mass simulation for "normal" conditions agrees well with these numbers.



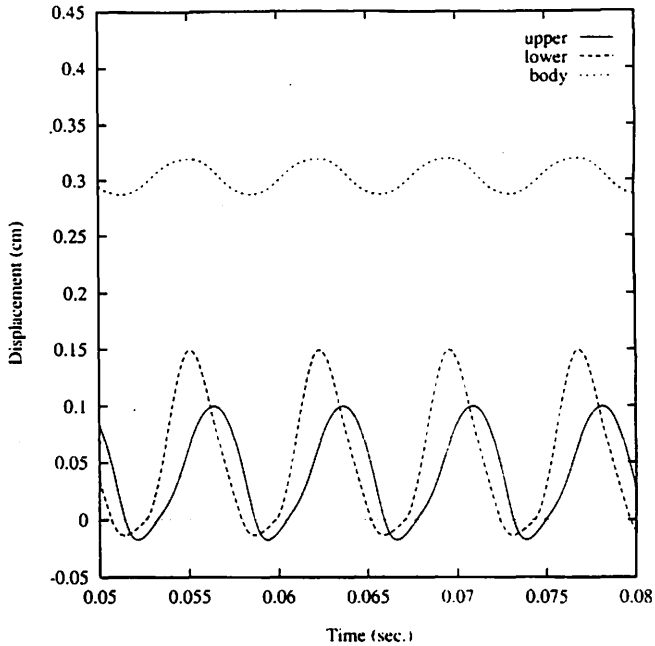


Figure 6 (left). Displacements of the three masses for case B in Table 1. Figure 7 (right). Displacements of the three masses for case C in Table 1.

Figure 4d shows the conditions that Hirano (1974) suggests are used for falsetto voice. A large CT contraction imposes a large amount of passive stiffness on both the cover and the body. This means that k_u, k_l and , k_b are all very large, but k_b is large because the deep layer of thelamina propria (part of the vocal ligament) now assumes the longitudinal tension rather than the muscle. The active tension in the muscle is nearly zero, the amplitude of vibration is very small and there is little or no vertical phase difference. A recent study of the stress-strain properties of the human vocal ligament has shown that a stress of approximately 130 kPa is developed by a strain of 40-50% (Min et al., 1994). The calculation of the stiffness, k_b, for the falsetto case was based on this information. Also, the effective vibrating mass of the body is primarily the ligament and so m_b has been reduced.

The simulation in Figure 8 shows very little movement in both the body and cover. The displacements of both of the cover masses are shown in this graph but they are nearly "locked" together in vibration, hence they have almost identical displacement waveforms. This means that there is very little vertical phase difference and a correspondingly high mucosal wave velocity. It should also be noted that the cover masses do not collide with the opposite vocal fold which is quite typical of human falsetto production. The fundamental frequency has increased to 333 Hz which is an increase in frequency of more than a factor of 2.5 over the previous cases.

Table 2.
Simulation results of the four cases of Hirano (1974) in terms of resulting F_o, vertical phase differences (VPD), lower to upper amplitude ratio (AR), and mucosal wave velocity (MWV).

Case	F_o(Hz)	VPD (deg/mm)	AR	MWV (m/s)
A	125	17	1.03	2.7
B	113	65	3.0	0.2
C	138	44	1.6	1.1
D	333	5.0	1.0	23.7

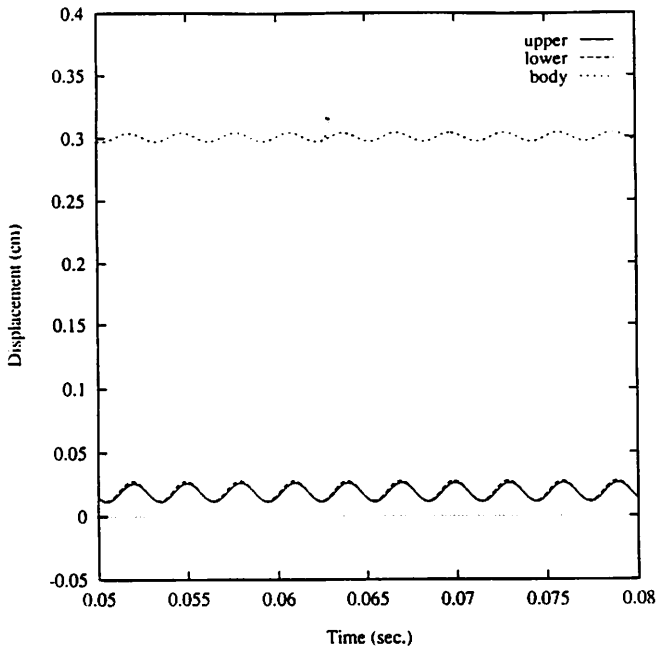
Table 2 summarizes the results of these four cases in terms of resulting F_o, vertical phase difference (VPD), lower to upper amplitude ratio (AR), and mucosal wave velocity (MWV).

Intraglottal Pressure

As a further test of how the three-mass model can simulate characteristics of real vocal fold vibration, intraglottal pressure calculated from the model was compared to recent measurements ofthis quantity by Jiang and Titze (1993, in press).

In the three mass model the intraglottal pressure was defined to be the average pressure acting on the two cover masses. Using the parameters of case "c" (normal phonation) the intraglottal pressure pattern shown in Figure 9 is produced.

The two peaks at the beginning of one of the intraglottal pressure cycles (at about 0.058 seconds for



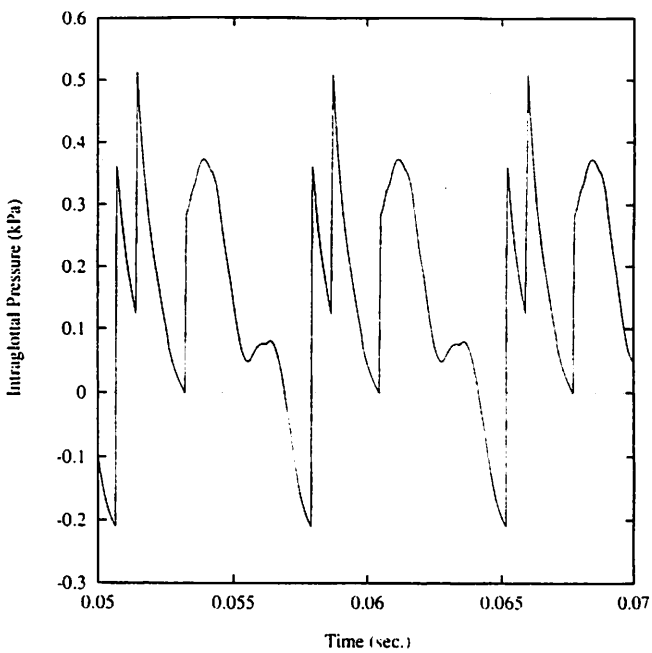


Figure 8 (left). Displacements of the three masses for case D in Table 1. Figure 9 (right). Intraglottal pressure computed by the three mass model for case C in Table 1. Includes a simulated human sized supraglottal and subglottal vocal tract.

example) are due to the successive collisions of the lower and upper masses. The collision forces then begin to subside as the folds start their opening phase. When the lower masses (left and right) break apart, the aerodynamic pressure dramatically increases. The aerodynamic pressure then declines as the folds once again move toward closure. When the right and left folds are very close to collision the pressure sinks to negative values. This is the so-called Bernoulli effect.

Jiang and Titze (1993) measured intraglottal pressure in an excise canine hemilarynx configuration. They obtained pressure curves like that shown in Figure 10. Owing to the fact that we are comparing a real continuous system to a discretized model we will certainly see significant differences in these waveforms. Specifically, the "peakiness" seen the modeled intraglottal pressure comes from discretizing the vertical aspect of the vocal folds into two sections. However, a difference that is not explained by the discretization is the presence of the large negative Bernoulli pressure in the model and its virtual absence in the experiment. It was suspected that this might be due to the fact that the model simulation was performed with a supra- and subglottal vocal tract that simulated the human system.

The model was run again using a simulation of the supraglottal and subglottal setup used in the experiment. The supraglottal system was almost completely removed except for one section (0.875 cm) because a typical larynx dissection leaves cartilage about 0.5 to 1 cm higher than the vocal folds. The subglottal system consisted of a brass

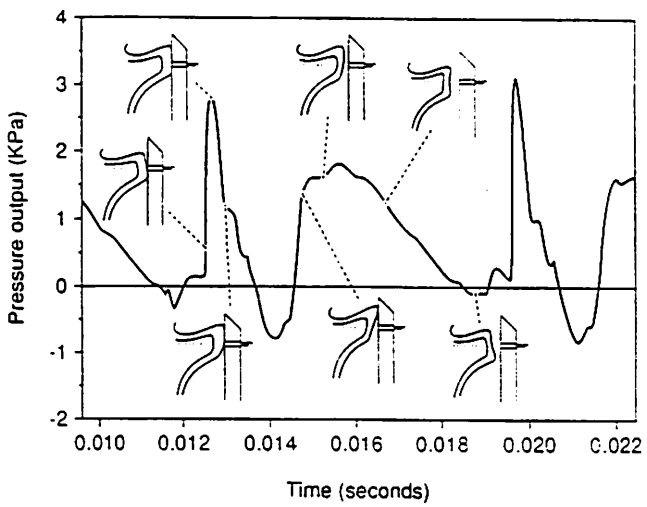
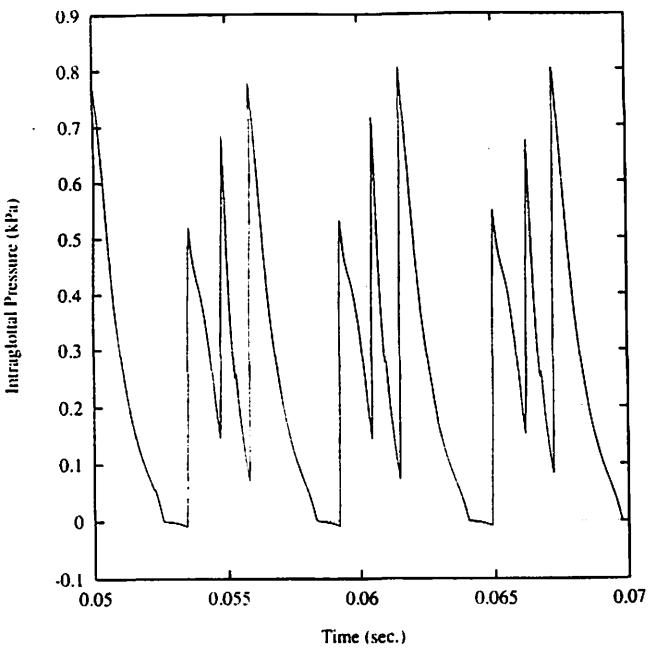


Figure 10. Intraglottal pressure measure with excised canine hemilarynx setup (from Jiang and Titze, 1993 in press).

"tracheal" tube that was connected to "pseudo-lung" (chamber of volume comparable to that of lungs). This was simulated by lengthening the subglottal system in the model and creating the chamber volume. The resulting simulation is shown in Figure 11 (next page).

Now there is just a slight dip below zero in the Bernoulli phase which is quite compatible with the experimental curve shown in Figure 10. It should also be noted that the fundamental frequency increased to 175 Hz in the simulation of the experimental setup. The fact the F_0 will



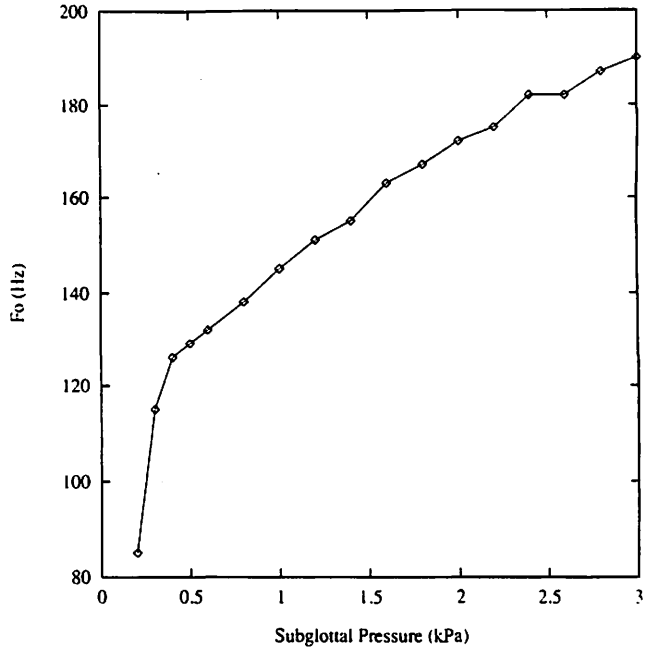


Figure 11 (left). Intraglottal pressure computed by the thress mass model for case C in Table 1. Included a simulation of the supraglottal and subglottal systems used in the experiment shown in Figure 10. Figure 12 (right). Fundamental frequency vs subglottal pressure for the three mass model. Slope is approximately 1.6 Hz/cm H20.

be higher without a vocal tract load is discussed in Ishizaka and Flanagan (1972) and in Titze (1988).

F_0 vs. Subglottal Pressure

As one final test of the three mass model, we compute resulting fundamental frequency that is produced by variations in subglottal pressure. Once again the parameters of case C were used and the subglottal pressure was varied from 200 to 3000 Pa (2 to 30 cmH_2O). A graph of F_0 versus subglottal pressure is shown in figure 12. It was found that the F_0 increased by approximately 0.016 Hz/Pa (1.6 Hz/cmH_2O). This falls in the range of 0.005 - 0.06 Hz/Pa (0.5 - 6 Hz/cmH_2O) found by Titze (1989b).

Conclusion

The three mass model offers a simple, low dimensional system in which various laryngeal configurations can be simulated. Contractions of the CT and TA muscles are incorporated in the values used for the stiffness parameters of the body and cover springs. Simulations have shown reasonable similarity to observed vocal fold motion, measured vertical phase difference and mucosal wave velocity, as well as experimentally obtained intraglottal pressure.

Acknowledgements

This research was partially funded by grant No. P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders.

References

Alipour-Haghighi, F., and Titze, I.R., "Elastic models of vocal fold tissues," 90(3), 1326-1331, 1991.

Alipour-Haghighi,F.,Titze,I.R., and Perlman, A.L.,"Tetanic contraction in vocal fold muscle,"JSHR, 32,226-231, 1989.

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

Baer, T., "Investigation of phonation using excised larynges," Doctoral Dissertation, Cambridge, Mass: Massachusetts Institute of Technology, 1975.

Berry, D.A., Herzel, H., Titze, I.R., and Krischer, K., "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," JASA, in press, 1994.

Flanagan, J.L., and Landgraf, L., "Self-oscillating source for vocal-tract synthesizers," IEEE Trans.Audio Electroacoustics, AU-16(1), 57-64, 1968.

Hirano, M., Kurita, S., and Nakashima, T., "The structure of the vocal folds," in Vocal Fold Physiology, edited by K. Stevens and M. Hirano, Univ. of Tokyo Press, 33-41, 1981.

Hirano, M., "Structure and vibratory behavior of the vocal folds," In *Dynamic Aspects of Speech Production*, M. Sawashima and F.S. Cooper (Eds.), Univ. of Tokyo Press, Tokyo, Japan, 13-27, 1977.

Hirano, M.,"Morphological structure of the vocal cord as a vibrator and its variations," Folia Phoniat.,26,89-94, 1974.

Ishizaka, K. and Isshiki, N., "Computer simulation of pathological vocal-cord vibration," JASA, 60(5), 1193-1198, 1976.

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

Ishizaka, K., and Matsudaira, M., "Fluid Mechanical Considerations of Vocal Cord Vibration," Monogr. 8, Speech Commun. Res. Lab., Santa Barbara, CA, 1972.

Isshiki, N., *Functional Surgery of the Larynx*, Kyoto University, Kyoto, Japan, pp. 62-67, 1977.

Jiang, J.J., and Titze, I.R., " Measurement of vocal fold intraglottal pressure and impact stress," J. Voice, in press.

Kaneko, T., Asano, H., Naito, J., Kobayashi, N., Hayashi, K., and Kitamura, T., "Biomechanics of the vocal cords - on damping ratio," J. Jpn. Bronchoesophagol. Soc., 25(3), 133-138 (in Japanese), 1972.

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

Koizumi, T., Taniguchi, S., and Hiromitsu, S., "Two-mass models of the vocal cords for natural sounding voice synthesis," JASA, 82(4), 1179-1192, 1987.

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

Min, Y., Titze, I.R., Alipour-Haghighi, F., and Hoffman, H., "Biomechanics of the human larynx: Stress-strain response of the vocal ligament," (in preparation), 1994.

Perlman, A.L., "A technique for measuring the elastic properties of vocal fold tissue," unpublished doctoral disseration, The University of Iowa, Iowa City, Iowa, 1985.

Scherer, R., and Titze, I.R., "Pressure flow relationships in a model of the laryngeal airway with a diverging glottis," in *Vocal Fold Physiology: Current Research and Clinical Issues*, edited by D. Bless and J. Abbs (College Hill, San Diego), 179-193, 1983.

Sloan, S.H., Berke, G.S., Gerratt, B.R., Kreiman, J., and Ye, M., "Determination of vocal fold mucosal wave velocity in an in vivo canine model," Laryngoscope, 103(9), 947-953, 1993.

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

Titze, I.R., "The human vocal cords: A mathematical model, part II", Phonetica, 28, 129-170, 1973.

Titze, I.R., "The human vocal cords: A mathematical model, part I" Phonetica, 29, 1-21, 1974.

Titze, I.R., "On the mechanics of vocal-fold vibration," JASA, 60(6), 1366-1380, 1976.

Titze, I.R., and Talkin, D.T., "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation," JASA, 66(1), 60-74, 1979.

Titze, I., "Parameterization of the glottal area, glottal flow, and vocal fold contact area," JASA, 75(2), 570-580, 1984.

Titze, I.R., "The physics of small-amplitude oscillation of the vocal folds," JASA, 83(4), 1536-1552, 1988.

Titze, I.R., "On the relation between subglottal pressure and fundamental frequency in phonation," JASA, 85(2), 901-906, 1989b.

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

Titze, I.R., Jiang, J.J., and Hsiao, T.Y., "Measurement of mucosal wave propagation and verticalphase difference in vocal fold vibration," Ann. of Oto., Rhin., & Laryn., 102(1), 58-63, 1993.

Titze, I.R., and Alipour, F., *Myoelastic Aerodynamic Theory of Phonation*, (forthcoming book).

Wong, D., Ito, M.R., Cox, N.B, and Titze, I.R., "Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases," JASA, 89(1), 383-394, 1991.

NCVS Status and Progress Report - 6
May, 1994, 13-18

Phonation Threshold Pressure in a Physical Model of the Vocal Fold Mucosa

Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Sheila S. Schmidt, B.S.E.
Department of Speech Pathology and Audiology, The University of Iowa
Michael R. Titze
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

The vocal fold mucosa, which consists of the epithelium and the superficial layer of the lamina propria, has been modeled by a fluid encapsulated in a silicone membrane. The artificial mucosa was attached to a rigid (metal) vocal fold body and introduced into an airflow channel. Flow-induced oscillation of the mucosa was achieved at various flow pressures and glottal diameters. Phonation threshold pressure, the parameter of interest, was lowest for glottal diameters between 0.0 and 0.1 mm and for fluids with the lowest viscosity. There was a consistent hysteresis effect; that is, phonation threshold pressure was always lower when oscillation subsided than when oscillation was initiated.

Introduction

Phonation threshold pressure (P_{th}) has been defined as the minimum lung pressure that produces oscillation of the vocal folds (Titze, 1988; 1992). It can have both theoretical and practical importance. Theoretically, it establishes the conditions for flow-induced vibration, the sudden switch from a point attractor (damped oscillation) to a limit cycle (self-sustained oscillation). In nonlinear dynamics, this has been called a Hopf bifurcation. From a practical standpoint, P_{th} is a measure of the "ease" of phonation. It has the potential for becoming a diagnostic tool for vocal health.

The objective of the present study was to test the validity of an analytical expression derived earlier for phonation threshold pressure (Titze, 1988). The expression (reviewed below) contains several geometric and biomechanical parameters of the vocal folds. It was assumed that these parameters could be controlled in isolation by constructing a physical model of the vocal fold mucosa. Specific questions of interest were: (1) does P_{th} increase linearly with glottal width and fluid viscosity of the mucosa, and (2) is there a predictable hysteresis for P_{th} in terms of the direction in which the bifurcation approached?

Review of the Analytical Result

The vocal fold cover consists of a 50 m layer of stratified squamous epithelium and a 1-2 mm lamina propria (Hirano, 1975). This cover propagates a surface wave in the direction of airflow. The surface wave is commonly called a *mucosal wave* because the epithelium and the outer (superficial) layer of the lamina propria, which assume most of the vibration, together constitute the vocal fold mucosa. The mucosa resembles a balloon filled with a viscoelastic fluid, i.e., a gel. In previous analytical work, small oscillation conditions were investigated by assuming a mucosal wave on the surface of the vocal folds (Titze, 1988). It was shown that energy can be transferred from the glottal airstream to the vocal folds if the mucosal wave propagates from the bottom to the top. For an assumed rectangular pre-phonatory glottis, the key relation was

$$P_{th} = \left(\frac{k_t}{T}\right) Bc\,\xi_o \qquad (1)$$

where k_t is a transglottal pressure coefficient (about 1.1 according to Scherer,1981),T is the vocal fold thickness, B is the viscous damping coefficient for mean lateral displacement of the vocal fold,c is the propagation velocity of the mucosal wave,and ξ_o is the prephonatory glottal half-width.

Note that the equation predicts a linear relation between P_{th} and glottal half-width (or the total glottal width if symmetry exists across the glottal midline). A linear relation is also predicted between P_{th} and viscous damping in the vocal folds. The other variables are also of interest, but are not subject to present empirical investigation.

Apparatus and Methods

A physical model of the glottal airway was constructed to control the glottal width and viscosity of the mucosa (Figure 1). The airway consisted of a rectangular plexiglass encasement mounted on an air supply. As seen, this construction simulated a hemilarynx rather than a full larynx. The rationale for this simplification was that a hemilarynx behaves very similar to a full larynx (Jiang & Titze, 1993), but it eliminates many asymmetries that might occur due to inconsistent membrane construction. Such asymmetries will be the subject of future investigations.

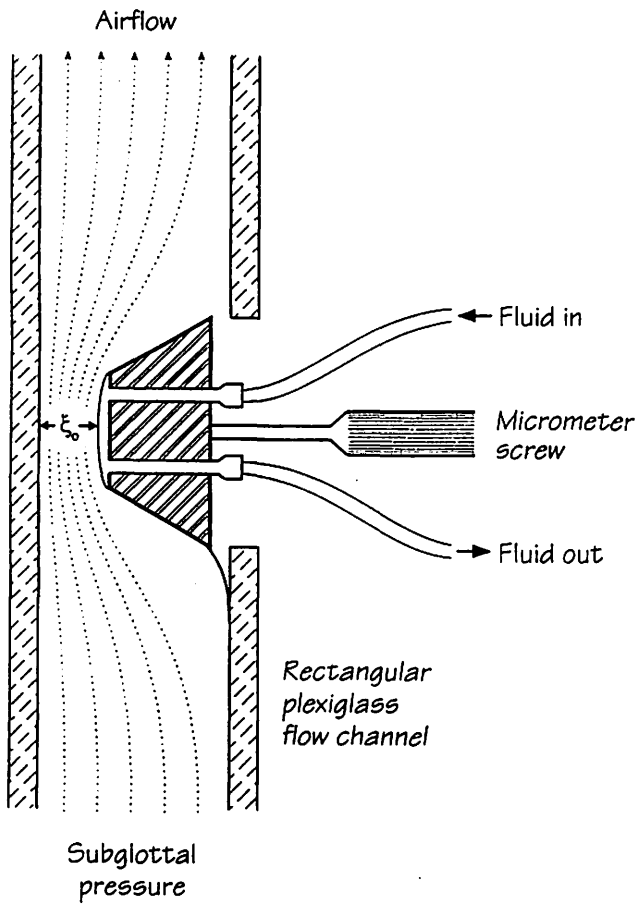


Figure 1. Sketch of the physical model of the vocal fold and the glottal airway.

The model incorporated a moveable stainless steel vocal fold body machined from a hexagonal rod (cut in half). This body offered a flat medial surface and tapered upper and lower surfaces that extended toward the walls of the airway. The vocal fold thus appeared as symmetrical trapezoid.

In the construction phase of each vocal fold, an additional stainless steel mold of similar size was topped with a clay layer approximately 1.5 mm thick. Vocal fold "epithelia" were created by coating this mold with a silicone dispersion liquid (A-1009 Factor II Supplies). The silicone thickness was controlled by motorized dipping. The membranes used for this study were consistently on the order of .2 mm in thickness, as measured by a Mitutoyo 0-1" digital micrometer. A gp-polystyrene release spray (Factor II Supplies) was sprayed on the mold prior to silicone application to ease future removal of the membrane. The silicone was then left to cure for 8-16 hours in a well ventilated area.

The membrane was subsequently removed from the mold and attached to the vocal fold with Dow Corning Medical Type-A Silicone Adhesive. This attachment produced an airgap on the medial surface of the vocal fold due to the previously added volume of clay on the mold. This airgap was filled with fluids of varying viscosities. The combination of membrane and fluid became the model of the mucosa. The fluid supply lines each consisted of a small hose adapter, 1/8" OD tubing, and a syringe. One syringe provided the input of any fluid, whereas the other syringe provided a means of extracting residual air. Thin gasket rubber material was cut and used wherever fluid tight washer-like seals were needed. The tubing could be clamped off whenever it became necessary to switch syringes, fluid, etc.

The vocal fold was pushed into the airway by a micrometer screw. This screw was located between the two fluid holes. The bottom edge of the moveable vocal fold was sealed to the airway wall with a flexible membrane and with Dow Corning Performance Plus white silicone sealant. When the vocal fold was properly sealed, the entire model was mounted onto the air supply.

The air supply was regulated by a Fairchild Model 10 0-2 psi pressure regulator. The air first traveled through a long helical subglottal tube, designed to decrease acoustic resonances. A pressure tap was placed directly below the attachment point of the model, which led to a Dwyer 60 cm H_20 manometer. The entryway into the model consisted of a 3/4" circular duct that was gradually tapered to the square walls of the model with clay. This tapering was intended to reduce turbulence as the airstream entered the glottis, but tests for turbulence were not conducted.

Prior to formal data collection, the oscillation tendencies of the mucosa were checked at various random gap sizes. Since it was desirable for oscillation to occur at various gap sizes, slight additions or subtraction of fluid would sometimes be necessary to provide a broad region of oscillation.

A video camera was mounted above the model so that the glottal gap could be watched and magnified on a nearby monitor, as well as recorded on tape if desired. A "zero point" was chosen and the corresponding reading on the micrometer was noted. This zero-point corresponded to the point when any part of the membranes was just about to touch (usually the bottom portion). A transparency was then used to trace the outline of the membranes from the monitor. This tracing served as a baseline volume measurement between fluids. If the fluid was changed in a membrane, i.e. to increase the viscosity, the tracings could be compared and the fluid amount adjusted so that the two configurations were as similar as possible in volume and shape.

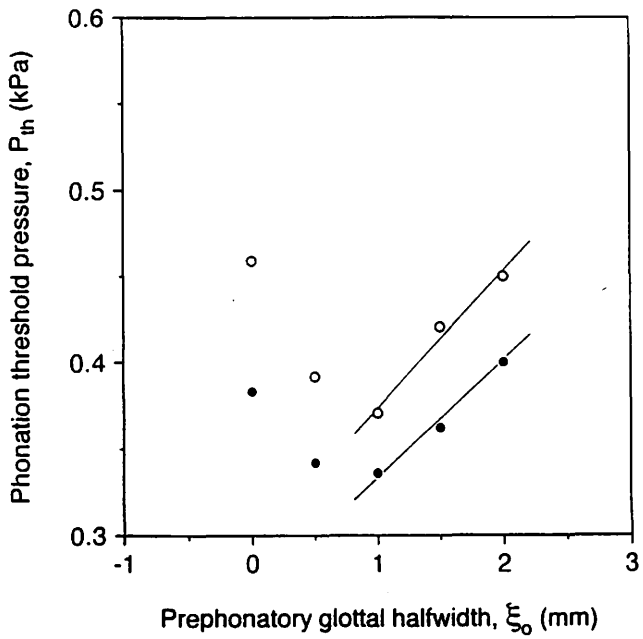
The air pressure was increased until oscillation occurred. The manometer pressure at this point was recorded as the ascending threshold pressure for the given gap size and viscosity. The pressure was then decreased until the oscillation ceased. The corresponding pressure was recorded as the descending threshold pressure. The measurements were repeated as necessary to ensure consistency and reproducibility. The gap size was changed with the micrometer, and similar measurements were gathered.

At the completion of a set of measurements, the fluid within a membrane was extracted, and the membrane flushed with clear water if necessary. A new fluid of different viscosity was injected, checking the volume and shape so as to closely match the tracing previously mentioned. The fluids used in this experiment were created by mixing graduated weights of sodium carboxymethyl cellulose powder (sodium CMC, type 7H from Aqualon). These mixtures provided a range of viscous liquids that had a similar solvent, in this case water.

Results

Figures 2 and 3 show data from two separate experimental sessions, repeated one day apart under the same control conditions. P_{th} is plotted against the prephonatory glottal half-width ξ_o. The fluid injected was pure water, which has a viscosity of 1.0 cp (centipoise). Differences between Figures 2 and 3 can be attributed to measurement errors (pressure, micrometer reading), errors in remounting and reinjecting fluid, and errors in possible "aging" of the membrane in a 24 hour period. All in all, the repeatability of the data was within about ± 0.02 kPa.

Each data point in Figures 2 and 3 represents an average of 3 to 5 readings as subglottal pressure was varied in ascending steps through threshold, followed by descend-



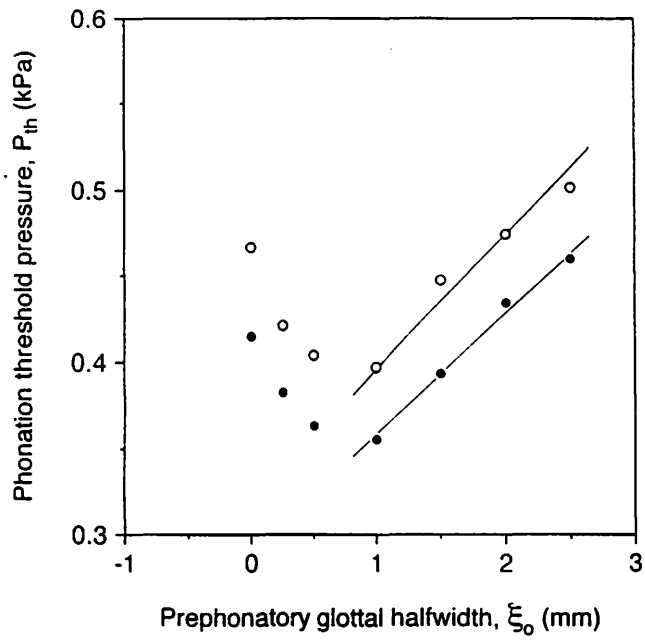


Figure 2 (left). Phonation threshold pressure for ascending subglottal pressure (open circles) and descending subglottal pressure (filled circles) as a function of glottal gap size. The fluid was pure water. Measurements are averages over one experimental session. Figure 3 (right). Phonation threshold pressure for ascending subglottal pressure (open circles) and descending subglottal pressure (filled circles) as a function of glottal gap size. The fluid was pure water. Measurements are averages for a second experimental session, one day later.

ing steps through threshold, in a cyclic manner. The pressure steps were 0.05 kPa. Note that P_{th} for ascending subglottal pressure is always higher than for descending subglottal pressure. Thus, the oscillator showed a significant hysteresis in its threshold pressure. Differences between onset and offset were on the order of 0.05 kPa, as Figures 2 and 3 indicate.

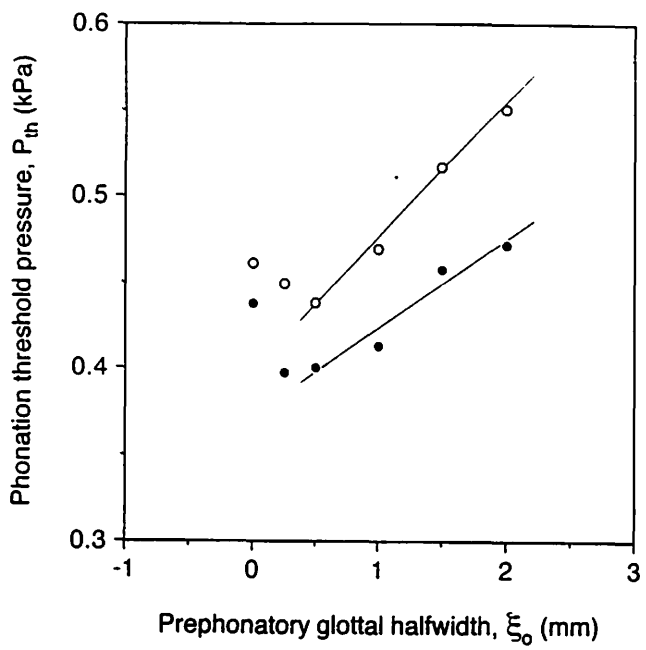
Perhaps the most noteworthy feature of the two graphs is that P_{th} does not change in a linear fashion over the entire range of ξ_o values chosen. A minimum threshold pressure occurred near ξ_o = 1.0 mm in these data sets. On the right side of the minimum, P_{th} increases monotonically (if not linearly) with ξ_o. This was predicted by equation (1). The sudden turnaround for ξ_o < 1.0 mm was not predicted theoretically, however. Inspection of the videotapes of the strobed oscillation revealed that collision between the membrane and the plexiglass wall occurred for these smaller diameters, suggesting that the small (infinitesimal) oscillation assumptions were not met. Hence, the theory upon which equation (1) was based cannot be expected to apply to these smaller diameters. This does not invalidate the small oscillation theory, but simply says the theory cannot be tested when ξ_o gets close to zero.

Figures 4 and 5 show results when the fluid viscosity was increased systematically. In Figure 4, the concentration of CMC powder was 0.088% (by weight) in an H_2O solution. The corresponding fluid viscosity was about 12 cp (centipoise) as determined by the manufacturer's calibration chart. The general shape of the P_{th} curves was not affected, but the curves are shifted upward. The minimum P_{th} is about 0.4 kPa (compared to about 0.35 kPa for pure H_2O in Figures 2 and 3). Note that this occurred at a smaller glottal width, about 0.5 mm instead of 1.0 mm. The highest P_{th} value in Figure 4 is about 0.55 kPa, which exceeds the highest values for pure H_2O by about 0.05 kPa. In Figure 5, the concentration of CMC was raised to 0.117% (by weight). The corresponding viscosity was about 20 cp. This had two effects. First, it changed the shape of the lower threshold curve such that no clear upturn in the pressure was seen for the smallest values of ξ_o. Second, it raised the overall pressures by another 0.05 to 0.10 kPa. The lowest pressure now measured was 0.45 kPa, and the highest was 0.65 kPa. Thus, it appears that there is a direct dependence of P_{th} on fluid viscosity.

To show whether this dependence is linear, the ξ_o=1.5 mm values were selected from each graph and replotted in Figure 6. It appears that there is an approximate linear relation between P_{th} and viscosity, as suggested by equation (1).

Conclusion

Results of this experimental study of the effects of prephonatory glottal width and mucosal fluid viscosity on phonation threshold pressure confirm the general predictions of the previously conducted small amplitude analysis. P_{th} rises with fluid viscosity in the mucosa. P_{th} also rises with



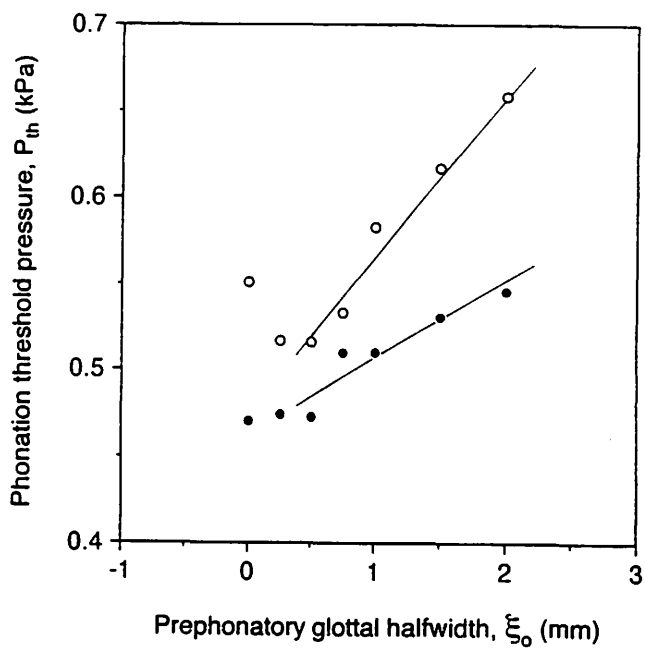


Figure 4 (left). Phonation threshold pressure for ascending subglottal pressure (open circles) and descending subglottal pressure (filled circles) as a function of glottal gap size. The fluid was a 0.088% mixture (by weight) of CMC powder with H_2O. Figure 5 (right). Phonation threshold pressure for ascending subglottal pressure (open circles) and descending subglottal pressure (filled circles) as a function of glottal gap size. The fluid was a 0.117% mixture (by weight) of CMC powder with H_2O.

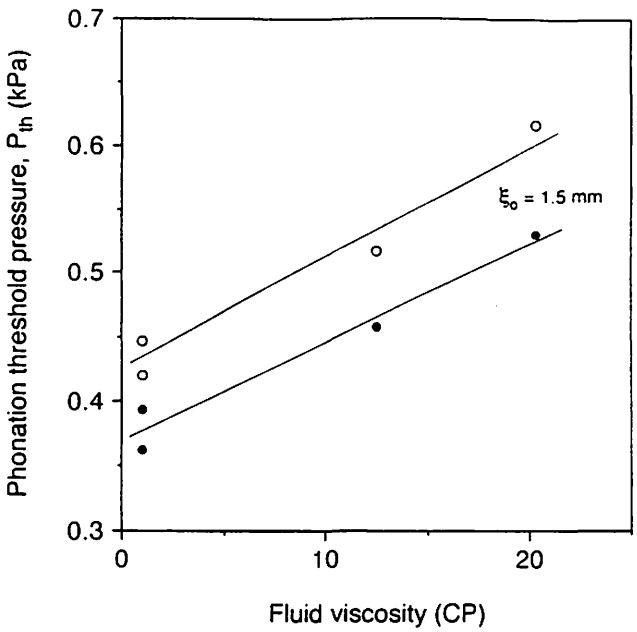


Figure 6. Phonation threshold pressure for ascending subglottal pressure (open circles) and descending subglottal pressure (filled circles) as a function of fluid viscosity. The prephonatory glottal half-width was 1.5 mm.

prephonatory glottal width. There are two findings, however, that did not emerge in the theoretical study. First, the glottal width is reduced to zero, the threshold pressure does not continue to drop to zero. Rather, its lowest value occurs for a positive value of glottal width (somewhere between 0.0 to 1.0 mm), probably because small amplitude oscillations cannot be maintained when the folds nearly touch. Collision takes over in portions of which the vibratory cycle, which invokes an entirely different set of nonlinearities and energy dissipation mechanisms. Secondly, the phonation threshold pressure is consistently higher when subglottal pressure is ascending than when subglottal pressure is descending. This hysteresis effect was also not predictable from the small amplitude linear analysis.

From a clinical and pedagogical point of view, the results suggest that an optimal glottal width may exist for "ease" of phonation, at least for soft voice. A slight separation of the vocal folds seems to have an advantage over firm approximation, or worse, a pressing together of the folds. In addition, the current results suggest that "ease" of phonation is facilitated by a reduction in viscosity of the fluid-like material in the vocal fold mucosa. This gives credibility to hydration treatments for vocal fatigue and tissue inflammation, or to augmentation of the mucosa with low-viscosity biomaterials.

Acknowledgements

This study was supported by Grant NO P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders.

References

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

Scherer, R. (1981). Laryngeal fluid mechanics: Steady flow considerations using static models. Unpublished dissertation, The University of Iowa, Iowa City.

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

Titze, I. (1992). Phonation threshold pressure: A missing link in glottal aerodynamics. Journal of the Acoustical Society of America, 91(5), 2926-2935.

Jiang, J., & Titze, I. (1993). A methodological study of hemilaryngeal phonation. Laryngoscope, 103(8), 872-882.

NCVS Status and Progress Report - 6
May, 1994, 19-31

Bifurcations in an Asymmetric Vocal Fold Model

Ina Steinecke, M.S.
Humboldt-Universität zu Berlin, Berlin, Germany
Hanspeter Herzel, Ph.D.
Technische Universität Berlin, Berlin, Germany

Abstract

A two-mass model of vocal fold vibrations is analyzed with methods from nonlinear dynamics. Bifurcations are located in parameter planes of physiological interest (subglottal pressure, stiffness of the folds). It is shown that sufficiently large tension imbalance of the left and right vocal fold induces bifurcations to subharmonic regimes, toroidal oscillations and chaos. The corresponding attractors are characterized by phase portraits, spectra and next-maximum maps. The relevance of these simulations for voice disorders such as laryngeal paralysis is discussed.

Introduction

The development of the theory of nonlinear dynamics in the past years provided the scientific world with a rich variety of methods for investigating dynamical systems. The principal understanding of the mechanisms in such systems was associated with the creation of diverse means for the description and analysis of nonlinear processes. Furthermore, many characteristic phenomena, such as bifurcations and chaos and their origins could be classified. The new view of the world attracted more and more interest from various sciences (see, e.g., the review (Lauterborn, 1988) on nonlinear dynamics in acoustics). In this way, the theory of nonlinear dynamics found an application in the analysis of voice signals, too.

Standard methods of voice analysis, such as the estimation of jitter and shimmer and the harmonics-to-noise ratio, are valuable for the characterization of normal phonation. However, in the case of newborn infant cries and voice disorders, the occurrence of abrupt changes to other regimes, for example subharmonic, diplophonic or irregular ones (Kelman, 1981; Robb and Saxman, 1988; Sirviö and Michelsson, 1976) indicates the suitability of the methods from nonlinear dynamics (Herzel and Wendler, 1991; Herzel et al., 1993; Mende et al., 1990; Titze et al., 1993). The theory predicts that in nonlinear systems, such as the voice source, for certain parameter variations periodic oscillations change to subharmonic regimes, oscillations with two frequencies and chaotic dynamics.

The application of nonlinear dynamics to time series has been demonstrated to be useful. The reconstruction of attractors and the estimation of their properties indicate low-dimensionality of the dynamics (Herzel and Wendler, 1991; Mende et al., 1990; Titze et al., 1993). The investigation of high-dimensional models of vocal fold vibration supports this assumption. It was shown by means of the empirical orthogonal functions that normal phonation is well represented by only two eigenmodes. The simulation of disordered voice has shown that the three strongest modes contain 90% of the variance (Berry et al., 1993).

Thus, it should be suitable to continue a more detailed analysis of the voice producing system at models representing only a few of the possible modes in vocal fold vibration. In fact, already the symmetric two-mass model, approaching each vocal fold by two coupled oscillators, shows bifurcations of various types and deterministic chaos (Herzel et al., 1991; Herzel, 1993; Herzel and Knudsen, 1994). Comparable phenomena are found in a high-dimensional model, representing the vocal folds as a nearly real shaped continuum and solving the partial differential equations by the finite element method (Berry et al., 1993; Titze et al., 1993). However, comprehensive bifurcation diagrams in such extensive simulations are not available yet.

Here we study a simplified two-mass model to get detailed bifurcation diagrams. The reduction of the system dimension and of the number of parameters allows us to interpret the effect of the most important parameters

quantitatively. We are aware that our study has to be supplemented by bifurcation analysis of more sophisticated models.

It has been emphasized that various mechanisms exist which lead to qualitatively changed sound. Our attention is directed to the effect of tension imbalance between both sides. This is, like in earlier studies, an approach to unilateral laryngeal paralyses (Ishizaka and Isshiki, 1976; Smith et al., 1992; Wong et al., 1991). However, the investigations in these papers were limited to specific simulations of a few representative parameter values.

In addition to the simplification of earlier models our analysis is supported by improved numerical tools. The available workstation Sparc 2 enabled extensive simulations. Furthermore, the continuation program AUTO has been used for the calculation of phonation onset under various conditions.

In the next two sections the symmetric version of our model is studied. Then, the effects of asymmetries related to different paralyses are analyzed. In both cases, the relevant bifurcations are located in parameter planes spanned by parameters of physiological meaning. In the symmetric case, the onset of phonation is discussed which corresponds to a Hopf bifurcation from a stable steady state to a limit cycle. In the asymmetric case, instabilities of the limit cycle due to tension imbalance are described by modeling the transition from normal to disturbed phonation.

The Simplified 2-Mass Model

The 2-mass model, developed and analyzed by Ishizaka and Flanagan (1972), evolved to a standard for exploring the voice producing system through the years. The main goal of the model was to synthesize voice by a self-oscillating mechanism. The basic principle is intimately related to the observed phase difference between the lower and the upper edge of the vocal fold. This effect can be modeled by representing each fold by two coupled oscillators (Fig. 1). The oscillations are driven by the lung pressure. The induced phase difference of the upper and the lower mass enables the energy transfer from the air stream to the vocal folds. For sufficiently large subglottal pressure the dissipative losses can be compensated, and phonation sets in.

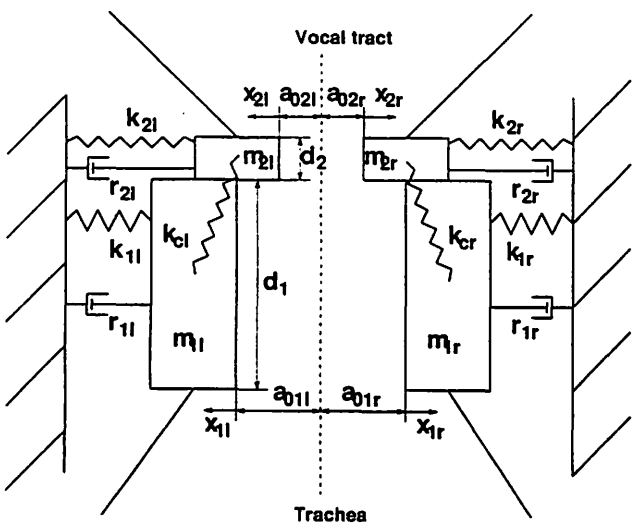


Figure 1. 2-mass model of the vocal folds.

In addition to this principal mechanism, some other facts known about the voice producing system are taken into consideration. Cubic terms describe the nonlinear elastic properties of the tissue. Furthermore, various pressure drops inside the glottis are considered, such as the pressure drop at the relatively abrupt glottal inlet according to measurements of van den Berg (van den Berg, 1957). Another kind of energy loss is the viscous damping of the air as it passes through the glottis. The most essential pressure changes due to glottal geometry are described by Bernoulli terms in the pressure equations.

In comparison with more realistic models, such as models being based on partial differential equations simulations (Titze and Talkin, 1979; Titze and Alipour-Haghighi, forthcoming book) or systems of more than two coupled oscillators (Story and Titze, 1994; Titze, 1973; Wong et al., 1991), the 2-mass model appears to be quite simple. Nevertheless, more than 10 coupled ordinary differential equations with many parameters are necessary to solve the mechanical and the aerodynamical problem (Ishizaka and Flanagan, 1972).

That is why we reduced the model to its basic principle, the phase delay between the lower and the upper edge of the vocal folds. It is emphasized that the interaction of the vocal fold geometry with the air flow according to the Bernoulli equation is the prerequisite to the self oscillating motion (Broad, 1979; Stevens, 1977).

Our aim is to analyze instabilities due to the dynamics of the coupled oscillators. Thus, we neglect all parts of description beside the basic mechanical equations. Particularly, we neglect the cubic nonlinearities of the oscillators and the viscous losses inside the glottis and separate the vocal fold vibration from the vocal tract dynamics. We are aware that neglecting the influence of vocal tract and subglottal resonances is a strong simplification of the voice source. However, bifurcations that we expect to occur are also observed in vibration patterns of excised larynxes. This fact supports our approach of analyzing extremely reduced equations, describing soleley the vocal fold dynamics.

In Ref. (Herzel, 1993; Herzel and Knudsen, 1994; Lucero, 1993) similar simplifications have been studied and it turned out that the waveforms were quite realistic and similar to simulations of the complete 2-mass model (Ishizaka and Flanagan, 1972). Moreover, the phonation onset in such

models is found at quite realistic parameter values (Herzel, 1993; Herzel and Knudsen, 1994; Lucero, 1993).

Another change regards the description of the aerodynamical problem. We neglect the pressure drop at the inlet and suppose, contrarly to Ishizaka and Flanagan, Bernoulli flow only below the narrowest part of the glottis. Above this constriction, the airstream is considered to be a jet (Titze and Alipour-Haghighi, forthcoming).

According to the idea of the 2-mass model and the modifications mentioned above, we obtain the following equations. The motion of the masses is described by equations of coupled oscillators (see also Fig. 1)

$$m_{i\alpha}\ddot{x}_{i\alpha} + r_{i\alpha}\dot{x}_{i\alpha} + k_{i\alpha}x_{i\alpha} + \Theta(-a_i)c_{i\alpha}\left(\frac{a_i}{2l}\right) + k_{c\alpha}(x_{i\alpha} - x_{j\alpha}) = F_i(x_{1l}, x_{1r}, x_{2l}, x_{2r}) \quad (1)$$

with

$$\Theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x \le 0 \end{cases}. \quad (2)$$

Besides the usual oscillator terms, the θ-function models another restoring force which acts only during the collision of both sides. The forces F_i describe the action of the pressure in the glottis.

The four variables $x_{i\alpha}$ are the displacements of masses $m_{i\alpha}$, the other variables and parameters are defined as follows:

$$\begin{aligned} a_{i\alpha} &= a_{0i\alpha} + l x_{i\alpha}, \\ a_i &= a_{il} + a_{ir}, \\ a_{0i} &= a_{0il} + a_{0ir}. \end{aligned} \quad (3)$$

$m_{i\alpha}$ - masses
$k_{i\alpha}$ - spring constants
$k_{c\alpha}$ - coupling constants
$c_{i\alpha}$ - additional spring constant during collision
$r_{i\alpha}$ - amping constants
a_{0i} - glottal rest area
l - length of the glottis

$$i, j = \begin{cases} 1 & \text{- lower mass} \\ 2 & \text{- upper mass} \end{cases}$$

$$\alpha = \begin{cases} l & \text{- left side} \\ r & \text{- right side} \end{cases}$$

The forces F_i are computed by the following equation:

$$F_i = l d_i P_i. \quad (4)$$

d_i - thickness of part i
P_i - pressure inside the glottis acting on part i

The Bernoulli equation, including the assumption of the build up of the jet, reads as follows

$$P_s = P_1 + \frac{\varrho}{2}\left(\frac{U}{a_1}\right)^2 = P_0 + \frac{\varrho}{2}\left(\frac{U}{a_{min}}\right)^2, \quad (5)$$

$$a_{min} = \min(a_{1l}, a_{2l}) + \min(a_{1r}, a_{2r}). \quad (6)$$

P_s - subglottal pressure
P_0 - supraglottal pressure
U - volume flow velocity
ρ - air density

Setting $P_0 = 0$ and taking into account that the Bernoulli equation is only relevant during the completely opened glottis, we can derive the equations

$$P_1 = P_s\left(1 - \Theta(a_{min})\left(\frac{a_{min}}{a_1}\right)^2\right)\Theta(a_1), \quad (7)$$

$$P_2 = 0, \quad (8)$$

$$U = \sqrt{\frac{2P_s}{\varrho}}\, a_{min}\Theta(a_{min}). \quad (9)$$

For numerical simulations the function θ(x) was approximated by

$$\Theta(x) = \begin{cases} \tanh\left(50\frac{x}{x_0}\right) & x > 0 \\ 0 & x \le 0 \end{cases}. \quad (10)$$

x_0 is a value to measure the gradation of the function θ(x). For computation of $\theta(a_1)$ we used $x_0 = a_1$.

In the case of $\theta(a_{min})$ we define an implicit θ-function by a new definition of a_{min}

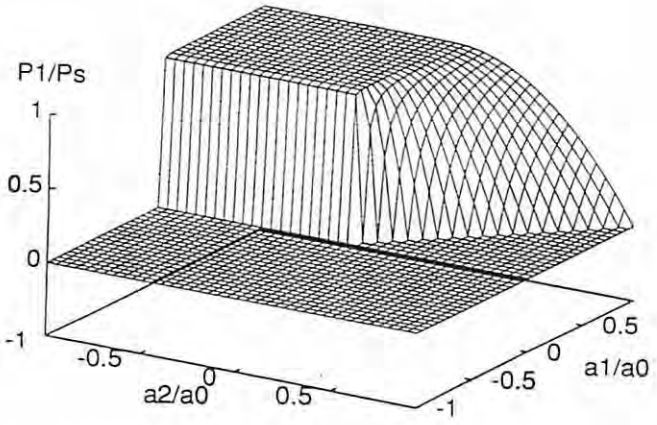
$$a_{min} = \max(0, a_{min}) \quad (11)$$

after calculation of equation 6.

This procedure results in a continuous dependence of P_1 from its arguments (Fig. 2 - next page). We note that no negative pressure values occur due to our assumption of a jet regime above the narrowest part. Thus, the asymmetric model is described by eight first order differential equations and one algebraic pressure equation.

Bifurcation Analysis of the Symmetric Model

The simplified symmetric model shows quite realistic behavior under similar parameters like the original



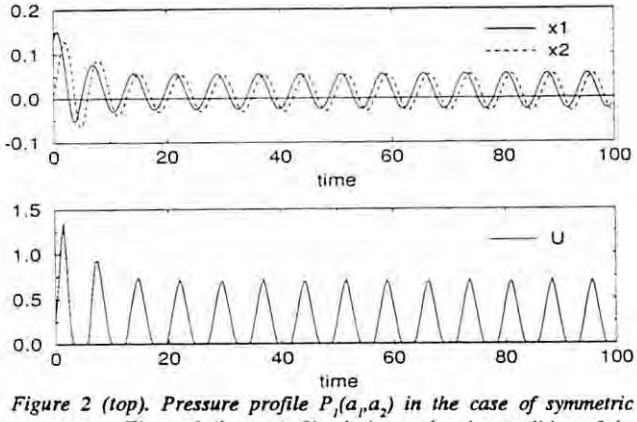


Figure 2 (top). Pressure profile $P_1(a_1,a_2)$ in the case of symmetric parameters. Figure 3 (bottom). Simulation under the condition of the standard parameter set showing oscillations of the lower and upper masses and glottal volume flow velocity U of "normal phonation".

model. The following parameters constitute our standard parameter set.

$$m_u = m_{1r} = m_1 = 0.125 \qquad r_{1l} = r_{1r} = r_1 = 0.02,$$
$$m_{2l} = m_{2r} = m_2 = 0.025, \qquad r_{2l} = r_{2r} = r_2 = 0.02,$$

$$k_{1l} = k_{1r} = k_1 = 0.08, \qquad c_{1l} = c_{1r} = c_1 = 3_{k1},$$
$$k_{2l} = k_{2r} = k_2 = 0.008, \qquad 2_{2l} = c_{2r} = c_2 = 3_{k2},$$

$$d_1 = 0.25, a_{01} = a_{01l} + a_{01r} = 0.05,$$
$$d_2 = 0.05, a_{02} = a_{02l} + a_{02r} = 0.05,$$

$$P_s = 0.008 \ (\approx 8 \text{ cm } H_2O),$$
$$\rho = 0.00113.$$

All units are given in cm, g, ms and their corresponding combinations. The equations of motion are written as

$$\dot{x}_1 = v_1, \tag{12}$$

$$\dot{v}_1 = \frac{1}{m_1}(P_1 l d_1 - r_1 v_1 - k_1 x_1 - \Theta(-a_1) c_1 \frac{a_1}{2l} - k_c(x_1 - x_2)), \tag{13}$$

$$\dot{x}_2 = v_2, \tag{14}$$

$$\dot{v}_2 = \frac{1}{m_2}(-r_2 v_2 - k_2 x_2 - \Theta(-a_2) c_2 \frac{a_2}{2l} - k_c(x_2 - x_1)), \tag{15}$$

$$P_1 = P_s \left(1 - \Theta(a_{min}) \left(\frac{a_{min}}{a_1}\right)^2\right) \Theta(a_1), \tag{16}$$

$$a_1 = a_0 + 2l x_1, \tag{17}$$

$$a_2 = a_0 + 2l x_2, \tag{18}$$

$$a_{min} = \begin{cases} a_1 & if \quad 0 < x_1 < x_2 \\ a_2 & if \quad 0 < x_2 \leq x_1 \\ 0 & \text{else} \end{cases}, \tag{19}$$

using the variables x_1, $v_1 = \dot{x}_1$, x_2 and $v_2 = \dot{x}_2$.

Fig. 3 shows the oscillations of the lower and the upper mass and the glottal volume flow for the standard parameters. The upper masses are only driven by the lower ones and oscillate with some delay which refers to the phase difference discussed above.

To describe the voice, one point of interest is the so-called voice range profile. It is defined as the part of the frequency-intensity plane in which phonation is possible. The phonation onset is associated with a Hopf bifurcation, i.e. a stationary point loses its stability and a stable limit cycle arises. In analogy to the voice range profile, we discuss the phonation onset in the P_s - k_1 plane. P_s is intimately related to the intensity and the stiffness k_1 is the principal parameter governing the frequency of the output.

Before we consider bifurcation diagrams, we want to discuss the existence and the stability of the fixed points of the system. For this simplified model an analytical treatment is possible. To get the equilibrium positions, we set the derivatives of the variables in Eqs. (12) through (15) to zero. From (12) and (14) we obtain

$$v_1 = 0, \tag{20}$$

$$v_2 = 0. \tag{21}$$

Eq. (15) is solved by

$$x_2 = \frac{k_c}{k_2 + k_c} x_1. \tag{22}$$

$$x_1 = -\frac{a_0}{2l} + \frac{ld_1P_s}{2}\frac{k_2(k_2+2k_c)}{(k_2+k_c)(k_1k_2+k_c(k_1+k_2))} \pm \sqrt{\left(\frac{ld_1P_s}{2}\frac{k_2(k_2+2k_c)}{(k_2+k_c)(k_1k_2+k_c(k_1+k_2))}\right)^2 + \frac{a_0d_1P_s}{2}\frac{k_2}{k_1k_2+k_c(k_1+k_2)}}. \quad (24)$$

Replacing x_2 by Eq. (22) in Eq. (13) we get a cubic equation for x_1. In the case of a rectangular shaped glottis, we find the trivial solution

$$x_1 = 0, \qquad x_2 = 0. \quad (23)$$

This result is independent of the elastic properties of both sides because the entire pressure is dropped at the glottal inlet. Due to our assumption of a jet above the smallest area, this is a solution for a divergent rest position of the glottis, i.e. $a_{10} < a_{20}$, too, since $P_1 = 0$ follows from $a_{min} = a_1$.

Knowing one solution of the cubic equation for the estimation of the equilibrium position of x_1, the equation can be reduced to a quadratic one the solutions of which are in equation 24 (above).

Ignoring our restriction $x_1 \geq x_2$ (Bernoulli equation applies in our model only to the convergent glottis), this result is comparable to expressions in Ref. (Lucero, 1993). However, due to our assumption of a jet, no negative pressure values occur and, hence, no steady state with $x_1 < 0$, leading to $a_1 < a_2$, exists. Below a critical value of P_s

$$P_{crit} = \frac{a_0}{4l^2d_1}\frac{k_1k_2+k_c(k_1+k_2)}{k_2} \quad (25)$$

both solutions of (24) are negative, i.e. only the trivial solution $x_1 = 0$ remains.

Note that the solution of (24) corresponding to the minus sign is always less than $-a_0/2l$ indicating a closed glottis. Thus, this fixed point does not exist in our model independent of assuming a Bernoulli flow or a jet. Lucero (1993) got a third (stable) equilibrium position for a small pressure interval. The existence of this fixed point is associated with the pressure drop describing the abrupt glottal entry. For our assumed idealized Bernoulli flow at this point, this equilibrium position does not occur.

Thus, in contrast to Refs. (Herzel, 1993; Herzel and Knudsen, 1994; Lucero, 1993), the assumption of the build up of a jet at the smallest glottal gap leads to only one equilibrium position for $P_s < P_{crit}$ (Steinecke, 1993). P_{crit} is the value at which the positive root of (24) becomes greater than zero, i.e. this branch crosses the trivial solution and becomes relevant.

Now we want to consider briefly the stability properties of the steady states. For this purpose, our system of n = 4 equations

$$\dot{x} = f(x) \quad (26)$$

is linearized at its fixed points x° with $f(x^\circ) = 0$. The evolution of a small deviation $y = x - x^\circ$ is described by the linearized equations

$$\dot{y} = Ay \quad , \quad A = \left(\frac{\partial f_i}{\partial x_k}(x^\circ)\right)_{i,k=1}^{n}. \quad (27)$$

A is termed Jacobi-matrix. If the real part of at least one eigenvalue is greater than zero, the steady state is unstable. Instead of calculating all the eigenvalues, it is possible to use the Ruth-Hurwitz criterion to get conditions for the signs of all real parts (Jetschke, 1989).

In the case of $k_c = 0$, i.e. if both masses are uncoupled, it is possible to calculate the critical values for an instability (bifurcation). Considering the equilibrium position $x_1 = 0$, $x_2 = 0$, we get as the result that it becomes unstable for

$$\frac{\partial F_1}{\partial x_1} > k_1. \quad (28)$$

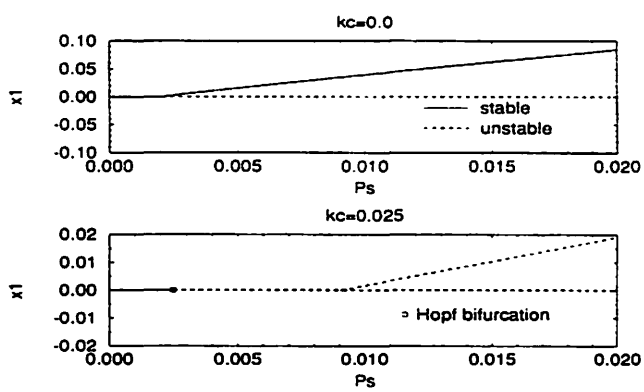
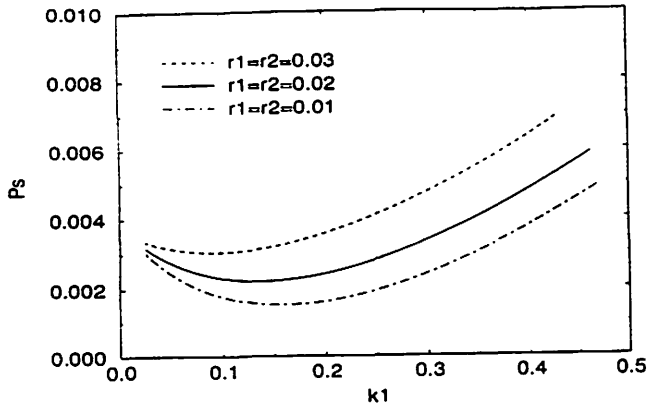


Figure 4. Equilibrium positions of x_1 under nonoscillatory ($k_c = 0$) and oscillatory conditions.



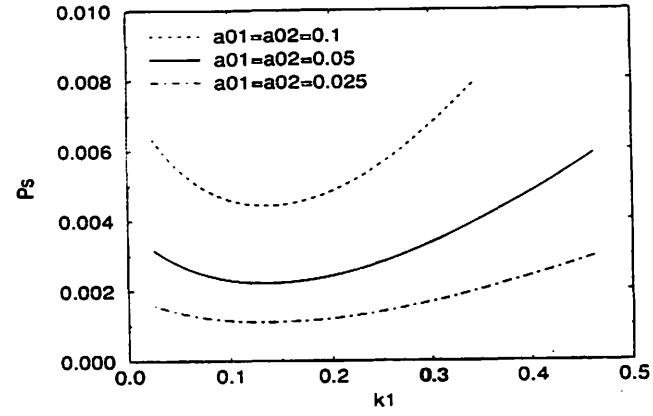


Figure 5 (top). Phonation onset in the k_1 - P_s plane via a Hopf bifurcation in dependence on damping parameter values. The solid line corresponds to the standard values. Figure 6 (bottom). Phonation onset in the k_1 - P_s plane in dependence on glottal rest area. The solid line corresponds to the standard values.

This is equivalent to the requirement (Steinecke, 1993)

$$P_s > P_{crit}. \tag{29}$$

The branch from Eq. (24) which crosses the trivial solution at this point becomes stable, i.e. an exchange of stability, termed transcritical bifurcation, occurs (see also Lucero, 1993).

All in all, there is in contrast to the analysis of Lucero (1993) below P_{crit} only one stable fixed point and for $P_s > P_{crit}$ one stable and one unstable fixed point exist (see Fig. 4). The transcritical bifurcation corresponds to a change of the glottal shape from a rectangular to a convergent one.

For $k_c > 0$ the bifurcations were determined by using the continuation program AUTO. Continuation techniques allow the location of bifurcation lines in parameter planes with moderate effort (Kubicek and Marek, 1983). At $k_c = 0.025$ (Fig. 4) the trivial equilibrium position loses its stability via a Hopf bifurcation and oscillations set in.

In Figs. 5 and 6 the phonation onset is drawn in dependence on damping parameters and glottal rest area, respectively. Below the lines a stable fixed point exists and oscillations are damped out. For increasing pressure phonation sets in. The effect of both of the parameters is to be seen. As expected, the threshold increases with the damping constants.

An increased glottal rest area leads to a significant increase of the phonation threshold for all stiffness values. Moreover, oscillations cease to exist for higher values of k_1. This result should be kept in mind in so far that a paralysis of the nervus laryngeus inferior is associated with a fixing of the affected vocal fold in the paramedian or the intermedian position (Wendler and Seidner, 1987).

Modeling Unilateral Paralyses

Unilateral laryngeal paralyses lead to changed glottal configurations and tension imbalance between both sides. It has been shown in experiments with excised canine larynxes that such imbalance results in diverse oscillatory patterns of the vocal folds (Isshiki et al., 1977). Depending on the glottal rest area three patterns could be distinguished. In the case of extremely narrow or wide glottal gaps the oscillations are quite regular, the tenser side remarkably precedes the flaccid one. In the case of moderate rest areas the oscillations show alternating amplitudes, and for highly increased imbalance the vibrations become irregular (Isshiki et al., 1977).

The degree of the disturbance of the voice quality depends on the localization of the nerve injury (Pahn et al., 1984; Wendler and Seidner, 1987). The laryngeal muscles are essentially innervated by the nervus laryngeus superior and the nervus laryngeus inferior. The superior laryngeal nerve paralysis results in the deficiency of the cricothyroid muscle. This muscle is responsible for the principal tension of the vocal fold. That is why, the voice is often limited in the phonation of higher tones. Furthermore, patients sometimes have difficulties in intonation (Wendler and Seidner, 1987).

In the case of an inferior laryngeal nerve paralysis (recurrent paralysis) all intraglottal muscles except the cricothyroid are involved. Therefore, the vocalis muscle cannot contract, and because of the absence of adductors the vocal cord cannot be brought into phonatory position. The active cricothyroid muscle partially compensates for this effect by imposing passive stress on the vocal fold which is mostly fixed in the paramedian position or just off midline. The voice quality is highly influenced by this fixing position and the actual position of the arytenoid cartilage (Wendler and Seidner, 1987).

Under conditions of a combined paralysis of both nerves, mostly caused by an injury of the vagal nerve, the vocal fold is often fixed in the intermedian position and the voice is highly disturbed (Pahn et al., 1984; Wendler and Seidner, 1987).

For a very large glottal gap breathiness due to turbulence is the dominant deviation from normal phonation. However, in several cases subharmonics, modulations and irregularities have been reported for patients with paralytic dysphonia (Hammarberg et al., 1986; Herzel et al., 1993).

Now, we want to investigate the simplified 2-mass model under the conditions of unilateral paralysis to find correlations between changed glottal parameters and the type of vibrations.

For this purpose we define two tension parameters Q and $\tilde{Q}$ as an attempt to describe the activity of the cricothyroid and the vocalis muscle, respectively. The cricothyroid muscle stresses the whole vocal fold. That means that the elastic properties of all parts of the vocal fold are affected. The corresponding parameters Q_α is defined in a similar way as in Ref. (Ishizaka and Flanagan, 1972)

$$k_{i\alpha} = Q_\alpha k_{i\alpha 0}, \tag{30}$$

$$k_{c\alpha} = Q_\alpha k_{c\alpha 0}, \tag{31}$$

$$c_{i\alpha} = Q_\alpha c_{i\alpha 0}, \tag{32}$$

$$m_{i\alpha} = m_{i\alpha 0}/Q_\alpha. \tag{33}$$

The index 0 denotes the standard parameter set. The tension parameters Q_α describe the change of the eigenfrequencies as follows:

$$\omega_{i\alpha} = Q_\alpha \omega_{i\alpha 0}, \tag{34}$$

$$\omega_{i\alpha}^2 = \frac{k_{i\alpha}}{m_{i\alpha}}. \tag{35}$$

The contraction of the vocalis muscle only effects its own tension. Supposing that the properties of the vocalis muscles are represented by the masses m_α, $\tilde{Q}_\alpha$ may describe their tensions in the following way (Hirano, 1981; Smith et al., 1992):

$$k_{1\alpha} = \tilde{Q}_\alpha k_{i\alpha 0}, \tag{36}$$

$$c_{1\alpha} = \tilde{Q}_\alpha c_{i\alpha 0}, \tag{37}$$

$$m_{1\alpha} = m_{i\alpha 0}/\tilde{Q}_\alpha. \tag{38}$$

In the next sections we discuss bifurcations due to varying parameters Q and $\tilde{Q}$.

Bifurcation Diagrams of Asymmetric Vocal Fold Vibration

Modeling Superior Nerve Paralysis

Unilateral superior nerve paralysis is modeled by tension imbalance $Q_l \neq Q_r$. Inspections of Eqs. (1) reveals that the transformation of the time $t \rightarrow Q_l t$ and of the subglottal pressure $P_s \rightarrow P_s/Q_l$ reduces the problem in a way that the left side remains uninvolved. Hence, the asymmetric properties are described solely by $Q = Q_r/Q_l$ as the ratio of the eigenfrequencies of both sides.

$$\begin{aligned} m_{il}&=m_{i0}, & k_{il}&=k_{i0}, & c_{il}&=c_{i0}, \\ m_{ir}&=m_{i0}/Q, & k_{ir}&=Qk_{i0}, & c_{ir}&=Qc_{i0}. \end{aligned} \tag{39}$$

The scaling of the time corresponds to higher frequencies at higher tension, the scaling of the subglottal pressure refers to the necessity of an increased pressure to displace the stressed vocal folds. The transformation implies that we can restrict our analysis to $Q \leq 1$ since any situation with $Q_l > Q_r$ can be rescaled to a case with $Q < 1$.

In addition to Q as a measure of asymmetry, the subglottal pressure P_s takes an important part. P_s is a measure of the energy that is pumped into the system. Furthermore, the higher P_s, the stronger both sides are coupled by the induced airflow. For $P_s \rightarrow 0$ both sides are separated and cannot oscillate.

Along this line, we want to investigate the vibratory behavior in the Q - P_s - parameter plane. According to the transformation and physiological relevance, it is sufficient to consider only values with $0.4 < Q < 1$.

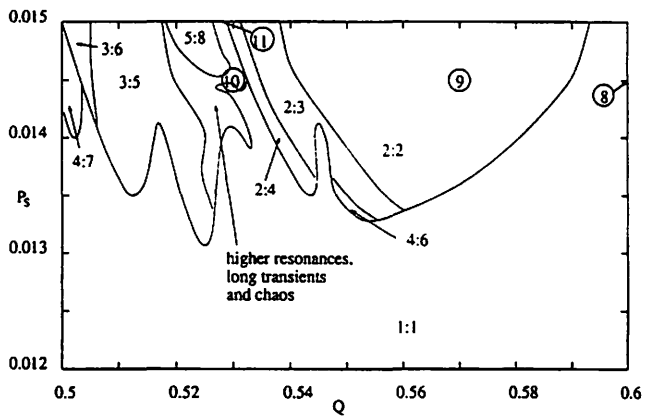


Figure 7. Bifurcation diagram in the Q - P_s plane showing different types of attractors labeled by the ratio of maxima of x_{1r}:x_{1l} during one oscillatory cycle.

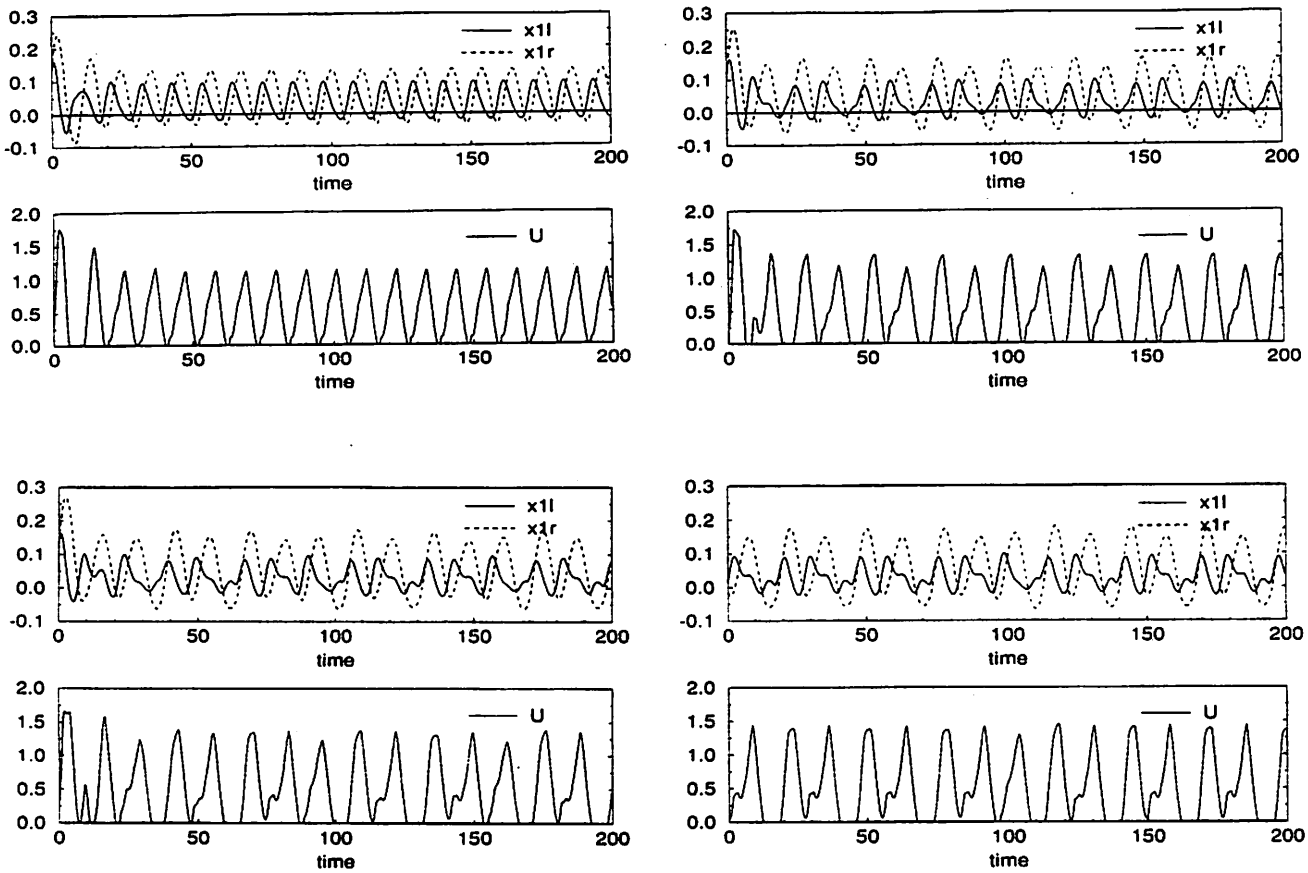


Figure 8 (upper left). Simulations for $P_s = 0.0145$ and $Q = 0.6$ showing 1:1 locking of both folds. Figure 9 (upper right). Simulations for $P_s = 0.0145$ and $Q = 0.57$ showing 2:2 locking of both folds. Figure 10 (lower left). Simulations for $P_s = 0.0145$ and $Q = 0.53$ showing 5:8 locking of both folds. Figure 11 (lower right). Simulations for $P_s = 0.0150$ and $Q = 0.529$ showing chaos.

Because bifurcations primarily occur in the case of significant asymmetry, in Fig. 7 only a part of the Q - P_s plane is shown. The bifurcation diagram is based on 400 ms simulations of the equations of motion on a grid in the plane ($\Delta Q = 0.001$, $\Delta P_s = 0.0002$). Initial conditions are the same for each simulation ($x_{1l}(0) = x_{1r}(0) = 0.1$, $\dot{x}_{1l}(0) = \dot{x}_{1r}(0) = 0.1$, $x_{2l}(0) = x_{2r}(0) = 0$ and $\dot{x}_{2l}(0) = \dot{x}_{2r}(0) = 0$).

In the shown region, $0.5 < Q < 0.6$, a variety of bifurcations can be detected. Regions of different attractors are labeled with the ratio of the number of the maxima of x_{1r} and x_{1l} during one oscillatory cycle. Normal phonation corresponds, for example, to one maximum of the elongations $x_{i\alpha}$ labeled by 1:1. Alternating amplitudes, related to period-doubling, imply a ratio 2:2. The circled numbers in Fig. 7 refer to the parameter values studied in the following Figures 8 through 11. In each figure the oscillations of the lower masses and the glottal volume flow velocity are represented. In Fig. 12 the Fourier spectra of the corresponding time series $x_{1r}(t)$ are presented and Figs. 13 and 14 display phase portraits and next-maximum maps of these simulations.

The frequency of the motion is governed by the frequency of the flaccid side, i.e. the lower Q, the lower fundamental frequency. This result is related to the inability of patients to phonate higher tones and the difficulty to hit the intended tone.

Fig. 8 shows the locking of both sides in a 1:1 ratio. The healthy vocal fold, x_{1l}, precedes the affected side, which oscillates with a greater amplitude. In Fig. 9 we can find period-doubling behavior, leading to alternating pulses in the air flow. Such a subharmonic regime has been often observed in voice pathology and is associated with such terms as "octave jump" or "dicrotic dysphonia". In Fig. 10 the right vocal fold oscillates with period 5. The left side with its higher eigenfrequency is locked in such a way that its motion has 8 maxima during one cycle. This behavior leads to an increased intensity of subharmonic components (see Fig. 12). For the parameters in Fig. 11 the vibrations become irregular. The corresponding Fourier spectrum in Fig. 12 shows besides the peaks of the basic frequency and the first subharmonic a quite unstructured contour.

