NCVS Status and Progress Report
Volume 1/June 1991

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

Editorial and Distribution Information

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

Distribution of this report is not restricted.
However, production was limited to 500 copies.

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

Sponsorship

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

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

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

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

University of Utah
Department of Otolaryngology - Head and Neck Surgery

Memphis State University
Department of Audiology and Speech Pathology

University of Illinois
Department of Speech and Hearing Science

NCVS Personnel

Administration

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

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

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

Investigators and Support Staff

Fari Alipour-Haghighi, Ph.D.
Patricia Benjamin, B.S.
Florence Blager, Ph.D.
Diane Bless, Ph.D.
James Brandenburg, M.D.
Larry Brown, D.M.A.
Stefanie Countryman, M.A.
Heather Dove, M.A.
David Druker, B.S.
John Folkins, Ph.D.
Charles Ford, M.D.
Steven Gray, M.D.
Chen Guo, M.S.
Susan Hensley, M.A.
Margaret Hoehn, M.D.
Henry Hoffman, M.D.
Yoshiyuki Horii, Ph.D.
Joel Kahane, Ph.D.
David Kuehn, Ph.D.
Jon Lemke, Ph.D.
Julie Lemke
Erich Luschei, Ph.D.
Katherine Verdolini-Marston, Ph.D.
Brian McCabe, M.D.
Martin Milder, B.S.
Paul Milenkovich, Ph.D.
Kenneth Moll, Ph.D.
Hughlett Morris, Ph.D.
Jerald Moon, Ph.D.
John Nichols, B.A.
Neil Pinto, Ph.D.
Lorraine Olson Ramig, Ph.D.
Julie Ostrem, B.S.
Pam Rios
Donald Robin, Ph.D.
Ronald Scherer, Ph.D.
Elaine Smith, Ph.D.
Leatitia Thompson, Ph.D.
Ingo Titze, Ph.D.
Nancy Tye-Murray, Ph.D.
Vern Vail, B.S.
Sandy Wiley
William Winholtz, A.A.S.
Raymond Wood, M.D.
Patricia Zebrowski, Ph.D.

Doctoral Students

Julie Barkmeier, M.S.
Jack Jiang, M.D.
Alice Smith, M.A.
Brad Story, B.S.
Kenneth Tom, M.A.

Postdoctoral Students

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

Visiting Scholars

Carlin Hegeman, Ph.D., University of Northern Iowa
Steve Hsiao, M.D., Taiwan
Aliaa Khidr, M.D., Egypt
Kyoshi Makiyama, M.D., Japan
Marwa Saleh, M.D., Egypt

Advisory Board

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

Contents

Forward

The National Center for Voice and Speech was organized in October 1990 as a consortium of institutions with common and complementary interests in voice and speech problems. Within the consortium--which consists of The University of Iowa, The Denver Center for the Performing Arts, University of Wisconsin-Madison, and University of Utah--emphasis is on a broad spectrum of applications that include not only normal and abnormal voice and speech production, but also production by professional voice and speech-users. This requires multiple sites and varied populations for investigation. At The Denver Center for the Performing Arts, for example, developing actors and accomplished actors can be studied in residence at the National Theater Conservatory and the Denver Theater Company, respectively. On the other side of the spectrum, the combined otolaryngology departments at The University of Iowa, University of Wisconsin, and University of Utah offer a variety of populations, procedures, clinicians, and disorders for clinical trials. Finally, basic studies on acoustics, biomechanics and physiology of voice and speech are intensified by the combined doctoral and postdoctoral training programs in speech and hearing science departments at the consortium institutions. Some trainees have the flexibility to study at multiple sites.

The NCVS has four basic objectives: 1) to conduct independent research in voice and speech; 2) to provide thorough training in the biomedical sciences; 3) to disseminate information about voice and speech to the general public and to special "at risk populations"; and 4) to provide continuing education to practitioners in the field (otolaryngologists, speech language pathologists, singing teachers, vocal coaches and others involved in speech production).

The Status and Progress Report is divided into two parts. Part I contains preprints of articles submitted for publication in peer-reviewed, archival journals by Center investigators. There is a lag between ongoing activities and completed reports. Thus, much of the work reported here was conducted prior to the beginning of Center support in October 1990. These papers will appear elsewhere with minor modifications. Part II contains a variety of less formalized papers, reports, and summaries of activities conducted by Center personnel. Given that training, dissemination, and continuing education activities do not lend themselves to homogenous publication and reporting procedures, the length, depth and style varies considerably.

The Center is highly appreciative of the financial support and advice it receives from the National Institute on Deafness and Other Communication Disorders.

Ingo R. Titze
Director

Part I

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 1
June 1991, 1-14

Phonation Threshold Pressure: A Missing Link in Glottal Aerodynamics

Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Recording and Research Center, The Denver Center for the Performing Arts

Abstract

Phonation threshold pressure has previously been defined as the minimum lung pressure required to initiate phonation. By modeling the dependence of this pressure on fundamental frequency, it is shown that relatively simple aerodynamic relations for time-varying flow in the glottis are obtained. Lung pressure and peak glottal flow are nearly linearly related, but not proportional. For this reason, typical power law relations that have previously been proposed do not hold. Glottal impedance for time-varying flow must be defined differentially rather than as a simple ratio between pressure and flow. It is shown that the peak flow, the peak flow derivative, the open quotient, and the speed quotient of inverse-filtered glottal flow waveforms all depend explicitly on phonation threshold pressure. Data from singers are compared with those from nonsingers. The primary difference is that singers obtain two to three times greater peak flow for a given lung pressure, suggesting that they adjust their glottal or vocal tract impedance for optimal flow transfer between the source and the resonator.

Introduction

Glottal airflow is considered to be the source of sound in voice production. Temporal and spectral features of the glottal flow pulses are of primary importance in both speech analysis and synthesis. For studies in which physiologic control of glottal airflow is of interest, the dependence of pulse amplitude, pulse duty ratio (open quotient), and pulse skewing (speed quotient) on such variables as lung pressure, fundamental frequency, and vocal fold adduction is sought in an explicit way.

It is generally understood that the peak airflow increases with subglottal pressure and decreases with fundamental frequency and adduction, but exact analytical or empirical relations have not been determined. Likewise, expressions for open quotient and speed quotient and other temporal features have yet to be developed. Glottal impedance during vocal fold vibration remains a conceptual term that has neither a firm definition nor an analytic expression.

Whenever mean airflow or peak airflow in phonation is plotted against one of the physiologic variables named above, much scatter in the data is observed (Isshiki, 1964; Sawashima et al, 1988). Subjects use multiple strategies for control of the glottal air pulse, giving rise to this scatter. The purpose of this paper is to sort out some of the complex dependencies of glottal airflow on lung pressure, fundamental frequency, and vocal fold adduction. In particular, the phonation threshold pressure is built into a set of empirical equations relating the peak flow and the peak flow derivative. This has been a missing link. It takes a finite pressure to establish an infinitesimal oscillating flow. Simple power law relations between pressure and flow are therefore in error because they do not include this zero offset between the two variables. It will be shown that by including the phonation threshold pressure, a more consistent set of aerodynamic laws can be established.

I. Phonation Threshold Pressure

A modest amount of data has recently been published on phonation threshold pressure (P_{th}). On purely theoretical ground, (Ishizaka & Matsudaira, 1972; Titze 1988) it appears that in order to sustain small oscillation around an equilibrium pre-phonatory glottis, the mean subglottal pressure has to exceed the threshold value

$$P_{th} = k_t B c \xi_o / T, \tag{1}$$

where k_t is a transglottal pressure coefficient, c is the mucosal wave velocity in the vocal fold cover, T is the vocal fold thickness, B is the mean damping coefficient for mechanical vibration in the tissue, and ε_o is the pre-phonatory glottal halfwidth. Above this threshold value of pressure, small oscillations can build up, with energy being transferred from the airstream to the tissue. Below this threshold value, excessive energy loss in the tissue causes the oscillations to be damped out.

Note that the threshold pressure in equation (1) is lowered by decreasing any of four variables: transglottal pressure coefficient (i.e., the aerodynamic loss in the glottis), tissue damping, mucosal wave velocity, or pre-phonatory glottal half-width. The threshold pressure is also lowered by increasing the vocal fold thickness. Equation (1) is a special case for a rectangular pre-phonatory glottis. For a convergent or a divergent glottis, more complex expressions were obtained by Titze (1988). Threshold pressures ranged between 0.1 and 1.0 KPa (approximately 1-10 cm H_2O).

The theoretical treatment did not include the effect of fundamental frequency (F_o) on any of the variables, although it was pointed out that c is quite likely to increase with F_o, and T is known to decrease with F_o (Hollien & Curtis, 1960). Both of these changes would raise P_{th} with increasing F_o. An experimental study by Finkelhor *et al* (1987) showed that this was indeed the case. Although this study was primarily intended to show the effect of tissue hydration on P_{th}, the frequency effect was quite evident. Four excised canine larynges were phonated with an artificial air supply. Adduction was simulated by positioning the vocal processes with micro-manipulators, and vocal fold length was varied similarly to change F_o. Three levels of elongation were used to estimate the 10%, 50%, and 80% levels of F_o, respectively. A composite graph for the four larynges is shown as the curve labeled F in Figure 1. Note the rise in P_{th} with increasing F_o, from about 0.3 KPa at 10% of the F_o range to about 1.1 KPa at 80% of the F_o range.

Two other experimental studies, and perhaps the most direct ones pertaining to P_{th}, were conducted by Gramming (1988) and Verdolini-Marston *et al.* (1990) on human subjects. The

Verdolini-Marston *et al.* study related closely to the Finkelhor *et al.* study in that hydration effects were included. Three males and three females produced barely audible sounds at three pitches (low, medium, and high) and three conditions of hydration (normal, dry, and wet; the hydration was induced by environmental humidity, medication, and ingested water). The threshold pressure was measured indirectly by the technique described by Smitheran and Hixon (1981) for mean subglottal pressure.[1] P_{th} again increased systematically with F_o, as shown in Figure 1 by the curves labeled V (solid line for males and dotted line for females) for the normal hydration case. (For dry and wet conditions, the same general pattern of P_{th} as a function of F_o was observed, but the threshold shifted up for the dry and down for the wet condition.) In the Gramming (1988) study, hydration was not altered, but the subject pool was larger (10 normal males and 10 normal females). Results are shown by the curves labeled G in Figure 1. Note the similarity between the V and G curves, obtained on different subjects in different laboratories, but with similar experimental methods.

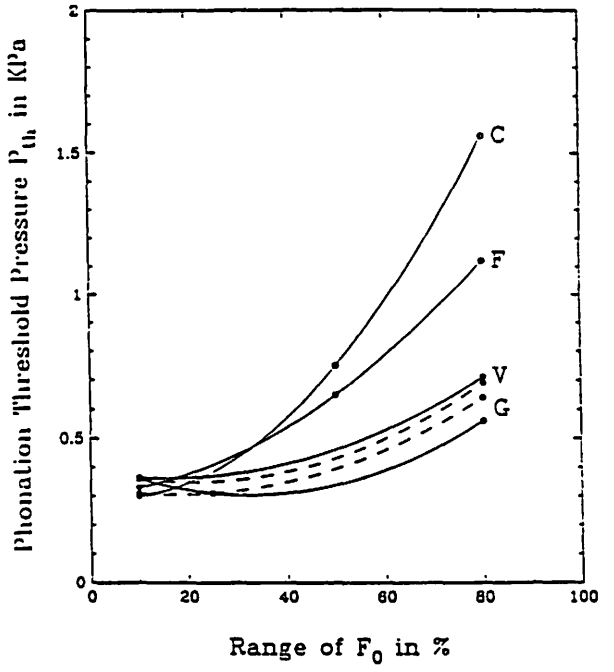

Figure 1. Change of phonation threshold pressure with fundamental frequency. Data set G is from Gramming (1988), F is from Finkelhor et al. *(1988), V is from Verdolini-Marston* et al. *(1990), and C is from Cleveland and Sundberg (1985).*

One other study by Cleveland and Sundberg (1988) is noteworthy. In this study, phonation threshold pressure was not targeted directly, but "soft" phonation was contrasted with "medium" and "loud" phonation for a variety of pitches. Subglottal pressure was again estimated by measuring oral pressure during a stop consonant in consonant-vowel repetitions. Results are also plotted in Figure 1 (curve labeled C). The subjects were three male singers, a baritone, a bass, and a tenor. It is seen that soft phonation pressures for these singers are much greater than for the V and G groups. The probable reason for this is that they rejected pure falsetto as a legitimate mode of phonation. Singers often phonate in a mixed register and do not target the threshold of phonation. At low pitches, however, the pressure values merge with those of other data sets. In all of the studies, low frequency pressures cluster around 0.3 to 0.4 KPa (about 3-4 cm H_2O).

The question remains whether any of the measured pressures represents oscillation threshold in the theoretical sense (as described by equation 1). Experimentally, phonation tends to "kick in" and "kick out" in a more abrupt way than small amplitude theory would predict. Furthermore, the "kicking in" may occur at a higher value of pressure than the "kicking out," suggesting a hysteresis (memory) for oscillation having previously been on or off. More theoretical and empirical studies are needed to resolve this issue. At present, we suspect that even the V and G curves in Figure 1 may overestimate the small amplitude threshold by about 0.1-0.2 KPa (1-2 cm H_2O), especially at low F_o.

II. Empirical Relations Involving Threshold Pressure

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

All subjects in the Sundberg *et al* study repeated the syllable [pæ] five times on each of five pitches, spanning about 1½ to 2 octaves. There were four loudness conditions (*p, mp, mf,* and *f*) on each pitch. In the Holmberg *et al.* study, there was one pitch level (normal speaking pitch) and three loudness levels (soft, normal, and loud). Subjects were 25 males and 20 females, a larger group than the singers. In addition to oral airflow, subglottal pressure was estimated from oral pressure during mouth occlusion. All signals were recorded on a VCR/PCM combination recorder for a later analysis.

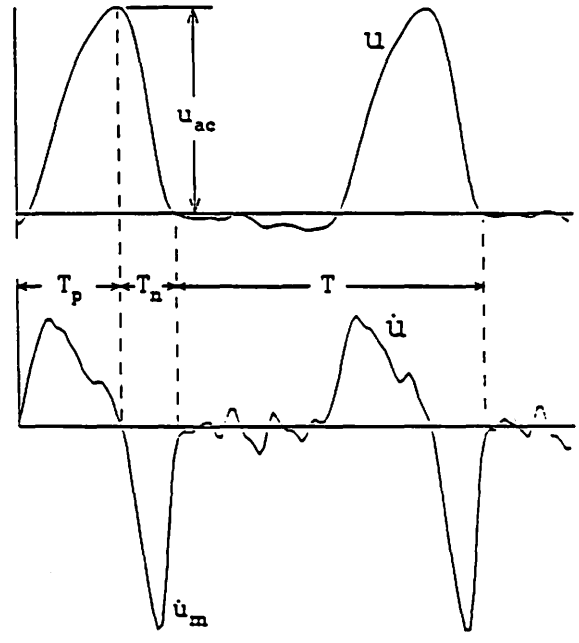

Figure 2. Two cycles of an inverse filtered glottal flow waveform u (top) and its derivative $\dot{u}$ (bottom). Time and amplitude scales are arbitrary. Note the definitions of the alternating flow u_{ac}, the peak flow derivative $\dot{u}_m$, the duration of positive flow derivative T_p, the duration of negative flow derivative T_n, and the period T.

After inverse filtering and digitization (20 kHz, 16 bits), each glottal flow waveform and its time derivative were displayed on a computer terminal. An example of two cycles of a flow waveform u and its derivative $\dot{u}$ is shown in Figure 2. (The leakage flow, often observed in inverse filtered waveforms and attributed to a posterior glottal chink, was subtracted to obtain the time-varying waveform shown. This leakage flow, also called DC flow or minimum flow, has no direct bearing on the time-varying pressure flow events discussed here.) From the inverse filtered waveforms, the following measures were extracted: the alternating flow u_{ac}, the peak negative flow derivative $\dot{u}_m$, the time for which the flow derivative is positive T_p, the time for which the flow derivative is negative T_n, and the period T. From these measures, the open quotient Q_o and the speed quotient[2] Q_s were calculated as

$$Q_o = \frac{T_p + T_n}{T} \qquad (2)$$

$$Q_s = \frac{T_p}{T_n} \tag{3}$$

The quantities u_{ac}, $\dot{u}_m$, Q_o, and Q_s will now be used to develop some empirical pressure-flow relations.

A. Peak Flow and Glottal Impedance

The alternating flow u_{ac} in Figure 2 is neither a peak flow nor a peak-to-peak flow in the sinusoidal sense. For glottal impedance calculations, where sinusoidal quantities are assumed, the flow pulse in the open phase can be approximated by a raised cosine of the form

$$u = u_{ac} - u_m(1 - \cos\omega t) \tag{4}$$

where u_m is the peak flow and ω is the angular frequency. This simple sinusoidal approximation obviously does not take into account pulse skewing or any other nonlinear pressure-flow effects, but for a time-independent (average) glottal impedance calculation in the open phase, these effects can be neglected.

At the instants of glottal opening and closing, the flow goes to zero, which allows a relation between u_{ac} and u_m to be determined,

$$u_m = \frac{u_{ac}}{1 - \cos\pi Q_o}, \tag{5}$$

where Q_o is the open quotient as defined above. The derivation is simple algebra when it is recognized that opening and closing occurs at $t = \pm Q_o T/2$, with $t = 0$ being centered at the peak value of the raised cosine. For $Q_o = 1$, the waveform is raised entirely above the baseline and $u_m = u_{ac}/2$, half of the peak-to-peak value. For $Q_o = 0.5$, the waveform is a half-sinusoid and $u_m = u_{ac}$.

All of the u_{ac} values in the Sundberg *et al.* and Holmberg *et al.* studies were adjusted according to equation (5) to obtain a reasonable estimate of the equivalent sinusoidal peak flow u_m in the open phase. Figure 3 shows a plot of averaged u_m values versus P_L for various groups of subjects. Results from the five tenors are shown with open triangles, circles, and squares, respectively, for three frequencies: a low frequency (117 Hz), a midrange frequency (234 Hz), and a high frequency (383 Hz). The low frequency represented an ensemble average of 105, 110, 110, 131, and 131 Hz, the frequencies actually produced by the five tenors. Likewise, the midrange frequency was the result of averaging 207, 220, 220, 262 and 262 Hz, and the high frequency was the result of averaging 349, 392, 392, 392 and 392 Hz. (The tenors used somewhat different musical pitches at which they could comfortably perform the vowel-consonant repetitions at different loudnesses.)

For comparison with control subjects, the filled triangles and asterisks represent data from Holmberg *et al.* for male nonsingers and female nonsingers, respectively. Note the much smaller

peak airflows for comparable lung pressures. Apparently, the singers are able to adjust the flow impedance, which includes the vocal tract input impedance, in such a way as to optimize the peak airflow. At a lung pressure of 0.5 KPa, for example, u_m is about 5 l/s for the tenors at 117 Hz and only about 1.5 l/s for the male nonsingers at 116 Hz. This is a ratio of more the 3:1 and must be regarded as truly remarkable.

The approximate 2:1 ratio between male and female u_m is expected on the basis of vocal fold scaling, as previously discussed by Holmberg *et al* (1988) and by this author (Titze, 1989b). A primary factor is the membranous vocal fold length, which is about 1.6 times greater in males than females. An additional scale factor for vibrational amplitude (estimated to be about 1.2-1.5) makes the time-varying glottal area about twice as large for males as for females. The time-varying flow would be expected to follow this glottal area scaling.

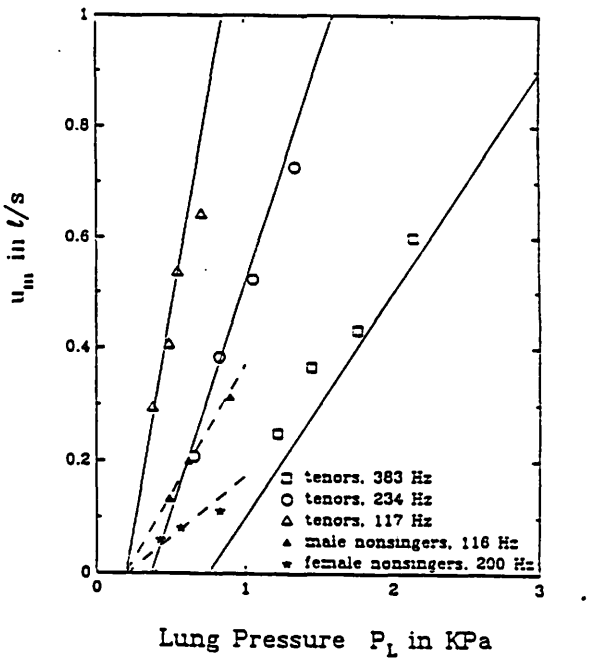

***Figure 3.** Peak flow u_m versus lung pressure. Data points are averages over 5 tenors, 25 male nonsingers, and 20 female nonsingers. Solid lines and dashed lines are empirical models as described in the text.*

Another major point of interest is the linear (but non-proportional) relation between u_m and P_L. This was not expected, but can perhaps be explained as follows. Unlike the static pressure-flow relation $u = (2P/\rho)^{1/2}A$, in which P is the transglottal pressure, is the air density, and A is the minimum glottal area, the time-varying pressure flow relation involves a yielding wall, i.e., a glottal area that increases with pressure. By recognizing that the vibrational amplitude increases approximately as the square root of lung pressure (Titze, 1989a), the glottal area A can also be judged to increase as $P^{1/2}$, which establishes the linear relation between u and P in Figure 3. The non-proportionality is explained on the basis of phonation threshold pressure, which is the projected intercept onto the x-axis for every data set. An empirical model, represented by the straight lines in Figure 3 (solid lines for tenors and dashed lines for nonsingers), is

$$u_m = k_1 \frac{P_L - P_{th}}{P_{th}} \tag{6}$$

where k_1 is 0.3 l/s for the tenors, 0.09 l/s for the male nonsingers, and 0.045 l/s for the female nonsingers. Any frequency variation of the flow is carried entirely by P_{th}, a finding that will prove to be highly significant.

Owing to the linearity of u_m and P_L, a meaningful glottal impedance can be defined differentially from equation (6) as

$$|z| = \frac{dP_L}{du_m} \tag{7}$$

$$= P_{th}/k_1 \tag{8}$$

Thus, the magnitude of the glottal impedance is proportional to P_{th} (which increases with F_o) and inversely proportional to k_1. Table 1 shows some examples of k_1, P_{th}, and |z| calculations for frequencies produced by the various subject groups.

Table 1.
Magnitude of glottal impedance and related variables

Subject groups	F_o (Hz)	k_1 (l/s)	P_{th} (KPa)	|z|(KPa s/l)
tenors	117	0.3	0.20	0.67
tenors	234	0.3	0.37	1.23
tenors	383	0.3	0.75	2.50
male nonsingers	116	0.09	0.20	2.22
female nonsingers	200	0.045	0.21	4.67

Note that the impedance is lowest for the tenors at low F_o and highest for the female nonsingers. This impedance, which includes the input impedance of the vocal tract, can apparently vary an order of magnitude across various subjects. Within subjects, it can vary by a factor of 5 or so, as seen by the low versus high F_o conditions in the tenors.

B. Open Quotient and Speed Quotient

Consider next the open quotient Q_o, shown as a function of P_L in Figure 4. The first observation is that Q_o ranges from about 0.5 to 0.8, averaged over subject groups. Ranges of individual subjects were from about 0.4 to 1.0. As expected, females have the highest open quotients, followed by singers at high F_o, then singers and nonsingers for progressively lower F_os. The difference between male singers and male nonsingers is not remarkable.