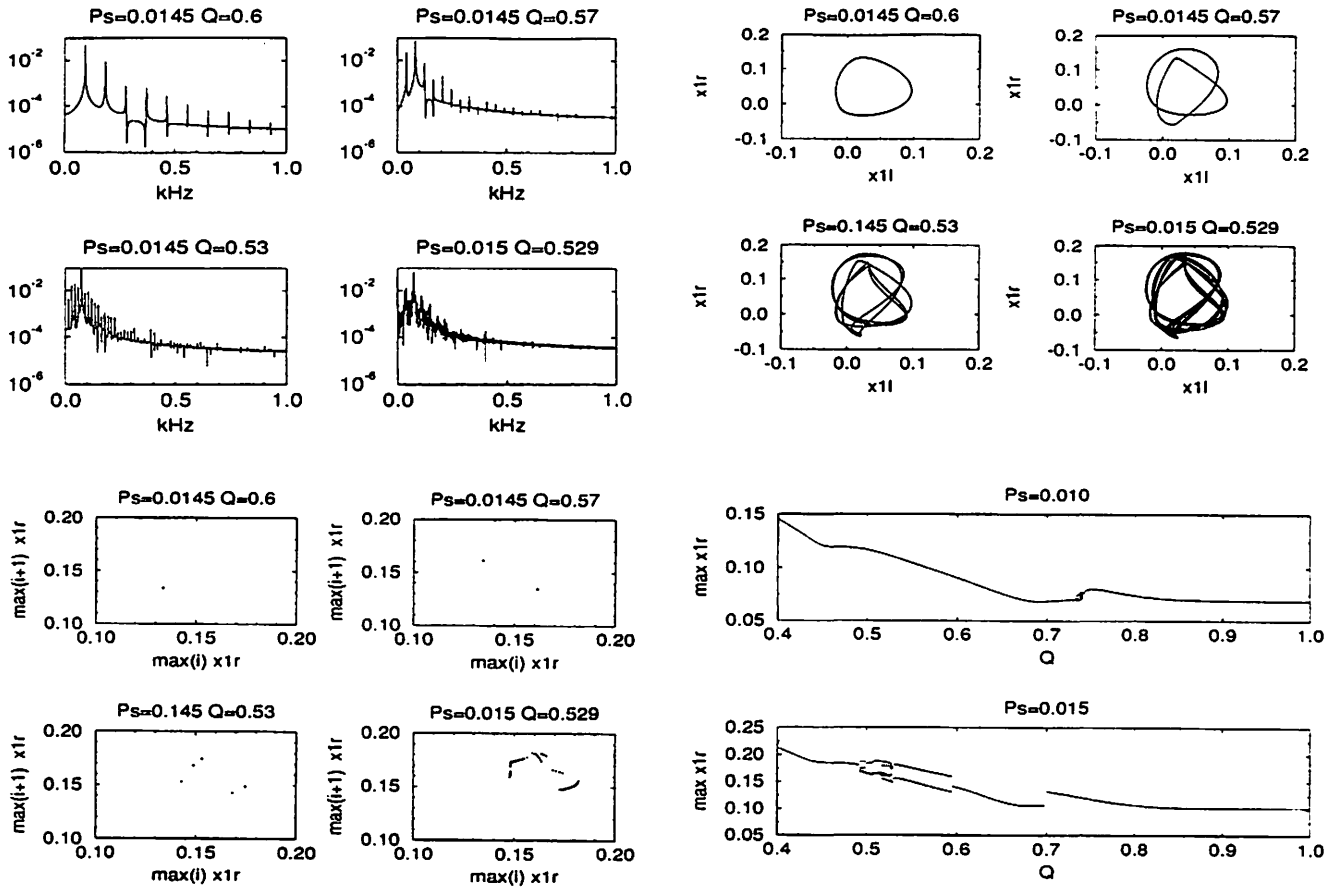


Figure 12 (upper left). Fourier spectra of the time series of $x_{1r}(t)$ showing the occurrence of subharmonics of the 2:2 and the 5:8 attractor and a more unstructured spectrum for the chaotic oscillations. Figure 13 (upper right). Attractors in the x_{1l} - x_{1r} plane. Figure 14 (lower left). Plot of consecutive maxima of x_{1r}. Periodic oscillations of with n maxima result in n points. Figure 15 (lower right). 1-parameter-bifurcation diagrams showing maxima of x_{1r} of the attractors.

Another tool for the analysis of complex motion are phase portraits and next-maximum maps. Fig. 13 displays the projection of the attractors into the x_{1l} - x_{1r} plane. The plot of consecutive maxima of x_{1r} is shown in Fig. 14. These figures support our claim that the observed complex oscillations corresponds to subharmonic regimes and chaos, respectively.

It is emphasized that qualitative changes in the voice signal which results from the flow U are governed mainly by the multiple period of the behavior of x_{1r}, i.e. a period-doubling, -tripling etc. in the motion of $x_{1r}(t)$ leads to the bifurcation in U(t), whereas the period of x_{1l} influences the sound, i.e. the intensity of harmonics in the spectrum by forming the shapes of the glottal flow pulses.

All in all, transitions to oscillations with a multiple period dominate in the Q - P_s plane, while chaos seldom occurs. Furthermore, these bifurcations are located at low Q values and relatively high P_s. Note, that for tensed vocal folds the subglottal pressure has to be increased even more to obtain the same bifurcations according to the transformation mentioned above. Therefore, instabilities of the voice are to be expected mostly at high intensity and low fundamental frequency.

In Fig. 15 the maxima of x_{1r} are plotted versus asymmetry Q. At $P_s = 0.010$ a small parameter region shows period-doubling and chaos. The plot for higher pressure $P_s = 0.015$ is characterized by abrupt transitions to other regimes, indicating hysteresis, i.e. the bifurcations are associated with the coexistence of two different attractors.

In such a case it is worth studing the basins of attraction of both attractors. Fig. 16 shows the basins of the 1:1 and the 2:2 attractor at parameters Q = 0.585 and P_s = 0.0145 for varying initial displacements of the masses. The structure of the regions is relatively complicated. To see if there is a more differentiated structure, a small parameter region has been tested with a higher precision. In Fig. 16 it is marked by a dark area of the basin of the 1:1 attractor (due to the higher density of points) surrounding a mitten-like area of the 2:2 basin. However, the borderline between the basins seems to be smooth. Intertwined basins of attraction

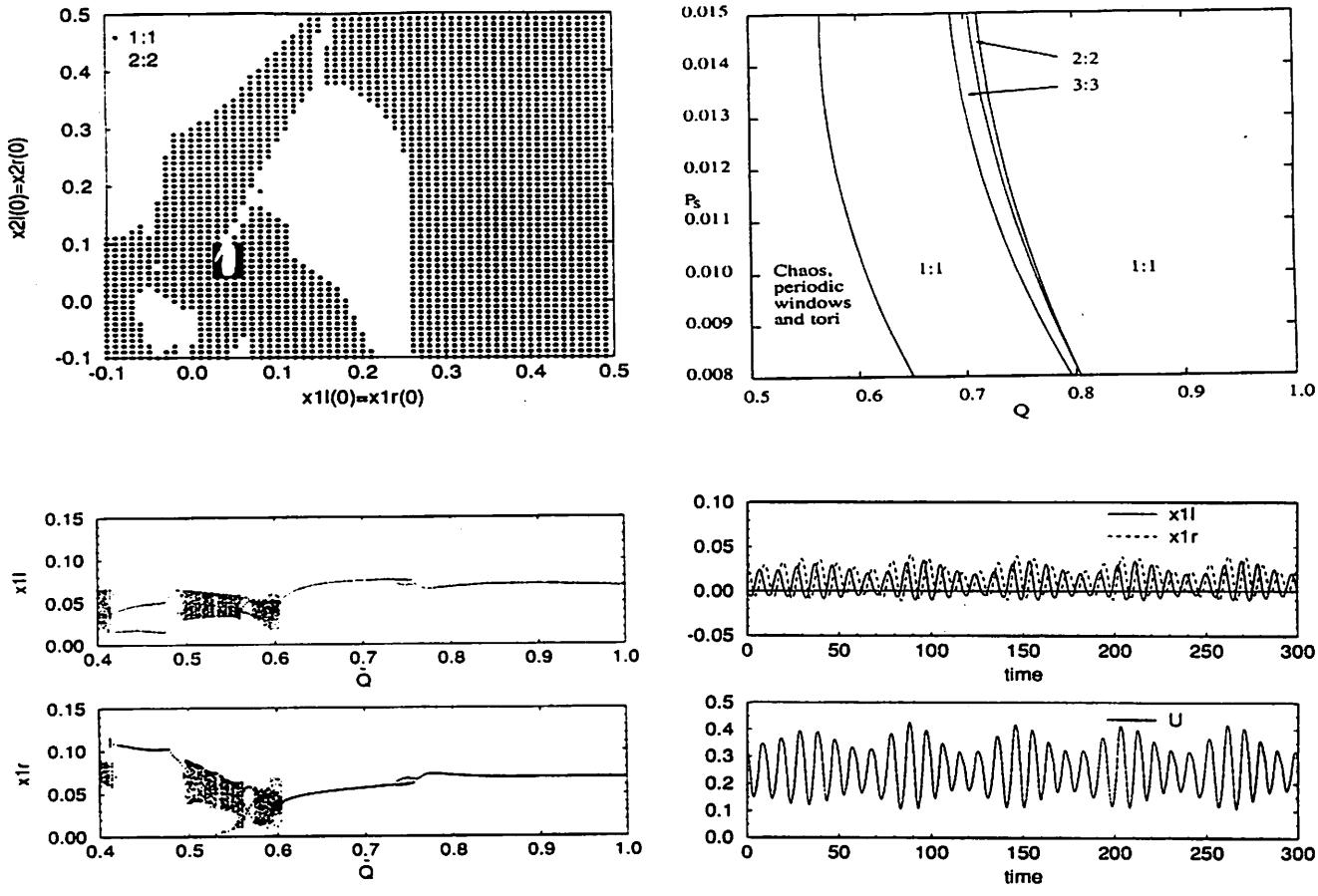


Figure 16 (upper left). Basins depending on initial displacements for Q = 0.585 and P_s = 0.0145. Simulations are carried out on a grid in the considered plane ($\Delta x_i(0)$=0.01). If the motion finished in the 1:1 attractor a star was plotted. White areas correspond to the basin of the 2:2 attractor. Figure 17 (upper right). Bifurcation diagram in the $\tilde{Q}$ - P_s plane. Figure 18 (lower left). 1-parameter-bifurcation diagrams for P_s = 0.010 showing maxima of x_{1l} and x_{1r} of the attractors, respectively. Figure 19 (lower right). Simulations for P_s = 0.008 and $\tilde{Q}$ = 0.64 showing toroidal oscillations.

imply that rather weak perturbations may induce abrupt jumps to the other regime as observed in acoustic signals (Herzel et al., 1991; Herzel et al., 1993; Kelman, 1981; Mende et al., 1990; Ramig et al., 1988; Sirviö and Michelsson, 1976).

Modeling Recurrent Nerve Paralysis

Recurrent nerve paralysis is associated with a change of the glottal rest area. Considering the rectangular shaped glottis, a fixed rest position of only one side does not show any asymmetric effect. The glottal gap affects mainly the phonation onset as discussed in section 3.

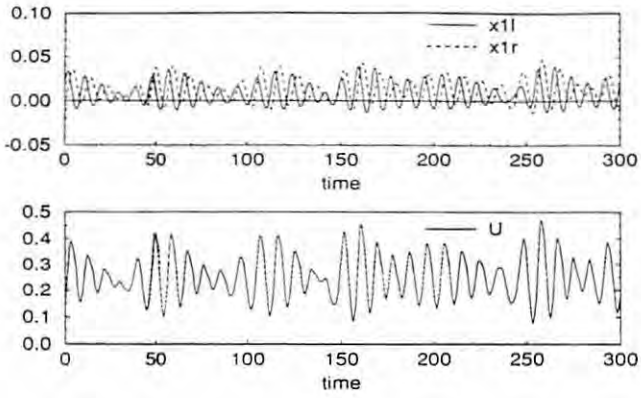
In this section we want to investigate bifurcations in the $\tilde{Q}$ - P_s plane. Despite the absence of a transformation as in preceding section, we restrict the analysis of the $\tilde{Q}$ dependence to the same interval as for Q to underscore the different effects of these parameters.

The bifurcation diagram in the $\tilde{Q}$ - P_s plane ($\Delta\tilde{Q}$ = 0.01, ΔP_s = 0.0002) in Fig. 17 shows another structure as Fig. 7. There are significant differences to be seen.

The small $\tilde{Q}$ interval with period-doubling or -tripling at $\tilde{Q} \approx 0.7$ exists over the whole P_s scale. Bifurcations already occur at relatively large values of $\tilde{Q}$, i.e. less asymmetry. The region at $\tilde{Q} \approx 0.5...0.6$ shows a variety of different regimes, containing tori and chaos. The 1-dimensional bifurcation diagram in Fig. 18 displays the occurrence of large regions of irregular motion with interspersed periodic windows. Contrarly to the behavior in the Q - P_s plane, standard subglottal pressure 8cm H^2O is sufficient to reach these bifurcations.

Figs. 19 and 20 show toroidal and chaotic oscillations, respectively. In Figs. 21 and 22 the corresponding spectra, attractors and consecutive maxima are plotted.

In addition to the occurrence of other types of attractors, bifurcations of the voice signal are not to be seen in the behavior of the affected side. In the case of frequency locking the period of $x_{1r}(t)$ does not necessarily agree with that of the glottal volume flow velocity as in the case of simulations due to the Q asymmetry.



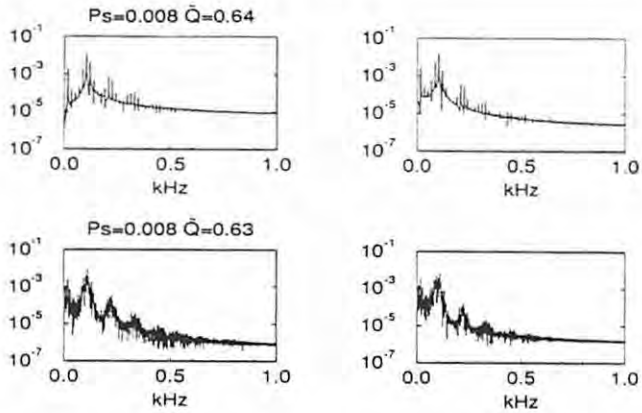


Figure 20 (top). Simulations for $P_s = 0.008$ and $\tilde{Q} = 0.63$ showing chaotic oscillations. Figure 21 (bottom). Fourier spectra of x_{1l} and x_{1r} display the difference between the toroidal and the chaotic regime.

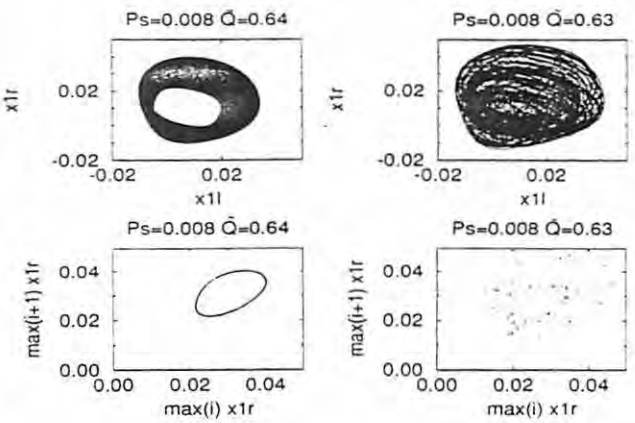


Figure 22. Attractors in the x_{1l} - x_{1r} plane and plot of consecutive maxima of x_{1r}, respectively.

Summary and Discussion

The aim of the paper was to analyze bifurcations which are closely related to observations in voice pathology. For example, in Ref. (Herzel et al., 1993) period-tripling and low-frequency modulation have been analyzed for patients with laryngeal paralysis which are qualitatively comparable to the simulations above. In order to get comprehensive bifurcation diagrams in appropriate parameter planes we had to simplify the classic Ishizaka-Flanagan model. However, inspection of the time-series and the study of the phonation onset revealed that the simplified model retained some physiological relevance.

The main focus of the paper was the location of instabilities due to different stiffness values of the left and right vocal fold being characteristic for laryngeal paralysis. It was found that overcritical tension imbalance leads to subharmonic regimes, two frequency oscillations and chaos. In many cases, the transitions are associated with hysteresis, i.e. in certain parameter regions different attractors coexist. For such a case the basins of attraction of two coexisting limit cycles have been analyzed. It turned out that these basins are intertwined which implies an additional source of unpredictability. For slightly different initial conditions rather distinct asymptotic dynamics results.

Our paper is embedded in continuing studies on implications of nonlinear dynamics for voice research. In papers devoted to signal analysis three different kinds of attractors have been discussed (Herzel et al., 1993; Titze et al., 1993): subharmonic regimes, toroidal oscillations and chaotic attractors. These regimes are often observed in newborn cries and voice disorders. They have been termed, for example "octave jumps", "double harmonic break", "diplophonia", "dicrotic dysphonia" or "creaky voice".

In our model these three attractor types have been identified with the aid of phase portraits, next-maximum maps and power spectra. The origin of the instabilities in our simplified model can be traced back to the desynchronization of two oscillators, the left and right fold. For sufficiently large effects of nonlinearities (related to large subglottal pressure) and overcritical detuning of the eigenfrequencies, complex oscillation patterns are found. Hence, the hypothesis is substantiated that many observed voice instabilities are due to the desynchronization of a few principal modes of vocal fold vibrations (Berry et al., 1993).

Earlier studies devoted to asymmetric vocal fold models (Ishizaka and Isshiki, 1976; Smith et al., 1992; Wong et al., 1991) were restricted to simulations at a few parameter values or, at most, one-parameter variations. As a major methodical result of our study we claim that the complex bifurcation scenarios require the analysis of (at least) two-parameter variations. Continuation techniques, such as AUTO may support direct simulations of the model equations. An already widely used two-parameter "bifurcation diagram" is the voice range profile. The location and

characterization of instabilities in such planes of accessible parameters may be the appropriate strategy to compare experimental observations with models quantitatively.

Future studies will be devoted to a quantitative comparison of voice disorders and simulations along this direction. As a first step, bifurcation diagrams, such as Fig. 7 in this paper, based on more sophisticated models and excised larynx studies are desirable.

Acknowledgement

We gratefully acknowledge the colleagues from the National Center for Voice and Speech for illuminating and stimulating discussions and Katharina Krischer for advice regarding continuation techniques. This work was supported by the Deutsche Forschungsgemeinschaft and by Grant No. P6000976 from the National Institute on Deafness and Other Communication Disorders.

References

Berry, D.A., I.R. Titze, H.Herzel, and K.Krischer. "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am., submitted.

Broad, D.J. "The new theories of vocal fold vibration," In N.J. Lass, editor, Speech and Language: Advances in Basic Research and Practice Vol.2. Academic Press.

Hammarberg, B., B.Fritzell, J.Gauffin, and J.Sundberg (1986). "Acoustics and perceptual analysis of vocal dysfunction," J. of Phonetics, 533—547.

Herzel, H., I.Steinecke, W.Mende, and K.Wermke (1991). "Chaos and bifurcations during voiced speech," In Complexity, Chaos and Biological Evolution (Plenum Press), 41—50.

Herzel, H.and J.Wendler (1991). "Evidence of chaos in phonatory signals," In Proc. EUROSPEECH Genova, 263—266.

Herzel, H. (1993). "Bifurcation and chaos in voice signals," Appl. Mech. Rev., 46, 399—413.

Herzel, H., D.Berry, I.Titze, and M.Saleh (1993). "Analysis of vocal disorders with methods from nonlinear dynamics," J. Speech Hearing Research, submitted.

Herzel, H.and C.Knudsen (1994). "Bifurcations in a vocal fold model," Nonlinear Dynamics, in press.

Hirano,M. (1981). "Structure of the vocal fold in normal and desease states anatomical and physical studies," Technical Report 11, ASHA, 11—30.

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

Ishizaka, K. and N.Isshiki (1976). "Computer simulation of pathological vocal-cord vibration," J. Acoust. Soc. Am., 60, 1193—1198.

Isshiki, N., M.Tanabe, K.Ishizaka, and D.Broad (1977). "Clinical significance of asymmetrical vocal cord tension," Am. Otol., 86, 58—66.

Jetschke, G. (1989). Mathematik der Selbstorganisation. (Deutscher Verlag der Wissenschaften, Berlin).

Kelman, A.W. (1981). "Vibratory pattern of the vocal folds," Folia Phoniatr., 33, 73—991.

Kubivek, M. and M.Marek (1983). Computational Methods in Bifurcation Theory and Dissipative Structures (Springer-Verlag, New York).

Lauterborn, W. and U.Parlitz (1988). "Methods of chaos physics and their application to acoustics," J. Acoust. Soc. Am., 84, 1975—1993.

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

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

Pahn, J., R.Dettmann, and Sram (1984). "Zur Verteilung und funktionellen Auswirkung von Paresen der Stimmlippenbewegungs- und -spannungsmuskulatur anhand elektromyographischer Untersuchungen," Folia phoniatr., 36, 273—283.

Ramig, L.A., R.C. Scherer, I.R. Titze, and S.P. Ringel (1988). "Acoustic anaylsis of voices of patients with neurologic disease: Rationale and preliminary data," Ann. Otol. Rhinol. Laryngol., 97, 164—172.

Robb, M.P. and J.H. Saxman (1988). "Acoustic observations in young children's non-cry vocalizations," J. Acoust. Am., 83, 1876—1882.

Sirvio, P. and K.Michelsson (1976). "Sound-spectrographic cry analysis of normal and abnormal newborn infants," Folia phoniat., 28, 161—173.

Smith, M.E., G.S. Berke, B.R. Gerratt, and J.Kreiman (1992). "Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration," J. Speech. Hear. Res., 35, 545—554.

Steinecke, I. (1993). "Untersuchungen an einem vereinfachten Stimmlippenmodell," Diplomarbeit, Humboldt-Universitat zu Berlin.

Stevens, K.N. (1977). "Physics of laryngeal behavior and larynx modes," Phonetica, 34, 264—279.

Story, B. and I.R. Titze (1994). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am., in review.

Titze, I.R. (1973). "The human vocal cords: a mathematical model. Part I," Phonetica, 28, 129—170.

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

Titze, I.R., R.Baken, and H.Herzel (1993). "Evidence of chaos in vocal fold vibration," In I.R. Titze, editor, Vocal Fold Physiology: New Frontiers in Basic Science (Singular Publishing Group), 143—188.

Titze, I.R. and F.Alipour-Haghighi (1994). Myoelastic Aerodynamic Theory of Phonation. (forthcoming book).

vanden Berg, J., J.T. Zantewa, and R.Doornenbal, Jr (1957). "On the air resistance and the bernoulli effect of the human larynx," J. Acoust. Soc. Am., 29, 626—631.

Wendler,J. and Wolfram Seidner (1987). Lehrbuch der Phoniatrie, (Georg Thieme Verlag Leipzig).

Wong, D., M.R. Ito, N.B. Cox, and I.R. Titze (1991). "Observation of perturbation in a lumped-element model of the vocal folds with application to some pathological cases," J. Acoust. Soc. Am., 89, 383—394.

NCVS Status and Progress Report - 6
May, 1994, 33-41

The Effect of Subglottal Resonance Upon Vocal Fold Vibration

Stephen F. Austin, M.M., Ph.D.
Laboratory for Research of the Singing Voice, Louisiana State University
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology,The University of Iowa

Abstract

An excised larynx experiment was undertaken to investigate the interaction between acoustic pressures in a subglottal tube and the amplitude of vibration of the vocal folds. Pressure was measured beneath the vocal folds during three specific moments of the vibratory cycle: (1) when the superior margin of the vocal folds began to separate, (2) when the vocal folds were maximally apart and (3) when the inferior margin of the vocal folds began to touch. Results indicate that the amplitude of vibration increases when the acoustic pressure beneath the vocal folds are positive at the moment of opening and negative when the vocal folds are maximally apart. The implications of these experiments to involuntary register transitions in the human voice are discussed in light of a previously proposed register theory.

Introduction

Subglottal pressure has two components the mean pressures that result from expiratory effort and an acoustic component that results from a resonance in the subglottal airway[1,2,3]. This paper addresses the effect that the acoustic component has on the amplitude of vibration of the vocal folds.

It appears that resonance in the subglottal airway can influence the way the vocal folds vibrate. According to Large[4], Nadoleczny-Millioud and Zimmerman[5] concluded that the influence of resonances in the subglottal airway undoubtedly had some effect upon the vocal registers in singing. Van den Berg[6] explained that register breaks are experienced by singers at the frequencies of the subglottal resonances and therefore must be influential in their occurrence. Even with very limited evidence, these ideas have been compelling enough to influence pedagogical theories of voice training. Vennard[7] describes the importance of subglottal resonance in explaining why the chest to falsetto (or head) register transition occurs in both men and women at about the same frequency (around E_4). He concludes that it could only be the similarity in the subglottal resonances that could account for this because anatomical differences in the size and shape of the larynges between the sexes are too large.

The subglottal airway has been a difficult place to measure acoustic pressures on normal subjects. For this reason, early estimates of subglottal pressure were made on cadavers of large dogs and humans[8] and laryngectomies[9]. The development of miniature pressure transducers, however, has enabled investigators to measure subglottal pressures in a few special subjects. Small transducers can now be mounted on the end of a catheter and placed in the subglottal airway via the glottis. This makes it possible to measure the acoustic pressure changes that occur within the vibratory cycle and accurately determine the resonant frequencies[10,11]. Subglottal formant frequency data gathered in this manner differ significantly from early estimates, as can be seen in Table 1. Note, in particular, that the measurements by van den Berg et al. on a cadaver underestimated the first three formant frequencies, and the measurement by Ishizaka et al. overestimated the first three formant frequencies, in comparison to more recent measurements.

Table 1.
Subglottal formant frequencies and band widths reported in the literature.

	F_1'	F_2'	F_3'	B_1'	B_2'	B_3'
van den Berg (1960)	300	870	1427	120	150	
Ishizaka et al. (1967)	640	1400	2100	155	140	
Boves (1984)	475	1175	1945	275	330	340
Cranen & Boves (1987)	510	1355	2290	104	154	358

Influence of Subglottal Resonance Upon Aerodynamic Driving Pressure

Titze proposed a hypothesis of vocal registers that considers the subglottal acoustic pressure as part of the driving pressure of the vocal folds during phonation[12]. Based on the first subglottal formant frequency F_1' being 510 Hz[13], Titze described conditions of constructive and destructive interference between vocal fold movement and formant pressures. The amplitude of vibration of the vocal folds, he claimed, could be significantly affected by these interactions, so that register transitions in the voice may be triggered by this interference phenomenon.

Figure 1 shows several conditions for which constructive and destructive interference was proposed to occur.The resonance of the first subglottal formant is represented as a decaying sinusoid at 510 Hz.The dashed line for each condition is the glottal area curve (idealized with an open quotient of 0.5) for one period of vocal fold vibration. Glottal opening begins on the horizontal line and maximum amplitude of the glottis occurs at the apex of the dashed curve.

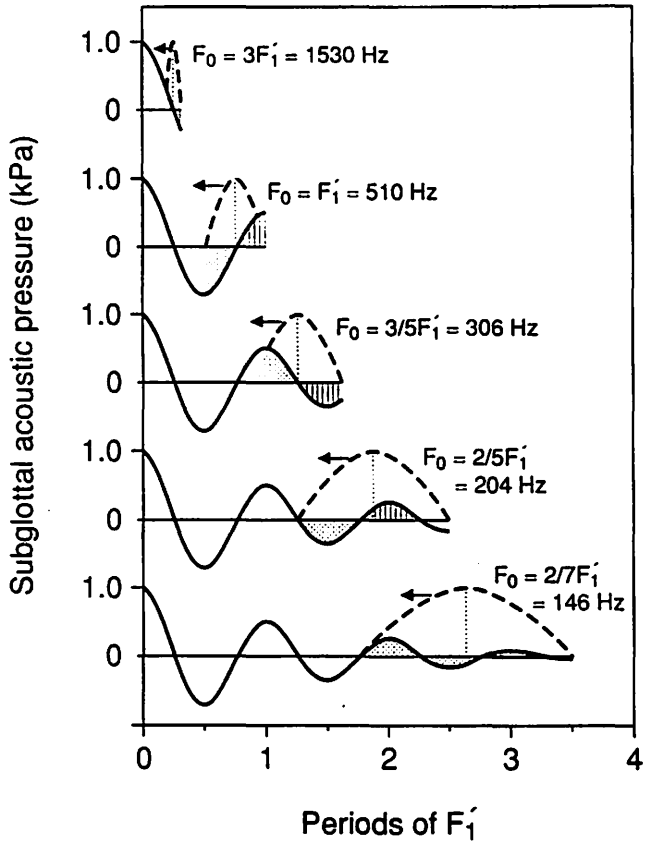


Figure 1. Phase relationships between the acoustic pressure of a fixed first subglottal formant F_1' (solid lines) and the glottal area waveform (dashed lines) for systematically increasing fundamental frequency (bottom to top). From Titze, I.R. (1988). A framework for the study of vocal registers. J. Voice, 2,(3) 189.

Constructive interference should occur when the positive peak of the pressure waveform coincides with glottal opening. In this case, the positive pressure adds to the driving force of the mean subglottal pressure. If, in addition, the pressure is negative when the vocal folds are returning to the midline, then optimum reinforcement should occur; the pressure in the glottis provides a force that is applied in the direction of the velocity of the vocal folds (outward and inward). The optimum relationship should occur when $F_0 = 3/5F_1'$, or about 306 Hz, as shown in the middle of the figure.

Cases of destructive interference can be seen when $F_0 = F_1'$, or when $F_0=2/5F_1'$; the acoustic pressure in these cases is negative when the vocal folds are moving apart and positive when they are moving together. Several other interference relationships at higher and lower F_0 are also presented in Figure 1, but these are deemed to be of lesser importance.

At the optimum constructive interference condition ($F_0 = 3/5F_1'$=306 Hz), it was predicted that the amplitude of vibration could increase by as much as 0.8 mm. This is a significant change when compared to a typical amplitude of vibration of 1.0 - 2.0 mm[14].

To follow up on this prediction, an experiment was undertaken in this study to investigate the interaction between subglottal acoustic pressure and vocal fold vibration. Excised larynges were used instead of human subjects to allow for quantification of vibrational amplitude.

The previous analytical predictions were based on relationships between a constant subglottal formant frequency and a broad range of F_0 in the human. Due to the limited frequency range of the excised larynx, it was not possible to recreate this same condition in the excised larynx. Therefore, the experiment was designed so that the F_0 of the excised larynx would remain constant and the subglottal resonances could be tuned through a range of frequencies.

Experimental Apparatus

A pseudo-subglottal system was built using copper tubing acquired from a musical instrument manufacturer. The pseudo-trachea consisted of straight and curved copper tubing similar to the slide mechanism on a musical brass instrument (Figure 2). Formant frequencies were changed by sliding the telescoping lengths of tubing in or out of one another. Six different sets of slides of various length were necessary to provide the range of resonance frequencies required. This tuning apparatus was integrated into a previously described excised larynx bench set-up[15].

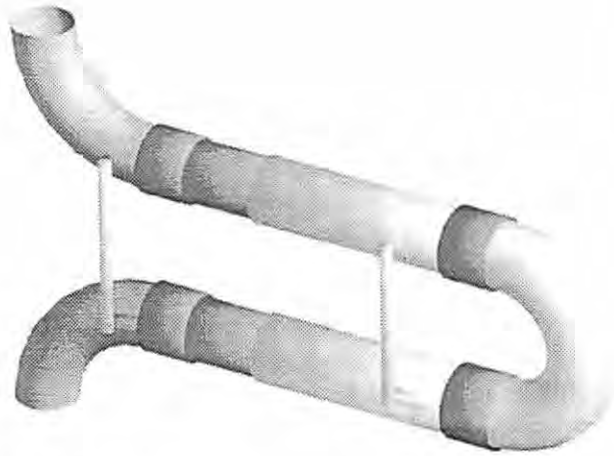


Figure 2. Pseudo-trachea consisting of a series of close tolerance copper tubes and elbows that change the tube length by sliding in and out.

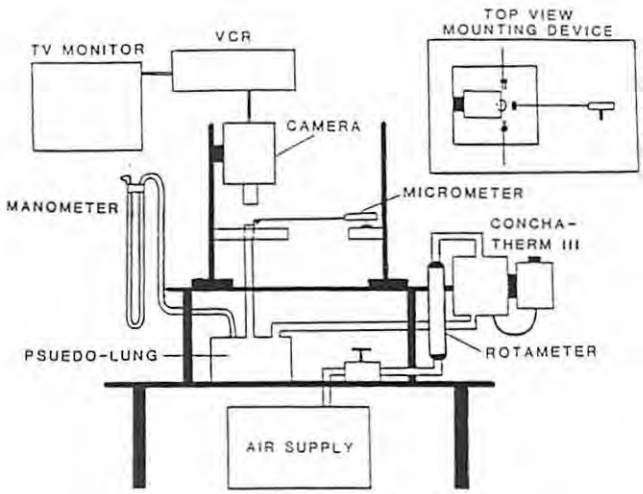


Figure 3. Illustration of the excised larynx experimental setup.

Procedures

Larynges

Excised canine larynges were procured from the cardio-vascular research laboratory located in the Medical Research Laboratories at the University of Iowa. The dogs were typically of mixed breed and included both male and female animals weighing between 25 and 30 kg. Each animal had been used in cardio-vascular research projects prior to the harvesting of the larynx. No animals were euthanized for the purposes of this experiment. The larynges were excised within a few minutes post mortem, placed in a 0.9% saline solution and chilled for at least 24 hours before being used in the experiment. The medial surfaces of the vocal folds were carefully inspected for abnormalities. Preparation of the tissue for the excised experiment is fully described by Durham[15], but a brief review of procedures will follow.

Tissue Mounting and Data Collection

The larynx was mounted by placing the tracheal tissue over the end of the pseudo-trachea and securing it with a clamp. The larynx was stabilized with a customized mounting apparatus consisting of an aluminum framework with needle prongs attached to micrometers (Figure 3). The needles attached to the micrometers were inserted into the muscular processes of the arytenoid cartilages and/or associated tissues, and were used to provide the adductory force which closed the glottis.

The length of the vocal folds was fixed with a suture at the thyroid angle to a secure attachment on the experimental apparatus. Optimum length of the vocal folds was subjectively determined for each larynx by judging the quality and stability of phonation. Optimum length was

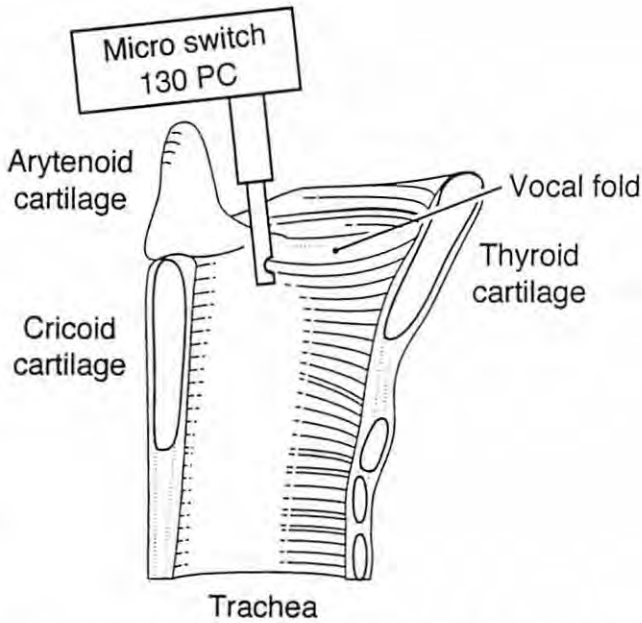


Figure 4. Subglottal section of a larynx showing the placement of the sensing tube of the pressure transducer between the arytenoid cartilages in the posterior larynx.

usually within 1 or 2 mm of the rest length (defined to be the in situ length of the vocal folds in the excised state). Length remained constant during the course of the experiment. Electrodes from a Synchrovoice Inc. electroglottograph (EGG) were attached to the thyroid lamina and held in place with small pins. The ground electrode was attached to the cricoid lamina.

A Micro Switch gauge type pressure transducer (130PC) was used to record the pressure (P_{sg}) underneath the

vocal folds. The housing of the transducer was fitted with a small extension tube for accessing the subglottal space. The transducer was mounted above the larynx, directly superior to the arytenoid cartilages (Figure 4). The extension tube was placed between the arytenoid cartilages with the open end extending below the level of the vocal folds. The distal end of the tube was sealed and a window was opened on the side of the tube for monitoring pressure[16]. The tissue was sutured closed around the tube to prevent air leakage.

Video images of the vocal folds were acquired with a Sony CCD video camera (model DXC-102) using a 90 mm Rokunar lens for measurement of glottal amplitude. The camera was mounted approximately 20 cm above the larynx and focused on the superior plane of the vocal folds.

The EGG signal was processed by a special phase delay circuit that was used to trigger a Pioneer DS330-ST stroboscopic light source[17]. The phase of the strobe was controlled manually to 'fix' the visual image at three specific points of the vibratory cycle: (1) maximum excursion from the midline of the glottis (henceforth referred to simply as *maximum*), (2) moment of opening of the superior margins of the vocal folds (*opening*) and (3) moment of first contact of the inferior margins of the vocal folds (*closing*). The image was 'fixed' at the three parts of the cycle by observing the video image on a TV monitor. Calibration of the image for measuring glottal dimensions was achieved by recording 1 mm grids placed on the superior surface of the vocal folds prior to the experiment. Video images were recorded on a Panasonic AG 1960 pro-line SVHS video cassette recorder, which has accurate frame by frame capabilities for analysis purposes. Maximum glottal amplitude was measured directly off the screen of a Sony Trinitron video monitor screen. Maximum glottal amplitude measurements were made when the glottis was rectangular, i.e. when the lower and upper lip of the vocal folds were vertically lined up.

Each time the strobe light flashed, it produced a large voltage pulse which was significant enough to contaminate the signal from the Microswitch pressure transducer. The pulse of the strobe was used as an event marker in the P_{sg} signal, indicating the place to measure the pressure which was associated with the fixed visual image. Measurements were made on the pressure signal just prior to the superimposed pulse from the strobe at each of the three points of interest during the vibratory cycle (determined from the visual image).

All signals transduced from the larynx (except the video image) were recorded with a digital Sony Instrumentation recorder (PC-108M) at a bandwidth of 5kHz (10 kHz sampling frequency). All signals were time synchronous within a small percent of error with this device.

A constant-flow air supply was provided by an Ingersoll-Rand Type 30 air compressor. Levels of air flow were monitored electronically with the use of a Statham differential pressure transducer (PM5). The air was conditioned for temperature and humidity by two Concha-therm III devices (Respiratory Care Inc.) to approximately 37 - 40° C and greater than 95% humidity prior to entering the pseudo-tracheal tube system. Some cooling of the air took place in the tubing of the pseudo-trachea between the pseudo-lung exit and the vocal folds.

After the larynx was mounted and all transducers attached, a suitable subglottal pressure for stable phonation was determined and utilized throughout the experiment. The full length range of the pseudo-trachea tubes (in 1 cm increments) was used for each larynx in a randomized order. A U-tube manometer was employed during the experiment to monitor the mean subglottal pressure for each condition.

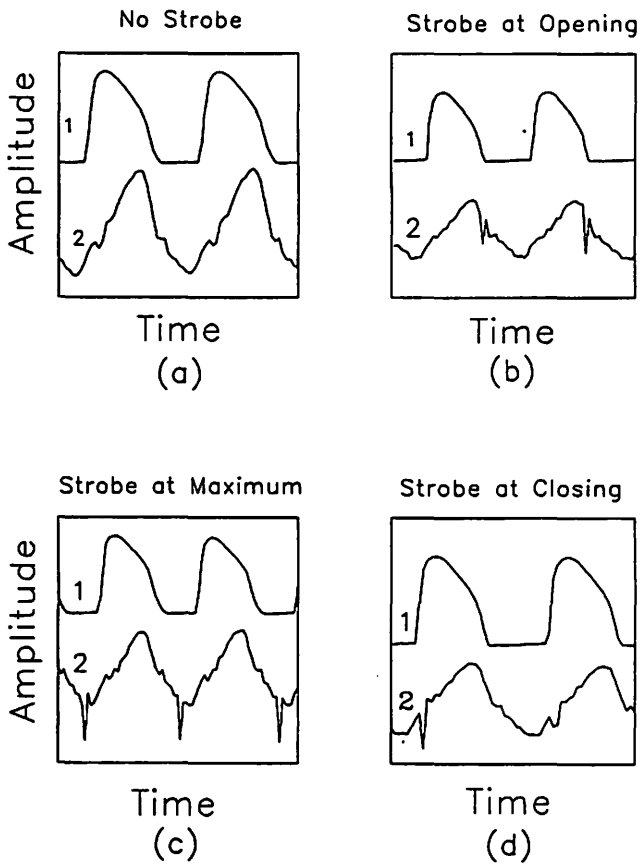


Figure 5. Points of measurement on the subglottal pressure waveform. Top waveform (1) is the EGG signal and (2) is the pressure waveform; (a) without the strobe pulse, (b) strobe pulse superimposed on the pressure waveform at the moment of glottal opening, (c) strobe pulse at moment of maximum glottal width and (d) strobe pulse at the moment of glottal closing.

Phonation was initiated and all signals were recorded for each trial. The strobe was turned on and the video image used to tune the trigger/delay circuit to each of the three moments of the vibratory cycle. Several seconds of all data signals were recorded at each of the three vibratory moments. The next length condition was set and the process repeated until the full range of the pseudo-trachea tubes had been used. The vocal folds were periodically bathed in a 0.9% saline solution to prevent desiccation of the tissue.

Results

Figure 5 shows typical EGG and P_{sg} waveforms from the experiment. An example with no strobe pulse is shown in part (a). The strobe pulse can be seen as the negative-going spike in the other three graphs. In part (b) the spike occurs at the moment of opening (as determined by the visual image), in part (c) during the maximum amplitude of vibration and in part (d) at the moment of closure of the inferior margin of the vocal folds. The point on the P_{sg} waveform immediately before the spike occurred was taken to represent the pressure at that point in the vibratory cycle. By comparing the shape of the waveform with and without the strobe pulse, it was determined that the pulse introduced no significant distortion of the waveform other than at the point of occurrence.

Figure 6 presents waveforms for a case in which the length of the pseudo trachea was approximately 38 cm and F_0 was 181 Hz. The upper trace is the EGG signal and the lower trace is the P_{sg} signal. Coronal images of the vocal folds[18] are placed along the waveform to describe vocal fold movement along with subglottal pressure and contact area. Four events were determined by comparing pressure waveforms with and without the strobe pulse and inserting the sketches in the approximate positions. Referring to Titze's theoretical relationships between F_0 and F_1' in Figure 1, this case represents nearly destructive interference between subglottal acoustic pressure and vocal fold movement. Specifically, subglottal pressure drops in the closed glottis phase and rises in the open glottis phase, similar to what is seen for $F_0 = F_1'$ in Figure 1. Pressure is lowest just prior to glottal opening. This retards vocal fold movement at a moment when a strong positive subglottal pressure is needed to blow the vocal folds apart at the bottom.

To the contrary, the waveforms in Figure 7 show a strong positive pressure at glottal opening. This acoustic

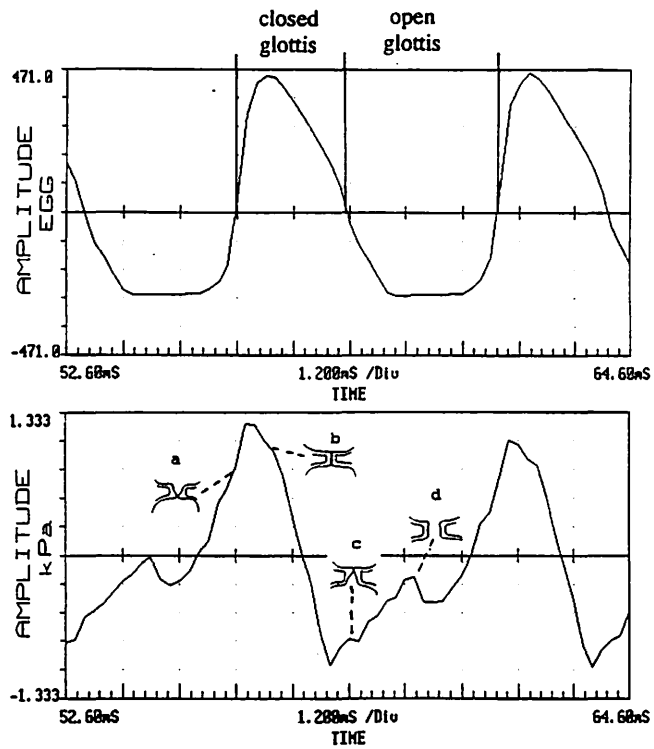


Figure 6. Waveforms from electroglottograph (top) and subglottal pressure transducer (bottom) with schematic drawings of coronal sections of the vocal folds at four points in the vibratory cycle. This case represents low interaction between subglottal acoustic pressure and vocal fold movement.

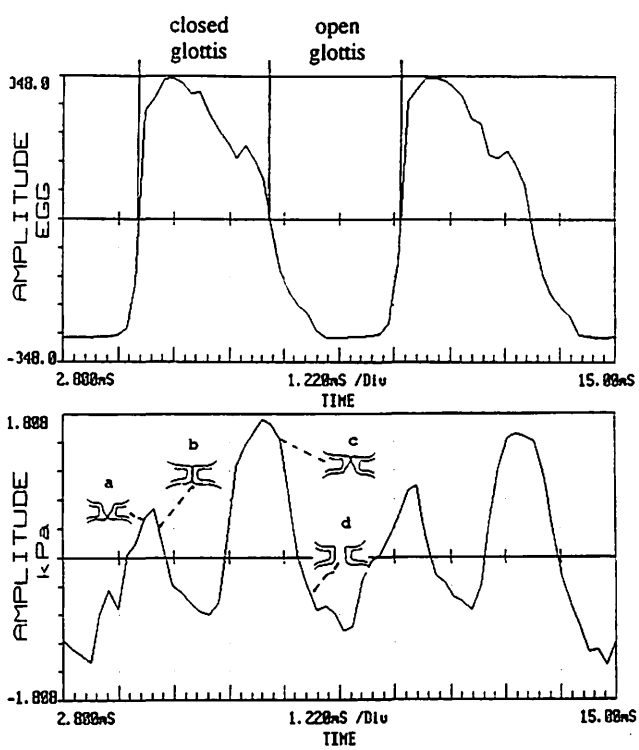


Figure 7. Waveforms from electroglottograph (top) and subglottal pressure transducer (bottom) with schematic drawings of coronal sections of the vocal folds at four points in the vibratory cycle. This case represents high interaction between subglottal acoustic pressure and vocal fold movement.

pressure helps to drive the vocal folds apart in the opening phase. The waveforms are from the same larynx, with the pseudo-trachea adjusted to a length of approximately 80 cm, with the F_o at approximately 175 Hz. Presumably, the waveforms in Figure 6 and Figure 7 differ only as a result of the change of length of the pseudo-trachea (from 38 cm to 80 cm). A constructive relationship now exists between subglottal pressure and vocal fold movement, which is similar to the $F_o = 3/5F_1'$ case depicted in Figure 1.

A quantitative relation was sought between vibrational amplitude and some effective acoustic driving pressure on the vocal folds. Since vibrational amplitude is dependent upon the mean subglottal pressure as well as the acoustic pressure, the relevant amplitude will be reported as a ratio of the maximum halfwidth of the glottis normalized to the mean subglottal pressure. This will be called the *glottal amplitude ratio* R_a defined as

$$R_a = \frac{\text{1/2 maximum glottal width (at mid-membranous vocal fold)}}{\text{mean subglottal pressure}}$$

Defining an effective acoustic driving pressure is less straightforward because subglottal pressure does not drive

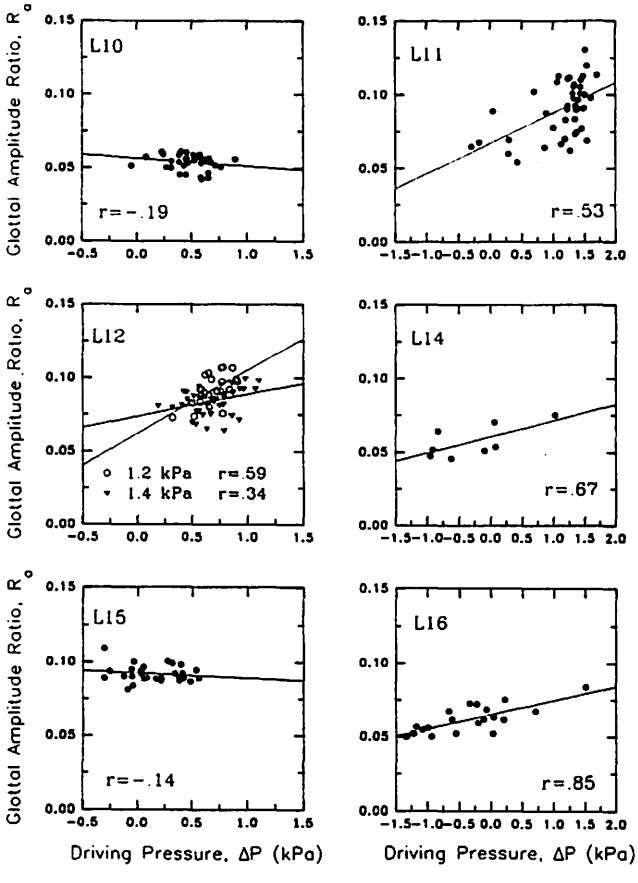


Figure 8. Glottal amplitude ratio R_a as a function of the acoustic pressure p measured below the vocal folds for six larynges.

the vocal folds over the entire glottal cycle. In the open part of the cycle, the net driving pressure on the medial surface of the vocal fold can be approximated as:

$$P = P_{in}\frac{a_2}{a_1} + P_{sg}\left(1-\frac{a_2}{a_1}\right)$$

where P_{in} is the input pressure to the vocal tract (supraglottal pressure), P_{sg} is the subglottal pressure, a_1 is the duct area at glottal entry, and a_2 is the duct area at glottal exit[12]. For an excised larynx phonating into free space, $P_{in} \approx 0$ and the driving pressure is the subglottal pressure modified by the factor $1 - a_2/a_1$. When $a_1 \gg a_2$ (a highly convergent glottis), the full subglottal pressure is applied to the vocal folds. This occurs just after glottal opening (see the sketches at point c in Figures 6 and 7). For a rectangular glottis, however, the driving pressure is basically zero because $a_2/a_1 = 1$. This occurs at the maximum glottal width (point d in Figures 6 and 7). For a divergent glottis, the driving pressure is also near zero because the flow detaches from the wall and the glottal airstream is in the form of a jet. Thus, we conclude that the subglottal pressure has its main driving effect in the time interval between glottal opening and maximum glottal width. We therefore define

$$\Delta P = P_{sg}(\textit{at glottal opening}) - P_{sg}(\textit{at maximum glottal width})$$

to be the relevant acoustic pressure change that should influence the amplitude of vibration. Again, in Figures 6 and 7, this pressure change occurs between points c and d. It is small and negative in Figure 6, but large and positive in Figure 7.

Scatter plots for data gathered on six larynxes are presented in Figure 8 by plotting the glottal amplitude ratio R_a for each condition as a function of driving pressure ΔP. Each data point represents an averaged value of at least 20 measurements taken from individual cycles. The lines through the data indicate the best fit first order linear regression. The regression coefficient is indicated in each plot. For the three larynges on the left side (L10, L12, and L15), the data indicate no strong relationship between acoustic driving pressure and the amplitude ratio. Correlation coefficients are generally low except for the case in L12 where subglottal pressure was 1.2 kPa. For the three larynges on the right side (L11, L14, and L16), a positive relation is seen between the amplitude ratio and the acoustic driving pressure P. The amplitude ratio changes by a factor of 2 (100%) in L11, for example.

The two larynges for which subglottal acoustic pressure affects vibrational amplitude the least (L10) and the most (L16) are examined further in Figures 9 and 10. Acoustic pressure is plotted for the three major events in the glottal cycle: glottal opening, maximum glottal width, and glottal closing. Different data points are for different tube lengths. Note the relatively constant pressure near zero at

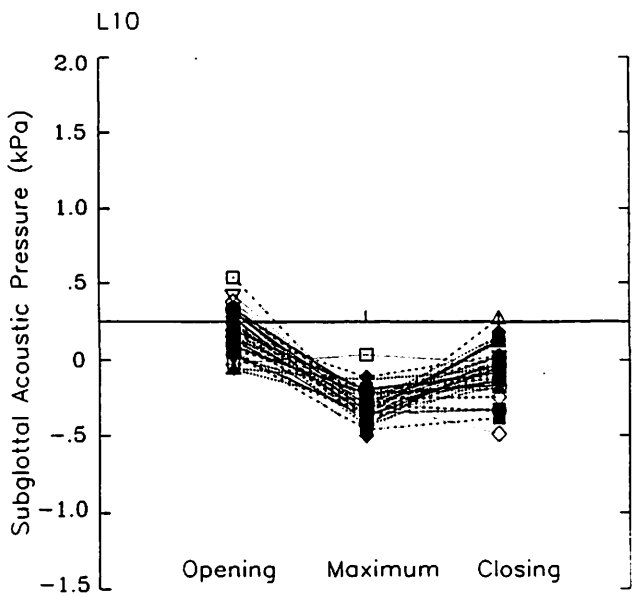


Figure 9. Subglottal acoustic pressure at three moments in the glottal cycle for a case of low interaction a) just as the superior margins of the folds separate, b) when the folds are at their maximum excursion and c) as the inferior margins of the vocal folds come together at closure.

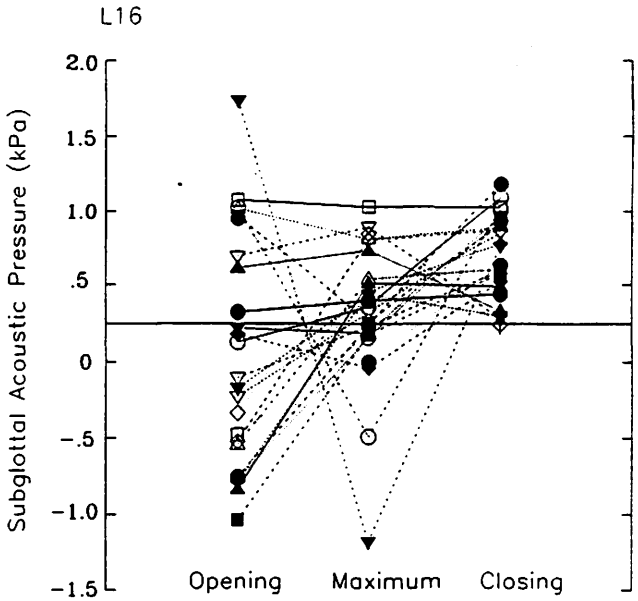


Figure 10. Subglottal acoustic pressure at three moments in the glottal cycle for a case of high interaction a) just as the superior margins of the folds separate, b) when the folds are at their maximum excursion and c) as the inferior margins of the vocal folds come together at closure.

opening for L10 in Figure 9. On the contrary, note the wide fluctuations in pressure at opening for L16 in Figure 10. There are positive acoustic pressures as high as 1.8 kPa and negative acoustic pressure as low as -1.1 kPa in Figure 10. Furthermore, in several cases the pressures changes direction in the short interval from glottal opening to maximum glottal width. These cases were always the ones for which the amplitude ratio changed the most. A sudden decrease in pressure (large positive ΔP) caused a large amplitude increase and a sudden increase in pressure (large negative ΔP) caused a large amplitude decrease, as was previously discussed in connection with Figures 6 and 7.

Discussion

The data presented here indicate that the subglottal pressure in excised larynges contains not only a mean (dc) component, but also a high frequency (ac) component that is the result of resonance in the subglottal airway. The resonance frequencies were at least in part under experimental control and primarily dependent upon the length of the pseudo-trachea.

The magnitude of the peak to peak pressures in these data are similar to those reported elsewhere. Peak to peak values in the literature vary from 40 to 50% of the mean value[19] to as much as 100% of the mean pressure[20,21]. In the present study, the pressure at a specific moment in the vibratory cycle did vary over a range equal to or greater than the mean subglottal pressure.

As is common with excised larynges, there was considerable variability in the data. Conditions for stable oscillation were established according to the response of the individual larynx; no attempt was made to control the adductory force, longitudinal tension, or any other dimension of the vocal folds in any uniform manner across all larynges. We suspect that the larynges for which the amplitude of vibration was least affected by subglottal resonance had a different prephonatory glottal shape.

The data presented support the theory presented by Titze[12] that modulating pressure in the subglottal vocal tract can have an influence upon amplitude of vibration of the vocal folds. When the vocal folds were exposed to high pressures at the moment that the vocal folds were opening, an increase in the glottal amplitude ratio occurred. In those larynges in which no significant amplitude changes were found, the pressures measured beneath the vocal folds were not of sufficient magnitude to contribute to the driving pressure of the vocal folds.

Application to Human Data

At this point we may ask, do these data on excised canine larynges apply in any way to human phonation? Figure 11 shows data from Miller and Schutte[22] indicating subglottal pressure variations (labeled P_{sub}) in a human subject. The pressure signals were recorded in the vocal tract with miniature pressure transducers. The EGG signal was used to mark closing and opening of the glottis (vertical

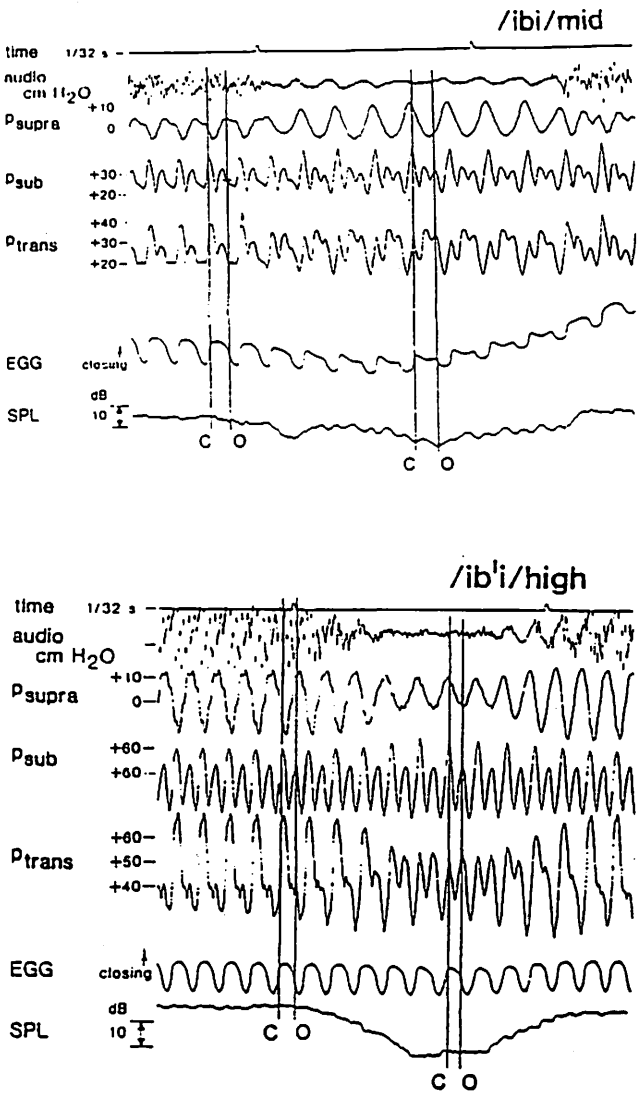


Figure 11 (top). Pressure signals recorded at several points along the vocal tract, including the subglottal space. EGG signal is used to mark moment of opening and of closure. F_0 is around 215 Hz. (From: Miller, D.G. and Schutte, H.K. (1991). Effects of downstream occlusions on pressures near the glottis in singing. In Gauffin J, Hammarberg B. eds. Vocal Fold Physiology (pp. 91-98). San Diego: Singular. Used with permission. Figure 12 (bottom). Pressure signals recorded at several points along the vocal tract, including the subglottal space. EGG signal is used to mark moment of opening and of closure. F_0 is around 307 Hz. From: Miller, D.G. and Schutte, H.K. (1991). Effects of downstream occlusions on pressures near the glottis in singing. In J. Gauffin and B. Hammarberg (Eds.), Vocal Fold Physiology (pp. 91-98). San Diego: Singular. Used with permission.

lines). The waveforms show the pressure events associated with the subject singing /ibi/ at an F_0 of 225 Hz. The marked segment on the left is prior to the plosive consonant and during the initial vowel. A pressure maxima in the P_{sub} waveform occurs just after glottal closure, and is followed by several negative and positive pressure peaks, each of decreasing amplitude (due to damping). The point marked as the opening moment ("O") occurs coincidentally with the third positive pressure peak and indicates a low pressure on the P_{sub} waveform, which the authors identify as 2.6 kPa. The mean value for subglottal pressure was not given for these data, but it seems from the overall shape of the waveforms that the pressure at this point is approximately the same as the mean pressure in magnitude. This frequency (225 Hz) in the male singing voice does not represent one of the major involuntary register transitions. No interference, either constructive or destructive, between P_{sub} and vocal fold vibration would be anticipated here.

Figure 12 shows data from the same subject with the F_0 at 307 Hz. Glottal opening is now coincident with the secondary pressure peak following the maximum at closure. The pressure is reported as 5.3 kPa, significantly higher than the pressure at opening in Figure 11. The F_0 is approaching the range at which an involuntary register adjustment occurs (especially in the untrained voice). If we assume that the subglottal formant frequency was 510 Hz[13], the ratio between the F_0 (307 Hz) and the subglottal formant is approximately .6 (F_0=3/5F_1); this is the ratio that Titze describes as the condition for optimum positive interference.

Conclusion

Interactions between acoustic pressures in the subglottal airway and amplitude of vibration of the vocal folds have been quantified with the use of excised larynges. Maximum changes in amplitude of vibration (on the order of 50 - 100%) agree with theoretical predictions made earlier[12]. The potential for similar interactions in human subjects exists on the basis of recordings made by other investigators.

For many centuries, singers have been dealing with these interactions as register transitions. Various laryngeal and articulatory maneuvers have been taught that allow singers to modify and utilize these interactions to the benefit of vocal style and vocal control. The better these interactions are understood, the better voice trainers will be able to lead students through the somewhat perilous journey of voice registration.

Acknowledgement

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

References

1. Kitzing P, Carlborg B, Löfqvist A. Aerodynamic and glottographic studies of the laryngeal vibratory cycle. *Folia phoniat* 1982;34:216-224.

2. Koike Y. Sub and supraglottal pressure variations during phonation. In: Stevens K, Hirano M, eds. *Vocal Fold Physiology*. San Diego: College Hill, 1981:181-191.

3. Miller DG, Schutte HK. Characteristic patterns of sub and supraglottal pressure variations within the glottal cycle. *Transcripts of the XIIth Symposium: Care of the Professional Voice*. New York: The Voice Foundation, 1984.

4. Large J. Towards and integrated physiologic-acoustic theory of vocal registers. *The NATS Bulletin* 1972;29:18-25;30-36.

5. Nadoleczny-Millioud M, Zimmerman R. Categories et registres de la voix. *Rev. fr. phoniat* 1938;23:21-31.

6. Berg JW van den. Register Problems. *Ann. N.Y. Acad. Science* 1968;155(1):129-135.

7. Vennard W. *Singing...the Mechanism and the Technique*. Revised Ed. New York: Carl Fischer, Inc. 1967.

8. Berg JW van den. An electrical analogue of the trachea, lungs and tissues. *Acta Physiol. Pharmacol. Neerlandica* 1960;9:361-385.

9. Ishizaka K, Matsudaira M, Kaneko T. Input acoustic-impedance measurement of the subglottal system. *J. Acoust. Soc. Am.* 1976;69(1):190-197.

10. Boves L. *The Phonetic Basis of Perceptual Ratings of Running Speech*. Cinnaminson N.J.: Foris Publications, 1984.

11. Cranen B, Boves L. Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production. *J. Acoust. Soc. Am.* 1985;77(4):1543-1551.

12. Titze IR. A framework for the study of vocal registers. *J. Voice* 1988;2:183-194.

13. Cranen B, Boves L. On subglottal formant analysis. *J. Acoust. Soc. Am.* 1987;81(3):734-746.

14. Titze IR. On the relation between subglottal pressure and fundamental frequency in phonation. *J. Acoust. Soc. Am.* 1989;85(2):901-906.

15. Durham PL, Scherer RC, Druker DG, Titze IR. Development of excised larynx procedures for studying mechanisms of phonation. *Technical Report No. VABL-1*. Voice Acoustics & Biomechanics Lab, U. of Iowa 1987.

16. Baken RJ. *Clinical Measurement of Speech and Voice*. Boston: College-Hill, 1987.

17. Jiang J. A methodological study of hemi-laryngeal phonation and the measurement of vocal cord intraglottal pressure and impact stress. Unpublished doctoral dissertation, 1991; University of Iowa.

18. Hirano M, Kurita S, Nakashima T. The structure of the vocal folds. In: Stevens KN, Hirano M, eds. Vocal Fold Physiology. Tokyo: University of Tokyo Press, 1981:33-43.

19. Garret JD, Luschei ES. Subglottic pressure modulation during evoked phonation in the anesthetized cat. In: Baer T, Sasake K, Harris K, eds. Laryngeal Function in Phonation and Respiration. Boston: College Hill, 1987:130-153.

20. Baer T. Investigation of phonation using excised larynxes. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1975.

21. Koike Y. Sub and supraglottal pressure variations during phonation. In: Stevens K, Hirano M, eds. Vocal Fold Physiology. San Diego: College Hill, 1981:181-191.

22. Miller DG., Schutte HK. Effects of downstream occlusions on pressures near the glottis in singing. In: Gauffin J, Hammarberg B, eds. Vocal Fold Physiology. San Diego: Singular, 1991:91-98.

NCVS Status and Progress Report - 6
May, 1994, 43-52

Direct Training of Velopharyngeal Musculature

Julie M. Liss, Ph.D.
Department of Communication Disorders, The University of Minnesota
David P. Kuehn, Ph.D.
Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign
Kathy P. Hinkle, B.A.
Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign

Abstract

Behavioral techniques for the remediation of velopharyngeal insufficiency appear to be widely used in the clinical domain, but their efficacy has not been demonstrated empirically. This paper provides a physiologic rationale for the use of palatal training procedures by discussing literature on strength training and neural adaptation, and the mechanisms for sensorimotor control in the velopharyngeal mechanism. A direct palatal training method that has shown clinical promise, CPAP therapy, is described.

There has been little empirical support for the use of direct or indirect palatal training procedures to reduce hypernasality (Peterson-Falzone, 1984; Ruscello, 1982, 1989). A systematic literature review conducted by Ruscello (1989) led him to conclude that, "palate training methods remain within the realm of experimental study" (p. 347). Pannbacker (1992), in an effort to identify and correct common clinical "myths," stated that, "palatal and pharyngeal muscle training are of no value for improving velopharyngeal closure" (p. 13). However, nearly 64% of the respondents in a survey (Pannbacker, Lass, and Stout, 1990) reported the occasional use of speech therapy (indirect palatal training) to address velopharyngeal deficits. Nearly 24% of the respondents reported its frequent use. This frequency of usage may reflect, as Pannbacker (1992) suggests, the unwarranted potency of clinical history. Another view is that there are circumstances under which clinicians believe that palatal training procedures may be efficacious.

It would seem that collective clinical experience has generated certain assumptions about the sensorimotor control of the velopharyngeal mechanism despite the absence of corroborating empirical evidence. The first assumption is that behavioral methods can modify the degree and possibly the timing of velopharyngeal closure by increasing muscle strength and endurance. A second assumption is that palatal training procedures increase the speaker's sense or awareness of velopharyngeal closure by pairing the acoustic/perceptual event with feedback from sensory receptors along the vocal tract, and that this awareness leads to enhanced sensorimotor control. These assumptions appear reasonable, particularly for patients who have no prohibitive structural anomalies. Moreover, they parallel those that might be applied to other impaired motor systems (e.g., Hartigan, Persing, Williamson, Morgan, Muir, and Edlich, 1989; Lohi, Lindberg, and Andersen, 1993; Milner-Brown, 1993).

The purpose of this paper is to present a physiologic rationale for the use of palatal training procedures by discussing literature on strength training and neural adaptation, and the mechanisms for sensorimotor control in the velopharyngeal mechanism. Finally, a direct palatal training method that has shown clinical promise is described.

Physiologic Rationale

Velopharyngeal insufficiency (VPI) is associated with a variety of etiologies (Kuehn, 1986). Typically, VPI resulting from structural anomalies, such as cleft palate, are considered separately from velopharyngeal deficits associated with nervous system lesions, such as stroke or traumatic brain injury. It can be proposed, however, that direct palatal training procedures are warranted whenever VPI is

thought to be the result of inefficient neuromuscular function, regardless of the etiology.

Motor Units

Structurally, striated muscle contractions result from the shortening of specific groups of muscle fibers within an individual muscle. The contraction of these motor fibers is determined by the functional unit of the striated muscle, the motor unit, which is composed of a nerve cell body, an axon, a motor nerve and the muscle fibers it supplies with its terminal branches. Fibers within individual motor units may be intermingled with fibers of other motor units. Therefore, a motor unit is not characterized by a single action potential, but by a multitude of action potentials at different sites in the motor unit (Basmajian and DeLuca, 1985).

Strength, Power, Endurance, Fatigue

The integrity of the contractile properties of muscle fibers can be measured in a number of ways. Muscle **strength** refers to the maximum force or tension generated by a muscle or muscle groups. Strength is not time dependent and refers only to the maximal performance during an episode of voluntary contraction. Muscle **power** also refers to the force or tension generated by a muscle or muscle group, but power is time dependent. That is, it refers to maximum performance during some time unit of voluntary contraction. Muscle **endurance** refers to the ability of muscle fibers to sustain some level of submaximal contraction over a given time period. A muscle exhibits **fatigue** if it does not maintain a given level of tension without increasing neural drive. Bigland-Ritchie and Woods (1984) defined **threshold for fatigue** as "that level of exercise which just cannot be sustained indefinitely" (in their example, greater than 45 minutes).

Muscle Fiber Types

Strength, power, and endurance have as their physical correlates muscle fiber-type composition and cross-sectional area of fibers. All mature striated muscles contain a mixture of muscle fiber types whose properties vary along several dimensions: contraction speed, fatigability, and metabolic characteristics. The composition of muscle fiber types in any muscle provides information about the function of the muscle. For example, muscles important in posture maintenance typically contain a preponderance of fibers that have a slow contraction speed, that are fatigue resistant, and that have an oxidative metabolism. These types of fibers are classified as "red" or Type I. Muscles that involve rapid burst movements typically contain fibers that are fast contracting and have anaerobic metabolism. These are called "white" or Type II fibers. Type II fibers are further classified into Type IIA, which are fast twitch but fatigue resistant, and Type IIB, which are fast twitch and fatigue sensitive (Brooke and Kaiser, 1973; Peter, Barnard, Edgerton, Gillespie, and Stempel, 1972). Another fiber, Type IIC, can be found in embryonic, degenerating, and regenerating muscle and has functional characteristics between Type I and Type II fibers.

In human fetal development, muscle fibers differentiate into types based upon the functional connection between the muscle and its motoneurons. Postnatal changes of metabolic activity, fiber-type distribution and cross-sectional area can be achieved through changes in muscle use (Cadefau, Casademont, Grau, Fernandez, Balaguer, Vernet, Cusso, and Urbano-Marquez, 1990; Takekura and Yoshioka, 1990). The precise mechanisms of these changes are currently being explored (Adams, Hather, Baldwin, and Dudley, 1993; Wang, Hikida, Staron, and Simoneau, 1993).

Neural Adaptation

Neural factors that play a role in strength performance are thought to be of particular importance in motor control. According to Sale (1986), "strength and power training may cause changes within the nervous system that allow an individual to better coordinate the activation of muscle groups..." (p. 289). Thus, the goal of strength training is not simply to make muscles stronger, but to make them more efficient for the targeted tasks. These changes within the nervous system that correspond to muscle use and function are referred to as **neural adaptation.**

The most direct evidence of neural adaptation has been afforded by measures of electromyographic activity in strength training studies. Using EMG, Fleck and Kraemer (1987) found that any level of submaximal force could be produced with less neural activity after a resistance training program, suggesting a more efficient pattern of motor unit recruitment. Other studies reported that increased EMG activity was correlated with improved voluntary strength (Sale, 1986; Sale, 1988). Increases in strength had been observed as rapidly as the first training session, a phenomenon that simple muscle adaptation fails to explain. However, the majority of EMG increase was found in the first three to four weeks of training, indicating that initial increases in strength likely are due to neural adaptation (Sale, 1986). Contralateral training studies also have reported both increased EMG activity and improved voluntary strength without changes in either the metabolic capacity of the muscle or increased fiber size (Housh and Housh, 1993; Houston, Froese, Valeriote, Green, and Ranney, 1983; Sale, 1986; Sale, 1988).

Although the mechanisms that underlie neural adaptation are not well understood, several possibilities have been explored. Specifically, neural adaptation may be the result of increased activation of prime movers, more efficient activation of synergists and antagonists, or some

combination of the two mechanisms (Sale, 1986; Sale, 1988). The first of these possibilities implies that full activation of the motor units is somehow inhibited or underutilized in normal circumstances. The idea that there is a typically untapped reserve of force production in muscle has been tested, and largely supported, by techniques that have examined muscle force and motor unit activity during maximal voluntary contraction (Grimby, Hannerz, and Hedman, 1981; Sale, 1986; Sale, McComas, MacDougall, and Upton, 1982). The proposal, then, is that neural adaptation improves voluntary strength by enabling the subject either to recruit these previously untapped motor units or to increase the firing rate of already activated units (Kraemer, Deschennes, and Fleck, 1988; Sale et al., 1982; Sale, 1986).

The second possible mechanism of neural adaptation involves changes in the motor unit activation patterns of synergists and antagonists, resulting in more efficient force production. In many tasks, especially those involving strong, rapid muscle contractions or precise movements, contraction of the antagonists accompany that of the agonist as a stabilizing, braking mechanism (Sale, 1988). However, this also reduces the effective force output of the agonist. Strength training and neural adaptation may reduce excessive contraction of the antagonists, thereby allowing uninhibited motor unit activation in the agonists.

Although co-contraction of antagonists inhibits the net force produced by the agonist, some studies attribute certain advantages to this method of force production. A precontraction of the antagonist muscle seems to augment motor unit activation in subsequent agonist contractions (Sale, 1986). In a similar manner, a brief decrease in the voluntary effort of motor units just prior to a maximal voluntary contraction allows subjects to maintain higher levels of force production for extended periods of time (DeLuca, LeFever, McCue, and Xenakis, 1982; Grimby et al., 1981). The mechanism by which individual muscles create and maintain this extra force is based on a highly ordered, muscle-specific recruitment pattern (Freund, 1983).

Specificity of Training

There is a preponderance of evidence to suggest that neural changes that accompany increments in muscle strength are task-specific. That is, each activity requires a specific pattern of motor unit recruitment, involving type of contraction, appropriate angle, and proper velocity during execution of a specific activity (Kraemer et al., 1988). The implication is that training techniques should exploit the causal link between task parameters and neural adaptation. Sale and MacDougall (1981) identified four kinds of specificity of training: movement pattern, velocity, contraction type and contraction force, and these are discussed in turn.

Specificity of movement pattern refers to the fact that increased voluntary strength will be greatest in the position that is trained. In a study conducted in 1957 by Rasch and Morehouse, subjects trained elbow flexion in the standing position. Results indicated a considerable increase in strength in this position, but only a slight increase if the subject performed the task in a supine position. Since then, studies have shown responses to training specificity even to the trained joint angle (Sale, 1986; Sale and MacDougall, 1981). This specificity of movement pattern found even in simple movements likely is due to the role of neural adaptation in strength training. Because an increase in voluntary muscle strength has been exhibited, the specificity of task performance may be caused by certain groups of motor units being preferentially recruited for particular tasks (Sale 1988; Tanaka, Costille, Thomas, Fink, and Widrick, 1993; Tax, van der Gol, Geilen, and Kleyne, 1990).

The second kind of specificity of training mentioned by Sale and MacDougall (1981) is specificity of velocity. As early as 1970, studies were indicating that training at low velocities increased voluntary strength at low velocities but not at high. In the same way, high velocity training caused greater improvement at high velocities than at low. Other studies, while confirming these results, also indicate that high velocity training has a better transfer effect to low velocities than vice versa (Coyle and Feiring, 1980). Therefore, the results of a high velocity training program are more effective when later tested on both high and low velocity tasks. This specificity of velocity can most aptly be explained by the mechanism of neural adaptation. For one, there is evidence that the brain stores and induces fast movements differently than slow movements. In addition, strength training has been shown to cause hypertrophy in both slow and fast twitch muscle fibers. Therefore, specificity of training in regards to velocity is due likely to movement organization by the brain, not selective recruitment of particular motor unit types (Sale and MacDougall, 1981).

The third type of specificity of training concerns the type of contraction used in the training program. Tension can be developed in a muscle while the muscle is shortening (concentric contraction), while the muscle is lengthening (eccentric contraction), or when there is no change in the length of the muscle (isometric contraction). Although the motor unit activation is similar in all three types of contractions, utilizing the same form in training that occurs during performance allows the appropriate neural adaptation to occur (Sale and MacDougall, 1981). The final type of specificity of training involves contraction force. If the task to be performed involves a limited number of maximal contractions, then training should imitate this by centering on only a few repetitions of nearly maximal voluntary contractions. Similarly, training for tasks applying numerous submaximal contractions should focus on several repetitions of only 75%-80% maximal voluntary

contraction. As Sale and MacDougall (1981) concluded, training exercises should "simulate the sport movement as closely as possible, in relation to movement pattern, velocity of movement, muscular contraction type, and contraction force" (p. 91).

According to DiNubile (1991), muscle training can induce the recruitment of more muscle fibers, "with resultant increased neural activation including selective recruitment and modulation of the fast- and slow-twitch motor units as well as differential control of the motor unit firing pattern" (p. 45). The implication is that well-chosen training activities can induce changes not only in aspects of muscle force generation, but also in the coordination of activation among muscle groups. These types of changes are expected to be beneficial for speakers who exhibit VPI secondary to inefficient muscle function because the velopharyngeal mechanism is one that requires certain levels of coordinated muscular force generation to attain the complex closure patterns in speech production. In the following section, sensory and motor components of the velopharynx will be addressed.

Sensorimotor Control of Velopharyngeal Closure

It is known that aperture variations of the velopharyngeal mechanism for the production of speech are achieved by synergistic activity of a number of muscles (Kuehn, Folkins, and Cutting, 1982; Moon, Smith, and Folkins, 1993; Seaver and Kuehn, 1980). Strength training of the velopharyngeal musculature possibly could be used to modify the degree and timing of velopharyngeal closure, and to increase or develop the speaker's awareness of velopharyngeal closure (Moon and Jones, 1991). Strength training might enhance (or, in some cases, re-establish) sensorimotor control of the velopharyngeal mechanism. To consider the possible effects of palatal training procedures in more detail, it is first necessary to describe what is known about the sensory and motor components of the velopharyngeal region.

Sensory Components

Mucosa of the hard and soft palates and the pharynx contain sensory fibers from the vagus nerve (CN X). The soft palate also receives fibers from the glossopharyngeal nerve (CN IX). Reports of the presence of tastebuds on portions of the soft palate suggest that fibers of the facial nerve (CN VII) are also available for mediation of this special sense (Holland, 1984).

The types of sensory terminations located in the human oral mucosa, and their potential functions, are matters of some controversy. Free nerve endings, Merkel's discs, Krause end bulbs, and Meissner's corpuscles have been identified in the mucosa of the hard palate and other oral structures (Botezat, 1907; Dixon, 1961; Gairns, 1955; Gairns and Aitchison, 1950; Kadanoff, 1928). Only free nerve endings have been identified in the mucosa of the soft palate and uvula, and these endings appear to be somewhat less complex than those found in more anterior sites (Gairns, 1955). In general, a greater number and variety of sensory receptors have been found in anterior as compared with the posterior oral cavity (Grossman and Hattis, 1964; Kanagasuntherum, Wong, and Chan, 1969). This corresponds with behavioral and psychophysical investigations that report greater sensitivity for a variety of stimuli in the anterior as compared to the posterior oral cavity (Barlow, 1987; Dubner, Sessle and Story, 1978; Eccles and Tolley, 1987; Grossman, Hattis and Ringel 1965; Ringel and Ewanowski, 1965).

Sensory receptors located in **muscle** that contribute to proprioceptive (position) and kinesthetic (movement) sensations include Golgi tendon organs, free nerve endings, and muscle spindles. Muscle spindles in particular are thought to play a significant role in the control of some types of movement. Muscle spindles are neuromuscular structures that lie in parallel with muscle fibers. They are "neuromuscular" because they contain afferent and efferent nerve fibers as well as contractile fibers. Because they lie in parallel with the surrounding muscle fibers, they are subject to stretch when the surrounding muscle fibers are stretched, and relaxation or compression when the muscle fibers contract. The presence of muscle spindles in a muscle is an indication that, under the proper circumstances, proprioceptive and kinesthetic information related to muscle movement, particularly muscle stretch, is available to the central nervous system. At least two muscles involved in velopharyngeal movement for speech, the levator veli palatini and the palatoglossus, contain muscle spindles (Kuehn, Templeton and Maynard, 1990; Liss, 1990). The potential contribution of muscle spindles to velopharyngeal control has been described by Liss (1990, pp. 743-744).

Motor Components

Many muscles interact to achieve velopharyngeal port modulation for speech production and swallowing (Bell-Berti, 1976; Fritzell, 1969; Kuehn, 1979; Kuehn et al., 1982; Moon et al., 1993). The levator veli palatini (LVP) muscle serves as the primary elevator of the velar mass. Its contraction results in superoposterior movement of the soft palate. Cranial nerve innervation of the LVP has not been conclusively determined. Evidence suggests that the vagus and glossopharyngeal (CN IX, and X) are responsible for most LVP activity, with a potential contribution from fibers of the facial nerve (CN VII) (Broomhead, 1951; Furusawa, Yamaoka, Kogo and Matsuya, 1991; Ibuki, Tamaki, Matsuya and Miyazaki, 1981; Keller, Saunders, van Lovern, and Shipley, 1984; Nishio, Matsuya, Machida and Miyazaki, 1976; Sedlakova, 1967).