Q_o decreases with P_L in all cases, but it reaches an asymptote of around 0.4 to 0.6. This is quite understandable on the basis of increasing amplitude of vibration. As the amplitude of vibration of the vocal folds gets large, the glottis is open about half of the time and closed about half of the time, regardless of the spacing between the vocal processes. A moderate amount of abduction of the vocal processes does not keep the tissue from colliding over nearly the full length of the folds. For small amplitudes, however, any separation of the processes can prevent part of the glottis from closing, thereby increasing the open quotient.

The open quotients are projected (theoretically) to reach 1.0 at phonation threshold (top of the figure). Here the amplitude of vibration is so small that there is no glottal closure (as long as the vocal processes are slightly apart). For pressed voice or vocal fry, this reasoning does not hold. None of the subjects were asked to produce pressed voice or vocal fry, and none of them apparently did.

An empirical model for Q_o is shown in Figure 4 by the solid lines for singers and dashed lines for nonsingers. This model has the form

$$Q_o = k_2 + (1 - k_2)\frac{P_{th}}{P_L} \qquad (9)$$

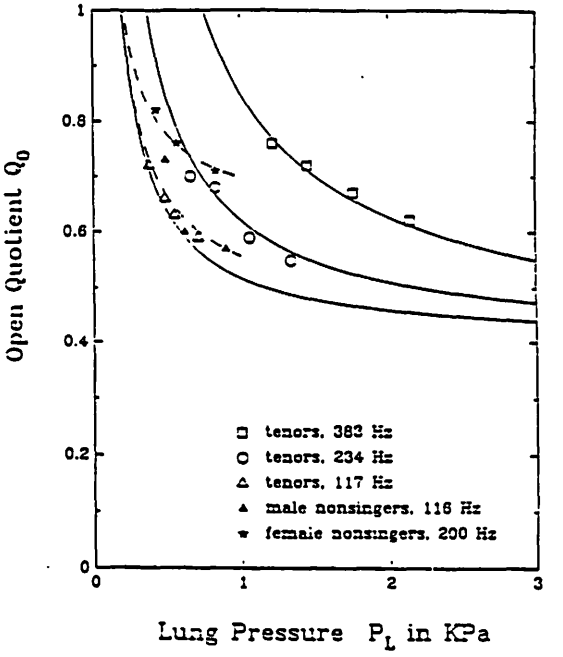

Figure 4. Open quotient Q_o versus lung pressure. Data points are averages over 5 tenors, 25 male nonsingers, and 20 female nonsingers. Solid lines and dashed lines are empirical models as described in the text.

for all subjects. The parameter k_2 is the asymptotic value of Q_o for large P_L. In Figure 4, it was set to 0.4 for male singers, to 0.45 for male nonsingers, and to 0.62 for female nonsingers. This simple one-parameter model gives a satisfactory match to the data. Note that phonation threshold pressure is again explicitly involved in equation (9), and that it accounts for the F_o variation of Q_o.

Consider now the speed quotient Q_s, shown in Figure 5 as a function of P_L. Averaged over subject groups, Q_s ranges from about 1.3 at low pressures to slightly greater than 2.0 at larger pressures. Speed quotients for individual subjects have a greater range, of course. Holmberg *et al* (1988) reported $Q_s > 3.0$ for some males phonating loudly, and some singers in the Sundberg *et al* (in press) study had $Q_s > 4.0$ for loud phonation.

Again, using only P_{th} and a few empirically determined coefficients, a model for Q_s is

$$Q_s = 1.0 + k_3 (2P_m - P_L - P_{th})(P_L - P_{th}) \quad . \qquad (10)$$

The quantity P_m is the pressure for which Q_s reaches a maximum. For males, the best value is 2.0 KPa and for females 0.65 KPa. The constant k_3 was set to 0.5 for the tenors, 0.55 for the male nonsingers, and 3.0 for the female nonsingers to achieve the fit to the data shown in Figure 5. The dependence of Q_s on P_{th} is highlighted by the pressure values at the projected intercepts between the solid lines and the dot-dashed horizontal line ($Q_s = 1$). These are the same intercepts that are found along the bottom of Figure 3 and along the top of Figure 4.

A question might be raised about the mechanism of Q_s increase with P_L. Usually, pulse skewing is attributed to vocal tract inertance (Rothenberg, 1981), which should not increase with P_L. But skewing of the flow pulse can also come from skewing of the glottal area function. This can be

pressure dependent. Subglottal pressure, for example, drives the bottom of the vocal folds more effectively than the top of the folds, the latter being nearly at supraglottal pressure (Ishizaka & Matsudaira, 1972; Stevens, 1977). Thus, if the vibrational amplitude at the bottom grows disproportionally in relation to the amplitude at the top, the speed of closure at the bottom becomes progressively greater because more distance is covered in less time (the oscillation period is the same). To complete the argument, since glottal area follows the top edge at opening and the bottom edge at closing, the speed quotient of the glottal area increases with P_L.

But what causes the Q_s increase to saturate, and ultimately even reverse again? At high frequencies, Q_s may decrease because vocal tract inertance is less effective in skewing the pulse (as the first formant is approached, the vocal tract becomes more resistive). This may be why Q_s is *generally* not higher for the tenors at 383 Hz in Figure 5 (open squares), but it cannot explain the saturation with increased pressure. More importantly, the large drop in Q_s for the highest pressure value in the female data (asterisks in Figure 5) is impossible to explain on the basis of vocal tract inertance. Given that F_o is only 200 Hz, any argument about a major reduction in vocal tract inertance seems weak. Other hypotheses need to be investigated.

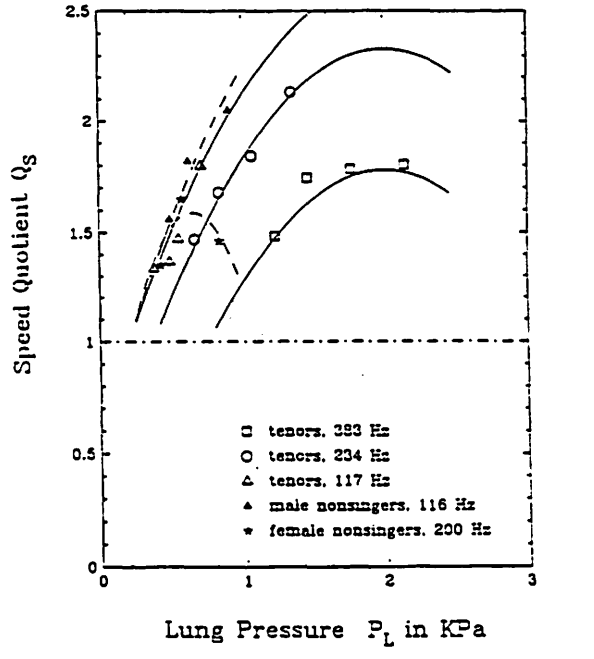

Figure 5. Speed quotient Q_s versus lung pressure. Data points are averages over 5 tenors, 25 male nonsingers, and 20 female nonsingers. Solid lines and dashed lines are empirical models as described in the text.

C. Maximum Flow Derivative

Consider now the maximum flow derivative $\dot{u}_m$, identified first in the bottom waveform of Figure 2. If the flow declination were a straight line, as in a triangular waveform, then

$$\dot{u}_m = u_{ac} / T_n \tag{11}$$

for this idealized case. Eliminating T_p in equations (2) and (3) and solving for T_n, we get

$$T_n = \frac{Q_o T}{Q_s + 1} \tag{12}$$

Now substituting (12) into (11) and replacing the period T with its inverse F_o, the expression

$$\dot{u}_m = u_{ac} F_o \frac{Q_s + 1}{Q_o} \tag{13}$$

is obtained. This expression is important because it shows the different ways in which the maximum negative flow derivative, which largely determines the source power, can be affected by waveform changes. Note that $\dot{u}_m$ is proportional to F_o, inversely proportional to Q_o, and proportional to $Q_s + 1$.

For a more sinusoidal waveform, $\dot{u}_m$ is underestimated by the straight line approximation. A correction fact k_4 can then be introduced such that

$$\dot{u}_m = k_4 u_{ac} F_o \frac{Q_s + 1}{Q_o} \tag{14}$$

Knowing that

$$\dot{u}_m = 2\pi F_o u_m \tag{15}$$

for any sinusoid (by time differentiation), it is clear that the correct result is obtained from equation (14) by letting k_4 approach $\pi/2$ when Q_s and Q_o both approach 1.0 (recall also that $u_{ac}=2u_m$ for a sinusoid). Thus, it is expected that k_4 ranges at least between 1.0 and 1.57, assuming that glottal flow waveforms have features that range between triangular and sinusoidal. For highly skewed sinusoids, however, k_4 could be expected to go higher, perhaps reaching 2.0 or more.

A test of the self-consistency of the empirical models developed thus far is to predict the $\dot{u}_m$ subject data with equation (14). This would contain an accumulation of all the errors of equations (6), (9), (10) and (14). Figure 6 shows the results of this prediction. A value of $k_4 = 1.5$ was chosen for the tenors to provide a best overall fit to the measured $\dot{u}_m$, suggesting that the waveforms were closer to sinusoidal than triangular, on average, in the falling portion. Inspection of the wave-

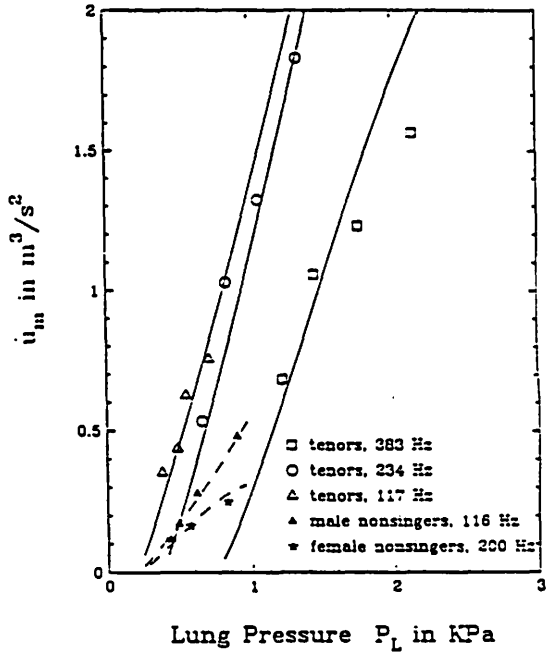

Figure 6. Peak flow derivative $\dot{u}_m$ versus lung pressure. Data points are averages over 5 tenors, 25 male nonsingers, and 20 female nonsingers. Solid lines and dashed lines are empirical models as described in the text.

forms confirmed this (see Figure 2). For the male nonsingers, k_4 = 1.9 gave the best fit, and for the female nonsingers, a slightly smaller value of k_4 = 1.8 matched the data reasonably well. It is not clear at this point why a single value of k_4 did not predict all of the data for all of the subject groups. Perhaps there were some systematic measurement errors (u_{ac} being too low for the nonsingers or $\dot{u}_m$ being too high for the singers), or perhaps some modeling errors accumulated in different directions. If there were no modeling or measurement errors, we would conclude that singers produce glottal waveforms that are more sinusoidal in the falling phase than those of nonsingers.

Before leaving the discussion of Figure 6, it should be pointed out that the singer-to-nonsinger ratio for $\dot{u}_m$ is somewhat less than that for u_m. The peak flow derivative is two to three times greater for the tenors than the male nonsingers (at a comparable F_o). Also, female nonsingers have more than half as great a $\dot{u}_m$ as male nonsingers. Nevertheless, the resemblance of Figure 6 to Figure 3 is quite remarkable, suggesting that the maximum flow derivative is influenced primarily by the maximum flow, and to a much lesser extent by waveform shape. This can be further appreciated by realizing that $Q_s + 1$ ranges only from about 2.3 to 3.3 in Figure 5, an increase of less than 50% over all frequencies, loudnesses, and subject groups. Q_o ranges only from about 0.6 to 0.8, a change of about 30%. By contrast, u_m ranges from 0.05 l/s to more than 0.7 l/s, a range of a full order of magnitude.

III. Frequency Dependence of Phonation Threshold Pressure

With specific values of P_{th} determined as x-axis intercepts in Figures 3-6, it is logical to attempt to construct an equation for P_{th} as a function of F_o. This equation would ideally predict the curves shown in Figure 1. But what kind of a function should this be, a linear, quadratic, or some higher-order polynomial? The curvatures in Figure 1 suggest at least a second order function.

Returning to equation (1), a second order function can indeed be postulated. The inverse relation between vocal fold thickness and F_o has already been mentioned. With regard to the mucosal wave velocity c, if we assume that the phase lag between the upper and lower margins of the vocal folds remains nearly constant with F_o (about 90°, or a quarter wavelength) then the well-known relation

$$c = F_o \lambda \tag{16}$$

for waves propagating in elastic media suggests that c is proportional to F_o. This gives at once a quadratic relation between P_{th} and F_o on the basis of c and T alone. The contributions of the remaining variables, k_t, B, and ε_o are less clear and cannot be assessed here.

One difficulty in postulating a relation between P_{th} and F_o is the behavior near the origin. As F_o approaches zero, does P_{th} also approach zero, stay constant, or rise above some minimum value? Figure 1 does not give a clear picture about that. The G and V curves appear to have a broad minimum at around 20-30% of the F_o range, whereas the C and F curves project a minimum value near F_o = 0. It is conceivable that phonation threshold pressure is minimum at "most comfortable pitch

level", but there is presently no theoretical backing for that. For the models presented in Figures 3-6, the best fit to the x-axis intercepts was obtained with the equation

$$P_{th} = 0.14 + 0.060\left(\frac{F_0}{\bar{F}}\right)^2 KPa \qquad (17)$$

where $\bar{F}_o$ is the mean F_o for conversational speech (120 Hz for males and 190 Hz for females). This yielded the P_{th} values shown in Table 1. Equation (17) is now represented graphically in Figure 7 (the lowest of the four curves) for male subjects. (For female subjects the shape is the same). Multiples of phonation threshold pressure ($2P_{th}$, $3P_{th}$, and $4P_{th}$) are also shown as a function of F_o to clarify the range of pressures used by the subject groups. Note that the tenors and male nonsingers used a pressure somewhat less than $2P_{th}$ for soft phonation and 3-$4P_{th}$ for loud phonation. Nonsingers, on the other hand, used a pressure of more than $2P_{th}$ for soft phonation and more than $4P_{th}$ for loud phonation.

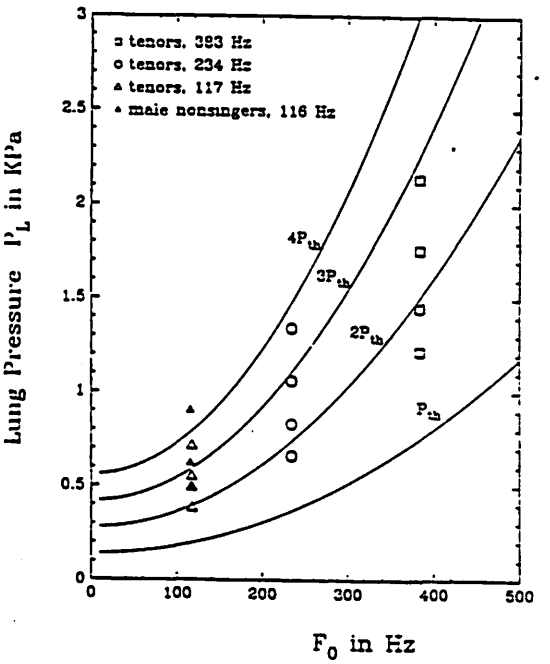

Figure 7. Lung pressures used by subjects in relation to phonation threshold pressure. Data points are for male singers and male nonsingers.

IV. Conclusions

This study has pointed out the need to include phonation threshold pressure as a key variable in glottal aerodynamics when vocal fold vibration is involved. Phonation threshold pressure is obtainable as an intercept along the pressure axis because relations between lung pressure and time-varying flow tend to be linear. Flow impedance can then be defined differentially as a constant over wide ranges of pressure. This impedance increases with phonation threshold pressure.

An important secondary result of this study was the discovery that singers are apparently able to lower their flow impedance by more than three-fold over nonsingers. This increases their peak flow (and peak flow derivative) dramatically, therewith increasing their glottal source power. It is speculated that this is partly a vocal fold adjustment and partly a vocal tract adjustment, but the exact contribution of each is not yet known.

There is no claim here that the numerical coefficients used in the empirical relations are final values. The number of subjects and diversity of subject groups need to be increased to build more confidence in the predictions of maximum flow, open quotient, speed quotient, and maximum flow

derivative as a function of lung pressure. More importantly, direct measurements of phonation threshold pressure need to be repeated, especially with mode of phonation and register being somewhat better defined.

Acknowledgements

This work was supported by a grant from the National Institutes of Health, P60 DC 00976-01. The author wishes to thank Professor Johan Sundberg, who visited The Recording and Research Center during a sabbatical leave from the Royal Institute of Technology, Stockholm, Sweden, and inspired some of this work. Special thanks also go to C.G. Guo for assistance in computer graphics and Pamela Rios and Linnie Southard for manuscript preparation.

References

Cleveland, & Sundberg, (**1988**). "Acoustic analysis of three male voices of different quality." In *SMAC 83, Proceedings of the Stockholm Music Acoustics Conference* , edited by A. Askenfelt, S. Felicetti, E. Jansson, & J. Sundberg, (Stockholm: Royal Swedish Acad. of Music), pp. 143-156.

Finkelhor, B.K., Titze, I.R., & Durham, P. (**1987**). "The effect of viscosity changes in the vocal folds on the range of oscillation." J. of Voice, 1(4), 320-325.

Gramming, P. (**1988**). *The phonetogram. An experimental and clinical study.* (Department of Otolaryngology, University of Lund, Malmö, Sweden).

Hollien, H., & Curtis, J. (**1960**). "A laminagraphic study of vocal pitch." J. Speech Hear. Res., 3(4), 361-371.

Holmberg, E., Hillman, R., & Perkell, J. (**1988**). "Glottal airway and pressure measurements for soft, normal, and loud voice by male and female speakers." J. Acoust. Soc. Am., 84(2),

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

Ishizaka, K., & Matsudaira, M. (**1972**). *Fluid mechanical considerations of vocal cord vibration.* (Monograph No. 8). Santa Barbara, CA: Speech Communication Research Laboratory.

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

Rothenberg, M. (**1981**). "Acoustic interaction between the glottal source and the vocal tract." In *Vocal fold physiology,* edited by K. Stevens & M. Hirano, (Tokyo, Japan: University of Tokyo Press. pp. 305-323.

Sawashima, M., Niimi, S., Horiguchi, S., & Yamaguchi, H. (**1988**). "Expiratory lung pressure, airflow rate, and vocal intensity: Data on normal subjects." In *Vocal fold physiology* (New York: Raven Press). pp. 415-422.

Smitheran, J., & Hixon, T. (**1981**). "A clinical method for estimating laryngeal airway resistance during vowel phonation." J. Speech Hear. Dis., 46, 138-146.

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

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

Titze, I. (**1989a**). "On the relation between subglottal pressure and fundamental frequency in phona tion." J. Acoust. Soc. Am., 85(2), 901-906.

Titze, I. (**1988**). "The physics of small-amplitude oscillation of the vocal folds." J. Acoust. Soc. Am., 83(4), 1536-1552.

Titze, I. (**1989b**). "Physiologic and acoustic differences between male and female voices." J. Acoust. Soc. Am.

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

Footnotes

[1] From this point on, the shorter term *lung pressure* will be used for simplicity to refer to *mean subglottal pressure as estimated by the Smitheran and Hixon technique.* It is recognized that a difference of at least 0.1 kPA (1 cm H_2O) can exist between these measured pressures and true lung pressure, but that does not defeat the arguments presented here. The term lung pressure is preferred conceptually because it is the control variable of interest.

NCVS Status and Progress Report - 1
June 1991, 15-37

Acoustic Interpretation of the Voice Range Profile (Phonetogram)

Ingo R. Titze, Ph.D.

Department of Speech Pathology and Audiology, The University of Iowa and
Recording and Research Center, The Denver Center for the Performing Arts

Abstract

The Voice Range Profile (VRP) is a display of vocal intensity range versus fundamental frequency (F_o). Past measurements have shown that the intensity range is reduced at the extremes of the F_o range, that there is a gradual upward tilt of the high and low intensity boundaries with increasing F_o, and that a ripple exists at the boundaries. The intensity ripple, which results from tuning of source harmonics to the formants, is more noticeable at the upper boundary than the lower boundary because higher harmonics are not energized as effectively near phonation threshold as at maximum lung pressure. The gradual tilt of the intensity boundaries results from more effective transmission and radiation of acoustic energy at higher fundamental frequencies. This depends on the spectral distribution of the source power, however. At low F_o, a smaller spectral slope (more harmonic energy) produces greater intensity. At high F_o, on the other hand, a shift of energy toward the fundamental results in greater intensity. This dependence of intensity on spectral distribution of source power seems to explain the reduced intensity range at higher F_o. An unrelated problem of reduced intensity range at low F_o stems from an inherent difficulty by a subject to keep F_o from rising when subglottal pressure is increased.

Introduction

The Voice Range Profile (VRP), also called the phonetogram (Damsté, 1970; Schutte & Seidner, 1983), is a display of vocal intensity versus fundamental frequency. Intensity (I) is usually plotted vertically as sound pressure level in dB and fundamental frequency (F_o) is plotted horizontally in linear or logarithmic units. A primary feature of the display is the manner in which the *range* of intensity co-varies with the *range* of fundamental frequency; hence, the term *Voice Range Profile*, which has recently been adopted by the Voice Committee of the International Association of Logopedics and Phoniatrics (IALP). Other labels for this display have been *Stimmfeld* (voice area) (Klingholz & Martin, 1983; Seidner, Krüger, & Wernecke, 1985),

phonogram (Komiyama, 1972), and *fundamental frequency-sound pressure level profile* (Coleman, Mabis, & Hinson, 1977).

The VRP has a characteristic shape, although many variations are found across individuals, across vowels, and across gender (Gramming, 1988; Klingholz, 1990). Male and female VRPs, adapted from Gramming, are shown in Figure 1(a) for the vowel /ɑ/ and in Figure 1(b) for the vowel /i/. Solid lines are for the male subject and dotted lines for the female subject. It is seen that the intensity range is usually greatest at some intermediate value of F_0 and diminishes toward the extremes of the F_0 range, although Gramming did not have her subjects phonate all the way to the extremes. The boundaries are not smooth, but display a fluctuating pattern. This suggests that there are some physiologic and acoustic constraints on what might otherwise be two parallel lines for "softest" and "loudest" phonation. In addition, there is often a gradual upward tilt in the "softest" and "loudest" contours, especially for /ɑ/, suggesting that there is a systematic increase in the average I with F_0. For the vowel /i/, there is a more restricted low-frequency I-range and a less restricted high frequency I-range.