Fiber type composition of the normal human LVP is virtually unknown. One Japanese study (Yamaguchi and Takumida, 1989) reported data from histochemical fiber-typing of biopsied LVP muscles of six VPI patients (two cleft palate, two submucosal cleft, two congenital VPI). They found fewer Type I fibers in the cleft side LVP's (in the overt and submucosal cleft subjects) than in the noncleft side muscle tissue. Type I fiber diameters in congenital VPI were smaller than those in the noncleft side LVP's of the overt and submucosal cleft palate subjects. These investigators speculated that the results reflected lack of appropriate LVP activity in these populations. Obtaining such information about non-VPI speakers will be useful for understanding the potential role of muscle spindles in LVP as well as the relationship between biomechanical function and metabolism.

In summary, a variety of mucosa and muscle sensations are transduced by cranial nerves and peripheral sensory structures located in the velopharynx. Likewise, cranial nerve innervation subserves contraction of the muscles of the velopharyngeal complex. Velopharyngeal muscle inefficiency, then, is the result of deficits among these peripheral efferent and afferent components, possibly in combination with inappropriate central nervous system drive. Barring anatomic barriers, task-specific strength training should serve to modify not only the contractile properties of the muscles proper, but also improve the time-dependent synergies that rely on the pairing of movement and sensory feedback. A technique to accomplish this goal is described in the following section.

Continuous Positive Airway Pressure (CPAP) Therapy

Resistance Exercises

Inadequate velopharyngeal closure is generally regarded as the primary physical manifestation underlying hypernasal speech. Inadequate closure may involve a spatial problem, temporal problem, or both. Examples of a spatial problem include insufficient velar elevation or insufficient inward displacement of the pharyngeal walls. Examples of a temporal problem include closure that is not maintained long enough or closure that occurs at inappropriate times. Increasing strength of the muscles of velopharyngeal closure might reduce these problems, especially if the problem is spatial. It may reduce a temporal problem as well, particularly if velar velocity is affected.

It is well known that resistance exercises increase muscle strength (Atha, 1981; DiNubile, 1991; Komi, 1986). Therefore, it is logical that appropriate resistance exercises might increase strength of the velopharyngeal closure muscles as well. Although the concept of increasing velopharyngeal muscle strength to reduce hypernasality is not new, achieving that goal has been difficult. Non-speech exercises such as blowing have been used for many years with no documented success. Although such tasks may in fact cause the muscles of velopharyngeal closure to increase level of activity (Kuehn and Moon, submitted) and, by inference, to work "harder," there may be little or no generalization of that increased effort to speech. This lack of generalization might relate to the specificity of training phenomenon discussed earlier in this report. That is, it is possible that during speech, different neuromuscular control mechanisms are used compared to that for nonspeech tasks such as blowing. Thus even if muscle strength is increased in relation to blowing tasks, that increased strength may not be carried over for speech tasks.

In addition to the specificity of training issue, a problem of access exists. Unlike that for limb muscles and corresponding appendages, it would be difficult logistically to attach a freestanding solid weight to the velum especially for routine and repeated clinical purposes. Kuehn (1991) introduced a new approach for resistance training that loads the velopharyngeal closure mechanism not with a solid mass but with a column of air. Using this approach, a mask is placed over the subject's nose and elevated air pressure is introduced using a commercially available device commonly called continuous positive airway pressure or CPAP. The trade name for one such device is REMstar, available from Respironics, Inc., Murrysville, PA. CPAP machines have been used successfully to treat patients with obstructive sleep apnea by keeping the upper airway patent during sleep (Schmidt-Nowara, 1984). CPAP machines are available in Sleep Labs at most major hospitals.

Therapy Procedure

A brief description of the CPAP therapy procedure will be given here. Additional information is provided elsewhere (Kuehn, 1991). The procedure being used is still experimental and a clinical trials project is underway that involves patients from seven different cleft lip and palate centers in the Midwestern United States. Therapy is conducted in the patient's home, one session per day, six sessions per week, usually over an eight-week period. Each session in Week 1 lasts 10 minutes. Two-minute increments are added to sessions each week up to the eighth week during which each session lasts 24 minutes.

An air source is delivered to the nasal cavities via a mask that covers only the nose. The mask is attached to a flexible hose which, in turn, is attached to the CPAP machine. The machine generates an airflow and a heightened air pressure. For safety reasons, the pressure generated by the machine does not exceed 20 cm H_2O, about the same level generated by the speaker's respiratory system during loud speech. In an informal study of nasal pressures generated during nose-blowing in three male and two female normal adults, we found that the males generated pressures between 50-60 cm H_2O and females generated

pressures between 40-50 cm H_2O (Moon and Kuehn, 1993). Therefore, the maximum pressures generated by the CPAP machine appear to be at a safe level.

In the current protocol, the starting CPAP pressure during the first session is 4.0 cm H_2O, and the maximum pressure during Week 8 is 8.5 cm H_2O. An incremental pattern is used between these extremes over the course of therapy. These intranasal pressures are within the range of typical intraoral air pressure values during the production of obstruent speech sounds for men, women, and children (Subtelny, Worth, and Sakuda, 1966). If the velum is lowered, the two pressure heads summate, thus exceeding the usual intraoral pressure to which the speaker is accustomed. This produces an uncomfortable feeling of fullness in the oral cavity. Warren argued that intraoral air pressures during speech normally are "regulated," that is, kept fairly constant. Thus, the summation of the two pressure heads would cause the intraoral air pressure to vary significantly and it would be difficult for the speaker to keep the intraoral air pressure constant.

An intraoral air pressure overload might be perceived by the individual as physiologically stressful and in need of compensation. To overcome the summation of the two pressure heads described above, the velopharyngeal port could be kept tightly closed thereby isolating the CPAP pressure from the respiratory pressure. The neuromuscular involvement might be recruitment of higher threshold motor units in response to a physiologically stressful situation (Sale, 1986). In the context of neural adaptation, it may be that more forceful velopharyngeal closure movements are achieved not because of physical changes in the muscles proper but because of the manner in which the nervous system utilizes the muscles.

Theoretically, the column of air impinging on the velopharyngeal port is a form of loading or resistance exercise in that the muscles of velopharyngeal closure must exert more force to close the velopharyngeal port. In support of such a loading phenomenon, Kuehn, Moon, and Folkins (1993) showed that increases in intranasal air pressure were accompanied by significant increases in LVP muscle activity in subjects with and without cleft palate.

Each therapy session includes speech drill-work consisting of /VNCV/ utterances in which V = any vowel, N = any nasal consonant, and C = and obstruent consonant (stop, fricative, affricate). Emphatic stress is placed on the second syllable. The /VNCV/ utterances are produced in sets of 50. Each set is followed by six short sentences in which sentences without nasal consonants alternate with sentences containing nasal consonants. The sentences are followed by a different set of 50 /VNCV/ utterances and a different set of six sentences until the time for the day's session expires.

The rationale for the /VNCV/ utterances is that velopharyngeal open position for the /N/ is immediately followed by a consonant that normally requires a high intraoral air pressure and thus a rigorous velopharyngeal closure effort especially in the context of a stressed syllable. This activity theoretically enhances the resistance that opposes velopharyngeal closure. Sentences with and without nasal consonants are used for generalization to the more typical conversational mode of production. Heightened intranasal air pressure is maintained during the sentences as well as the /VNCV/ utterances.

Kuehn (1991) has reported several patients for whom hypernasality was decreased following CPAP therapy and for whom reduced hypernasality was maintained for months following the cessation of therapy. Theoretically, maintenance of reduced hypernasality might be expected on the basis of sensory-motor linkages. Given the motoric ability to close the velopharyngeal port, continued closure ability would be expected to reduce a nasally directed breath stream during speech. A nasally directed breath stream dampens the acoustic output and would appear to be naturally avoided if the speaker has the ability to do so.

The Future

A standard protocol has been used for CPAP therapy thus far and because the procedure currently is being tested rigorously in clinical trials, the experimenters have not deviated from that standard protocol. However, modifications might be made in future applications of the CPAP procedure. For example, it may be that longer or more frequent treatment sessions produce better results. Higher CPAP pressures might also lead to better results. These are obvious potential changes that are in keeping with established principles of resistance exercises for strengthening other muscles in the body. Increases in time, frequency, or intensity relate to the well-known axiom of "no pain, no gain." However, trade-offs exist. A patient must be motivated to spend the time and effort to justify the level of improvement. Of course, this is also true for a parent on behalf of his or her child. Thus, patient compliance might be jeopardized if the exercise regimen is too demanding.

In addition to possible changes in the quantity of therapy measures, that is, time, frequency, and intensity (pressure), it is possible that qualitative changes might be tested as well. For example, there might be a means other than the use of CPAP devices to implement resistance exercises in the velopharyngeal region. CPAP is appealing because of its simplicity, but any procedure that loads the muscles of velopharyngeal closure might also be effective.

CPAP has been used in isolation without on-line feedback to the patient. It is possible that CPAP, or some other resistance exercise, might be coupled with procedures that provide information to the patient about ongoing velopharyngeal activity. For example, it is possible to drill a small hole in the nasal mask to receive an endoscope or

some other device to feed back information about the status of the velopharyngeal port to the patient. He or she could then be instructed to attempt modifications in velopharyngeal activity while the closure muscles are being loaded by the CPAP device. In this way, both sensory and motor systems would be manipulated and sensory-motor linkages might be strengthened. This might be especially valuable for carryover outside the therapy environment.

Much remains to be learned about the anatomy and physiology of the velopharyngeal region. Very basic information is lacking such as knowledge about muscle fiber types and angles of muscle insertions into the velum. Factors underlying velopharyngeal fatigue in relation to strength or endurance are not well understood. Nor are these factors understood in relation to possible anatomical constraints. Information of this type will help direct treatment approaches and guide our thinking in new avenues of research, both basic and applied.

Acknowledgements

This work was supported in part by PHS Research Grants DC-00976 and DC-01015 from the National Institute on Deafness and other Communication Disorders, and DE-10436 from the National Institute of Dental Research.

References

Adams, G.R., Hather, B.M., Baldwin, K.M., and Dudley, G.A. (1993). Skeletal muscle myosin heavy chain composition and resistance training. **Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology, 74**, 911-915.

Atha, J. (1981). Strengthening muscle. **Exercise and Sport Science Review, 9**, 1-73.

Barlow, S. (1987). Mechanical frequency detection thresholds in the human face. **Experimental Neurology, 96**, 253-261.

Basmajian, J.V. and DeLuca, C.J. (1985). **Muscles alive: Their functions revealed by electromyography** (5th edition). Williams and Wilkins: Baltimore.

Bell-Berti, F. (1976). An electromyographic study of velopharyngeal function in speech. **Journal of Speech and Hearing Research, 19**, 225-240.

Bigland-Ritchie, B. and Woods, J.J. (1984). Changes in muscle contractile properties and neural control during human muscular fatigue. **Muscle and Nerve, 7**, 691-699.

Botezat, E. (1907). Beitrage zur Kenntnis der Nervenenden in der Mundschleimhaut. **Anatomischer Anzeiger, 31**, 575-594.

Bowman, J.P. (1971). **The muscle spindle and neural control of the tongue: Implications for speech.** Springfield: Charles C. Thomas.

Brooke, M.H. and Kaiser, K.K. (1973). The use and abuse of muscle histochemistry. **The Trophic Functions of the Neuron, Annals of the New York Academy of Science, 228**, 121-144.

Broomhead, I.W. (1951). The nerve supply of the muscles of the soft palate. **British Journal of Plastic Surgery, 4**, 1-15.

Cadefau, J., Casademont, J., Grau, J.M., Fernandez, J., Balaguer, A., Vernet, M., Cusso, R. and Urbano-Marquez, A. (1990). Biochemical and histochemical adaptation to sprint training in young athletes. **Acta Physiologica Scandinavia, 140**, 341-351.

Cooper, S. (1960). Muscle spindles and other receptors. In G.H. Bourne (Ed.) **Structure and function of muscle.** Academic Press: New York.

Coyle, E.F. and Feiring, D. (1980). Muscle power improvements: specificity of training velocity. In Sale, D. and MacDougall, D. Specificity in strength training: a review for the coach and athlete. **Canadian Journal of Applied Sports Science, 6**, 87-92.

DeLuca, C.J., LeFever, R.S., McCue, M.P., and Xenakis, A.P. (1982). Behaviour of human motor units in different muscles during linearly varying contractions. **Journal of Physiology, 329**, 113-128.

DiNubile (1991). Strength training. **Clinics in Sports Medicine, 10**, 33-62.

Dixon, A.D. (1961). Sensory nerve terminations in the oral mucosa. **Archives of Oral Biology, 5**, 105-114.

Dubner, R., Sessle, B.J., and Storey, A.T. (1978). **The neural basis of oral and facial function.** New York: Plenum.

Eccles, R. and Tolley, N.S. (1987). The effect of nasal airflow stimulus upon human alae nasi E.M.G. activity. **Journal of Physiology, 394**, 78p.

Fleck S.J., and Kraemer, W.J. (1987). Designing resistance training programs. Champaign: Human Kinetics.

Freund, H. (1983). Motor unit and muscle activity in voluntary motor control. **Physiological Reviews, 63,** 387-436.

Fritzell, B. (1969). The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. **Acta Otolaryngologica, Supplement, 250,** 1-81.

Furusawa, K., Yamaoka, M., Kogo, M. and Matsuya, T. (1991). The innervation of the levator veli palatini muscle by the glossopharyngeal nerve. **Brain Research Bulletin, 26,** 599-604.

Gairns, F.W. (1955). The sensory nerve endings of the human palate. **Quarterly Journal of Experimental Physiology, 40,** 40-48.

Gairns, F.W. and Aitchison, J. (1950). Preliminary study of the multiplicity of nerve endings in the human gum. **Dental Records, 70,** 180-194.

Grimby, L., Hannerz, J., and Hedman, B. (1981). The fatigue and voluntary discharge properties of single motor units in man. **Journal of Physiology, 316,** 545-554.

Grossman, R.C., Hattis, B.F., and Ringel, R.L. (1965). Oral tactile experience. **Archives of Oral Biology, 10,** 691-705.

Grossman, R.C. and Hattis, B.F. (1964). Oral mucosal sensory innervation and sensory experience. In J.F. Bosma (Ed.), **First symposium on oral sensation and perception** (pp. 5-62). Charles C. Thomas: Springfield, Il.

Hartigan, C., Persing, J.A., Williamson, S.C., Morgan, R.F. Muir, A., and Edlich, R.F. (1989). An overview of muscle strengthening. **Journal of Burn Care and Rehabilitation, 10,** 251-257.

Holland, G.R. (1984). In J. Meyer, C.S. Squier and S.J. Gearson (Eds.). **The structure and function of oral mucosa.** Pergamon Press.

Housh, D.J. and Housh, T.J. (1993). The effects of unilateral velocity-specific concentric strength training. **Journal of Orthopedic Sports and Physical Therapy, 17,** 252-256.

Houston, M.E., Froese, E.A., Valeriote, St.P., Green, H.J., and Ranney, D.A. (1983). Muscle performance, morphology, and metabolic capacity during strength training and detraining: a one leg model. **European Journal of Applied Physiology, 51,** 25-35.

Ibuki, K., Tamaki, H., Matsuya, T., and Miyazaki, T. (1981). Velopharyngeal closure in patients with facial paralysis: A fiberscopic examination of the velopharyngeal movements. **Cleft Palate Journal, 18,** 100-109.

Kadanoff, D. (1928). Die innervation des Zahnfleisches beim menschen. **Zeitschrift fur Zellforschung, 6,** 637-646.

Kanagasuntheram, R., Wong, W.C., and Chan, H.L. (1969). Some observations on the innervation of the human nasopharynx. **Journal of Anatomy, 104,** 361-376.

Keller, J., Saunders, M., van Lovern, H., and Shipley, M. (1984). Neuroanatomical considerations of palatal muscles: Tensor and levator veli palatini. **Cleft Palate Journal, 21,** 70-75.

Komi, P.V. (1986). Training of muscle strength and power: Interaction of neuromotoric, hypertrophic, and mechanical factors. **International Journal of Sports Medicine, 7 (Suppl.),** 10-15.

Kraemer, W.J., Deschenes, M.R., and Fleck, S.J. (1988). Physiological adaptations to resistance exercise: implications for athletic conditioning. **Journal of Sports Medicine, 6,** 246-256.

Kuehn, D.P. (1979). Velopharyngeal anatomy and physiology. **Ear, Nose, and Throat Journal, 58,** 316-21.

Kuehn, D.P. (1986). Causes of velopharyngeal incompetence. **Journal of Childhood Communication Disorders, 10,** 17-30.

Kuehn, D.P. (1991). New therapy for treating hypernasal speech using continuous positive airway pressure (CPAP). **Plastic and Reconstructive Surgery, 88,** 959-966.

Kuehn, D.P. and Moon, J. (submitted). Levator veli palatini muscle activity in relation to intraoral air pressure variation. **Journal of Speech and Hearing Research.**

Kuehn, D.P., Moon, J.B., and Folkins, J.W. (1993). Levator veli palatini muscle activity in relation to intranasal air pressure variation. **Cleft Palate-Craniofacial Journal,30,** 361-368.

Kuehn, D.P., Folkins, J.W., and Cutting, C.B. (1982). Relationships between muscle activity and velar position. **Cleft Palate Journal, 19,** 25-35.

Kuehn, D.P., Templeton, P.J., and Maynard (1990). Muscle spindles in the velopharyngeal musculature in humans. **Journal of Speech and Hearing Research, 33,** 488-493.

Liss, J.M. (1990). Muscle spindles in the human levator veli palatini and palatoglossus muscles. **Journal of Speech and Hearing Research, 33,** 736-746.

Lohi, E.L, Lindberg, C., and Andersen, O. (1993). Physical training effects in myasthenia gravis. **Archives of Physical Medicine and Rehabilitation, 74,** 1178-1180.

Lucas-Keene, M.F. (1961). Muscle spindles in human laryngeal muscles. **Journal of Anatomy, 95,** 25-29.

Milner-Brown, H.S. (1993). Muscle strengthening in a post-polio subject through a high-resistance weight-training program. **Archives of Physical Medicine and Rehabilitation, 74,** 1165-1170.

Moon, J.B. and Jones, D.L. (1991). Motor control of velopharyngeal structures during vowel production. **Cleft Palate Journal, 28,** 267-273.

Moon, J.B. and Kuehn, D.P. (1993). Nasal pressures generated by normal adult male and female subjects during nose-blowing. Unpublished study.

Moon, J.B., Smith, A., and Folkins, J. (1993). Coordination of velopharyngeal muscle activity during submaximal positioning of the palate. Paper presented at the annual meeting of the American Cleft Palate-Craniofacial Association, Pittsburgh, 1993.

Nishio, J., Matsuya, T., Machida, J., and Miyazaki, T. (1976). The motor nerve supply of the velopharyngeal muscles. **Cleft Palate Journal, 13,** 20-30.

Pannbacker,M., Lass, N.J., and Stout, B.M. (1990). Speech-language pathologists' opinions on the management of velopharyngeal insufficiency. **Cleft Palate Journal, 27,** 68-71.

Pannbacker, M. (1992). Some common myths about voice therapy. **Language, Speech, and Hearing Services in Schools, 23,** 12-19.

Peter, J.B., Barnard, R.J., Edgerton, V.R., Gillespie, C.A., and Stempel, K.E. (1972). Metabolic profiles of three fiber types of skeletal muscle in guinea pigs and rabbits. **Biochemistry, 11,** 2627-2633.

Peterson-Falzone, S. (1984). Hypernasality: Comments on the article by Andrews, Tardy, and Pasternak. **Language, Speech, and Hearing Services in Schools, 15,** 222-223.

Rasch, P.J. and Morehouse, C.E. (1957) Effect of static and dynamic exercises on muscular strength and hypertrophy. **Journal of Applied Physiology, 11,** 29-34.

Ringel, R.L. and Ewanowski, S.J. (1965). Oral perception. I. Two-point discrimination in the mouth. **Journal of Speech and Hearing Research, 8,** 389-398.

Ruscello, D. (1982). A selected review of palatal training procedures. **Cleft Palate Journal, 19,** 181-192.

Ruscello, D. (1989). Modifying velopharyngeal closure through training procedures. In K. Bzoch (Ed.). **Communicative disorders related to cleft lip and palate.** (pp. 338-350). Boston: College Hill Press.

Sale, D. G. (1986) Neural adaptation in strength and power training. In N.L. Jones, N. McCartney, and A.J. McComas (Eds). **Human Muscle Power.** (pp. 281-305). Champaign: Human Kinetics.

Sale, D.G. (1988) Neural adaptation to resistance training. **Medicine and Science in Sports and Exercise. 20,** Supplement 5, S135-S145.

Sale, D. and MacDougall, D. (1981). Specificity in strength training: a review for the coach and athlete. **Canadian Journal of Applied Sports Science, 6,** 87-92.

Sale, D.G., McComas, A.J., MacDougall, J.D., and Upton, A.R.M. (1982). Neuromuscular adaptation in human thenar muscles following strength training and immobilization. **Journal of Applied Physiology: Respiration, Environment, Exercise Physiology, 53,** 419-424.

Schmidt-Nowara, W.W. (1984). Continuous positive airway pressure for long-term treatment of sleep apnea. **American Journal of Diseases of Children, 138,** 82-92.

Seaver, E.J. and Kuehn, D.P. (1980). A cineradiographic and electromyographic investigation of velar positioning in non-nasal speech. **Cleft Palate Journal, 17,** 216-226.

Sedlakova, E. (1967). The syndrome of the congenitally shortened velum: The dual innervation of the soft palate. **Folia Phoniatrica, 19,** 441-450.

Subtelny, J.D., Worth, J.H., and Sakuda, M. (1966). Intraoral pressure and rate of flow during speech. **Journal of Speech and Hearing Research, 9,** 498-518.

Takekura, H. and Yoshioka, T. (1990). Different metabolic responses to exercise training programmes in single rat muscle fibers. **Journal of Muscle Research and Cell Motility, 11,** 105-113.

Tanaka, H., Costill, D.L., Thomas, R., Fink, W.J., and Widrick, J.J. (1993). Dry-land resistance training for competitive swimming. **Medicine and Science in Sports and Exercise, 25,** 952-959.

Tax, A.A.M., van der Gon, J.J.D., Geilen, C.C.A.M. and Kleyne, M. (1990). Differences in central control of m. biceps brachii in movement tasks and force tasks. **Experimental Brain Research, 79,** 138-142.

Wand, N., Hikida, R.S., Staron, R.S, and Simoneau, J.A. (1993). Muscle fiber types of women after resistance training--quantitative ultrastructure and enzyme activity. **European Journal of Physiology, 424,** 494-502.

Winckler, G. (1964). L'equipment nerveux du muscle tenseur du voile du palais. **Archives d'Anatomie, d'Histologie, d'Embryologie,** (Strasb), **47,** 311-316.

Yamaguchi, S., Takumida, M., Taira, T., and Hiramoto, M. (1989). Morphological and histochemical observation of human levator muscles. **Practica Otolaryngologica Kyoto, 82,** 1779-1784.

NCVS Status and Progress Report - 6
May, 1994, 53-60

Measurement of Velopharyngeal Closure Force During Vowel Production

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

Abstract

The purposes of this study were to a) design and test a new velopharyngeal closure force sensing bulb, and b) use the closure force bulb to gather additional information on the variations in closure force associated with different vowels. The closure force sensing bulb possessed a flat frequency response to 30 Hz. Its output was highly linear relative to applied gram force. Reliable placement of the bulb in human subjects was achieved following prescribed placement criteria. The bulb was sensitive to small variations in velopharyngeal closure force. In agreement with some previous reports, high vowels were associated with greater velopharyngeal closure forces than low vowels. The results of this investigation support the notion that articulatory goals are specified for vowels. This vowel specificity was observed during production in both isolation and in context.

Speech is produced using a number of different articulators (i.e. lips, jaw, tongue, velum). The velopharyngeal mechanism, particularly the velum, has received much less attention in the literature as an articulator than have the lips, jaw, and tongue. Describing characteristics of its control, however, is no less important to our understanding of the speech production process. Development of a model of speech production requires a body of data regarding control properties of all of the articulators. Previous studies of control of the lips, jaw, and tongue have focused on displacement (Kuehn and Moll, 1976; Barlow, 1984), movement velocity (Kuehn and Moll, 1976; Munhall et al., 1985), force control (Barlow and Netsell, 1986; Barlow and Burton, 1990), and closure force or pressure (Lubker and Parris, 1970; Barlow and Rath, 1985; Williams et al., 1988; Hinton and Luschei, 1992; Robin et al., 1992; Wood et al., 1992).

Various aspects of velopharyngeal control also have been assessed, each providing valuable information about the performance of that valve. Descriptions of velar displacements (Moll and Daniloff, 1971; Kent et al., 1974; Kuehn, 1976) and of velar movement velocity (Hoopes et al., 1970; Kuehn, 1976) have been reported. However, few reports of velopharyngeal closure force or pressure have appeared in print. Nusbaum et al. (1935) studied the firmness of velar-pharyngeal occlusion during production of various vowels. This was accomplished by applying air pressure to the subject's nasal cavity during vowel production until the seal between the velum and pharyngeal wall was broken, allowing the impounded pressure to be vented to the oral cavity. Average pressures ranged from 7 cm H_20 for [æ] to 23 cm H_2O for [u].

However, as noted by Linville et al. (1984), one cannot assume that pressures required to force the velum open from the nasal side are equal to pressures associated with the maintenance of velopharyngeal closure during speech.

Others have attempted to record velopharyngeal closure pressures by inserting transducers into the velopharyngeal space, with mixed results. Goto (1977) designed a tube-like pressure sensing bulb to record velopharyngeal port closure force. The balloon-like bulb was coupled to a pressure transducer outside the nose. The

device was calibrated against known circumferential gram forces, and found to be linear. Goto reported velopharyngeal port forces for 21 normal speakers and 63 speakers with repaired palatal clefts. Greater closure force values were observed for the stop consonants (average ranging from 74.4 - 79.1 grams) compared to vowels (average ranging from 45.1 - 65.8 grams). Within the vowels tested, high vowels were produced with greater closure forces than low vowels.

Cohn et al. (1986) utilized a circumferential pressure transducer typically used to probe esophageal sphincteric pressures. Many potential subjects could not tolerate insertion or retention of the probe in their nasal passages. In addition, intrasubject variability was judged to be excessive. The authors attributed this either to unreliable instrument performance, instability in probe placement, variations in speech performance, or a combination of these factors. Given the circular shape of the probe, it is likely that positioning was at least one factor. That is, probe shifting to the left or right on closure caused by the elevating velum may well have produced variable closure pressure levels as pressure was recorded at a different place. Acknowledging these difficulties, Cohn et al. reported closure pressure data for 8 normal speakers and 6 speakers with repaired palatal clefts. Pressures in excess of 40 mm Hg were observed for the fricative [s]. Lower pressures (20 - 30 mm Hg) were observed for vowels. High vowels were produced with greater closure pressures than low vowels.

The ability to record closure forces between the velum and posterior pharyngeal wall reliably during speech would add a new dimension to our level of understanding of velopharyngeal dynamics. It would allow us to test previously unanswered questions regarding muscular control of that articulator. It would also allow us to evaluate the effects of palatal clefting on the ability to achieve appropriate velopharyngeal closure forces during speech and nonspeech tasks. Finally, it would allow us to evaluate the effects of therapies (both surgical and behavioral) on the ability of patients to isolate the oral from the nasal cavity.

The purpose of this study was twofold. The first aim was to design and test a new velopharyngeal closure force sensing device. The device was constructed according to a number of specified design criteria. The second aim was to use the closure force bulb with a normal subject population to a) gather data regarding reliability of placement, and b) report additional findings on the variations in closure force associated with different vowel sounds.

Methods

Part A Device Design and Calibration

A number of design and performance criteria were adhered to during the construction and testing of the closure force bulb. These included device size and shape, frequency response, and response linearity.

Device Size and Shape. A number of size and shape issues were considered. The device needed to be large enough to provide an accurate estimate of velopharyngeal closure force against a posterior pharyngeal wall that is seldom flat. That is, in the normal individual the nasal surface of the velum is convex with a superimposed bulge, the musculus uvulae eminence. In individuals with cleft palate, the uvular bulge may be lacking and a slight depression evident. Thus, the nasal surface of the velum in individuals with and without cleft palate is generally uneven. In the adult, the surface of the posterior pharyngeal wall is somewhat concave and is relatively more even compared to the velum. In children, the surface of the posterior pharyngeal wall is uneven owing to the presence of adenoid tissue that may be quite variable in size and location. The contact surface of the bulb needed to be relatively wide to ensure that it did not lodge in regions of high or, more likely, low closure force and sense force in a small area that might not be representative of overall closure force. At the same time, however, the device needed to be small enough to be passed through the nasal passage into position at the velopharyngeal port. The device needed to be flexible enough to conform to the topography of the posterior pharyngeal wall and velum during velopharyngeal closure, and to interfere with velar movements as little as possible. Flexibility was also thought to be an important design attribute when considering insertion through the nasal passage. Finally, a rounded or cylindrical shape led to slippage from a region of high closure force to a lower force region. A flattened shape was found to be less susceptible to such instability.

The force sensing bulb designed and used in this study was a teardrop shaped silastic bulb that was flattened on two sides. The bulb was produced by dipping a teardrop shaped metal mold into a silicone rubber dispersion three times, with sufficient time between each dip to allow air cooling. The finished bulb was removed from the mold and its open, smaller end attached to silastic tubing (approximately 2 mm inside diameter and 3 mm outside diameter) with medical grade liquid adhesive. The finished bulb was 5 mm thick, and 10 mm wide at its widest point (Photo 1 - see center plate).

The finished bulb is larger than the typical 3 mm diameter endoscope used to view the velopharynx. However, it is also much more flexible and collapsible. Most subjects tested to date have undergone insertion of the bulb with little discomfort following a light application of a topical anesthetic to the nasal passage and a light coating of surgical lubricant to the bulb. Successful insertion in many of the speakers with smaller nasal passages was accom-

plished by suctioning the bulb. This was accomplished by attaching a 1 cc syringe in series with the bulb and pressure transducer and withdrawing the syringe plunger. Following insertion of the bulb, the plunger was pushed back in until transducer output offset returned to zero. Attempts to insert the bulb have not been successful in some speakers with smaller nasal passages, even following bulb suctioning.

Frequency Response. The frequency response characteristics of the bulb and associated differential pressure transducer were assessed using a position controlled servo system (Ling-Altec model 408 servo). The force bulb was connected to a Honeywell Microswitch (model 162PC01D) transducer. The transducer signal was amplified and digitized on-line using a commercial software package. The force bulb was first positioned against a rigid wall. The servo controlled plunger was positioned perpendicular to the flattened surfaces of the force bulb such that a small closure force signal was recorded when the plunger was positioned at its minimum horizontal excursion point. The contact surface of the plunger was circular in shape, with a surface area of 78 mm^2.

A sine wave generator (Krohn-Hite model 1400) was used to drive the servo at 6 frequencies ranging from 1 to 30 Hz. Horizontal plunger displacements of 0.64 and 1.30 mm were used. The frequency response characteristics of the closure force bulb and pressure transducer combination were assessed by measuring the peak - to - peak amplitude of the closure force bulb output as a function of servo frequency. The results are displayed in Figure 1. With the exception of a small elevation in output at 1 Hz for the higher amplitude condition, the response of the force bulb and transducer was flat to 30 Hz for both displacement conditions. Opening and closing movements of the velopharyngeal valve typically do not occur in less than about 60 msec (Kuehn, 1976), or about 17 Hz. Therefore, the frequency response characteristics of the closure force bulb appear to be sufficient for dynamic measurement of velopharyngeal closure force.

Response Linearity. As is the case with any transduction system, response linearity is highly desirable. This aspect of bulb performance was tested by simulating velar and pharyngeal wall contact on opposite sides of the bulb. The bulb was positioned on a wood platform 10 mm in width to simulate contact with the posterior pharyngeal wall. The platform extended over the edge of a table to facilitate application of gram weights. Velar contact was simulated by placing a rigid wood cap 17 mm wide and 10 mm long on top of the force bulb. Weights were then applied to the wood cap by hanging them on a loop of thread draped over and secured to the wood cap. Five weights ranging from 20 to 100 grams were applied to the wood cap in randomized order. Each weight was applied five times. The response linearity characteristics of the force bulb are displayed in Figure 2. It is clear that the force bulb and associated pressure transducer output are highly correlated to applied force.

Part B Closure force Associated with Vowel Production

Instrumentation. The closure force bulb was coupled to a Microswitch (model PC16201D) pressure transducer. The transducer output was amplified using a Biocommunications Electronics (model 205) amplifier.

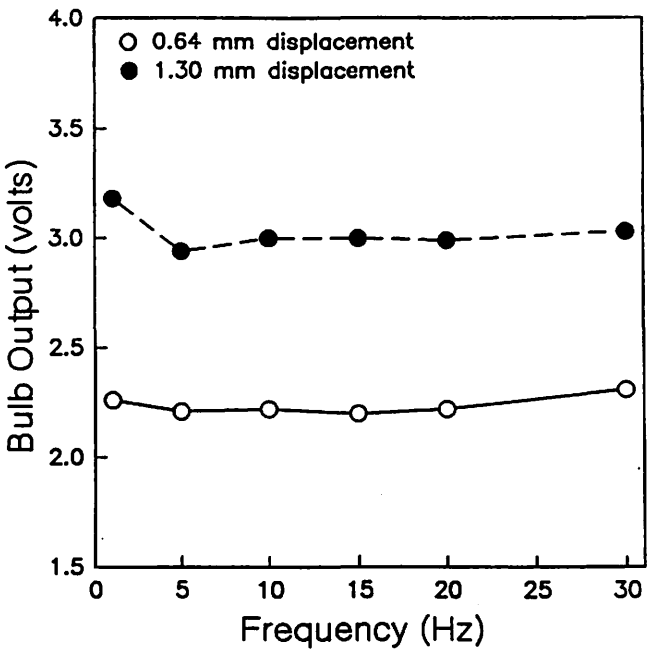


Figure 1. Frequency response characteristics of closure force bulb and pressure transducer.

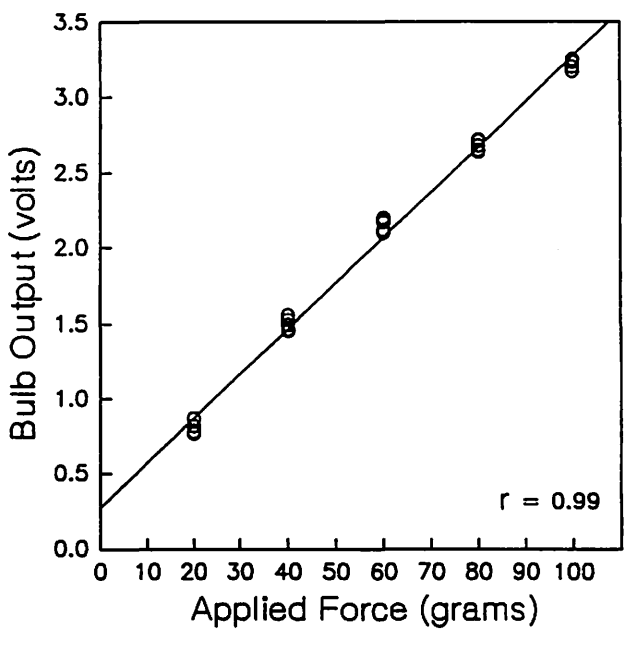


Figure 2. Response linearity characteristics of closure force bulb and pressure transducer.

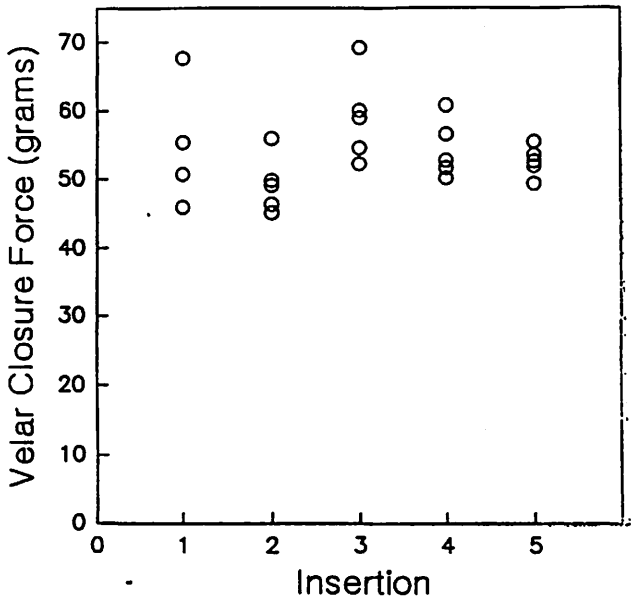


Figure 3. Closure force values associated with five productions of the vowel [ɑ] recorded during five separate bulb insertions.

Each subject's voice signal was transduced using a Sony microphone coupled to a Nakamichi (model MX-100) mixer and Tascam (model 22-4) amplifier. The voice and velopharyngeal closure force signals were recorded on two channels of a Sony (model PC108M) digital instrumentation recorder for subsequent analysis.

Calibration of the bulb was accomplished as described in the preceding section on response linearity. However, a second calibration was also conducted. Previous work (e.g. Cohn et al., 1986) has reported velopharyngeal closure pressure instead of closure force. To directly compare our data with these previous studies, the pressure transducer output was also calibrated for pressure. Specifically, pressure transducer outputs were calibrated against known air pressures (in cm H_2O) applied to the pressure transducer using a U tube water manometer.

Subjects. Velopharyngeal closure force data were collected from seven adult speakers (two male and five female). Average age was 32 years 9 months (range = 22:3 to 43:0). All subjects were judged by the experimenters to demonstrate normal velopharyngeal function as assessed by oral examination and perceptual judgment.

Speech Sample. Each subject was asked to prolong the vowels [ɑ], [i], [u], and [æ] for approximately 5 seconds each. Ten repetitions of each vowel were recorded. Subjects were then asked to produce ten repetitions of the phrase "say s_t again" using each of the four vowels listed above. Order of production of each of the vowels within the isolated and context conditions was counterbalanced.

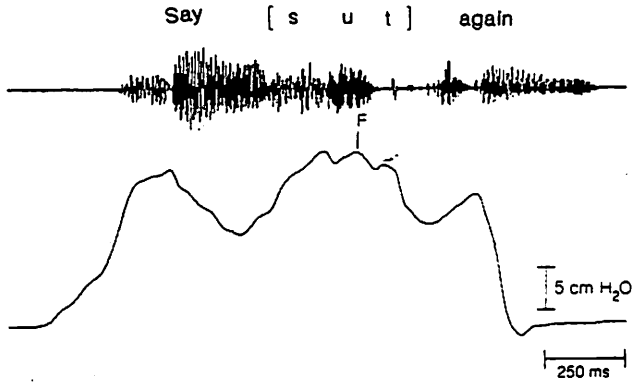


Figure 4. Voice and closure force bulb signal associated with vowel [ɑ]. Average closure force measured between T_1 and T_2

Procedure. Prior to insertion of the velopharyngeal force bulb, 4 % Lidocaine was applied lightly to the more patent side of the nose using a hand held atomizer. In addition, the bulb was coated with a surgical quality lubricant. The bulb was then inserted and positioned between the velum and posterior pharyngeal wall. Vertical bulb position was adjusted by pushing and pulling on the tube extending from the nose until a maximum force signal was observed during repeated productions of [s]. The silastic tubing was then taped to the nose to anchor the bulb in position.

Placement Reliability. The procedure of bulb placement described in the previous section was assessed in a preliminary study. Specifically, the closure force bulb was inserted and positioned in a single subject on five separate occasions using the bulb placement criteria described in the previous section. Figure 3 shows closure force values measured during five repetitions of the vowel [ɑ] as a function of insertion number. A single factor analysis of variance revealed no significant differences ($p > 0.05$) in velopharyngeal closure force as a function of insertion. It was also observed in the pilot study, using an endoscope inserted nasally, that the bulb tends to seat itself in a vertical position at the junction between the velum and pharyngeal wall.

Data Analysis. Recorded voice and closure force signals were digitized at a 1000 Hz sampling rate using commercially available software. Closure force signals then were smoothed digitally using a 30 ms time constant.

For each vowel produced in isolation, average closure force was measured over the steady state portion of the vowel production (e.g. between T_1 and T_2 in Figure 4). For each vowel produced in context, the corresponding vowel closure force observed between the closure force peaks associated with the [s] and [t] in [s_t] was recorded (Figure 5 and 6).

A mixed model analysis of variance was used to assess the relationship between vowel produced and velopharyngeal closure force, and between velopharyngeal closure forces associated with vowels produced in isolation versus [s_t] context. The model included condition (isolated versus context) and vowel as fixed factors. Random factors included subjects and the subject-by-condition, subject-by-vowel, and subject-by-condition-by-vowel interactions. Post hoc analyses involved Bonferroni multiple comparisons. The significance level used for these 16 comparisons was 0.003 (0.05/16).

Reliability. Measurement reliability was assessed by having a second observer a) identify T_1 and T_2 and derive average closure force for isolated vowels produced by one subject (N = 40 tokens)and b) remeasure closure force associated with vowels produced in context for one subject (N = 40 tokens). Intraclass correlation coefficients of 0.99 were obtained in both cases. The average difference in closure force recorded by the two observers was 0.50 grams for vowels in isolation and 0.95 grams for vowels in context.

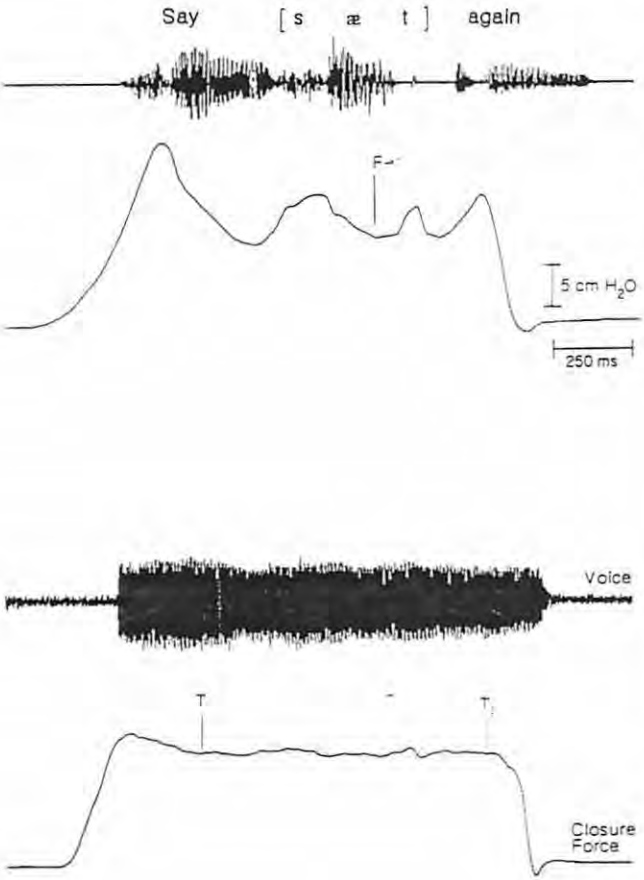


Figure 5 (top). Voice and closure force bulb signal associated with phrase "Say [sæt] again" (F - velopharyngeal closure force associated with vowel [æ]). Figure 6 (bottom). Voice and closure force bulb signal associated with phrase "Say [sut] again" (F - velopharyngeal closure force associated with vowel [u]).

Results

Mean and standard deviation values for velopharyngeal closure force associated with vowels produced in isolation and in context are shown in Table 1. It was observed that intraspeaker closure force variability (expressed as standard deviation) as measured over ten productions of the speech sample was small in most cases. The higher standard deviation levels associated with the overall values in Table 1 reflect both intra and interspeaker variability, and were not surprising. One factor that may contribute to interspeaker variability is gender. For example, McKerns and Bzoch (1970) reported that amount of contact between the velum and posterior pharyngeal wall, measured radiographically, was greater for females. Gender differences in velopharyngeal closure force were not systematically addressed in this study. However, visual analysis of Table 1 did not reveal a gender effect.

The condition main effect was non-significant ($p > 0.05$). That is, no overall differences in closure force were observed between the isolated and context conditions. Further, no significant differences ($p > 0.003$) in closure force were observed for a given vowel compared between the two conditions.

Within the isolated vowel condition, all comparisons were significant ($p < 0.003$) except for the vowels [u]

Table 1.
Mean and standard deviation values for velopharyngeal closure force (grams) associated with vowels produced in isolation and in the context "Say s_t again". Standard deviation values are in parentheses. Values may be converted to cm H_2O by multiplying by 0.36.

Isolation

Subject	u	i	ɑ	æ
1M	42.84 (2.43)	48.89 (3.51)	41.52 (4.63)	27.28 (4.43)
2M	64.37 (4.72)	66.28 (2.71)	35.47 (10.08)	41.41 (9.61)
3F	54.86 (10.06)	25.20 (2.40)	13.32 (4.06)	16.36 (3.98)
4F	65.11 (10.51)	44.09 (2.97)	43.67 (4.90)	39.52 (4.08)
5F	62.08 (7.38)	51.45 (7.16)	37.88 (6.01)	34.09 (5.68)
6F	50.70 (3.83)	30.84 (2.78)	18.56 (5.45)	25.54 (6.71)
7F	49.03 (1.76)	42.23 (3.75)	29.95 (5.89)	26.88 (3.43)
Overall	55.57 (10.22)	44.42 (13.18)	31.30 (12.35)	30.15 (9.86)

Context

Subject	u	i	ɑ	æ
1M	50.67 (2.37)	54.40 (1.98)	41.68 (5.38)	35.60 (3.81)
2M	38.60 (6.18)	67.92 (6.25)	41.07 (9.78)	39.74 (11.14)
3F	66.96 (2.18)	31.30 (6.09)	43.80 (7.33)	11.41 (11.50)
4F	68.53 (3.03)	77.29 (10.82)	47.22 (3.37)	47.10 (5.85)
5F	49.14 (5.75)	44.46 (5.22)	29.55 (6.29)	28.84 (10.24)
6F	31.92 (4.09)	27.26 (2.36)	11.67 (3.73)	11.89 (6.88)
7F	54.43 (4.01)	47.65 (4.37)	26.23 (4.18)	22.79 (4.43)
Overall	51.47 (13.21)	49.71 (17.56)	34.46 (13.14)	28.20 (15.00)

versus [i] and [ɑ] versus [æ]. However, the difference between [u] and [i] did approach significance ($p = 0.006$). Within the context condition, the same pattern was observed. That is, all comparisons were significant ($p < 0.003$) except for the vowels [u] versus [i] and [ɑ] versus [æ].

Discussion

This study was performed to a) design and test a new velopharyngeal closure force sensing device, b) test its reliability, and c) report closure force data associated with various vowels.

As demonstrated by the linearity and frequency response data obtained, the velopharyngeal closure force bulb possesses the performance properties required to transduce accurately the force of velopharyngeal closure during speech. It has also been demonstrated that, following the placement criteria outlined, reliable and consistent positioning of the bulb can be achieved across individuals and on repeated insertions in the same individual.

The results obtained for various vowels produced in isolation and in [s_t] context are of interest, and may be interpreted in light of previously published work concerning velar control. The absolute values of velopharyngeal closure force compare favorably with those reported by Goto (1977) for Japanese speakers, although mean values obtained in the present study were about 10 grams lower. Similarly, if converted to units of pressure, values obtained in the present study are lower than those observed by Cohn et al. (1986). In addition, variability (expressed as standard deviation) was about one half the level reported by Goto (1977).

There are a number of possible explanations for the differences noted above. It may be that Japanese speech is produced with greater levels of velopharyngeal closure force. Another explanation may lie in the design of the device. Our closure force bulb was designed to transduce a large portion of the horizontal extent of velar contact. This increased the potential for measuring typical closure force in the region. It also minimized the measurement of localized high or low closure forces that might occur due to tissue surface geometry and reduced measurement variability due to variations in bulb position, especially in the medial-to-lateral dimension, from one production to the next. The transducers used by Goto (1977) and Cohn et al. (1986) were cylindrical in shape. It may be that the positioning protocol used by these investigators resulted in placement of the bulb in a localized region of higher closure force (e.g. lateral). The greater variability reported by Goto (1977) could be related to the potentially greater susceptibility of his system to localized pressure maxima and minima. Regardless of the absolute differences, the same pattern of higher closure force for high vowels than for low vowels was observed in all three studies.

Bell-Berti et al. (1979) observed variable velar positions as a function of vowel produced. This is consistent with many other studies showing that vowels, particularly in isolation, vary in relation to tongue height (e.g. Moll, 1960). Velar height is greater for high vowels than for low vowels. However, the specification of velar position for vowels within a connected speech context is less clear. Moll and Daniloff (1971) argued that velar position for vowels in context is "unspecified" and depends on the specification (velopharyngeal port open versus closed) of the surrounding speech sounds. Contrary to the suggestion of Moll and Daniloff, the results of the present investigation support the notion that articulatory goals are specified for vowels. Bell-Berti et al. (1979) demonstrated vowel specific variations in velar height. Vowel specific variations in velopharyngeal closure force were observed in the present study. This vowel specificity in closure force was observed in both conditions (isolation versus [s_t] context). Graphic evidence for variation in velopharyngeal closure force observed across different vowels produced in context appears in a comparison of Figures 5 and 6.

The fact that systematic variations in velopharyngeal closure force were observed in the isolated condition and also in context is strong evidence for preprogrammed articulatory goals for vowels. However, it is not clear whether the goal is a target velar height depending on the vowel, a target velopharyngeal closure force, or some other target that results in the variations in vowel height or closure force observed in this and other studies. Based on the findings reported by Kuehn et al. (1982), the target does not appear to be a given level of activity in an individual muscle. In agreement with Seaver and Kuehn (1980), Kuehn et al. (1982) found that levator muscle activity did not appear to be directly related to velar position. Kuehn et al. did suggest that combined activity in more than one velar muscle (i.e. levator, palatoglossus, palatopharyngeus, superior constrictor) might be more related to velar position. The findings of Moon et al. (1993) demonstrated that such a coordinative structure framework does account for variations in velar position.

Control of force has, however, been accepted as a central component of speech motor control theory (Barlow and Bradford, 1992). Therefore, it is of some interest to interpret the velopharyngeal closure force measures obtained in the present investigation in relation to the force generating properties of other articulators during speech. Muller et al. (1985) estimated that labial closure for bilabial consonants was associated with about 0.5 Newtons (50 grams) of contractile force. Lubker and Parris (1970) and Hinton and Luschei (1992) recorded labial contact pressures ranging from about 10 to 30 cm H_2O during speech.

McGlone et al. (1967) recorded lingual-alveolar pressures ranging from about 5-50 g/cm^2 (5 - 50 cm H_2O) during production of [t,d,n]. We recorded average velopharyngeal closure forces ranging from 11-77 grams. Calibrated for pressure, average closure pressures ranged from 4 - 28 cm H_2O. Therefore, velopharyngeal closure forces (or pressures) recorded in the present study are comparable to those values reported previously for other articulators.

Of additional interest is the determination of where speech function lies within the overall operating range of articulator performance. Kuehn and Moon (1993) used a nonspeech task (blowing) to elicit maximal levator activation and reported that, on average, levator activation for speech tends to range from about 5 - 30 % of the maximal activation level observed during blowing. Similarly, Robin et al. (1993) measured maximal tongue - hard palate contact pressures of about 900 cm H_2O in normal speakers during a nonspeech task. Given the tongue-hard palate values expressed above in the McGlone et al. (1967) study (5 - 50 cm H_2O), it appears that tongue activity for speech also occurs at a relatively low level of physiologic effort. Future studies will determine the performance limits of velopharyngeal closure force in relation to other articulators, and in relation to force levels typically observed during speech.

In summary, a new velopharyngeal closure force sensing bulb has been developed, bench tested, and applied to a study of velopharyngeal closure forces associated with vowel production by normal speakers. Further delineation of maximal and typical closure forces associated with normal speech are expected to provide valuable information about the speech production process. In addition, study of the abnormal velopharyngeal mechanism using this device may be expected to provide new insights into the nature of velopharyngeal dysfunction and may prove useful in the evaluation of both behavioral and surgical intervention techniques.

Acknowledgements

We would like to acknowledge the statistical support of the Biostatistics Core of the National Center for Voice and Speech. This study was supported by PHS Research Grant DC-00976 from the National Institute on Deafness and Other Communication Disorders.

References

BARLOW S. Fine force and position control of select limb and orofacial structures in the Upper Motor Neuron Syndrome. Doctoral Dissertation: University of Wisconsin, 1984.

BARLOW S, BRADFORD P. Measurement and implications of orofacial muscle performance in speech disorders. J Human Muscle Perform 1992; 1: 1031.

BARLOW S, BURTON M. Ramp-and-hold force control in the upper and lower lips: developing new neuromotor assessment applications in traumatically brain injured adults. J Speech Hear Res 1990; 33: 660-675.

BARLOW S, NETSELL R. Differential fine force control of the upper and lower lips. J Speech Hear Res 1986; 29: 163-169.

BARLOW S, RATHE. Maximum voluntary closing forces in the upper and lower lips in humans. J Speech Hear Res 1985; 28: 373-376.

BELL-BERTI F, BAER T, HARRIS K, NIIMI S. Coarticulatory effects of vowel quality on velar function. Phonetica 1979; 36: 187-193.

COHN E, FISHER S, McWILLIAMS B, SETH A, FERKETIC M, BEERY Q, VALLINO L. An investigation of velopharyngeal sphincteric pressures during speech and non-speech tasks. Presented at annual meeting of the American Cleft Palate Assoc., 1986.

GOTO T. Tightness in velopharyngeal closure and its regulatory mechanism. J Osaka Univ Dental Soc 1977; 22: 1-19.

HINTON V, LUSCHEI E. Validation of a modern miniature transducer for measurement of interlabial contact pressure during speech. J Speech Hear Res 1992; 35: 245-251.

HOOPES J, DELLON A, FABRIKANT J, EDGERTON M, SOLIMAN A. Cineradiographic definition of the functional anatomy and pathophysiology of the velopharynx. Cleft Palate J 1970; 7: 443-454.

KENT R, CARNEY P, SEVEREID L. Velar movement and timing: evaluation of a model for binary control. J Speech Hear Res 1974; 17: 470-488.

KUEHN D. A cineradiographic investigation of velar movement variables in two normals. Cleft Palate J 1976; 13: 88-103.

KUEHN D, MOLL K. A cineradiographic study of VC and CV articulatory velocities. J Phonetics 1976; 4: 303-320.

KUEHN D, MOON J. Levator veli palatini muscle activity in relation to intraoral air pressure variation. Presented at annual meeting of American Cleft Palate-Craniofacial Association, 1993

KUEHN D, FOLKINS J, CUTTING C. Relationships between muscle activity and velar position. Cleft Palate J 1982; 19: 25-35.

LINVILLE R, SCHERER R, FOLKINS J. A preliminary study of velopharyngeal kinetics. Paper presented at annual meeting of American Speech-Language-Hearing Assoc., 1984.

LUBKER J, PARRIS P. Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates [p] and [b]. J Acoust Soc Amer 1970; 47: 625-633.

McGLONE R, PROFFIT W, CHRISTIANSEN R. Lingual pressures associated with alveolar consonants. J Speech Hear Res 1967; 10: 606-615.

McKERNS D, BZOCH K. Variations in velopharyngeal valving: the factor of sex. Cleft Palate J 1970; 7: 652-662.

MOLL K. Cinefluorographic techniques in speech research. J Speech Hear Res 1960; 3: 227-241.

MOLL K, DANILOFF R. Investigation of the timing of velar movement during speech. J Acoust Soc Amer 1971; 50: 678-684.

MOON J, SMITH A, FOLKINS J, LEMKE J, GARTLAN M. Coordination of velopharyngeal muscle activity during positioning of the soft palate. Cleft Palate-Cranio J 1994; 31: 45-55.

MULLER E, MILENKOVIC P, MacLEOD G. Perioral tissue mechanics during speech production. In Eisenfeld J and DeLisi C, eds. Mathematics and Computers in Biomedical Application. Amsterdam: Elsevier, 1985: 363-371.

NUSBAUM E, FOLEY L, WELLS C. Experimental studies of the firmness of the velar-pharyngeal occlusion during the production of the English vowels. Speech Monographs 1935; 2: 71-80.

ROBIN D, GOEL A, SOMODI L, LUSCHEI E. Tongue strength and endurance: relation to highly skilled movements. J Speech Hear Res 1992; 35: 1239-1245.

SEAVER E, KUEHN D. A cineradiographic investigation of velar positioning in nonnasal speech. Cleft Palate J 1980; 17: 216-226.

WILLIAMS W, VAUGHN A, CORNELL C. Bilabial compression force discrimination by human subjects. J Oral Rehab 1988; 15: 269-275.

WOOD L, HUGHES J, HAYES K, WOLFE D. Reliability of labial closure force measurement in normal subjects and patients with CNS disorders. J Speech Hear Res 1992; 35: 252-258.

NCVS Status and Progress Report - 6
May, 1994, 61-65

Long-term Result of Vocal Cord Augmentation with Autogenous Fat

Carol A. Bauer, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Joseph Valentino, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Henry T. Hoffman, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa

Abstract

The use of autogenous fat for augmentation of the paralyzed vocal fold is a promising substitute to alternate injectable material such as Teflon® (polytef paste, Ethicon/ polytetrafluoroethylene) and Gelfoam® (absorbable gelatin sponge, Upjohn). Long-term histological evaluation of fat grafts to the larynx have not previously been reported in the literature. We present a case report of autogenous fat augmentation of a paralyzed vocal fold with documentation of persistent fat graft five months after transplant.

Introduction

The functional disabilities of dysphonia, aspiration and a weak cough associated with unilateral recurrent laryngeal nerve paralysis result from inefficient glottic closure due to abnormal position, bulk and muscle tone of the affected vocal cord. Controversy persists regarding the relative advantages of treatment with injection laryngoplasty, laryngeal framework surgery, or reinnervation procedures. Different investigators have variably identified each of these procedures as the preferred technique in the surgical treatment of symptomatic laryngeal paralysis.

Reports suggest that successful laryngeal reinnervation results in muscle contraction characterized by synkinesis which confers tone but not movement to the vocal fold[1]. It is theorized that reinnervation also maintains bulk of the vocal fold by providing neurotrophic factors to the thyroartyenoid muscle thereby preventing atrophy.[2] Reinnervation has yet to become generally accepted as a common treatment of laryngeal paralysis because of questions regarding its efficacy and because of the availability of less invasive procedures that have high success rates.

Laryngeal framework surgery including thyroplasty type I and its many modifications currently is the standard treatment for laryngeal paralysis in many centers. A low complication rate and a high rate of success in improving cough, swallow and strength of the voice account for the popularity of this procedure. However, improved quality of voice as distinguished from increased loudness is not consistently seen with this procedure.[3] It is theorized that vocal quality following type I thyroplasty may be compromised by distortion of the vocal fold substance through compression by the medialized cartilage segment or silastic block.[3,4,5]

Arytenoid adduction provides an alternative approach to improving glottic closure by repositioning the paralyzed vocal fold without lateral compression. This technique distorts the substance of the vocal fold only by the presence of suture passing deep within the paraglottic space as it passes from the muscular process of the arytenoid cartilage to the thyroid cartilage. Unlike type I thyroplasty, arytenoid adduction is capable of positioning the vocal process of the arytenoid into a medial and inferior position consistent with vocal fold position as it occurs during normal phonation (Woodson, personal communication). Although arytenoid adduction consistently improves posterior glottic closure, adequate medial displacement of the

membranous vocal fold does not always occur and may require supplementation with additional procedures including type I thyroplasty or injection laryngoplasty.

Injection laryngoplasty has been used successfully in the treatment of symptomatic laryngeal paralysis for over 80 years. A variety of substances have been placed into the vocal fold and paraglottic space in an effort to improve glottic closure both by repositioning the vocal fold and by providing bulk to it. Despite trials with many types of materials, problems with foreign body reaction and migration from the site of injection have limited the current practice of vocal fold augmentation with alloplastic materials to silicone and teflon.

Although silicone is widely used in Japan for laryngeal injection, teflon has been the primary alloplastic material used for this purpose in the United States. Foreign body reaction with granuloma formation uniformly develops after teflon injection, but generally is not associated with infection, airway obstruction, or deterioration in voice suffiently severe to require operative removal. Significant difficulties have been reported in the removal of teflon which, for practical purposes, makes teflon vocal fold augmentation essentially an irreversible procedure. Although strength of voice and cough are generally improved following treatment with teflon, associated vocal fold stiffness and irregular contour with unpredictable bulk effect commonly result in vocal roughness and aperiodicity.

Preliminary reports regarding the use of autologous fat for injection laryngoplasty support its use as an efficacious and minimally invasive method of providing both medial displacement and bulk to the vocal fold.[6] As a soft pliable substance, autogenous fat may be considered an ideal substance for augmentation if its transfer does not change its character and result in vocal fold distortion from fibrosis associated with inflamation and fat necrosis.

Concerns regarding necrosis of fat at the time of transfer as well as gradual resorption of successfully transferred fat have precluded its common use for vocal fold augmentation. To date there have been no short- or long-term histologic evaluations of human larynges after autogenous fat implantation. We present the first histologic confirmation of persistent viable fat after injection laryngoplasty in a patient who required total laryngectomy five months after treatment.

Case Report

A.W. is an 82 year old male who presented with a right Wallenburg syndrome after suffering a brainstem infarction in December, 1991. His primary symptoms were significant aspiration and dysphagia secondary to paralysis of the right cranial nerves 9, 10 and 11. Because of difficulty managing secretions, the patient was tracheotomy-dependent with nutrition provided by a gastrostomy feeding tube. Fat augmentation of the right vocal fold was performed in November 1992 with improvement in voice quality. Postoperatively there was a persistent posterior glottic chink with significant aspiration demonstrated on video swallow examination. A right arytenoid adduction was performed one month later, again with further improvement in voice quality and cough. However, repeat swallow studies showed continued aspiration. The patient elected to undergo total laryngectomy which was performed five months after the fat injection.

Methods

Laryngeal fat injection

The patient was administered oropharyngeal topical anaesthesia using Dyclone gargles, followed by bilateral glossopharyngeal and superior laryngeal lidocaine nerve blocks. The abdomen was sterilely prepared and after local anaesthesia was achieved, fat was harvested from the periumbilical area using a 15 french lipectomy cannula and a Wells-Johnson Aspirator II system under 1 atm of negative pressure. The fat was irrigated with normal saline to remove blood and liquified debris. A Bruning syringe and a 19 gauge needle were used to inject the fat transorally into the right vocalis muscle in the mid- and posterior-cord regions.

Tissue preparation

The larynx was excised using a narrow-field laryngectomy technique. The larynx was then fixed in formalin, decalcified, and processed for histology. After decalcification, the larynx was embedded in celloidin and cut in 20-micron sections. Hematoxylin and eosin stains were applied.

Results

Examination of the whole larynx preparation without magnification shows the presence of three separate fat graft boluses within the right vocalis muscle (Photo 2, see center-bound photo plate). Under magnification the fat grafts had normal appearing architecture. Each graft bolus contained a mass of intact adipocytes with minimal degeneration or lipid-filled cystic cavities. There was minimal evidence of surrounding inflammation (Photo 3). In each of the discrete graft sites blood-filled capillaries can be seen extending into the graft from the surrounding recipient sites (Photo 4).

Discussion

After the successful use of autogenous fat to repair an orbital defect by Neuber[7] in 1893, the technique of soft tissue fat augmentation was widely employed in a number of innovative ways. Fat grafts for soft tissue augmentation

have been used in a variety of contexts including cosmetic body contouring,[8] repair of dural defects,[9,10] orbital reconstruction,[11] palatal augmentation,[12] scar revision,[13] and sinus ablation.[14] The primary limitation of the use of fat for augmentation has been the unpredictable final result with long-term follow up. The difficulty in accurately estimating the amount of overcorrection necessary to compensate for partial graft loss is still a major factor in this technique.

In an early human study Dedo and Rowe[15] used fat grafts to augment post-traumatic soft tissue glottic defects, secondary to scarred atrophic vocal folds, and obtained good results. Subsequent work by Wexler and Gray,[16] using a dog model, demonstrated excellent phonatory characteristics in larynges following fat augmentation of surgically created vocal cord soft tissue defects. As promising as this technique is, important questions remain to be answered. The ultimate fate of the transplanted adipocytes, the optimal graft harvesting technique, the long-term survival of the graft, and the amount of overcorrection necessary during augmentation, are concerns which have yet to be answered.

The landmark work of Peer[11] in 1956 confirmed the cell survival theory of fat grafts. Prior to this work, fat transplants were thought to involute, providing a substrate for migrating fibroblasts which eventually replaced the transplanted adipocytes. Peer histologically examined fat grafts that were implanted into the rectus sheaths of human volunteers and subsequently retrieved at various stages post-implantation. He observed that microvascular connections occurred between the host and graft capillary circulation as early as day 4 post-implantation. He also reported that a connective tissue capsule developed around the graft, but that replacement of adipocytes by scar and fibrous ingrowth was not observed. Furthermore an inflammatory infiltrate of lymphocytes and macrophages was observed to occur within the first week of implantation, presumably in response to the cystic degeneration and vacuolization of traumatized, non-surviving adipocytes. Peer reported that this inflammatory infiltrate subsided as the graft was cleared of degenerating cellular debris. Additional work by Peer has shown that grafts harvested and transplanted atraumatically have a reduced inflammatory response and survive with a greater volume than traumatized specimens.

The optimum graft size for re-implantation has not yet been established. Neuber maintained that grafts the size of a "bean" were optimal, since larger grafts tended to resorb. Peer advocated a larger graft size to maintain the graft's native vasculature and enhance early revascularization. The technique of using fat "pearls", surgically excised adipose tissue divided into 5 mm minigrafts, has been used extensively in the areas of oculoplastics and facial cosmetic surgery.[17]

Shorr[18] in 1988 reported excellent cosmetic results using fat pearls to augment upper and lower eyelid soft-tissue defects. Multiple 5 mm pearls were placed beneath skin-muscle flaps in 15 patients, with follow-up from 3 months to 3 years. They estimated graft survival to be 50 to 90 percent. Histological study was available in two patients requiring subsequent graft reduction. Excised graft tissue showed normal fat architecture and a typical lipogranulomatous reaction surrounding the graft at six months post-operatively.

The development of the technique of liposuction for fat harvesting by Illouz[19] in 1977 furthered the interest in using fat grafts for augmenting soft tissue defects. However some controversy exists regarding the fate of adipocytes harvested with this technique. Some authors claim that up to 90 percent of harvested adipocytes are disrupted using standard negative pressures of 1 atm.[20] In a study of long-term fat graft survival after transplantation using the rabbit ear model, Kononas[21] demonstrated that surgically excised fat maintained greater volume than suction aspirated fat grafts, 42 percent vs 31.6 percent, respectively. Histologic evaluation of fat grafts 9 months after transplantation into subdermal pockets of rabbit ears showed greater cystic degeneration, lipid cavities, and fibrosis in the transplants harvested by aspiration. Similar results were obtained in a study by McFarland[22] comparing the survival of fat "pearl" grafts and aspirated fat "cell" grafts reimplanted beneath skin-muscle flaps of rabbit eyelids. Both gross and histological evaluation of the grafts 12 weeks post-transplantation demonstrated nearly complete resorbtion of the grafts harvested by aspiration while the fat "pearls" retained approximately 50% of the original grafted volume.

Nguyen[20] compared the effects of three harvesting techniques on adipocytes: Harvesting by suctioning at standard 1 atm negative pressure, harvesting by aspiration using a syringe and a 14 ga needle, and harvesting by surgical excision. Histological examination immediately after harvesting demonstrated that 90 percent of the suctioned adipose tissue was fragmented with extensive cell rupture, and only 10 percent of the suctioned specimens contained intact adipocytes. In contrast, 95 percent of the specimens harvested by low-pressure aspiration, or by surgical excision, showed intact fat cells with fragments of blood vessels and collagen fibers. But disappointingly, when the harvested fat was transplanted into either rectus sheaths or ear pockets of rabbits, histologic examination at 9 months post-transplantation showed only minimal survival of adipocytes, regardless of the harvesting technique.

It is difficult to reconcile the differences in the literature regarding the ultimate fate of transplanted adipocytes. Certainly a critical aspect of these transplant models is the vascularity of the recipient bed. The subdermis

of the rabbit ear is considered to be relatively avascular and therefore perhaps an inhospitable environment, even for the low metabolic requirements of fat. More relevant to the issue of satisfactory vocal fold augmentation in this regard is the work of Wexler and Gray[16]. They demonstrated in dogs, that using a laryngofissure technique, a 0.2 to 0.4 ml volume of autogenous abdominal fat placed beneath the leading vocal fold edge survived and had histologically normal appearing adipocytes at six weeks post-transplantation. Using an excised larynx preparation, the larynges that received fat augmentation after vocal fold mucosal stripping showed phonation threshold pressures within the normal range. These pressures were significantly less than in larynges after mucosal stripping without vocal fold augmentation. In addition, the fat-augmented cords showed improved vibratory amplitudes on videostroboscopic examination in comparison to the mucosal excision only larynges. Moreover, the mean amplitudes of vibration were not significantly different from the control larynges.

To our knowledge, there have been no reports of the histological features of autogenous fat grafts after transplantation in human larynges. The present case demonstrates the features of normal appearing adipocytes in discrete organized masses within the vocalis muscle five months after transplantation. A moderate inflammatory infiltrate and surrounding fibrous reaction was present adjacent to the graft boluses. In several areas within each graft, blood-filled capillaries were observed. These histological features are similar to those originally described by Peer. These findings indicate that autogenous fat can be successfully used to augment vocal fold deficiencies and that substantial retention of graft volume can occur for up to five months. Furthermore, the fat grafts are not replaced by fibrous ingrowth nor do they incite any significant local trauma at the recipient site. In the present case, the technique of liposuction did not appear to adversely affect graft survival. However, even greater volume retention may be noted with surgically harvested grafts and this potential improvement should be further investigated. In addition, a distinct and critical improvement over the use of teflon is the tolerance of the recipient vocal fold bed to grafting with autogenous material. In this case, we observed very little distortion or rigidity of the vocal fold cover in response to the underlying fat graft. The literature indicates that loss of graft volume can be expected to occur with time. We would expect that even with fibrous tissue replacement of the fat graft, augmentation with fibrous tissue would result in phonatory characteristics superior to that obtained with a foreign substance such as Teflon® with its associated inflammatory reaction. Graft resorbtion is only a relative limitation of this technique since repeated fat injections to titrate the degree of augmentation may be performed with little morbidity. The procedure is easy to perform, well-tolerated by patients, and graft material is usually in abundance. In addition, the use of combined techniques such as arytenoid adduction in conjunction with autologous fat injection is possible, as demonstrated in this case report.

Acknowledgements

The authors gratefully acknowledge Grant P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. They thank Penny Harding for her assistance in histological processing.

Bibliography

1. Crumley RL. Update: ansa cervicalis to recurrent laryngeal nerve anastamosis for unilateral laryngeal paralysis. Laryngoscope 1991 101(4): 384-7.

2. Crumley RL. Treatment of vocal cord paralysis. West J Med 1991 154 (6):715-6.

3. Gray SD, Barkmeier J, Jones D, Titze I, Druker D. Vocal evaluation of thyroplastic surgery in the treatment of unilateral vocal fold paralysis. Laryngoscope 1992 102(4):415-21.

4. Hirano M, Kakita Y. Cover-body theory of vocal cord vibration. In: Daniloff RG, ed. Speech science. San Diego: College-Hill Press 1985:1-46.

5. Titze IR, Jiang J, Drucker DG. Preliminaries to the body-cover theory of pitch control. J Voice 1988 1:314-319.

6. Brandenburg JH, Kirkham W, Koschkee D. Vocal cord augmentation with autogenous fat. Laryngoscope 1992 102:495-500.

7. Neuber G. Fett transplantation. Chir Kongr Verhandt Deutsche Gesellsch Chir 1893 22:66.

8. Watson J. Some observation on free fat grafts: with reference to their use in mammoplasty Br J Plast Surg 1959 12:263.

9. Kiviluoto O. Use of free fat transplants to prevent epidural scar formation: an experimental study. Acta Orthop Scand Suppl 1976 164:3.

10. Halstead AE, Caylor HD. Repair of dural and brain defects by free fat and fascial transports. JAMA 1924 82:13.

11. Peer LA. The neglected "free fat graft", its behavior and clinical use. Am Jour Surg 1956 92:40-47.

12. Von Gaza W. Free retropharyngeal fat grafting in cleft palate. Arch Klin Chir 1926 142:590.

13. Ellenbogen R. Free autogenous pearl fat grafts in the face - a preliminary report of a rediscovered technique. Ann Plast Surg 1986 16(3):179-194.

14. Denneny JC, III. Frontal sinus obliteration using liposuction. Otolaryngol Head Neck Surg 1986 95:15.

15. Dedo HH, Rowe LD. Laryngeal reconstruction in acute and chronic injuries. Otolaryngol Clin North Am 1983 16:373-389.

16. Wexler DB, Jiang J, Gray SD, Titze IR. Phonosurgical studies: Fat-graft reconstruction of injured canine vocal cords. Ann Otol Rhino Laryngol 1989 98:668-673.

17. Silkiss RZ, Baylis HI. Autogenous fat grafting by injection. Ophthalmic Plast Reconstr Surg 1987 3(2):71-75.

18. Shorr N, Christenbury JD, Goldberg RA. Free autogenous "pearl fat" grafts to the eyelids. Ophthalmic Plast Reconstr Surg 1988 4(1):37-40.

19. Illouz YG. Surgical remodeling of the silhouette by aspiration lipolysis or selective lipectomy. Aesthetic Plast Surg 1985 9:7-21.

20. Nguyen A, Pasyk KA, Bouvier TN, Hassett CA, Argenta LC. Comparative study of survival of autologous adipose tissue taken and transplanted by different techniques. Plast Reconstr Surg 1990 85(3):378-386.

21. Kononas TC, Bucky LP, Hurley C, May JW. The fate of suctioned and surgically removed fat after reimplantation for soft-tissue augmentation: A volumetric and histologic study in the rabbit. Plast Reconstr Surg 1993 91:763-768.

22. McFarland JE. The free autogenous fat graft. A comparison of the fat "pearl" and fat "cell" graft in an animal model. Ophthalmic Plast Reconstr Surg 1988 4(1):4147.

NCVS Status and Progress Report - 6
May, 1994, 67-72

A Preliminary Study of the Prognostic Role of Electromyography in Laryngeal Paralysis

Young B. Min, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Eileen M. Finnegan, M.A.
Department of Speech Pathology and Audiology, The University of Iowa
Henry T. Hoffman, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Erich S. Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Timothy M. McCulloch, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa

Abstract

Confidence in the reliability of laryngeal electromyography (EMG) to predict recovery is critical if this tool is to be used to select the type and timing of surgical intervention. The characteristics of EMG of 14 patients with unilateral vocal fold paralysis were assessed to determine which factor or combination of factors would be most useful in determining prognosis. We examined the duration, amplitude, waveform morphology, root-mean-square, and time interval from onset to EMG recording. The results supported the concept that EMG recordings are valuable in determining prognosis if it is performed prior to 6 months and preferably within 6 weeks of onset of laryngeal paralysis. A positive prognosis for laryngeal recovery was indicated when the following EMG features were present in the immobile vocal fold: 1) normal motor unit waveform morphology, 2) overall EMG activity characterized by an RMS value greater than 40 uV in any one task, and 3) no electrical silence during voluntary tasks. Based on this criteria our overall correct prognostic rate was 89%.

Introduction

Laryngeal electromyography (EMG) performed on patients with vocal fold paralysis (VFP) has been reported as useful in predicting return of normal laryngeal function. Identification of an accurate test to predict laryngeal recovery has important implications for counseling and for directing therapy. Confidence in the reliability of laryngeal EMG to predict recovery is critical if this tool is to be used to select the type and timing of surgical intervention.

Several studies in the current literature have evaluated the capacity of EMG to predict laryngeal recovery based on qualitative analysis of electrical activity in the intrinsic laryngeal muscle. Hirano et al.[1] based his predictions on the presence or absence of action potentials during voluntary activity. He found laryngeal EMG performed more than 6 months after the onset of paralysis was not useful for prognosis. None of these patients had recovery regardless of EMG findings. He reported that when the EMG was performed within 6 months post onset, the overall predictive rate was 69% (20/29) with 7 false positive and 2 false negative predictions.

In addition to considering the presence of EMG activity, Thumfart [2] analyzed the morphology of EMG waveforms present in 114 patients with VFP in a qualitative fashion describing electrical silence, fibrillations, positive sharp waves, and/or motor unit potentials (MUPs) during voluntary activity. His predictive rate was 71% (81/114) with 13 false positive and 20 false negative projections. In a study similar to Thumfart's [2], Parnes et al. [3] determined prognosis based on waveform morphology among 20 pa-

tients with VF paralysis. Since there was a question of bilateral paralysis in some cases, 26 vocal folds were evaluated in all. His predictive rate was 88% (23/26) with 3 false positive and no false negative predictions.

Although these investigators found laryngeal EMG to be useful in prognosis, two issues were raised. Unlike Hirano et al. [1], Parnes et al. [3] did not exclude those EMGs which were performed more than 6 months after onset of paralysis. In fact, the majority of EMGs performed by Parnes et al. [3] exceeded this designated time period. Nonetheless, Parnes et al.'s [3] prognostic rate was higher than Hirano et al.'s [1], 88% and 69%, respectively. It is unclear if the difference in the prognostic rate between the Hirano et al. [1] and Parnes et al.[3] studies was due to the decision to include patients greater than 6 months post onset, or to some other difference in their diagnostic criteria.

Another issue is Parnes et al.'s [3] definition of "decreased motor unit potentials" which he classified as a poor prognostic sign. Although other investigators [4,5] have also noted fewer than normal number of MUPs present in the paralyzed vocal fold (VF), this judgement is subjective. Furthermore, prognostic rates based on qualitative judgements are not easily replicable. To establish a standard prognostic criteria for evaluation of laryngeal EMG, a critical assessment of quantification of these characteristics should be explored.

Quantitative differences between EMG activity recorded from the mobile and the paralyzed VF have been reported in patients diagnosed with VFP. Faaborg-Andersen in 1957 [6] compared the EMG signal recorded from the normal VF to the paralyzed side and observed that the maximum amplitude was decreased, and the duration was increased. Haglund et al. [7] investigated the duration and amplitude of potentials detected in patients with VFP and compared the distribution of these measurements to that of persons with normal VFs. Most of the amplitude and duration measurements of the paralyzed VF were at the extremes of the measurement distribution of normal VF. These results suggest quantitative measurements of amplitude and duration may be useful adjuncts to the prognostic criteria.

The purpose of our study was to evaluate both the qualitative and quantitative characteristics, such as duration, amplitude, waveform morphology, root-mean-square, and time interval from onset to EMG recording to determine which factor or combination of factors would be most useful in delineating an accurate prognostic indicator.

Methods

Subjects: 14 subjects selected were evaluated at the University of Iowa between January, 1992 and May, 1993. All were diagnosed with unilateral vocal fold paralysis following visualization and recording of an immobile VF by either an oral or nasopharyngeal videoendoscopy. Additional studies were performed to assess the etiology of the immobile vocal fold based on the clinical presentation. All patients were evaluated with a standardized approach to laryngeal EMG.

Procedure: Bipolar hooked-wire electrodes were inserted into each thyroarytenoid (TA) muscle bilaterally in all patients. The needle was inserted percutaneously via a submucosal route to the TA muscle. The verification procedures for electrode placement were performed according to the standards outlined by Hirano and Ohala.[8] Simultaneous EMG recording of the mobile VF, as well as the paralyzed VF, was performed. Thus, each subject served as their own control for comparison of EMG activity in the normal VF to the immobile VF.

The protocol consisted of three activities: (1) one 25 second period of rest during which the patient was instructed to relax and breathe quietly, (2) three repetitions of sustained /i/ at normal pitch and loudness, and (3) three repetitions of the valsalva maneuver.

Signals detected by the electrodes were differentially amplified by a 4 channel preamplifier (custom-made by Bioengineering at the University of Iowa). The audio signal was transduced by a dynamic microphone placed 12 cm in front of the subject. The EMG signals and the audio signal were further amplified by an 8 channel bioamplifier and then recorded onto an 8 channel Sony Digital Audio Tape (DAT) recorder at a sampling rate of 5000 Hz/ channel. The EMG signal was simultaneously transmitted to a speaker for auditory monitoring of the EMG signal during the procedure and also to an oscilloscope and a computer monitor for on-line visual monitoring.

Study patients were seen for follow-up evaluation with visualization of vocal folds to note any change in their mobility.

Data Analysis: A nine second segment of one token of each tasks was digitized into a computer with DATAQ Instruments CODAS (computer-based oscillograph and data acquisition system) software at a sampling rate of 4000 Hz/ channel. For sustained /i/ and valsalva, the entire token was included in the digitized segment. Analysis of these segments included measurements of: 1) duration, 2) amplitude, 3) waveform morphology, and 4) root-mean-square. Utilizing the CODAS program, the compression rate of the digitized EMG signals was decreased to visualize the signal in an expanded form. A maximum of 8 individual MUPs were identified visually from both channels of the recording (i.e. from both the mobile and paralyzed TA). The duration of each MUP was manually

determined as well as the peak to peak measures of amplitude. This expanded view of the EMG signal was also used in the examination of waveform morphology in the paralyzed VF.

The root-mean-square (RMS) value, a measure of overall EMG activity level, was calculated in a 3-step process. First, a 1 second segment which was free from artifact was selected from each channel for all three tasks. The sections selected from sustained /i/ and valsalva were taken from a period of steady EMG activity. Second, a baseline RMS measure was obtained from a segment of rest which contained no EMG activity, and this was considered a measure of signal noise. Finally, the RMS value was converted to microvolts (uV) by subtracting the baseline RMS from the measured RMS and dividing by the gain.

Results

Of the 14 patients with vocal fold paralysis (VFP), 3 patients had return of VF mobility and 11 patients did not (Table 1).

Table 1.
Patient Profile

Patient	Time Interval Onset to EMG (month)	Time Interval EMG to F/U (month)	Return of Vocal fold Function?	Etiology of vocal fold paralysis
A	1	2	Y	parathyroidectomy
B	1	2	Y	R anterior cervical fusion
C	5	14	Y	voice symptoms after URI
D	1	9	N	subtotal thyroidectomy
E	4	8	N	resection of acoustic neuroma
F	4	5	N	R thyroidectomy
G	4	10	N	ligation of L ant corroidal aneurysm
H	5	9	N	R thyroidectomy
I	5	11	N	skull base tumor resection
J	7	10	N	voice symptoms after URI
K	8	15	N	R exploratory thoracotomy
L	13	4	N	R thyroidectomy
M	15	8	N	anterior cervical fusion
N	94	10	N	B carotid bypass surgery

Patients above dashed line had return of vocal fold mobility
R- right, L- left, B- bilateral, URI- upper respiratory infection
F/U- follow-up

Duration: There were 8 patients in whom the EMG of the paralyzed thyroarytenoid (TA) muscle showed motor unit potentials (MUPs) which permitted measurement of duration and amplitude (Table 2). Measurements could not be obtained from patients who showed electrical silence. The mean duration of the MUPs from the normal TA muscle of all patients was 3.17 msec (S.D. +/- 1.06 msec) and that of all the paralyzed VFs was 4.50 msec (S.D. +/- 1.55 msec). The mean duration of the MUPs of the paralyzed TA muscle in patients with laryngeal recovery were longer than that of the mobile TA. With one exception, all the patients who did not have return of VF mobility also had mean durations which were longer than that of the normal (i.e. mobile) TA muscle. Therefore, differences in mean durations were not significant between the two groups. The 10th and 90th percentiles of the total duration distribution of waveform duration in normal VFs were 1.93 msec and 4.71 msec, respectively. None of the patients had greater than 10% of their durations shorter than 1.93 msec. Although there were 6 of 8 patients whose distribution of durations was greater than 10% above 4.71 msec, some of these patients had return of VF mobility and some did not.

Table 2.
Duration of MUPs from Paralyzed TA compared to Normal TA

Patient **	Duration of MUPs-Paralyzed TA N (219)*	Mean msec (3.17)*	S.D. msec (1.06)*	Frequency Distribution <= 1.93msec % (10%)*	>= 4.71msec % (10%)*
A	14	3.44	1.84	7	21
B	15	5.01	2.10	0	27
E	24	5.31	1.40	0	63
G	9	3.87	.86	0	33
J	19	5.21	.83	0	84
L	15	3.77	1.14	7	7
M	15	4.05	.75	0	20
N	6	3.14	.12	0	0

* Corresponding data from normal thyroarytenoid muscles
** Patients with motor unit potentials for measurement
Patients above dashed line had return of vocal fold mobility
MUPs- motor unit potentials, TA- thyroarytenoid muscle,
N- number of motor unit potentials

Amplitude: The mean amplitude of MUPs from the normal TAs of all patients was 257 uV with a standard deviation of +/- 256 uV (Table 3). The mean amplitude for the paralyzed VF was 294 (S.D. +/- 159). The mean amplitude of the paralyzed TA in one of the recovered patients and several of the nonrecovered patients was greater than the mean of the normal TA. Similarly, some of the mean amplitudes of nonrecovered patients and one of the mean amplitudes of recovered patient were less than the mean of normal TA.

When distribution of the amplitude was evaluated, the 10th and 90th percentiles of the total amplitude distribution of normal VF were 63 uV and 688 uV, respectively. None of the patients with VF recovery showed amplitudes less than 63 uV or greater than 688 uV. Although there were some nonrecovered patients whose amplitudes were less or greater than the 10th and 90th percentiles of normal amplitude, there were also patients whose amplitudes were between 63-688 uV.

Table 3.
Amplitude of MUPs from Paralyzed TA compared to Normal TA

Patient **	Amplitude of MUPS-Paralyzed TA N (219)*	Mean uV (257)*	S.D. uV (256)*	Frequency Distribution <=63 uV % (10%)*	>= 688 uV % (10%)*
A	14	322	109	0	0
B	15	241	146	0	0
E	24	371	121	0	0
G	9	203	40	0	0
J	19	367	233	0	16
L	15	253	174	27	0
M	15	212	68	7	0
N	6	161	20	0	0

*Corresponding data from normal thyroarytenoid muscle
**Patients with motor unit potentials for measurement
Patients above dashed line had return of vocal fold mobility
MUPs- motor unit potentials, TA- thyroarytenoid muscle,
N- number of motor unit potentials

Table 4.
Root-mean-square of EMG activity and waveform analysis

Patient	Root-mean-square (uV) NORMAL TA Rest	/i/	Vls	PARALYZED TA Rest	/i/	Vls	Waveform Analysis*
A	20	127	146	53	52	40	M
B	5	24	18	9	42	20	M
C	1	7	157	0	0	1	S
D	19	187	239	0	1	0	S
E	10	148	154	18	32	22	M+P
F	33	89	37	0	0	0	S
G	15	26	11	0	25	0	S+M
H	11	37	37	0	1	0	S
I	6	83	ND	0	0	ND	S
J	0	0	13	3	49	144	M+P
K	23	223	225	0	0	0	S
L	7	34	103	3	17	79	M
M	5	112	145	6	21	3	M
N	15	212	81	0	8	0	S+M

/i/- phonation of sustained /i/, Vls- valsalva maneuver, ND- not done,
TA- thyroarytenoid muscle
* M- motor unit potential, S- electrical silence, P- polyphasic potential

Waveform Morphology: In 2 of the 3 patients who experienced return of function, normal motor unit potentials were identified in the EMG signals (Table 4). However, one patient who had recovery of function had electrical silence. Among those patients with persistent vocal fold immobility, 6 of the 11 patients had normal MUPs and two of these had polyphasic potentials as well. However, there were more MUPs in those patients who had laryngeal recovery. No evidence of fibrillations or positive sharp waves were found in the recorded signals.

Root-mean-square: The RMS values of the paralyzed TA were generally lower than those values of the TA from the mobile VF (Table 4). In both the paralyzed and mobile vocal fold, EMG activity during sustained /i/ and valsalva was greater than during rest, consistent with motor unit recruitment. RMS values obtained from the mobile side during voluntary tasks ranged from 7 to 239 uV, except for a single patient (J) who exhibited no EMG activity in the normal TA during two of the three tasks. It is possible that this finding was due to inappropriate electrode placement.

Two of the 3 patients who had return of VF mobility had an RMS value greater than 40 uV during at least one task. One of the patients demonstrated electrical silence. In the 11 patients with persistent immobility, 7 demonstrated electrical silence during at least one task as represented by the RMS values less than 1 uV. In 2 of the remaining 4 patients, electrical activity was detected but the RMS values were less than 40 uV. In the remaining two patients, who were both greater than 6 months post onset, the RMS value was greater than 40 uV.

Time Post Onset: At the time of EMG, five patients were more than 6 months post onset. None of these patients had recovery of function (Table 1). All three recovered patients had EMGs performed within 6 months post onset.

Discussion

The contribution of duration and amplitude measures to the prognostic criteria continues to be unclear. Consistent with the findings of Faaborg-Andersen [6], we found the mean duration of the recorded potentials from the paralyzed vocal folds (VF) to be significantly greater than that of the mobile VFs. However, the mean duration of potentials in patients with return of VF mobility was not substantially different from the mean of those patients who did not have recovery. Therefore, the mean duration had no clinical correlation to recovery. Similarly, the mean amplitude was not helpful in distinguishing the recovered group from the nonrecovered group. Furthermore, assessment of frequency distributions of the duration and amplitude, as described by Haglund et al. [7], showed no correlation to prognosis. A possible explanation for the variable findings in duration and amplitude is that these measurements are sensitive to numerous factors such as distance between the electrode and the motor unit potential (MUP), size of the muscle fibers, and patient age.[9] Greater num-

bers of patients in the recovered and nonrecovered groups may be needed to detect significance between the two groups.

The waveform morphology and the timing of its appearance provide an information about the possible pathophysiology of the damaged nerve. This information has implications for prognosis. The detection of MUPs close to the time of injury were found in two patients who had return of VF mobility. The presence of normal MUP waveforms without evidence of spontaneous denervation potentials and decreased voluntary recruitment of activity are the usual findings in intact axon such as in neurapraxia.[10] Therefore, these two patients most likely had neurapraxia of some recurrent laryngeal nerve fibers during surgery in light of their prompt recovery.

Root-mean-square (RMS) values provided a means of differentiating between decreased EMG activity associated with a poor prognosis and moderate activity indicative of a good prognosis. In our series, when EMG was performed within 6 months of onset, an RMS value of 40 uV distinguished between those patients who recovered function and those who that did not.

Consistent with Hirano et al. [1], we found that time of onset to EMG recording was a critical factor in predicting laryngeal recovery. The 5 patients who had EMG recordings more than 6 months after the onset of paralysis had no return of VF mobility regardless of the findings of their EMGs. Two of these patients had normal MUPs and/or polyphasic potentials with some RMS values equal to or greater than those of patients who had return of VF mobility. It is likely that reinnervation had occurred in these patients but not in a manner capable of restoring normal function.

Hiroto et al. [11] and others [1,6,12,13] reported EMG activity in patients without vocal fold motion. These investigators have suggested that simultaneous contraction of antagonistic muscles (i.e. synkinesis) during phonation could result in a clinically immobile VF.[3,11,12] The notion of synkineses is supported by Crumley[12] who has reported electromyographic evidence of misdirected regeneration of recurrent laryngeal nerve.

It is difficult to differentiate between synkinesis, associated with poor prognosis, and appropriately directed regeneration, which has a potential for laryngeal recovery based on the EMG signals. Hiroto et al.[11] have reported reinnervation potentials detected in patients as early as 6 weeks. Ueda[1], mentioned in Hirano et al.'s [1] study, reported that the earliest sign of regeneration was 50 days post injury in patients, as evidenced by polyphasic potentials. The results of these studies suggest that even at 2 months, the results of the EMG recording may be misleading. These reports suggest that EMG should be performed within 6 weeks of onset of paralysis when possible before reinnervation has occurred to minimize false positive prediction.

Rather than a single EMG feature, a combination of features may provide the most accurate predictive rate. A positive prognosis for recovery of vocal fold function was indicated in our study when the following features were present: 1) normal motor unit potential waveform morphology, 2) a root-mean-square value greater than 40 uV in any one task, and 3) no electrical silence during voluntary tasks. When these features were noted in EMG performed within 6 months post onset, our overall predictive rate was 89% (8/9) with one false negative prognosis. However, if these features were obtained in EMGs performed within 6 weeks post onset, the predictive rate was 100% (3/3).

EMG is not infallible tool in determining prognosis. The patient (C) had a return of vocal fold mobility without any detection of EMG activity. It is unclear why no EMG was detected. Possible reasons include inappropriate placement of electrodes, the tips of the two wires may have short circuited, and/or areas of the muscle sampled were not reinnervated.[8] The patient may have experienced a partial recovery or reinnervation which extended to other adductor muscles, such as lateral cricoarytenoid, but did not include the TA. Thumfart [12] suggests that less than 10% incidence of axonotmesis may result in regeneration despite abnormal EMG results. Some of these possible errors may be avoided by our recent use of a bifilar hooked-wire electrode with the two wires affixed together to prevent short circuiting and the placement of two separate bipolar electrodes within the same muscle to sample additional areas. In addition, serial follow-up EMG recordings especially for false negative cases, should minimize the inappropriate placement of electrodes and sampling errors.

Although the number of EMGs performed within 6 weeks post onset is small, the study suggests that the predictive rate would be considerably improved if this criteria is applied to EMGs performed within 6 weeks. The most common etiology of vocal fold paralysis is surgery. Therefore, as a part of our continued investigation, we will encourage the surgeons to refer patients within 6 weeks of suspected laryngeal nerve injury to determine if this does indeed improve our ability to provide a more accurate prognosis.

Acknowledgments

Supported by the grant #P60 DC00976 from the National Institute on Deafness and Other Communication Disorders.

References

1. Hirano M, Nosoe I, Shin T, Maeyama T. Electromyography for laryngeal paralysis. In: Hirano M, Kirchner J, Bless D, eds. Neurolaryngology: recent advances. 1st ed. Boston: A College-Hill Publication, 1987:232-248.

2. Thumfart, W. Electromyography of the Larynx. In: Samii M, Gannetta PJ, eds. The Cranial Nerves. 1st ed. Berlin: Springer, 1981:597-606.

3. Parnes S, Satya-Murti SM. Predictive value of laryngeal electromyography in patients with vocal cord paralysis of neurogenic origin. Laryngoscope 1985;95:1323-1326.

4. Blair RL, Berry H, Briant TD. Laryngeal electromyography: techniques and application. Otolaryngol Clin North Am 1978;11:325-46.

5. Rodriquez AA, Myers BR, Ford CN. Laryngeal electromyography in the diagnosis of laryngeal nerve injuries. Arch Phys Med Rehabil 1990;71:587-590.

6. Faaborg-Andersen, K. Electromyographic investigation of intrinsic laryngeal muscles in humans. Acta Physiol Scand [Suppl] 1957(suppl 140).

7. Haglund S, Knutsson E, Martensson A. An electromyographic analysis of idiopathic vocal cord paresis. Acta Otolarygnol 1972;74:265-270.

8. Hirano M, Ohala J. Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. J Speech Hear Res 1969;12:362-73.

9. Kimura J. Electrodiagnosis in diseases of nerve and muscle: principles and practice. 2nd ed. Philadelphia: F. A. Davis Co., 1989:227-248.

10. Seddon HJ. Three types of nerve injury. Brain 1943;66:238-288.

11. Hiroto I, Hirano M, Tomita H. Electromyographic investigation of human vocal cord paralysis. Ann Otol Rhinol Laryngol 1968;77:296-304.

12. Crumley RL. Laryngeal synkinesis: its significance to the otolaryngologist. Ann Otol Rhinol Laryngol 1989;98:87-92.

13. Thumfart W. Electrodiagnosis of the larynx and related technics. Acta Oto Rhino Laryngol 1986;40:358-76.

NCVS Status and Progress Report - 6
May, 1994, 73-75

The Portable Telemetry System for Electromyography

Young B. Min, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Erich S. Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Eileen M. Finnegan, M.A.
Department of Speech Pathology and Audiology, The University of Iowa
Timothy M. McCulloch, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Henry T. Hoffman, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa

Introduction

In 1990, the National Institutes of Health Consensus Development Conference on Clinical Use of Botulinum Toxin (Botox) recommended Botox as safe and effective for the treatment of strabismus, blepharospasm, hemifacial spasm, adductor spasmodic dysphonia, jaw-closing oromanibular dystonia, and cervical dystonia.[1] In treatment of adductor spasmodic dysphonia, four approaches for Botox injection to thyroarytenoid muscles were developed; they are transoral with indirect laryngoscopic guidance,[2] percutaneous with flexible nasopharyngeal endoscopic guidance,[3] endoscopic through the flexible nasal endoscope (K. Rhew, personal communication with a co-author, 1993) and percutaneous with electromyography (EMG) guidance.[4] EMG-guided percutaneous injection is well tolerated by patients and takes only a few minutes to administer. However, the EMG- guided percutaneous injection has not come into common use by many otolaryngologists because of the large, immobile, and expensive equipment generally used to detect EMG activity. In addition, some expertise in EMG is needed. Hence, these injections are usually performed in large university hospital laboratories with EMG equipment which is not available to most otolaryngologists. In most settings, standard EMG equipment requires a special space outside the general office exam room. Patients are thus moved from one location to another to perform a 3 minute Botox injection.

Currently, there are three commercially available portable EMG-detecting systems, which can be modified for EMG-guided Botox injection: 1) Brackman EMG system (WR Medical Electronics Co., Stillwater, MN), 2) Neurosign 100 (Smith & Nephew Richards Inc., Memphis, TN), and 3) Nerve Integrity Monitor-2 (Xomed-Treace Inc., Jacksonville, FL).

These units can be used for Botox injection with the minor modification of a lead with a clip to attach to the Teflon-coated Botox needle. However, they have visual and audio monitors and are designed for intraoperative facial and other cranial nerve stimulation, as well as EMG detection. With this additional capability, the cost of these units are $4,600.00, $9,775.00, and $17,100.00, respectively. In addition, some of these units require expertise in visual and/or acoustic identification of EMG signal and assistance for utilization.

An inexpensive, easy to use, and portable system specifically for EMG-guided Botox injections was developed for use in our hospital clinic. This system employs a single-channel-battery operated EMG amplifier for detection of EMG activity which is then FM transmitted to an inexpensive FM receiver for monitoring purposes.

Equipment

Our system consists of two main components (Fig. 1). One component is a compact single-channel EMG differential amplifier and FM transmitter. The second component is an FM radio receiver with speakers (a "boom box"). Two surface electrodes are affixed to the patient's forehead; one functions as a reference and the other as a differential to the injection needle electrode. Generic electrocardiograph electrode pads are used to attach the surface electrode leads. The needle electrode is completed by attaching the lead to a clip designed for use with the Teflon-coated hollow bore needles.

The amplifier was designed by one of the co-authors and constructed by the University of Iowa Bioengineering Laboratory. The differential amplifier has a gain of 10,000 with a frequency response of 30-5k Hz which is adequate to detect EMG activity. It contains a 60 Hz notch filter to attenuate 60 Hz artifact, a modulation control knob, and a transmitter with a flexible wire antenna.

An improved model will also have a test button which will send a test signal into the amplifier and transmit a simulated EMG signal. This will provide information to the user regarding EMG acoustic signal characteristics and will verify the functional integrity of the system. This current unit measures 7.5 X 13 X 18.5 cm and weighs 0.7 kg. The power is supplied by two 9 volt batteries.

The electrical activity detected in the muscle by the monopolar injection needle is differentially amplified, and the signal is transmitted via radio waves to the second component, a portable FM receiver. The FM radio receiver is tuned to the output of the telemetry unit, and consists of an amplifier, a volume control, an equalizer/filter, and speakers. The acoustic signal of the EMG activity broadcasts over the speakers of the FM receiver. This unit can be purchased from any commercial audio store. The FM receiver we purchased has a variable filter, commonly known as an equalizer, which can be adjusted to accentuate the high frequency characteristics of the EMG signal, the "crisp", "crackle" quality that we prefer in listening to EMG activity. This system also has an important safety feature. Since the amplifier is battery operated and the system involves radio transmission of the signal, the patient is electrically isolated from earth ground. This feature also greatly increases the rejection of 60 Hz artifact.

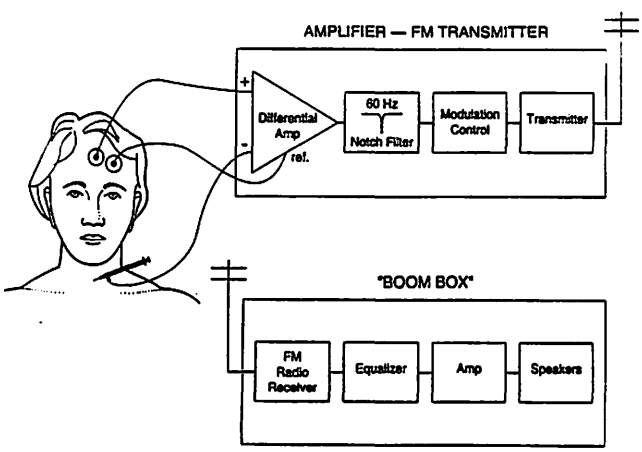


Figure 1 . Schematic diagram of the portable telemetry system.

Validity

To demonstrate the validity of the telemetry system, the activity from the dorsal interosseus muscle was simultaneously recorded with the standard EMG equipment and the telemetry system. The motor unit potential waveforms were visualized by digitizing the EMG signal into a computer with DATAQ Instruments CODAS (computer-based oscillograph and data acquisition system) software. The motor unit potentials were comparable and identifiable (Fig. 2). However, there is a slight difference in the waveform characteristics because of the effect of the equalizer. The filter setting for amplification of the high frequencies produces a subtle modification, so that the telemetry EMG waveform contains the derivative of higher frequency components of the standard EMG equipment waveform.

Discussion

During the past 9 months we have performed more than 100 Botox injections using the telemetry system on patients with dystonia (spasmodic dysphonia, oromandibular dystonia, and hemifacial spasm).

The telemetry system is specifically designed for EMG-guided percutaneous Botox injections and has numerous advantages: easy to use, high quality acoustic output, portability, increased resistance to 60 Hz artifact,

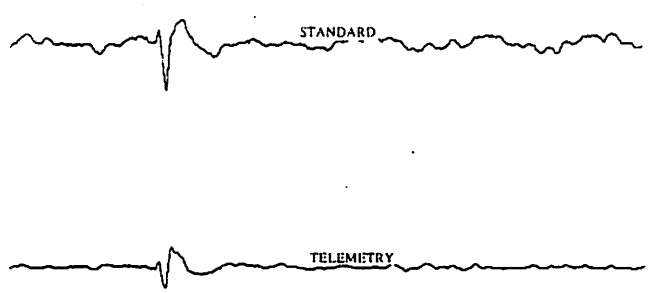


Figure 2. The motor unit potential waveforms recorded from the standard EMG equipment and the telemetry system.

and low cost. Unlike the commercially available EMG detectors with numerous controls, this system is easy to use with minimal number of knobs. Since the controls are a part of the radio, most users are already familiar with the them. Once the filter/equalizer controls are adjusted to broadcast the most recognizable acoustic quality of EMG activity to the user, the volume control is the only adjustment modified during the injection. Therefore, a single user can perform the Botox injection without assistance in operating the equipment.

Since a visual monitor is not needed to identify the EMG activity, the user can quickly guide the Botox injection needle to the target muscle solely by listening to the acoustic signal without stopping to view the monitor. Since the FM receiver with speakers is built specifically for a high quality audio output, as compared with the three commercially available systems which are not designed for sound quality, the telemetry system is far superior for distinctive EMG sound characteristics.

Another advantage of this system is the portability of the two units. The size and weight of the FM receiver is dependent on the model purchased. The compact size of the telemetry unit allows it to be used in a small clinic room where the unit is placed on a standard otolaryngology exam table. Hence, the pre- and postoperative examinations, as well as the Botox injection, are performed in the same room.

The telemetry system is resistant to the 60 Hz artifact. Its portable size and its reduced sensitivity to artifacts make this system ideal for use in an operating room.

The telemetry system is also low in cost. The cost of the prototype amplifier is approximately $1,500.00. The price of the FM receiver with speaker is dependent on which model one chooses to purchase, but our model cost only $60.00. The cost of the three leads is $10.00 each. Thus, the total cost of the telemetry system is approximately $1,590.00.[1]

During the time this telemetry system has been in use in our clinic the following problems were encountered: 1) inability to check the integrity of the battery and 2) loose component in the amplifier. However with the addition of a test button to check the integrity of the system in the improved model and attaching all components more securely, these problem should be more easily detected.

Although the telemetry system does not replace the direct system for recording or analyzing EMG activities, its many advantages make it easier for physicians to perform EMG-guided Botox injections.

[1] The custom-built amplifier is available for purchase through the University of Iowa Bioengineering Laboratory.

Acknowledgment

Supported by the grant #P60 DC00976 from the National Institute on Deafness and Other Communication Disorders.

References

1. Consensus Statement. Clinical use of botulinum toxin. NIH Consensus Dev Conf Consens Statement 1990;8:1-20.

2. Ford CN, Bless DM, Lowery JD. Indirect laryngoscopic approach for injection of botulinum toxin in spasmodic dysphonia. Otolaryngol Head Neck Surg 1990;103:752-8.

3. Green DC, Berke GS, Ward PH, Gerratt BR. Point-touch technique of botulinum toxin injection for the treatment of spasmodic dysphonia. Ann Otol Rhinol Laryngol 1992;101:883-887.

4. Blitzer A, Brin MF, Fahn S, Lovelace RE. Localized injections of botulinum toxin for the treatment of focal laryngeal dystonia (spastic dysphonia). Laryngoscope 1988;98:193-7.

NCVS Status and Progress Report - 6
May, 1994, 77-80

Conversion of a Head-Mounted Microphone Signal Into Calibrated SPL Units

William S. Winholtz, A.A.S.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

A method is reviewed for conversion of a microphone signal into calibrated SPL units. The method follows ANSI standard S1.4 (SPL meters) published by the Acoustical Society of America, and requires an accurate SPL meter and a stable calibration sound source for conversion. Accuracy and validation data from test signals and human phonation are provided. The results indicate that under typical speech conditions, an absolute accuracy of plus or minus 1.6 dB (Type 1 SPL meter) can be obtained with a miniature head-mounted microphone.

Introduction

The SPL meter is a recognized standard instrument for quantifying the absolute sound pressure level of a sung or spoken utterance. For high ambient voice, excessive head movement, or low intensity sound production, however, placement of the SPL meter at close and constant distance to the mouth can be a problem. For this reason, many investigators have gone to a miniature head-mounted microphone to convert recorded signals into calibrated SPL units. A head-mounted microphone can maintain a close (and constant) distance to the source and can be miniaturized to reduce weight and facial obstruction. The method described here reviews and documents the procedure for SPL calibration of a popular miniature head-mounted microphone.

Conversion Method

The conversion of microphone pressure to SPL is done with a computer program following ANSI standard S1.4, which gives definitions and characteristics for three different grades of sound level meters. The method requires that each time the gain in the signal conditioning equipment changes a calibration tone be recorded, along with its corresponding SPL reading. For ease of calibration, a sound source is positioned as close to the mouth as possible to produce a constant tone lasting a minimum of one second in duration (Figure 1). The sound source has two preset levels, corresponding to high and low subject SPL levels. It is justifiable to use a sinusoidal sound source for calibration because such a source is minimally affected by sampling and conditioning electronics. In this study, a portable sound source (Wintronix WSS-5) was used. (It has been brought to our attention that occasionally an electrolarynx has been used as a calibration source. It is perhaps unwise to use such a source because of its variable duty ratio when not in contact with tissue.)

A contaminating factor in calibration can be acoustic noise below 60 Hz. Some acoustically treated rooms do not eliminate structural noise effectively. A linear phase 60 Hz High Pass filter is recommended as part of the conditioning circuitry. (An attenuation of 24 dB/octave has been needed in our case, but this may vary with building design).

Two sampled data files are created in the computer, one of the calibration tone and the other of the vocal utterance. The program analyzes the data file for the calibration tone to establish an error difference between the reading taken from the meter and a value calculated from the calibration data (Figure 2). Inverse square loss in dB for the distance between the meter and the source is included. The difference is then added to the vocal signal for absolute calibration. Once the relationship is established, calibration of entire signal chain is accounted for.

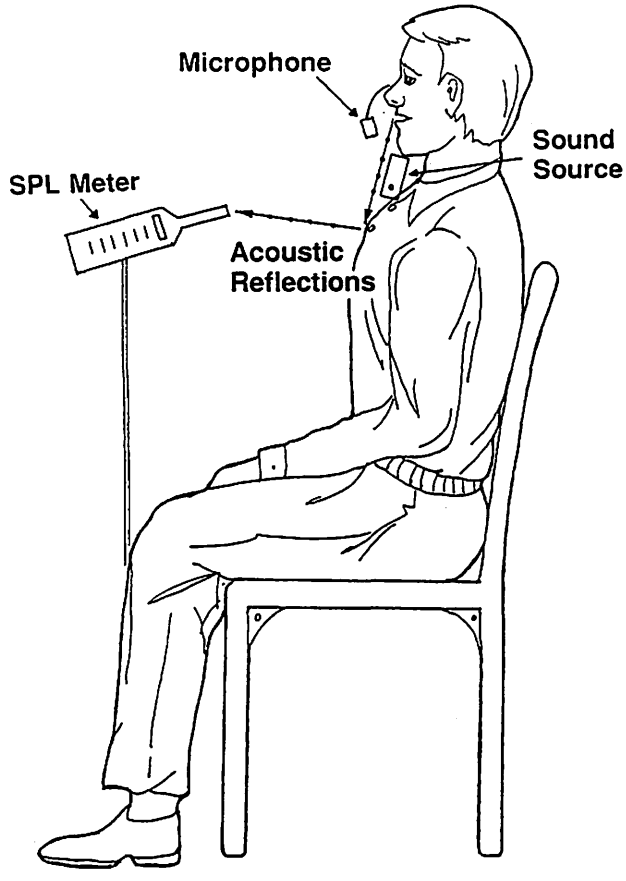


Figure 1. Diagram of subject with placements of microphone, SPL meter, and sound source.

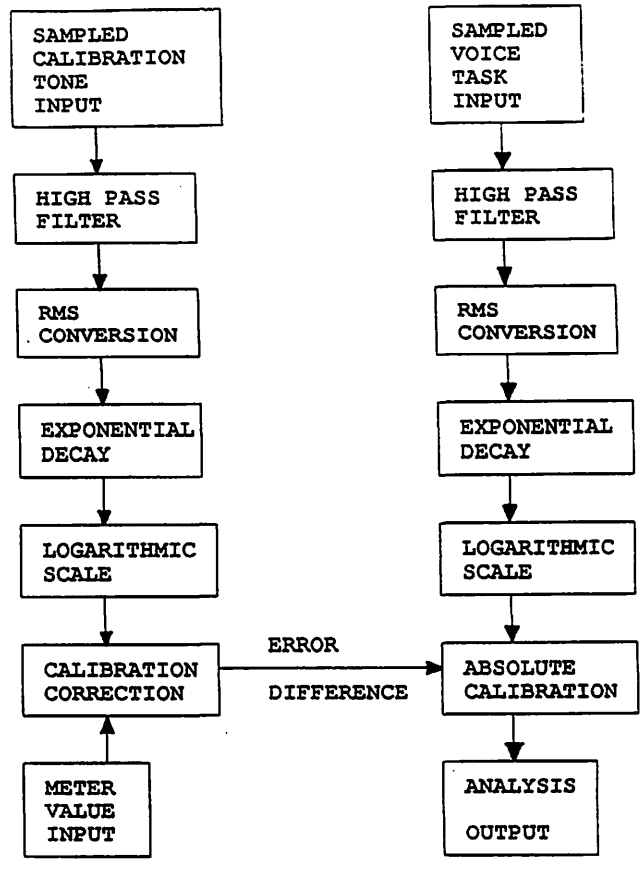


Figure 2. Block diagram of computer program for conversion of microphone signal into calibrated SPL units.

Some correction may be required to account for transmission path errors related to unwanted acoustic reflections from the subject (Figure 1). Of the two critical paths for calibration, to the microphone and to the meter, the meter path is the most susceptible because of the greater distance. In our experiments, errors of several dBs were produced by varying acoustic reflections that depend on the physical size of the subject. These errors can be compensated for by taking a measurement at the desired operating distance without any reflective objects in the path; this measurement is then subtracted from the average values taken under normal conditions.

Next a root mean square (RMS) calculation is performed on the sampled data. The ANSI standard requires a 125 ms time constant for the fast response mode, which we used as the default mode here; this amounts to a window size of 625 points at a sampling frequency of 5000 Hz. After the RMS operation, an exponential decay is imposed on the signal. The ANSI standard suggests a fixed decay time of .5 s per 10 dB

Next, the data are converted to the logarithmic (base 10) scale and the corrected calibration value is added to the vocal utterance. At this stage the data are in calibrated SPL units and ready for statistical calculations, or plotting routines to display the data.

Validation

The accuracy of this method was verified under static and dynamic conditions. All recordings were made in an acoustically treated sound booth (IAC, 10.5 ft x 11.5 ft x 8 ft) with an ambient noise level of 50 dB. A direct comparison was made between the program output and the DC output of the SPL meter (B&K 2230 type 1) for the same input stimulus. The meter was set for RMS detection, FAST response mode, and C filter response. An AKG C-410 head mounted microphone was used to transduce the acoustic signals for the program. (It is important to use a microphone with linear amplitude response and a flat frequency response down to approximately 100 Hz). Prior to sampling (DCS-200, two channel, 5 kHz each), the microphone signal was preamplified (ATI M-1000), and further amplified (Tektronix 502). The DC output of the meter was also amplified (Tektronix 502).

Table 1.
Comparison of SPL meter display versus conversion program output for sinewaves of 400Hz and five levels.

TEST SIGNAL	ADJUSTED METER DISPLAY	PROGRAM OUTPUT	ERROR	MEAN ERROR	ERROR STANDARD DEVIATION
1	81.5	81.6	+0.1		
2	91.8	91.0	-0.8		
3	101.8	101.2	-0.6	0.0	0.69
4	112.0	112.8	+0.8		
5	122.0	122.5	+0.5		

For the static test, a constant level sinewave of 400 Hz was produced with a loudspeaker placed at 30 cm (Auratone 5C) at five different target SPL levels, 60 dB, 70 dB, 80 dB, 90 dB, and 100 dB to correspond with levels found in human phonation. In addition, four subjects, two male and two female, were asked to sustain the vowel /a/ at three levels, low, medium, and high. They were seated in a chair with a headrest and instructed to maintain a constant head position. Mouth to microphone distance was 8 cm and mouth to meter distance was 30 cm.

For the dynamic test, two subjects (one male and one female) were asked to read the rainbow passage at two loudness levels, low and medium. A modified computer program was used to simultaneously plot and calculate the mean levels for the dynamic task. For this test the meter was positioned at the same distance (8 cm) as the microphone to the source to allow direct comparison of the outputs.

Table 2.
Comparison of SPL meter output versus conversion program output for human sustained vowels.

SUBJECT	ADJUSTED METER OUTPUT	PROGRAM OUTPUT	ERROR	MEAN ERROR	ERROR STANDARD DEVIATION
		SESSION 1			
F1	91.0	91.7	+0.7		
TRIAL 1	110.8	111.2	+0.4		
	121.3	119.6	-1.7		
F1	89.8	91.3	+1.5		
TRIAL 2	112.9	113.2	+0.3		
	120.7	121.1	+0.4		
				+0.56	0.9
M1	88.5	90.2	+1.7		
TRIAL 1	108.9	109.5	+0.6		
	120.8	121.9	+1.1		
M1	89.2	90.7	+1.5		
TRIAL 2	110.6	110.4	-0.2		
	121.1	121.5	+0.4		
		SESSION 2			
F2	92.5	94.0	+1.5		
TRIAL 1	101.1	101.3	+0.2		
	112.1	110.4	-1.7		
F2	95.2	94.0	-1.2		
TRIAL 2	104.5	103.6	+0.2		
	111.7	111.3	-0.4		
				-0.55	0.9
M2	92.5	92.2	-0.3		
TRIAL 1	110.2	108.6	-1.6		
	120.5	119.6	-0.9		
M2	99.2	98.4	-0.8		
TRIAL 2	107.9	107.6	-0.3		
	121.0	120.8	-0.1		

Results

Table 1 presents the data for the static test using sinewaves. As expected, the mean error was low because of the highly stable input signal, 0.00 dB with a standard deviation of .69 dB. Table 2 shows the data for the static test using sustained vowels. These data are more variable and indicated a basic program accuracy of ±0.9 dB. In addition to this system error, a meter error of ±.6 dB must be included. This error was determined by comparing the mean error for session 1 versus session 2 on two different recording days showing the drift in meter accuracy. Therefore, using a type 1 SPL meter (specified accuracy of ±.7 dB in the ANSI standard) in conjunction with this program one can expect an absolute accuracy of ±1.6 dB. Using a type 0 meter (accuracy of ±.4 dB) would improve the overall accuracy slightly.

Figures 3 and 4 (next page) illustrate the results of the dynamic test. Figure 3 is for the male subject at low loudness level and Figure 4 is for the female subject at low loudness level. The top trace is the output from the program and the bottom trace is the output from the meter for the same stimulus. The horizontal line is an arbitrary cutoff point such that data points below the line are not included in the mean value. This exclusion is justified because of the larger dynamic range of the program over the meter, as illustrated by the lower noise floor in the top trace. The increase comes from the shorter source to microphone distance of the program; if unaccounted for, this noise floor would bias the mean calculation lower.

The subtle differences in the waveforms may be due in part to a loss of energy above 2.5 KHz that is filtered out because of the 5 kHz sampling rate. However, this waveform difference has a small affect on the SPL values in Table 3. In the table, the data are presented for two subjects at two reading levels. The mean error for the dynamic test was low, -0.05 dB with a standard deviation of .74 dB.

Table 3.
Comparison of the means for SPL meter output versus conversion program output for two subjects reading approximately 10 secondsof the rainbow passage at low and medium levels.

SUBJECT	METER OUTPUT	PROGRAM OUTPUT	ERROR	MEAN ERROR	ERROR STANDARD DEVIATION
F3					
LOW	68.6	68.5	-0.1		
MEDIUM	77.8	76.9	-0.9		
				-0.05	0.74
M1					
LOW	73.2	74.1	+0.9		
MEDIUM	83.2	83.0	-0.1		

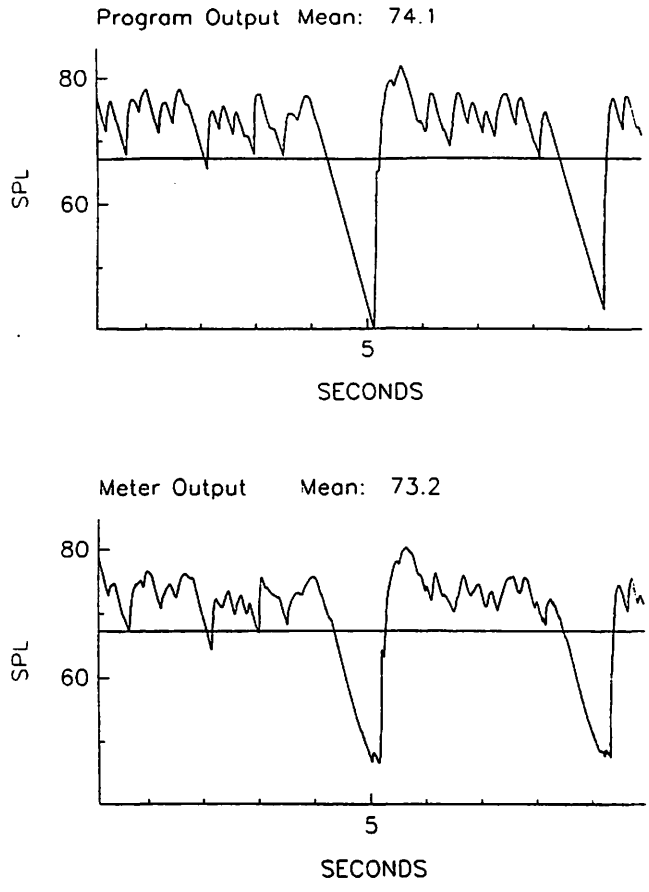


Figure 3. Plot of program output and meter output for male subject low level reading of the rainbow passage.

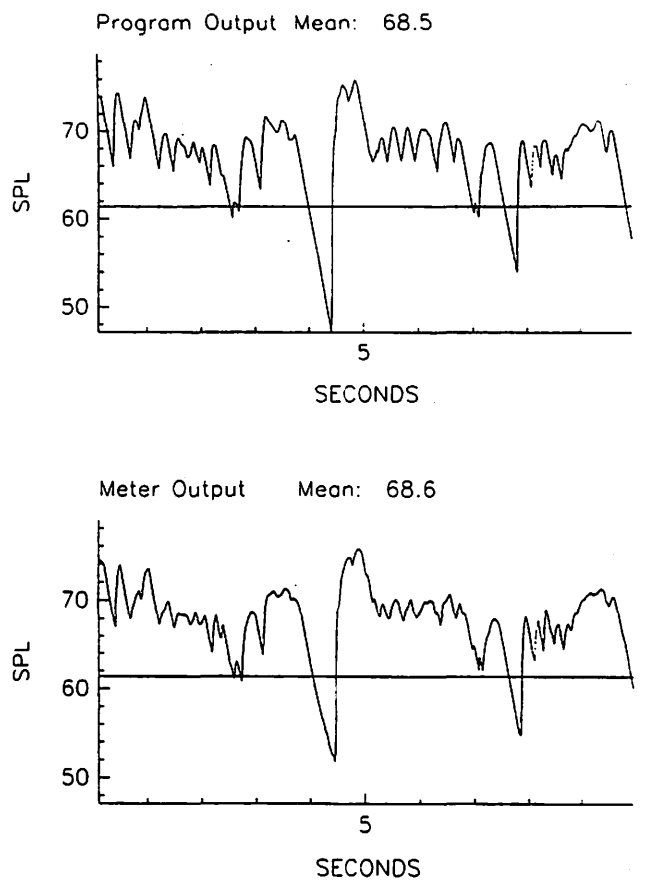


Figure 4. Plot of program output and meter output for female subject low level reading of the rainbow passage.

Conclusion

A method for converting a microphone signal into calibrated SPL units has been reviewed. Under typical speech conditions, the absolute accuracy is ±1.6 dB, but can improve with the accuracy of the SPL meter. Potential drawbacks with this method are an increase in complexity and the need for some experimentation to establish consistent conversion. If the gain is inadvertently changed in the system post calibration, or the calibration value is in error, then accurate conversion cannot take place. Furthermore, it is important to account for SPL errors resulting from acoustic reflections of objects in the sound field between the source and SPL meter. Under identical operating conditions, errors of several dB can result from differences in the size of one subject to another.

When making SPL measures under favorable acoustic conditions, such as in an acoustically treated sound booth, or when head motion needs to be considered, the method presented here can be used to obtain a higher accuracy for some SPL measures.

Acknowledgment

This study was supported by a grant from the National Institutes of Health, grant Nos. R01 DC00387-04 and R01 DC01150. The authors thank CG Guo and Darrell Wong for their assistance.

References

ANSI standard 1.4. Specification for Sound Level Meters. American Institude of Physics. Rev 1983

Bruel & Kjaer (1983). Precision integrating sound level meter type 2330. Insruction manual.

Titze, I.R., & Winholtz, W.S. Miniature Head Mount Microphone for Acoustic Analysis. Journal of Speech and Hearing Research, (in review).

NCVS Status and Progress Report - 6
May, 1994, 81-88

Acoustic Analysis of the Pre- and Post-Therapy Fluent Speech of Children Who Stutter: Preliminary Observations

Hyun-Sub Sim, M.S.
Department of Speech Pathology and Audiology, The University of Iowa
Patricia M. Zebrowski, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Previous studies of pre- and post-therapy fluent speech of individuals who stutter have reported longer durations of voiced speech segments and shorter durations of silent intervals, unvoiced speech segments/subsegments. These post-therapy changes in temporal acoustic durations may be associated with by-products of such "target behaviors" as prolonged syllables and reduced articulatory rate rather than with the speech motor control processes underlying fluency enhancement. The current study aims to uncover acoustic events of post-therapy fluent speech which are related to speech production behaviors underlying fluency enhancement. For this purpose, rather than using prolonged speech, slow initiation of voicing, or a reduced speech rate to facilitate fluency, the children who stutter in the current study received "Normal Talking Process" therapy (Williams, 1979), which teaches both purposeful change of individual speech parameters to increase fluency and behavioral awareness of fluent ("easy") versus stuttered ("hard") speech. Further, in order to improve interpretations of such an analysis, pre- and post-therapy fluent speech obtained from stuttering children who stutter was compared to similar speech samples produced by normal children.

This study compared absolute durations of temporal acoustic parameters at the subsegmental level of fluent speech produced by stuttering children (N=4) before and after "Normal Talking Process" therapy. These measures were then compared to an age and gender-matched control group of normally fluent children. The findings indicated that voice onset time (VOT) and CV (consonant-vowel) transition duration decreased significantly after therapy. Children who stutter produced significantly longer vowel durations than the control group both before and after therapy. The results was discussed in terms of differential aspects of this treatment.

A number of acoustic studies have compared the pre- and post-therapy fluent speech of both children and adults who stutter. The primary motivation for such investigations is to ascertain whether decreases in stuttering as a result of intervention are associated with alterations in various temporal speech events as represented within the acoustic waveform. According to Metz, Samar, & Sacco (1983), observed changes in the acoustic properties of fluent speech following stuttering therapy may be important indices of "...changes in the operation of motor control process that underlie fluency enhancement, whereas others may be systematic by-products of the particular fluency-enhancing condition. (p. 531)"

Results from previous work have shown that following therapy, the fluent speech of adults who stutter is characterized by both longer vowel and phrase durations and increased proportion of voicing during closure for stop consonants (Mallard & Westbrook, 1985; Metz, Onufrak, & Ogburn, 1979; Metz, Samar, & Sacco, 1983), increased proportions of voiced segments (Franken, 1987), decreased periods of silence during the intervocalic interval for voiced stops (Metz, Samar, & Sacco, 1983), and longer voice onset times (VOT) (Metz, Schiavetti, & Sacco, 1990). The relatively few studies which have examined the pre- and post-therapy fluency of children who stutter have reported increased vowel durations and percent vocalized time

(Robb, Lybolt, & Price, 1985), decreased vowel durations and VOT (Zebrowski, 1991) or no significant change in any of these parameters (Onslow, Doorn, & Newman, 1992). Although some discrepancy among findings exists, the majority of results from these studies suggest that the post-treatment fluency of individuals who stutter can be characterized by a general pattern of increased duration of voiced speech segments or subsegments, and decreased duration of silent intervals or unvoiced speech segments or subsegments.

With the prior conclusion in mind, an important question remains: That is, as Metz et al. (1983) suggested, are the increases in the proportion of voicing, and decreases in the proportion of unvoiced segments associated with fluent speech necessary conditions for fluency, or are they merely residual effects of the specific strategies taught by various stuttering treatment approaches (Guitar, 1991). For example, in most of the previously cited studies, participating subjects received stuttering therapy which emphasized either (1) "stretched" or prolonged syllables (e.g. Franken, 1987; Onslow et al., 1992), (2) reduced articulatory rate (e.g. Metz et al., 1979), (3) slow initiation of phonation and prolonged voicing during speech (e.g. Metz et al., 1979; Robb et al., 1985) or some combination of these strategies. Further, it seems reasonable that techniques or program "target" behaviors such as these result in fluent speech characterized by, among other things, increased vowel and syllable duration, and decreased speech rate.

Presently, there is a need to continue to try and uncover those features of post-treatment fluent speech which represent certain speech production behaviors underlying fluency, as opposed to features which are by-products of specific fluency-enhancing strategies taught in therapy. One reasonable method for attempting to answer this question is to analyze the post-therapy fluent speech of individuals who stutter who have not received therapy which teaches the use of lengthened or prolonged syllables ("prolonged-speech"), slow initiation of voicing, or reduced speech rate. Further, in order to improve interpretation of such an analysis, pre- and post-therapy fluency obtained from these individuals should be compared to similar samples produced by normal speakers. With these considerations in mind, the purpose of the present study was to compare pre- and post-therapy measures of (1) Stop-gap duration, (2) Voice onset time (VOT), (3) Consonant-vowel (CV) transition duration, (4) Vowel duration, and (5) Consonant-vowel (CV) segment duration (stop-gap duration + CV transition duration + vowel duration) in the fluent speech of children who stutter after completing an intensive stuttering therapy program. A second purpose of this study was to compare the pre- and post-therapy perceptually fluent speech of these children with that of age and gender-matched normally fluent children.

Method

Subjects

Subjects for this study were four boys, with a mean age of 12:11 (years:months; range: 11:5 to 14:0) who were enrolled in a six- week intensive residential stuttering therapy program at the University of Iowa. A control group consisting of four age and gender-matched nonstuttering boys was included (M = 12:1; range: 11:0 to 14:2).

The boys who stuttered were referred to the intensive program by both school speech-language pathologists and their parents. Each child had a history of stuttering therapy which was provided in the schools, although in varying amounts. Further, no subject displayed additional speech or language problems, and all subjects had normal hearing. For each subject the presenting complaint upon referral for intensive treatment was that prior therapy had been unsuccessful.

On the first and last day of therapy, each stuttering subject was videotaped while he engaged in conversation with a graduate-level clinician in speech-language pathology. These videotaped conversational samples were subsequently analyzed in order to examine (1) the mean frequency of overall speech disfluency in 100 words, as well as (2) the three most frequently produced disfluency types, and (3) mean duration of within-word (stuttered) disfluencies. Further, the Stuttering Severity Instrument for Children and Adults (SSI) (Riley, 1980) was administered to each stuttering subject before and after therapy. The four nonstuttering subjects were recruited from the community through newspaper advertisement. All subjects were paid volunteers who were naive as to the purpose of the study.

Description of Treatment

Treatment procedures generally followed those described by Williams (1957; 1979), and elaborated upon by Conture (1990), which emphasize the processes underlying "normal talking." In this approach, the emotional, psychosocial, and speech production behaviors associated with an individual's fluent speech are taught, and compared to those which underlie and perpetuate his stuttered disruptions. According to Williams (1979), the main goal of the "normal talking" therapy model is to teach the person who stutters how to identify and attend to both the things he is doing which interfere with talking, as well as those things he is doing which facilitate speaking.

In order to accomplish this aim, therapy involved (1) discussion and identification of the motoric processes underlying both fluent speech and within-word (stuttered) disfluencies (e.g. sound/syllable repetitions, sound prolongations), (2) discussion of the main parameters of speech production, including: airflow, physical tension, movement, timing, and sound, (3) behavioral awareness and

comparison of fluent ("homebase" or "easy") speech versus stuttered ("hard") speech, (4) "off-line" and "on-line" identification and description of the behaviors associated with instances of stuttered speech, (5) "on-line" changing of stuttered speech disfluencies to "homebase" or "easy" (fluent) speech, (6) using anticipatory cues to "stay at homebase", and (7) identifying relationships between stuttering and different emotional states, speaking situations and partners, and practicing strategies for "staying at homebase" when talking under these various conditions (Zebrowski, 1991).

Data Collection

Data were collected from the subjects who stuttered on the first and last day of therapy. Data from the nonstuttering subjects were collected in the interim between the first and last day of therapy for the stuttering subjects. Because this acoustic study was part of a larger investigation of articulator movement and interarticulator coordination, all subjects' lower lip and jaw movements were simultaneously transduced via strain gauge while recording speech for subsequent acoustic analysis. For the purposes of the present study, each subject was asked to repeat four different test words ten times each within the carrier phrase "That's a (test word) a day." Test words consisted of consonant-vowel-consonant (CVC) or consonant-vowel (CV) sequences containing word-initial /p/ and /b/ followed by the vowels /a/ and /i/ (for example, pop, pea, bob, and bee). These test words and phrases were part of a larger repertoire of syllables, phrases and sentences which the subjects were asked to repeat. All speech samples were recorded on a Sony (PC-108) digital audio tape recorder.

Data Analysis

All subjects' perceptually fluent productions of word initial /p/ and /b/ were acoustically analyzed. These included the pre- and post-therapy fluent productions of the children who stutter, and those of their nonstuttering peers. Only those utterances in which the test word and each word in the carrier sentence were judged to be fluent by both the first and second author were used in this study. Wide-band spectrograms (CSpeech, Version 3.1) displayed on a computer monitor (Gateway 2000) were made of each of the fluent test words, and cursors were used to delineate onset and offset points for measurements of stop-gap duration, voice onset time (VOT), vowel duration, and CV transition duration, along with the duration of the initial consonant-vowel (CV) segment in each of the fluent test words. Operational definitions of these measures have been reported elsewhere (Zebrowski, Conture, & Cudahy, 1985; Zebrowski, 1991), and are described below.

Stop-gap duration. Measured from cessation of acoustic energy for /ə/ immediately preceding the test word (e.g. "a pop") to the onset of acoustic energy ("burst") associated with the release of oral closure for either /p/ or /b/.

Voice onset time (VOT). Measured from the beginning of acoustic energy associated with supraglottal release for /p/ or /b/, to the onset of voicing (appearing as the first regularly appearing vertical striation in the region of the second formant).

CV transition duration. Measured from the onset of supraglottal release for the stop-plosive to the onset of the steady-state portion of the vowel at the level of the second formant.

Vowel duration. Measured from the offset of the CV transition to the point where the second formant ceases its horizontal orientation.

CV segment duration. Measured from the offset of acoustic energy for /ə/ preceding the test word, to the point where the second formant in the vowel of the test word ceases its horizontal orientation (i.e. end of steady-state portion). This measure reflects the sum of the total duration for stop-gap, VOT, CV transition and vowel duration combined.

Reliability

To assess intrajudge reliability, the first author remeasured four randomly selected test words from each of the nonstuttering subjects, and eight from each of the children who stutter; four pre-therapy and four post-therapy words. There were 48 remeasured words in total. Twenty-eight of these 48 words consisted of /p/ in the initial position, and 20 consisted of /b/ in the initial position. The mean intrajudge measurement error for each of the five acoustic measures for /p/ and /b/ respectively was as follows: (1) stop-gap duration: 3.8 msec and 5.3 msec; (2) VOT: 3.8 msec and 3.5 msec; (3) CV transition duration: 5.4 msec and 6.4 msec; (4) vowel duration: 3.8 msec and 3.5 msec; (5) CV segment duration: 6.2 msec and 5.0 msec.

For interjudge reliability, a doctoral student in speech-language pathology familiar with methods of acoustic analyses separately measured this same subset of 48 test words. The average interjudge measurement error for each of the five acoustic measures for /p/ and /b/ respectively was as follows: (1) stop-gap duration: 2.3 msec and 2.2 msec; (2) VOT: 1.4 msec and 1.6 msec; (3) CV transition duration: 4.3 msec and 3.9 msec; (4) vowel duration: 2.4 msec and 3.3 msec; (5) CV segment duration: 4.5 msec and 3.7 msec.

Results

Pre- vs Post-Therapy Comparisons of Stuttering and Disfluency

Table 1 provides a comparison of pre- and post-therapy percentages of speech disfluency. Following

Table 1.
Pre- and Post-Therapy Measures of Speech Disfluency

	Pre-therapy				Post-therapy			
	1	2	3	4	1	2	3	4
mean percentage of speech disfluency per 100 words	37%	35%	12%	26%	28%	31%	8%	15%
*most frequently produced speech disfluency type per 100 words	ISP	ISP	ASP	ISP	ISP	ASP	INT	ISP
mean duration of disfluencies (seconds)	.68	1.5	.43	.94	.46	1.0	.35	.59
** Stuttering Severity Instrument	SV	VS	MO	VS	MO	VS	MI	SV

* ISP=Inaudible prolongation ASP=Audible sound prolongation
SSR=Sound/Syllable repetition INT=Interjection

** VS=Very Severe SV=Severe MO=Moderate MI=Mild

therapy, each subject exhibited a decrease in the percentage of speech disfluency and mean duration of disfluency in 100 words of conversational speech. However, there were individual variations in the most-frequently observed within-word disfluency type. For example, while no change in most frequent disfluency type for subjects 1 and 4 was found, the most frequent disfluency type for subject 2 changed from inaudible sound prolongation to audible sound prolongation. Also, for subject 3, the most frequent disfluency type changed from audible sound prolongations to interjections.

Pre- vs Post-Therapy Comparisons of Temporal Acoustic Measures

Figures 1 and 2 show the means and standard deviations of all measures obtained from the perceptually fluent productions of word-initial /p/ (Figure 1) and /b/ (Figure 2) of the nonstuttering subjects, as well as means of pre- and post- therapy measures produced by the children who stutter. For the data from the stuttering subjects, separate one-way repeated ANOVA's were conducted for each acoustic measure to ascertain any pre- vs post-therapy differences. Results of this analysis revealed more post-therapy differences for /p/, showing a significant decrease in four of the five measures following treatment. Specifically, as shown in Figure 1, there was a significant decrease in stop-gap duration following therapy, ($F(1, 106) = 6.46$, $p < .01$), as well as significant decreases in VOT ($F(1, 105) = 5.19$, $p < .02$), CV transition duration ($F(1, 106) = 7.97$, $p < 0.005$), and CV segment duration for /p/ ($F(1, 106) = 5.77$, $p < .02$).

While Figure 2 shows pre vs post-therapy differences in all temporal measures for /b/, only the post-therapy decrease in VOT was significant ($F(1, 93) = 9.57$, $p < .003$), while a decrease in CV transition duration approached, but did not reach, significance ($F(1, 91) = 2.87$, $p < .09$).

Comparison of Temporal Acoustic Measures Obtained from Stuttering and Nonstuttering Children

Fluent speech of nonstuttering children and children who stutter pre-therapy. Recall that a second goal of this study was to compare both the pre- and post-therapy fluent speech of the children who stutter with that of their nonstuttering counterparts. Again, Figures 1 and 2 provide means and standard deviations for the five acoustic measures obtained from the fluent productions of both groups of children for word-initial /p/ (Figure 1) and /b/ (Figure 2).

Results of separate Wilcoxon-Mann-Whitney analyses, with adjusted alpha levels of .01 for each of five comparisons (.05 for all five comparisons as a family for each sound) revealed that prior to therapy, the children who stutter produced fluent speech subsegments and segments which were longer in duration than those of their normally fluent peers. Specifically, the children who stutter exhibited significantly longer vowel durations ($p < .003$), CV transition durations ($p < .004$), and CV segment durations

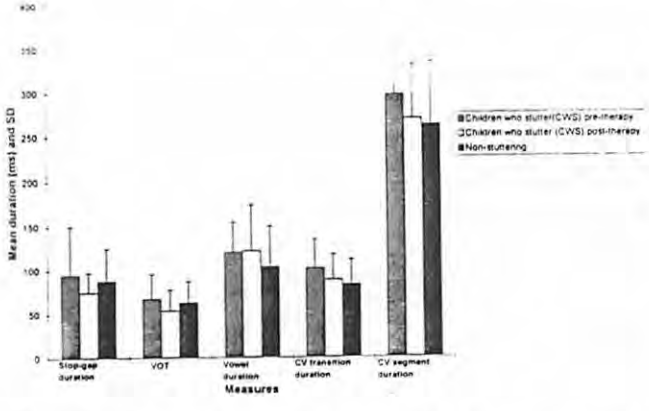


Figure 1. Means and standard deviations of five temporal acoustic measures from fluent productions of word-initial /p/ obtained from children who stutter (CWS) (pre- and post-therapy) and nonstuttering (NS) children.

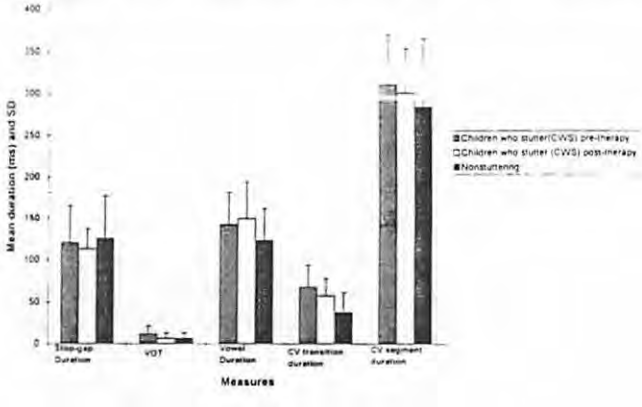


Figure 2. Means and standard deviations of the five temporal acoustic measures from fluent productions of word-initial /b/ obtained from children who stutter (CWS) (pre- and post-therapy) and nonstuttering (NS) children.

($p < .0001$) in their pre-therapy fluent productions of /p/, when compared to the nonstuttering children (Figure 1).

Similar results were observed for /b/; that is, the pre-therapy fluent productions of the children who stutter were characterized by significantly longer vowel durations ($p < .01$) and CV transition durations ($p < .0001$). In addition, while longer CV segment durations were produced by the stuttering children for /p/ but not /b/, a finding unique to /b/ for the children who stutter is that they produced significantly longer VOT's ($p < .01$) than the nonstuttering children during their pre-therapy fluent productions of this sound.

Fluent speech of nonstuttering children and children who stutter post-therapy. In order to provide insight into the extent to which stuttering therapy might influence fluent speech production, we compared the post-therapy fluent productions of word-initial /p/ and /b/ produced by the stuttering children to those produced by the nonstuttering children. Again, separate Wilcoxon-Mann-Whitney analyses with adjusted alpha levels of .01 for each of five comparisons (.05 for all five comparisons as a family for each sound) were performed. Results for vowel duration were similar to those observed in the pre-therapy comparison, that is, following intensive therapy the children who stutter continued to show significantly longer vowel durations ($p < .01$) for /p/ as compared to the nonstuttering children (Figure 1). However, as shown in Figure 1, no other significant between-group differences were found among the remaining four temporal acoustic measures for post-therapy fluent productions of /p/.

For /b/ (Figure 2), the children who stutter as a group showed significantly longer CV transition duration ($p < .0001$) and vowel duration ($p < .0005$) after therapy, similar to the pre-therapy comparison. However, unlike the pre-therapy analysis, following treatment there were no other significant between-group differences for any of the remaining three acoustic measures obtained from fluent productions of /b/.

Discussion

The principal finding of this study is that following therapy, children who stutter exhibited decreased duration of a variety of temporal measures related to the perceptually fluent productions of /p/ and /b/. Further, these durational decreases were seen primarily in measures representing dynamic, rather than static aspects of speech production, those requiring fairly rapid, precise coordination between or within supralaryngeal and laryngeal systems (e.g. VOT, CV transitions as opposed to steady-state portion of vowels).

Results from this investigation are inconsistent with those from the majority of published acoustic studies of the pre- vs post-therapy fluent speech of children and adults who stutter. As described earlier, these studies have shown a general tendency for the post-treatment fluency of people who stutter to be characterized by longer voice onset times (e.g. Metz, Onufrak, & Ogburn, 1979; Metz, Schiavetti, & Sacco, 1990; longer vowel duration (e.g. Metz et al., 1979; Metz, Samar, & Sacco, 1983; Robb, Lybolt, & Price, 1985), and speech segments (e.g. intervocalic interval) containing either increased proportion of voicing (e.g. Franken, 1987), decreased periods of silence (Metz et al., 1983), or both. Present findings do, however, support preliminary work (Zebrowski, 1991) which reported decreased durations of stop-gap, VOT and steady-state vowel production in the fluent speech of two children who stutter enrolled in the same therapy program as that reported here.

Discrepancies between the present findings and most prior studies are most likely explained by differences in both subject and treatment characteristics. For example, with the exception of the studies by Robb et al. (1985) and Zebrowski (1991), all of the previously discussed investigations examined the pre- and post-therapy fluent speech of adults who stutter, as opposed to that produced by children. Several researchers have made the case that studying the fluency of adults who stutter is problematic in that any anomalous or disrupted speech behaviors observed are most likely related to a history of stuttering (effect) rather than its etiology (cause) (e.g. Conture, Colton, & Gleason, 1988; Hulstijn, Starkweather, & Peters, 1991; Zebrowski, Conture, & Cudahy, 1985). For example, in a recent review of studies comparing the fluent speech of children who stutter to that of their nonstuttering peers, Armson & Kalinowski (1994) concluded that older children (and, therefore adults as well) with a longer history of stuttering, as opposed to young children relatively close to stuttering onset, are more likely to exhibit fluency which is "contaminated" or influenced by this history. Specifically, these older children and adults have most likely established strategies for coping with or compensating for their instances of stuttering, and such strategies are most likely manifest in speech production behaviors underlying the entirety of their speech, both stuttered and fluent. Therefore, it is more likely to observe between-group differences in the fluency of stuttering and nonstuttering adults than it would be when comparing the fluent speech of children who stutter and their nonstuttering counterparts.

The second reason for disparate findings between the present and previous research is related to treatment effects. One issue here of course is the history of treatment itself, with regard to amount and duration of therapy received. However, a second, but equally important issue relates to the type of therapy received and the techniques or strategies employed to produce fluent speech. Most of the pre- vs post-treatment studies discussed here obtained

acoustic measures from subjects who underwent therapy which emphasized either (1) reduced speech rate (e.g. Metz et al., 1979; 1983), (2) "stretched", prolonged or increased syllable duration (e.g. Franken, 1987), (3) a slow, prolonged phonation pattern (e.g. Robb et al., 1985), (4) a combination of these strategies. In contrast, as in Zebrowski (1991), the children in the present study received a form of "stuttering modification" or "stutter more fluently" therapy (see Gregory, 1979, and Guitar & Peters, 1985 for discussions of "speak more fluently" vs "stutter more fluently" approaches to the treatment of stuttering). In the approach used here, a general identification and "relearning" of the normal (fluent) speech process was the focal point of therapy, and slowed speech rate or prolonged syllables were not emphasized (see Method section for a discussion of specific goals of therapy). One of the important goals of this approach is "acceptable stuttering", in which the speaker produces "noticeable but not severe disfluency and feels comfortable speaking despite his disfluency (Guitar & Peters, 1985, p. 15)." While arguably "noticeable but not severe disfluency" is ambiguous, for the program used here it is interpreted to describe stuttered speech characterized by either sound, syllable and word repetitions, and audible sound prolongations of relatively short duration (500 ms and below), or with no or minimal accompaniment of associated (non)speech behaviors (e.g. visible physical tension, consistent loss of eye contact, extraneous facial, articulator, head, torso or limb movement, audible inhalations or exhalations, vocal pitch rise or break), or both.

Considering the focus of these different therapy approaches, it seems reasonable to conclude that findings from a number of studies are most likely the result of specific targets taught in therapy, and do not necessarily represent speech production behaviors which are essential for fluent speech. For example, the consistent finding of fluency characterized by longer vowel durations and increased phonation time or proportion of voiced segments probably stems from the fact that the individuals for whom these changes were observed were taught to "stretch" or prolong syllables in order to speak fluently. Further support for this speculation can be found in a recent study of lip and jaw movement associated with fluent speech (McClean, Kroll, & Loftus, 1990). In this investigation, McClean et al. observed that stuttering adults with no recent history of therapy were not significantly different from their nonstuttering peers with regard to 15 parameters of articulator movement during lip closure associated with fluent speech. However, an additional experimental group of stuttering adults enrolled in an intensive fluency shaping program showed significantly longer jaw movement duration and time to peak velocity of the upper and lower lips, and jaw following treatment, when compared to the control group. Further, differences in articulatory behavior were found post-therapy for the individuals who stutter when compared to their nonstuttering counterparts. In addition, the treated experimental subjects more frequently showed "reversals" in the typical sequencing of upper lip, lower lip and jaw velocity peaks accompanying lip closure. Further, such "reversals" were observed in relation to increases in jaw movement duration during the closing gesture. The authors deduced that in general, the lip and jaw movements associated with stuttering and nonstuttering adults' instances of fluent speech do not differ "unless the stutterers have completed a speech therapy program that targets prolongation of speech segment durations (p. 758)." Results from the present investigation offer support for this conclusion, in that the subjects participating in this study did not receive therapy which focused on fluency shaping through the modification of specific aspects of speech timing. Consequently, their post-therapy fluency was generally characterized by segment and subsegment durations which were shorter in duration than observed pre-therapy, but not significantly different from the same measures produced by nonstuttering children.

Significance of Post-Therapy Changes in Duration

The findings from this study may be interpreted in a variety of ways. First, when combined with those from similar preliminary work (Zebrowski, 1991), present findings suggest that stuttering therapy which emphasizes a "relearning" of the normal processes of speech, as opposed to focusing on rate reduction, prolonged speech, or both, can help children who stutter produce fluent segment and subsegment durations which more closely approximate those produced by nonstuttering children. Further, as previously discussed, the post-therapy temporal changes seen here were primarily related to dynamic as opposed to static speech behaviors, associated with sound-to-sound transitions (e.g VOT and CV transition duration. The nature of these changes, namely a reduction in duration, suggests that as a group the children who stutter showed post-treatment fluency characterized by faster or more efficient or coordinated transitioning between speech sounds (Wingate, 1976; 1988). More specifically, as discussed by Metz et al. (1983), it may be the case that certain changes in acoustic parameters reflect a "regularization" of underlying timing relationships between articulatory and phonatory gestures, which facilitates smooth, fluent transitions between speech segments and sounds. Finally, as Guitar (1991) speculated, it may be the case that "at least for young stutterers, slowing rate and increasing duration of voicing may be not be necessary. Speech may become more normal if the stutterer stops reacting to the stuttering so catastrophically (p. 555)."

Issues Related to Post-Therapy Fluency

A broader issue related to the findings from this and similar studies has to do with the identification of treatment approaches for children who stutter which are most likely to result in the highest degree of maintenance of both fluent speech and "acceptable stuttering." Obviously, this is an empirical question for which data are long overdue. However, results from acoustic studies of pre- and post-therapy fluency, along with observations made in studies of speech naturalness, may provide a basis for some conjecture.

For example, as previously discussed, studies have shown that children and adults who stutter enrolled in therapy programs emphasizing rate reduction and prolonged syllables and vowels, exhibit post-therapy fluency which is characterized by longer segment and subsegment durations. These findings are most likely the result of the specific targets or behaviors taught in therapy, and may not represent behaviors necessary for fluent speech production by people who stutter. Further, as the present study suggests, and as supported by other work (e.g. McClean et al., 1990; Zebrowski, 1991) the post-treated fluency of people treated for stuttering through the use of these "fluency shaping" procedures is less like the fluent speech of nonstuttering individuals than either their pre-therapy fluency or the fluency of people enrolled in other kinds of therapy programs (for example, stuttering modification). Finally, a number of studies have consistently indicated that the perceptually fluent speech of stuttering individuals treated through fluency shaping techniques is judged to sound "unnatural" and "monotonous" by listeners (e.g. Ingham, Gow, & Costello, 1985; Franken, Boves, Peters, & Webster, 1991).

Taken together, findings from both perceptual and speech production studies of post-treatment fluency suggest that the use of stuttering modification or "stutter more fluently" approaches with school-aged children who stutter may result in fluent and disfluent speech which more closely approximates the fluent speech of nonstuttering children, both in the way it is produced, and the way it sounds to listeners. An obvious advantage here is that fluency is not contrived through the pervasive use of slowed rate, prolonged syllables, or reduced prosody, and therefore it is likely that children can retain both spontaneity and expressiveness in speaking (Franken et al., 1991; Guitar, 1991). Therefore, while this kind of approach requires a fairly high degree of cognitive ability, behavioral awareness and desensitization to stuttering and speaking, it may be the case that long-term results for school-aged children with a history of stuttering are more consistent when compared to fluency shaping. As previously discussed, data on treatment results for children who stutter are sorely needed before the validity of this and similar speculations can be addressed.

Limitations

There are limitations to this study. Because of the small number of the subjects in this study, our ability to interpret and generalize our results is limited. One of the ways in which the study could be strengthened would be to report individual data, which will be presented at a later date. Also, Onslow et al. (1992) suggests that if we use only repetitious speech samples (i.e., multiple repetitions of the same sentence), the external validity of the results is questionable. Therefore, given that the phonetic context of the tokens analyzed was selectively limited, our ability to generalize our results to other phonetic contexts remains limited.

Future Research

Because of the inherent limitations of acoustic analysis (i.e. that it provides only indirect measurement of physiological events), a more complete understanding of how the fluency of children who stutter is achieved following participation in a particular treatment program will be offered following the use of acoustic analysis in conjunction with analysis of respiratory, laryngeal, and supralaryngeal events. Future attempts to better understand the production of fluent speech following treatment might include a larger number of subjects, and the use of more diverse phonetic contexts. Such study will illuminate our understanding of how treatment goals are realized at the physiological level and which treatment goals contribute most to the production of fluent speech in chidden who stutter.

Acknowledgments

This work was supported by Grant No. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders.

References

Armson, J. and Kalinowski, J. (1994). Interpreting results of the fluent speech paradigm in stuttering research: Difficulties in separating cause and effect. Journal of Speech and Hearing Research, 37, 69-82.

Conture, E. (1990). Stuttering (2nd ed.). NJ: Prentice Hall.

Conture, E., Colton, R., & Gleason, J. (1988). Selected temporal aspects of coordination during fluent speech of young stutterers. Journal of Speech and Hearing Research, 31, 640-653.

Franken, M.C. (1987). Perceptual and acoustic evaluation of stuttering therapy. In H.F.M. Peters and W. Hulstijn (Eds.) Speech Motor Dynamics in Stuttering (pp. 285-294). Wien/NY: Springer-Verlag.

Franken, M.C., Boves, L., Peters, H.F.M., & Webster, R.L. (1991). Prosodic features in the speech of post-therapy stutterers compared with the speech of nonstutterers. In H.F.M. Peters, W. Hulstijn, & C. W. Starkwether (Eds.) Speech motor control and stuttering (pp. 527-534). Amsterdam: Excerpta Medica.

Gregory, H.H. (Ed.) (1979). Controversies about stuttering. Baltimore: University Park Press.

Guitar, B. (1991). The implications of research findings in speech motor control for the treatment of stuttering. In H.F.M. Peters, W. Hulstijn, & C. W. Starkwether (Eds.) Speech motor control and stuttering (pp. 547-557). Amsterdam: Excerpta Medica.

Guitar, B. & Peters, T.J. (1985). Stuttering: An Integration of Contemporary Therapies. Memphis, TN: Speech Foundation of America.

Hulstijn, W., Starkweather, C.W., & Peters, H.F.M. (1991). Speech motor control and stuttering: An introduction. In H.F.M. Peters, W. Hulstijn, & C.W. Starwether (Eds.) Speech motor control and stuttering (pp. XVII-XXIII). Amsterdam: Excerpta Medica.

Ingham, R.J., Gow, M., & Costello, J.M. (1985). Stuttering and speech naturalness: Some additional data. Journal of Speech and Hearing Disorders, 50, 217-219

McClean, M.D., Kroll, R.M., & Loftus, N.S. (1990). Kinematic analysis of lip closure in stutterers' fluent speech. Journal of Speech and Hearing Research, 33, 755-760.

Mallard, A. and Westbrook, J. (1985). Vowel duration in stutterers participating in Precision Fluency Shaping. Journal of fluency Disorders, 10, 221-228.

Metz, D.E., Onufrak, J. & Ogburn, R. (1979). An acoustical analysis of stutterers' speech prior to and at the termination of therapy. Journal of Fluency Disorders, 4, 249-254.

Metz, D.E., Samar, V.J., & Sacco, P.R. (1983). Acoustic analysis of stutterers' fluent speech before and after therapy. Journal of Speech and Hearing Research, 26, 531-536.

Metz, D.E., Schiavetti, N., & Sacco, P.R. (1990). Acoustical and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutterers. Journal of Speech and Hearing Disorders, 55, 516-525.

Onslow, M., Doorn, J.V., & Newman, D. (1992). Variability of acoustic segment durations after prolonged-speech treatment for stuttering. Journal of Speech and Hearing Research, 35, 529-536.

Riley, G. (1980). Stuttering Severity Instrument for Young Children. Austin, TX: Pro-Ed.

Robb, M.P., Lybolt, J.F., & Price, H.A. (1985). Acoustic measures of stutterers' following an intensive therapy program. Journal of Fluency Disorders, 10, 269-279.

Wingate, M.E. (1976). Stuttering: Theory and treatment. NY: Irvington-Wiley.

Wingate, M.E. (1988). The structure of stuttering: A Psycholinguistic analysis. NY: Springer-Verlag.

Williams, D. (1957). A point of view about "stuttering". Journal of Speech and Hearing Disorders, 22, 390-397.

Williams, D. (1979). A perspective on approaches to stuttering therapy. In H. H. Gregory (Ed.), Controversies about stuttering. Baltimore: University Park Press.

Zebrowski, P.M. (1991). Preliminary acoustic analysis of the perceptually fluent speech of stuttering children before and after therapy (pp. 423-432). In H.F.M. Peters, W. Hulstijn, & C.W. Starkweather (Eds.), Speech motor control and stuttering. Amsterdam: Excerpta Medica.

Zebrowski, P.M., Conture, E. & Cudahy, E.A. (1985). Acoustic analysis of young stutterers' fluency: Preliminary observation. Journal of Fluency Disorders, 10, 173-192.

NCVS Status and Progress Report - 6
May, 1994, 89-97

Acoustic and Glottographic Voice Analysis During Drug-Related Fluctuations in Parkinson's Disease

Kristin K. Larson, Ph.D., CCC-SP
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Lorraine Olson Ramig, Ph.D., CCC-SP
Department of Communication Disorders and Speech Science, The University of Colorado-Boulder
and
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts
Ronald C. Scherer, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts

Abstract

Acoustic and electroglottographic measures were used to quantify the effects of drug cycle fluctuations, specifically the "on-off" effect and dyskinesias, on phonation in two subjects with Parkinson's disease. Analysis of maximum sustained and comfortable vowel phonations indicated phonatory instabilities and variability and increased glottal adduction were present in both subjects. Also, a systematic and consistent relationship between drug cycle fluctuations and phonatory measures was not observed, but there was evidence to suggest that "short term" perturbation measures (shimmer and jitter) vs. "long-term" perturbation measures (coefficient of variation of amplitude and frequency) were more sensitive to drug treatment effects and that phonatory function was more similar to limb responses during "on-off" vs. dyskinetic states. Potential mechanisms underlying these findings are discussed.

Introduction

Neuropharmacologic treatment, such as levadopa, has been proven effective in reducing rigidity and akinesia in Parkinson's disease (PD) (Cotzias, Papavasiliou & Gellene, 1969; Diamond, Markham & Treciokas, 1976). However, rapid fluctuations in motor response, ranging between severe PD symptoms to random, involuntary movements, sometimes occur after long-term levadopa use (Shoulson, Glaubiger & Chase, 1975; Yahr, 1975; Marsden & Parks, 1976). These fluctuations, referred to as the "on-off" phenomena and dyskinesias, are readily apparent in limb musculature, but their influence on phonation is less obvious. Several perceptual studies, in which phonatory function was assessed prior to and after the administration of levadopa, suggest that vocal quality, pitch, loudness and prosody improve subsequent to levadopa therapy (Audelman, Hoel & Lassman, 1970; Mawdsley & Gamsu, 1971; Wolfe, Garvin, Bacon & Waldrop, 1975). Although these subjective reports indicate a trend toward improvement, objective data quantifying this effect throughout a subjects drug cycle are lacking. Since 89% of PD subjects experience voice problems (Logemann, Fisher, Boshes, Blonsky, 1978) and approximately half experience adverse drug-related side effects (McDowell & Sweet, 1976; Granerus, 1978), it would be important to quantify the effects of drug cycle modulations on phonation for reasons related to accurate diagnoses and treatment effects. Also, determining if the phonatory mechanism is modulated similarly to other motor systems may provide insight into the neural mechanisms underlying phonatory control.

This study was performed to quantify potential changes in phonatory function related to the "on-off" phenomena using acoustic and electroglottographic (EGG) measures. These types of analyses provide an objective means to measure phonatory stability and glottal adduction, via measures such as the coefficient of variation of amplitude and frequency (long-term perturbation measures), jitter and shimmer (short-term perturbation measures) (Scherer, Gould, Titze, Meyers & Sataloff, 1988) and the abduction quotient (a glottal adduction measure) (Titze,

1984). Since acoustic measures may reflect subtle perturbations that occur during phonation and the abduction quotient may provide information regarding vocal fold closure, these measures combined provide a noninvasive means to document phonatory changes in laryngeal structure or function that may occur as a result of pharmocologic, behavioral or surgical interventions.

Method

Subjects

Two males (H.W. and N.P) diagnosed with idiopathic PD, aged 59 and 37, participated in this study. Each experienced the "on-off" effect at least twice daily that ranged from dyskinesias to Parkinsonian symptoms, primarily tremor and rigidity. Both displayed a hypokinetic dysarthria that was characterized by reduced vocal loudness. In addition, H.W. experienced breathiness and N.P. experienced vocal tremor. These symptoms reportedly fluctuated within drug cycles. Neurologic characteristics of the subjects are summarized in Table 1.

Table 1.
Neurologic characteristics of subjects H.W. and N.P.
Disease stage based on Hoehn and Yahr (1967).
SCR = controlled-release carbidopa-levadopa (Sinemet CR);
B = bromocriptine.

	H.W.	N.P.
Years post onset	16	12
Disease stage "on"	II	II
Disease stage "off"	III	IV
Medications	SCR, B	SCR, B

Voice and Electroglottographic (EGG) Recordings

The voice was recorded with a AKG EMC-21 condensor microphone that was positioned at a constant 8 cm from each subject' s mouth via a headset. EGG recordings were obtained by securing two electrodes of a Teany electroglottograph on opposite sides of the thyroid lamina near the level of the thyroid notch and a third electrode against the skin of the neck. The amplified voice and EGG signals were digitized, video encoded (Sony Digital Processor PCM-501ES) and recorded on a Magnovox VCR-VR8417.