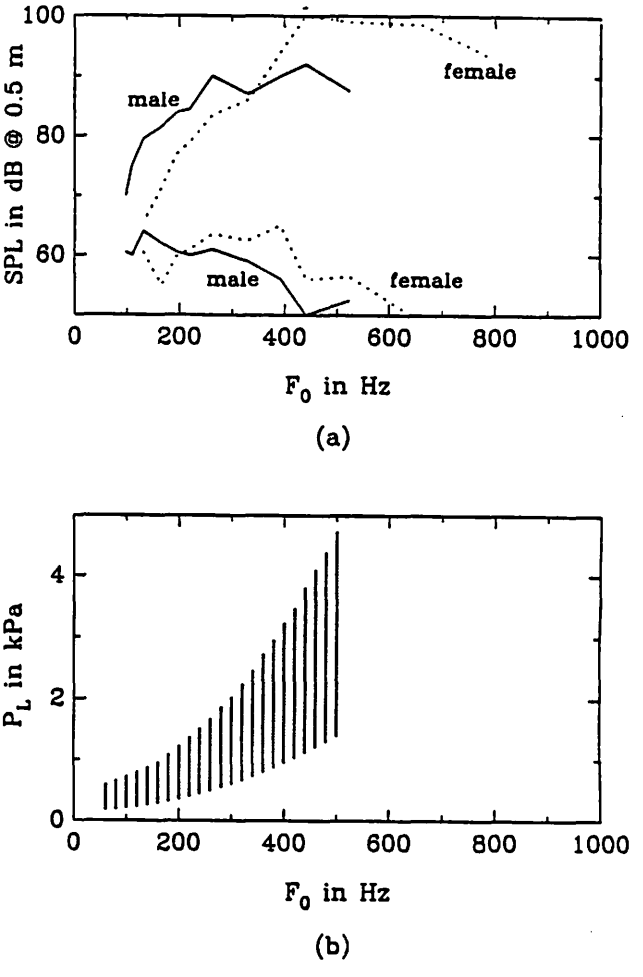

Figure 1. (a) Typical VRP from a male subject (solid lines) and a female subject (dotted lines) phonating an /ɑ/ vowel. (b) Same for an /i/ vowel. (After Gramming, 1988, by permission.)

The purpose of this paper is to develop an acoustic interpretation of the shape of the VRP. A model for co-variation of I and F_0 has been constructed. This model, partly theoretical and partly empirical, will be used to provide at least partial answers to the questions, (1) what determines the upper and lower boundaries of the VRP, (2) what determines the average I-F_0 slope, (3) what is responsible for the broad I-range at some mid-range frequency, (4) what constitutes the ripple at the boundaries, and (5) what are the primary physiologic mechanisms for changing I and F_0 between the boundaries? Much of the detail of the model is yet to be worked out because critical data on humans are missing. At present, some initial inferences will be made from measurements obtained on human subjects and excised canine larynges.

It is anticipated that a better understanding of the VRP will be helpful in guiding strategies for diagnosis and treatment of voice disorders. Voice training should also benefit from a better understanding of the acoustic interactions between F_0 and I. In time, vocal performance measures may be designed that utilize features of the VRP for training and rehabilitative purposes, giving rapid and accurate assessment of strengths and weaknesses of a particular voice. Specifically, three-dimensional VRPs, in which some aspects of voice quality (Pabon & Plomp, 1988) or muscle

activity (Klingholz, Martin, & Jolk, 1986) are being developed. These more experimental displays are promising for the future, but will not be discussed in this paper.

Section I. A Model for Co-Variation of Intensity and Fundamental Frequency

It is well known that voice fundamental frequency and voice intensity often co-vary in speech and singing. Independent control can be achieved, but sometimes with difficulty. In speech, a word or syllable stress is generally accompanied by an F_0 rise. People raise their voices in pitch when they raise their voices in loudness, and vice versa. In singing, pianissimo production can be problematic at high pitches, and fortissimo production is difficult at very low pitches. An acoustic interpretation of this co-variation begins with a review of the power transfer and radiation characteristics of the vocal tract.

A. Mouth Radiation

One major factor in the I-F_0 dependence is radiation of sound from a localized source. It is known from acoustic theory (Morse, 1948) that an isotropic radiator (a pulsating sphere, for example) radiates acoustic power proportional to the square of the frequency. This 6 dB/octave rise, which also applies to radiation from a piston in a circular baffle, has been used as a model for radiation at the mouth of a talker (Fant, 1960; Flanagan, 1972). Flanagan, for example, showed that the transfer function between radiated pressure and mouth flow is

$$P_r = \frac{sR_r}{s + R_r/I_r} U \quad , \tag{1}$$

where P_r is the Laplace transform of radiated pressure, U is the Laplace transform of mouth flow, s is complex frequency, R_r is the radiation resistance, and I_r is the radiation inertance. The radiation resistance and inertance are combined in parallel as shown in Figure 2 and have the values

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

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

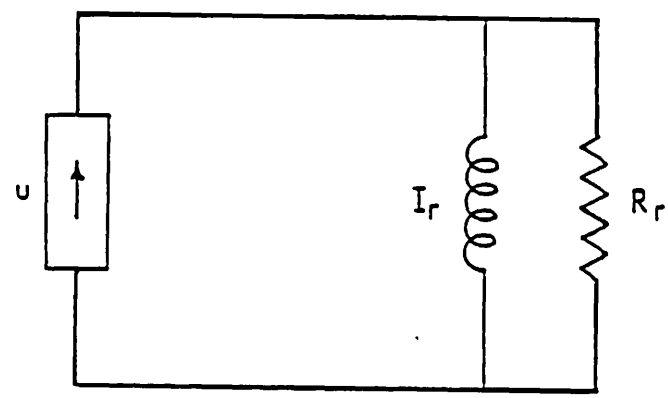

Figure 2. Equivalent electrical circuit for radiation at the mouth. I_r is the radiation inertance, R_r the radiation resistance, and u the mouth flow.

where ρ is the density of air (1.14 kg/m³), c is the sound velocity (350 m/s for warm, moist air), and r is the mouth radius (on the order of 1 cm for an /ɑ/ vowel). Note that for low frequencies (s→ 0),

the radiated pressure in equation (1) is proportional to frequency, whereas for high frequencies ($s \to \infty$), the pressure does not vary with s. The cutoff region is on the order of 3000-5000 Hz (depending on r), making the low frequency case applicable more often for speech than the high frequency case. This gives rise to the usual 6 dB/octave gain mentioned earlier.

By definition, the instantaneous radiated power is

$$\rho(t) = p_r(t)\,u(t) \tag{4}$$

$$= p_r^2/R_r \quad , \tag{5}$$

where p_r and u are the instantaneous radiated pressure and mouth flow, respectively (the inverse Laplace transforms of P_r and U). If p_r is steady-state and has Fourier components of the form

$$p_r = P_o \sin(\omega t + \theta) \quad , \tag{6}$$

where P_o is a frequency-dependent amplitude factor and θ is a frequency-dependent phase, then the average radiated power over one cycle at ω is

$$\overline{\rho}(\omega) = \frac{1}{T}\int_o^T P_r^2/R_r \, dt \tag{7}$$

$$= \frac{1}{2} P_o^2/R_r \quad . \tag{8}$$

The same result can be obtained in the frequency domain by letting $s = j\omega$ and performing the complex conjugate operation.

$$\overline{\rho}(\omega) = \frac{1}{2} P_r P_r^*/R_r \quad . \tag{9}$$

Recall that P_r in equation (1) is generally a complex quantity and that radiated power must be a real quantity; hence, the complex conjugate operation.Equation (9) allows the time-averaged power to be computed from frequency domain expressions, provided that U is known. This can be computed from a model of the glottal flow and the vocal tract transfer function.

B. Vocal tract transfer function

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

$$\frac{U(s)}{U_g(s)} = \frac{1}{\prod_{i=1}^{\infty} (1-s/s_i)(1-s/s_i^*)} \quad , \tag{10}$$

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

$$s_i = -\frac{\omega_i}{2Q_i} + j\omega_i \quad , \tag{11}$$

where ω_i is the i-th radian formant frequency and Q_i is the quality of the resonance. Wall losses are neglected in this all-pole cascade model.

Combining the vocal tract and the radiation transfer functions (equations 1 and 10), the ratio of radiated pressure to glottal flow becomes

$$\frac{P_r(s)}{U_g(s)} = \frac{sR_r}{(s+R_r/I_r)\prod_{i=1}^{\infty} (1-s/s_i)(1-s/s_i^*)} \quad , \tag{12}$$

$$= H(s) \quad . \tag{13}$$

This is the desired transfer function between the source flow and the radiated pressure. From H(s), the power can be calculated by equation (9) if the source spectrum is known.

C. The source spectrum

The simplest way to approximate the source spectrum is to assume a uniformly decaying spectral envelope with increasing harmonic number. This has the form

$$U_g(nF_o) = k(nF_o)^{-\alpha} \quad , \tag{14}$$

where n is the harmonic number, F_o is the fundamental frequency of the source, and k and α are constants to be evaluated. This source spectrum neglects several of the subtleties discussed recently by Ananthapadmanabha (1984) and by Fant and Lin (1988). Primarily, a broad, low frequency roll-off and a series of spectral zeros are omitted, but these do not substantially affect a first-order power calculation attempted here. In the future, however, a more realistic source model should be adopted.

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

$$\alpha = \frac{\textit{spectral slope in dB/oct.}}{20 \log_{10} 2} \quad . \tag{15}$$

For a typical slope of 12 dB/octave, the value of α is

$$\alpha = \frac{12}{6.0206} \approx 2 \quad . \tag{16}$$

For normal phonation, the spectral slope ranges from about 10 dB/octave to about 15 dB/octave (Sundberg, 1987), with lower values being possible for pressed voice and higher values for falsetto or breathy voice.

D. Total power radiated

The simple source spectrum $U_g(nF_o)$ in equation (14) is a real quantity and can be separated from the rest of the transfer function H(s). In other words, no source-tract interaction (Fant & Lin, 1987) needs to be considered for this simplified case. The radiated power of the n-th harmonic is then

$$\overline{P}_n = \frac{k^2}{2R_r} (nF_o)^{-2\alpha} |H(nF_o)|^2 \tag{17}$$

according to equations (9), (12)-(13), and (14). Note that the magnitude-squared operation in equation (17) is the effect of the complex conjugate multiplication in equation (9).

All quantities in equation (17) are known except the constant k. This can be evaluated by considering a low-frequency asymptotic condition for the radiated power. As F_o approaches zero, the vocal tract is a small fraction of an acoustic wavelength. For practical purposes, the vocal tract has negligible dimensions and the glottis is connected directly to the mouth. Glottal flow and mouth flow are then identical, as can be seen from equation (10) with $s \to 0$. For this asymptotic case, the instantaneous power delivered to the load R_r is, according to equation (5),

$$\mathcal{P}(t) = p_r^2/R_r = (I_r\dot{u})^2/R_r \quad , \tag{18}$$

where the low frequency condition $p_r = I_r\dot{u}$ has been applied.

Consider now a flow pulse as shown in Figure 3. Solid lines depict the flow u and dashed lines the flow derivative $\dot{u}$. In the second pulse shown, the flow derivative has been approximated by straight line segments. The quantities T_p and T_n refer to the durations for which the flow derivative is positive and negative, respectively, and $\dot{u}_m$ is the maximum negative derivative. With the straight-line approximations, it is easily shown (by time integration, as in equation 7) that the time-averaged low frequency power according to equation (18) is

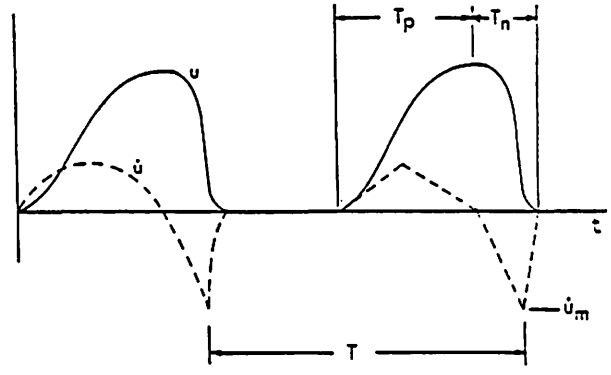

Figure 3. Two cycles of a glottal flow pulse (solid lines) and its flow derivative (dashed lines). In the second pulse, the flow derivative is approximated by straight lines for a power calculation.

$$\bar{\mathcal{P}}_o = \frac{\rho}{6\pi c}\frac{Q_o}{Q_s}\dot{u}_m^2 \quad . \tag{19}$$

Here Q_o is the open quotient, defined as $(T_p+T_n)/T$, with T being the period. The quantity Q_s, defined as T_p/T_n, is sometimes referred to as the *speed quotient*, but this author prefers *skewing quotient* because speed of closure is not necessarily correlated with rate of decrease of flow. The quantity $\bar{\mathcal{P}}_o$ will henceforth be called the low-frequency glottal reference power, or simply *reference power*.