Procedures

Subjects participated in two full-day recording sessions, one week apart. During each day, data were collected every hour for ten consecutive hours. Prior to each hourly recording session, subjects rated their perceived physiologic response to drug treatment using two visual analog rating scales (Boeckstyns & Backer, 1989). One scale measured "on-off" symptoms, with the "on" end of the continuum representing the absence of PD symptoms and the "off" end representing the presence of extreme PD symptoms. The second scale measured perceived degree of dyskinesia, with one end of the scale representing no dyskinesia and the other end representing extreme dyskinesia. The subjects were instructed to place a vertical slash through each scale at a point that was most representative of their current response to drug treatment. It was accepted that the subjects were able to accurately document their physiologic response to drug treatment since both had been doing so as part of their year long involvement in a double-blind cross over study that investigated the efficacy of different drug treatment types for PD. Furthermore, each subjects' neurologist confirmed their ability to correctly document drug treatment responses and each patient expressed confidence in their ability to perform this task.

Each hourly recording session occurred in a sound-treated booth and consisted of having the patient produce 6 maximum sustained vowel phonations (MSVPs), 3 at the beginning and 3 at the end of each session, and 15 comfortable vowel phonations (CVPs). Since speech task may influence acoustic perturbation values (Ramig & Ringel, 1983), both MSVPs and CVPs were performed. For the MSVPs, the patient was instructed to take a deep breath and sustain the vowel /a/ for as long and steadily as possible. Instructions to the subjects for CVPs were to sustain the vowel /a/ for approximately 5 seconds. Both tasks were performed at comfortable pitch and loudness levels. Each hourly recording session lasted approximately 15 minutes.

Data Analysis

Acoustic analysis. A coustic analysis was performed on data obtained for all trials on day 2 and for trials on day 1 that corresponded to the subjects' physiologic rating of their most extreme "on", "off" and dyskinetic state. The data obtained on day 2 data was selected for primary analysis in an attempt to minimize any learning effect (King, Ramig, Lemke & Horii, in press). Two seconds near the middle of the MSVPs and CVPs were filtered at 10 KHz and digitized (16-bit Digital Sound Corporation A/D converter at 20,000 samples/s) to a VAX 11/750 computer. The middle of each phonation was sampled to maximize stability and minimize vocal onset and offset effects (Ramig, Scherer, Burton, Titze & Horii, 1990). Attempts were made to obtain the following long- and short-term acoustic perturbation measures for each MSVP and CVP using the computer analysis software GLIMPSES (Glottal Imaging by Processing External Signals, Titze, 1984): the coefficient of variation of amplitude (CVA), the coefficient of variation of frequency (CVF) and jitter and shimmer with linear trend removed (See

Table 2.
The amount of analyzable data (s) for maximum sustained vowel phonations (MSVPs) and comfortable vowel phonations (CVPs)for H.W. on day 2 (D2) and day 1 (D1). Trial numbers and perceived physiologic state ratings ("on-off"/dys) are also shown.

Trial	On-off/dys rating D2	H.W. MSVPs (s out of 12 possible s)	H.W. CVPs (s out of 30 possible s)	On-off/dys rating D1	H.W. MSVPs (s out of 12 possible s)	H.W. CVPs (s out of 30 possible s)
3	91/8	6	14			
4	110/5	4	10	115/5	4	12
5	68/16	10	13			
6	4/20	4	24			
7	7/82	2.5	18	3/65	2.5	21
8	55/4	8	30	47/6	2	20
9	39/5	3	29	66/0	7	18
10	5/45	11	28			

Table 3.
The amount of analyzable data (s) for maximum sustained vowel phonations (MSVPs) and comfortable vowel phonations (CVPs) for N.P. on day 2 (D2) and day 1 (D1). Trial numbers and perceived physiologic state ratings ("on-off"/dys) are also shown.

Trial	On-off/dys rating D2	N.P. MSVPs (s out of 12 possible s)	N.P. CVPs (s out of 30 possible s)	On-off/dys rating D1	N.P. MSVPs (s out of 12 possible s)	N.P. CVPs (s out of 30 possible s)
1	24/94	6	28			
3	110/4	12	30			
4	104/6	12	24	99/8	8	26
5	97/3	6	15			
6	28/104	5	16	25/94	5	22
8	95/5	6	10			
9	100/4	6	22			
10	108/2	8	6			

Appendix for formulas). The latter two measures were selected in order to reduce long-term effects on cycle-to-cycle measures. Since some acoustic segments were extremely aperiodic, perturbation measures could not be obtained for each vowel phonation. Thus, the amount of analyzable data varied across trials, ranging between 2 - 12 s/trial of the 12 seconds available for MSVPs and 6 - 30 s/trial of the 30 seconds available for CVPs. Segments less than 1 s in duration (100 successive cycles) were excluded. For the specific amount of analyzable data for each subject across trials, refer to Tables 2 and 3.

Electroglottographic (EGG) Analysis. EGG analysis was performed on the fifteen CVPs obtained for each trial on day 2. Two seconds of the EGG signal were digitized simultaneously (20,000 samples/s) with the two second portions of the acoustic signals to the VAX 11/750 computer. Abduction quotients were obtained by analyzing five successive cycles of the EGG signal with GLIMPES (Titze, 1984).

Perceptual Ratings. As an adjunct to acoustic and EGG measures, the CVPs of each subject's most "on", "off" and dyskinetic trials were blindly evaluated by two trained speech-language pathologists for the presence/absence of mucus interference, breathiness, glottal fry, vocal tremor and vocal tract fluctuations. For reliability purposes, each trial was rated twice.

Otolaryngological Ratings. A videoendopscopic screening was performed on each subject during a maximum "on", "off" and dyskinetic state using either a flexible fiberscope or rigid endoscope. The tapes of these screenings were blindly evaluated by three examiners (an otolaryngologist and two speech scientists), each of whom were asked to characterize what they observed. These observations were used to supplement acoustic and electroglottographic results.

Results

Data Exclusion

Data were excluded when: (a) only one 2-second vowel segment or less was analyzable for a trial and (b) when the validity of a subjects rating for a trial was questionable. Using the first criterion, N.P.'s trial 7 MSVP data ("off" trial) and trial 2 CVP data ("off" trial) on day 2 were eliminated as were H.W.'s trial 8 MSVP data ("on" trial) on day 1. Data obtained for H.W.'s trial 1 and 2 on day 2 were also eliminated because H.W.'s visual analog scale ratings of perceived physiologic state were questionable. For example, H.W.'s ratings for trial 2 suggested that he was experiencing extreme dyskinesia while "off" his drug. This is contradictory to literature that suggests dyskinesias reflect an over-medicated state or maximum "on" state (Duvoison, 1989; Obeso et al., 1989). With the exception of these two trials, the remainder of the subject's physiologic ratings were considered accurate and were consistent with the experimenter's observations of the subject.

Subjects Ratings of Perceived Physiological State

The visual analog scale data that represented the subject's perceived physiologic state were converted to a numerical scale ranging from 0 to 125, with 0 representing the absence of Parkinsonian symptoms and dyskinesias and 125 representing the presence of extreme symptoms. As shown in Figure 1 a and b, perceived physiologic state ratings for H.W. on day 2 were well distributed along the "on-off" and dyskinetic continuums, while the majority of N.P.'s ratings on day 2 were clustered in either the "off" or extreme dyskinetic state. Distributions similar to these were observed for each subject on day 1.

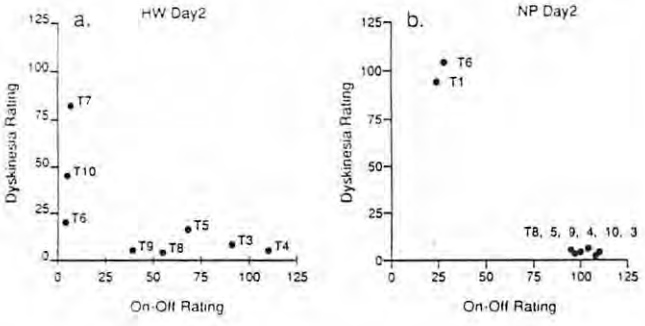


Figure 1. Numerical ratings for perceived physiologic condition for H.W. (a) and N.P. (b) on day 2. Dyskinesia ratings are shown on the y axis and "on-off" ratings are shown on the x axis. Trial numbers are adjacent to the data points.

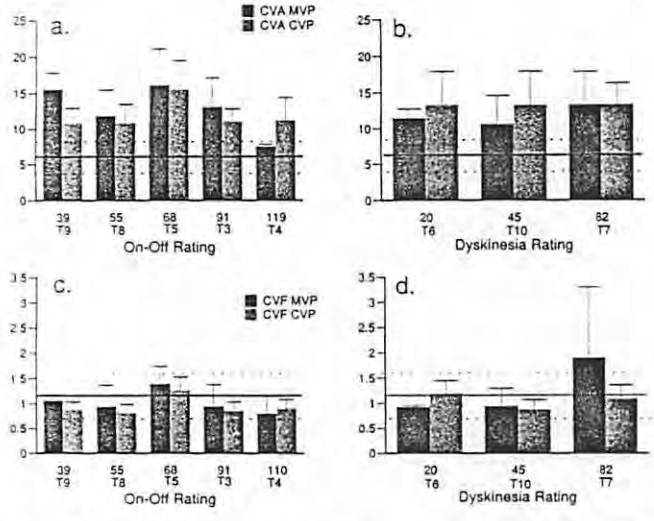


Figure 2. Means and standard deviations for the coefficient of variation of amplitude (CVA) (a,b) and the coefficient of variation of frequency (CVF) (c,d) for "on-off" and dyskinetic states for H.W. on day 2 during maximum sustained vowel phonations (dark bars) and comfortable vowel phonations (light bars). Numerical values for physiologic state ratings and the corresponding trial number are shown on the x axis. The solid horizontal line represents the normative mean and the dashed horizontal lines represent +/- 1 standard deviation.

Acoustic Perturbation Results

Means and standard deviations for the coefficient of variation of amplitude (CVA), coefficient of variation of frequency (CVF) and shimmer and jitter with linear trend removed were derived from the analyzable MSVPs and CVPs for each trial. Given the small number of subjects, statistical analyses were not performed on these data. Rather, acoustic perturbation results are discussed relative to normative data obtained for these measures from 48 male speakers, aged 25-75 years. Half of these subjects were classified as being in good physiologic condition while the other half was classified as being in poor physiologic condition (Ramig & Ringel, 1983).

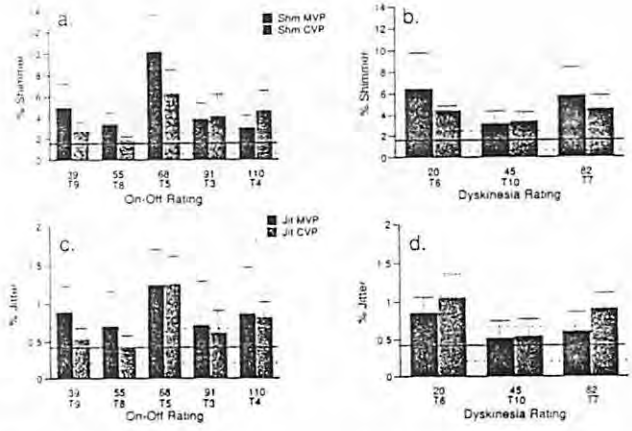


Figure 3. Means and standard deviations for the shimmer (a,b) and jitter (c,d) with linear trend removed for "on-off" and dyskinetic states for H.W. on day 2 during maximum sustained vowel phonations (dark bars) and comfortable vowel phonations (light bars). Numerical values for physiologic state ratings and the corresponding trial number are shown on the x axis. The solid horizontal line represents the normative mean and the dashed horizontal lines represent +/- 1 standard deviation.

H.W. Day 2

Coefficient of variation of amplitude (CVA) and coefficient of variation of frequency (CVF) results for H.W.'s "on-off" and dyskinetic states and for MSVPs and CVPs on day 2 are shown in Figure 2 a,b,c,d. CVA values for MSVPs and CVPs were above the normal range with the exception of MSVPs for trial 4 ("off" state), while CVF values for these phonatory tasks were within the normal range with the exception of MSVPs for trial 7 (most dyskinetic state). A relationship between CVA and CVF and perceived physiological condition was not observed for either phonatory task.

H.W.'s shimmer and jitter results for "on-off" and dyskinetic states and for MSVPs and CVPs are shown in Figure 3 a,b,c,d. Mean shimmer values for H.W.'s MSVPs were consistently above the normal range and did not appear to vary with drug-induced physiologic changes. In contrast, mean shimmer values for CVPs fell within or close to normal limits for trial 8 and 9 ("on" states) but were 2 to 3 standard deviations above the normal range for trial 3 and 4 ("off" states). Also, shimmer variability for "on" states was smaller than the variability for "off" states (Figure 3a). Shimmer values for MSVPs and CVPs along the dyskinetic continuum fell 1 to 4 standard deviations above the normal range regardless of perceived physiologic response to drug treatment (Figure 3b).

Jitter values for MSVPs along the "on-off" continuum were consistently above the normal range and variable, whereas jitter means and standard deviations for CVPs tend to reflect greater phonatory stability for the "on" trials compared to the "off" trials. However, "on-off" jitter

Table 4.
Means (M) and standard deviations (SD) of perturbation measures obtained for maximum sustained vowel phonations (MSVPs) and comfortable vowel phonations (CVPs) for H.W.'s most "on", "off" and dyskinetic state on day 1.
CVA = coefficient of variation of amplitude; CVF = coefficient of variation of frequency; * = normal ranges were exceeded.

	H.W. MSVPs Day 1			H.W. CVPs Day 1		
	"On" T9	"Off" T4	Ext Dys.T7	"On" T8	"Off" T4	Ext Dys.T7
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
CVA	13.5 (4.5)*	16.2 (4.2)*	23.4 (9.9)*	15 (5.7)*	15 (3.8)*	15 (10.5)*
CVF	1.1 (.20)	.55 (.35)	1.37 (.26)	1.3 (.40)	.92 (.22)	1.2 (.17)
Shimmer	7 (1.2)*	6.2 (.9)*	6.4 (.7)*	5.2 (1.5)*	3.9 (1.6)*	5.2 (.71)*
Jitter	.85 (.12)*	.56 (.34)	1.22 (.05)*	1.1 (.36)*	.65 (.20)	1.1 (.29)*

values for CVPs stay within or close to normal limits regardless of perceived physiologic condition (Figure 3c). Jitter values for H.W.'s dyskinetic trials were variable across phonatory tasks and did not appear related to severity of dykinesia. For example, trial 6 jitter values (least dyskinesia) were above normal limits, while trial 10 jitter values (moderate dyskinesia) fell within normal limits (Figure 3d).

H.W. Day 1

For reliability, acoustic perturbation measures were obtained for H.W.'s most "on" (trial 8 and 9), "off" (trial 4) and dyskinetic states (trial 7) on day 1. Results are shown in Table 4.Results for CVA and CVF for both MSVPs and CVPs were similar to results observed for these measures and tasks on day 2. Specifically, CVA values were above normal limits, while CVF values were within normal limits. A trend related to drug-induced fluctuations was not apparent for either of these measures.

Day 1 shimmer and jitter values for both MSVPs and CVPs tended to fall above normal limits regardless of the subjects perceived physiologic state. The exception being jitter values obtained for trial 4 ("off" state), which were within the normal range. Also, the trend observed for shimmer and jitter on day 2 for CVPs was not observed on day 1. Specifically, shimmer and jitter values did not appear to vary with changes in H.W.'s response to drug treatment.

N.P. Day 2

As shown in Figure 1b, N.P.'s ratings were clustered in either the "off" or extreme dyskinetic state. Thus, it was difficult to assess drug treatment effects on phonatory stability because N.P. never perceived himself as being "on" with no dyskinesia. Nevertheless, results for long and short-term amplitude and frequency measures are shown in Figure 4 a,b,c,d and Figure 5 a,b,c,d.

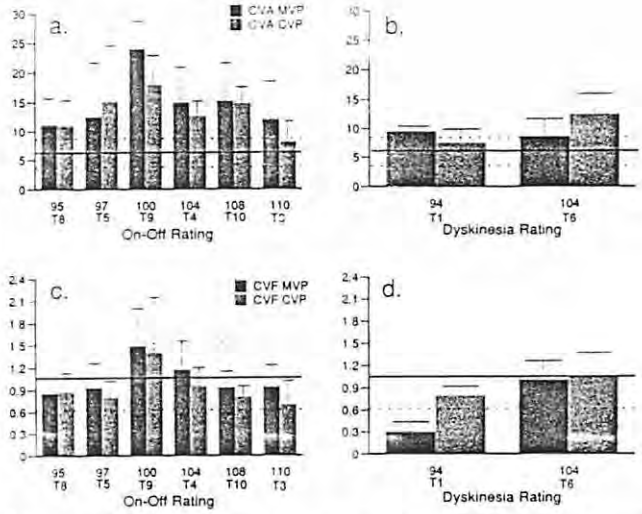


Figure 4. Means and standard deviations for the coefficient of variation of amplitude (CVA) (a,b) and the coefficient of variation of frequency (CVF) (c,d) for "on-off" and dyskinetic states for N.P. on day 2 during maximum sustained vowel phonations (dark bars) and comfortable vowel phonations (light bars). Numerical values for physiologic state ratings and the corresponding trial number are shown on the x axis. The solid horizontal line represents the normative mean and the dashed horizontal lines represent +/- 1 standard deviation.

CVA values for N.P.'s MSVPs and CVPs were generally above the normal range for both "off" and dyskinetic periods. Also, more variability was apparent for this measure during "off" trials compared to dyskinetic trials (Figure 4 a,b). CVF values consistently fell within the normal range regardless of perceived physiological condition (Figure 4 c, d). Shimmer values fell above normal limits for MSVPs and CVPs with the exception of CVPs for trial 10 ("off" trial) (Figure 5 a, b), while jitter values for both phonatory tasks were within or close to the normal range regardless of N.P.'s perceived response to drug treatment (Figure 5 c, d).

N.P. Day 1

Similar to day 2, N.P. did not experience any "on" states during day 1. Nevertheless, results for N.P.'s most "off" (trial 4) and dyskinetic state (trial 6) on day 1 are shown in Table 5. Results for CVA and shimmer showed increased phonatory stability during dyskinetic states compared to "off" states. This contrasts results for these measures on day 2. CVF and jitter values fell within normal limits regardless of perceived physiologic response. This is consistent with results from day 2.

Electroglottographic (EGG) Results

It has been suggested that abduction quotients of -1.0 or less may correspond to laryngeal hyperadduction, while abduction quotients of 0.5 or greater may be indicative of laryngeal hypoadduction (Scherer et al., 1988). As

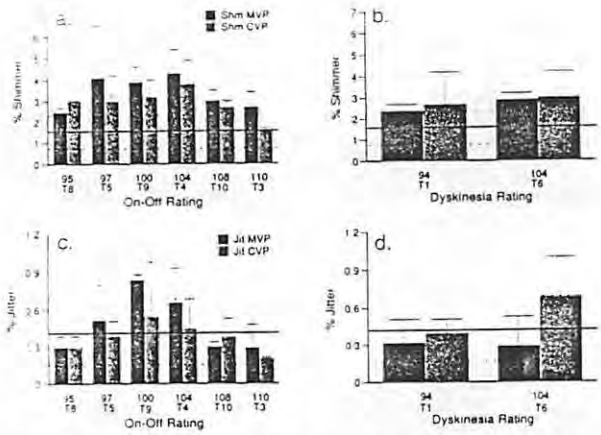


Figure 5. Means and standard deviations for the shimmer (a,b) and jitter (c,d) with linear trend removed for "on-off" and dyskinetic states for N.P. on day 2 during maximum sustained vowel phonations (dark bars) and comfortable vowel phonations (light bars). Numerical values for physiologic state ratings and the corresponding trial number are shown on the x axis. The solid horizontal line represents the normative mean and the dashed horizontal lines represent +/- 1 standard deviation.

indicated in Figure 6a, abduction quotients for H.W. showed a slight trend toward increased adduction as he progressed from "on" to "off" states. However, variability was prevalent. In Figure 6b, it can be seen that abduction quotients for H.W.'s least and most dyskinetic trials indicate laryngeal hyperadduction, while trial 10's (moderate dyskinesia) abduction quotient indicates normal laryngeal function (Figure 6b). Most of N.P.'s abduction quotients were within the normal range, but showed a tendency toward hyperadduction for both "off" and dyskinetic trials (Figure 6c, d). In both subjects EGG data, a "bump" was intermittently observed on the opening phase of EGG waveforms (Figure 7).

Table 5.
Means (M) and standard deviations (SD) of perturbation measures obtained for maximum sustained vowel phonations (MSVPs) and comfortable vowel phonations (CVPs) for N.P's most "off" and dyskinetic state on day 1. CVA = coefficient of variation of amplitude; CVF = coefficient of variation of frequency; * = indicate normal limits were exceeded.

	N.P. MSVPs Day 1		N.P. CVPs Day 1	
	"Off" T4	Ext Dys T6	"Off" T4	Ext Dys T6
	M (SD)	M (SD)	M (SD)	M (SD)
CVA	6.9 (.4)	12.3 (3.7)*	7 (2.9)	17.1 (4.4)*
CVF	.8 (.25)	1.1 (.55)	1.2 (.44)	1.4 (.23)
Shimmer	2 (.93)	3 (1.9)*	2.2 (.32)	2.9 (.9)*
Jitter	.38 (.19)	.5 (.17)	.52 (.35)	.56 (.18)

Perceptual Results

Perceptual ratings revealed that H.W.'s voice was characterized by breathiness, glottal fry, mucus flow interference and vocal tract fluctuations, while N.P.'s voice was characterized by vocal tremor, glottal fry and mucus flow interference. These characteristics were observed randomly and did not appear to fluctuate with drug-induced changes in physiologic response. Perceptual ratings were 75 - 94% reliable across descriptors and trials.

Otolaryngological Results

Videoendoscopic results revealed random mucus stranding for H.W., and supraglottal and glottal fluctuations, for example, vertical laryngeal movements, for N.P. These observations occurred independent of any drug-related phenomena. It was not possible to judge fine laryngeal movement patterns, for example, amplitude of vocal fold vibration, due to the reduced visual quality of the video tapes.

Discussion

This study investigated changes in phonatory function throughout the drug cycles of two subjects with PD using acoustic and electroglottographic measures. The findings indicate that phonatory instabilities and increased glottal adduction were present regardless of the subject's response to drug treatment. Also, phonatory measures did not appear to be systematically or consistently modulated by drug-induced changes in physiologic state, with the exception of shimmer and jitter for one subject on one day. These measures indicated greater cycle-to-cycle phonatory stability for "on" vs. "off" states. These observations suggest that "short-term" vs. "long-term" perturbation

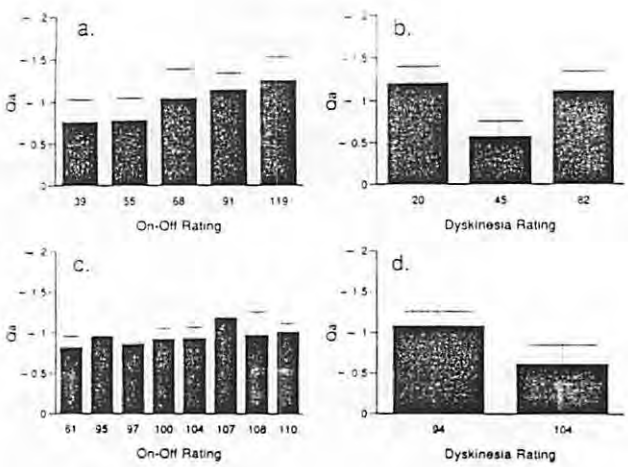


Figure 6. Abduction quotients (Qa) for H.W. (a,b) and N.P. (c,d) during "on-off" and dyskinetic states. Numerical values for physiologic state ratings are shown on the x axis.

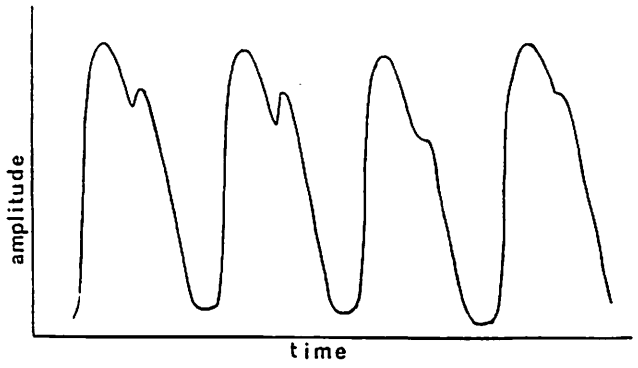


Figure 7. Sample of four electroglottographic waveforms demonstrating the "bump" on the opening phase.

measures may be more sensitive to medication effects and that phonatory responses may be more similar to limb motor responses during drug-induced "on-off" states rather than dyskinetic states.

Acoustic studies have demonstrated vocal instabilities in PD (Ramig, Scherer, Titze & Ringel, 1988; Ludlow & Bassich, 1988). For example, Ramig et al. (1988) reported cycle-to-cycle instabilities in intensity and frequency and increased spectral noise in the sustained phonations of PD subjects. In the current study, amplitude perturbation measures (CVA and shimmer) were typically above normal limits, while frequency perturbation measures (CVF and jitter) fluctuated within and above normal limits. Scherer et al. (1988) suggested that high CVA values may indicate relatively slow innervation fluctuations to laryngeal abductor/adductor muscles and/or fluctuations in subglottal pressure and supralaryngeal structures. It seems unlikely that subglottic pressure variations contributed to the abnormal CVA values because CVF did not parallel CVA across trials. Rather, it may be that "slow" neuromuscular fluctuations influenced laryngeal abductory/adductory movements and/or supraglottic movements. Research has recently shown that the posterior cricoarytenoid, a laryngeal abductor, can contribute to slow-varying vocal oscillatory behaviors (Koda & Ludlow, 1993). Also, perceptual evaluations revealed that vocal tremor and vocal tract fluctuations were present in these subjects. Thus, CVA and CVF results in conjunction with subjective impressions suggest that these subject's experienced a pervasive long-term amplitude abnormality, potentially related to laryngeal and/or supralaryngeal oscillations, but unrelated to perceived physiologic response to drug treatment ("on", "off" or dyskinesia).

In contrast, there appeared to be some relationship between shimmer and jitter and drug cycle modulations. Specifically, these measures increased as one subject progressed from "on" to "off" states. Also, the standard deviations for shimmer and jitter were smaller and twice as much data were analyzable for "on" (59 s) vs. "off" (23 s) trials, further suggesting there was greater phonatory stability for "on" vs. "off" states. Several researchers have suggested that shimmer and jitter perturbations may be related to neuromuscular fluctuations in laryngeal and respiratory muscles, asymmetrical vocal fold movements, mucosal movement on the vocal folds and/or aerodynamic factors (Baer, 1979; Larson, Kempster & Kistler, 1987; Titze, 1991). Although it is difficult to discern the specific sources contributing to the trend observed for shimmer and jitter, it is possible that neurologic input to laryngeal muscles changed as a result of pharmocological intervention and as such influenced vocal fold abduction, adduction and/or tension. However, if the mechanisms underlying shimmer and jitter were truly influenced by drug-induced fluctuations, a similar trend would have been expected across days and speech tasks (MSVPs and CVPs), but this was not observed. Since MSVPs showed greater phonatory instability and variability compared to CVPs, particularly for "on-off" vs. dyskinetic states, it may be that differences in perturbation measures related to drug treatment were masked by the variability during MSVPs. Greater instability for MSVPs compared to CVPs have also been reported by Ramig & Ringel (1988). They suggested this may be related to the fact that MSVPs are a maximal effort task and as such may be more susceptible to subtle breakdowns in a system. It is also possible that the trend observed for shimmer was related to mucus stranding. However, it seems unlikely that mucus contributed solely to the shimmer differences between "on" and "off" states, mainly because perceptual and fiberoptic observations showed that mucus appeared independent of changes in physiologic state. Also, mucus would not influence jitter, which remained within normal limits, but did show more phonatory stability for "on" vs. "off" trials.

For dyskinetic trials, high shimmer values and fluctuating (within and above normal limits) jitter values were observed. As previously mentioned, high shimmer may be related to "fast" acting fluctuations in the laryngeal abd/adductory system and/ or tissue assymetries, while jitter variability may be the result of abnormal neural input to laryngeal muscles involved in pitch-regulation. However, random and irregular contractions of, for example, the cricothyroid and thyroarytenoid, would have produced variable voice fundamental frequency (VF_0) patterns, especially for trials characterized by high jitter. This was not observed. Thus, it may be that jitter reflected laryngeal muscle degeneration, potentially as a result of PD or aging (Ramig and Ringel, 1983). If subject's were sometimes able to compensate for this degeneration, normal jitter values may have occurred.

Abduction quotients indicated a tendency toward increased laryngeal adduction. Hyperadduction may reflect an attempt by the subject to overcome the symptoms

of decreased loudness and breathiness, or as Hanson, Gerratt & Ward (1984) suggested it may be the result of a breakdown in laryngeal agonist and antagonist muscle coordination. Negative abduction quotients may also be an artifact of mucus stranding. The "bump" observed on some EGG waveforms indicates the presence of mucus. Since this "bump" would also widen the EGG waveform, it could have biased abduction quotients toward the negative (hyperadducted) range. Although laryngeal abduction and adduction seemed affected, it did not appear to be modulated by drug treatment.

With the exception of the aforementioned shimmer and jitter findings, results of this study tend to suggest that phonatory function is not modulated systematically and is not modulated similarly to the limbs throughout the drug cycles of these two subjects. These findings are similar to those reported by Solomon and Hixon (1993), in which only slight differences in the respiratory function of PD subjects were reported during different phases (middle and end) of the subject's drug cycle. Similarly, Sharpe, Fletcher, Lang & Zacker (1987) found that the eye tracking abilities (eye movement velocities) of PD subjects did not fluctuate with drug-induced changes in physiologic state. In contrast, research quantifying labial and jaw kinematics as a function of pharmocologic treatment during dynamic speech tasks (syllable repetitions) showed that speech movements improved during "on" vs. "off" levadopa states (Caliguri, 1988; Svensson, Henningson & Karlsson, 1993). Thus, it appears that dopaminergic drugs may differentially influence motor systems. Also, given the fact that there was a considerable amount of within subject variability for perturbation values and relatively little variability for physiologic state ratings, the idea that distinct dopaminergic mechanisms mediate drug-related fluctuations in phonatory and skeletomotor systems seems even more tenable.

Other factors that may have influenced the results relate to methodology, for example, phonatory task and subject selection. Phonation may have been more susceptible to drug-related fluctuations if dynamic vs. isometric (sustained vowels) phonatory tasks were produced. Also, both of these subject's experienced severe drug-dose related fluctuations and one subject was typical of early-onset PD. Therefore, these subject's may not be the most representative of idiopathic PD. These characteristics may also partially explain the intra- and intersubject variability that was observed for perturbation measures during the subject's drug cycle.

Given the inconsistency and variability in these results, it is apparent that further investigations are necessary to elucidate the influence of drug-related fluctuations on phonation. More direct physiologic measures combined "short-term" acoustic variables and different speech tasks may help clarify this issue.

Acknowledgements:

This research was supported by the research and training grant (University of Iowa) DC-00976 and NIH-NIDCD grant #R01 DC01150. Dr. Dave Hill, Dr. Arlen Meyers, Vern Vail, William Winholtz and Neal Pinto are gratefully acknowledged.

References:

Audelman, J.U., Hoel, R.L. & Lassman, F.M. (1970). The effect of L-dopa treatment on the speech of subjects with Parkinson's disease. Neurology, 20 (4), 410-1.

Baer, T. (1979). Vocal jitter: a neuromuscular explanation. In V. Lawrence and B. Weinberg (Eds.), Transcripts of the eighth symposium. Care of the professional voice (pp. 19-22). New York: The Voice Foundation.

Boeckstyns, M.E. & Backer, M. (1989). Reliability and validity of the evaluation of pain in subjects with total knee replacement. Pain, 38(1), 29-33.

Caliguri, M.P. (1989). Short-term fluctuations in orofacial motor control in Parkinson's disease. In K. Yorkston & D. Beukelman (Eds.), Recent Advances in Clinical Dysarthria (pp. 199-212). Boston: Little Brown & Co.

Cotzias, G.C., Papavasiliou, P.S. & Gellene, R. (1969). Modification of parkinsonism-chronic treatment with L-dopa. New England Journal of Medicine, 280, 337-45.

Diamond, S.G., Markham, C.H. & Treciokas, L.J. (1976). Long-term experience with L-dopa: efficacy, progression and mortality. In W. Birkmayer & O. Hornykiewicz (Eds.), Advances in Parkinsonism: Vth international symposium on Parkinson's disease. Basel: Editones Roche.

Duvoisin, R.C. (1989). New strategies in dopaminergic therapy of Parkinson's disease: the use of a controlled-release formulation. Neurology, 39 (suppl 2), 4-6.

Granerus, A.K. (1978). Factors influencing the occurrence of "on-off" symptoms during long-term treatment with L-dopa. Acta Medica Scandinavica, 203, 75-85.

Hanson, K., Gerratt, B. & Ward, R. (1984). Cinegraphic observations of laryngeal function in Parkinson's disease. Laryngoscope, 94, 348-353.

King, J.B., Ramig, L.O., Lemke, J.H. & Horii, Y. (in press). Parkinson's disease: longitudinal changes in acoustic parameters of phonation. Journal of Medical Speech-Language Pathology.

Koda, J. & Ludlow, C.L. (1992). An evaluation of laryngeal muscle activation in subjects with voice tremor. Otolaryngology Head Neck Surgery, 107, 684-96.

Larson, C.R., Kempster, G.B. & Kistler, M.K. (1987). Changes in voice fundamental frequency following discharge of single motor units in cricothyroid and thyroarytenoid muscles. Journal of Speech and Hearing Research, 30, 552-58.

Logemann, J., Fisher, H., Boshes, B. & Blonsky, E. (1978). Frequency and co-occurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and Hearing Disorders, 43, 47-57.

Ludlow, C.L. & Bassich, C. (1988). The results of acoustic and perceptual assessment of two types of dysarthria. In W. Berry (Ed.), Clinical dysarthria (pp. 121-154). San Diego: College-Hill Press.

Marsden, C. & Parkes, J. (1976). "On-off" effects in patients with Parkinson's disease on chronic levadopa therapy. The Lancet, 1, 292-96.

Mawdsley, C. & Gamsu, C.V. (1971). Periodicity of speech in Parkinsonism. Nature, 231, 315-6.

McDowell, F.H. & Sweet, R.D. (1976). The on-off phenomena. In W. Birkmayer & O. Hornykiewicz (Eds.), (pp. 603-12). Basel: Editones Rache.

Obeso, J., Grandas, F., Vaamonde, J., Luquin, M., Artieda, J., Lera, G., Rodriguez, M.& Martinez-Lage, J. (1989). Motor complications associated with chronic levadopa therapy in Parkinson's disease. Neurology, 39, 11-19.

Ramig, L.O., Ringel, R. (1983). Effects of physiological aging on selected acoustic characteristics of voice. Journal of Speech and Hearing Research, 26, 22-30.

Ramig, L.O., Scherer, R.C., Titze, I. and Ringel, S. (1988). Acoustic characteristics of voice of patients with neurological disease: a preliminary report. Annals of Otology, Rhinology and Laryngology, 97, 2, 164-172.

Ramig, L.O., Scherer, R., Burton, E., Titze, I. & Horii, Y. (1990). Acoustic analysis of voice in amyotrophic lateral sclerosis: a longitudinal study. Journal of Speech and Hearing Disorders, 55, 2-14.

Scherer, R.C., Gould, W.J., Titze, I.R., Meyers, A.D. & Sataloff, R.T. (1988). Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. Journal of Voice, 2, 230-44.

Solomon, N.P. & Hixon, T.J. (1993). Speech breathing in Parkinson's disease. Journal of Speech and Hearing Research, 36, 294-310.

Sharpe, J., Fletcher, W., Lang, A & Zacker, D. (1987). Smooth pursuit during dose-related on-off fluctuations in Parkinson's disease. Neurology, 37, 1389-91.

Shoulson, I., Glaubiger, G.A. & Chase, T.N. (1975). On-off response. Neurology, 25, 1144-48.

Svensson, P., Henningson, C. & Karlsson, S. (1993). Speech motor control in Parkinson's disease: a comparison between a clinical assessment protocol and a quantitative analysis of mandibular movements. Folia Phoniatrica (Basel), 45, 4, 157-64.

Titze, I. (1984). Parameterization of the glottal area, glottal flow and vocal fold contact area. Journal of the Acoustical Society of America, 75, 570-80.

Titze, I. (1991). A model for neurologic sources of aperiodicity in vocal fold vibration. Journal of Speech and Hearing Research, 34, 460-72.

Wolfe, V.I., Garvin, J.S., Bacon, M. & Waldrop, W. (1975). Speech changes in Parkinson's disease during treatment with L-Dopa. Journal of Communication Disorders, 8, 271-79.

Yahr, M. (1975). Levadopa: diagnosis and treatment. Annals of Internal Medicine, 83, 677-682.

Appendix

Formulas for Short- and Long-term Stability Measures

The coefficient of variation (for frequency or amplitude) is defined as the standard deviation divided by the mean and expressed in percentage.

$$\frac{100}{\bar{x}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

Jitter (shimmer) with linear trend removed is defined as the average absolute difference in frequency (amplitude) between the following: a cycle's frequency (amplitude) minus the average frequency (amplitude) and expressed in percentages.

$$\frac{100}{(N-2)\bar{x}}\sum_{i=2}^{N-1}|.5(x_{i+1}+x_{i-1})-x_i|$$

NCVS Status and Progress Report - 6
May, 1994, 99-111

Speech and Voice Deficits in Parkinsonian Plus Syndromes: Can They be Treated?

Stefanie Countryman, M.A.
Wilbur James Gould Voice Research Center, The Denver Center for The Performing Arts
Lorraine O. Ramig, Ph.D.
Department of Communication Disorders and Speech Science, The University of Colorado-Boulder
and
Wilbur James Gould Voice Research Center, The Denver Center for The Performing Arts
Annette A. Pawlas, M.A.
Wilbur James Gould Voice Research Center, The Denver Center for the Performing Art

Abstract

Three cases of Parkinsonian plus syndrome with associated moderate to severe speech and voice deficits are presented. Each subject completed an intensive 1 month course of the Lee Silverman Voice Treatment. Acoustic analysis was completed on experimental variables collected pre-treatment, immediately post-treatment and 6 months post-treatment. Measures of intensity and maximum duration of sustained vowel phonation and intensity and fundamental frequency variability during reading and speaking improved in all three subjects following voice therapy. These changes were reflected in perceptual ratings of loudness, monotonicity, slurring and overall intelligibility completed by the patients and family members. The effects of intensive voice therapy on functional communication skills in this population is discussed.

Parkinsonism is a progressive neurological movement disorder that effects over one million individuals in the United States (Duvosin, 1984). Approximately 80-100% of people with Parkinsonism develop speech and voice deficits that negatively influence their ability to communicate with family and friends and/or limit employment opportunities (Otoxby, 1982; Streifler & Hofman, 1984).

There are several different subgroups of Parkinsonism. The largest is idiopathic Parkinson's disease (IPD). IPD affects the pigmented brainstem nuclei, particularly the substantia nigra (Forno, 1982) and is not a result of drugs, trauma, or encephalitis. Approximately 80% of the Parkinsonism group have IPD. The second largest group of Parkinsonian disorders, approximately 12.2% (Stacy & Jancovic, 1992), is classified clinically as Parkinson plus syndromes (PPS) and pathologically as multiple system degenerations (Jancovic, 1989a; 1989b). This population may have symptoms of IPD as well as additional neurological anomalies. Because of the early stage neurological similarities of PPS to IPD, these syndromes are often misdiagnosed as such (Jellinger, 1987; Stadlan, Duvoisin, & Yahr, 1966; Tygstrup & Norholm, 1963). Frequently, PPS patients maintain a diagnosis of "possible" IPD throughout a large part of their disease. In some cases, patients become severely disabled or die before a diagnosis of a PPS is conclusive.

Typically, the speech and voice deficits that accompany PPS are more severe and deteriorate faster than those observed in IPD (Quinn, 1989). Yet no efficacious approach to speech therapy has been documented for this population. In addition, referrals for speech therapy usually occur very late in the disease or when the patient is so debilitated that an augmentative communication device is the only option. Hence the patients may endure months or even years of poor oral communication that can significantly diminish their quality of life.

Recently, an effective intensive voice treatment approach [Lee Silverman Voice Treatment (LSVT)] was developed for patients with IPD (Ramig, Horii & Bonitati, 1991; Ramig, Mead, DeSanto, & Horii, 1988; Ramig, Mead, Scherer, Larson, & Kohler, 1988). This therapy specifically focuses on increased phonatory effort. Increased phonatory effort in IPD has been shown to increase vocal loudness and generalize to and improve other speech deficits (e.g. articulation, rate, and intonation) not directly addressed in therapy (Ramig, 1992; Ramig, Bonitati, Lemke, & Horii, 1994; Ramig, Horii, & Bonitati, 1991). Consequently, speech intelligibility and functional communication significantly improves in patients with IPD receiving the LSVT approach.

Since many of the PPS syndromes have similar speech and voice deficits as those seen in IPD (Hanson, Ludlow, & Bassich, 1983; Kluin, Foster, Berent, & Gilman, 1993; Metter & Hanson, 1991; Quinn, 1989), the LSVT program is a possible option for treatment. A positive response to the LSVT would provide the benefit of effective speech treatment to a large population of patients who normally are overlooked and are generally considered non responsive to speech therapy.

In addition, because of the similar neurological and physical pathologies in PPS and IPD, PPS patients are often referred to a speech clinic with the diagnosis of IPD. A differential response on the part of the PPS patient to the LSVT (when compared to the IPD patient) may offer useful differential diagnostic information. If this differential response is combined with other suspicious neurological symptoms, it may support a possible non-IPD diagnosis.

This study was developed in order to document the effect of the LSVT on the speech and voice deficits in three cases of PPS and to evaluate the maintenance of those changes up to 6 months post-treatment.

Methods

Subjects

All patient characteristics are summarized in Table 1.

Table 1.
Patient Characteristics for the Parkinsonian Plus Syndrome Patients

	SEX	AGE	STAGE (Hoehn & Yahr, 1967)	*YSD	SPEECH DISABILITY
Patient 1	M	59	IV	3	severe
Patient 2	F	64	III	4	moderate-severe
Patient 3	F	73	IV	3	moderate-severe

* years since diagnosis

Patient 1 was a 59 year old male in Parkinson's stage IV (Hoehn & Yahr, 1967) at the time of enrollment (September 1991) in the LSVT program. Parkinson medication included carbidopa-levodopa (Sinemet) and amantadine hydrochloride (Symmetrel). He was first diagnosed with abnormal neurological symptoms in 1988. The patient was diagnosed with PPS in 1990 and conclusively with multiple system atrophy (MSA) in 1993.

MSA is a progressive disease which involves the autonomic and central nervous systems (Graham & Openheimer, 1969; Polinsky, 1984; Quinn, 1989) and represents approximately 5.3% of cases referred to movement disorder clinics. Quinn (1989) suggested that a diagnostic indicator of MSA is more severe speech symptoms than those observed in IPD. He described the speech of MSA patients as severely hypophonic with a severe slurring dysarthria which may be accompanied by respiratory stridor while sleeping. Symptoms of Parkinson disease may be a dominant characteristic of MSA and frequently these patients are misdiagnosed as such (Rivest, Quinn, & Mardsen, 1990).

A structural function speech mechanism examination in patient 1 revealed tremors in the tongue, reduced tongue strength, reduced speed and range of motion of the tongue, and reduced range of motion and coordination of the obicularis oris muscle. Perceptual speech and voice characteristics included significantly reduced loudness, monotony of loudness, extreme breathiness with frequent aphonia during both spontaneous speech and sustained phonation. Vocal flutter (Aronson, Ramig, Winholtz, & Silber, 1992), low pitch, monotonicity, rough vocal quality, a wet or gurgle vocal quality, reduction in normal stress patterns, reduced articulatory precision and rapid rate were also observed. The speech pathologist rated overall speech impairment as severe. The patient and family member reported limited functional communication skills. Problems with swallowing both liquids and solids were reported by the patient.

Pre-treatment otolaryngological examination revealed redness and swelling of both vocal cords, excess watery secretions in the laryngeal area, tremor, mild bowing, severely decreased amplitude of mucosal wave, and consistent irregularity of vocal fold movement.

Extensive neuropsychological examination revealed mild to moderate impairment learning new verbal (*Story Memory Test*) (Heaton, Grant, & Matthews, 1991) and nonverbal information (*Figure Memory Test*) (Heaton et al., 1991). Performance was within normal limits on all other tests of cognitive functioning. These results indicated mild cerebral dysfunction in this patient, primarily affecting learning and memory. Psychosocial testing revealed significant depression (*Profile of Mood States*) (McNair, Lorr, & Droppleman, 1971) as well as considerable fatigue

and emotional lability, (*Sickness Impact Profile*) (Bergner, Bobbit, Carter, & Gilson, 1981).

Patient 2 was a 64 year old female in late Parkinson's stage III (Hoehn & Yahr, 1967) at the time of enrollment (March 1992) in this program. Parkinson medication included carbidopa-levodopa, selegiline hydrochloride (Eldyprel), and pergolide mesylate (Permax). Abnormal neurological symptoms were first noticed in 1988. The patient was diagnosed with atypical Parkinson's disease in 1990, likely Shy-Drager syndrome.

Shy-Drager syndrome (SDS), is a neurological disorder of multiple systems atrophy which consists of progressive generalized autonomic failure and extrapyramidal and/or cerebellar dysfunction. Vocal cord paralysis (specifically the abductor muscles), swallowing deficits, and speech deficits have been reported in Shy-Drager syndrome (Bassich, Ludlow & Polinsky, 1984; Briskin, Lehrman, & Guilleminault, 1978; Hanson, Ludlow, & Bassich, 1983; Keww, Gross, & Chapman, 1990; Kwentus, Auth, & Foy, 1984; Linebaugh, 1979; Martinovits, Leventon, Goldhammer, & Sadeh, 1988; Rizzoli, 1986; Thomas & Schirger, 1970; Williams, Hanson, & Calne, 1979). Specifically, Hanson et al. (1983) described the following speech and voice deficits in SDS: breathy, strained voice quality, reduced loudness, monopitch, monoloudness, imprecise articulation, variations in rate, a slow deliberate speaking rate, excess vocal hoarseness, and intermittent glottal fry.

A structural function speech mechanism examination for patient 2 revealed a slight tremor in her tongue, reduced speed and range of lateral tongue movements, rigidity in the obicularis oris muscle, and inability to elicit gag reflex. Perceptual speech and voice characteristics included reduced loudness, monotonicity, a slight vocal tremor, vocal fry, increased speaking rate, reduced articulatory precision and reduction of normal stress patterns. The speech pathologist rated overall speech impairment as moderate to severe. The patient and family member reported reduced functional communication, with limited initiation of conversation. No swallowing difficulties were reported.

Pre-treatment otolaryngological examination revealed redness and swelling on the right vocal cord posterior to the vocal fold process. This was concluded to be a result of gastric reflux. (The reflux was treated and resolved before the initiation of the LSVT). Decreased abduction of the left vocal cord, a slight vocal tremor, moderate degree of bowing, slight decrease in amplitude of mucosal wave, and an infrequent irregularity of periodicity were also noted.

Neuropsychological examination revealed mild impairment on a test that requires efficiency in following new sequential procedures (*Trail Making Test*) (Reitan & Wolfson, 1985) and mild to moderate impairment in sustained attention and concentration (*Digit Vigilance Test*) (Heaton et al., 1991). Patient 2 also showed mild to moderate slowing in learning new verbal (*Story Memory Test)* (Heaton et al., 1991) and nonverbal information (*Figure Memory Test*) (Heaton et al., 1991) and on a test assessing visual spatial analysis (*Judgement of Line Orientation Test*) (Benton, Hannay, & Varney, 1975). All other test results were within normal limits. These results indicated mild, fairly generalized cerebral dysfunction. Psychosocial testing revealed mild depression (*Beck Depression Inventory*) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) (*Montgomery-Asberg Depression Rating Scale*) (Montgomery & Asberg, 1979). The patient also reported nervousness, tension, and anxiety in her day-to-day activities (*Profile of Mood States*) (McNair et al., 1971), (*Sickness Impact Profile*) (Bergner et al., 1981).

Patient 3 was a 73 year old female in Parkinson's stage IV (Hoehn & Yahr, 1967) at the time of enrollment (October 1992) in this program. Parkinson medication included carbidopa-levodopa, selegiline hydrochloride, and bromocriptine mesylate (Parlodel). The patient's first abnormal neurological symptoms were noticed in 1989. She was diagnosed with a Parkinsonian plus syndrome in 1991, most likely progressive supranuclear palsy (PSP).

PSP is a progressive neurological disease centered predominantly in the brainstem (Steele, Richardson, & Olszewski, 1964) and accounts for approximately 3.5 - 7.5% of patients referred to movement disorder clinics with a possible diagnosis of IPD (Agid, Javoy-Agid, & Ruberg, 1986; Duvoisin, Golbe, & Lepore, 1987; Jackson, Jancovic, & Ford, 1983; Stacy & Jancovic, 1992). Often, dysarthria is an early symptom of PSP and occurs in the majority of PSP cases (Kluin, Foster, Berent, & Gilman, 1993; Stacy & Jancovic, 1992; Steele, Richardson, & Olszewski, 1964). Speech difficulties are reported as a spastic, hypokinetic or mixed dysarthria (Kluin et al., 1993; Metter & Hanson, 1991). The specific speech symptoms associated with PSP are similar to IPD in that they include imprecise articulation, reduced stress, monopitch, monoloudness, reduced loudness, disturbance of rate, and a harsh voice (Kluin et al., 1993; Metter & Hanson, 1991). However, Metter and Hanson (1991) reported that imprecise articulation is usually the most pre-dominant speech characteristic in PSP whereas this is not usually true for IPD.

A structural function speech mechanism examination revealed an edentulous woman with a large tongue. All other musculature was within normal limits. Perceptual speech and voice characteristics included reduced loudness, monotonicity, a breathy weak voice, low pitch, reduced rate of speaking, reduced respiratory support and imprecise articulation. During counting and spontaneous speech there was a noticeable deterioration in the subject's voice and a need for frequent inhalations. The speech pathologist rated the patient's overall speech impairment as

moderate to severe. The patient and family member reported significantly reduced communication ability both in initiating and maintaining conversation. No swallowing problems were reported by the patient.

Pre-treatment otolaryngological examination revealed moderate bowing of the vocal cords and a moderate to severe degree of false fold hyperadduction.

Neuropsychological testing was not completed due to the patient's poor vision and digital manipulation. Psychosocial testing revealed moderate depression (*Beck Depression Inventory*) (Beck et al., 1961) and significant emotional stress as a result of her disease (*Profile of Mood States*) (McNair et al., 1971).

Treatment

The Lee Silverman Voice Treatment (LSVT) (Ramig, Mead, DeSanto, & Horii, 1988; Ramig, Mead, Scherer, Horii, Larson, & Koehler, 1988; Ramig, 1994) was administered to these patients. Treatment techniques focused on increasing vocal loudness by targeting the hypothesized underlying laryngeal pathophysiology and maximizing phonatory and respiratory effort. The program was intensive (16 sessions in one month) and specifically designed to maximize patient motivation and immediate carry-over of increased vocal loudness into functional communication. The LSVT method of treatment has been described in detail elsewhere (Countryman & Ramig, 1993; Ramig, 1994; Ramig, Countryman, Winholtz, Horii, Thompson, 1994).

Variables

The objective variables chosen to evaluate treatment related change were: intensity, maximum duration, and mean fundamental frequency of sustained vowel phonation and intensity, mean fundamental frequency, and fundamental frequency variability during reading and speaking. The following subjective perceptual variables were selected for measurement: single word and overall intelligibility and loudness, monotonicity, and slurring.

Maximum duration of sustained vowel phonation was chosen for measurement and training for several reasons (Ramig, Bonitati, Lemke, & Horii, 1994). Patients with Parkinsonism have a well-established difficulty in simultaneously executing two different movements (Breneckel, Rothwell, Dick, Day, & Marsden, 1986; Yanagisawa, Fujimato, & Tamaru, 1989) and may experience impaired cognitive function (Loranger, Goodell, & Lee, 1972; Pollack, 1966; Stacy & Jankovic, 1992). This phonatory task may be taught easily to most cognitively impaired patients and can be analyzed simply in clinical and experimental situations. Analysis can be completed in instances of severely dysphonic voices. Maximum duration of sustained vowel phonation is related to laryngeal (Yanagihara, Koike, & Leden, 1966) and respiratory function (Boone, 1977). Stimulation of these two mechanisms is important for improving overall vocal loudness in patients with IPD (Ramig, 1992).

Intensity in sustained vowels, reading and speaking was chosen for measurement because all three patients in this study presented with reduced loudness. Vocal loudness improvement is the principal treatment goal of the LSVT program. It was hypothesized that increasing vocal loudness in all three patients would increase their overall functional communication.

Fundamental frequency variability during reading and speaking as well as mean fundamental frequency in sustained vowel phonation, reading and speaking were measured in order to determine if changes in phonatory and respiratory effort and vocal intensity generalized to other aspects of the patients' speech. Changes in fundamental frequency often accompany changes in vocal intensity (Linville & Korabic, 1987; Jacob, 1968). In addition, it has been noted anecdotally that IPD patients who complete the LSVT program appear more animated and have enhanced affect. Improved affect may also influence measures of intensity and frequency. Changes in these measures have been previously documented in patients with IPD following a course of the LSVT (Ramig, Mead, Scherer, et al., 1988; Ramig et al., 1994).

Single word intelligibility was measured because it was hypothesized that increased vocal loudness and overall phonatory effort would enhance speech intelligibility. Improvements in single word intelligibility have been previously documented in patients with IPD following a course of the LSVT (Maclay, Ramig, Scherer, Jancosek, 1994).

Ratings of loudness, monotonicity, slurring and overall intelligibility were selected as the perceptual counterparts of the objective acoustic variables measured. Patient and family member perceptual ratings were completed in order to evaluate the impact of treatment on everyday functional communication of each patient. Family members were included because of the known problems with Parkinsonian patient self-perception of communication deficits (Scott, Caird, & Williams, 1984; Yorkston, Bombardier, & Hammen, 1993).

Data Collection

Experimental data were collected twice pretreatment to establish a baseline and to account for potential variability associated with Parkinson disease (King, Ramig, Lemke, & Horii, 1994) and once immediately post-treatment and 6 months post-treatment. The two pretreatment sessions were collected within one week of the start of treatment. The post-treatment session was collected within one week immediately following therapy. All data collec-

tion sessions were scheduled at approximately the same time of day for each subject and were completed by the same experimenter. All patients were determined to be stable on their medication pretreatment and did not change medication throughout the therapy period.

To obtain the objective acoustic data the subjects were seated in an IAC sound-treated booth with a headset microphone (AKG 410) positioned 8 cm in front of the lips (Titze & Winholtz, 1994). After preamplification through an ATI-1000, the microphone signal was recorded onto a Sony Digital PC-108M (DAT) eight-channel recorder. To collect sound pressure level data, the signal from a Bruel and Kjaer 2230 sound level meter (SLM), placed 50cm from the patient's mouth, was recorded onto the eight-channel DAT recorder.

Immediately preceding each of the patient's voice recording sessions, both the patient and family member rated perceptual variables of loudness, monotonicity, slurring and overall intelligibility of the patient using a visual analog scale (Kempster, 1984). The scale was custom designed to be a clinically feasible tool for both the patient and family member.

To attain measures of maximum duration and intensity of sustained vowel phonation, the subjects were asked to take a deep breath and sustain phonation of the vowel /a/ for as long as possible. A timer with a second hand was within the subjects' view to encourage them to monitor their performance and sustain phonations longer with each repetition. Four to six maximally sustained vowel phonations were collected during each recording session. In order to obtain measures of intensity, fundamental frequency, and fundamental frequency variability during speech, the subjects were asked to read the "Rainbow Passage" (Fairbanks, 1960) and speak for 30 seconds on a topic of interest to them (monologue) at a comfortable rate and loudness. In order to measure single-word intelligibility, the subjects were asked to read, at a comfortable rate and loudness, a closed-set word list (Kent et al., 1989). Words were presented individually on a television monitor at 5 second intervals. Single-word intelligibility is reported only for patient 2. Patient 1's data were invalid due to considerable fatigue (falling asleep) while completing this task. Patient 3's data were invalid due to her pronounced Eastern accent, which appeared to confound listeners' rating of intelligibility.

An otolaryngological history and examination was completed in each patient prior to the beginning of therapy. Laryngeal imaging and videolaryngostroboscopic examination were conducted utilizing well-described techniques (Bless, Hirano, & Feder, 1987). The nasal passage was anesthetized with 4% lidocain spray. Endoscopic examination was conducted with both an Olympus ENF-P3 fiberscope and Nagashima SFT-70 rigid telescope. Images were recorded with a CCD camera, using a 35mm lens for the fiberscope and a 60 mm lens for the rigid telescope, and a SVHS tape recorder.

Forced vital capacity was collected using a Collins wet spirometer (Model RS 2785). Subjects were asked to take their deepest breath and blow out "as hard, fast and as long as you can." This task was repeated three times at the beginning and end of the recording session and the best performance was taken as forced vital capacity.

Data Analysis

Intensity measures during sustained phonation, the "Rainbow Passage," and conversational speech were calculated using a custom built software analysis program. The SLM signal was pre-amplified and then digitized at 5000 samples per second into a VAX system computer through a 16 bit resolution DSC-200 A/D converter. The signal was then analyzed using the software program which displayed the signal in decibels. This program also measured maximum duration of sustained vowel phonation by cursor marking the first negative going peak and the last negative going peak of the SLM signal. The duration was then calculated by the program to within a millisecond.

To calculate measures of mean fundamental frequency during sustained phonation, the microphone signal was passed through the Vocal Demodulator MVD 100 (Winholtz & Ramig, 1992) and the average F_o during the midpoint of each phonation was calculated. The mean for all tokens was then determined.

To determine measures of mean fundamental frequency and fundamental frequency variability (F_oSD) during reading and speaking, the microphone signal of the reading of the "Rainbow Passage" and conversational speech were digitized at 5000 samples per second and analyzed on a 486 computer using C-Speech software (Milenkovic, 1987). The F_oSD was then converted to express frequency variability in semitones (ST).

Single word intelligibility was derived by playing the previously mentioned closed word list at a comfortable loudness level to naive listeners. The listeners were seated in a quiet room approximately 48" from the speaker and were asked to transcribe each word spoken by the patient. Each listener transcribed only one data collection session to avoid familiarity of the speech samples (Beukelman & Yorkston, 1980; Maclay et al., 1994; Yorkston & Beukelman, 1978). Single word intelligibility was calculated in percent from the number of correctly transcribed utterances.

Standard procedures for analysis of visual analog scales (Boeckstyns & Backer, 1989) were used to obtain perceptual data.

Results

Due to the small number of samples and the individual presentation of each case, statistical analyses

Table 2.
Means (boldface) and standard deviations of patients' performance on all experimental measures, pre, immediately post-, and 6 months post-treatment.

	Pre 1	Pre 2	Post 1	6 months post
PATIENT 1				
Vital Capacity (L)	3.0	2.4	2.5	2.8
Maximum Duration (sec)	12.5 (1.76)	13.85 (1.36)	19.52 (1.95)	10.13 (1.33)
Intensity (dB) 50cm				
Sustained phonation	65.27 (0.25)	65.60 (3.00)	71.03 (3.14)	64.22 (2.56)
Rainbow	61.00 (4.68)	62.18 (5.63)	69.62 (2.63)	60.50 (2.36)
Conversation	DNT	65.24 (3.21)	70.50 (3.92)	<60
Semitone Standard Deviaton (stsd)				
Rainbow	1.69	1.81	2.84	1.25
Conversation	DNT	1.54	2.53	1.15
Fundamental Frequency (Hz)				
Sustained phonation	120.17 (9.20)	107.5 (7.58)	113.3 (4.97)	116.50 (4.28)
Rainbow	123.67 (12.04)	118.93 (12.43)	119.87 (19.61)	111.51 (8.10)
Conversation	DNT	122.75 (10.91)	115.14 (16.77)	116.31 (7.72)
PATIENT 2				
Vital Capacity (L)	2.0	2.0	2.2	2.0
Maximum Duration (sec)	26.27 (1.58)	26.34 (1.12)	28.27 (2.01)	23.85 (1.95)
Intensity (dB) 50cm				
Sustained phonation	62.40 (1.42)	63.55 (1.06)	77.10 (2.14)	73.16 (1.52)
Rainbow	65.22 (3.10)	66.76 (2.36)	74.10 (3.18)	69.46 (2.86)
Conversation	66.16 (3.19)	65.87 (3.04)	68.13 (3.01)	68.45 (2.89)
Semitone Standard Deviation (stsd)				
Rainbow	1.26	1.11	2.16	1.56
Conversation	1.49	1.32	1.91	1.25
Fundamental Frequency (Hz)				
Sustained phonation	195.00 (19.06)	189.75 (16.98)	226.60 (13.33)	169.17 (10.40)
Rainbow	177.38 (12.89)	188.86 (12.14)	194.31 (23.47)	181.87 (16.37)
Conversation	215.63 (18.60)	199.99 (15.28)	202.49 (21.25)	185.45 (13.44)
PATIENT 3				
Vital Capacity (L)	1.2	1.0	1.4	1.2
Maximum Duration (sec)	27.93 (2.10)	19.04 (1.75)	20.11 (1.81)	24.38 (2.12)
Intensity (dB) 50cm				
Sustained phonation	59.02 (1.47)	59.66 (0.77)	66.89 (1.70)	64.10 (1.28)
Rainbow	60.42 (2.16)	63.42 (3.46)	69.30 (3.90)	69.14 (3.34)
Conversation	61.54 (2.34)	58.20 (3.66)	65.40 (3.59)	66.03 (3.68)
Semitone Standard Deviation (stsd)				
Rainbow	1.88	2.30	2.60	1.88
Conversation	2.33	1.90	2.61	2.26
Fundamental Frequency (Hz)				
Sustained phonation	175.33 (39.40)	182.60 (6.07)	171.83 (8.66)	170.60 (13.40)
Rainbow	129.55 (14.10)	160.75 (21.31)	170.80 (25.50)	161.75 (14.10)
Conversation	137.94 (18.57)	127.76 (14.01)	150.69 (22.68)	145.63 (18.17)

were not carried out on these data (McReynolds & Kearns, 1983). The results are presented through individual means, standard deviations and description. Pre, immediately post-, and 6 month post-treatment data for each patient for the variables forced vital capacity (L), maximum duration of sustained vowel phonation (sec), intensity (dB), fundamental frequency variability (stsd), and average fundamental frequency (Hz) are presented in Table 2. Intensity and fundamental frequency variability are plotted in Figure 1 for each patient. Pre to immediately post-treatment and immediately post- to 6 month post-treatment changes are presented in the text in parentheses. Pre to post-treatment changes were derived by determining the mean of pre 1 and pre 2 and then subtracting the post-treatment data from this mean.

Twenty percent of the acoustic and perceptual data were reanalyzed to assess reliability. Pearson product-moment correlation coefficients ranged from 0.96 to 0.99

for all acoustic measures analyzed (intraexaminer). Intrarater reliability for repeated perceptual ratings by patient subjects and families ranged from 0.95 to 0.99 for all perceptual variables. Interexaminer reliability for single word intelligibility ratings by two listeners was reflected in a correlation of 1.0.

Patient 1

As summarized in Table 2 and plotted in Figure 1, patient 1 improved on the following variables pre to post-treatment: maximum duration of sustained vowel phonation (6.35 sec), intensity during sustained vowel phonation (5.59 dB), reading (8.03 dB), and speaking (5.26 dB) and fundamental frequency variability during reading (1.09 stsd) and speaking (0.99 stsd).

Both the patient (P) and family member (F) rated improvement pre to post-treatment on the following perceptual variables: loudness (P=31%, F=69%), monotonicity (P=22%, F=69%), slurring (P=48%, F=63%), and overall intelligibility (P=22%, F=82%), (see Table 3).

By 6 months post-treatment, patient 1 deteriorated from immediately post-treatment levels on the following acoustic measures: maximum duration of sustained vowel phonation (-9.39 sec), intensity during sustained vowel phonation (-6.81 dB), reading (-9.12 dB), and speaking (-10.5 dB) and fundamental frequency variability during reading (-1.59 stsd) and speaking (-1.38 stsd). At 6 months post-treatment these acoustic variables were at or slightly below pretreatment levels (see Table 2).

The patient and family member rated a decline on the following perceptual variables at 6 months post-treatment compared to immediately post-treatment: loudness (P=-22%, F=-7%), monotonicity (P=-19%, F=-23%), slurring (P=-60%, F=-19%), and overall intelligibility (P=-19%, F=-66%) (see Table 3). However, all 6 month perceptual ratings were above pretreatment levels (P=3-9%, F=16-42%) with the exception of the patient's self-rating of slurring.

Due to physical and transportation constraints the patient was unable to complete an immediately post-treatment ENT evaluation. However, videostroboscopic data obtained from patient 1 at 10 months post-treatment revealed bowing of the vocal cords during sustained vowel phonation. However, when requested to increase loudness, he completely adducted his vocal folds. Complete adduction was not achieved during his pretreatment evaluation for the same task.

Patient 2

Patient 2 improved on the following variables pre to post-treatment: maximum duration of sustained vowel phonation (1.96 sec), intensity during sustained vowel phonation (14.12 dB), reading (8.11 dB), and speaking (2.11 dB) and fundamental frequency variability during

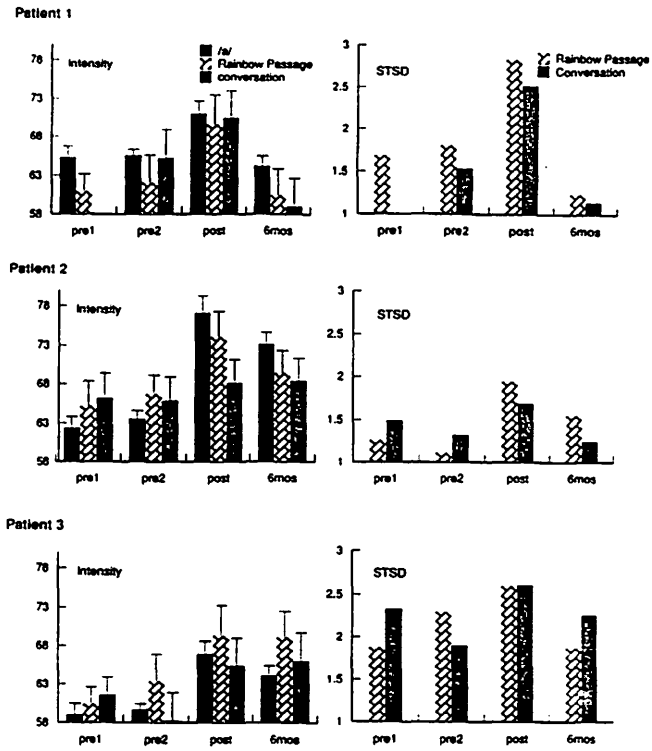


Figure 1. Pre, post- and 6 months post-treatment measures of intensity (dB) during sustained phonation and intensity and fundamental frequency variability (stsd) during reading and conversation for each patient.

reading (0.97 stsd) and speaking (0.50 stsd) (Table 2 and Figure 1).

Both the patient and family member rated improvement pre to post-treatment on all perceptual variables: loudness (P=60%, F=10%), monotonicity (P=43%, F=12%), slurring (P=42%, F=30), and overall intelligibility (P=23%, F=22%) (see Table 3).

By 6 months post-treatment patient 2 deteriorated from immediately post-treatment levels on the following acoustic variables: maximum duration of sustained vowel phonation (-4.42 sec), intensity during sustained vowel phonation (-3.94 dB) and reading (-4.64 dB) and fundamental frequency variability during reading (-0.60 stsd) and speaking (-0.66 stsd). The patient maintained post-treatment levels for the variable intensity during speaking. Despite these declines, intensity during sustained phonation and reading and fundamental frequency variability during reading were above pretreatment levels (Table 2).

Patient 2 and the family member rated a decline on the following perceptual variables from immediately post-treatment to 6 months post-treatment: loudness (P=-55%), monotonicity (P=-11%, F=-10%), slurring (P=-64%, F=-20%), and overall intelligibility (P=-36%, F=-12%). Patient 2's family member rated her loudness 5% improved from post-treatment to 6 months post-treatment. All 6

month post-treatment perceptual ratings were above pre-treatment levels (P= 5-32%, F=11-65%) except for the patient's self-rating of slurring and overall intelligibility (Table 3).

Single word intelligibility for Patient 2 was 72.86 and 81.43% for pre-treatment 1 and 2 respectively. Post-treatment single word intelligibility was 92.75%. Single word intelligibility 6 months post-treatment was 91.43%.

Videostroboscopic data collected on patient 2 immediately post-treatment revealed an improvement in overall vocal function from pretreatment. Complete closure of the vocal folds was observed during sustained phonation as well as improved regularity and periodicity of vocal fold movement. The decreased abduction of the left vocal fold that was noted pretreatment was not observed post-treatment.

Patient 3

Patient 3 improved on the following variables pre to immediately post-treatment: intensity during sustained vowel phonation (7.55 dB), reading (7.38 dB), and speaking (5.53 dB), and fundamental frequency variability during reading (0.51 stsd) and speaking (0.49 stsd) (see Table 2 and Figure 1).

Both patient 3 and the family member rated improvement pre to post-treatment on the following perceptual variables: loudness (P=34%, F=57%), monotonicity (P=76%, F=89%) slurring (P=62%, F=93%), and overall intelligibility (P=55%, F=92%) (Table 3).

By 6 months post-treatment (Table 2) patient 1 deteriorated from immediately post-treatment on the following variables: intensity during sustained vowel phonation (-2.79 dB) and fundamental frequency variability during reading (-0.72 stsd) and speaking(-0.35 stsd). The patient maintained immediate post-treatment improvements by 6 months post-treatment for the variables maximum duration of sustained vowel phonation and intensity during reading and speaking (see Table 2).

The patient and family member rated a decline on the following perceptual variables from immediately post-treatment to 6 months post-treatment: loudness (F=-9%), monotonicity (P=-13%), slurring (P=-16%, F=-19%), overall intelligibility (P=-12%, F=-18%). The patient and family member rated improvement immediately post-treatment to 6 months post-treatment on two variables: loudness (P= 19%) and monotonicity (F= 4%). All 6 month perceptual ratings were above pretreatment levels (P=43-63%, F=25-80%).

Due to physical constraints and illness, patient 3 did not complete an immediately post-treatment ENT evaluation. Videostroboscopic data collected six months post-treatment revealed mild bowing during sustained phonation. However, complete closure was achieved when the patient was requested to increase loudness during sustained phonation. This was not achieved during her pre-treatment evaluation.

Table 3.
Patient and family member evaluation of percent change (improvement) pre to immediately post-treatment and percent change (decline) immediately post-treatment to 6 months post-treatment of 4 perceptual characteristics.

Perceptual characteristic	Post-treatment		Six months post-treatment	
	patient	family member	patient	family member
Patient 1				
Loudness	31	69	-22	-7
Monotonicity	22	69	-19	-23
Slurring	48	63	-60	-19
Intelligibility	22	82	-19	-66
Patient 2				
Loudness	60	10	-55	+5
Monotonicity	43	12	-11	-10
Slurring	42	30	-64	-20
Intelligibility	23	22	-36	-12
Patient 3				
Loudness	34	57	+19	-9
Monotonicity	76	89	-13	+4
Slurring	62	93	-16	-19
Intelligibility	55	92	-12	-18

Discussion

Parkinsonian plus syndromes (e.g. multiple system atrophy, Shy-Drager syndrome and progressive supranuclear palsy) comprise approximately 12% of patients referred to movement disorder clinics. While 80-100% of these patients display speech and voice deficits, the effect of speech treatment in this population has not been evaluated. This is due, in part, to the expectation that speech therapy would be futile for these patients given the rapid progressive deterioration of these diseases. Therefore, it is important to evaluate the efficacy of the Lee Silverman Voice Treatment for this population.

The objective and perceptual data reported here support improvement of the speech and voice deficits in three patients with Parkinsonian plus syndrome following one month of intensive voice treatment (LSVT). Increased intensity during the patients' sustained phonation, reading and conversational speech contributed to improved intelligibility and functional communication as reported by the attending speech pathologist, patients and family members. In addition, increased intensity was associated with improved vocal fold adduction as confirmed by videostroboscopic examination. Improved vocal fold adduction following the LSVT has been documented in IPD patients as well (Smith, Ramig, Dromey, Perez, Samandari, 1994).

By 6 months post-treatment, the patients' objective and perceptual data declined from immediately post-

treatment levels. However, the patients and families reported that overall functional communication skills remained above pre-treatment performance. For example, the family of patient 1 (multiple system atrophy) reported that at 6 months post-treatment he was able to increase his loudness level on cue which immediately improved his intelligibility. Before treatment, the patient was unable to increase loudness spontaneously or on cue. Thus, despite the apparent 6 month post-treatment deterioration of patient 1's objective and perceptual speech data, his overall functional communication skills were superior to his pretreatment abilities.

The multiple system atrophy patient (patient 1) showed the least amount of improvement when compared to patients 2 and 3. Considering the extent of neural degeneration occurring in MSA, this is not surprising. We speculate that if the MSA patient was referred earlier in the course of his disease, when speech, voice and physical deficits were less severe, treatment benefits may have been greater. Despite severe motor deterioration seen in MSA, these patients have a remarkable preservation of intellect well into the advanced stages of the disease (Quinn, 1989). Therefore, effective communication skills would be essential in this population in order to maintain a high quality of life.

Of the three syndromes presented, the patient with Shy-Drager syndrome (patient 2) demonstrated the greatest pre to post-treatment improvement. SDS usually does not progress as rapidly as the other syndromes studied here (Black, 1982). Thus it is possible that SDS patients may have a better prognosis for voice treatment when compared to patients with multiple system atrophy and/or progressive supranuclear palsy.

Although the speech of an individual with SDS may not appear dysarthric until the later stages of the disease, phonatory disorders have been documented in the early stages (Bassich, Ludlow, & Polinsky, 1984). Early intervention for a speech and voice problem that worsens with disease progression, may help the SDS patient maintain a longer period of functional/optimal communication. As reported by this patient and her family, improved functional communication contributed to improved quality of her life.

The progressive supranuclear palsy patient (patient 3) was able to incorporate techniques from the LSVT program into her spontaneous speech post-treatment. This improved her overall speech intelligibility. Hanson and Metter (1980) reported the use of a delayed auditory feedback device to increase loudness, decrease speaking rate and improve overall intelligibility in a progressive supranuclear palsy patient. However, they found that the patient could not spontaneously increase loudness or reduce speaking rate without the device. The PSP patient in the present study spontaneously used treatment techniques in her speech up to 6 months post-treatment.

Since dysarthria is often an early symptom of progressive supranuclear palsy, an efficacious speech and voice treatment program is needed. However, the speech and voice deficits of a PSP patient may have a spastic, hypokinetic, ataxic and/or mixed component (Kluin et al., 1993). The LSVT may be an effective voice treatment approach in this disease if voice symptoms are consistent with those present in patient 3.

The post-treatment levels of the SDS patient (patient 2) and PSP patient (patient 3) closely resembled those of the IPD group previously studied (Ramig et al., 1994). However, given the pretreatment stimulability of patient 2 and 3, it was expected that post-treatment levels and long-term maintenance would have been greater. Overall, the PPS patients responded less well to the treatment and maintained improvement for a shorter duration than the IPD population. Since PPS are frequently misdiagnosed as IPD, an unexpected poor response to the LSVT may support a diagnosis of PPS if other suspicious neurological symptoms are present.

The objective and perceptual data in this study indicate that long-term voice treatment efficacy for this population may be questionable beyond 6 months post-treatment. PPS patients typically have more severe speech and voice deficits and more rapid physical deterioration than patients with IPD. However, for the purpose of this study, these PPS patients received identical treatment as the IPD population. It might be postulated that modifying the LSVT for the PPS patient may improve his/her overall post-treatment results as well as increase long-term efficacy. We suggest that the PPS patients complete the same intensive treatment but receive an extended version, possibly 5 weeks instead of 4. In addition, PPS patients would likely benefit from follow-up therapy sessions since all three patients in this study maintained some functional communication ability above pretreatment levels by 6 months post-treatment.

Due to the cognitive problems and rapid progression observed in PPS syndromes, it is important for a speech therapy program to be simple and enhance functional communication as quickly as possible. We suspect that the LSVT was effective in this population because the therapy concept is simple. The concept "think loud" is taught primarily by focusing on phonatory effort drills. Even the most severely involved patient was able to complete these drills and achieve immediate success in some aspect of the program. Furthermore, despite the PPS patients more severe deficiencies in articulation, prosody and rate when compared to IPD patients, these findings support that intensive therapy focusing on voice only can improve functional communication in this population.

It is important to recognize that in many cases cognitive deficits in PPS patients are less severe than would be expected by initial impression. Lack of facial expression, reduced eye mobility and severe speech impairment give the impression of a greater physical and cognitive impairment than is actual (Kedas, Reed, & Lux, 1989). Consequently, these patients may not be referred for speech therapy and/or the speech therapist may deny treatment. This study demonstrates that PPS patients should not be excluded from intensive voice treatment solely due to physical condition, cognitive deficits, lack of affect and/or energy.

Treatment efficacy data in the field of speech pathology is lacking, especially in the areas of motor speech disorders (Yorkston, Beukleman, & Bell, 1988). Although this study is limited to three cases, it documents the efficacy of voice treatment in a population where no efficacious speech or voice treatment previously has been identified. This voice treatment approach should seriously be considered for PPS patients well before augmentative communication devices are indicated.

Acknowledgments

The authors gratefully acknowledge the efforts of the patients who participated in this study for their courage and continuous challenges. We gratefully acknowledge the contributions of Dr. Marshall Smith, Dr. Laetitia Thompson, Dr. Christopher O'Brien, Dr. Lawrence Brown, William Winholtz and Christopher Dromey.

The research presented here was supported in part by NIH Grant no. R01 DC01150 and The National Center for Voice and Speech Grant no. P60 DC00976.

Bibliography

Agid, Y., Javoy-Agid, F., & Ruberg, M., et al. (1986). Progressive Supranuclear Palsy: Anatomical and biochemical considerations. *Advances in Neurology*, 45, 191-206.

Aronson, A., Ramig, L. O., Winholtz, W., & Silber, S. (1992). Rapid voice tremor or "flutter" in amyotrophic lateral sclerosis. *Annals of Otology , Rhinology, and Laryngology*, 101(6), 511-518.

Bassich, C.J., Ludlow, C.L., & Polinsky, R.J. (1984). Speech symptoms associated with early signs of Shy-Drager syndrome. *Journal of Neurology, Neurosurgery, and Psychiatry*, 47, 995-1001.

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.

Beneckel, R., Rothwell, J. C., Dick, J. R., Day, B. L., & Marsden, C. D. (1986). Performance of simultaneous movements in patients with Parkinson 's disease. *Brain*, 109, 739-757.

Benton, A.L., Hannay, H.J., & Varney, N.R. (1975). Visual perception of line direction in patients with unilateral brain disease. *Neurology*, 25, 907-910.

Bergner, M., Bobbit, R.A., Carter, W.B., & Gilson, B.S. (1981). The sickness impact profile: Development and final revision of a health status measure. *Medical Care*, 19, 787-805.

Black, I.B. (1982). Idiopathic autonomic insufficiency (idiopathic orthostatic hypotension, Shy-Drager syndrome). In Cecil (Ed.),*Textbook of Medicine*, (pp. 2034-2035). Philadelphia, PA: WB Saunders Co.

Boeckstyns, M.E., & Backer, M. (1989). Reliability and validity of the evaluation of pain in patients with total knee replacement. *Pain*, 38(1), 29-33.

Boone, D. R. (1977). *The Voice and Voice therapy*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Briskin, J.G., Lehrman, K.L., & Guilleminault, C. (1978). Shy-Drager syndrome and sleep apnea. In: Guilleminault & Dement (Eds.), *Sleep apnea syndromes* (pp. 317-322) New York: Alan R. Liss.

Beukelman, D. R. & Yorkston, K. M. (1980). Influence of passage familiarity on intelligibility estimates of dysarthric speech. *Journal of Communication Disorders*, 13, 33-41.

Countryman, S., & Ramig, L.O. (1993). Effects of intensive voice therapy on speech deficits associated with bilateral thalamotomy in Parkinson's disease: A case study. *Journal of Medical Speech Pathology*, 1(4), 233-249.

Duvosin, R. C. (1984). *Parkinson's disease: A guide for patient and family*. New York: Raven Press.

Duvosin, R.C., Golbe, L.I., & Lepore, F.E. (1987). Progressive supranuclear palsy. *The Canadian Journal of Neurological Science*, 14, 547-554.

Fairbanks, G. (1960). *Voice and articulation drill book*. New York: Harper & Brothers.

Forno (1982). *Pathology of Parkinson's disease. Movement Disorders*.. In C.D. Marsden, & S. Fahn (Eds.) (pp. 25-40). London: Buttersworth.

Graham, J.G., & Openheimer, D.R. (1969). Orthostatic hypotension and nicotine sensitivity in a case of multiple system atrophy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 32, 28-34.

Hanson, D.G., Ludlow, C.L., & Bassich, C.J. (1983). Vocal cord paresis in Shy-Drager syndrome. *Annals in Otology, Rhinology and Laryngology*, 92, 85-90.

Heaton, R.K., Grant, I., & Matthews, C.G. (1991). *Comprehensive norms for an expanded Halsted-Reitan Battery.* Odessa, FL: Psychological Assessment Resources, Inc.

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

Jackson, J.A., Jancovic, J., & Ford, J. (1983). Progressive supranuclear palsy: Clinical features and response to treatment in 16 patients. *Annals of Neurology*, 13, 273-278.

Jacob, L.A. (1968). A normative study of laryngeal jitter. Unpublished master's thesis. University of Kansas.

Jancovic, J. (1989a). Parkinson's plus syndromes. *Movement Disorders*, 4, S95.

Jancovic, J. (1989b). The relationship between Parkinson's disease and other movement disorders. In D.B. Calne (Ed.), *Handbook of Experimental Pharmacology* (p. 227), vol. 88, Berlin: Springer-Verlag.

Jellinger, K. (1987). The pathology of parkinsonism. In: C.D. Marsden & S. Fahn (Eds.), *Movement Disorders 2* (pp. 124-165) London: Butterworths.

Kedas, A., Reed, M. L., Lux, W. E. (1989) Parkinson's Mime. *Geriatric Nursing*, 10(4), 182-183.

Kent, R.D., Weismer, G., Kent, J., & Rosenbek, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, 54, 482-499.

Keww, J., Gross, M., & Chapman, P. (1990). Shy-Drager syndrome presenting as isolated paralysis of vocal cord abductors. *The British Medical Journal*, 300, 1441.

Kempster, G. B. (1984). A multidimensional analysis of vocal quality in two dysphonic groups. An unpublished doctoral dissertation, Northwestern University, Evanston, IL.

King, J.B., Ramig, L.O., Lemke, J.H., & Horii, Y. (1994). Variability in acoustic and perceptual parameters of phonation in patients with Parkinson's disease. *Journal of Medical Speech-Language Pathology*, in press.

Kluin, K.J., Foster, N.L., Berent, S., & Gilman, S. (1993). Perceptual analysis of speech disorders in progressive supranuclear palsy. *Neurology*, 43, 563-566.

Kwentus, J.A., Auth, T.L., & Foy, J.L. (1984). Shy-Drager syndrome presenting as depression: Case report. *Journal of Clinical Psychiatry*, 45, 137-139.

Linebaugh, C. (1979). The dysarthrias of Shy-Drager syndrome. *Journal of Speech and Hearing Disorders*, 44, 55-60.

Linville, S. E., & Korabic, E. (1987). Fundamental frequency characteristics of elderly women's voices. *Journal of Acoustical Society of America*, 81, 1196-1199.

Loranger, A.W., Goodell, H., & Lee, J.E. (1972). Levadopa treatment of Parkinson's syndrome. *Archives of General Psychiatry*, 26, 163-168.

Maclay, S., Ramig, L.O., Scherer, R., & Jancosek, E. (1994). Speech intelligibility in Parkinson's disease following intensive voice therapy. *Clinical Linguistics and Phonetics*, (in submission).

Martinovits, G., Leventon, G., Goldhammer, Y., & Sadeh, M. (1988). Vocal cord paralysis as a presenting sign in the Shy-Drager syndrome. *The Journal of Laryngology and Otology*, 102, 280-281.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1971). *EITS Manual for the Profile of Mood States.* San Diego, CA: Educational Testing Service.

McReynolds, L.V., & Kearns, K.P. (1983). *Single subject experimental design in communication disorders.* Baltimore: University Park.

Metter, E. J., & Hanson, W. R. (1991). Dysarthria in Progressive Supranuclear Palsy. In C. A. Moore & K. M. Yorkston (Eds.), *Dysarthria and Apraxia of Speech: Perspective on Management* (pp. 127-136) Baltimore: Paul H. Brookes Publishing Co.

Milenkovic, P. (1987). Least mean square measures of voice preturbation. *Journal of Speech and Hearing Research*, 30, 529-538.

Montgomery, S.A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, 134, 382-389.

Otoxby, M. (1982). *Parkinson's disease patients and their social needs*. London: Parkinson's Disease Society.

Pollack, M., & Hornabrook, R.N. (1966). The prevalence, natural history and dementia of Parkinson's disease. *Brain*, 89, 429-228.

Polinsky, R.J. (1984). Multiple system atrophy. *Neurological Clinics*, 2, 487-498.

Ptacek, P. H., Sander, E.K., Maloney, W.H., & Jackson, C.R. (1966). Phonatory andrelated changes with advanced age. *Journal of Speech and Hearing Research*, 9, 353-360.

Quinn, N. (1989). Multiple system atrophy-the nature of the beast. *Journal of Neurology, Neurosurgery, and Psychiatry*, 52 (suppl), 78-89.

Ramig, L.O., (1992). The role of phonation in speech intelligibility: a review and preliminary data from patients with Parkinson's disease. In R. D. Kent (Ed.) *Intelligibility in Speech Disorders: Theory, Measurement and Management* (pp.119-156). Amsterdam: John Benjamin.

Ramig, L. O. (1994). Speech Therapy for Parkinson's Disease. In W. Koller & G. Paulson (Eds.) *Therapy of Parkinson's Disease*. New York: Marcel Dekker.

Ramig, L.O., Bonitati, C., Lemke, J., & Horii, Y. (1994). The efficacy of voice therapy for patients with Parkinson's disease. *Journal of Medical Speech Pathology* (in submission).

Ramig, L.O., Countryman, S., Winholtz, W., Horii, Y., Thompson, L. (1994). Intensive voice treatment for Parkinson's disease. Unpublished manuscript.

Ramig, L.O., Horii, Y., & Bonitati, C. M. (1991). The efficacy of voice therapy for patients with Parkinson's disease. *National Center for Voice and Speech Status and Progress Report*, 1, 61-86.

Ramig, L.O., Mead, C., DeSanto, L. S., & Horii, Y. (1988). Voice therapy and Parkinson's disease. *Journal of the American Speech and Hearing Association*, 30(10), 128.

Ramig, L.O., Mead, C., Scherer, R., Larson, K., & Kohler, D. (1988, February). *Voice therapy and Parkinson's disease: A longitudinal study of efficacy*. Paper presented at the Clinical Dysarthria Conference, San Diego, CA.

Reitan, R.M., & Wolfson, D. (1985). The Halstead-Reitan Neuropsychological Test Battery. Tucson, Arizona: Neuropsychology Press.

Rivest, J., Quinn, N., & Mardsen, C.D. (1990). Dystonia in Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy. *Neurology*, 40, 1571-1578.

Rizzoli, A.A. (1986). Psychiatric disturbances in the Shy-Drager syndrome (letter). *British Journal of Psychiatry*, 148, 484.

Scott, S., Caird, F., & Williams, B. (1984). Evidence of an apparent sensory speech disorder in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 47, 840-843.

Smith, M., Ramig, L.O., Dromey, C., Perez, K., and Samandari, R. (1994). Intensive voice treatment in Parkinson's disease: laryngostroboscopic findings. *Journal of Voice*, (in submission).

Stacy, M., & Jancovic, J. (1992). Differential diagnosis of Parkinson's disease and the Parkinsonism plus syndromes. *Parkinson's Disease*, 10, 341-359.

Stadlan, E.M., Duvoisin, R.C., & Yahr, M.D. (1966). The pathology of parkinsonism. In: Proceedings of the Vth International Congress of Neuropathologists. Amsterdam: *Excerpta Medica*, 569-571.

Steele, J.C., Richardson, J.C., & Olszewski, J. (1964). Progressive supranuclear palsy. *Archives of Neurology*, 10, 333-359.

Streifler, M., & Hofman, S. (1984). Disorders of verbal expression in Parkinsonism. In R.G. Hassler & J.F. Christ (Eds.), *Advances in Neurology*, 40, 385-393.

Stoicheff, M. L., (1981). Speaking fundamental frequency characteristics of non-smoking female adults. *Journal of Speech and Hearing Research*, 24, 437-441.

Thomas, J.E., & Schirger, A. (1970). Idiopathic orthostatic hypotension. A study of it's natural history in 57 neurologically affected patients. *Archives of Neurology*, 22, 289-293.

Titze & Winholtz (1994). Effect of microphone type and placement on voice perturbation measurements. *Journal of Speech and Hearing Research*, 36, 1177-1190.

Tygstrup, I., & Norholm, T. (1963). Neuropathological findings in 12 patients operated for Parkinsonism. *Acta Neurology Scandinavia*, 39 (Suppl 4), 188-195.

Williams, A., Hanson, D., & Calne, D.B. (1979). Vocal cord paralysis in the Shy-Drager syndrome. *Journal of Neurology, Neurosurgery and Psychiatry*, 42, 151-153.

Winholtz, W., & Ramig, L.O. (1992). Vocal tremor analysis with the Vocal Demodulator. *Journal of Speech and Hearing Research*, 35(3), 562-573.

Yanagihara, Koike, & Leden (1966). Phonation and respiration. Function study in normal subjects. *Folia Phoniatrica*, 18, 323-340.

Yanagisawa, N., Fujimato, S., & Tamaru, F. (1989). Bradykinesia in Parkinson's disease: Disorders of onset and execution of fast movement. *European Neurology* (suppl.), 29, 19-28.

Yorkston, K.M., & Beukelman, D.R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders*, 11, 499-512.

Yorkston, K.M., Bombardier, C., & Hammen, V. (1993) Dysarthria from the viewpoint of dysarthric individuals. In J. Till, K. Yorkston, & D. Beukelman (Eds.), *Motor speech Disorders: Advances in assessment and treatment* (PP. 19-36). Baltimore: Paul Brookes.

NCVS Status and Progress Report - 6
May, 1994, 113-125

Speech Intelligibility in Parkinson's Disease Patients Following Intensive Voice Therapy

Sharon Maclay, M.A.
Fairfax, Virginia
Lorraine Olson Ramig, Ph.D., CCC-SP
Department of Communication Disorders and Speech Science, The University of Colorado-Boulder
and
Wilbur James Gould Voice Research Center, The Denver Center For The Performing Arts
Ronald C. Scherer, Ph.D.
Wilbur James Gould Voice Research Center, The Denver Center For The Performing Arts
Elizabeth Jancosek, Ph.D.
Department of Communication Disorders and Speech Science, The University of Colorado-Boulder

Abstract

This study implements two published intelligibility tests to measure the gains in intelligibility made by Parkinson's disease patients after a regimen of intensive vocal therapy. In the first study, ten subjects with idiopathic Parkinson's disease were rated using a paired-word intelligibility test. Results showed that this test was too simple to document subtle intelligibility changes in these patients resulting from voice therapy. Notwithstanding, several error categories emerged as being vulnerable to distortion in the Parkinson's disease patients tested. A second, more phonemically challenging published test was implemented for Experiment 2. This test incorporated the frequently-distorted error categories of the first study such as glottal versus null, initial voicing, and fricative versus nasal distortions, as well as other categories historically of interest, in a more difficult corpus of test words. Both speech pathologists familiar with dysarthric speech and naive professionals were included as listeners in Experiment 2. Although the second test highlighted more errors than the first, statistical analysis showed that the structure of the second test introduced confounds which seriously impacted its validity. The greatest source of error was determined to be due to the introduction of test words of interest without corresponding test foil words. After repeated scorings, both naive and informed listeners made predictions based on former presentations. Recommendations are made regarding the structure of a future intelligibility test to avoid the noted confounds.

The improvement of speech intelligibility is the cornerstone of clinical intervention. In many cases, it is the patient's sole criterion for judgment of success or failure of his or her therapy. Evaluation of intelligibility improvement must be an integral part of the assessment of the efficacy of speech treatment. An accurate and reliable intelligibility measure should objectively ascertain pre-therapy intelligibility and substantiate claims of improvement after treatment.

Intelligibility assessment for communication disorders has been addressed primarily in literature related to speakers with hearing impairment (Markides, 1983; McGarr, 1983; Mencke, Ochsner, & Testut, 1983; Metz, Schiavetti, & Sitler, 1980; Monsen, 1978; Samar & Metz, 1988) and dysarthria (Ansel & Kent, 1992; Beukelman & Yorkston, 1980; Cullinan, Brown, & Blalock, 1986; Kent, Weismer, Suffit, Rosenbek, Martin & Brooks, 1990; Ramig, 1992; Sheard, Adams, & Davis, 1991; Yorkston & Beukelman, 1980). Historically, two types of intelligibility testing have

been proposed. One requires the use of interval rating scales, with a listener appraising intelligibility on a continuum from unintelligible to intelligible. The other involves the listener either writing down what he heard or choosing what he heard from a closed set of possibilities. For clinical purposes, rating scales have an advantage in that they are easily administered and compiled. However, the lack of validity of rating scales continues to pose problems (Samar & Metz, 1988). In several studies the assertion has been made that tests requiring the listener to show an objective understanding of the test word, either by writing it down or selecting it from a choice list, yield more valid and reliable assessments (Metz, Schiavetti, & Sitler, 1980; Samar & Metz, 1988; Yorkston & Beukelman, 1978; and Yorkston & Beukelman 1980). Kent, Weismer, Kent and Rosenbek (1989) proposed a closed-set intelligibility test based upon minimal contrasts for assessing intelligibility and providing for underlying acoustic-correlate interpretation in dysarthric patients. Their study involved amyotrophic lateral sclerosis patients, and implemented a test word list with three foils for each test word.

Another dysarthric population widely recognized as having reduced intelligibility is patients with Parkinson's disease. Recently a speech treatment was developed for Parkinson's disease by Ramig and colleagues (Ramig, Mead, Scherer, Larson & Kohler, 1988; Ramig, Horii & Bonitati, 1991; Ramig, 1992). This treatment, the Lee Silverman Voice Treatment (LSVT), focuses on increasing phonatory effort to enhance output of the entire speech production system. The goal of the LSVT is to address the underlying pathophysiological concerns (such as bowed vocal folds and rigidity, hypokinesia, and tremor in the laryngeal and respiratory muscles of Parkinson's disease patients) to improve perceptual voice characteristics. Specific exercises geared toward improving the muscular and aerodynamic sources of voice may generalize to the entire speech system, thus increasing overall competence and intelligibility. One of the primary objectives of the LSVT is to increase speech intensity. This component of improvement has been underemphasized in other intelligibility studies, where procedures to normalize speaker parameters (Ellis & Fucci, 1991) or to present speakers at a comfortable listening level (Kent et al., 1989; Sheard et al., 1991) have historically lead to the obscuring of this key variable. The study reported here was undertaken for three purposes: to evaluate the usefulness of the method of dysarthric intelligibility assessment proposed by Kent et al. (1989) applied to Parkinson's disease patients, to modify the method of intelligibility assessment as necessary to specifically address Parkinson's disease concerns, and to assess the impact of the Lee Silverman Voice Treatment on the speech intelligibility of patients with Parkinson's disease.

Experiment 1

In Kent et al. (1989), two word lists of differing levels of phonemic difficulty were presented for use in intelligibility measurement. In the present study, Experiment 1 was designed to evaluate which of these two lists would be more appropriate for the Parkinson' disease patients under consideration, and to note response possibilities that might be refined to specifically address Parkinson's disease patient intelligibility testing.

Intelligibility Test Development

When discussing the concept of intelligibility, it is important to distinguish the difference between articulatory difficulties and their effects on intelligibility. A constellation of misarticulations cannot be said to be directly related to intelligibility; a listener may be able to understand even a severely impaired speaker's utterances if the errors are consistent and predictable (Peterson & Marquardt, 1981). Therefore, it is important to devise a functional intelligibility test which allows for the difference between purely articulatory incompetence and an intelligibility breakdown. A well-conceived test would objectively document individual Parkinson's disease patients' pre-therapy intelligibility, substantiate claims of improvement after treatment, and elucidate the physiological correlates of this improvement.

A paradigm using a closed set of minimal contrast alternatives, rather than a pure write-down procedure, has the benefit of allowing for both intelligibility quantification and correlated acoustic analysis. For instance, if the test word "pea" is mistaken for "bee", acoustic analysis will help infer the physical aspects of the speech signal which led the listener to decide upon the initial /b/ rather than /p. Many studies have shown that perceptual and acoustic data can be analyzed in tandem to provide a comprehensive picture of speech output, including clues as to the underlying deficits in the speech mechanism (Forrest, Weismer & Turner, 1989; Kent et al., 1989; Kent et al., 1990; Ludlow & Bassich, 1983, 1984; Monsen, 1978; Weismer, Kent, Hodge, & Martin, 1988; Weismer, Martin, Kent & Kent, 1992; Yorkston, Beukelman, Minifie, & Sapir, 1984; Ziegler et al., 1988).

In 1989, Kent, Weismer, Kent and Rosenbek outlined two closed-set word lists based upon minimal contrasts for assessing intelligibility and providing for acoustic-correlate interpretation. The first, a phonemically less challenging test comprised of primarily open syllables, includes forty-eight words minimally paired across different phoneme contrasts. Each word is presented once, so that each word in each pair is targeted as a stimulus word. The second test is composed of seventy stimulus words. These words have also been chosen to test phonemic errors common in dysarthrics, but are more difficult in that they

include more challenging, closed syllable words such as "likes" and "blends". Instead of a choice of two, as in the first list, this test asks the listener to choose from among four minimally paired alternatives. The two different tests are useful in that the examiner can select the list which more closely fits the articulation ability of the population of interest. For instance, in amyotrophic lateral sclerosis (ALS) patients, the less challenging test can be useful for severely involved patients. If, however, the test words are too easily produced, results will not be as discerning and may prevent the opportunity to document subtle impairments. It is for this reason that Experiment 1 was conducted to ascertain which word list would more closely fit the ability levels of the Parkinson's disease patients under observation. It was hypothesized that the less challenging list would prove to be fairly easily articulated by most of the patients involved.

Method

Subjects

Subjects in Experiment 1 consisted of ten patients from the Lee Silverman Voice Treatment for Parkinson's Disease program at The Wilbur J. Gould Voice Research Center of The Denver Center For The Performing Arts (DCPA). All patients were subjected to neurological, psychological and otolaryngological examinations before commencement of the therapy. Nine patients were diagnosed as having idiopathic Parkinson's disease (IPD), and one patient had neurological symptoms of uncertain origin.

Table 1.
Parkinson's Disease Patient Profile - Experiment 1

Patient #	Age	Sex	Diagnosis	PD-related Medication	Years Since Diagnosis	Primary Speech Complaints
1	50	M	IPD Stage II	Sinemet Eldepryl Parlodel	19	slurring
2	65	M	IPD Stage III	Sinemet Artane Symmetrel	9	hesitation, reduced rate
3	69	M	IPD Stage III	none	2	hoarseness, reduced volume
4	73	M	IPD Stage III	Sinemet Artane	11	reduced volume
5	63	F	n.a.	n.a.	3	slurring
6	82	M	IPD Stage II	Sinemet Eldepryl	3	unclear speech
7	62	M	IPD Stage III-IV	Sinemet Deprynel Symmetrel	5	reduced volume, raspiness
8	61	M	IPD Stage III	Sinemet Parlodel Eldepryl	5	slurring, increased rate
9	65	F	IPD Stage III Bilateral Thalamotomy	none	7	reduced volume, slurring, slow rate
10	52	M	IPD Stage IV	Symmetrel Sinemet Eldepryl Parlodel	19	slurring, mumbling, stuttering

Patients ranged in age from 50 to 82 years old, with an average age of 64.2 years. Patient characteristics are summarized in Table 1.

Speaker Task

A videotaped recording was made of the forty-eight minimally paired words suggested by Kent et al. (1989). (See Table 2). The format consisted of a contrasting word pair at the top of the screen, followed by the stimulus word centered below:

bee pea
pea

This configuration was chosen because Parkinson's disease performance is known to be highly effort-dependent (England & Schwab, 1959; Hallet & Khosbin, 1980; McDowell, Lee & Sweet, 1986). It was rationalized that demonstrating the goal contrast would maximize patient effort toward achieving the target phoneme. Each word of each pair served as a stimulus word for a total of ninety-six elicitations. Consisting for the most part of single syllable consonant-vowel (CV) or vowel-consonant (VC) constructs, the words were relatively easily produced by the subjects.

In pilot trials, patients were asked to say the pair out loud, then to pause and say the third word again separately and "as clearly as possible." It was determined that this method confounded the production of the stimulus word, largely because of the patients' tendency to phrase the three words in a sentence-like utterance. Additionally, it was frequently noted in pilot work that during the listeners' task of intelligibility scoring, the dual elicitation presented a contrast useful for word discrimination. Therefore, in Experiment 1, patients were instructed to read aloud only the centered stimulus word. Because no treatment effect was being investigated, only words elicited pre-treatment were included in this portion of the study.

Table 2.
Word Pairs for Experiment 1 Categorized by Feature Contrast

Category of Feature	Pair 1	Pair 2	Pair 3
Initial voicing	bee-pea	do-two	goo-coo
Final voicing	add-at	buzz-bus	need-neat
Vowel duration	eat-it	gas-guess	pop-pup
Stop vs. fricative	see-tea	sew-toe	do-zoo
Glottal vs. null	high-eye	hit-it	has-as
Fricative vs. affricate	shoe-chew	shop-chop	ship-chip
Stop vs. nasal	dough-no	bee-me	buy-my
Alveolar vs. palatal	see-she	sew-show	sip-ship
Tongue height	eat-at	soup-soap	eat-eight
Tongue advancement	hat-hot	tea-two	day-dough
Stop place	pan-can	dough-go	bow-go
Diphthong	buy-boy	high-how	aisle-oil
/r/ vs. /l/	ray-lay	rip-lip	raw-law
/w/ vs. /r/	way-ray	row-woe	won-run
Liquid vs. vowel	string-stirring	spring-spurring	bring-burring
Cluster w/intrusive vowel	blow-below	plight-polite	claps-collapse

Recording Procedure

All recording was done in an IAC sound booth at the DCPA. Patients were scheduled for two sessions, usually on continuous days. Care was taken to time the sessions for equal intervals post-medication to control for any related fluctuations in performance. During these sessions, a battery of tests was conducted, including respiratory, phonatory, and articulatory tasks. The tests specific to this study were carried out near the end of each session.

Acoustic recordings of patients' readings were made at a constant mouth-to-microphone distance of eight centimeters using a head-mounted microphone (AKG-410), and recorded on one channel of a Sony PC-108M digital audio tape recorder. Sound pressure level measurements at a distance of 50 cm [Bruel & Kjaer Type 2230 Precision Integrating Sound Level Meter (SLM)] were recorded on the digital audio tape recorder. Intensity levels were also manually noted from the SLM data collection.

Listener Task

Audio tapes of the patients' responses were presented via earphones to an informed listener (a graduate student in speech science) who was given a list of two or three minimally-paired possibilities from which to identify each stimulus word. Because no treatment effect was being investigated, the listener was free to manipulate the headphone volume as needed to achieve a comfortable listening level. Word presentation occurred at approximately ten-second intervals. The listener was instructed to highlight his choice from the possibilities given. In any case, when this listener heard a word that was not listed, he was instructed to choose the best alternative, as well as to write in the word he thought he heard.

Results and Discussion

Three out of ten patients attained intelligibility scores of one hundred percent, and five others were rated higher than ninety-two percent, for an average intelligibility score of 97.1% across the ten speakers. This outcome conflicted with that expected from listening to the conversational speech of the patients. In 1965, Canter observed that some Parkinsonian patients performed normally on simple single-word articulation tests although considerable articulation difficulties were evident in more complex conversational tasks. Similarly, it was concluded that this test did not challenge the Parkinson's disease patients involved. However, several trends were noted. The results by phonetic feature category are summarized in Table 3.

By far, most errors (13 out of 30 occurrences) were made in the glottal versus null comparisons (e.g., high-eye). Less frequently, but still worthy of note, were confusions in initial voicing. Other errors were made in the fricative vs. affricate, stop vs. nasal, /w/ vs. /r/, liquid vs. vowel, and cluster vs. cluster + intrusive vowel categories. Additionally, it was observed that the patients had difficulty with the initial voicing contrastive pair of /dz /versus /ts/, as evidenced by the fact that "gyp" was written in for "chip" on several occasions. These errors were recorded in order to assure that these types of words were included in the Experiment 2. Additionally, it was noticed that no confusions between vowels (e.g., "ham" for "him") were registered in this study.

The conclusion of this preliminary study was that this intelligibility word list was too easy for the Parkinson's disease patients involved, and thus did not provide opportunities to document subtle impairments.

Table 3.
Error Responses (by Feature Category) of Ten Parkinson's Disease Patients in Experiment 1

Glottal vs. Null	13
Initial Voicing	5
Cluster + Intrusive Vowel	2
Liquid vs. Vowel	2
/w/ vs. /r/	1
Stop vs. Nasal	1
Fricative vs. Affricate	1

Experiment 2

The second experiment was undertaken to establish an intelligibility test incorporating the results of Experiment 1, as well as to evaluate the effectiveness of the Lee Silverman Voice Treatment on speech intelligibility in patients with Parkinson's disease.

Intelligibility Test Development

Due to the conclusion of Experiment 1 that the more easily produced list was too simple for the Parkinson's disease patients, it was logical to administer the second, more phonemically challenging word list developed by Kent et al. (1989) for this phase of the study. However, it was first necessary to address several considerations of the more difficult word list and their impact on the test used in this experiment. The words are more complex than the simple CV and VC words of the first list. Words such as "bunch" and "shoot" are included to test clusters and closed syllables. As in the first test, all the syllables included are actual (rather than nonsense) words to preserve a minimal pragmatic component necessary to make this a functional intelligibility test rather than a pure articulatory exercise. (The complete list of test words can be found in the Appendix.)

Kent et al. (1989) concur with Yorkston and Beukelman (1980) that random selection from a list of test words is preferable over an unchanging stimulus set because this procedure "gives the potential for several equivalent lists that can be used for repeated assessments in the individual..." (Kent et al., 1989). However, analysis of their list leads to the realization that the words do not satisfy this criterion. The words do not fall into an interchangeable matrix where any one of several words can be chosen as a stimulus word, because each is only minimally paired across one phoneme at a time. For example, consider the stimulus word "bad", with foils "bed", "bat", and "pad". "Bad" is the test word, but no other word in the set can be a stimulus word because it varies from the others by more than one phoneme, e.g., "bed" and "pad" vary from each other by both initial voicing and vowel duration. Therefore, the only way the test can be changed for subsequent assessments is by changing the choice word order, not the stimulus words themselves. If a truly random list was desired, the matrix for "bed" as a test word could include for example, "bet", "bad", and "bid". This action would result in an unwieldy list, however.

A more important concern was how re-presentation would affect familiarity for the listeners. In the present experiment, two tests using the same test words in different orders were developed. Because of the similarity of the test words, it was hypothesized that the two-test option would preclude retest confounds.

Adjustments in the following categories were also made:

initial voicing: the contrasting pairs of "chip-gyp" and "cheer-jeer" were added to quantify a contrast difficulty observed in the earlier study.

stop-fricative: two contrasts are included in the original "hand-pant" pair - the /h/ vs. /p/ and /d/ vs. /t/. Changing the second word to "panned" alleviates this concern.

stop-affricate: /tir/ and /tɛr/ are both spelled "tear". Changing the spelling to tier for the former is clearer. (The measured contrasts are between "chair-tear" and "cheer-tier".)

initial consonant vs. null: "blend-end" was included in this category, but "blend" is actually an initial cluster. It would be more correct to list "bend-end" and "lend-end" in this category and leave "blend" for the initial cluster vs. singleton contrast category.

final consonant-null: "bun-bunch" was included in this category, but a reclassification into the final cluster vs. singleton is more appropriate.

/r/ vs. /l/: "wax" vs. "lax" was listed in this category, although it is a /w/ vs. /l/ contrast. Testing "lax" against both "racks" and "wax" will keep these words in established categories.

/r/ vs. /w/: "rock" vs. "walk" is listed as only one phonemic contrast. Dialectically, "walk" can have a more rounded vowel coloring than does "rock." For these situations, "wok" can be considered a purer contrast to the stimulus word "rock." Both "wok" and "lock" were listed as choice alternatives to assess the initial target phoneme without allowing the vowel confound potentially found in "walk".

Method

Subjects

Subjects in Experiment 2 consisted of five patients drawn from the Lee Silverman Voice Treatment for Parkinson's Disease program at the DCPA. Four of the five were diagnosed as having idiopathic Parkinson's disease, the other with Parkinsonian symptoms of uncertain origin. All participated in neurological, psychological, and otolaryngological exams before commencement of therapy.

For a profile of the patients selected for Experiment 2, see Table 4. Ages of the patients fell within a range of 64 to 77, with an average age of 71.2 years. All complained of decreased vocal volume, which was their prime motivation for seeking voice therapy.

Table 4.
Parkinson's Disease Patient Profile - Experiment 2

Patient #	Age	Sex	Diagnosis	PD-Related Medication	Years Since Diagnosis	Primary Speech Complaints
1	71	M	IPD-Stage III	Eldepryl	1	soft voice, hoarse
2	72	M	IPD-Stage II	Sinemet	5	slurring, huskiness, power and volume loss
3	64	F	Progressive akinetic rigid syndrome w/ Parkinsonism	Sinemet	4	soft voice, slurring, mild stuttering
4	72	M	IPD-Stage III	Sinemet	6	soft voice
5	77	M	IPD-Stage III-IV	Sinemet	17	soft voice

Speaker Task

In this task, patients were asked to read a word "as clearly as possible" from a videotaped presentation. Two different videotapes were presented. Each included the same 70 test words, but in adifferent random order (Test Order 1 and 2). The word list was composed the more phonologically difficult list proposed by Kent et al. (1989), in combination with the new words suggested by the preliminary study. Words were presented one at a time in easily seen white letters on a blue background at approximately five-second intervals.

Both Test Order 1 and 2 were presented both pre- and post-treatment. Either Test Order 1 or 2 was randomly chosen for administration on the first pre-treatment recording session, then the other was given on the second pre-treatment session. The same procedure was followed for

the post-treatment sessions. Each videotape presentation was approximately six and a half minutes in length.

All patients completed a one-month sixteen-session course of the Lee Silverman Voice Treatment, which incorporates phonatory and respiratory exercises to enhance phonatory effort and vocal loudness. Details of the treatment have been summarized elsewhere (Ramig et al., 1991; Ramig, in press). Following this treatment program, all patients participated in post-treatment recordings at the DCPA. The task for assessment of speech intelligibility was the same as pre-treatment. Patients were asked to read each word "as clearly as possible" from the videotape presentation. This yielded a body of two pre-treatment and two post-treatment tests for each patient.

Recording Procedure

Recording methods were identical to those in Experiment 1.

Preparation of Listening Tapes

Audio tapes of one pre-treatment and one post-treatment test order were prepared from these samples for presentation to listeners. Particular test orders were selected for use only on the basis of completeness. If Test Order 1 was used on the pre-treatment sample, then Test Order 2 was used on the post-treatment sample. Thus, the first dubbed tape consisted of each patient speaking a complete list of words pre-treatment, then a complete list of words from the opposite post-treatment. In addition, each list was started randomly on the dubbed listener tape, with the balance of the seventy-word list then added to the end. This assured that no listener would come to know the first word of each list. A second tape was then dubbed to also include re-presentations of fourteen words (20% of the sample) to measure reliability of the judges. A third tape was also generated to randomize the patients so they were no longer in pre-treatment, post-treatment order.

A critical variable in assessment of speech intelligibility is intensity (Ramig, 1992). Often in listening tasks of this nature, presentation intensity is normalized to a set intensity or presented "at a comfortable listening level" (Kent et al., 1989) to compensate for possible differences in elicitation and learning levels. Some previously reported listening studies make no mention of the presentation intensity (Ansel & Kent, 1992; Samar & Metz, 1988). Others mention presenting at a comfortable listening level (Kent et al., 1989; Sheard et al., 1991). A recent study by Ellis and Fucci (1991) stipulated presentation at 70dB SPL, regardless of elicitation level. However, none of these procedures is entirely appropriate in the present experimental situation. One of the main underpinnings of speech intelligibility is adequate intensity. The ability to generate intensity declines with the reduction in inspiratory and expiratory volume and increased bowing of the vocal folds associated with Parkinson's disease progression (Critchley, 1981; Hansen et al. 1984). Since the current therapy (Ramig, in press) seeks to increase intensity, normalizing the intensity on the listener presentation tapes would confound one of the main effects of the therapy. For this reason, a compromise solution was developed. A fourth tape was made which attended to volume differences of the original tests. This was done by measuring the maximum intensity on each patient's first word using computer analysis of the digitized signal (Guo & Winholtz, 1993), then dubbing the tape by setting record levels to reflect the differences between patient's volumes. To preserve the audibility of all the tapes, the patient exhibiting the lowest volume was raised from 68 dB SPL at (50 cm) to 75 dB SPL. All other patients' intensities were similarly raised 7 dB for the final listener presentation tape. In this way the actual intensity difference among the patients was preserved.

Listener Task

Two groups of five listeners participated. One group was composed of naive listeners who had had no exposure to Parkinson's disease and worked in non-related areas. All had earned master's degrees from varying fields: business administration, biology, and three in aerospace engineering. The naive group included three females and two males. The other group was made up of speech-language pathologists with experience listening to dysarthric speech. They also had master's degrees and the certificate of clinical competence (from ASHA). This group included four females and one male. All listeners passed hearing screenings.

The listeners were asked to listen to the presentation tapes, then highlight the word they thought they heard from a list of four alternatives. All listeners were seated at a constant distance of four feet from the speaker. Choice lists were fully randomized; no alternative grouping was in the same order either within one test or across tests. Immediately following each word choice, listeners were asked to highlight a number from one to five which corresponded to their confidence that the word they had chosen was in fact the word the patient had uttered. The following instructions were given to each participant:

> You will hear a series of words. Your job is to listen to each carefully, then highlight the printed word you think the patient said. When you have done that, highlight the number to the right which corresponds to how confident you are in your answer. The numbers are scaled so that five is the most confident, and one is the least confident. Five corresponds to "I am absolutely sure my answer is correct". Four is "I am nearly sure my answer is correct". Three - "I am more or less sure my answer is correct". Two - "I am not so sure my answer is correct", and One - "I am not at all sure my answer is correct". Reminders are printed at the bottom of your answer sheets. You will only hear each word once. These words move along quickly, so try not to get behind thinking over the confidence levels. If at any time you get lost, stop me and we'll note where, then re-do the words you missed at the end of the session. Let's do a few practice words.

Four practice words from a tape of a patient who did not participate in the study were then administered.

The use of confidence levels was based upon the work of Costermans, Lories, and Ansay (1992), and measures the "feeling of knowing". Their applications tend more toward comprehension or memory retrieval tasks, rather than judgment tasks, but the option list provides a satisfactory means for measuring the conviction in listener choice. Gauging this confidence provides another yardstick of intelligibility improvement. Even if the raw intelligibility scores do not improve with therapy, if the listener confidence levels tend to increase, improvement might be inferred.

Feature contrasts and error possibilities for both the Kent et al. (1989) list and the intelligibility list generated for Experiment 2 are summarized in Table 5.

Results and Discussion

This section will address the following results of Experiment 2: reliability of the listeners, the performance of the naive versus informed listeners, measured improvements from pre- to post-treatment, the contribution of confidence levels, and considerations regarding the intelligibility test itself.

Table 5.
Feature Contrasts and Error Possibilities in Experiment 2

	Number of Error Possibilities	
Contrast Category	Kent et al. (1989) List	Current List
front - back vowel contrast	11	11
high - low vowel contrast	12	13
vowel duration contrast	11	11
voicing contrast - initial	9	11
voicing contrast - final	11	12
alveolar - palatal	7	9
consonant place	9	8
fricative - affricate	9	9
other fricative	16	16
stop - fricative	19	15
stop - affricate	6	7
stop - nasal	9	10
initial glottal - null	11	10
initial consonant - null	14	15
final consonant - null	9	11
initial cluster - singleton	12	14
final cluster - singleton	12	11
/r/ - /l/	9	10
/r/ - /w/	8	8

Note: "row" versus "woe" tests the /r/ vs. /w/ contrast. There are eight incidences of this contrast being measured in both tests.

Intrajudge Reliability

Intrajudge reliability for the naive and informted listeners was 92.2% and 93% respectively. As suggested by Kearns and Simmons (1988), the formula used to calculate overall reliability (R) was:

$$R = \frac{\text{Total Number of Agreements}}{\text{Total Number of Agreements + Disagreements}} \times 100$$

where agreements are between scoring of each initial and re-presented stimulus word. These comparable averages indicate that the task is straightforward enough that these naive listeners could be counted upon to be nearly as consistent as the informed listeners in this study with these speakers.

Interjudge Reliability

Interjudge percentage-agreement levels were 91.98% for naive listeners and 94.74% for informed listeners. Overall, the average reliability levels of 94.86 percent are acceptable. However, if the words which were scored 100% correctly are removed, leaving only those which had mistaken responses by at least one judge, the reliability levels drop to 71.47% for naive listeners and 74.15% for informed listeners. Kearns and Simmons (1988) made a similar observation on a perceptual rating test of ataxic dysarthric patients, wherein their "occurrence [of scored aberrances] reliability" levels decreased their overall reliability ratings by ten to thirty percent. They concluded that "additional experience with this approach to reliability assessment...is needed before we can firmly establish acceptable agreement levels for clinical research in speech-language pathology...The occurrence reliability data revealed potentially serious problems relating to agreement levels for deviant characteristics that were not revealed by the overall reliability analysis." This may also be the case in the present study. If the test attempts to gauge intelligibility for this population, but is overloaded with easily discriminated words, then overall reliability levels will not be a valid measure of the test's reliability.

Summary of Intelligibility Data

The overall intelligibility changes of the patients can be found in Table 6 (following page). As can be seen, the intelligibility test implemented in this study indicated measurable gains for three of the five patients.

In deciding which words merited analysis, a criterion was used of at least three incorrect responses in either the informed or naive listener groups for a given word to be judged as significantly unintelligible pre-treatment, with a reduction to one or zero incorrect responses post-treatment for the same group to indicate improvement. Summary data for these specific words only are shown in Table 7 (following page).

Table 6.
Intelligibility Scores for Parkinson's Disease Patients Pre- and Post-Treatment in Experiment 2.

Patient #	Overall Intelligibility		Percent Change
	Pre	Post	
1	94.0%	95.7%	1.7%
2	94.0%	95.2%	1.3%
3	93.6%	98.7%	5.4%
4	87.1%	85.6%	-1.7%
5	96.1%	89.9%	-6.5%

The categories in which perceived errors improved pre- to post-treatment are listed by frequency in Table 8. As can be seen, some of the contrasts expected from the preliminary study occurred relatively frequently, such as the glottal versus null and the affricate versus fricative groups. It should be noted that these were not the only confusions registered by the listeners; they are only those which attained a significant frequency of occurrence in the pre-treatment conditions and significant improvement in the post-treatment productions (as defined above). Thus, the data presented in Experiment 2 should not be interpreted as providing an all-encompassing profile of speech errors made by Parkinson's patients.

Table 7.
Summary Data for Words with Post-Treatment Intelligibility Improvement in Three Patients with Parkinson's Disease in Experiment 2.

Patient #	Stimulus Word		Duration	Ave. dB (50 cm)	Incorrect Choice(s)	Type of Error
1	cheer	pre	533.4	82.8	sheer	Fricative - Affricate
		post	646.8	92.1		
	harm	pre	484.2	80.3	arm	Initial Glottal - Null
		post	778.0	89.5		
	pad	pre	563.4	84.0	bad	Initial Voicing
		post	619.6	90.7		
2	at	pre	531.2	69.4	hat	Glottal-Null
		post	474.0	68.3		
	cake	pre	527.0	69.4	ache	Initial Consonant - Null
		post	454.4	69.5	cape	Consonant Place
	him	pre	471.2	70.2	hem	High - Low Vowel
		post	431.9	72.3		
	seed	pre	625.0	74.6	see	Final Consonant - Null
		post	587.8	72.0	seeds	Final Consonant - Singleton
3	at	pre	607.8	66.9	add	Final Voicing
		post	665.5	77.5	hat	Initial Glottal - Null
	chop	pre	447.4	69.7	shop	Fricative - Affricate
		post	775.5	79.5		
	feed	pre	502.0	71.6	fee	Final Consonant - Null
		post	765.4	73.5		
	him	pre	522.8	73.0	hem	High - Low Vowel
		post	797.2	76.4	ham	
	leak	pre	548.2	70.4	reek	/l/ - /r/
		post	997.0	77.0		
	side	pre	782.3	65.6	sigh	Final Consonant - Null
		post	997.6	79.4		
	sigh	pre	732.1	69.6	thigh	Other Fricative
		post	920.8	82.7		

Table 8.
Perceptual Error Frequency for all Parkinson's Disease Subjects in Experiment 2.

Category	Errors
Glottal - Null	13
Final Consonant - Null	12
Fricative - Affricate	9
High - Low Vowel	7
Initial Voicing	6
/l/ versus /r/	3
Other Fricative	3
Initial Consonant - Null	2
Consonant Place	1
Final Cluster - Singleton	1

Remembering that the focus of the Lee Silverman Voice Treatment was laryngeal, rather than articulatory, it would be expected that improvements in intelligibility would be related to phonemes involving laryngeal and possibly timing concerns. This is confirmed with the types of errors that most frequently improved, especially in the glottal versus null, fricative versus affricate, and initial voicing categories. However, the treatment concept of increased vocal effort improving general speech production is also supported here with the unexpected rise in intelligibility for vowels and final consonant versus null categories. These improvements imply that not only did the larynx achieve more appropriate closure post-treatment (useful for fricatives, affricates, and stops), but also that the increased overall effort was associated with the shaping of the vocal tract toward more target-appropriate movements as well.[1] Additionally, the improvement in the final-consonant versus null category supports the idea that enhancing the vocal source lends greater energy overall for the duration of the word, so that it can be completed intelligibly. Although treatment efficacy was not the prime motivation for this experiment, these data provide support for the treatment principles.

Confidence Levels

Confidence level data are presented in Table 9. The confidence scale ranged from one to five, with five being the most confident. Naive listeners were less confident in their choices than the informed listeners, as evidenced by the naive overall average of 4.4 versus the informed overall average of 4.9. It should be noted that the higher confidence ratings given by the speech pathologists left less room for variation; instead of using the whole scale, they tended to confine their choices to a level of either four or five. By effectively compacting the scale, this tendency reduced the usefulness of the measure.

Table 9.
Average Confidence Levels for Naive and Informed Listeners in Experiment 2 (Scale of 1-5).

Patient #	Informed Average		Naive Average		Avg. Pre Score	Avg. Post Score
	Pre	Post	Pre	Post		
1	4.94	4.94	4.75	4.80	4.84	4.87
2	4.82	4.85	4.35	4.46	4.58	4.65
3	4.84	4.93	4.32	4.82	4.58	4.87
4	4.90	4.89	4.56	4.40	4.73	4.65
5	4.74	4.71	3.87	4.11	4.31	4.41

When this study was conceptualized, it was anticipated that single word intelligibility changes would be subtle pre- to post-treatment. As a backup measure of improvement, confidence levels were included in an attempt to document subtle changes in intelligibility. An increase in listener confidence levels could indicate nuances of improvement. This idea has been applied in psychological studies related to memory and knowledge testing. Two of our patients did not attain measurable improvement scores. Patient #4 demonstrated deteriorating and fluctuating speech along with dementia and attention difficulties during therapy and experimental measurement tasks. This was reflected in the fact that his average intensity was lower post-treatment than on the initial measures. It would be expected that the confidence levels would show a corresponding decrease. They did in fact decrease, but not by a significant amount. It should be noted that this patient's pre-treatment sample was the first heard by the listeners, and his post-treatment group was the third that was presented. Therefore, it is unlikely that presentation order or familiarity with the test played a role in the confidence intervals failing to more closely mirror the pre-post treatment decline.

The results for the other patient who did not significantly improve, Patient #5, may demonstrate a flaw in the test. Unlike Patient #4 above, Patient #5's presentation was given in reverse order. That is, the listeners heard his post-treatment test second and his pre-treatment test seventh of those presented. Although this patient's intensity levels increased significantly, improvement pre-to post-treatment was not documented with this procedure. His confidence levels were also higher for the pre-treatment condition group, which importantly was presented nearer the end of the test. Since these results were not expected relative to the increased intensity post-treatment, it was considered important to explore this discrepancy.

Considerations Regarding the Intelligibility Test

In view of the outcome of Patient #5's pre- and post-treatment results, an analysis of variance (two factor with replication) was carried out to determine whether any measurable gain in scores was made as the test progressed. That is, if the later scores were more likely to be higher because of some consideration related to the test rather than due to the stimulus words themselves. As was suspected, order was a significant factor ($p<.001$) in this analysis. Given the fact that some factor (or factors) led to better scores as the test progressed, the sources of this problem were explored.

One reason why the scores would be better on one part of the test versus another is that the average intensity might be greater. If this proved to be the case, then the words would be "easier" to hear, so scores might be higher. However, measurements logged during this test indicated a mean difference of only one decibel between the first half of the samples and the second half of the samples. This one decibel difference is not likely to account for the significantly better scores found in the second half.

Another consideration relates to the format of the test itself. In the preliminary study, the corpus of test words was composed only of minimally-paired contrasts. Both words in every pair were included as stimulus words somewhere on the test, so it was impossible to predict which word would be chosen purely by reading the choice list, even for a listener who had been given the samples many times before. In the current test, given the lengthy duration of the listening task, and the similarity of the words, it was rationalized that the lack of a paired-word corpus would not affect the validity of the test. That may not have proved true.

In the original article by Kent et al. (1989), it is stated that "in practice, dummy words are inserted into the word list read by the patient to reduce listener's expectations based on word familiarity." Interviews with the listeners in this experiment, however, lead to the conclusion that this action would not furnish the control necessary to alleviate the problem. In both the informed and naive groups, the examiner noted that despite randomization, by the fifth or sixth patient presentation, listeners had their highlighters poised as soon as they saw a word they recognized from a previous presentation. Nine out of ten listeners spontaneously commented on this predictability when asked for their input on the test validity. Since the goal of this test is reliable repeated clinical usage, this predictability represents an important concern.

The findings of this experiment must be considered in relation to this observation. If, for example, a listener had made an erroneous judgment early in the task, he might again select the same word on a subsequent trial based solely on his expectation. But more in keeping with the significantly better performances later in the task, he or she may have unconsciously used the preceding information to hone his or her perception of the test words through the trials. This would tend to distort the experimentally significant findings by artificially lowering the post-treat-

ment scores, if they were presented earlier in the test. Similarly, the pre-treatment scores could be artificially raised if they were presented late in the task. Either way, the pre- to post treatment discrepancies would be masked, thus lowering the validity of the test. In practice, also, the intelligibility gains registered on the test did not match the predictions of the professionals involved in conducting the study. The magnitude of gains measured on this intelligibility test were less significant than would have been expected from the clinical interaction with the patients.

General Discussion

Despite the methodological concerns of the intelligibility test, this study makes several important contributions. Analysis of the patients' pre- and post-treatment productions show measurable differences which provide clues to improvements in intelligibility. Categories which show the most improvement include the glottal versus null, final consonant versus null, fricative versus affricate, high versus low vowels, initial voicing, /l/ versus /r/ distinction, and other fricative contrasts.

Hypothesizing the origin of these improvements would suggest the changes brought due to the therapy. For instance, Benjamin (1992) showed that the Lee Silverman Voice Treatment increased laryngeal valving capabilities with similarly affected Parkinson's disease patients. It would be logical to assume that the patients in the current study also made laryngeal gains. Spectrographic analysis (Maclay, 1992) suggested increased laryngeal abilities to improve the production of phonemes that depend upon adequate closures to build up needed air pressure. Increasing inspiratory and expiratory volume also may have helped in this regard. If the patient has adequate respiratory control, he can maintain the necessary aerodynamic requirements throughout the duration of the word. These factors help to explain the improvement in the final consonant versus null category, as well as the fricative and glottal - related classifications.

The hypothesis of the therapy is ambitious: that enhancing phonatory effort would generalize to enhanced supralaryngeal and articulatory function. The changes that were seen in the vowel category and the /l/ versus /r/ category support that claim. Without any attention to vocal tract shaping by the clinicians, production altered to increase the intelligibility of these phonemes. One aspect of production that did change measurably was word durations. It is a reasonable conjecture that these increased durations gave the patients additional time needed to position the articulators toward appropriate targets.

Regarding the fundamental issues of the construction of the intelligibility test, one option seems reasonable. Returning to the original test design presentation of a paired-word test with stimulus words and foils would alleviate the predictability concern. This solution does increase the length of the test in that every stimulus word must have a corresponding minimal contrast counterpart.

It would seem, then, that a choice must be made between an inherently longer test, a test with fewer error potential categories, or fewer opportunities for error in a given category. Moreover, it is recommended that the alternative chosen would depend upon the goal of the administrator. That is, a full intelligibility battery for all types of dysarthria may include all the words in a longer version. If a more quickly administered test were desired, categories with low probability of occurrence for the particular patient's impairment could be eliminated. Finally, a spot-diagnostic may include only a few words for each error category. As a general test protocol for Parkinson's disease patients, the most efficient choice would probably be the second option. Eliminating the least likely error categories to decrease the length of the test would provide the most benefit with the least risk of missing important information. To achieve this result, it would be important to be certain of the patient's motor speech disorder. An intelligibility test geared for hypokinetic dysarthria may not be appropriate for a different condition.

This study has attempted to elucidate some of the concerns surrounding the improvement in and measurement of intelligibility at the single-word level in patients with Parkinson's disease. The study has shown that in Parkinson's disease, intelligibility can be improved with treatment directed toward increased phonatory effort. Further research should focus on advancing the intelligibility test design to counteract the concerns outlined herein, as well as further quantifying the reasons for the improved performances of the patients post-therapy.

Footnote 1

An in-depth treatment of acoustic and perceptual correlates to the intelligibility gains can be found in Maclay (1992). She found that post-treatment improvement variables leading to increased intelligibility included raised intensity levels, longer phoneme durations, more accurate voicing, and more nearly normal F2 production. Spectrographic analysis showed that the vowels (and the /l/ in the "leak-reek" pair) reflected disordered F2 production pre-treatment, and showed measurable improvement toward expected frequency values post-treatment.

Acknowledgements

The authors gratefully acknowledge the contribution of William Winholtz. This research was supported in part by NIH grant No. R01 DC01150 and P60 DC-00976.

Appendix
Test Words and Foils for Experiment 2

	Stimulus Word	Foil 1	Foil 2	Foil 3
1	air	are	hair	fair
2	at	add	hat	fat
3	ate	aid	hate	fate
4	bad	bed	pad	bat
5	beat	boot	bit	meat
6	bill	dill	gill	mill
7	blend	end	lend	bend
8	blow	bloat	beau	low
9	bunch	punch	munch	bun
10	cake	take	ache	cape
11	cash	gash	catch	cat
12	chair	share	tear	air
13	cheer	jeer	sheer	tier
14	chip	gyp	ship	tip
15	chop	chap	shop	top
16	dock	mock	knock	docks
17	dug	tug	duck	bug
18	ease	is	cheese	peas
19	fat	feet	hat	at
20	feed	feet	seed	fee
21	feet	fat	fit	sheet
22	fill	full	hill	pill
23	fork	cork	four	forks
24	geese	goose	gas	guess
25	had	hid	pad	add
26	hail	sail	tail	ail
27	hall	ball	tall	all
28	hand	sand	panned	and
29	harm	farm	charm	arm
30	hat	fat	pat	at
31	heat	hate	feet	eat
32	him	hum	hem	ham
33	hold	fold	cold	old
34	knew	knee	know	gnaw
35	knot	nut	nod	dot
36	leak	Luke	lick	league
37	leak	lee	leaks	reek
38	lip	leap	lit	rip
39	mat	mats	bat	meat
40	much	much	muck	mutt
41	nice	knife	night	dice
42	pad	bad	pat	had
43	pat	pot	pit	pad
44	pit	pet	bit	spit
45	rake	ray	rakes	lake
46	read	rid	lead	weed
47	reap	rip	leap	weep
48	rise	eyes	lies	wise
49	rock	rocks	lock	wok
50	row	owe	low	woe
51	see	she	he	seed
52	seed	feed	see	seeds
53	sell	shell	fell	tell
54	sew	show	foe	toe
55	sheet	shoot	seat	eat
56	ship	sheep	sip	chip
57	shoot	suit	sheet	shot
58	side	sight	sign	sigh
59	sigh	shy	thigh	side
60	sin	shin	tin	in
61	sink	pink	ink	sing
62	sip	zip	ship	slip
63	slip	sleep	sip	lip
64	spit	it	pit	sit
65	steak	snake	sake	take
66	sticks	six	ticks	stick
67	tile	dial	pile	mile
68	wax	wack	lax	racks
69	witch	wish	wit	rich
70	write	ride	light	white

References

ANSEL, B. M., and KENT, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research*, **35**, 296-308.

BENJAMIN, P. A. (1992). *The effect of intensive speech therapy on vocal intensity in patients with Parkinson's disease* (Unpublished master's thesis. University of Colorado, Boulder, Colorado).

BEUKELMAN, D.R., and YORKSTON, K.M. (1980). The influence of passage familiarity on intelligibility estimates of dysarthric speakers. *Journal of Communication Disorders*, **13**, 33-41.

CANTER, G. (1965). Speech characteristics of patients with Parkinson's Disease: III. Articulation, diadochokinetic, and overall speech adequacy. *Journal of Speech and Hearing Disorders*, **30**, 211-234.

COSTERMANS, J., LORIES, G., and ANSAY, C. (1992). Confidence level and feeling of knowing in answering: the weight of inferential processes. *Journal of Experimental Psychology*, **18**, 142.

CRITCHLEY, E.M.R. (1981). Speech disorders of Parkinsonism: A review. *Journal of Neurological & Neurosurgical Psychiatry*, **44**, 751-758.

CULLINAN, W.L., BROWN, C.S., and BLALOCK, P.D. (1986). Ratings of the intelligibility estimates of dysarthric speakers. *Journal of Communication Disorders*, **19**, 185-195.

ELLIS, L. W., and FUCCI, D. J. (1991). Magnitude-estimation scaling of speech intelligibility: Effects of listeners' experience and semantic-syntactic context. *Perceptual and Motor Skills*, **73**, 295-305.

ENGLAND, A. C. and SCHWAB, R. S. (1959). The management of Parkinson's disease. *AMA Archives of Internal Medicine*, **104**, 439-468.

FORREST, K., WEISMER, G., and TURNER, G. S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. *Journal of the Acoustical Society of America*, **85**, 2608-2622.

GUO, C., and WINHOLTZ, W. (1993). Custom Software, The Denver Center For The Performing Arts; The Wilbur James Gould Voice Research Center.

HALLET, M. and KHOSBIN, S. (1980). A psychological mechanism of bradykinesia. *Brain*, **103**, 301-314.

HANSEN, D. G., GARRATT, B. R., and WARD, P. H. (1984). Cinegraphic observations of laryngeal function in Parkinson's disease. *Laryngoscope*, **94**, 348-353.

KEARNS, K.P., and SIMMONS, N. N. (1988). Interobserver reliability and perceptual ratings: More than meets the ear. *Journal of Speech and Hearing Research*, **31**, 131-136.

KENT, R. D., WEISMER G., KENT J. F., and ROSENBEK, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, **54**, 482-489.

KENT, R.D., KENT J.F., WEISMER, G., SUFFIT, R.L., ROSENBEK, J.C., MARTIN, R.E., and BROOKS B.R. (1990). Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Disorders*, **55**, 721-728.

LUDLOW, C. L., and BASSICH, C. J. (1983). The results of acoustic and perceptual assessment of two types of dysarthria. In W. R. Berry (Ed), *Clinical Dysarthria* (San Diego: College-Hill Press).

LUDLOW, C. L., and BASSICH, C. J. (1984). Relationships between perceptual ratings and acoustic measures of hypokinetic speech. In M. R. McNeil, J. C. Rosenbek, and A. Aronson (Eds), *The Dysarthrias: Physiology, Acoustic, Perception, Management* (San Diego: College-Hill Press).

MACLAY, S. L. (1992). Speech intelligibility gains in Parkinson's disease patients post voice treatment: Perceptual and acoustic correlates. (Unpublished master's thesis, University of Colorado, Boulder, Colorado).

MARKIDES, A. (1983). *The Speech of Hearing-Impaired Children* (Oxford: Manchester University Press).

McGARR, N.S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research*, **26**, 451-458.

McDOWELL, F. H., LEE, J. E. and SWEET, R. D. (1986). Extrapyramidal disease. In A. B. Baker and R. J. Joynt (Eds), *Clinical Neurology* (Philadelphia: Harper and Row).

MENCKE, E.O., OCHSNER, G.J., and TESTUT, E.W. (1983). Listener judges and the speech intelligibility of deaf children. *Journal of Communication Disorders*, **16**, 175-180.

METZ, D. E., SCHIAVETTI, N., and SITLER, R. W. (1980). Toward an objective description of the dependent and independent variables associated with intelligibility assessments of hearing-impaired adults. In J. D. Subteiny (Ed), *Speech Assessments and Speech Improvement for the Hearing Impaired* (Washington, DC: Alexander Graham Bell Association for the Deaf).

MONSEN, R. B. (1978). Toward measuring how well hearing-impaired children speak. *Journal of Speech and Hearing Research*, **21**, 197-219.

PETERSON H. A., and MARQUARDT, T. P. (1981). *Appraisal and Diagnosis of Speech and Language Disorders* (Englewood Cliffs, NJ: Prentice-Hall).

RAMIG, L. O. (in press). Speech therapy for patients with Parkinson's disease. In W. Keller and G. Paulson (Eds), *Therapy of Parkinson's Disease* (New York: Marcel Dekker).

RAMIG, L. O. (1992). The role of phonation in speech intelligibility: A review and preliminary data from patients with Parkinson's disease. In R. Kent (Ed), *Intelligibility in Speech Disorders: Theory, Measurement and Management* (Amsterdam: John Benjamin).

RAMIG, L. O., HORII, Y. and BONITATI, C. M. (1991). The efficacy of voice therapy for patients with Parkinson's disease. *NCVS Status and Progress Report*, June, p. 61-86.

RAMIG, L.O., MEAD, C., SCHERER, R., LARSON, K., and KOHLER, D. (1988). *Voice therapy for Parkinson's disease*. (Paper presented at the Clinical Dysarthria Conference, San Diego, California).

SAMAR, V. J. and METZ, D. E., (1988). Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. *Journal of Speech and Hearing Research*, **31**, 307-316.

SHEARD, C., ADAMS, R., D., and DAVIS, P. J. (1991). Reliability and agreement of ratings of ataxic dysarthric speech samples with varying intelligibility. *Journal of Speech and Hearing Research*, **34**, 285-293.

WEISMER, G., KENT, R. D., HODGE, M., and MARTIN, R. (1988). The acoustic signature for intelligibility test words. *Journal of the Acoustical Society of America*, **84**, 1281-1289.

WEISMER, G., MARTIN, R., KENT R. D., and KENT, J. F. (1992). Formant trajectory characteristics of males with amyotrophic lateral sclerosis. *Journal of the Acoustical Society of America*, **91**, 1085-1098.

YORKSTON, K., and BEUKELMAN, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders*, **11**, 499-512.

YORKSTON, K., and BEUKELMAN, D. R. (1980). *Intelligibility as an overall measure of dysarthric speech* (Paper presented at the annual convention of the American Speech-Language-Hearing Association, Detroit, Michigan).

YORKSTON, K., BEUKELMAN, D. R., MINIFIE, R. D., and SAPIR, S. (1984). Assessment of stress patterning in dysarthric speakers. In M. R. McNeil, J. C. Rosenbek, and A. Aronson (Eds), *The Dysarthrias: Physiology, Acoustic, Perception, Management* (San Diego: College-Hill Press).

ZIEGLER, W., HOOLE, P., HARTMANN, E., and VON CRAMON, D. (1988). Accelerated speech in dysarthria after acquired brain injury: Acoustic correlates. *British Journal of Disorders of Communication*, **23**, 215-228.

NCVS Status and Progress Report - 6
May, 1994, 127-133

Intensive Voice Treatment in Parkinson's Disease: Laryngostroboscopic Findings

Marshall E. Smith, M.D.
Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts
and
Department of Otolaryngology/Head and Neck Surgery, University of Colorado Health Sciences Center
Lorraine Olson Ramig, Ph.D.
Department of Communication Disorders and Speech Science, University of Colorado-Boulder
and
Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts
Christopher Dromey, M.A., CCC-SLP
Department of Communication Disorders and Speech Science, University of Colorado-Boulder
and
Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts
Kathe S. Perez, M.A., CCC-SLP
Department of Communication Disorders and Speech Science, University of Colorado-Boulder
and
Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts
Ráz Samandari, M.D.
Department of Otolaryngology/Head and Neck Surgery, University of Colorado Health Sciences Center

Abstract

As part of ongoing research to investigate and document the efficacy of intensive voice therapy to improve functional communication in patients with idiopathic Parkinson's disease, forty-four patients were enrolled in a controlled, randomized, prospective study. Pre-to-post-treatment comparisons are presented here on twenty-two of those patients with usable laryngeal imaging data. Of the twenty-two patients, thirteen patients received intensive therapy aimed at increasing vocal and respiratory effort (VR), while nine received intensive therapy aimed at increasing respiratory effort (R) only. All patients had a pre-treatment evaluation which included two (but sometimes only one) voice recordings and an otolaryngologic examination with laryngostroboscopy. At the completion of four weeks of therapy (sixteen sessions) two voice recordings were made and laryngostroboscopy was again performed. The pre- and post-therapy videolaryngostroboscopy tapes were then randomized and reviewed by four expert judges. Raters' findings were then compared with vocal intensity measured before and after therapy. The VR therapy group showed pre-post improvements on laryngostroboscopic variables: less glottal incompetence and a decrease in supraglottal hyperfunction following therapy. No differences were observed in the R only group. The mean intensity increase in the VR therapy group was 12.5 dB, compared with a decrease of 1.9 dB in the R only group. These findings suggest that in patients with Parkinson's disease, intensive therapy focusing on phonatory effort improves adduction of the vocal folds as assessed by laryngostroboscopy. Differences in laryngeal function in these patients observed with fiberoptic laryngoscopy and rigid telescopic laryngoscopy are discussed.

Introduction

Parkinson's disease afflicts 1.5 million Americans. It is estimated that 75% of these patients have problems with speech and communication with 89% having voice disorders (1) . Four percent of these patients report speech problems as an initial symptom of the disease (2). Darley, Aronson, and Brown (3) stated that the laryngoscopic examination of the Parkinson's disease patient does not reveal any characteristic abnormalities. However, several findings are commonly seen. Hanson, Gerratt, and Ward (4), studied 32 patients with Parkinson's disease by laryngoscopy. They identified bowing of the vocal folds in 30 patients. Tremor of the supraglottic structures was seen in 10 patients, and during relaxed breathing 4 patients showed tremulous movement of the arytenoids. Vocal tremor was perceived only in patients who had associated head tremor during phonation. They also noticed an extreme degree of supraglottic constriction of the larynx in 5 patients.

Despite the disordered speech and voice in Parkinson's disease, efforts at speech therapy have generally proven ineffective (5,6). Recently, Ramig and associates (7,8,9) developed a treatment program for Parkinson's disease focusing on phonation. This program was based on the findings of Scott & Caird (10) and Robertson & Thompson (11), and was designed to increase phonatory effort, reduce bowing and enhance vocal loudness.

In order to study the interactive roles of respiratory and laryngeal systems in treatment related changes two forms of therapy were developed: respiratory (R) and voice and respiratory (VR). A focus of each treatment was increased effort; details of treatment procedures are summarized elsewhere (12). As part of this study to investigate the efficacy of intensive voice therapy to improve speech communication in Parkinson's disease patients, laryngostroboscopy was incorporated into pre- and post-therapy assessment. This examination allows the documentation of laryngeal abnormalities that may accompany impaired voice, and changes in laryngeal function that occur with therapy. Laryngostroboscopic observations may contribute to our understanding of the effects of behavioral voice therapy on underlying phonatory mechanisms. It was hypothesized that increases in laryngeal adduction and sound pressure level would be measured in patients in the VR treatment group.

Methods

Forty-four patients with idiopathic Parkinson's disease were the subjects in this study. Twenty-two of these patients had usable laryngeal imaging data and will be the focus of this report. There were 17 males and 5 females, ages ranged from 49 to 76 years. Symptoms of PD had been present for an average of 9 years in these patients. They had all been examined by a neurologist and given a rating on the Hoehn and Yahr scale (13), with average rating as Stage III. Age, sex, and stage of the disease were comparable across treatment groups. Patients did not change anti-Parkinson medications during the course of the study. They were randomly assigned to one of two treatment groups: respiratory (R) or voice and respiratory (VR).

Before beginning voice therapy, patients were examined with videolaryngostroboscopy. This examination usually took place in the month before treatment, but not on the same date as a voice recording session. An otolaryngologic history and an examination were obtained on all patients prior to beginning therapy. The laryngeal imaging examination was then repeated within 1-2 weeks after finishing intensive voice therapy.

Laryngeal imaging and videolaryngostroboscopic examination were conducted utilizing well-described techniques (14). The nasal passage was topically anesthetized with 4% lidocaine spray. Endoscopic examination was conducted with both an Olympus ENF-P3 fiberscope and Nagashima SFT-1 70° rigid telescope. Light sources included both an Olympus halogen constant light source and a Nagashima LS-3A laryngostrobe. Images were recorded with a CCD camera, using a 35 mm lens for the fiberscope and 60 mm lens for the rigid telescope, and a VHS or SVHS tape recorder. A separate video tape was used for each patient recording.

The examination protocol is briefly summarized as follows: with the fiberscope the larynx was visualized under constant light during quiet respiration, sustained phonation of the vowel /i/, and repetition of short phrases. The strobe light source was then used to visualize the larynx and vocal folds during phonation of sustained vowel /i/ under several conditions, including 1) normal pitch, normal loudness, 2) high and low pitch, 3) soft and loud phonation, and 4) during several vocal fold adduction "pushing" exercises. Trials were repeated until adequate samples were obtained, as judged by observation on the video monitor. The fiberscope was withdrawn and the stroboscopic examination of the larynx was repeated with the rigid telescope for the sustained vowel tasks.

The data for this study were prepared by creating a "master" study videotape that included segments from the videotapes of each patient, pre- and post-treatment. The order of the patient videotapes was determined by a computer random number generator. Six patients' recordings were repeated to assess intra-rater reliability. The audio signal from each patient tape was removed to eliminate auditory perceptual cues. Instead, a "prompter" overdubbed onto the audio track information that identified the tasks performed during examination of each patient to allow raters to identify them.

A study rating form was designed to record the specific variables of interest in assessing treatment efficacy (figure 1). The variables assessed on the rated videotapes included glottal configuration, degree of glottal incompetence, and laryngeal hyperfunction. Glottal configuration was observed as complete, bowing, hourglass, anterior chink, posterior chink, or incomplete. Degree of glottal incompetence was rated during the most closed portion of the glottal cycle on stroboscopy. A five point scale was used: 1 (slight) = folds just not touching, 3 (moderate) = approximately 50% of the length of the folds not touching with 1-2 mm gap, 5 (severe) = no vocal fold contact throughout the length of the folds and a large (3-4) mm glottal gap. Laryngeal hyperfunction was rated for both false fold compression (FF) and anterior-posterior (AP) compression on a five point scale. The raters judgments were standardized for this rating system by prior consensus on interpretation of the scale. For FF hyperfunction 1 = no false fold overclosure, laryngeal ventricles easily seen, 2 = one or both false folds obscure laryngeal ventricles, 3 = one or both false folds obscure ventricles and a portion of the true vocal folds, 4 = one or both false folds obscure the entire true vocal folds but are not touching, 5 = false folds touch and cover the entire glottis. For AP hyperfunction; 1 = no shortening of glottal length compared with resting, non-phonating state, 2 = glottal length shortened by 25%, 3 = glottal length shortened by 50%, 4 = glottal length shortened by 75%, 5 = arytenoids touch laryngeal surface of epiglottis, obscuring glottal view.

The twelve variables assessed for the study from the laryngostroboscopic examinations included each of the three parameters above (glottal phonatory configuration, glottal incompetence, supraglottal function) seen under four observation conditions. These were 1) flexible telescope view of vowel /i/ produced at the patient's self-judged normal pitch normal loudness (NPNL), 2) flexible telescope view of vowel /i/ produced at the patient's loud phonation effort, 3) rigid telescope view of vowel /i/ produced at the patient's self-judged NPNL, and 4) rigid telescope view of vowel /i/ produced at the patient's loud phonation effort.

Four raters with experience in voice research assessed the recordings. Three were speech pathologists and one was an otolaryngologist. They independently and as a group viewed a practice tape, and met several times together to develop standard rating criteria. 95% inter-rater agreement was achieved on the practice tape. Following the practice procedures, they independently viewed and rated the "master" tape.

Sound pressure levels were separately recorded as part of a broader protocol, summarized by Ramig, et al (12). Sound pressure level data were obtained from a B&K sound level meter, positioned 50cm from the lips during six maximum duration sustained "ah" phonations for each subject. These recordings were made at a different time than the laryngoscopy examinations, usually within one week and the findings will be summarized in detail elsewhere (15).

Results

Intra- and inter-rater reliability measures were obtained through a comparison of ratings from duplicated patient tape recordings on the master tape. 25% of the patient recordings were repeated randomly on the master tape. A correlation was performed for each rater's original and his or her rechecked data. Pearson correlation coefficients obtained ranged from .78 to .94, indicating good intra-rater reliability. For the group inter-rater assessment, the Pearson r coefficient for each of the 12 measurement variables ranged from .87 to 1.0.

The observations of glottal configuration prior to treatment for both groups (VR = voice and respiration therapy group, R = respiration therapy only group) are given in Table 1. These are for phonation of /i/ during normal pitch, normal loudness (NPNL), and loud phonation, using flexible, then rigid telescope. A graph was made of normal vs. any abnormal glottal configuration seen on flexible

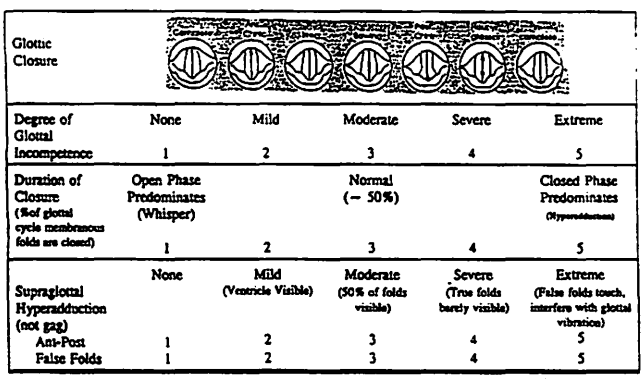


Figure 1. Laryngostroboscopy rating form used in this study.

Table 1.
Glottal closure patterns from laryngostroboscopy ratings of all patients prior to therapy, using both flexible and rigid telescopes, at both normal pitch/normal loudness (NPNL) and loud phonation of vowel /i/.

Glottal Closure Pattern	Flexible		Rigid	
	NPNL (n=21)	Loud (n=19)	NPNL (n=15)	Loud (n=9)
Normal Closure	9	15	4	6
Bowing	8	0	8	3
Anterior Commissure Chink	2	1	0	0
Posterior Commissure Chink	2	3	2	0
Incomplete Closure	0	0	1	0

endoscopy and the two groups were separated by treatment type (Figure 2). Pre-to post-treatment comparisons show a trend for improvement to normal glottal configuration post-therapy in the VR group. Analysis of variance showed a significant effect for treatment type in both flexible (F=8.45; df=1,19; p<.01) and rigid (F=4.35; df=1,12; p<.05) views. The observations taken at patient's loud phonation effort did not show a significant difference in pre-post treatment effect for either the flexible or rigid telescope (p=.630; p=.208).

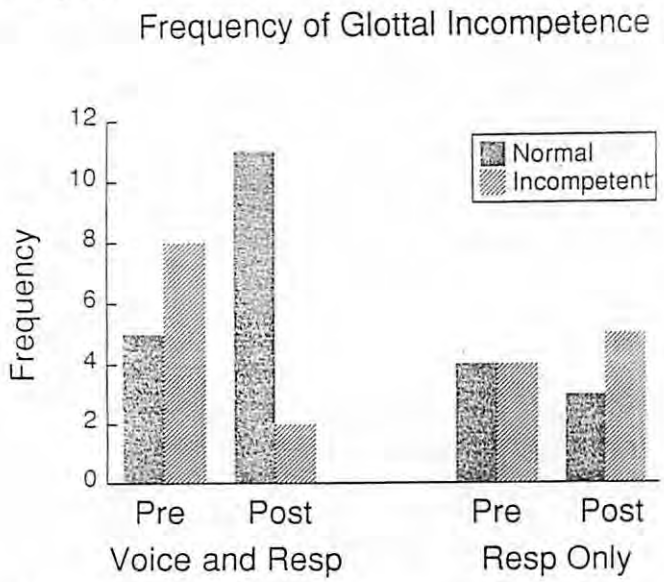


Figure 2. Comparison of frequency of glottal incompetence before and after two types of voice therapy for Parkinson's disease. Views rated from laryngostroboscopy examinations done with flexible telescope, vowel /i/, NPNL.

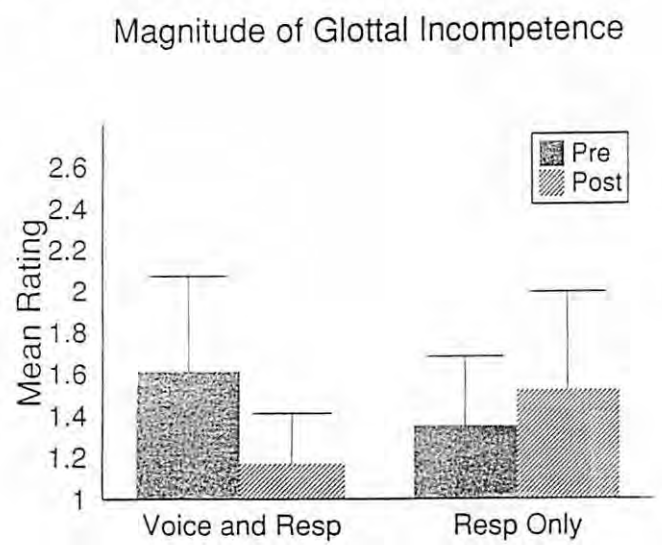
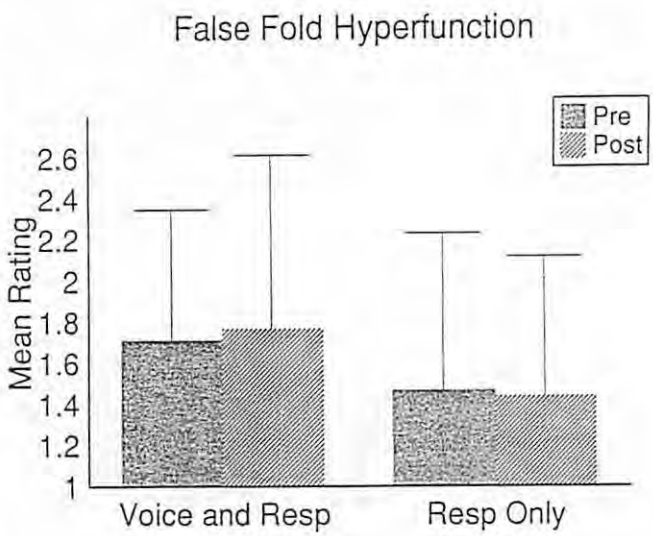


Figure 3. Group mean ratings of magnitude of glottal incompetence before and after two types of voice therapy for Parkinson's disease. Views rated from laryngostroboscopy examinations done with flexible telescope, vowel /i/, NPNL.

The magnitude of glottal incompetence for the non-normal glottal closure patterns was rated on a five point scale. There was a significant difference pre to post treatment between the two therapy types for both the flexible (F=8.25; df=1,19,p<.010) and rigid (F=8.67; df=1,11; p<.010) endoscopy views at NPNL. The group data are presented in figure 3. The observations taken at patients' loud phonation effort did not show a significant difference in pre-post-treatment effect for either the flexible or rigid telescope (p=,496; p=.233).



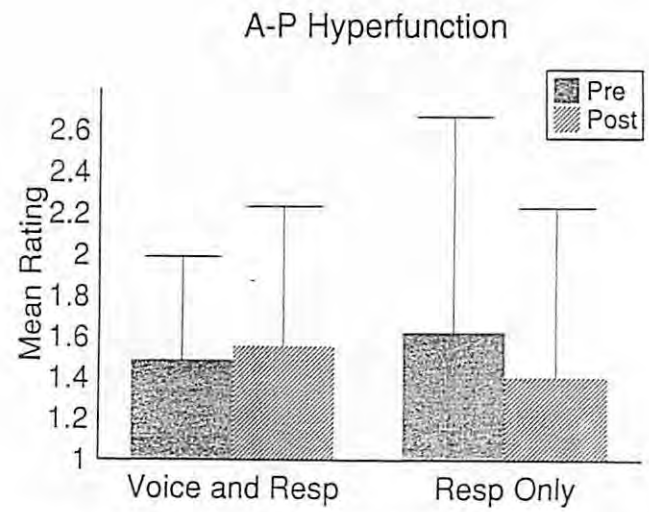


Figure 4a (top). Group mean ratings of false fold hyperfunction before and after two types of voice therapy for Parkinson's disease. Views rated from laryngostroboscopy examinations done with flexible telescope, vowel /i/, NPNL. Figure 4b (bottom). Group mean ratings of anterior-posterior supraglottal hyperfunction before and after two types of voice therapy for Parkinson's disease. Views rated from laryngostroboscopy examinations done with flexible telescope, vowel /i/, NPNL.

Supraglottal hyperfunction scores before and after therapy were separated by treatment type. Ratings were done for both false fold hyperfunction and anterior-posterior hyperfunction. The group mean results are displayed in figure 4a,b. Results indicated no statistically significant change in the supraglottal hyperfunction after therapy in either treatment group.

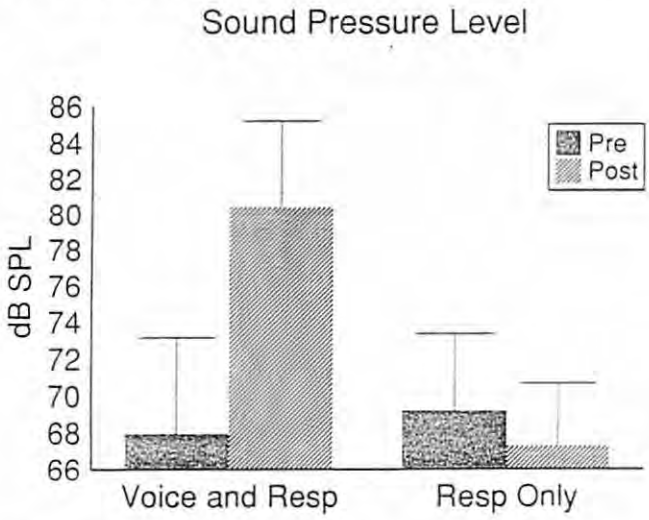


Figure 5. Sound pressure level group mean data for vowel /a/ pre-post treatment measures for two treatment types. Mouth to sound level meter distance: 50 cm.