No claim is made here that the waveform shape shown in Figure 3 is represented precisely by the spectrum assumed in equation (14). As mentioned earlier, an exact Fourier expansion of a

specific waveform yields spectral peaks and valleys that depend on the values of Q_o, Q_e, and $\dot{u}_m$, but the overall power of the pulse is less sensitive to the precise location of the peaks and valleys. Hence, an exact relationship between and the waveform parameters is not needed for this stage of development of the Voice Range Profile.

Equating the low frequency power in (19) to $\Sigma\overline{p}_n$, the summation of all of the harmonic components in (17), the constant k^2 can be evaluated as

$$k^2 = \overline{p}_o (2R_r) (2\pi F_o I_r)^{-2} F_o^{2\alpha} / \sum_{n=1}^{\infty} n^{2-2\alpha} \quad . \tag{20}$$

Strictly speaking, the summation of harmonic terms cannot go to infinity because the low-frequency approximation would not hold. However, only a few terms in the summation are significant for $\alpha \geq 2$ (i.e., for source spectra that fall off more than 12 dB/octave). The harmonic amplitudes decrease by $1/n^2$ or more, making only a few low harmonics significant in the sum. This is obviously not the case at normal to high fundamental frequencies, where harmonics amplitudes are boosted by the formants, but this is a low-F_o asymptotic calculation.

Substituting k^2 back into equation (17) and summing over all n, the total radiated power at any fundamental frequency becomes

$$\overline{p} = \overline{p}_o G(F_o) \quad , \tag{21}$$

where $G(F_o)$ is the power gain over the low frequency reference power,

$$G(F_o) = \frac{\sum_{n=1}^{\infty} n^{-2\alpha} |H(nF_o)|^2}{(2\pi F_o I_r)^2 \sum_{n=1}^{\infty} n^{2-2\alpha}} \quad . \tag{22}$$

Note that when F_o is small, $|H(nF_o)|$ approaches $2\pi nF_o I_r$ (equation 12 for real frequencies $s=2\pi nF_o$) and $G(F_o)$ approaches unity.

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

pulse.

E. Conversion from radiated power to intensity

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

$$SPL = 10 \log_{10} \frac{I}{I_o} \tag{23}$$

$$= 10 \log_{10} \frac{10^{12} \overline{\rho}}{4\pi R^2} \tag{24}$$

$$= 120 - 10 \log_{10} (4\pi R^2) + 10 \log_{10} \overline{\rho} \quad , \tag{25}$$

where R is the mouth-to-microphone distance in meters and ρ is the power in watts (equation 21). The reference intensity I_o is the standard 10^{-12} watt/m^2. Some correction to this far field, isotropic radiation expression is necessary if the mouth-to-microphone distance is small or if the microphone is highly directive. For R=0.5 m and a typical cardioid or omnidirectional pattern, the expression is expected to be accurate to within ± 2 dB (Flanagan, 1972), and equation (25) simplifies to

$$SPL = 115 + 10 \log_{10} \overline{\rho} \quad . \tag{26}$$

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

This concludes the acoustic derivations of power and intensity. Critical source parameters in expressions (19), (21), and (22) are $\dot{u}_m$, Q_o, Q_s, F_o, and α. Critical vocal tract parameters are r, the mouth radius, together with the formant frequencies ω_i and the resonant qualities Q_i. These parameters cannot all be varied independently by a speaker. Some further empirical relationships are needed to explore the combined control strategies used by individuals in intensity regulation. The

remainder of the paper will be devoted to an investigation of how the parameters affect each other and the VRP.

II. Parameter Variations and Empiracal Relations

Experimentally, vocal intensity has been shown to vary not only with temporal features of the glottal waveform (Sundberg & Gauffin, 1979; Gauffin & Sundberg, 1989; Sundberg, Scherer, & Titze, in press), but also with spectral features (Ananthapadmanabha, 1984; Fant & Lin, 1988). Thus, it is appropriate to investigate the manner in which spectral slope can be utilized in the control of vocal intensity. This is presumably accomplished by changing glottal adduction and other physiologic mechanisms that affect waveform shape.

A. Variation of intensity with spectral slope and F_0

Figure 4 shows a modeled power spectrum (equation 17) converted to SPL for a fundamental frequency of 130 Hz and a spectral slope of 12 dB/octave. The source envelope is depicted by short horizontal bars across the harmonics. Asterisks above the spectral lines indicate the accumulated radiated power (converted to SPL) with the addition of successive harmonics (the summation in equation 22). The vowel is a simulated male /ɑ/with ten cascaded formants (730, 1090, 2440, 3550, 4000, 5500, 6500, 7500, 8500, 9500 Hz), the first three of which are modeled after Peterson and Barney (1952) and the remainder are estimates. The formant Qs were all set to 10 and the mouth radius was 1.0 cm.

Note that the fifth and sixth harmonics, which are closest with F_1, add about 3-4 dB each to the SPL. The second, third, and fourth harmonics all add about 1 dB each, and the eighth harmonic, which is in the vicinity of F_2, adds about 1 dB. All higher harmonics add negligible power for this spectral slope. In Figure 5, where the spectral slope is 18 dB/octave, the sixth harmonic adds only about 1 dB and all other

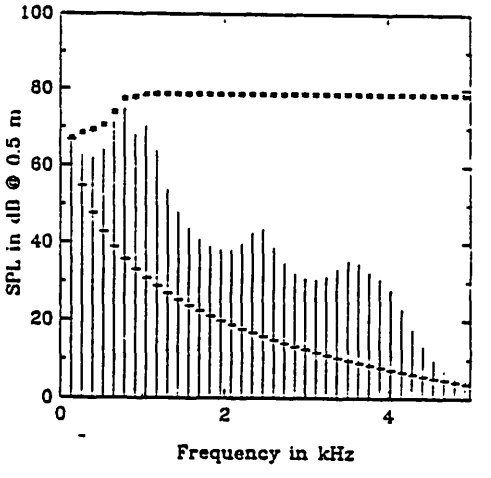

Figure 4. Harmonic power spectrum of the VRP model (equation 17) for an /ɑ/ vowel and F_0=130 Hz. Short horizontal bars indicate the source spectrum envelope with a 12 dB/octave roll-off. Asterisks indicate total power accumulation with the addition of successive harmonics.

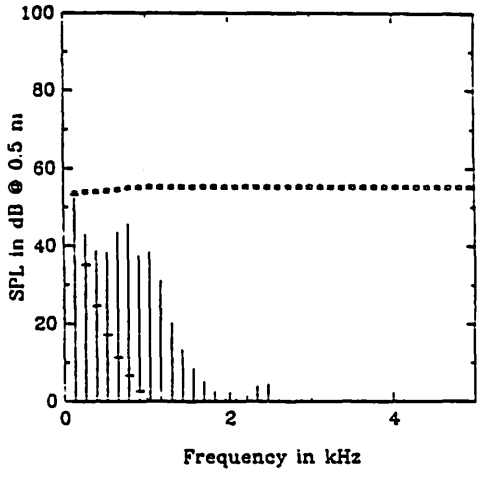

Figure 5. Harmonic spectrum of the VRP model (equation 17) for an /ɑ/ vowel and F_0=130 Hz. Short horizontal bars indicate the source spectrum envelope with an 18 dB/octave roll-off. Asterisks indicate total power accumulation with the addition of successive harmonics.

harmonics contribute no significant power. Thus, as Gramming and Sundberg (1989) have pointed out, the contribution of harmonic power to the SPL is a strong function of the spectral slope at the source.

As F_o is raised, the value of adding higher harmonics to the source is diminished. This is seen in Figure 6, where F_o has been raised to 600 Hz (near the vocal range limit for most females and perhaps beyond the limit for most males). Here the spectral slope is again 12 dB/octave, but the higher harmonics (n>1) add almost no power at this F_o. They are all above the first and second formants, thereby picking up little reinforcement from the vocal tract. Harmonics in the third, fourth, and fifth formant regions are also too low in energy to contribute to the overall power. (An exception may be if the subject has a vocal ring, i.e., a singer's formant, which boosts the gain in the 3000 Hz region. This special case will not be discussed here, however.) At high F_o, therefore, the intensity is determined almost solely by the fundamental, regardless of the spectral slope.

The combined variation of intensity with spectral slope and F_o is shown in Figure 7. F_o is along the abscissa and spectral slope is the parameter for the two fluctuating curves. The fluctuations result from harmonics being tuned to the formants. For the /ɑ/vowel, the largest peak occurs when the fundamental lines up with the first formant (F_o = 730 Hz). The smaller set of peaks to the immediate left corresponds to the tuning of the second harmonic to the second formant ($2F_o = F_2$ = 1090 Hz), and the third peak from the right corresponds to the tuning of the second harmonic to the first formant ($2F_o = F_1$ = 730 Hz). Further peaks correspond to third, fourth, and higher harmonics being tuned to F_1. Fluctuations in intensity with formant tuning can exceed ± 5 dB at higher F_o, but are usually not more than ± 3 dB in the male speech range, where only harmonics higher than $n \geq 4$ sweep through the formants (recall Figure 4).

Figure 7 begins to answer some important

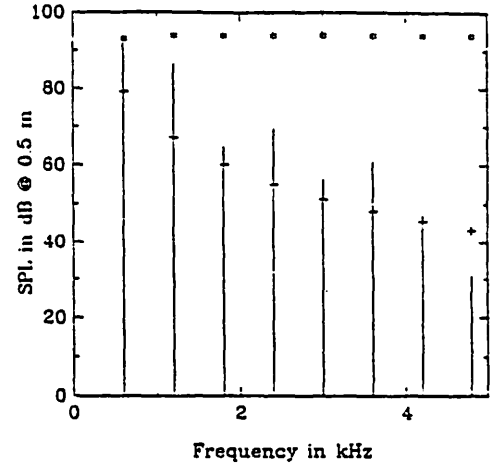

Figure 6. Harmonic spectrum of the VRP model (equation 17) for an /ɑ/ vowel, with F_o raised to 600 Hz. Short horizontal bars indicate the source spectrum envelope with a 12 dB/octave roll-off. Asterisks indicate total power accumulation with the addition of successive harmonics.

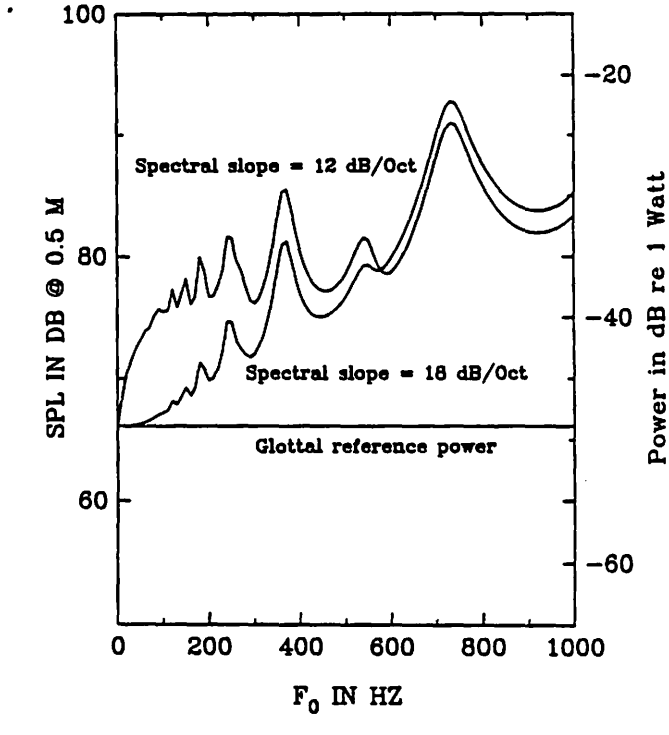

Figure 7. Variation of sound pressure level (SPL) with fundamental frequency when the glottal reference power is set to a constant. The parameter is spectral slope, and the ordinate on the right represents acoustic power radiated.

questions about intensity regulation. Aside from the fluctuations, intensity increases gradually with F_o when compared to the glottal reference power (straight horizontal line), which has deliberately been held constant with F_o. The SPL-F_o slope is a strong function of the spectral slope. For a spectral slope of 12 dB/octave, the intensity rises quickly at low F_o and then becomes rather flat (aside from the ripples). For a spectral slope of 18 dB/octave (a virtually sinusoidal flow pulse), the intensity rises slowly at first, but then approaches a steady 6 dB/octave rise over the 200 Hz to 800 Hz range. Note that the two curves cross at about 600 Hz. From this point upward, the near sinusoidal pulse is more effective in boosting the intensity than a pulse with a rich harmonic spectrum. The explanation is that the fundamental experiences a steadily decreasing impedance from the vocal tract as F_o approaches F_1. While the impedance for the higher harmonics fluctuates (because these harmonics pass through the formants), the impedance for the fundamental is monotonically decreasing (except for a few vowels such as /i/ or /u/, which have F_1 low enough that it may be surpassed by F_o). The more the spectral distribution is shifted to the fundamental, therefore, the greater the average SPL-F_o slope will be for F_os that are high, yet below F_1.

This leads to a prediction that the bottom curve of the VRP should rise faster than the top curve, assuming that soft phonation produces a more sinusoidal glottal waveform than loud phonation. The two curves should approach each other at the high end of F_o. This is, in fact, a characteristic of most VRPs that have been obtained experimentally (see Figure 1a, for example; see also the data of Coleman *et al.*, 1977). The curves also approach each other at the low end of the F_o range, but for different reasons, as will be discussed later.

The space between the two curves in Figure 7 (from about 100 Hz to 500 Hz) may be considered the beginnings of the male VRP for the vowel /ɑ/, but much more goes into the complete picture. In particular, the glottal reference power will be shown to change with vocal effort and with F_o. It is important to recognize at this point, however, that considerable intensity variation can be achieved by manipulation of the spectral slope alone. Nearly 10 dB increase in SPL is predicted as the spectral slope changes from 18 dB/octave to 12 dB/octave in the low F_o region (100-200 Hz). This suggests that intensity regulation by spectral means may be a particularly viable option for speech and low-pitched singing. By contrast, in the 600-800 Hz region, intensity is lost by the same spectral slope maneuver.

B. Control of intensity by lung pressure

Nothing has been said so far about lung pressure,[1] the primary control variable for vocal intensity. In order to complete the VRP model and elucidate further the I-F_o dependence physiologically, some empirical relations are needed to tie together to co-variation of $\dot{u}_m$, Q_o, Q_s, F_o, and spectral slope through lung pressure. This will also affect the glottal reference power.

In a concurrent publication (Titze, in press), the variations of $\dot{u}_m$, Q_o, and Q_s with lung pressure and F_o were studied on the basis of existing data on human subjects (Holmberg *et al.*, 1988). The following empirical relations were developed for 25 normal male subjects and 20 normal female subjects who had no vocal training:

$$Q_o = 0.40 + 0.60\frac{P_{ch}}{P_L} \qquad \textit{for males} \qquad (27a)$$

$$Q_o = 0.45 + 0.55\frac{P_{th}}{P_L} \qquad \textit{for females} \qquad (27b)$$

for males (28a)

$$Q_s = 1.0 + 0.55(4.0 - P_L - P_{th})(P_L - P_{th})$$

$$Q_s = 1.0 + 3.00(1.3 - P_L - P_{th})(P_L - P_{th}) \qquad \textit{for females} \qquad (28b)$$

In these equations, P_L is the lung pressure in kPa and P_{th} is the phonation threshold pressure, defined as the minimum lung pressure required to initiate vocal fold vibration (Titze, 1988b). An empirical model for P_{th} will be given below, but for the present it suffices to say that a good fit to the Holmberg *et al.* data was obtained by setting $P_{th} = 0.2$ kPa for both males and females.

Both of the waveform quotients asymptote to 1.0 as P_L approaches P_{th}, but Q_o decreases with P_L while Q_s increases with P_L. For example, at $P_L = 0.4$ kPa, Q_o has decreased to 0.70 for males and to 0.73 for females, while Q_s has increased to 1.37 for males and 1.42 for females. This corresponds to relatively soft phonation. For louder phonation ($P_L = 0.8$ kPa), Q_o has decreased to 0.55 for males and 0.59 for females, while Q_s has increased to 1.99 for males and 1.54 for females. These values agree well with those reported by Holmberg *et al.*, both in terms of overall range and in terms of trends with changing P_L.

Equations (27a,b) and (28a,b) are not considered to be final in their development. In particular, equations (27a,b) should ultimately involve the adductory process explicitly. This is not controlled here and may vary significantly from person to person. Adductory control could be included by replacing the constants 0.40 and 0.45 with variables to cover the range of adduction from pressed voice to breathy voice.

The maximum flow derivative $\dot{u}_m$ was also modeled after the Holmberg *et al.* data:

$$\dot{u}_m = 0.00017\left(\frac{P_L - P_{th}}{P_{th}}\right) F_o \frac{Q_s+1}{Q_o}(1-\cos\pi Q_o) \quad m^3/s^2 \qquad \textit{for males} \qquad (29)$$

For females, the constant 0.00017 is reduced to half its value. Otherwise, the formula is the same.

With the use of these empirical equations, the reference power $\bar{\rho}_o$ in equation (19) can be computed as a function of lung pressure,

$$\bar{\rho}_o = \frac{\rho}{6\pi c}(.00017)^2\left(\frac{P_L - P_{th}}{P_{th}}\right)^2 F_o^2 \frac{(Q_s+1)^2}{Q_o Q_s}(1-\cos\pi Q_o)^2 \qquad \textit{for males} \qquad (30)$$

Again, for females the constant 0.00017 is divided by 2. This keeps the reference power roughly the same for males and females since F_o is scaled by a factor of 1.7 (Titze, 1989b). Note that this power would increase with the square of (P_L-P_{th}) if F_o, Q_o, and Q_s were all held constant. It would then yield 6 dB of intensity increase per doubling of the pressure difference (P_L-P_{th}). It would yield a 9.5 dB increase if absolute pressure P_L were doubled from $2P_{th}$ to $4P_{th}$. This 9.5 dB per doubling is a value reported repeatedly in the literature (e.g., Ladefoged & McKinney, 1963; Isshiki, 1964; Iwata, 1988).

The factor $(Q_s+1)^2(1-\cos\pi Q_o)^2/(Q_oQ_s)$ changes little with lung pressure. For example, in going from 0.4 kPa ($2P_{th}$) to 0.8 kPa ($4P_{th}$), this factor decreases from 14.8 to 10.9 (1.3 dB) for males, giving a total of 8.6 dB increase in intensity for this doubling of pressure. For females, the factor changes from 15.6 to 11.6 over the same pressure change, a 1.2 dB decrease. Thus, waveform changes have relatively little effect on the low frequency reference power. $\bar{\varrho}_o$ is regulated primarily by the amplitude of the flow pulse, which changes with lung pressure.

The threshold pressures vary with F_o according to the relations,

$$P_{th} = 0.14 + 0.060\left(\frac{F_o}{120}\right)^2 \qquad \textit{for males} \tag{31a}$$

$$P_{th} = 0.14 + 0.060\left(\frac{F_o}{190}\right)^2 \qquad \textit{for females} \tag{31b}$$

These relations were developed from measurements by Gramming (1988) and Verdolini-Marston *et al.* (1990), and will be discussed later in conjunction with physiologic constraints on the VRP.

A final empirical relation for spectral slope was developed by matching SPL values reported by Holmberg *et al.* to those computed from equation (21) over a range of available pressures. The results were the same for males and females:

$$\textit{Spectral slope} = 11.0 + 5.0P_{th}/(P_L-P_{th}) \quad \textit{dB/octave} \quad . \tag{32}$$

This yields a range of 21.0 dB/octave to 12.7 dB/octave as lung pressure is varied from $1.5P_{th}$ to $4P_{th}$. As will be seen later, a three to four-fold increase in lung pressure from soft to loud phonation is usually an upper limit in phonation.

D. Physiologic constraints on lung pressure

Constraints on lung pressure are difficult to determine, especially for the "loudest" condition of phonation. Bouhuys *et al.* (1968) reported pressures as high as 6 kPa (about 60 cm H_2O) in one singer at frequencies around 400 Hz, but more typical maximum pressures in loud phonation are

between 1.0 kPa and 2.5 kPa when data of ten or more subjects are averaged at various frequencies (Kunze, 1962; Shipp & McGlone, 1971; Schutte, 1980; Gramming, 1988). Pressures increase consistently with increasing F_o for every subject. The largest pooled range for subjects was reported by Schutte, who had one normal subject out of 45 phonate with a pressure as low as 0.15 kPa in soft phonation and another with a pressure as high as 10.0 kPa in loud phonation. Individual pressure ranges were typically about 0.2-2.5 kPa from maximally soft to maximally loud, however.

The question arises, what constrains the subglottal pressure in phonation? Why is the range not simply zero to the maximum physiologic pressure a person can produce, and why does it change with F_o? First of all, there is the threshold effect mentioned earlier. A nonzero pressure is required to initiate phonation. This phonation threshold pressure has been shown to vary with mechanical tissue properties of the vocal folds, and with glottal geometry (Titze, 1988b). Specifically, P_{th} increased with increasing glottal convergence, with increasing mucosal wave velocity, with increased damping, with decreasing vocal tract inertance, and with decreasing vocal fold thickness. Threshold pressures on the order of 0.1-1.0 kPa were predicted, depending on these parameters. It was not possible, however, to predict the variation of P_{th} with changing vocal fold length (or, alternatively, with changing fundamental frequency). Generally, the mechanical driving impedance of the tissue increases as vocal fold stiffness increases, which is likely to account for the larger pressure requirements at greater lengths. In a vibrating string, for example, transverse driving impedance is inversely proportional to longitudinal tension (Kinsler & Frey, 1950).

Three experimental studies are available to quantify P_{th} with changing length or F_o. In a study by Finkelhor *et al.* (1987), direct measurements of P_{th} were made as a function of vocal fold elongation on excised canine larynges. Threshold pressures ranged from about 0.2 kPa to 1.5 kPa with increasing length, but the results varied considerably across four larynges.

In two additional studies, P_{th} was measured on human subjects as a function of F_o. In the Gramming (1988) study, subglottal pressure was estimated from oral pressure as 10 male and 10 female subjects phonated at low intensity. Threshold pressures ranged from about 0.3 kPa to 0.6 kPa over 10% to 80% of the F_o range. In another study by Verdolini-Marston *et al.* (1990), subjects phonated at 10%, 50%, and 80% of their F_o range. With the same measurement technique, results were very similar to the Gramming results. The combination of the two studies led to the empirical equations (31a,b).

The second important P_L constraint, maximum pressure in phonation, is theoretically more difficult to predict and practically more difficult to measure. The maximum lung pressure that a subject is willing to use in phonation probably depends on (1) a sense that phonatory instability will result, (2) a sense that vocal abuse may result, and (3) the appropriateness for the environment and condition of the speaker. Probably the most important of these is phonatory instability. Adult subjects are unwilling to produce screams that sound harsh, like a baby's intense cry. In life-threatening situations, they may use maximum physiologic pressure, but rarely in a clinical or conversational setting. Hence, the "steady" phonations elicited in diagnostic procedures usually imply relatively noise-free, unperturbed vocal fold vibrations for most subjects. We will therefore label the pressure for "loudest" phonation the phonatory instability pressure (P_i). In a study by Pabon & Plomp (1988), some evidence is seen that phonatory instability increases near the upper boundary of the VRP. They included voice perturbation as a third parameter in their displays.

The source model for the VRP is now complete if constraints on lung pressure are specified. In the simulations of the VRP to follow, P_L will always be kept above P_{th} and P_i will be varied as a parameter.