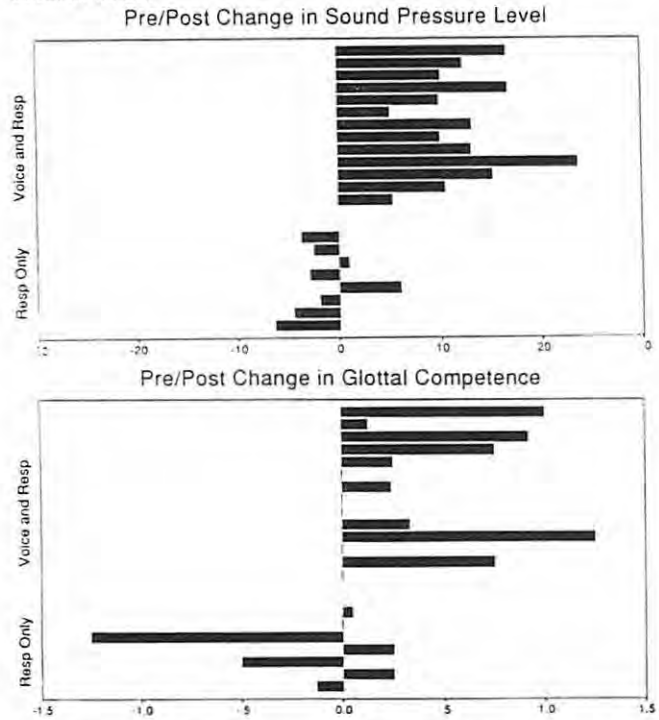


Figure 6a (top). Sound pressure level individual pre-post treatment differences for two treatment types. (Vowel /a/, mouth to sound level meter distance: 50 cm) Figure 6b (bottom). Glottal incompetence rating of individual pre-post treatment comparisons. (Flexible telescope, vowel /i/, NPNL).

The average sound pressure levels (SPL) during sustained vowel phonation were measured before and after therapy. Group mean results, separated by treatment type, are given in Figure 5. These indicate a significant increase ($F=48.4$; $df=1,19$; $p<.01$) in mean SPL for the VR treatment group. Individual data displaying the change in SPL and change in glottal configuration for the treatment types are presented in figure 6a,b.

Discussion

Laryngeal imaging examination is generally used as a diagnostic tool. It is also commonly used to document treatment changes, such as the resolution of vocal nodules or voice improvement following treatment of vocal fold paralysis. Laryngeal pathology in this study did not involve the presence or absence of a lesion, or neuromotor abnormality, such as paralysis. Phonatory pathology generally involved laryngeal hypofunction, resulting in decreased vocal intensity. The major treatment variable assessed in this study was vocal intensity. Therefore, laryngostroboscopy was used to investigate the changes in laryngeal vibration and configuration that resulted from behavioral intervention to increase vocal intensity. Our results indicated that patients with Parkinson's disease who underwent VR therapy demonstrated improved laryngeal adduction, as evidenced by the trend toward normal glottal closure and less glottal incompetence without increased supraglottal hyperfunction. The findings were correlated with increased vocal intensity in the VR therapy group.

Several observations may be made to place these results in context. These include consideration of the spectrum of laryngeal and phonatory dysfunction in Parkinson's disease, and the examination methods that affect these observations.

The laryngeal and phonatory pathology spectrum observed in patients with Parkinson's disease has been reported by several authors. This study was not designed to provide a comprehensive description of laryngeal characteristics in Parkinson's disease, but rather to evaluate pre-post treatment change in glottal closure. did not specifically survey this issue in our patient population, Consequently, the principal factors focused on in this study were vocal fold bowing and resulting glottal insufficiency. The percentage of patients with vocal fold bowing in our series was less than has been reported elsewhere (30 of 32 in Hanson's study). Reasons for these discrepancies may include use of stroboscopic vs. non-stroboscopic imaging, and flexible vs. rigid telescopes.

It would be expected that stroboscopic imaging would improve the assessment of glottal closure patterns. The glottic configuration may be studied during the most closed portion using the freeze frame capabilities of video

tape recorders. Cinegraphic or still photographic imaging cannot capture such details of vocal fold vibration and glottal configuration. The differences in laryngeal imaging with flexible and rigid telescopes may not be so apparent. First, differences in telescopes and their use affected data collection. 31% of patients in this group did not tolerate the rigid telescope examination to obtain an adequate view of the larynx for all assessment variables in the protocol. During the rigid telescope exam the vocal folds were visible for high pitched /i/ phonation but not always visible due to overhanging epiglottis for tasks of normal pitch and loud vowels which were those of interest in this study. Other factors involving this study group likely affected the use of the rigid telescope. Oral topical anesthesia was not usually given to these patients. Many of them complained of swallowing problems, and it was felt that oral topical anesthesia might temporarily exacerbate this. A 70° telescope required placement at an angle toward the tongue base that increased the gag reflex. Use of a 90° rigid telescope may have improved laryngeal visualization in these patients, but was not available. Many patients had head and neck tremors. This precluded stabilization of the rigid telescope to visualize the larynx. Likewise, the vocal tremor in many patients created pitch tracking problems for the stroboscope with both flexible and rigid views.

The flexible fiberscope had advantages when compared with rigid endoscopy. All patients tolerated the flexible fiberscope examination. It allowed visualization of the larynx and speech structures during connected speech and a variety of phonatory and non-phonatory gestures. It's disadvantages related to the optical distortion of the image, and the decreased brightness of the image, especially with strobe light. Nonetheless, the strobe images using the Nagashima strobe and Olympus ENF-P3 fiberscope were generally bright enough to allow excellent visualization of mucosal waves.

Of interest to us was the difference observed in glottal configuration and supraglottal activity between flexible and rigid views. It would be expected that laryngeal images from flexible endoscopy would be more realistic of actual laryngeal activity than from the rigid telescope. Phonation with the rigid telescope in place often requires an unnatural laryngeal posture with the larynx elevated, the tongue extended. The pitch must often be raised to bring the epiglottis forward to see the glottis. Our data support that vocal fold bowing was seen more frequently in our patients with the rigid telescope view than with the flexible endoscope. The observation in differences in glottal closure patterns using flexible vs. rigid telescopes on normal speakers has also been reported (14). This may account for an "overestimation" of the true incidence of vocal fold bowing in Parkinson's disease patients reported in other studies (4). We frequently observed that in the same patient, the glottal configuration viewed with flexible scope appeared to be complete closure or a small anterior chink, while on the rigid telescope view bowing was seen.

The effort level and ability during strobe examination to accomplish the tasks of normal pitch, low pitch, high pitch, soft/loud phonation were defined by the patients and varied widely. This could not be controlled for, except by comparing the patient's own pre and post-treatment efforts. Also, the influence of a possible learning effect in adapting to the laryngeal examination was not accounted for in this study.

Our findings provide physiologic support for increased vocal fold adduction accompanying increased vocal intensity following intensive voice treatment in Parkinson's disease. These findings offer evidence of the underlying physiologic changes that accompany successful voice treatment in Parkinson's disease.Legends

References

1. Logemann JA, Fisher HB, Boshes B, Blonsky ER. Frequency and ocurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson's disease patients. Journal of Speech and Hearing Disorders. 1978;42:47-57.

2. Aronson AE. Clinical Voice Disorders (3rd Ed.) New York: Thieme-Stratton.1990.

3. Darley FL, Aronson A, Brown J. Clusters of deviant speech dimensions in the dysarthrias. Journal of Speech and Hearing Research.1969;12:462-496.

4. Hanson DG, Gerratt BR, Ward PH. Cinegraphic observations of laryngeal function in Parkinson's disease. Laryngoscope.1984;94:348-353.

5. Sarno MT. Speech impairment in Parkinson's disease. Journal of Speech and Hearing Disorders.1968;49:269-275.

6. Weiner WJ, Singer C. Parkinson's disease and pharmacologic treatment programs. J Am Geriat 1989;37:359-363.

7. Ramig LO, Scherer RC, Speech Therapy for Neurologic Disorders of the Larynx in: Blitzer, A., Brin, M., Sasaki, C., Fahn, S., and Harris, K. (eds.) Neurologic Disorders of the Larynx. Thieme Medical Publishers, Inc., New York.1992; pg 248-278.

8. Ramig LO. The role of phonation in speech intelligibility: A review and preliminary data from patients with Parkinson's disease. Intelligibility in Speech Disorders:

Theory, Measurement and Management. John Benjamin:Amsterdam.(in press).

9. Ramig LO, Bonitati C, Horii Y. the efficacy of voice therapy for patients with Parkinson's disease. National Center for Voice and Speech Status and Progress Report. 1991;1:61-86.

10. Scott S, Caird FL. Speech therapy for Parkinson's disease. J Neurol Neurosurg Psychiat. 1983;46:140-144.

11. Robertson S, Thompson F. Speech therapy in Parkinson's disease: a study of the efficacy and long-term effect in intensive treatment. Br J Disord Comm 1984;19:213-224.

12. Ramig LO. Speech therapy for patients with Parkinson's disease. In: Koller W, and Paulson, G. (eds.) Therapy of Parkinson's disease. Marcel Dekker, Inc. New York (in press).

13. Hoehn M, Yahr M, Parkinsonism: onset, progression and mortality. Neurol 1967;17:427.

14. Bless DM, Hirano M, Feder R. Videostroboscopic examination of the larynx. Ear, Nose, and Throat J 1987;66:289-296.

15. Ramig LO, Countryman S, Winholz WS. The effect of intensive voice therapy on acoustic, aerodynamic, and kinematic variables in Parkinson's disease. (in preparation).

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

NCVS Status and Progress Report - 6
May, 1994, 135-146

A "Profiles" Approach to Clinical Voice Measurement: Indirect Physiological Measures

Katherine Verdolini, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Phyllis M. Palmer, M.S.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Two sets of indirect physiological measures were evaluated for their ability to generate profiles for selected diagnostic categories of voice disorders. Forty-five adults participated as subjects, including persons with nodules, granulomas, peripheral vocal fold paralysis, Parkinson's disease, non-organic voice disorders, and control subjects. The two sets of measures made for each subject were: (a) measures thought to reflect membranous vocal fold closure during phonation (maximum phonation time on /a/ or MPT, the S:Z ratio, and high-quiet singing), and (b) measures thought to reflect vocal fold adduction characteristics (laryngeal diadochokinetic or L-DDK rate, strength, and consistency). The results indicated that the combination of the S:Z ratio and L-DDK rate, strength, and consistency distinguished among the diagnostic categories in predictable ways. (MPT and high-quiet singing were insensitive to diagnostic condition.) The findings point to the possibility that physiological profiles might be developed for a range of voice disorders.

Recently, a large number of potentially useful physiologic and acoustic measures have been added to the existing armamentarium relevant to voice and laryngeal function. Although the collective measures permit more subtle and complete descriptions of voice and of the larynx than heretofore possible, in the clinical situation the measures are often used in an ad hoc, empiric manner, without a unifying framework. Stated differently, measures are sometimes made simply because they are available, not because a cohesive set of questions has been formulated. The present study represents an attempt to develop an organizing framework for voice measures, for clinical purposes.

The fundamental premise for the proposed framework is that when combined, a restricted set of perceptual, acoustic, and physiological measures should produce a characteristic "profile" or signature of a voice disorder, consistent with the diagnostic category in most cases. The utility of a profiles approach to voice evaluation would be twofold: (a) to contribute to the differential diagnosis of conditions affecting voice, and (b) to provide quantitative information about the specific ways that voice is affected in different diagnostic conditions. With its emphasis on differential diagnosis and specificity, the approach is conceptually similar to Darley, Aronson, and Brown's "Differential Diagnostic Speech Patterns of the Dysarthrias" (1969) and to Kent and Rosenbek's (1982) acoustic profiles of prosodic disturbances in neurologically lesioned subjects.

Perceptual, acoustic, and indirect physiological measures of voice were made in this study; however, only the latter are described here. For this study, we examined measures that could be made quickly, with inexpensive equipment. In this way the findings would be applicable to clinical settings in which speed and expense are constraints (see also Robin, Somodi, and Luschei, 1991, for similar comments). The framework could be re-evaluated later with more technologically sophisticated approaches, if warranted.

Framework and Predictions

The framework rests on the assumption that physiological distinctions among common voice disorders can be described by two basic parameters: structural and neuro-

logical integrity of the larynx. The equally important corollary assumption is that although structure and neurological functioning in the larynx can be impaired in many ways, in most clinical cases their status is reflected by (a) cycle-to-cycle membranous vocal fold closure characteristics during phonation, which grossly reflects the structural integrity of the membranous folds especially if overall adduction level is taken into account, and (b) vocal fold adduction characteristics, or long-term medial posturing and re-posturing capabilities of the vocal folds, which should usually reflect neural laryngeal integrity.

Taking a closer look at these assumptions, clinically, the most common structural abnormalities in the larynx involve mass lesions of the free margins of the membranous vocal folds (see for example, FitzHugh, Smith, & Chiong, 1958). Regardless of their histological make-up, the lesions protrude into the glottal lumen and interfere with full membranous vocal fold closure. Thus, measures of membranous vocal fold closure should be poor in these cases. Among neural abnormalities affecting the larynx, the most likely nerve to be affected is the recurrent laryngeal nerve (see for example, Crumley, 1994), which regulates vocal fold adduction and abduction. Thus, in cases of neurological impairment, measures of vocal fold adduction and abduction should usually be poor.[1]

Two levels of question were asked in the present study, based on the foregoing assumptions. The first - and conceptually less important question - regarded the ability of selected measures to indirectly reflect the physiological parameters of interest (i.e., the measures' sensitivity). The measures selected to possibly reflect cycle-to-cycle membranous vocal fold closure were maximum phonation time on /a/ (MPT) (Hirano, Koike, & von Leden, 1968; Ptacek & Sander, 1963; Ptacek, Sander, Maloney, & Jackson, 1966), the S:Z ratio (Eckel & Boone, 1981), and high-quiet singing (Bastian, Keidar, & Verdolini, 1990). The measures selected to reflect long-term vocal fold ab- and adduction characteristics (further referred to as vocal fold adduction, for brevity), were laryngeal diadochokinetic (L-DDK) or repeated glottal plosive rate, strength, and consistency. The second and more important question was whether the measures identified as sensitive would produce characteristic profiles for selected diagnostic categories.

The predictions for this second, conceptually more important question are shown in Figure 1 for a range of diagnostic conditions. This figure distinguishes among structural, neurological, and non-organic conditions affecting voice. Further subdivisions indicated in Figure 1 distinguish between structural abnormalities affecting the membranous vocal folds (nodules in the present study) versus structural abnormalities of the arytenoid surface (granulomas), peripheral neurological deficits (unilateral adductory vocal fold paralysis with the affected fold away from midline, in the present study) versus central neurological deficits (Parkinson's disease, in this study), and non-organic (or "functional") voice disorders versus a control condition.

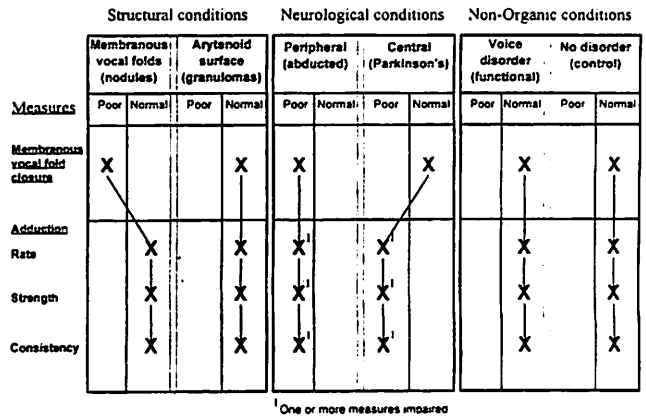


Figure 1. Experimental predictions.

With nodules, measures reflecting membranous vocal fold closure during phonation should be poor. However, measures of vocal fold adduction should be normal because neurological laryngeal functioning is not affected in this condition.

With granulomas restricted to the arytenoid surface, measures of membranous vocal fold closure during phonation should be normal because the membranous folds are not involved. Also measures of vocal fold adduction should be unimpaired, as for more anterior mass lesions.

For unilateral adductory paralysis with the affected fold in the abducted position, measures of vocal fold adduction rate, strength, or consistency might be impaired. Reduced muscle strength is anticipated with paralytic conditions in general, and rate might also be impaired because a decreased shortening velocity in impaired muscle fibers actually underlies muscle paralysis (Luschei, 1991; Robin, Goel, Somodi, & Luschei, 1992). Measures of adduction consistency might be poor because of the dependence of central timing functions on peripheral neurological functions (see for example, Netsell & Rosenbek, 1985). Measures of membranous vocal fold closure during phonation should also be poor because of the generally abducted laryngeal posturing with this condition.

For Parkinsons's disease, the central nervous system disorder evaluated in this study, the predictions would vary depending on disease stage and treatment. However, in general we reasoned that as for peripheral paralyses, measures of adduction rate, strength, or consistency might be impaired. These predictions are based on a mixture of empirical and theoretical reports. Studies on diadochokinetic rates for oral and laryngeal articulators in persons with Parkinson's disease are contradictory, with some indicating

normal rates (Ewanowski, 1964; Ludlow, Connor & Bassich, 1987), others indicating increased rates (for example, Canter 1965), and still others indicating slow rates (Kreul, 1972). Measures of adductory strength might be poor because of vocal fold bowing commonly present in persons with Parkinson's disease (Hansen, Gerratt, & Ward, 1984), or hypometria due to acceleration phenomena (Hirose, Kiritani, Ushijima, & Sawashima, 1981; Netsell, Daniel, & Celesia, 1975) or rigidity (Hunker, Abbs, & Barlow, 1982; Weismer, 1984). Finally, also measures of adduction consistency might be poor, based on reports of acceleration during diadochokinetic task performance in subjects with Parkinson's disease (Hirose et al., 1981; Netsell et al., 1975) or more generally, dysrhythmia (Kreul, 1972). Measures of membranous vocal fold closure during phonation should usually be normal. This last prediction might seem curious, given the common finding of vocal fold bowing - and thus incomplete membranous vocal fold closure - during speech in patients with Parkinson's disease (Hansen et al., 1984). However, persons with Parkinson's disease often override bowing for isolated vowel tasks such as those used in the present study (see for example, Ramig, in press).

Finally, in conditions for which there is no known or suspected organic basis for a voice disorder, membranous vocal fold closure as well as vocal fold adduction characteristics should be normal. This pattern of results should be obtained for non-organic (or "functional") voice disorders as well as for control conditions.[2]

Comparing all the predictions noted in Figure 1, it is apparent that the anticipated profiles for granulomas, functional voice disorders, and the control condition are the same. Perceptual and acoustic measures might help to differentiate these conditions. The point at this juncture is to emphasize that the indirect physiological profiles assessed here would represent a "first pass" at differentiating the diagnostic conditions examined in this study.

Methods

Subjects

Subjects with nodules, granulomas, peripheral paralyses, and functional voice disorders were recruited from our clinical records. Clinical charts from the past year and a half were reviewed, and patients with these diagnoses were contacted by letter and follow-up telephone call to inquire if they were willing to participate in the study for pay ($10). Subjects with Parkinson's disease were recruited from a list of subjects for a previous research protocol conducted in our center. The criteria for contacting these subjects were that they have mild to moderate disease severity (Stages 2-3, Hoehn & Yahr, 1967), evidence of speech problems (by self or clinician report), and no dementia (based on The Iowa Screening for Mental Decline, Eslinger, Damasio, & Benton, 1984), based on screenings conducted no more than a few months before the current protocol. All subjects with Parkinson's disease were medicated,[3] without reported general motor or specific speech fluctuations over the drug cycle. Control subjects were mostly recruited among friends or spouses of other subjects. Control subjects were included not as specific controls for those in other diagnostic categories, because age and gender-based normative data served as direct controls for these groups. Control subjects were rather included in the present study as a check on the validity of the normative data base that we would use.

The data for subjects who were discovered during the experiment to have a different diagnostic condition than the one originally thought (for example, normal larynx as opposed to nodules), for subjects discovered to have a mixed diagnosis (for example, granulomas and nodules), and for subjects discovered to have other significant medical conditions (for example, heart disease) were excluded from analysis. Also the data for subjects who had undergone surgical management were excluded, with the exception of patients with granulomas (two of the three subjects with granulomas had previously undergone surgical management). Subject recruitment continued until 10 subjects had been evaluated for each diagnostic category of interest, or until the maximum number of available subjects had participated (less than 10).

The result was a set of 45 subjects whose data could be analyzed. The number of subjects and the age, gender, and educational characteristics for each diagnostic category are indicated in Table 1.

Table 1.
Subject Characteristics by Diagnostic Category

	NODULES	GRANULOMAS	PARALYSIS	PARKINSON'S DISEASE	"FUNCTIONAL" VOICE DISORDER	CONTROL
TOTAL N	8	3	4	10	10	10
N Female	8	0	4	3	6	6
N Male	0	3	0	7	4	4
Age (yrs)	17 - 39	44 - 50	26 - 77	62 - 78	19 - 50	19 - 58
Education (yrs)	15	16	13	13	16	18

General Procedures

Subjects were first informed about the general purpose of the study ("...[to evaluate] a series of voice tests"), and they completed Informed Consent Forms. Then subjects underwent voice testing by an experimenter who was a speech/language pathologist with 11 years' experience in clinical management of voice disorders, and who was technically uninformed about subjects' diagnostic conditions. Of course, clues about likely diagnostic categories were available based on subjects' surface characteristics such as apparent age and gait. An elderly, festinating

subject was more likely to have Parkinson's disease than nodules. However, surface clues were not particularly telling for most subjects, and thus some degree of experimenter "blindness" was preserved.

The evaluation started by the examiner engaging the subject in brief conversation. ("So, where did you come from this morning?") Then the experimenter recorded demographic information (subject's name, age, gender, and years of education, by subject report). On the basis of these verbal exchanges, the examiner indicated a label to describe the subject's voice (normal, hoarse, weak, etc.), and perceived severity (normal, mild, moderate, moderately-severe, severe) on a data sheet. These evaluations constituted the basis for an auditory-perceptual measure of voice, not further discussed here. The evaluation proceeded with the elicitation of a series of acoustic and indirect physiological measures, and with a measure of phonatory effort. Acoustic measures included fundamental frequency in speech, semitone pitch range, and phonetogram (intensity over frequency range). The measure of phonatory effort involved a perceptual scaling of phonatory effort by the subject. Acoustic and effort measures are not further described here. Indirect physiological measures, which are the focus of the present report, included MPT on /a/, the S:Z ratio, high-quiet singing (as possible measures of membranous vocal fold closure during phonation), and laryngeal diadochokinetic (L-DDK) rate, strength, and consistency (as measures of laryngeal adduction).[4]

Measures were elicited in quasi-random order across subjects. The exception to random ordering was that the auditory-perceptual rating was made first, measures requiring maximum phoneme prolongations were never elicited successively (or recovery time was allowed), the order of /s/ and /z/ prolongation trials was varied within and across subjects, and the phonetogram was always elicited last.

After voice measures were completed (requiring about 15 minutes for most subjects, but usually longer for subjects with Parkinson's disease), the same experimenter or another one conducted a videoscopic examination of the larynx. General medical history was obtained from the patient at this time, and the data for subjects with current or prior significant medical conditions beyond those of interest (nodules, granulomas, etc.) would be excluded from further analysis. The laryngoscopic examination was conducted in the standard clinical fashion (see next section). Two experimenters, speech-language pathologists with extensive experience in videoscopic examinations, viewed the examination and conferred about the findings. In the event of a disagreement or uncertainty about diagnostic status, or when the results appeared inconsistent with prior medical records, a physician specialized in voice was consulted (three times). As noted, the data for subjects with different laryngeal conditions than originally thought, or mixed conditions, were excluded from analysis.

Specific Procedures and Equipment for Voice and Laryngeal Measures

For all voice measures, the procedures were similar to those indicated in the studies generating the normative data that we used. For MPT, the reference study was by Ptacek and colleagues (Ptacek, Sander, Maloney, & Jackson, 1966). Subjects were instructed: "Now I'm going to have you say the vowel /a/ for as long as you can. In just a moment, you will take a deep breath, and say /a/ for as long as you can, at this pitch." A pitch of G#3 (210 Hz) was provided for females, and a pitch of C3 (130 Hz) was provided for males, using a Casio portable keyboard. Subjects were further instructed to "keep your voice at about the same loudness, so that this little needle stays right around this mark" (between 0 and 1 on the Volume Unit Meter of a portable Sony Digital Audio Tape Recorder, Model TCD-D3, corresponding to 82 dB measured at two inches). Subjects were then given an opportunity to practice the criterion performance by the clinician saying, "You go ahead and try this now, just to get the hang of it." The subject was allowed to repeat the task, without continuing to actual maximum phonation time, over one or more trials until the clinician was convinced that the subject understood the task and was able to perform it. Then the clinician continued with, "OK, now we're ready. Take a deep breath and say /a/ as you did before, for as long as you can." The experimenter timed subjects' performance using a Micronta LCD quartz stopwatch, and recorded the performance in seconds to two decimal places. Three trials were used, with similar instructions before each trial (further practice trials were not used). Trials for which pitch or loudness specifications were violated were interrupted and repeated (approximately 5% of trials.)

For the S:Z ratio, the reference study was by Eckel and Boone (1981). Subjects were instructed, "Now I'd like you to hold out the sound /sss/ (or /zzz/)." A brief example of the phoneme was provided. "Take a deep breath and hold out /sss/ (or /zzz/) as long as you can. Go ahead..." The trial was timed with the stopwatch, and recorded in seconds to the second decimal place. Two prolongations of /s/ and two of /z/ were performed by each subject, varying the order of /s/ and /z/ within and across subjects.

For high-quiet singing, the reference study was by Bastian and colleagues (1990). For this task, subjects were instructed, "Now I'm going to have you sing the first two lines of 'Happy Birthday' at a high pitch, as quietly as you can. Don't worry if you think you can't sing. We are not listening for a beautiful voice. Do it like this..." The clinician provided a model of high-quiet singing, starting the "Happy Birthday" song at E4 (329 Hz for males) and A4

(440 Hz) for females. If the subject was male, the additional instruction was added, "You should do this in falsetto, if you can." Then the clinician proceeded with further explanations. "Now I'm going to set up this little instrument here." A Radio Shack Realistic Sound Level Meter set on the A-weighted scale in slow response mode was positioned with a microphone-to-mouth distance of three feet relative to the subject. "Now you go ahead and try the task." The subject was provided the starting pitch with the Casio keyboard, and repeated the task until the experimenter was convinced that he or she could understood it and could perform it within the range of his or her better abilities. Adequate performance was defined as the subject starting the task on the correct pitch, as perceived by the clinician, and continuing to sing the "Happy Birthday" song as quietly as possible (again as perceived by the clinician) in the same general pitch range, even if the correct pitches were not produced throughout singing. The experimenter then rated the subject's performance on a scale from 1-10. A score of 1 was used to indicate "extremely delayed phonatory onset, discontinuous phonation, and a perception that performance was 'not very quiet.'" A score of 10 was used to indicate "immediate phonatory onset, completely continuous phonation, and the perception of very quiet voice." Although one trial was used for most subjects, repeated trials were performed when the subject's performance was inadequate, or when the experimenter was not confident about the rating she should indicate.

For L-DDK, the reference study was again by Ptacek and colleagues (1966). For this task, the subject was told, "Now I'd like you to make this sound as quickly as you can" (the clinician showed the subject a card indicating "uh") "with a little click, like this..."[5] The clinician then demonstrated rapid glottal plosives for a few seconds. Then the subject was given an opportunity to practice the task: "Now you try it..." The subject practiced the task over one or more trials until the clinician was convinced that the subject understood the task and could perform it in the range of his or her ability. Then the subject was told, "Now take a deep breath, and make these sounds as quickly as you can until I stop you. It will be about 7 seconds. Ready ... begin..." Glottal stops produced by the subject were then indicated with the clinician dotting a pencil or pen on the data collection sheet for each production. The clinician stopped the pencil dotting when 7 seconds had elapsed on the stopwatch, and further indicated to the subject to stop the trial. If the subject could not do the task for the full 7 seconds, the trial was nonetheless accepted (approximately 2% of cases). The rate of glottal stops per second was calculated by dividing the number of glottal stops produced by the number of seconds actually performed (to the second decimal place). The clinician also rated the strength of the glottal stops (auditorially perceived rise-time and loudness) and their consistency over time (rhythmia), using a dichotomous measure ("good" versus "poor").

After all the voice measures had been elicited, a videoendoscopic examination of the larynx was performed. For this examination, a Kay Elemetrics videostroboscopic unit was used, with a 70 degree rigid scope, a Kay RLS 9100 Rhinolaryngeal stroboscope light source, a Panasonic GP-KS152 camera, and a Super NEC VHS, RS-232C. The examination was conducted in the standard clinical fashion, using a 1% Pontacaine or a 4% Lidocaine topical anesthetic spray for subjects with a strong gag reflex (about 10% of subjects).

The reliability of the measures was checked by the experimenter's re-doing four of each of the measures, randomly selected from tape recordings of the sessions. Inter-test correlations and difference scores were then computed to evaluate the stability of the measures.

Calculations of Normalcy/Abnormality

For MPT, both the longest and the average performance across three trials for each subject were used for analysis. The reason for the dual calculations is that it is not clear from the normative study by Ptacek and colleagues (1966) which of these values was used. The result for each measure was then transformed to a z-score for each subject, with reference to the subject's age and gender (see Appendix A). A z-score ≤ -1 was considered an indication of poor performance, and $z > -1$ was considered normal (for this study, distinctions were not made between normal and superior performance, for any of the measures). For example, a z-score of -1.5 for MPT would have been considered poor performance, and a z-score of 1.5 would have been considered normal.

For the S:Z ratio, the longest /s/ was divided by the longest /z/ to calculate the ratio. A z-score for the S:Z ratio was then calculated for each subject (see Appendix A). For this measure, a z-score of +1 or greater was considered an indication of poor performance, and $z < 1$ was considered normal.For high-quiet singing, a score of 7 or less on the single trial was considered indicative of poor performance, and a score of 8 or more was considered normal (Appendix A).

For L-DDK performance, the rate of glottal stops per second on the single trial was used to calculate a z-score relative to age and gender (Appendix A). A z-score of ≤ -1 was considered poor, and $z > -1$ was considered normal. Strength and consistency of glottal stops was considered "poor" or "good" straightaway by the clinician during data collection.

The only "measure" of the larynx was a consensus about the current diagnostic condition by two speech/language pathologists and, when required, also a physician specialized in voice.

Results

The results were evaluated in two stages. The first stage involved a sensitivity analysis (i.e., an evaluation of measures' ability to reflect what they were supposed to reflect). The measures' reliability was also assessed. Insensitive measures were discarded. In the second stage of analysis, the ability of the surviving measures to generate characteristic profiles consistent with predictions for the diagnostic categories was evaluated.

Stage I Analysis: Exclusion of Insensitive Measures (Validity and Reliability)

Measures proposed to reflect cycle-to-cycle membranous vocal fold closure during phonation were assessed for their ability to correctly reflect likely vocal fold closure characteristics (validity), given the results of the videoscopic examination. That is, z-scores for average MPT, maximum MPT, and the S:Z ratio, and the ordinal score for high-quiet singing were considered for each subject. Each score that correctly reflected likely poor membranous vocal fold closure during phonation according to our scoring criteria (see preceding section), given the videoscopic results, was considered a "hit" (correct detection of an abnormality; for example, $z = -2.0$ for MPT, in the case of nodules). Each score that implied poor membranous vocal fold closure when the videoscopic examination indicated good closure was considered a "false alarm" (F.A., incorrect detection; for example, $z = -1.2$ for MPT, in the case of granulomas). Each score implying good membranous vocal fold closure when the videoscopic examination implied poor closure was considered a "miss" (failure to detect; for example, $z = 1.5$ for MPT, in the case of nodules). Finally, each score that indicated good membranous vocal fold closure, consistent with the videoscopic impression, was considered a "correct rejection" of pathology (C.R.; for example, = 1.5 for MPT, for granulomas).[6] The percent of "false alarms" and "misses" were of particular interest, because they indicated the likelihood of a clinically incorrect value. The percents were calculated with the formulas:

$$\text{Percent F.A.} = \frac{\text{N of F.A.}}{\text{Total N}} \times 100$$

$$\text{Percent Miss} = \frac{\text{N of Miss}}{\text{Total N}} \times 100$$

The overall error rate was the percent of F.A. plus the percent of Misses. The results are shown in Table 2. The overall error rates for average MPT, maximum MPT, and high-quiet singing were between about 30% - 40%. Most of the errors were attributable to a high false alarm rate, especially for MPT. In contrast, the overall error rate for the S:Z ratio was 13%. Differences in the sensitivity of these measures was unrelated to interjudge scoring-rescoring reliability, which based on a random re-sampling of measures was equally good for all measures ($r = .99$ for MPT, with an average difference between first and second scorings of 0.12 s; $r = .95$ for high quiet singing, with an average difference between scorings of -0.50; $r = .99$ for the S:Z ratio, with an average difference between scorings of -0.0025). Based on these results, of MPT, the S:Z ratio, and high-quiet singing, only the S:Z ratio was retained in further analyses.

Table 2.
Percent False Alarms (% FA), percent Misses (% Miss), and Percent Overall Error for Voice Measures.

MEASURE	% FA	% MISS	% OVERALL ERROR
MPT average	33%	4%	38%
MPT maximum	27%	4%	31%
S:Z ratio	11%	2%	13%
Happy Birthday Test	14%	14%	28%

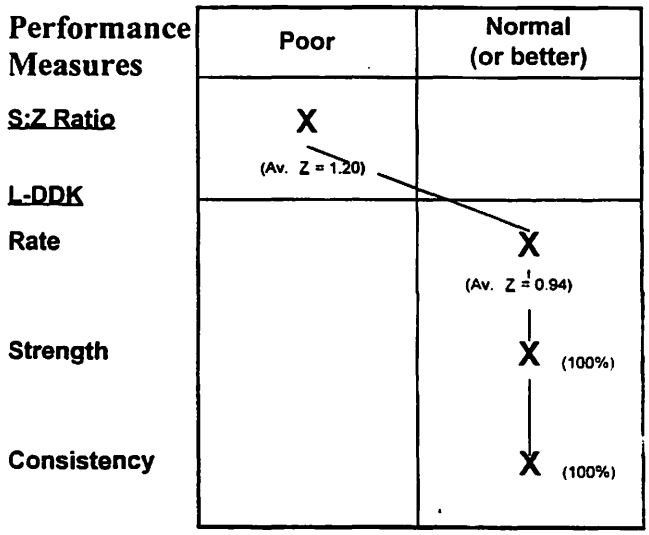
L-DDK values were not evaluated for their sensitivity in reflecting laryngeal adduction characteristics in this stage of the analysis, because there was no good standard against which to assess them. However, they were assessed for interjudge scoring-rescoring reliability. Rate and strength measures were stable across first and second measurings (for rate, $r = 0.99$, with an average difference of -0.50 between scorings; for strength, $r = 1.0$, with an average difference of 0.00 between scorings). Consistency measures were somewhat less stable, because in one of four cases re-evaluated, consistency was judged poor on the first trial and good on the second trial ($r = .58$).

Stage II analysis: Ability of surviving measures to generate distinctive profiles

The results for the S:Z ratio and L-DDK measures are shown in Figures 2-7, as compared with predicted results. For nodules, seven out of eight subjects produced the predicted profile of poor S:Z ratio (shown by seven of eight subjects, average z-score = 1.20) and normal L-DDK rate (produced by all subjects, average z-score = 0.94), strength (all subjects) and consistency (all subjects; Figure 2). The number of subjects showing this general profile exceeded chance levels ($p < .0001$), using a binomial test.[7]

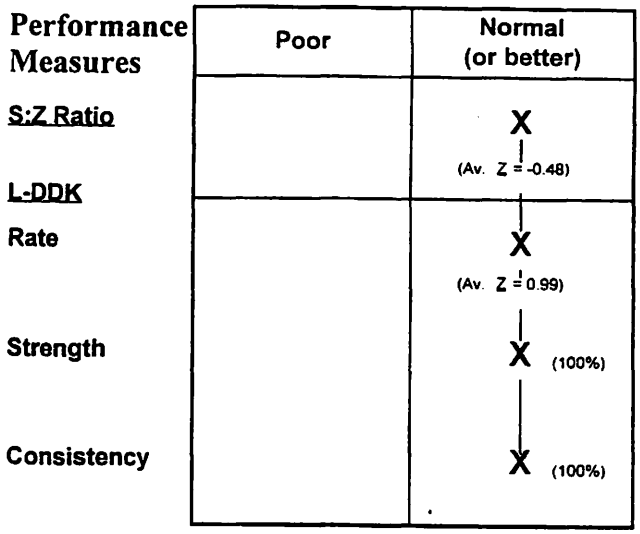
For subjects with granulomas (Figure 3), the results were also consistent with the predictions. Three of three subjects ($p < .001$) produced the anticipated pattern of normal S:Z ratio (average z-score = -0.48), normal L-DDK rate (average z-score = 0.99), and good L-DDK strength and consistency.

Nodules



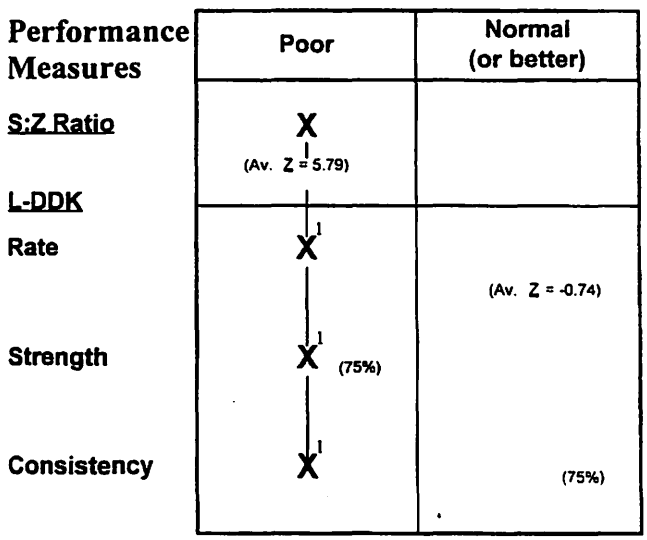
7/8 subjects showed predicted pattern.
($p < .0001$)

Granulomas



3/3 subjects showed predicted pattern
($p < .001$)

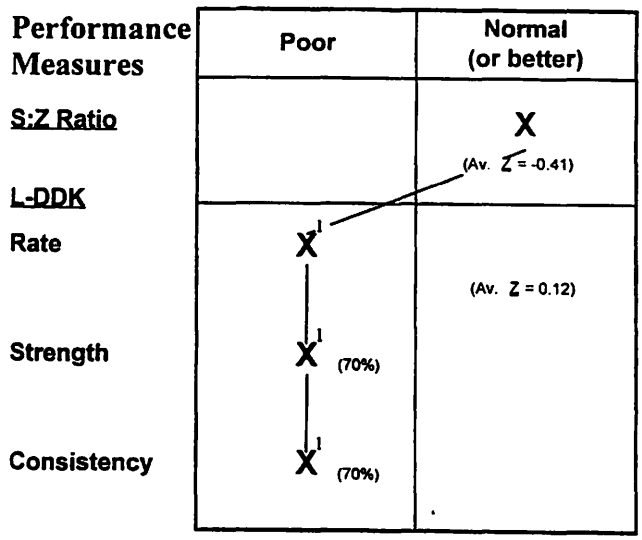
Paralysis



4/4 subjects showed predicted pattern.
($p < .01$)

[1] One or more measures impaired

Parkinson's Disease

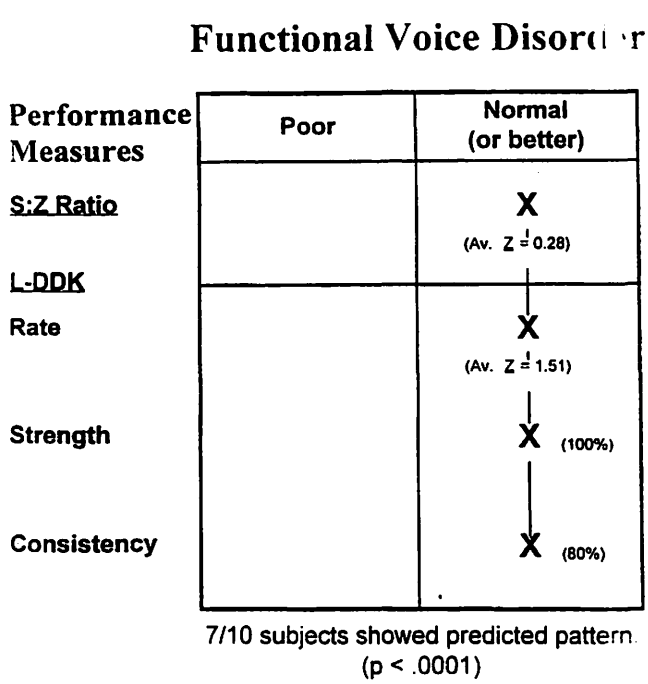


8/10 subjects showed predicted pattern.
($p < .0001$)

[1] One or more measures impaired

Figure 2 (upper left). Predicted results for subjects with nodules, with obtained results in parentheses. (Number of subjects producing the predicted pattern also indicated.) Figure 3 (upper right). Predicted results for subjects with granulomas, with obtained results in parentheses. (Number of subjects producing the predicted pattern also indicated.) Figure 4 (lower left). Predicted results for subjects with paralysis, with obtained results in parentheses. (Number of subjects producing the predicted pattern also indicated.) Figure 5 (lower right). Predicted results for subjects with Parkinson's disease, with obtained results in parentheses. (Number of subjects producing the predicted pattern also indicated.)

For subjects with unilateral peripheral paralyses (Figure 4), again, the results were as predicted. All subjects produced an extremely poor S:Z ratio (average z-score = 5.79), and poor L-DDK rate, strength, or consistency (two of four subjects had poor rate, although the average z-score for rate was z = -0.74; three of four subjects had poor L-DDK strength, and one of the four subjects had poor consistency). Overall, four of four subjects produced the predicted profile ($p < .01$).



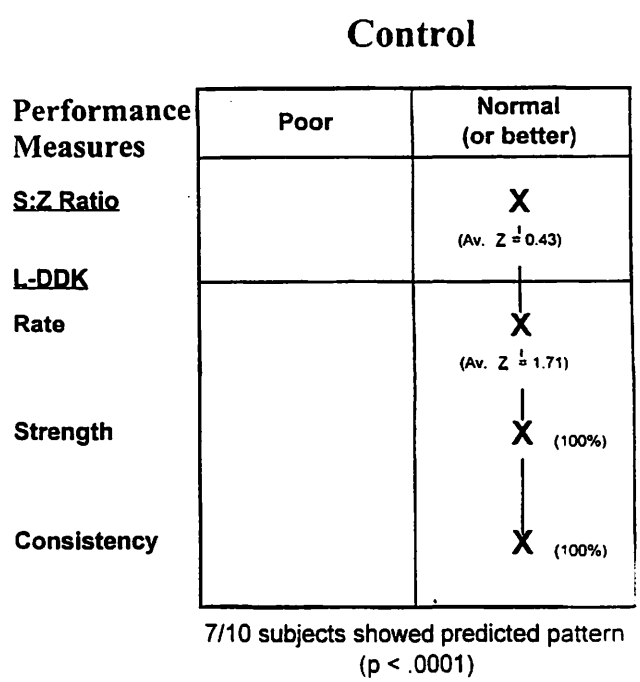


Figure 6 (left). Predicted results for subjects with a functional voice disorder, with obtained results in parentheses. (Number of subjects producing the predicted pattern also indicated.) Figure 7 (right). Predicted results for control subjects, with obtained results in parentheses. (Number of subjects producing the predicted pattern also indicated.)

Subjects with Parkinson's disease also generally showed the predicted pattern (Figure 5). Eight out of ten subjects in this diagnostic category showed the anticipated profile of a normal S:Z ratio (average z-score = -0.41, with 9/10 subjects having a normal ratio), and impairment in one or more aspect of L-DDK performance. L-DDK rate was usually, but not always normal (average z-score = 0.12, with a normal rate shown by 9/10 subjects). L-DDK strength was impaired for most subjects (7/10), and L-DDK consistency was also impaired for most subjects (7/10). The likelihood of 8/10 subjects showing the predicted profile exceeded chance levels ($p < .0001$).

The results for subjects without known organic conditions ("functional voice disorders" and control subjects) were similar, and were also as anticipated (Figures 6 and 7). In both groups, the average S:Z ratio was normal (z-score = 0.28 for subjects with functional disorders and z-score = 0.43 for control subjects, with 9/10 and 7/10 subjects producing a normal ratio in each group, respectively). L-DDK rate and strength were normal for all subjects in both groups (for rate, the average z-score = 1.51 for subjects with functional voice disorders, and 1.71 for control subjects, with normal rate shown by all subjects; for strength, again, performance was normal for all subjects). L-DDK consistency was normal for 8/10 subjects with a functional voice disorder and for 10/10 control subjects. The likelihood of the number of subjects who showed the anticipated pattern exceeded chance levels for both groups ($p < .0001$).

The final result regarded a pertinent clinical question: Given a characteristic (predicted) profile, what was the likelihood of the corresponding diagnosis? The answer is shown in Table 3. Seven out of 11 subjects with a "nodule" profile had nodules (64%). The only subjects with this profile who did not have nodules were one subject with a functional voice disorder, and 3 control subjects. Seventeen of 19 subjects with a granuloma/functional voice disorder/control profile had one of these conditions (89%). The exceptions were one subject with nodules, and one with Parkinson's disease. Four of five subjects who produced a "unilateral paralysis" profile had a unilateral paralysis (80%), with a fifth subject having Parkinson's disease. Finally, 8/10 subjects with a Parkinson's profile had known Parkinson's disease (80%). The two subjects who did not had a "functional" voice disorder.

Discussion

In the present study, two sets of indirect physiological voice measures tended to generate profiles reflecting diagnostic categories for a voice condition, as predicted. Specifically, the combination of the S:Z ratio (reflecting apparent membranous vocal fold closure during phonation) and L-DDK rate, strength, and consistency (proposed to

Table 3.
Likelihood (Expressed in Percent P) of Corresponding Diagnosis, Given Characteristic Profile.

DIAGNOSIS	NODULES	GRANULOMA/ FUNCTIONAL/ CONTROL	PARALYSIS	PARKINSON'S DISEASE	P
PROFILE		0/			
Nodule	7	1/ 3	0	0	7/11=64%
Granuloma/ Functional/ Control	1	3/ 7/ 7	0	1	17/19=89%
Paralysis	0	0	4	1	4/5=80%
Parkinson's Disease	0	0/ 2/ 0	0	8	8/10=80%
	8	23	4	10	

generally reflect neurological control of the larynx) largely distinguished between subjects with nodules, granuloma/ functional voice disorders/control conditions, unilateral adductory paralysis, and Parkinson's disease.

These findings are encouraging in terms of their potential for characterizing and quantifying disruptions to voice functions that are distinctive to a diagnostic category, for the categories examined. The results are also encouraging because profiles were generated with two simple tests (S:Z ratio and L-DDK) that can be conducted in a few minutes, and that do not require any equipment beyond a stopwatch. However, several limitations and caveats need to be noted.

First, as predicted, three conditions resulted in the same physiological profile: granulomas, functional voice disorders, and control condition. Clearly, other measures are required to distinguish these conditions and many others that are encountered clinically. Additional measures might be physiological, perceptual, or acoustic, or a combination of all these. Also, finer-grained analyses might be useful in distinguishing diagnostic categories, in some cases. A case in point is the S:Z ratio. Not only the valence (poor/normal) but also the magnitude of this measure might contribute to a profile. An extremely poor ratio might point to an adductory paralysis, as compared with a marginally poor ratio, which would be more consistent with mass lesions. The point is that the present approach represents a framework for thinking about distinctive patterns in voice analysis, but the approach is anything but complete.

A second comment is that it would be valuable to better assess the measures examined for their validity in reflecting what they are supposed to reflect. That is, the measures should be assessed against stronger "golden standards" than those used in the present study, using direct physiological measures. For example, the S:Z ratio as a reflection of membranous vocal fold closure could be compared with minimum airflow during phonation, directly indicating the comparative persisting glottal area between the vocal folds. L-DDK rate and consistency ratings could be evaluated more precisely with graphic recorders or other instrumented devices, and L-DDK strength ratings could be evaluated by examining the rise-time and the amplitude of acoustic signals from L-DDK trials.

Another, specific point regards the results for subjects with peripheral paralysis. Although all four subjects produced the anticipated profile of poor S:Z ratio and poor rate, strength, or consistency of L-DDK performance, in fact the average performance for rate and the modal performance for consistency were normal for this group. Larger subject numbers will be required to determine whether these findings are characteristic of this subject population or not.

Also, for subjects with Parkinson's disease, nine of ten subjects showed impaired L-DDK rate, strength, or consistency, as predicted. However, L-DDK rate was in fact normal for most subjects (9/10), as was the average rate for the group. These results, together with previous contradictory reports about diadochokinetic rates for patients with Parkinson's disease (Canter, 1969; Ewanowski, 1964; Kreul, 1972; Ludlow et al., 1987) make us hesitant to draw any conclusions about L-DDK rate in this subject group.

Another comment regards the finding that two subjects with "functional" voice disorders had profiles similar to the one for Parkinson's disease. Few clinicians would make anything of this finding except to monitor the patients' status for other possible indications of organic involvement. The point at this juncture is that, as already argued in the introduction, one advantage of the profiles approach might eventually be to provide diagnostically sensitive information that may not always be uncovered by standard medical examination. In summary, the voice measurement approach evaluated in this study provided some evidence that characteristic physiological profiles might be developed for a range of voice disorders, using a fast, cheap, but principled procedure. At the same time, it is important to emphasize again what the approach is not, in its current or even more advanced stages. Although the results might contribute to in the differential diagnosis of voice disorders in some cases, the approach is not a substitute for medical diagnosis.

The approach is further not a complete approach to assessment. Case history, perceptual, and acoustic measures are needed to fully address the distinctive nature of the voice disorder and also other clinical questions such as the precipitating and maintaining factors for the disorder.

The approach is also not intended to promote a "cookbook" approach to voice evaluation. A thorough understanding of voice disorders, of the patient's history, and of the measures themselves is required in order to interpret the results. A case in point is the example of poor arytenoid adduction measures, which may be obtained with some mechanical conditions such as arytenoid fixation and

subluxation, as opposed to the more usually encountered neurological problems. The clinician who uses the measures described here must be insistent about considering the specific functions the measures reflect rather than interpreting the measures in any rote fashion.

Finally, the approach is not an exhortation to abandon more sophisticated measures for those who have access to the technology.

Appendix A.
Norms[a]

Maximum Phonation Time

From Ptacek et al., 1966

Age and gender	Average	SD
18-39 yr, M	24.6 s	6.7 s
18-38 yr, F	20.9 s	5.7 s
68-89 yr, M	18.1 s	6.6 s
66-93 yr, F	14.2 s	5.6 s

High quiet singing

From Bastian et al., 1990

Age and gender	Ordinal rating
17-72 yr, M/F	8-10, no laryngeal abnormality or voice disorder

S:Z ratio

From Eckel & Boone, 1981

Age and gender	Average	SD
8-88, M/F	0.99	0.36

L-DDK rate

From Ptacek et al., 1966

Age and gender	Average	SD
18-39, M	5.1/s	1.0
18-38, F	5.3/s	0.8
68-89, M	4.1/s	0.9
66-93, F	3.9/s	1.3

References

Bastian, R.W., Keidar, A.K., & Verdolini-Marston, K. (1990). Simple vocal tasks for detecting vocal fold swelling. Journal of Voice, 4, 172-183.

Canter, G.J. (1965). Speech characteristics of patients with Parkinsons' disease: III. Articulation diadochokinesis, and over-all speech adequacy. Journal of Speech and Hearing Disorders, 30, 217-224.

Crumley, R.L. (1994). Unilateral recurrent laryngeal nerve paralysis. Journal of Voice, 8, 79-83.

Darley, F.L., Aronson, A.E., & Brown, J.R.(1969). Differential diagnostic patterns of dysarthria. Journal of Speech and Hearing Research, 12, 246-269.

Eckel, F.C. and Boone, D.R. (1981). The s/z ratio as an indicator of laryngeal pathology. Journal of Speech and Hearing Disorders, 46, 147-149.

Eslinger, P.J., Damasio, A.R., & Benton, A.L. (1984). The Iowa screening for mental decline. Iowa City, Iowa: Department of Neurology (Division of Behavioral Neurology), The University of Iowa College of Medicine.

Ewanowski, S.J. (1964). Selected motor-speech behavior of patients with parkinsonism. Unpublished doctoral dissertation, University of Wisconsin, Madison.

FitzHugh, G.S., Smith, D.E., & Chiong, A.T. (1958). Pathology of 300 clinically benign lesions of the vocal cords. Laryngoscope, 68, 855-861.

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

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

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

Hirose, H., Kiritani, S., Ushijima, T., & Sawashima, M. (1981). Patterns of dysarthric movements in patients with Parkinsonism. Folia Phoniatrica, 33, 204-215.

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

Hunker, C.J., Abbs, J.H., & Barlow, S.M. (1982). The relationship between Parkinsonian rigidity and hypokinesia in the orofacial system: A quantitative analysis. Neurology, 32, 749-754.

Kent, R.D., & Rosenbek, J.C. (1982). Prosodic disturbance and neurologic lesion. Brain and Language, 15, 259-291.

Kreul, E.J. (1972). Neuromuscular control examination (NMC) for parkinsonism: Vowel prolongations and diadochokinetic and reading rates. Journal of Speech and Hearing Research, 15, 72-83.

Leanderson, R., Persson, A., & Öhman, S. (1970). Electromyographic studies of the function of the facial muscles in dysarthria. Acta Otolaryngologica, Supplement 263, 89-94.

Ludlow, C.L., Connor, N.P., & Bassich, C.J. (1987). Speech timing in Parkinson's and Huntington's Disease. Brain and Language, 32, 195-214.

Luschei, E.S. (1991). Development of objective standards of nonoral strength and performance: An advocate's view. In C.A. Moore, K.M. Yorkston, & D.R. Beukelman, Dysarthria and apraxia of speech: Perspectives on management (pp. 173-184). Baltimore: Paul H. Brookes.

Netsell, R., & Rosenbek, J.C. (1985). Understanding and treating the dysarthrias. In J. Darby (Ed.), Speech and language evaluation in neurology: Adult disorders (pp. 363-392). Orlando, Florida: Grune & Stratton.

Netsell, R., Daniel, B., & Celesia, G.G. (1975). Acceleration and weakness in Parkinsonian dysarthria. Journal of Speech and Hearing Disorders, 40, 170-178.

Ptacek, P.H., & Sander, E.K. (1963). Maximum duration of phonation. Journal of Speech and Hearing Disorders, 28, 171-182.

Ptacek, P.H., Sander, E.K., Maloney, W.H., & Jackson, C.R. (1966). Phonatory and related changes with advanced age. Journal of Speech and Hearing Research, 9, 353-360.

Ramig, L.O. (in press). Speech therapy for patients with Parkinson's disease. In W. Koller & G. Paulson (Eds.), Therapy of Parkinson's disease. New York: Marcel Dekker.

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

Robin, D.A., Somodi, L.B., & Luschei, E.S. (1991). Measurement of tongue strength and endurance in normal and articulation disordered subjects. In C.A. Moore, K.M. Yorkston, & D.R. Beukelman (Eds.), Dysarthria and apraxia of speech: Perspectives on management (pp. 173-184). Baltimore: Paul H. Brookes.

Weismer, G. (1984). Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In M. McNeil, J. Rosenbek, & A. Aronson (Eds.), The dysarthrias: Physiology, acoustics, perception, management (pp. 101-130). New York: College Hill Press.

Acknowledgements

An earlier version of this paper was presented at the Twenty-Second Annual Symposium: Care of the Professional Voice Philadelphia, Pennsylvania, June 7-12, 1993. The study was partly supported by Grant No. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders, and by the Department of Speech Pathology and Audiology, The University of Iowa. Hughlett Morris is thanked for his gracious facilitation of this protocol, and Lori Somodi and Karla Wedemeyer are acknowledged for pilot work. Kelvin Lehrman and Linnie Southard are thanked for assistance with graphics, and Erich Luschei, Donald Robin, and an anonymous reviewer are thanked for their valuable comments on earlier versions of the manuscript.

Endnotes

[1] The proposed mapping between neurological functions and vocal fold adduction characteristics is admittedly imperfect. Some neurological conditions, in particular those involving an isolated deficit to the superior laryngeal nerve, might result in normal adduction performance. Also, some non-neurological conditions such as arytenoid fixation or subluxation might result in poor adduction performance. Such cases are quite rare in most clinical practices. We consider this an acceptable caveat at this juncture, which does not compromise the overall line of reasoning.

[2] Spasmodic dysphonia is excluded as an example of a functional voice disorder in this study.

[3] Across all subjects with Parkinson's disease, the list of reported medications that we could verify in the Physician's Desk Reference included (in alphabetical order): Allopurinal, Artane, Capoten, Eldepryl, Inderal, Ethmozine, Flexeril, Hydrochlorothiazide, Macrobid, Pamelor, Pepcid, Permax, Symmetrel, Sinemet CR, Synthroid, and Vitamin E.

[4] These measures are considered indirect because although they may reflect the physiological parameters of interest, they do not provide direct measures of them.

[5] Note that from the report by Ptacek et al. (1966), it was not clear if glottal plosives were included at the outset of the vowel productions or not. We included plosives because we

considered that L-DDK strength ratings in particular would be facilitated by their inclusion. The performance of our control subjects indicated that the normative data by Ptacek et al. could be used for the task as we required it.

[6] In these analyses, subjects with Parkinson's disease were considered to have "good membranous vocal fold closure" even in the presence of bowing because of these subjects' ability to achieve good closure for prolonged, voluntary tasks such as MPT (see Ramig, in press). The proportion of false alarms and misses for all measures was almost identical when subjects with Parkinson's disease were included in the analyses and when they were not.

[7] For all results, a binomial test was conducted based on the question: "What is the likelihood that `X' out of `Y' subjects (or more) produced this specific pattern of results across the number of predicted measures?" Except for neurological conditions, the number of predicted measures was four (S:Z ratio, and adductions rate, strength, and consistency). For neurological conditions, the number of predicted measures was two (S:Z ratio, and adduction rate/ strength/or consistency).

[8] When a subject's age did not fall within the specific age ranges indicated, the nearest range was used.

NCVS Status and Progress Report - 6
May, 1994, 147-153

Physiological Study of "Resonant Voice"

Katherine Verdolini, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
David G. Druker, M.S.
Department of Speech Pathology and Audiology, The University of Iowa
Phyllis M. Palmer, M.S.
Department of Speech Pathology and Audiology, The University of Iowa
Hani Samawi, C.Ph.
Department of Preventive Medicine and Environmental Health,The University of Iowa

Abstract

The purpose of this experiment was to conduct a comparative examination of the laryngeal configuration for a voicing mode called "resonant" voice, recently proposed for the treatment of nodules (Verdolini-Marston, Burke, Lessac, Glaze, & Caldwell, in press). Twelve adult trained voice users participated in the experiment as subjects, six with normal larynges and six with nodules. Subjects produced repeated tokens of resonant, pressed, breathy, and normal voice during sustained vowels. For each voice type, laryngeal adduction was estimated using two measures: an indirect measure, the electroglottographic closed quotient (EGG CQ), and a direct measure, an ordinal perceptual adduction rating from videoscopic views of the larynx. The results indicated that, for both subject groups and for both measures of adduction, resonant voice was produced with an intermediate adduction level, between the level for pressed voice (hyperadducted), and breathy voice (hypoadducted). Discussion focuses on the possible value of an intermediate laryngeal adduction level in the behavioral treatment of nodules.

Physiological Study of Resonant Voice

Laryngeal nodules are space-occupying lesions that develop at the midpoint of the membranous vocal folds in the superficial layer of the lamina propria, apparently as the result of intraglottal trauma. Voice therapy is usually recommended as part of the treatment program for nodules, whether laryngeal microsurgery is also performed or not. Probably the most common approach to voice therapy, at least in the United States, is the use of a quiet, breathy voice for phonation (see for example, Colton & Casper, 1990), or more generally the use of voicing patterns involving a widely abducted laryngeal posturing (for example, "yawn-sigh" phonation; Boone & McFarlane, 1993, 1994; see also Casper, Colton, Woo, & Brewer, 1989, 1990). This approach seems entirely sensible, particularly given recent evidence confirming a direct relation between adduction level and intraglottal contact stress[1] (Jiang & Titze, in press).

Recently, an alternative approach to therapy for nodules has been proposed, called "resonant voice" (Verdolini-Marston, Burke, Lessac, Glaze, & Caldwell, in press). This approach is based on work by a theatre voice and speech trainer, Arthur Lessac (see for example, Lessac, 1967). The distinctive aspects of resonant voice include vibratory sensations on or near the alveolar ridge or other facial plates during phonation, relatively strong voice output, and the sense of "easy" phonation. A preliminary efficacy study indicated that resonant voice training was as likely to produce a benefit from therapy for nodules as quiet, breathy voice training (Verdolini-Marston et al., in press). However, little is understood about the mechanisms underlying such a benefit from resonant voice therapy because limited studies have investigated the physiology of this voice type. In the single extant study of which we are aware (Peterson, Verdolini-Marston, Barkmeier, & Hoffman, in press), laryngeally normal subjects produced resonant voice with an intermediate level of laryngeal adduction (barely adducted vocal folds), between the levels for a voice type

called "pressed voice" (as if pushing), which was hyperadducted, and breathy voice, which was hypoadducted (see also Casper et al., 1989, 1990 for further descriptions of laryngeal configurations with breathy voice).

If resonant voice is as likely to produce a benefit from therapy for nodules as breathy voice (Verdolini-Marston et al., in press), the finding of relatively greater vocal fold adduction for resonant voice (Peterson et al., in press) is difficult to understand in light of the known direct relation between adduction and intraglottal contact stress (Jiang & Titze, in press), and the likely relation between impact force and pathogenesis (see also Jiang & Titze, in press). The answer might be that subjects with nodules produce resonant voice with more widely abducted vocal folds as compared with the normal subjects in the previous physiological study, thus explaining a therapeutic benefit of resonant voice for nodules. The purpose of the present study is to address this possibility by examining the adduction characteristics of resonant voice, in comparison with pressed voice, breathy voice, and "normal voice" (a control condition) in subjects with laryngeal nodules as well as in laryngeally healthy subjects. The voice types are specifically described under Procedures.

Methods

Subjects

Twelve adult subjects participated in the experiment as volunteers. Six subjects were recruited from graduate student singers in the Department of Music at our university.[2] These subjects reported no history of a voice disorder, their voices were normal to the examiner's ears on the day of the experiment (the examiner was a professor of clinical voice), and videoscopic examination conducted as part of the protocol confirmed normal laryngeal status for these subjects. The remaining six subjects were recruited from our clinical caseload, among patients with a diagnosis of laryngeal nodules, and who described themselves as amateur or professional singers. These subjects complained of hoarseness, they were chronically hoarse by our prior clinical evaluation, and videoscopic examination on the day of the experiment confirmed the continued presence of nodules. There were three males and three females in each experimental group. The average age for the laryngeally healthy subjects was 27 yr (range = 22-30 yr), and the average age for the subjects with nodules was 29 yr (range = 20-39 yr). Trained subjects were used in this study because they were more likely to produce valid samples of the voice types evaluated, as compared with untrained subjects.

Measures of laryngeal adduction

Two independent measures of laryngeal adduction were used in this experiment. One was an indirect measure, the electroglottographic closed quotient (CQ), obtained during sustained vowels in pressed, resonant, "normal," and breathy voice (see Scherer & Vail, 1988, for a discussion of the relation between the EGG CQ and direct measures of adduction). The other measure directly reflected adduction as an ordinal adduction rating from videoscopic views of the larynx during production of the same four voice types. Further details regarding the measures are provided under Procedures.

Equipment and software

A SynchroVoice Electroglottograph was used to collect electroglottographic (EGG) signals. Signals were monitored on-line during data collection using a Data 6000 oscilloscope, and a PC-108M SONY Digital Audio Tape (DAT) recorder was used to record signals, which were stored on DAT tape for later analysis.[3]

A Casio keyboard was used to provide pitches for critical experimental trials (described shortly), and a Wittner Taktall Piccolo metronome was used to indicate time elapsed for the trials.

After data collection was completed, EGG signals were digitized at 12,500 Hz using 16-bit analog-to-digital converter board in a Gateway 2000 386 computer. A locally written computer program calculated the closed quotient for each cycle of EGG signal (see Figure 1). The program found each cycle minimum and maximum which corresponded to maximum vocal fold opening and maximum vocal fold closure (Childers et al., 1986). Because researchers do not agree on an orientation of EGG signals, waveforms were arbitrarily represented in the program such that the vocal fold closure occurred at the maximum point of each cycle. Then the program calculated the closing event of the nth cycle as using a 35% criterion (Rothenberg & Mahshie, 1988):

$$s_c(n) = \min(n) + 0.35\,(\max(n) - \min(n))$$

where s_c is the vocal fold closure signal level. Opening is calculated as

$$s_o(n) = \min(n+1) + 0.35\,(\max(n) - \min(n+1))$$

where $s_o(n)$ is the vocal fold opening signal level. By using the extreme around each opening and closing, the program tracked fluctuations in the overall signal level. This procedure was necessary because the average (DC) voltage of EGG varies significantly as the subject makes small neck movements while performing the utterance. By using the above algorithm, the program output was unaffected by those fluctuations.

Using s_c and s_o, the program located the times of closing and opening for each cycle and calculated closed quotient as

$$CQ = \frac{\text{closed duration}}{\text{total duration}} = \frac{t_c(n) - t_o(n)}{t_c(n+1) - t_c(n)}$$

where $t_c(n)$ is the closing event and $t_o(n)$ is the opening event for the nth cycle and cycles are considered to start with vocal fold closure. All times were interpolated for greater accuracy.

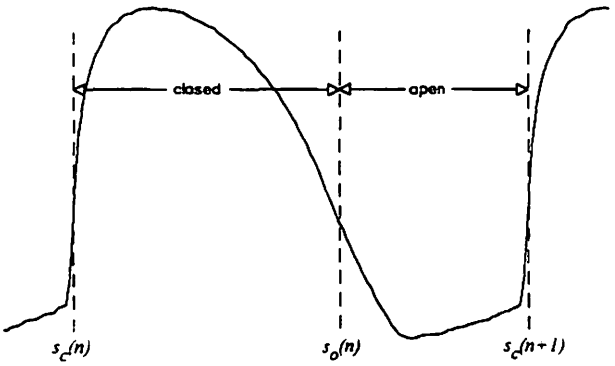


Figure 1. Typical EGG signal showing times used to calculate closed quotient (see Methods section for a description of symbols).

Finally, the program averaged the closed quotient for all of the EGG cycles in each file. in most cases, this was 300-600 cycles. A few utterances from some subjects were contaminated by noise and as few as 87 cycles were analyzed. The data from one subject was considerably more noisy than the data for other subjects, and only 30-100 cycles could be analyzed for each utterance from this subject. Videoscopic images of subjects' larynges were obtained using a 90 degree R. Wolf 4450.47 rigid telescopic endoscope, a Karl Storz 9000 Mini Solid State CCD Video Camera, and a Brüel and Kjaer Rhino-Larynx Stroboscope light source, Type 4914. Images were later played back to judges on a large-screen color television monitor, Sony Ideo Projection System KP-5000, using a Sony U-matic videocassette recorder.

Procedures

Subjects were seated throughout the experiment. Because all tokens for EGG signals would be produced at a predetermined conversational pitch individually determined for each subject, it was first necessary to determine that pitch. The subject counted out loud from one to five in a normal conversational voice, sustaining the vowel in the word "three." The fundamental frequency (f_o) for that vowel was determined from an oscilloscopic display, and used as the target f_o (converted to pitch) for all subsequent EGG trials for that subject.

Following the extraction of the conversational pitch, subjects received training and instructions for the experimental procedures. Specifically, subjects were instructed to produce the vowel /o/ (for training trials) at the identified conversational pitch (provided with a keyboard) for six seconds continuously following a "Ready...now...begin..." statement. A metronome set at 60 beats per minute was used to indicate the number of seconds elapsed. The subject practiced producing /o/ in this fashion using pressed, normal, resonant, and breathy voice, as demonstrated by the experimenter, with repeated trials until both the subject and the experimenter were satisfied that the target voice types were consistently produced (no more than a few trials were required for each voice type, for any subject).

Specific verbal descriptions of the voice types are somewhat challenging, and in fact the voice types were demonstrated (not described verbally) to subjects. For the purposes of this discussion, "pressed" voice corresponded to a high effort phonation mode, as if pushing with a relatively closed airway. "Normal" voice corresponded to a spontaneous conversational voicing mode. "Resonant" voice involved low effort voice production with anterior oral vibratory sensations. "Breathy" voice was also demonstrated as a low effort phonation mode, relatively quiet, with audible air escapage during phonation.

Following training (about 10 minutes), subjects produced the actual experimental tokens by sustaining the vowels /a/, /i/, and /u/ in blocks, counterbalanced across subjects, with three tokens each of pressed, normal, resonant, and breathy voice within each vowel, in randomized order. The order of vowel by voice-type trials was indicated on a sheet of paper individually supplied to each subject. Subjects were instructed to repeat any trial that they considered an invalid example of the target voice type. Including 3 vowels by 4 voice types by 3 repetitions, each subject produced a total of 36 valid tokens. A speech-language pathology doctoral student with an emphasis in clinical voice science, and who was uninformed about the order of the voice types, provided an independent estimate of the tokens' validity by indicating which voice type she perceived for each token. Subsequent analyses revealed that there was agreement between this listener's perceptions and the subject's intended voice type on 78% of trials. All tokens were retained for statistical analyses.

Although three vowels were produced, the focus for this study was the vowel /a/ because of its particular sensitivity to voicing mode, demonstrated in our previous study (Peterson et al., in press).

Following this first part of the experiment, which lasted no more than about 20 minutes including training and data collection, videoscopic images of subjects' larynges

were obtained during the production of the vowel /i/ produced with pressed, normal, resonant, and breathy voice, in randomized order across subjects. (The vowel /i/ was used because it facilitates full laryngeal views.) Videotaping continued until the examiner considered that an adequate view of the larynx had been obtained for each token. Again, a listener who was uninformed about the intended order of voice types (either a post-doctoral otolaryngology fellow or a speech pathology professor of voice disorders) indicated the voice type perceived for each trial. For these trials, there was 100% agreement between the intended and perceived voice types.

After the experiment was completed, four independent judges (two otolaryngologists and two speech pathologists, all with extensive experience in videoscopic imaging of the larynx) evaluated videotaped segments for each token produced under videoscopic examination. For each token, a 5-second segment was selected as the "best view of the larynx" by an experimenter who was uninformed about the order of the voice types and who was further naive to the experimental hypotheses. The selected segments were presented in random order to the judges, without audio, with tokens blocked by subject. After all tokens had been presented in this way once to the judges, the tokens were presented again without audio in a different (random) order. Judges were instructed to rate each segment viewed on an ordinal scale on which -5 = "extreme hypoadduction" and +5 = "extreme hyperadduction," using non-integer scoring values if desired. A rating of 0 was considered a "neutral" adduction rating on this scale. Judges were told to use their own internal criteria for the ratings. No other instructions were provided, nor were examples of different ratings provided in this study.

Results

EGG Closed Quotient

Average CQs for /a/ are shown in Table 1 as a function of voice type. This table shows that for both subject groups, the highest CQs were produced for pressed voice, intermediate CQs were produced for resonant (and normal) voice, and the lowest CQs were produced for breathy voice. A Multivariate Analysis of Variance (MANOVA) treating each vowel repetition as a separate outcome, and with group (2) as a between-subjects factor and voice type (4) as a within-subjects factor, was conducted to assess the interaction of group and voice type. The results of this analysis indicated that there was no interaction (Wilks' Lambda $F(9,92.63^4) = 0.66$, $p < .75$). A subsequent Repeated Measures Analysis of Variance (ANOVA) with group (2) and voice (4) as factors but without including the interaction factor in the equation (to avoid overfitting the model) indicated that the main effect of group was reliable[5] ($F(1,139) = 22.64$, $p < .0001$), as was the main effect of voice type ($F(3,139) = 10.43$, $p < .0001$). Thus, although CQs were generally lower for subjects with nodules (Table 1), a similar pattern of CQs across the voice types was obtained for both subject groups.[6] Post-hoc Tukey paired comparisons indicated reliable differences between the CQs for pressed voice and those for all other voice types, and further between CQs for breathy and resonant voice ($p < .05$). CQs for normal voice were not clearly distinguished from those for resonant and breathy voice in these analyses.

Table 1.
Average EGG Closed Quotients (CQs) for Pressed, Normal, Resonant, and Breathy Voice, for /a/. (Standard deviations indicated in parentheses.)

Group	Voice Type				
	Pressed	Normal	Resonant	Breathy	Average
Normal	0.58	0.53	0.55	0.45	0.53
	(0.06)	(0.07)	(0.07)	(0.10)	
Nodules	0.53	0.46	0.47	0.44	0.47
	(0.05)	(0.05)	(0.07)	(0.07)	
Average	0.55	0.50	0.51	0.44	

Laryngeal adduction ratings

Average adduction ratings are shown in Table 2. This table shows that, as for the EGG CQ, the highest average values were obtained for pressed voice (hyperadducted), intermediate values were obtained for resonant and normal voice, and the lowest values were obtained for breathy voice (hypoadducted). Of particular interest for this study, ratings for resonant voice indicated a barely adducted - or barely abducted laryngeal configuration (subjects with normal larynges and subjects with nodules, respectively).

Statistical analyses first involved a Cochran-Mantel-Haenszel Test of association to determine the level of agreement between judges for the ratings. The Row Means Scores Difference from this test indicated that there was good association between judges' scores ($p < .04$), indicating good agreement. Therefore, scores from all judges could be included together in subsequent analyses. A MANOVA used to assess the interaction between group (2, between-subjects factor) and voice type (4, within-subjects factor) indicated that there was no interaction ($F(3,184) = 0.28$, $p < .85$). A subsequent Repeated Measures ANOVA with group (2) and voice (4) further indicated that the group effect was unreliable ($F(1,187) = 0.63$, $p < .44$); however the main effect of voice type was reliable ($F(3,187) = 29.67$, $p < .0001$). Post-hoc paired comparisons with all subjects

pooled indicated that the adduction ratings for all voice types reliably differed from each other ($p < .05$), except for normal and resonant voice, which had equivalent ratings. Thus, a generally similar pattern of results was obtained for laryngeal adduction ratings as for the EGG CQ.

Table 2.
Average Laryngeal Adduction Ratings from Videoscopic Views, for Pressed, Normal, Resonant, and Breathy Voice, on /i/. (A value of +5 = extreme hyperadduction; -5 = extreme hypoadduction.) (Standard deviations indicated in parentheses.)

Group	Voice Type				
	Pressed	Normal	Resonant	Breathy	Average
Normal	1.23	-0.36	0.04	-1.33	-0.10
	(1.45)	(0.85)	(1.04)	(1.97)	
Nodules	1.31	-0.46	-0.40	-1.52	-0.27
	(1.43)	(1.64)	(1.36)	(1.36)	
Average	1.27	-0.41	-0.18	-1.43	

Discussion

Both subjects with nodules and subjects with healthy larynges produced "resonant voice" with an intermediate level of laryngeal adduction, between the level for pressed voice (hyperadducted) and breathy voice (hypoadducted). Specifically, EGG CQs for the vowel /a/ were greatest for pressed voice, intermediate for resonant voice, and smallest for breathy voice, and laryngeal adduction ratings from videoscopic views of the larynx (for /i/) followed the same pattern (see also Scherer & Vail, 1988, for a report of a similar covariance between the EGG CQ and direct measures of laryngeal adduction). Direct adduction ratings further indicated that resonant voice was produced with a barely abducted - or barely adducted laryngeal configuration, by both groups. These findings replicate those from an earlier study for vocally normal subjects (Peterson et al., in press). The findings, together with the earlier ones, also indicate that resonant voice is produced with a physiology similar to "flow" phonation described by Gauffin and Sundberg (1989).

Although the same pattern of results was obtained for both subject groups, EGG CQs were generally smaller for subjects with nodules. This result is anticipated, due to incomplete glottal closure and thus greater intraglottal resistance to electrical conductivity during phonation in these subjects. At a practical level, the point is that one cannot infer the voicing mode for an individual subject (or patient) from the average CQs in this study, due to variability in the absolute values of the CQs across subject types (normal and pathological). Stated differently, for the CQ, it is within subject variability that shows voice type. To identify the voicing mode with a given individual from the CQ, that individual's range of CQs should be evaluated. Confirmation of the specific laryngeal adduction level should be sought by direct examination, as in videoscopic viewing.

As noted, the EGG results were obtained for the vowel /a/. However, an inspection of the data for /i/ and /u/ for two randomly selected subjects (one with nodules and one without) revealed a similar pattern of results for these vowels. Thus, there was no indication that the findings were vowel-specific, based on these limited analyses.

It should also be pointed out that the results from this study were obtained with trained voice users. It is conceivable that untrained voice users might produce the voice types with somewhat different laryngeal configurations. However, based on our clinical experience we do not think that this is likely.

The pertinent clinical question regards the relevance of the intermediate laryngeal adduction level associated with resonant voice in the behavioral treatment of nodules. Why should this adduction level produce the same likelihood of a benefit from therapy as quiet, breathy voice (Verdolini-Marston et al., in press), which involves greater vocal fold abduction (Peterson et al., in press)? Based on a study with excised canine preparations, intraglottal contact stress appears to increase with increasing adduction (Jiang & Titze, in press), and thus breathy voice should produce a greater likelihood of a benefit than resonant voice, from a physiological standpoint. One possible explanation is that the critical level of vocal fold abduction required for lesion reversal is obtained with resonant voice, with further abduction producing little further likelihood of a benefit.

Another, more complex argument involves the relatively stronger voice output with resonant as compared with breathy voice. In our previous physiological study (Peterson et al., in press), resonant voice was characterized by high glottal airflow deceleration rates (maximum flow declination rates), large AC flows, and small minimum flows (Peterson et al., in press). These conditions, obtained with barely adducted or barely abducted vocal folds (see for example Gauffin & Sundberg, 1989), produce a relatively high intensity output spectrum (Titze, in press). Thus, resonant voice can generally be heard and understood more readily in a range of communication environments, as compared with quiet, breathy voice. There might be a trade-off between tissue protection (best for breathy voice, but adequate for resonant voice) and voice intensity, which in turn governs extra-clinical utilization of the voice technique outside the clinic. Patients with nodules may actually use resonant voice outside the clinic more readily than

quiet, breathy voice. Thus, the small losses in tissue protection with resonant as compared with breathy voice may be offset by more consistent application of this voicing type. The net result would be a similar likelihood of a cumulative therapeutic benefit for the two voice types: resonant voice produces a relatively smaller benefit for tissue protection, but is used more consistently because of its good intensity, increasing its cumulative therapeutic value; breathy voice produces a greater benefit for tissue protection, but is used less because of its low intensity, offsetting its potentially greater therapeutic value from a physiological standpoint.

Incidentally, the physiological characteristics of resonant voice described here and elsewhere (Peterson et al., in press) undoubtedly underlie not only the strong voice output typically perceived with resonant voice, but also the oral vibratory sensations. The glottal output spectrum is predicted to be most intense with barely abducted or barely adducted vocal folds (Titze, in press); this type spectrum would supply the intense oral air pressure variations resulting in the perceptible vibratory sensations that subjects report with successful resonant voice training.

In summary, resonant voice is produced with barely adducted - or barely abducted vocal folds, between the adduction level for pressed voice (hyperadducted) and breathy voice (hypoadducted). We speculate that the putative benefit of resonant voice in therapy for nodules is related to a trade-off between voice output intensity (relatively great) and intraglottal impact force (acceptably small, although not as small as for breathy voice). This speculation should be considered separately in another paper.

References

Boone, D.R., & McFarlane, S.C. (1994). The voice and voice therapy (5th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Boone, D.R., & McFarlane, S.C. (1993). A critical view of the yawn-sigh as a voice therapy technique. Journal of Voice, 7, 75-80.

Casper, J.K., Colton, R.H., Woo, P., & Brewer, D. (1989, 1990). Investigation of selected voice therapy techniques. An amalgamation of two papers presented at the Eighteenth and Nineteenth Symposia: Care of the Professional Voice, Philadelphia, PA.

Childers, D.G., Hicks, D.M., Moore, G.P., & Alsaka, Y.A. (1986). A model for vocal fold vibratory motion, contact area, and the electroglottogram. Journal of the Acoustical Society of America, 80, 1309-1320.

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

Editorial (1990). Journal of Experimental Psychology: General, 119, 3-4.

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

Jiang, J.J., & Titze, I.R. (in press). Measurement of vocal fold intraglottal pressure and impact stress. Journal of Voice.

Lessac, A. (1967). The use and training of the human voice: A practical approach to speech and voice dynamics. Mountain View, California: Mayfield Publishing.

Peterson, K.L., Verdolini-Marston, K., Barkmeier, J.M., & Hoffman, H.T. (in press). Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. Annals of Otology, Rhinology, and Laryngology.

Rothenberg, M., & Mahshie, J.J. (1988). Monitoring vocal fold abduction through vocal fold contact area. Journal of Speech and Hearing Research, 31, 338-351.

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

Titze, I.R. (in press). Control of vocal intensity and efficiency. In I.R.

Titze, Principles of voice production. Englewood Cliffs, N.J.: Prentice Hall.

Verdolini, K., Berry, D.A., & Titze, I.R. (1994). Laryngeal efficiency from a therapeutic perspective: A computational analysis. Manuscript submitted for publication.

Verdolini-Marston, K., Burke, M.K., Lessac, A., Glaze, L., & Caldwell, E. (in press). A preliminary study on two methods of treatment for laryngeal nodules. Journal of Voice.

Acknowledgements

An earlier version of this work was presented at the Twenty-Second Annual Symposium: Care of the Professional Voice, Philadelphia, June 7-12, 1993. The study was

supported by Grant No. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. The authors acknowledge Dr. Linnea Peterson for her assistance with the experiment, and Dr. Kenneth Moll for his comments on an earlier version of the paper.

Endnotes

[1] Contact stress is impact force per unit area of tissue contact.

[2] One of the singers in this group was actually a former graduate student in the same department, currently employed as a public school music teacher.

[3] Audio and aerodynamic signals were also collected, but are not further described in this report.

[4] This test uses adjusted degrees of freedom in order to evaluate the results against the F-distribution.

[5] In this article, we use the term "statistically reliable" in place of the usual term "statistically significant" because of a point made in an editorial in Journal of Experimental Psychology, March 1990. Reliability refers to the likelihood of replicating the results relative to the null hypothesis, given the amount of variability present in the data, whereas significance implies a conceptual importance in the results that may or may not be present with reliable data.

[6] Generally lower CQs are anticipated in subjects with nodules because of relatively poorer translaryngeal electrical conductivity in these subjects due to larger persisting intraglottal air spaces (around the lesions) during phonation.

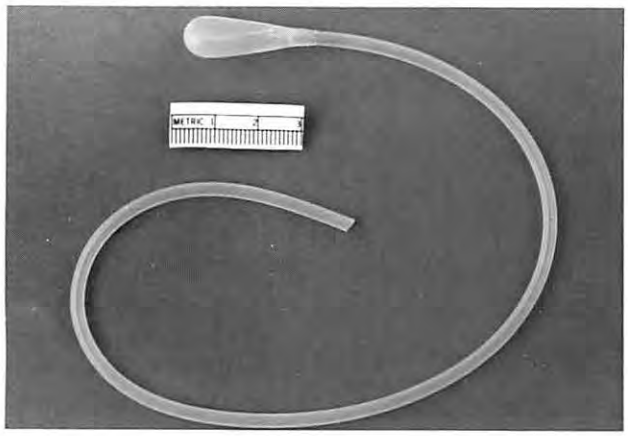


Photo 1. Velopharyngeal closure force sensing bulb

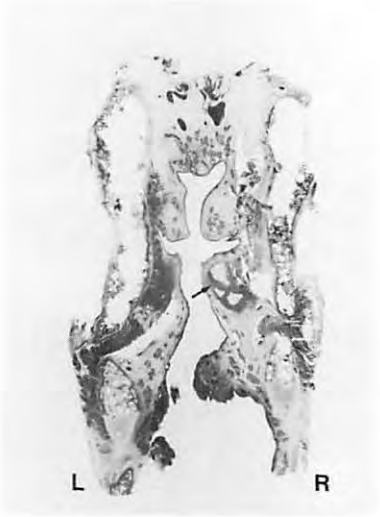


Photo 2. Whole larynx preparation showing three discrete sites of transplanted autologous fat in the vocalis muscle of the right vocal fold (arrow).

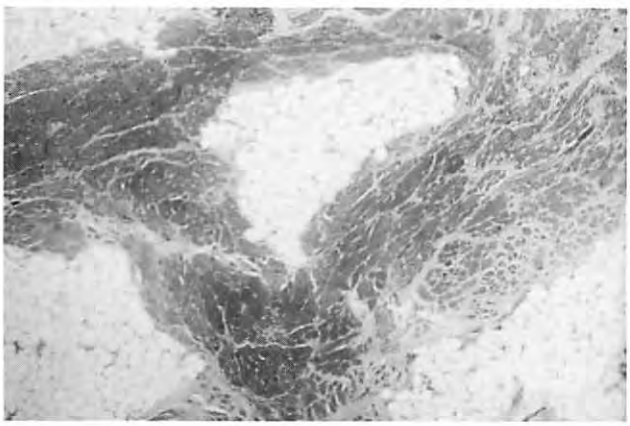


Photo 3. Magnified view of the right vocal fold showing a collection of transplanted adipocytes.

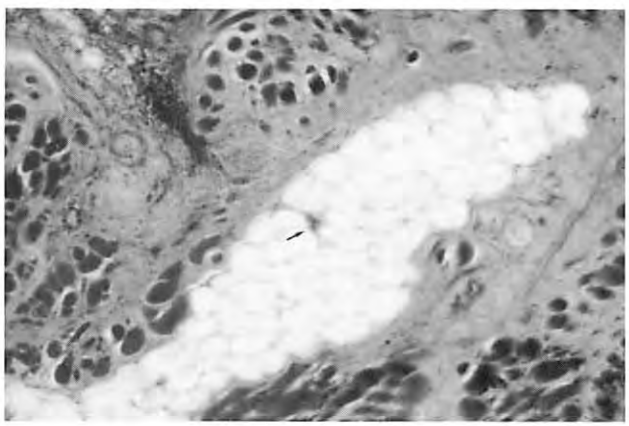


Photo 4. High power view of fat micrograft with blood filled capillary (arrow). Note there is minimal surrounding inflammation and no evidence of lipid degeneration.

Part II

Tutorial reports and summaries of Training, Continuing Education and Dissemination of Information

NCVS Status and Progress Report - 6
May, 1994, 155-163

Principles of Skill Acquisition Applied to Voice Training

Katherine Verdolini, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

As a young vocalist and voice trainer, I kept thinking that if I could only identify the "right technique," if I could only figure out with certainty which body part should go where and when during voice production, then I could sing well and I could also be a good voice trainer. Even beyond the rather frightful omission of emotions and art in this thinking, there was a serious fallacy: Knowing that you should hike up your larynx to your eyeballs when you sing (if that were the conclusion) is not equivalent to knowing how to do it or how to teach it. In fact, even as my knowledge of voice physiology increased, my ability to "do it" and to "teach it" was not necessarily improved proportionately. A gulf remained between my explicit, analytical knowledge about voice, and voice production itself. At first, the gap annoyed and perturbed me. Then it became interesting. What was the nature of "learning how"? How was it different from "learning that," or "book learning"? These and other questions prompted me to pursue studies in skill acquisition, in a department of psychology.

As luck would have it, during the same historical period in which I initiated those studies, a theoretical approach to learning and memory was developed that focused on the very distinction that intrigued me: the distinction between "knowing that" and "knowing how." Based on observations of amnesic subjects and also normal learners, it was proposed that there are two different memory systems or information processing modes[1] with distinct neuroanatomical and cognitive substrates: one system or mode that governs "knowing that," and one that governs "knowing how."

Little could have been more interesting. This essay is an attempt to summarize in a simple way what I consider to be among the most important findings on this topic and on skill acquisition in general, as they may apply to voice and voice training.

Some leaps will be necessary. Most of the experiments have been done using verbal or hand-eye coordination tasks. None have been carried out using voice tasks. Also, most of the experiments investigating the cognitive characteristics of the memory type we are interested in have been carried out not looking at skill acquisition per se, but rather at a parallel phenomenon called "priming," which will be discussed in some detail shortly. Despite these limitations, I think that it is reasonable to make some generalizations from these studies to skill acquisition in voice training, awaiting the proper studies.

Theoretical Background

Among the earliest contemporary observations leading to a distinction between "knowing that" and "knowing how" were those reported by Dr. Brenda Milner, a Canadian neurologist.[2] A patient of hers, a now-famous "H.M.," had undergone a bilateral resection of his temporal lobes (parts of the brain surface) as treatment for a debilitating seizure disorder. After his surgery, H.M. could not acquire new, conscious memories of post-surgical events.

[1] Throughout this essay, I repeatedly refer to "memory system" and "processing mode" side by side. The reason is that a debate has been going on for a few decades now about whether to consider memory structurally, as a "system," or whether to consider memory dynamically, in terms of its processing characteristics. My use of the terms "memory system" and "memory processing mode" in the same docment reflects my decision to acknowledge both theoretical views equally, essentially because I think both have merit, and I think that the distinction is an artificial one.

[2] Antecedent observations were made as early as Plato (see for example, Stumpf, 1975) and Descartes (see Haldane & Ross, 1967), and were later made by psychiatrists (Freud & Breuer, 1966; Janet, 1893, 1904), psychologists (Ebbinghaus, 1885; Hull, 1933; Thorndike & Rock, 1934), neurologists (Clarapede, 1911/1951; Korskoff, 1899), and others (see Schacter, 1987 for a review). However, Milner's reports more directly lead to the current formulations discussed in this essay.

He did not remember from one moment to the next that he had eaten or that he had met someone. Somewhat surprisingly, he could remember how to do new things. For example, he improved with practice on a task that required tracking a rotating target with a wand (Milner, 1962). So, he was indeed able to form new memories: those governing improved perceptual-motor performance with practice. He just did not know that he had acquired the memories. (In fact, a "memory" is any record of the past, regardless of whether the memory is "remembered" or experienced in consciousness or not.)

The theoretically important implication that eventually developed from this report and similar ones was that there must be a neuroanatomical system or processing mode that governs practice effects without conscious remembering, and that is spared in amnesia, as well as a system that governs memories that we experience and can talk about. The brain parts resected in H.M.'s surgery, and specifically the hippocampus and the amygdala in the temporal cortex, were clearly important for the latter but not the former type of learning and memory.

What are the characteristics of the memory system (or processing mode) preserved in amnesia, that appears to govern perceptual-motor skill acquisition? Information about this system and how it works might elucidate some useful principles for voice training. A look at some recent findings will shed some light on this question.

Characteristics of the memory system and operations preserved in amnesia, which appear to govern perceptual-motor skill acquisition ("implicit memory")

The memory system (or processing mode) preserved in amnesia governs not only skill acquisition for perceptual-motor tasks, but also other functions, including "priming" (see for example, Squire, 1986). Priming is facilitated performance on previously encountered stimuli, as compared with performance on new stimuli of the same class. For example, if a subject first studies a series of words and then is shown studied and non-studied words extremely rapidly (say 35 ms each), priming is shown by better identification of the studied versus the non-studied words, regardless of whether the subject remembers the studied words or not.

So, to reiterate, both skill acquisition and priming remain intact in amnesia; as such, they can be considered together as part of a set of memory functions called "implicit memory."[3] Specifically, implicit memory is reflected by any performance benefit from prior practice or exposure to stimuli, without learners explicitly remembering those stimuli (Graf & Schacter, 1985; see also Schacter, 1987, and Roediger, 1990, for excellent reviews). Not only amnesic learners but also neurologically intact subjects show implicit memory, as described next. In the discussion that follows, we will assume that the characteristics of implicit memory - most of which are known from studies on verbal priming - are consistent across all types of implicit memory, including perceptual-motor performance without awareness. This assumption may ultimately be shown to be incorrect, but I am willing to make it in the meantime.

Implicit memory is memory without awareness. Implicit memory, as it has been defined in theoretical research, is memory without awareness. That is, implicit memory involves a memory type without conscious remembering of events leading to memory development or conscious remembering of what has been learned. At best, conscious awareness is irrelevant for implicit memory. At worst, it may interfere with it (Verdolini-Marston, 1991).

Given this characteristic of implicit memory, it would seem that this memory type could be shown only in amnesic subjects, who don't remember much of anything. On the contrary. Implicit memory can also be shown in normal learners. Graf and his colleagues (Graf, Mandler, & Haden, 1982) reported a classic study in this regard. In that study, the investigators essentially simulated amnesia - and showed implicit memory - in neurologically intact learners.

Healthy adults were first asked to study a list of words (for example, they might have been asked to study "defend, repair, engage" etc.). Subjects were later asked to complete word stems (for example, "def. . .") with the first word that came to mind. Each stem could be completed with several different words ("defend, define, default..."). However, subjects tended to complete the stems with previously studied words, which they did not necessarily remember, at a greater than chance level (priming). Thus, the effect of prior exposure was seen on current performance, without the apparent assistance of conscious remembering. This finding constituted evidence of implicit memory or memory without awareness in normal learners.

Implicit memory appears fundamentally governed by perceptual processes. This point is the most important one to focus on regarding implicit memory, to my thinking. Implicit memory, at least as reflected by priming, fundamentally involves the mental processing of perceptual - or sensory information: sight, sound, smell, taste, and touch. As important, implicit memory generally does not appear to involve "associative processing," or the relating of perceptual information to other mental contents and operations that make it "meaningful" or symbolic.

The study cited above by Graf and colleagues is an example of the many, many experiments that have pointed to this conclusion (Graf et al., 1982). Recall that in the initial phase of that investigation, subjects studied a list of words. In fact, during the study phase of the experiment,

[3] Other terms have also been used to refer to the memory system or processing mode preserved in amnesia, for example, "procedural memory" (Squire, 1986).

subjects studied the words under different conditions. In one group, subjects studied the words by processing their meanings (subjects rated how much they "liked" each word, related to its meaning). In another group, subjects studied the words by processing their perceptual or surface characteristics (subjects indicated whether each word had any letters in common with the preceeding word). During the subsequent test phase of the experiment, it turned out that the likelihood of remembering or recalling the studied words was increased if the subject had processed the words' meanings during study. However, the likelihood of spontaneously completing a word stem with a previously studied word (or implicit memory) was unaffected by the earlier study task: words that had been examined for their meaning and words that had been examined for their surface or perceptual characteristics were equally likely to be produced as word stem completions, even when subjects did not recall which words they had studied. The conclusion was that implicit memory (reflected by word stem completion in this case) can develop any time one is exposed to a stimulus and processes its perceptual characteristics. The processing of stimuli's meanings, or associative processing, appears irrelevant for the development of this memory type.

Numerous other studies have pointed to similar conclusions, although qualifiers do need to be added. One qualifier is the topic of the next subsection.

Implicit memory requires attentional processing. Some studies have pointed to the likelihood that when new stimuli are encountered, for effective perceptual processing underlying implicit memory to occur, full attention must be directed to the stimulus. This principle was perhaps most clearly demonstrated in a series of studies by myself, for my dissertation (now in press in condensed form; Verdolini-Marston and Balota, 1994).

My studies involved a perceptual-motor task, "pursuit rotor." In this task, the object is to track a rotating stimulus with a wand. When there is contact, a counter is driven. If the subjects involved were amnesic subjects who did not remember that they had ever done the task before, improvements with practice would constitute evidence of implicit memory, i.e. memory without awareness. However, we used neurologically normal subjects in our studies, who did remember that they had done the task before from session to session. Thus, to investigate implicit memory, we had to come up with some measure of perceptual-motor learning without awareness in these normal learners. We essentially used a priming paradigm. Subjects first practiced on several different stimuli. Later, they returned for a test on the old stimuli, as well as on new stimuli of the same difficulty. Implicit memory would be shown by a performance benefit for the old as compared with the new stimuli, assuming that subjects did not remember which stimuli were old and which were new. In our first of several experiments, we demonstrated such priming for the pursuit rotor task.

The main objective of theoretical interest in our studies was then to investigate the type of mental processes that regulated priming for this perceptual-motor task. We thought that priming might be governed by perceptual processes, as had been found for verbal priming. Our experimental strategy was to give subjects in different groups different instructions about what to do mentally during the initial practice phase. In some groups, subjects were instructed to use images to assist their performance, such as stirring in a bowl or imagining a locomotive wheel as it turned. (According to earlier pilot work, subjects thought that these were good images to help performance.) In other groups, subjects were instructed to pay attention to the task, and to concentrate on it. Finally, in other groups, subjects were not given any instructions about what to do mentally during practice.

During the later test phase, it turned out that subjects showed priming (a performance benefit for old stimuli, without remembering which stimuli were old, or implicit memory) only if they had not been given any instructions about what to do mentally during practice. Subjects who had received instructions, whether they were imagery or concentration instructions, failed to develop priming or implicit memory.

We were a bit perplexed by the findings, but our explanation was ultimately as follows: perceptual processes underlie implicit memory in the perceptual-motor as in the verbal domain. However, for tasks that are novel to subjects - such as the pursuit rotor task - in order for perceptual processing to occur, subjects must devote their full attention to the task. (Perceptual processing might occur more automatically for familiar stimuli, such as words.) When instructions of any type were imposed, subjects paid attention to the instructions, and not to the perceptual array associated with the task. Thus, priming (implicit memory) failed to develop.

What we have said here may seem to be a paradox: implicit memory involves learning without awareness, yet attention is required. How can we be unaware of various aspects of a task, and yet attend to it? Thinking carefully, awareness and attention are actually different phenomena. Awareness involves knowing that one has done something specific (performed on a given stimulus, or perhaps moved the arm this way or that way). Attention refers to alertness, or the reception of information without any necessary conscious organizing of it in time or space. In a word, attention refers to "being in the moment," "being here, now," without judgment or comment or conclusions (awareness).

Another important finding in our studies, that may be of greater interest for this readership, was that not only implicit memory (priming, in this case) but also learning for the pursuit rotor task in general was poorest when subjects were given mental strategies to use during practice. Relatively poorer overall performance was seen for both imagery and concentration groups, as compared with the no-instruction group. Equally as interesting, subjects who used imagery and concentration strategies thought that those strategies were very helpful for learning and performing the pursuit rotor task. The implication is that imagery instructions or even concentration instructions may not be as useful as we sometimes think they are, regardless of students' informal impressions about their benefits.

Implicit memory depends on repetition. One of the strongest factors affecting implicit memory is repetition. The size of performance benefits for previously encountered or practiced stimuli increases as the number of exposures or repetitions increases. This principle seems obvious enough, so I won't drag you through the evidence (the interested reader is referred to Jacoby and Dallas, 1981). However, as noted in the preceding sections, for repetitions to be effective in producing implicit memory, apparently they must involve the processing of perceptual, or sensory information. The repetitions must also occur in the same modality as will be required for later tests of implicit memory. This issue is discussed next.

Implicit memory is modality and context-specific. Implicit memory appears to be a quite literal and inflexible memory type in that it is modality- and context-specific. If you change the performance modality between study (training) and test (for example from the auditory to the visual modality), or even if you change the environment in which study (or training) and test occur, (from a swimming pool gallery to a games arcade), implicit memory fails to develop normally.

An example of the modality-dependence of implicit memory was reported by Jacoby and Dallas (1981). In one of their experiments, words were presented to subjects either visually or auditorily during an initial study phase. In a later test phase, subjects were required to identify words that were presented visually, extremely rapidly (for 35 ms). Some of the test words were from the earlier study list and some were not. A better ability to identify studied as compared with non-studied words would constitute evidence of priming or implicit memory. In some cases the modality was consistent between study and test modality (visual-visual), and in some cases modality was inconsistent between study and test (auditory-visual). At test, on average subjects showed better identification of studied as compared with non-studied words (implicit memory), provided that the study and test modalities were consistent (visual-visual). When study and test modalities were inconsistent (auditory-visual), implicit memory was not shown. The implication is that implicit memory depends on modality consistency between training and later performance.

Implicit memory also fails to develop fully when the context (or environment) changes from study to test. An example of this principle was reported by Graf (1988). In his study, subjects studied the usual word list during an initial experimental phase, either in a swimming pool area or in a games arcade. Later, subjects performed on implicit memory tests (word stem completion and another conceptually similar test, a "category production" test), either in the same setting in which they had studied or in the other setting. Priming, or implicit memory performance was greater when the study and test environments were consistent. Thus, implicit memory depended on environmental consistency.

Other Principles of Skill Acquisition

There are numerous other important principles of skill acquisition that have been investigated within other frameworks besides implicit memory. Schmidt (1987) provides a comprehensive review of this literature and pertinent issues. For our purposes, three of what I consider the most important principles are discussed next.

Skill acquisition requires information about performance during training (knowledge of results). For skills to improve with practice, learners must have information ("knowledge of results," or KR) about how well they have done relative to the target performance. An example of this quite intuitive principle was illustrated in a study by Bilodeau and colleagues (Bilodeau, Bilodeau, & Schumsky (1959). In their study, subjects practiced what is called a "linear positioning task," which requires subjects to learn to displace an object to a criterion position. Some subjects received KR after each of 20 practice trials, some subjects received no KR for 20 trials, and other subjects received an intermediate amount of KR. The result was that no learning was observed in subjects who had not received information about their performance, or KR. Therefore, KR was apparently required for learning.

The interesting thing is that from this and other studies, although KR appears required for learning, more is not necessarily better. In another study by Johnson, Wicks, and Ben-Sira (1981), using a similar linear positioning task as in the study by Bilodeau and colleagues (Bilodeau et al, 1959), some subjects received KR after each of 10 trials. Some subjects received KR on 25% of 40 trials, and some subjects received KR on 10% of 100 trials. Thus, in each subject group, subjects received KR on 10 trials. However, there were differences in the number of intervening trials between KR provision. During the training phase of the experiment, subjects who received relatively infrequent

KR appeared to do somewhat worse than other subjects. However, when subjects were tested later, during a delayed test, subjects who had received the infrequent KR actually did the best.

The implication is that although frequent KR may be useful for immediate performance during training, less frequent KR may enhance actual learning, shown by and better performance at a later time. It may be that when learners are not pounded with KR during training, they can process information relevant for learning and "make it their own," with performance benefits at later follow-up.

To be able to perform acquired skills along with other tasks, consistent responding is required during training. One of the goals of training in many domains, including voice, is to skillfully produce the target behavior while at the same time performing other critical tasks, for example talking, moving, or dancing. Some work suggests that the ability to run a new, acquired behavior off skillfully along with other behaviors requires that during training, the behavior be consistently produced in the target fashion. This principle was evident from a series of studies by Schneider and Fisk (1982).

In their studies, subjects were first shown a series of cards with letters on them, in rapid succession, and were required to indicate in which position on a card a target letter had appeared. In some cases, a given letter (for example, "M") was always a "target" when it appeared. This was called a "consistent mapping" condition. In other cases, a given letter ("M") might be a target to watch for on one trial, but not a target on the second trial (i.e. "X" might have been the target, and "M" a distractor). This training condition was called "variable mapping". Later, the task was made more complicated by asking subjects to look for new letters as well as previous target letters. The result was that when this additional task was added, subjects were able to retain their previous skill level in position identification for the original target letters only if previous training had involved consistent identification of those letters, or "consistent mapping." If previous training was inconsistent, or variable (requiring the identification of different letters on different trials), earlier skill levels were not retained in the face of the new task.

An excellent way to summarize the conclusions from this study is found in a much earlier quote by a renowned American psychologist, William James:

"Never suffer an exception to occur till the new habit is securely rooted in your life. Each lapse is like the letting fall of a ball of string which one is carefully winding up; a single slip undoes more than a great many turns will wind again. Continuity of training is the great means of making the nervous system act infallibly right." (James, 1890).

Generalization to untrained variants of tasks is enhanced by practice under variable conditions. Practice under variable conditions appears to enhance the generalization of specific trained behaviors to other behaviors in the same class. An example demonstrates this principle. McCracken and Stelmach (1977) had subjects practice knocking over barriers with their hands within a given time-period on numerous trials. Some subjects received practice with a constant hand-to-barrier distance, and other subjects received practice with different hand-to-barrier distances on different trials. During the training phase, subjects with the constant distance did better. However, when all subjects were all tested later with a distance that was new for all of them, those who had previously received variable practice did better.

The point from this and other studies is that to maximize the generalization of a trained behavior to new, untrained situations, practice should involve varied practice conditions, perhaps with varied materials.

Applications to Voice and Voice Training

Having gone through some of the details, what do we do with this information in voice and voice training? In this next section, we take a second look at each of the principles discussed above, one by one, and try to make some links to voice and voice training. Then in the final section, we attempt to tie all the principles together into a cohesive whole.

You may discover that you already employ several of the suggested strategies. In this sense, the following section might provide support and encouragement for what you are already doing.

Specifics

Implicit memory is memory without awareness. Based on studies with amnesic subjects, implicit memory is a memory system or processing mode that appears to govern skill acquisition and does not involve conscious awareness of what has been learned. In fact, awareness may not only be indifferent for implicit memory. Explicit knowledge about learning contents may actually interfere with the development of implicit memory, in some cases (see for example, Verdolini-Marston, 1991).

The implications might be quite surprizing to some trainers: although knowledge about voice science may be extremely helpful to you as trainers, that knowledge is not necessarily helpful to your voice students. In particular, a mechanistic awareness about voice production in training, such as "expand your ribs here and drop your jaw there" may be fruitless or even counterproductive. I realize that this is a strong statement, and may not extend to all situations. However, I encourage you to explore it.

Implicit memory appears fundamentally governed by perceptual processes. From this principle, we understand that one of our main tasks as trainers is to promote our students' processing of perceptual - or sensory information during training. We should show students what good alignment for voice production looks like. We should let them feel a deep breath. We should let them hear the sound of a focused voice.

The emphasis on sensory information is not new to theatre. The Alexander Technique is profoundly based in this principle (see for example, Jones, 1976). Lessac also focuses on sensory information, going so far as to include "tasting" and "smelling" various aspects of voice production (Lessac, 1967). There are many, many other examples of theatre trainers for whom sensory information is a critical part of the training program.

A discomforting corollary is that metaphoric images ("imagine rowing a boat as you breathe"), which are commonly used in voice training and which involve meaningful or associational processes (as opposed to perceptual processes), may work against voice development in some cases.

Implicit memory requires attentional processing, where novel stimuli are concerned. The implication from this principle is that any operations that divert away attention from perceptual - or sensory information related to voice may interfere with memory development. Attention can be diverted by too many instructions about mental strategies, so that the learner attends to the strategies and not to sensory information. Attention from perceptual information can also be diverted by emotional responses. ("That was so terrible!!!" - foot stamping optional). Attention may be affected by low motivational levels. Most of you are empirically familiar with these and other apparent perils to attention.

As already mentioned, I think that the best way to describe what attention to perceptual information "looks like" is "being in the moment," "being here, now." As theatre trainers, you are all assuredly very familiar with this concept.

Implicit memory depends on repetition. This principle is quite straightforward. PRACTICE! Or, as Lessac (1967) suggests, "Don't practice, do it!" (repeatedly).

Implicit memory is modality and context-specific. Here we have an interesting point. There has been much talk about different "modality" strengths across learners, in theatre training. According to this view, some people would be "visual" learners. Others would be "auditory" learners. Still others would be "digital" or verbal learners. The suggestion is to modify your teaching input to match the student's modality strength (CITE). The finding that implicit memory is modality-specific suggests a different tack. Within this framework, the best modality for training is the modality required for task execution. For voice production as for any motor task, ultimately, perceptual (not verbal) information guides central nervous system output commands. Therefore, training should occur in the perceptual realm, including auditory, kinesthetic, and visual modalities, depending on the specific task. If we train a student in the "digital" or verbal mode, at best a translation of that information to a perceptual code will be required before voice can be produced; the transformation takes time and precious information may be lost or distorted in the process. The verbal mode is right if you want the student to learn to talk about voice production. But the verbal instructional mode is wrong for voice training, if you want the student to actually learn to produce voice better, regardless of the student's relative "modality" strengths.

Also, training benefits may be greatest when training occurs in the same or similar environments as will be encountered for performance. We all think about conducting classes on a stage, but I for one do not do it nearly enough.

Skill acquisition requires information about performance (knowledge of results). Students need information about how they are doing relative to what you want them to do. However, intensive feedback is not necessarily good. Witholding information about performance for several trials may enhance the student's own processing of information and later performance, or learning. This conclusion may be especially valid later (as opposed to earlier) in learning. So, not giving any feedback for several trials may help the student generalize from what you have worked on for one vowel to another vowel, from one monologue to another, and so forth. The translation of this principle to simpler language is, "Give your student space!"

To be able to perform acquired skills along with other tasks, consistent responding is required during training. I have heard theatre trainers talk about this principle: if the student is to use a certain voice production mode on stage while acting, dancing, emoting, etc., that same mode or a similar mode must be consistently used, even off stage. What it boils down to is that we can't summon a physiological operation in voice on stage, adding other task requirements, if we don't consistently use that mode in all or most situations when it is appropriate.

Generalization to untrained variants of tasks is enhanced by variable practice. The use of varied training materials will probably enhance the student's ability to generalize a target voicing mode from trained to untrained materials. For example, if you working on a focused voice, you should train it on many different vowels, phrases, monologues, etc. (so who doesn't?). That way, when the student encounters new materials, generalization of focused voice should transfer to those materials. Similarly,

you should include different performance materials in training: different monologues, dialogues, comic and tragic material, etc.

Tying it Into a Cohesive Whole

For a few years, I found myself incorporating many of the principles desribed here in a helter-skelter fashion in voice training. My mind lept from one to another to another of the principles, like the proverbial monkey in a tree. I was run ragged by the job, and my students and patients were probably stunned. What was the classic "banana" that would quiet the mental monkey? If we accept the principles discussed here, how can we tie them all together into a cohesive whole? Let me share what I have come up with so far.

The kingpin of what I call my "skill acquisition package" is the notion that skill acquisition fundamentally involves attention to perceptual information, in numerous repetitions, with varied tasks. If the student effectively processes perceptual information, several of the principles that we discussed will be satisfied as a by-product. We avoid a mechanistic "awareness" of voice production. We stay in the relevant modality, which is perceptual, not verbal. Numerous repetitions occur, with consistent performance required across different tasks. An additional notion is that I attempt to supply "some" but not "too much" feedback (knowledge of results).

So, fine! What does all this look like? I have a hierarchy of five steps that I think represent a start. Let's take breathing as an example. Say that we want the student to expand the abdomen on inspiration, and press it in during expiration (and voicing). How can we get the behavior trained, following the principles outlined? The steps that I use are:

Step 1. Direct the student's attention to the body in general, and ask him to notice any sensations. ("Scan your body with your mind's eye. Do you notice any sensations?") Do not ask the student to describe the sensations verbally, just to attend to them. Often, the desired physiological behavior will appear with this simple step, without further work. If not, proceed to Step 2.

Step 2. Direct the student's attention to the specific body part of interest, in this case the abdomen. ("Focus on your abdomen as you breathe.") Again, the desired behavior may appear with this step. If not, proceed to either Step 3 or Step 4 (they are interchangeable in order).

Step 3. Model the behavior for the student. For example, without saying anything, place the student's hand on your abdomen as you breathe in and out.

Step 4. Manipulate the student's body so that the target behavior is likely to occur. In this case, you might stabilize the shoulders and chest so that only the abdomen is available for movement during breathing.

If the student still does not produce the behavior, as an absolute last resort, proceed to Step 5, which violates the principle of perceptual training:

Step 5. Tell the student what to do.

This sequence, which I think I came up with on my own, has a similar flavor to what Timothy Gallwey describes in his book on tennis learning, Inner Game of Tennis (1974). In that book, he also emphasizes the processing of perceptual, or sensory information. The sequence that I outlined can be used with just about any behavior that you want to train. If a student is "squeezing" in the throat while talking, direct the student's attention to the body in general, and ask him to notice any tension. If the tension does not dispel, direct his attention to his throat. If that doesn't work, model tight versus free voice, and have him mimic you in both. Or, manipulate the neck or other body parts in a way that is likely to free the throat. Finally, if you must, tell him to let go of his throat, for heaven's sake.

Once the student has produced the behavior for simple tasks, you proceed (of course) with many, many repetitions, in different and increasingly complex contexts (for example, simple vowels, words, phrases, short discourse, conversation, monologues). Gradually fade models so that the student depends less and less on your input.

With this type of sequence as a part of your training technique, you will discover that you are talking and explaining less and less, and "doing" more and more. Bonnie Raphael recently referred to this principle at a recent VASTA workshop (1993): "Every year I promise myself, 'Half as many words...'"I agree with her, and consider minimized yammering an indication of principled - and I think effective training.

Summary

In this essay, we have discussed several principles of skill acquisition, mostly derived from a theoretical construct called "implicit memory." This type of memory, to which skill acquisition appears tightly linked, is a memory without awareness, it involves attention to perceptual or sensory information, and it develops with numerous and consistent, modality-and context-specific repetitions of target behaviors. Its optimal training for voice performance likely maximizes attention to sensory information, minimizes verbal analytic explanations, and includes training with a variety of materials and in physical environments that will be relevant for performance. The approach described here is, in sum, "something to experience."

References

Bilodeau, E.A., Bilodeau, I.M., & Schumsky, D.A. (1959). Some effects of introducing and withdrawing knowledge of

results early and later in practice. Journal of Experimental Psychology, 58, 142-144.

Clarapède, E. (1951). Recognition and 'me'ness.' In D. Rapaort (Ed.), Organization and pathology of thought (pp. 58-75). New York: Columbia University Press. (Reprinted from Archives de Psychologie, 1911, 11, 79-90).

Cohen, N.J., & Squire, L.R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of "knowing how" and "knowing that." Science, 210, 207-210.

Ebbinghaus, H. (1885). Uber das Gedächtnis [Memory]. Leipzig: Duncker and Humblot. Evarts (citation forthcoming)

Freud, S., & Breuer, J. (1966). Studies on hysteria. (J. Strachey, Trans.). New York: Avon Books.
Gallwey, W.T. (1974). The inner game of tennis. New York: Bantam Books.

Graf, P. (1988). Implicit and explicit memory in same and different environments. Paper presented at Psychonomics, Chicago.

Graf, P., Mandler, G., & Haden, D.E. (1982). Simulating amnesic symptoms in normal subjects. Science, 218, 1243-1244.

Graf, P., & Schacter, D.L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11(3), 501-518.

Haldane, E.S., & Ross, G.R.T. (Eds.) (1967). The philosophical works of Descartes. Cambridge: Cambridge University Press.

Hull, C.L. (1933). Hypnosis and suggestibility. New York: Appleton Century.Jacoby, L.L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 110(3), 306-340.

James, W. (1890). Principles of psychology (Vol. 1, p. 123). New York: Holt.

Janet, P. (1893). L'amnésie continue [Continuous amnesia]. Révue Générale Des Sciences, 4, 167-179.

Janet, P. (1904). L'amnésie et la dissociation des souvenirs par l'émotion [Amnesia and the dissociation of memories by emotion]. Journal de Psychologie Normale et Pathologique, 1, 417-453.

Johnson, R.W., Wicks, G.G., & Ben-Sira, D. (1980). Practice in the absence of knowledge of results: Acquisition and transfer. Unpublished manuscript, University of Minnesota.

Jones, F.J. (1976). Escape from the monkey trap: An introduction to the Alexander technique. In F.J. Jones, The Alexander technique: Body awareness in action. New York: Schocken Books.

Korsakoff, S.S. (1899). Etude médico-psychologique sur une forme des maladies de la mémoire [Medical-psychological study of a form of diseases of memory]. Révue Philosophique, 28, 501-530.

Lessac, A. (1967). The use and training of the human voice: A practical approach to speech and voice dynamics. Mountain View, CA: Mayfield Pub.

McCracken, H.D., & Stelmach, G.E. (1977). A test of the schema theory of discrete motor learning. Journal of Motor Behavior, 9, 193-201.

Milner, B. (1962). Les troubles de la mémoire accompagnant des lésions hippocampiques bilatérals [Disorders of memory accompanying bilateral hippocampal lesions]. In Physiologie de l'hippocampe. Paris: Centre National de la Recherche scientifique.Roediger, H.L. (1990). Implicit memory. American Psychologist, 45 (9), 1043-1056.

Schacter, D.L. (1987). Implicit memory: History and current status. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(3), 501-518.

Schmidt, R.A. (1987). Motor control and learning: A behavioral emphasis (2nd ed.). Champaign, Illinois: Human Kinetic Publishers.

Schneider, W., & Fisk, A.D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. Perception and Psychophysics, 31(2), 160-168.

Squire, L.R. (1986). Mechanisms of memory. Science, 232, 1612-1619.

Stumpf, S.E. (1975). Socrates to Sartre: A history of philosophy (2nd ed.). New York: McGraw-Hill.

Thorndike, E.L., & Rock, R.T., Jr. (1934). Learning without awareness of what is being learned or intent to learn it. Journal of Experimental Psychology, 17, 1-19.

Verdolini-Marston, K. (1991). Processing characteristics of perceptual-motor memories with and without awareness. Unpublished doctoral dissertation, Washington University, St. Louis.

Verdolini-Marston, K., & Balota, D.A. (in press). The role of elaborative and perceptual integrative processes in perceptual-motor performance. Journal of Experimental Psychology: Learning, Memory, and Cognition.
Cite for modality specificity in training

NCVS Status and Progress Report - 6
May, 1994, 165-170

Interface Between Theatre Voice and Speech Trainer and Speech-Language Pathologist

Katherine Verdolini, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

How might the theatre voice and speech trainer and the speech-language pathologist interact? In the case of an acting student with a voice or speech problem, a classic scenario has been that the theatre trainer refers the student to the speech pathologist, who works with the student until the problem is resolved and the student is once again "normal." Then, the speech pathologist student sends the student back to the theatre trainer for continued artistic development.

This type of interface essentially amounts to handing the proverbial stick from one practitioner to the other (hopefully not treating the student like an incidental piece of baggage in the process). Although there is an interface in this case, there is actually little interaction between the practitioners. Are there better ways to think about interactions? Might the mutual enrichment be greater - and the care of the student be better - if we could think more creatively? That is the topic of this essay: creative thinking about the ways in which theatre voice and speech trainers (further abbreviated as "theatre trainers") and speech-language pathologists (further abbreviated as "speech pathologists") might interact.

We will do some thinking about the historical relations between the practitioners, we will think about what theatre trainers might get from speech pathologists and what speech pathologists might get from theatre trainers, what the benefits of reciprocal interactions might be, and where theatre trainers can turn for contact with speech pathologists. We will also point out some cautions in interacting across disciplines.

Historical and Philosophical Background

Historically, theatre trainers and speech pathologists have interacted preciously little. There are probably at least two major reasons. First, theatre and speech pathology people have tended to focus on different issues in training. Theatre trainers have usually focused on the development of esthetic and expressive capabilities across a very wide range of human emotions and situations. Speech pathologists have usually focused on restoring impaired voice and speech to normal status, for a comparatively limited repertoire of tasks. With these different goals, it is easy to see how theatre and speech pathology trainers might not have found much to talk about or share.

I think that there is also another - and perhaps more significant reason why theatre and speech pathology trainers have interacted little historically. We tend to talk different languages. That is, we tend to approach training with fundamentally different thought-tools. Although the distinction is not absolute by any means, it is probably safe to say that the theatre trainer's tools are often predominantly intuitive, whereas speech pathologists' tools are often predominantly analytical. In many cases, this type of difference might make communicating quite a challenge.

In thinking about this a bit further, what do we actually mean by "intuitive" versus "analytical" tools? I define intuitive processes as those that heavily rely on non-conscious mental operations, whereby the intuitions are the surface (or conscious) manifestations of such operations. The advantage of this approach is that non-conscious operations are carried out by an enormous capacity mental processor that is equipped to handle formidable amounts of information all at once. The solutions that emerge might be truly powerful ones because they are based on so much information. Analytical thinking, as I am thinking about it in this context, involves the slow, deliberate, and rational assembly of individual thought elements in consciousness. The results can be inspected and described verbally. The advantage is that this type of thinking is flexible; it can extricate itself to a degree from old habits that are run off

automatically, allowing for novel solutions. (The interested reader is referred to Kahneman, 1973, and Posner & Snyder, 1975, for a discussion of conscious and non-conscious information processing.)

How might such considerations affect interactions between theatre trainers and speech pathologists in a positive way? This question is the focus of the next sections.

What Can Theatre Voice Trainers Learn From Speech Pathologists?

Theatre trainers can learn a lot of techniques and facts from speech pathologists. Among the many, many examples, it might be useful for the theatre trainer to know that an important aspect of injury prevention and treatment is probably vocal fold adduction, or "pressing" (see for example, Hillman, Holmberg, Perkell, Walsh, & Vaughn, 1989; Jiang & Titze, in press; Peterson, Verdolini-Marston, Barkmeier, & Hoffman, in press; Verdolini, Druker, Palmer, & Samawi, submitted; Verdolini, Berry, & Titze, submitted; Verdolini-Marston, Burke, Lessac, Glaze, & Caldwell, in press). Theatre practitioners have suspected this conclusion for a long time, but it is nice to have confirmations from controlled experiments. Technical information of this type is undoubtedly useful. However, I think that the real benefit of interacting with speech pathologists goes well beyond the reception of technical information. I think that a more important benefit is the exposure to a way of thinking that is not necessarily salient in theatre traditions - the analytical mode. The point is not that the analytical mode is superior to the intuitive one - not at all. The point is rather that exposure to a new thinking mode, whatever that mode may be, provides an opportunity to expand one's thought repertoire. And how can theatre people reject exposure to any new experience of this type, interested as they are in the range of human experiences?

With an analytical thinking mode as we have described it here, theatre trainers might learn to combine rational information in a new way, generating new solutions to novel or old problems. The solutions might be the same as the ones generated by speech pathologists, or they might be different. The point is that well-supported analytical thinking, in which speech pathologists have considerable training, produces flexibility. New training systems can be devised to fit individual cases, old systems can be altered, and the effectiveness of training systems in general can be evaluated in a way that minimizes the trainer's own bias.

If the theatre trainer further becomes involved in the research side of speech pathology, yet other benefits can be expected. Yes, technical information can be gained from scientific work, that might be useful in training. However, again, I think that the real benefit of an involvement with research is the exposure to a somewhat peculiar philosophy, on which contemporary science is based. According to this thinking, one can never "prove" something to be true. One can only falsify a hypothesis, leaving other hypotheses available as possibilities. In fact, to assert that some fact is "scientifically proven" is completely oxymoronic. The basic itinerary is to successively disprove a series of hypotheses, theoretically with some final product left available as a possibility. The underlying idea is that "truth" is fundamentally elusive. It cannot be touched nor handled nor seen. It can only be inferred. This philosophy explains why some speech pathologists, who have exposure to scientific methodology, may be more tentative than theatre trainers about what works in training. Properly understood, the scientific philosophy illuminates new possibilities for thinking and discovery that might be exciting and stimulating for many theatre people, properly understood.

What Speech Pathologists Can Learn From Theatre Voice Trainers

Speech pathologists have a tremendous amount to learn from theatre trainers. Theatre approaches to voice and speech have evolved over decades, if not centuries, emerging from ponderous "field-testing" in real-life situations. A theatre approach simply will not survive if its trainees are not getting - and keeping jobs in demanding performance situations. The ultimate result is a robust voice and speech production approach that actually works - in lots of situations. Being constrained by such tough ecological considerations, most approaches to voice and speech in theatre are physiologically more complex, expressively more rich, and pedagogically more sound than many approaches in speech pathology. Theatre trainers may not know why (they might learn why by working with speech pathologists), and they may not be able to adapt the techniques for individual situations (they might learn to do so by being exposed to speech pathologists' analytical thinking processes), but I do think it is true that theatre systems are more complex and in some ways, more sound. Let me give a few examples. Theatre trainers tend to consider breathing, phonation, speech, and postural factors all together in training, leading to complex physiological acts that are coordinated across physiological systems (see for example, Lessac, 1967, OTHER CITES). Speech pathologists tend to train component voice and speech parts separately, and depending on the disorder, may never combine them all into a cohesive whole. Theatre trainers address a wide range of human emotions and vocal expressions (sorrow and ecstasy, whispers and screams), rendering their work exquisitely expressive. Speech pathologists, on the other hand, tend to address relatively quiet voice and speech in controlled environments. And by tradition, theatre trainers have tended to emphasize sensory (perceptual) experience in training, consistent with what I consider sound pedagogical principles (see my other essay in this text), as compared with

speech pathologists, who have tended to emphasize rational and verbal explanations in training.

From these considerations, it is easy to see how speech pathologists could benefit from exposure to theatre trainers' voice and speech techniques. But as we have already discussed, an even greater benefit might be exposure to a mode of thinking that is not particularly salient in speech pathology, that involves a generally intuitive approach to training. This is not to say that the speech pathologist would abandon analytical thinking - hardly! However, by learning to trust intuitive elements in training, even more than he or she already does, and by incorporating more of such elements in the training repertoire, the speech pathologist might greatly enrich his or her training style. In theatre, the intuitive approach includes detecting and responding to very subtle aspects of voice and speech production in pre-verbal, spontaneous ways, "moment to moment." The speech pathologist who is exposed to this mode might incorporate it in therapy, resulting in therapy sessions that are more relevant to the learner, with more powerful (encompassing) solutions than those that are sometimes achieved with the predominantly analytical mode that is common to speech pathology.

And Now, What About Actual Reciprocal Interactions?

To this point, we have talked as if theatre trainers might learn from speech pathologists, and speech pathologists might learn from theatre trainers, as if in independent moments. The theatre trainer might take formal courses in speech pathology, and the speech pathologist might receive some formal training in theatre. These solutions would undoubtedly contribute to reciprocal learning, and are discussed in the next section. But what about a truly integrated interaction? What might that look like?

Integrated interaction between theatre trainers and speech pathologists might take place in a theatre training room, or in a speech pathology clinic, where actual learners are involved. Let's say, for example, that the theatre trainer wants to institute a prevention program as a part of the training curriculum. The speech pathologist and theatre trainer might develop a coordinated curriculum, so that theatre approaches and speech pathology approaches are combined into single activities. Or in the case of an acting student with a voice or speech problem, the theatre trainer and speech pathologist might conduct as least some sessions together, so that their approaches are convergent.

The outcome of this type of attempt can be quite interesting. Training sessions tend to take on a new look, that is not the "theatre" look nor the "speech pathology" look. A new entity emerges from the combination of the two traditions. The entity tends to transcend the historical distinction between esthetics and restoration. Beauty and health converge. The itinerary no longer involves a sequential restoration of injury to normalcy, followed by training in the development of superior skills. Restoration and development become one and the same process.

Where Might the Theatre Trainer go to Find a Speech Pathologist With Whom to Interact, or for Specific Training in Speech Pathology?

One of the best ways to find a speech pathologist with whom to interact is to determine if there is one in the community with a specific interest (and possibly, expertise) in working with performers. Such persons may still be rare, but they do exist in some communities. Another tack would be to determine if there is a speech pathologist with a particular interest in voice disorders. That speech pathologist might be more willing - and able - to work with a theatre trainer as compared with a speech pathologist specialized in, say, neurological problems of speech and language due to stroke or disease.

If you want more formal training in speech pathology than collaboration with a single speech pathologist would afford, you might consider doing some coursework in a speech pathology department. There are many courses that might be relevant, including courses in anatomy, physiology, and acoustics of voice and speech, voice disorders, articulation or phonology disorders, and stuttering (many of the same issues seem to arise in fluency disorders as in voice).

A third possibility would be to attend regional or national conferences on voice. Some of the most important ones are listed in Appendix A.

A fourth, much more demanding option would be to enroll in a Master's program in speech pathology, for the purpose of obtaining licensure and certification as a speech-language pathologist. There are many good programs in the country, some of which include a special emphasis in voice - and performing voice - for students who are interested in it (see Appendix A).

Cautions

There are some cautions that you might want to consider if you decide to interact with speech pathology. Speech pathologists are not in any way frightening or dangerous people - on the contrary. But one danger is that with ongoing interactions, one or both of you might start to think that you are equally competent in the other's area. You might inadvertently overstep your respective professional boundaries. Even after much work with you or even with some coursework in theatre, a speech pathologist is not a theatre trainer and should not claim to be one. In the same way, unless you complete a Master's degree in speech pathology and other certification and licensure requirements, you cannot claim to independently treat voice and

speech disorders. There are legal implications in both directions. What I consider more important are the ethical considerations, which require fair and honest self-appraisal and community representation of ourselves and our abilities.

Another issue may be irrelevant in some cases, but relevant enough in others to be worth mentioning. Theatre people and speech pathologists may have a different interaction style, and typically do have a different training style. The differences are occasionally upsetting to some people. On average theatre people tend to be exuberant and relatively uninhibited in their expressive style, sometimes using colorful language and what some might consider non-traditional dress. In training, theatre sessions tend to explore a range of voice and speech possibilities with mental and physical experiences including floor work and touching. In general, therapy sessions in speech pathology tend to involve a narrower range of voice and speech activities, and the tone is relatively more "reserved" and less "emotional." You may react to the speech pathologist's work as restricted, and some speech pathologists might respond to you and your work as, well, shocking. I encourage you to persist. After all, the differences provide the actual impetus for interacting. They are the very source of richness that we seek.

Conclusions

Theatre voice and speech trainers and speech-language pathologists have much to gain by interacting with each other. There is the potential for the exchange of valuable technical information, as well as exposure to a different way of approaching training in general. I hope that theatre trainers will explore such possibilities by seeking out speech pathologists interested in the performer, and possibly by doing some coursework in speech pathology. By interacting, with time we may see the emergence of a new generation training mode, that crosses traditional boundaries with an integration of health and beauty in training, and of restoration and development.

Teachers of singing and speech pathologists have already considered some of the same issues together, and a joint statement between the National Association of Teachers of Singing (NATS) and the American Speech-Language-Hearing Association (ASHA) is provided in Appendix B. The Voice and Speech Trainers' Association (VASTA) is currently seeking to develop a similar joint statement, that will hopefully further stimulate our thinking on this important topic.

References

Jiang, J.J., & Titze, I.R. (1994). Measurement of vocal fold intraglottal pressure and impact stress. Journal of Voice, 8 (2) (pages to be determined).

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

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Lessac, A. (1967). The use and training of the human voice: A practical approach to speech and voice dynamics. New York: Drama Publishers.

Peterson, K.L., Verdolini-Marston, K., Barkmeier, J.M., & Hoffman, H.T. (in press). Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. Annals of Otology, Rhinology, and Otolaryngology.

Posner, M.I., & Snyder, C.R.R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), Information processing and cognition: The Loyola symposium. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Verdolini, K., Berry, D.A., & Titze, I.R. (submitted). Vocal efficiency from a therapeutic perspective: A computational approach. Journal of the Acoustical Society of America.

Verdolini, K., Druker, D.G., Palmer, P.M., & Samawi, H. (submitted). Physiological study of "resonant voice." Journal of Speech and Hearing Research.

Verdolini-Marston, K., Burke, M.K., Lessac, A., Glaze, L., & Caldwell, E. (in press). A preliminary study on two methods of treatment for laryngeal nodules. Journal of Voice.

Appendix A

Conferences including work on voice and professional voice in speech pathology and Master's programs with the possibility of a voice specialization

Annual Symposia: Care of the Professional Voice. Sponsored by The Voice Foundation. Usually held in June each year, in Philadelphia. For more information contact The Voice Foundation, 1721 Pine Street, Philadelphia, PA 19103; Phone (215) 735-7999; Fax (215) 735-9293.

Pacific Voice Conference. Usually held in October each year, in San Francisco. For further information contact Dr. Krzysztok Izdebski at 350 Parnassus Avenue, Suite 501, San Francisco, CA 94117; Phone (415) 476-2792.

American Speech-Language-Hearing Association Convention. Sponsored by the American Speech-Language-Hearing Association (ASHA). Usually held the weekend before Thanksgiving, in November. Location changes each year. For information contact ASHA National Office, 10801 Rockville Pike, Rockville, MD 20852; Phone (301) 897-5700; Fax (301) 571-0457. Also ask ASHA office about regional conferences occurring throughout the year.

Clinical Master's program with possibility of a special emphasis in voice disorders: Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa, 52242. For more information contact Ingo R. Titze, Ph.D., or Katherine Verdolini, Ph.D., (319) 335-6601, or the department chair, Arnold Small, Ph.D., (319) 335-8718.

Appendix B

The Role of the Speech-Language Pathologist and Teacher of Singing in Remediation of Singers with Voice Disorders·

ASHA and NATS Joint Statement

Since the founding of the American Speech-Language-Hearing Association (ASHA) in 1925 and the founding of the National Association of Teachers of Singing (NATS) in 1944, there has been increasing awareness of (a) the importance of having healthy laryngeal function in both speech and singing, and (b) the existence of a connection between optimal vocal usage in speech and optimal vocal usage in singing. The fundamental mechanism for healthy phonation is essentially the same for both singing and speaking. Therefore, it is recognized by both ASHA and NATS that the etiology of a voice disorder can be related to improper singing as well as to improper speaking technique.

Historically, development of the speaking voice related to disorders of laryngeal maturation and function has been the province of qualified speech-language pathologists. Similarly, development of the voice to its maximum function for use in singing has been the province of teachers of singing. The speech-language pathologist has given special attention to remediation of voice disorders to restore and maintain normal voice function. The singing teacher has given special attention to the development of full pitch and dynamic range, artistic quality, and vocal endurance of the singing voice. This has resulted in separate and independent work and in a serial approach to the remediation of voice disorders in singers.

In recent years, there has been increasing awareness within both ASHA and NATS that this dichotomous approach may not be in the best interest of singers with voice disorders. Rather, both organizations acknowledge that the most effective path to vocal recovery will often include an integrated approach to optimal voice care and production that addresses both speech and singing tasks. ASHA and NATS therefore affirm the importance of interdisciplinary management of singers with voice disorders, with the management team ideally consisting of, but not restricted to, at least an otolaryngologist, a speech-language pathologist, and a singing teacher.

ASHA and NATS recognize that there are a variety of possible configurations for such teams: Some ASHA-certified speech-language pathologists may also be experienced teachers of singing who are members of NATS. Similarly, there may be some otolaryngologists who hold ASHA certification in speech-language pathology or who are members of NATS. However, both ASHA and NATS recognize that such dual specialization is rare, and that in most cases, the management team will consist of at least three individuals.

Although ASHA and NATS recognize the differences in both professional preparation and in the primary goals of their respective memberships, both organizations acknowledge the need for broader, interdisciplinary training of speech-language pathologists and teachers of singing who plan to work with singers with voice disorders. The following general guidelines are recommended:

(1)The preparation of the teacher of singing needs to be augmented by inclusion of training in anatomy and physiology and in clinical management of voice disorders.

(2)The preparation of the speech-language pathologist who works with singers needs to be augmented in a parallel manner to include instruction in vocal pedagogy (the art and science of teaching voice) and vocal performance.

Both ASHA and NATS affirm that the following areas remain the province of our individual organizations to act upon as desired: (a) to develop more specific training requirements for individual members who elect to work with singers with voice disorders, (b) to develop criteria for evaluation and subsequent recommendation and/or accreditation of training programs, and (c) to develop criteria for identification and/or certification of individual members who obtain the recommended specialty training.

Both ASHA and NATS recognize the existence of state licensure laws that govern delivery of services to persons with communication disorders, including voice disorders. All persons who work with singers with voice disorders are encouraged to become familiar with these laws. ASHA and NATS affirm that it remains the responsibility of the individual practitioner to ensure that his of her work with singers does not violate the scope of practice defined by these laws.

ASHA and NATS encourage their members to cooperate in the development and delivery of interdisciplinary programs and services for singers with voice disorders.

January 21, 1992
New York, NY

This statement was accepted by the American Speech-Language-Hearing Association (ASHA) in March, 1992, and by the National Association of Teachers of Singing (NATS) on July 4, 1992. Members of the Ad Hoc Joint Committee for development of this statement included the ASHA representatives Janet Graves-Wright (chair), Reinhardt Heuer, and Stan Dublinske (*ex officio*), and the NATS representatives Jean Westerman Gregg (chair) and Ingo Titze.

NCVS Status and Progress Report - 6
May, 1994, 171-172

Training Update

Erich S. Luschei, Training Coordinator
Department of Speech Pathology and Audiology, The University of Iowa

During the past year, three postdoctoral fellows and seven predoctoral trainees have received financial support from the National Center for Voice and Speech. In addition to their academic studies, they have all been engaged in collaborative research projects with various faculty members of the NCVS. Their efforts have resulted in authorship on 8 publications and 17 posters or presentations at national scientific meetings. Five additional manuscripts describing the results of these studies are currently being prepared.

Postdoctoral Fellows

Kristin Larson, Ph.D., just started her training at the Denver Center for Performing Arts this past fall, but has already presented two papers at the convention of the American Association of Speech-Language and Hearing on her study of the voice of patients with Parkinson disease. She has also submitted a paper with Drs. Lorraine Olson Ramig and Ronald Scherer on this topic to the Journal of Medical Speech Pathology. Dr. Larson has also been instrumental in the plan to establish a laboratory, in conjunction with Dr. Marshall Smith and others at the DCPA, for conducting laryngeal EMG recording.

Young Min, M.D., has completed two years of study with Drs. Harry Hoffman, Erich Luschei, Ingo Titze, and Katherine Verdolini, and has been accepted for a residency program in Otolaryngology. She has worked on a variety of projects, including quantitative study of laryngeal EMG of patients with vocal fold paralysis, use of telemetry of EMG signals, biomechanics of human laryngeal muscles, effects of hydration on phonation threshold pressure, and the use of MRI in the evaluation of laryngeal paralysis. Her work has already resulted in two first-authored publications that are currently in press.

Savita Prakash, M.D., has worked at The University of Wisconsin at Madison under the direction of Drs. Diane Bless and Paul Milenkovic. She will begin a residency in Otolaryngology this summer. She has set up a "bench" for conducting studies of vocal fold movement in excised dog larynges and begun experiments using an improved ionic solution for maintaining the natural state of the tissue. She has also travelled to The University of Iowa with Carmen Ramos, a graduate student in Speech Pathology at The University of Wisconsin, to work with Dr. Erich Luschei on studies of the dynamics of botulinum toxin paralysis of the thyroarytenoid muscle of anesthetized dogs.

Predoctoral Trainees

Julie Barkmeier, Ph.D., completed her dissertation this spring under the direction of Dr. Erich Luschei. Her study quantitatively described the coverings of the recurrent laryngeal nerve of the dog, in particular the epineurium, and showed that this covering is proportionally thicker and more fat-containing than a comparison nerve of the leg. This finding may have implications for understanding the etiology of vocal fold paralysis. She will leave this summer to take a staff position in the Voice and Speech Division of the National Institutes of Health, working with Dr. Christy Ludlow.

Eileen Finnegan, M.A., has completed her predissertation research project, which exposed some of the inherent difficulties of using indirect behavioral methods of estimating subglottic pressure during phonation to calculate the laryngeal resistance of many individuals with spasmodic dysphonia. She has also been performing videolaryngoscopy as a member of the voice clinic team at the University of Iowa Hospitals.

Phyllis Palmer, M.A., is currently studying laryngeal physiology, particularly as it relates to human swallowing. She is analyzing EMG data from oral, pharyngeal, and laryngeal muscles as it relates to bolus size--work done in collaboration with Drs. Timothy McCullough and Adrienne Perlman. She is also helping with videolaryngoscopy of voice patients at the University of Iowa Hospitals.

Alice Smith, M.A., has contributed to numerous publications and presentations of research from the laboratory of Drs. Jerald Moon and John Folkins. She has completed her Ph.D. comprehensive exams, and is embarking on her dissertation work, which will examine the relations between velopharyngeal opening and closing gestures that are produced during experimental manipulations of speech rate, stress, and phonetic contexts.

Julie Stierwalt, M.A., has worked with Dr. Donald Robin on measures of tongue strength and fatigue and their relation to measures of intelligibility and overall speech effectiveness of individuals who have sustained traumatic brain injury. She has also been involved in studies of the effects of traumatic brain injury on all aspects of language and cognition.

Brad Story, B.S., has completed his comprehensive Ph.D. exams, and worked with Ingo Titze on the final development of a lumped element body-cover model of the vocal folds. He has also worked on improving the computer model of the vocal tract with regard to a yielding wall (in collaboration with Dr. Paul Milenkovic), and development of a method of changing the vocal tract shape, as in running speech, to produce more natural-sounding and intelligible speech simulation.

Kenneth Tom, M.A., has completed his coursework at the University of Iowa, and begun developing his thesis project, which will examine the mechanism of intensity control in male falsetto phonation. He is currently working as an Assistant Professor at Emerson College in Boston.

NCVS Status and Progress Report - 6
May, 1994, 173-175

Continuing Education Update

Julie Ostrem, Continuing Education Coordinator
Department of Speech Pathology and Audiology, The University of Iowa

Which phonosurgical techniques most effectively repair vocal folds damaged by disease or trauma? Are there therapy methods to improve the weak, breathy voices of Parkinson patients? What is full range of voice habilitation and rehabilitation methods available to speech-language pathologists for their clients with voice disorders?

These types of questions are at the crux of the Continuing Education program. Investigators glean results from basic science and clinical research projects within the NCVS and make them practical and understandable to practitioners in the voice and speech disciplines. Three target groups for this information are otolaryngologists, speech-language pathologists, and voice professionals. For the purposes of this project, voice professionals are defined as those whose occupations involve vocal training of others: theatre coaches, choral conductors and singing teachers.

A Little History

Early on in the granting period, target groups were identified and their educational needs assessed, relying heavily upon the clinical expertise of NCVS members. Pairs of investigators--a speech-language pathologist and an otolaryngologist--were identified at each consortium site. With their assistance, "information gaps" were prioritized and appropriate mechanisms for conveying the messages discussed. Both traditional and non-traditional formats were selected.

With this information at hand, the CE group has developed an ever-evolving timeline detailing development of projects for both the current granting period (through August of 1995) and into the future competitive renewal (through the Year 2000).

CE Projects in the Works

A number of projects were launched immediately because the need for information in selected clinical areas was great, and a format to convey this knowledge was readily apparent. Other projects have required more extensive planning and are still being developed. The group's most comprehensive project, development of a CD-ROM, is currently in its early planning stages.

Expressed Need: *Speech-language pathologists and other voice professionals may be unaware of the full repertoire of voice therapy and training techniques. They may be unsure of which method is most effective for a particular disorder or to achieve desired results.*

"Vocologist's Guide to Voice Therapy and Training"

Kate Devore, a master's student in the vocology track at Iowa, and Dr. Katherine Verdolini, NCVS investigator, researched current voice therapy methods. Six clinical voice therapy techniques and four popular voice training techniques from the theatre realm are described with the following criteria: population of application, technique developer, general description, underlying mechanism of effectiveness, efficacy studies (if any), and references.

This information, in booklet form, is entitled "A Vocologist's Guide to Voice Therapy and Training." It was written for two audiences. The booklet provides a quick reference for practitioners who work with the voice. Additionally, the guide is for individuals with voice problems who want to make educated choices about the treatment they receive.

As Volume Six of the NCVS Status and Progress Report goes to press, final editing and design of the booklet are being completed. "A Vocologist's Guide to Voice Therapy and Training" will be available for distribution in early summer 1994. Individuals may contact the NCVS or the NIDCD Clearinghouse for a free copy.

Expressed Need: *Music educators and choral conductors have, in general, received very little formal education in the areas of vocal function and vocal health (including preventive care).*

"Bodymind & Voice: Foundations of Voice Education"

To convey useful and timely information about the voice and its care, the NCVS approached the VoiceCare Network. Directed by Dr. Leon Thurman, the organization has a long tradition of providing information to music educators through its summer workshops and written materials. Jointly, the NCVS and VoiceCare Network will re-write an earlier publication entitled, "Bodymind and Voice: Foundations of Voice Education." The second edition will be written on two levels: for those who have minimal or no prior education in the voice sciences, and more detailed information for those who want to know more about the voice. Clinical applications of new research findings and detailed graphics are other planned improvements to the earlier version.

NCVS investigators will edit sections of the book according to their areas of expertise to ensure that translations of scientific jargon to lay language maintains accuracy. A re-writing of the text has begun, and it is anticipated that the completed volume will be available in Spring 1995.

NATS Columns

Dr. Ingo Titze continues to write a bi-monthly column for the National Association of Teachers of Singing Journal. The subject matter varies from column to column, but there exists a common theme of an integration of scientific principles with the art of singing. Recent columns have focused on the Mariachi singing style, singer's annual "vocal meltdown" difficulties in the late fall season, raised versus lowered larynx singing, and "Boosting the Singer's Formant in a Concert Hall."

Expressed Need: *Until recently, speech-language pathologists had very little to offer their clients with Parkinson disease and other neurologic disorders that cause weak and breathy voices. The recently-developed Lee Silverman Voice Therapy Method is showing great promise as an effective technique. How can the NCVS investigators communicate this method to other speech-language pathologists?*

Lee Silverman Voice Therapy Manual

Dr. Lorraine Olson Ramig, originator of the LSVT, and her colleagues have been invited to a number of workshops and conferences to demonstrate the method. However, it soon became obvious that a manual describing the method would reach larger audiences of clinicians more efficiently and economically.

Annette Pawlas, a speech-language pathologist who works closely with Dr. Ramig at The Denver Center for the Performing Arts, began writing the text in Spring 1994. Topics that will be included in the manual are: development and rationale for LSVT; week-by-week therapy plans (including sample worksheets and record sheets that may be copied and used by the clinician); discussion of possible problems and solutions in using the method; and reference information. The reference section will include a complete bibliography, reprints of clinical research articles, and information about Parkinson disease support groups.

The text will be fully written by September 1994. Appropriate diagrams will be added, and the editing process will be expedited so that the completed manual will be available by late 1994.

Expressed Need: *It is difficult for otolaryngologists to keep up with advances in voice surgery and disorders.*

Phonosurgery Conferences

The third in an annual series of phonosurgery conferences is planned for July 15-16, 1994, at The University of Wisconsin-Madison. These two-day seminars have been attended by about 200 speech-language pathologists, otolaryngologists, educators, and doctoral and postdoctoral students in each of the previous years. Assessment and management of patients with voice disorders are discussed, as well as comparative phonosurgical techniques. Dr. Diane Bless and her staff have been very successful in attracting a roster of respected experts in the discipline to participate in this forum.

Expressed Need: *The epidemiology of voice disorders is the initial, and perhaps most important, piece of information that should be disseminated to otolaryngologists and other voice clinicians.*

Epidemiologic/Demographic Projects

Student workers at Iowa have been hired to determine how many individuals in the U.S. depend upon voice to perform their jobs. Figures from the U.S. Bureau of Labor Statistics indicate that almost 3.5 percent of the U.S. working population have occupations in which voice use is necessary for public safety. Another 24 percent have jobs in which voice use is essential. Thus, than more 27 percent of the working population in the U.S. depends on voice for their livelihood.

Dr. Elaine Smith, an epidemiologist with the NCVS, is currently applying for additional funding to uncover the incidence of voice disorders in the United States. Other studies have described laryngeal pathologies by age, gender and occupation.

It is anticipated that all these data will be integrated and published in a booklet or brochure in 1995 and made available to speech-language pathologists and otolaryngologists.

Expressed Need: *Principles of voice production are generally not taught in a manner relevant to clinical problems. Thus, speech-language pathologists and otolaryngologists need additional information about diagnosis and management of voice disorders; indications, techniques, expected results of instrumentation (videostroboscopy and laryngeal EMG, for example); and a deeper understanding of normal vocal function.*

"Voice Production: Research and Reality"

A CD-ROM with the working title, "Voice Production: Research and Reality" is proposed for the next granting period. A number of topics will be linked by a set of scientific principles that apply to all aspects of voice production. A major emphasis will be placed on defining terms and explaining basic concepts. It is anticipated that learners will use the CD-ROM as a springboard to utilize other sources of information, such as journal articles and conference presentations.

The first segment of the CD-ROM is currently in production. It describes modes of vibrations of the vocal folds during speech using text, animated wire-frame diagrams, and video clips of canine vocal folds. Question and answer sections follow a self-paced, interactive format.

National meetings and Conferences

Though certainly not innovative in format, conferences provide an excellent arena for colleagues to share ideas, question methodology, reconsider accepted hypotheses, and debate research results. NCVS investigators have built an impressive track record for their contributions to national meetings and major conferences featuring current topics in voice and speech. More than 9,000 individuals were reached by presentations from NCVS investigators over the past year.

Dr. Marshall Smith has organized a series of Voice Care Seminars at the Wilbur James Gould Research Center, Denver Center for the Performing Arts. Those interested in voice care from the local community are invited to participate in the monthly sessions. Selected topics include: "Difficult Professional Voice Abusers I Have Known," and "All (Or More Than) You Ever Wanted to Know About Vocal Nodules."

Special conferences are also co-sponsored by the NCVS from time to time. For example, Dr. Gayle Woodson was an invited guest to The University of Iowa Department of Otolaryngology in April 1994. During her visit, a day-long conference on the topic of phonosurgery was presented for 110 regional otolaryngologists, speech-language pathologists, medical residents and students.

NCVS Status and Progress Report - 6
May, 1994, 177-180

Dissemination of Information Update

Barbara G. Bustillos, Dissemination of Information Coordinator
Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts

The key purpose of the dissemination project is to distribute information to the general public on prevention, detection, and treatment of voice and speech disorders. The methods of reaching the public include the use of media (television, radio and print), educational presentations (workshops, seminars and lectures) and shared information with other professional organizations dedicated to voice care.

Outreach to Special Populations

Since the last report, dissemination activities have targeted ***Latino*** and ***Student*** populations.

Latino Groups

Video presentations screened by Latino vocalists and students in New Mexico, California, Nevada, and Colorado, created an opportunity to present voice care information and allow the NCVS to explore the Mariachi vocalist.

The Latino population is today the largest ethnic population in the United States. The NCVS has addressed the need for voice care information by creating a video tape entitled "Exploring the Mariachi Voice." This video is a 17-minute story highlighting historical information and demonstrations of the different styles of Mariachi singing. Most importantly, the video teaches vocal care and demonstrates scientific information about the function of the voice. Dr. Ronald Scherer, Senior Scientist, and Dr. Florence Blager, Speech Pathologist, shared their expertise with recording artist and Mariachi Director Jose Hernandez. The video targets young Mariachi vocalists as well, including an interview with a 15-year old student from the rural town of Avondale, Colorado. Scientific information, anatomical charts, and vocal fold footage are presented in the video. Voice care information is emphasized.

The NCVS and the Los Angeles *Mariachi Heritage Society*, a nationally-known organization dedicated to teaching traditional Mariachi music throughout the Southwest and Mexico, have joined efforts to distribute the voice care video. The distribution plan focuses on Mariachi workshops in New Mexico, California, Arizona and Texas and nation-wide Mariachi concerts attended by the general public. Endorsements featuring international recording artists Lolla Beltran, Lucha Villa, and Miguel Aceves Mejia have created ties between the celebrities and the NCVS in efforts to raise vocal health awareness.

Video and media presentations have played an important role in bringing knowledge of vocal health to Mariachi students in the Southwest. Thanks to Colorado KCNC/NBC TV, The University of New Mexico, and the Mariachi Heritage Society, NCVS dissemination efforts have reached nearly 31,000 people at a variety of events as shown in Table 1.

In summary, "Exploring the Mariachi Voice" video tape is now available for distribution. Auxiliary projects, "Mariachi Voice" audio tapes, and "Exploring the Mariachi Voice - Part II" video and handbook, are works in progress.

Student Groups

In Denver, the Student population was targeted by the presentation of the arts-versus-science play, "Voices of People and Machines." This drama stars an opera singer and computer hacker who develop a friendship based on their desire to understand the human voice. They explore the possibilities of creating a synthesized voice, and together they begin to unravel the workings of the human voice. The performance incorporates music and dance with scientific/educational segments such as video footage of the vocal folds in motion and demonstrations of a MIDI sound module (synthesized voice). The play encourages the audience to think and care more about their own voices.

"Voices of People and Machines" has been presented primarily to voice professionals and university

Table 1.

Date	Audience Size	Audience	Presentation
June 1993	30,000	KCNC/NBC TV-4 audience (Denver)	TV-story
June 1993	10	Mariachi vocalists-concert (Denver)	Interviews
July 1993	300	Mariachi students-workshops (Albuquerque)	"Voice Doctors"*
July 1993	45	Mariachi artists-workshop (Albuquerque)	"Voice Doctors"
July 1993	30	Mariachi artist-workshop (Albuquerque)	"Voice Doctors"
August 1993	12	Jose Hernandez, Vocalist (Denver)	Interview
September 1993	28	Mariachi voice professionals (Los Angeles)	Screening
September 1993	45	Mariachi artists-workshop (Los Angeles)	"Mariachi video"**
September 1993	4	Vicki Carr, Performing Artist (Los Angeles)	"Voice Doctors"
December 1993	52	Mariachi artists-workshop (Phoenix)	"Mariachi video"
December 1993	35	Mariachi students-workshop (Phoenix)	"Voice Doctors & Mariachi Video"
December 1993	11	Mariachi students-L. Solano-workshop (Pueblo)	"Voice Doctors & Mariachi video"
February 1994	18	Pop vocalists, Los Lobos video (Denver)	"Voice Doctors & Mariachi video"
March 1994	12	Beltran,Mejia,Villa vocalists-interview (Los Angeles)	"Exploring the Mariachi Voice"***
March 1994	48	Mariachi vocalist workshop (Los Angeles)	"Exploring the Mariachi Voice"
March 1994	8	UCLA: Dr. Honrubia & staff (Los Angeles)	"Exploring the Mariachi Voice"
March 1994	12	Mariachi Heritage Society-video/lecture (Los Angeles)	"Exploring the Mariachi Voice"

*"Voice Doctors": Dr. Wilbur James Gould talks about why and how The Recording and Research Center was founded. Ronald Scherer discusses and demonstrates the type of research done at the RRC. Important visuals of the vocal folds has become a good teaching tool for the viewers. (Producer, Barbara Bustillos)

**"Mariachi Video": First draft of the Mariachi video describing the need for vocal care. (Producer/Writer, Barbara Bustillos)

***"Exploring the Mariachi Voice": Final product on the care of the Mariachi vocalist. (Director/Producer, Barbara Bustillos)

Table 2.

Date	Audience, Location, Size
2/93	Texas Music Educators Association,San Antonio, TX (100)
2/93	University of N. Texas School of Music, Denton, TX (400)
6/93	Voice Foundation symposia, Philadelphia, PA (200)
10/93	Acoustical Society Conference, Denver, CO (980)
10/93	Northwestern University, Evanston, IL (100)

student populations. Local media stories were developed as the show crisscrossed the Midwest. A total of 31,240 people viewed the play, and perhaps three times that many became aware of the human voice through articles in the local media.

The "Voices of People and Machines" comic book was produced as a supplement to the play: a glossary, anatomical drawings, and a brief description of how the voice works were included. Over 2,000 comic books have been distributed; plans are to enclose the comic book with each video distributed.

The Wilbur James Gould Voice Research Center, formerly The Recording and Research Center, a division of the Denver Center for the Performing Arts, is currently producing a 30-minute video for distribution entitled "Voices of People and Machines." It is scheduled for release in the summer of 1994.

Media Partnerships

The Wilbur James Gould Voice Research Center has developed a media outreach partnership with the community affairs department at KCNC/NBC TV-4 in Denver, Colorado. Since August of 1992, six video stories have been produced about voice. Some of the video segments have been distributed to area high schools, screened at arts organization meetings, and made available for NCVS promotional distribution to other voice research centers, such as The Voice Foundation, The National Stuttering Project, The Gerard Souzay Vocal Arts Foundation, and The Mariachi Heritage Society.

"Ciria - Young Vocalist" (June 1993) This story is about an 18-year old vocalist who is in the beginning stages of a recording contract and singing career. After attending a voice workshop, Ciria and her mother are accompanied by singing coach Larry Brown, D.M.A., during a typical day of rehearsal at the recording studio. Brown gives specific voice care recommendations to the young singer and record producers during a recording session.

A full videotape is planned that will combine current footage with pop vocalists, record producers, and managers to explain special voice care needs of the young

vocalist. Since the record and entertainment industry is a leading occupation for young vocalists in the United States, the voice care video will remind vocalists the effect of vocal abuse in regard to their livelihood and how voice care will impact on their quality of life. Future plans for videotaping include an exploration of the gospel singer: style, spirituality and vocal health concerns.

"Mariachi Voice" (August 1993) This story demonstrates the Mariachi art form which has existed since the late 1800's. According to the University of New Mexico Continuing Education program, Mariachi workshops across the Southwest have increased due to the interest of young vocalists and musicians that want to retain their culture and expand the artform. The story presents interviews with the top Mariachi groups from Mexico and the United States explaining the history and the importance of this music to the Latino communities.

"Arts and Science" (October 1993) This story plants the seed that students should consider careers as voice scientists and acousticians. Voice scientist Ingo Titze, Ph.D and acoustical engineer, Dana Hougland, encourage students in the sciences and acoustics fields. Titze and Hougland talk about how these fields play a crucial role in the arts. The story highlights a singing robot and introduces an arts & science comic book.

Over 100 hours of footage have been retained for future video development and distribution. Many of the short pieces produced for television features will be expanded to full videotape projects.

Co-sponsored Projects

The NCVS and *The Voice Foundation* jointly will distribute two medical videos entitled "Dr. Robert Sataloff: Clinical Voice Seminar" and "Dr. Robert Sataloff: Recognizing Problems in the Professional Voice User." The second video will be ready for distribution in the summer of 1994.

The National Stuttering Project and the *University of Colorado, Department of Communication Disorders and Speech Science* have been approached to co-produce a video about people who stutter.

Print Journalism

To maximize the Dissemination of Information project's modest budget, investigators utilize established media whenever feasible to spread the word about vocal awareness and voice care. Eric Richard, a recent graduate of The University of Iowa School of Journalism, served as a writing intern for Spring Semester 1994. Eric's primary assignment was to write articles about NCVS research for newspapers, magazines and newsletters.

As a first step, Eric selected three major topic areas that are understandable, yet not generally known to the general public: vocal health, stuttering, and voice therapy for individuals with Parkinson disease. He has proposed articles to magazines whose readers likely would be interested in these topics.

Vocal Health: A "top ten" list of vocal health tips were prepared from conclusions of research studies underway at the NCVS. These reminders are helpful to anyone, but particularly those who use their voices vigorously in their occupations or hobbies (public speaking or singing). Thus far, voice care tips have appeared in the following publications:

"Communication Briefings" (February 1994), a monthly newsletter distributed to communication professionals. (Circulation 46,000).

"The American Salesman" (April 1994), a monthly magazine for sales professionals. (Circulation 1,500).

"The Toastmaster" (April 1994), a monthly magazine for members of Toastmasters, a national organization for professional and non-professional communicators. (Circulation 170,000).

"Good Health News" (scheduled for publication September 1994), a pamphlet distributed free through a corporate chain of pharmacies to its customers. (Circulation 500,000).

"Vibrant Life" (accepted for publication, date not yet announced), a bimonthly consumer interest magazine on general health (Circulation 50,000).

"Your Health" (accepted for publication, no date yet announced), a semi-monthly magazine focusing on health and fitness. (Circulation 50,000).

Stuttering: Although stuttering affects an estimated one million individuals in the United States, the general public understands little about the etiology, diagnosis and available treatment for this speech disorder. A series of proposed articles were submitted to a variety of magazines.

"Instructor" magazine responded immediately requesting publication of an article for school teachers. The article will describe ways that teachers can help their students who stutter. No publication date has been announced. (Circulation 275,000).

Parkinson disease: Until recently, speech-language pathologists had little to offer clients experiencing voice problems as a manifestation of Parkinson disease. However, a new therapy--the Lee Silverman Voice Treatment method--has shown great promise in increasing loudness and decreasing breathiness. The thrust of this dissemination effort was to inform the general public that this innovative treatment is available.

"Parent Care," a new, monthly publication with issues of home care of needy parents has accepted an article for inclusion in the newsletter. No publication data has been finalized. Its circulation size is 125.

Other articles are in the preparation stage and will be offered to three Parkinsonism newsletters: The American Parkinson Disease Association (APDA) Newsletter, the Colorado APDA newsletter, and the Parkinson Disease Update newsletter.

Future Videotape Projects

Due to the strong connection between *Colorado KCNC/NBC TV-4* and the production facilities of the *Denver Center Media*, Dissemination investigators have been able to record interviews with voice professionals, celebrities, and other individuals for future dissemination purposes.

These collections are earmarked for audio stories for public radio and future video products. Also, continued cooperation with local media has lead to connections with members of national media affiliates and the educational community. These tapes also serve as an archive of the work of the NCVS. A sampling of the footage is listed in Table 3.

Table 3.

Dr. Wilbur J. Gould, otolaryngologist
Mr. Bobby McFerrin, jazz vocalist
Dr. Ingo Titze, voice scientist
Mr. Daniel Valdez, actor, musician and director
Mr. Edward James Olmos, actor, director
Mr. Robert M. Young, independent film director
Mr. Donald Seawell, CEO, DCPA
Mr. Lester Ward, president, DCPA
Mr. John Herrera, stage/film actor, theater director
Ms. Anna Moffo, opera singer
Dr. Robert Sataloff, M.D., director, The Voice Foundation
Mr. Jose Hernandez, mariachi director, vocalist
Mr. Pepe Martinez, director, Mariachi Tecalitlan, Mexico
Ms. Lolla Beltran, vocalist, Mexico
Mr. Miguel Aceves Mejia, vocalist, (stutterer) Mexico
Ms. Ciria Arrellano, 18 year-old pop recording artist
Ms. Lydia Solano, 16 year-old Mariachi recording artist
Ms. Lynn Skinner, jazz vocalist

We have also made contact with mezzo-soprano Marilyn Horne and actor Hal Holbrook to create public service announcements about voice awareness and voice care. These, along with the above mentioned list, will be distributed to national media networks in the hopes of creating a national campaign about voice care. Issues such as voice treatment for those with disorders and voice research also will be addressed.