III. Characteristics of the Voice Range Profile

A first attempt at a male VRP model is shown in Figure 8(a). The open circles are averaged data from Gramming (1988) and the small dots (stipples) are calculations for various combinations of P_L and F_o in the foregoing equations. With regard to vocal tract parameters, the vowel was a simulated /ɑ/ with formant frequencies as given earlier. The resonant Qs for all of the formants were adjusted to be

$$Q = 3\frac{P_L}{P_{th}} \qquad (33)$$

which gives a value of 3 at threshold (where there is much glottal leakage) and 12 at $P_L = 4P_{th}$ (where there is little glottal leakage). The corresponding first formant bandwidths, for example, varied from 243 Hz for the high leakage case (at the bottom of the VRP) to 61 Hz for the low leakage case (at the top of the VRP). Since no data on variable bandwidth with changing intensity were available, this was at best an intelligent guess. But it does not carry with it a serious consequence. Bandwidth only affects the ripple in the VRP, not the average SPL values. Note that in the present model, there is no ripple at the bottom and a large amount of ripple at the top. This extreme variation was chosen to make the point of a possible bandwidth change with vocal loudness. A lesser variation in bandwidth would simply make top and bottom boundaries appear more similar (as in Figure 7, for example, where the bandwidth was not varied between the two curves).

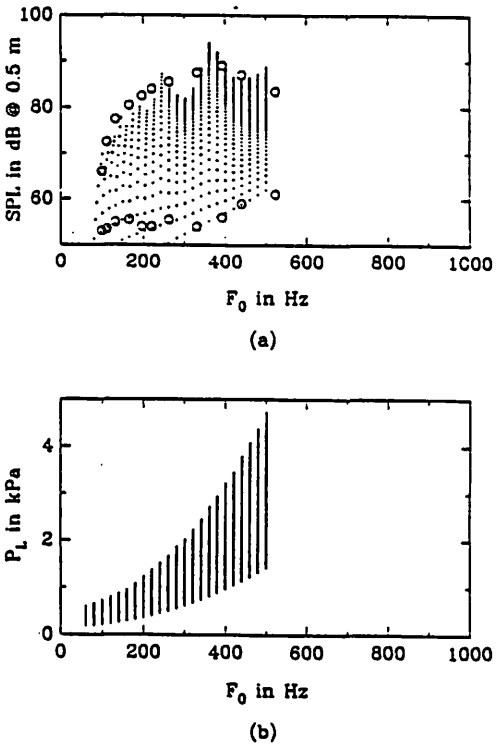

Figure 8. (a) Male VRP for an /ɑ/ vowel as produced by the model. Data circles correspond to averaged values for 10 males by Gramming (1988). (b) Range of lung pressure vs. F_o from which the model was computed. Stipples represent 0.05 kPa increments in lung pressure and correspond to stipples in VRP above.

The vertical distances between stipples represent 0.05 kPa increments in P_L and the horizontal distances are 20 Hz increments in F_o. This mapping can be related to Figure 8(b), where lung pressure is shown as a function of F_o. The two graphs together allow for direct comparisons between changes in P_L and SPL. Logarithmic increments of the stipples in Figure 8(a) correspond to linear increments in P_L in Figure 8(b). Note the large spacing between increments near the bottom of the VRP (near threshold). Between 100 and 300 Hz, for example, an increment in P_L of 0.05 kPa (½ cm H_2O) can give rise to as much as 5 dB variation in SPL. At higher frequencies and higher intensities, the changes are much smaller and asymptote to zero. This demonstrates that intensity range is likely to be gained or lost primarily at the low end (near threshold) of the VRP.

With regard to the data circles in Figure 8(a), ten male subjects in the Gramming (1988) study produced average SPLs as shown for highest and lowest intensities. (Gramming's SPL values

had to be lowered by 4.4 dB because the mouth-to-microphone distance in her study was 0.3 m instead of 0.5 m). The fact that both size and shape of the VRP model agree, for the most part, with those of the ten subjects gives some credibility to the model. Agreement in the formant ripple cannot be expected, of course, because such ripple would be washed out by averaging over the subjects' different formant frequencies.

Of particular interest is the downturn in SPLs of the upper boundary above 400 Hz. The model does not follow the human data precisely because P_i was chosen to be too large at high F_o. The simple relation $P_i=4P_{th}$ was used for the upper boundary, which forced lung pressures to be above 4.0 kPa in the high F_o region (see Figure 8b). Gramming's estimates of P_i were less (about 1.5 to 2.0 kPa), which would account for the downturn in the data points. Some further experimentation with P_i will be discussed later.

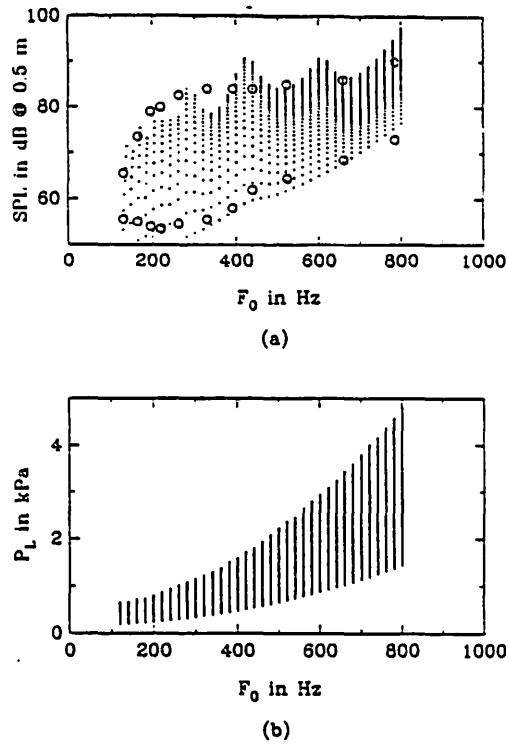

Figure 9. (a) Female VRP for an /ɑ/ vowel as produced by the model. Data circles correspond to averaged values for 10 females by Gramming (1988). (b) Range of lung pressure vs. F_o from which the model was computed. Stipples represent 0.05 kPa increments in lung pressure and correspond to stipples in VRP above.

Figure 9 shows measured and modeled VRPs for females. Circles are again data averaged over 10 subjects by Gramming (1988). All of the gender differences described in the previous sections were incorporated, and the F_o range was stretched to 800 Hz. The first three formant frequencies were changed to 850 Hz, 1220 Hz, and 2810 Hz (Peterson & Barney, 1952), all other formants remaining the same. Note that the intensities at high F_o in Figure 9(a) are again overestimated somewhat due to the steep rise in the upper boundary lung pressure in Figure 9(b).

In Figure 10, a different upper boundary on P_L has been selected for comparison with Figure 9. Rather than choosing $P_i=4P_{th}$, a quadratic trend, the maximum lung pressure has been chosen to vary linearly with F_o. Note that this brings up the maximum intensity in the midrange frequencies, but decreases it slightly at the high end. The match with the data is a little better, suggesting that a linear trend is perhaps a little more realistic. Sundberg *et al.* (in press) found a nearly linear relation between F_o and P_L in singers when all loudness conditions were averaged, which gives some clue that P_i perhaps does not rise as rapidly as P_{th}. As stated before, more work is needed to quantify the nature of the P_i-F_o relation for various subjects.

In Figure 11, the effect of the vocal tract has been eliminated. In other words, the high frequency gain in equation (22) has been set to 1 in order to view the source power variation without the effect of the vocal tract. This would be the situation if the larynx were to radiate into free-space, as in the Titze (1988a) study involving excised larynges on a laboratory bench. Ranges of P_L in

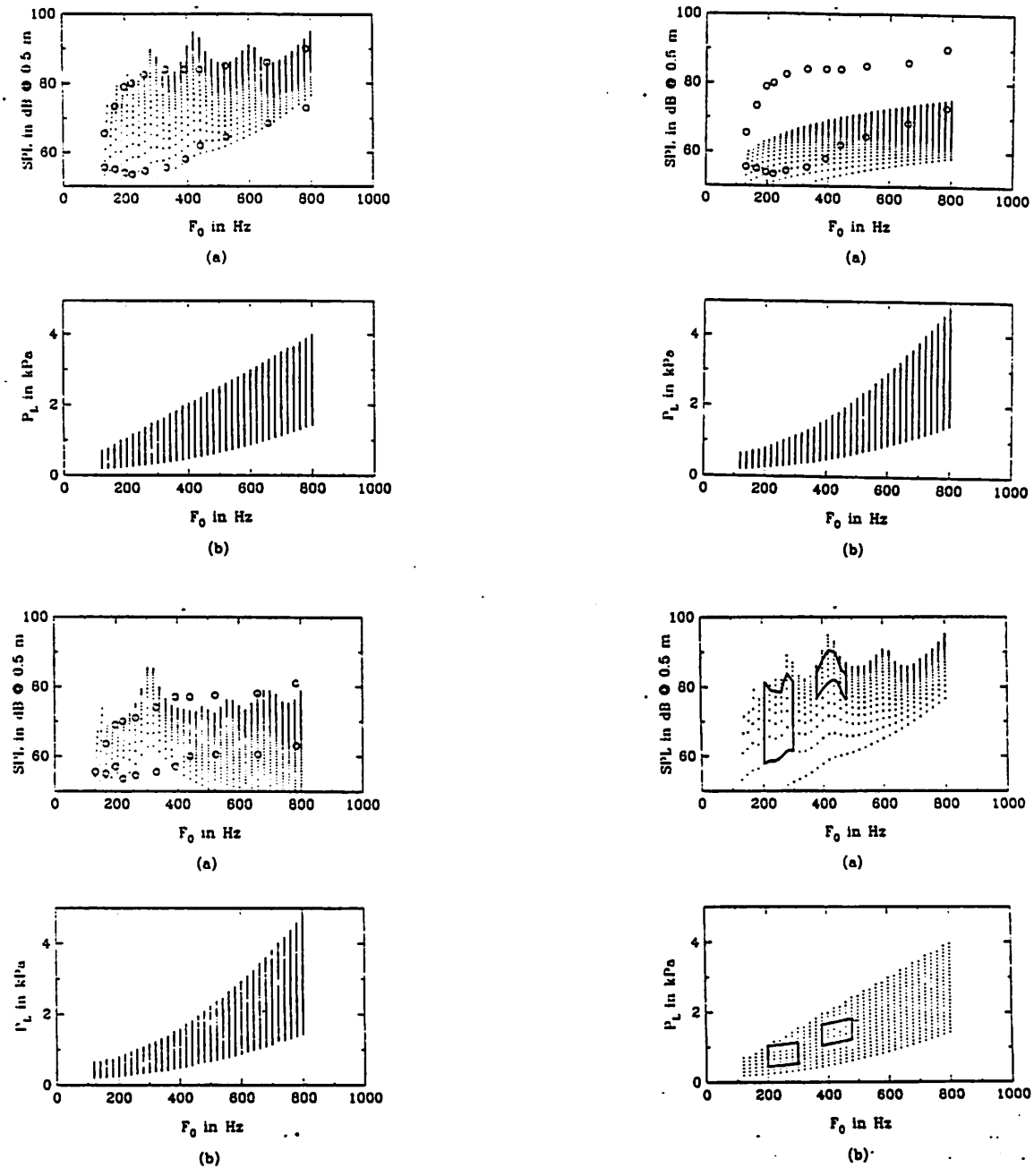

Figure 10. (upper left) (a) Female VRP for an /ɑ/ vowel as produced by the model. The phonation instability pressure has been changed to raise SPL at the center and lower it at high F_o. (b) Range of lung pressure vs. F_o from which the model was computed. The upper boundary has been linearized as shown. Stipples represent 0.05 kPa increments in lung pressure and correspond to stipples in VRP above. Figure 11. (upper right) (a) Female VRP as produced by the model, but with the vocal tract disconnected. Data circles are kept the same as in Figures 9 and 10 for comparison. (b) Range of lung pressure vs. F_o from which the model was computed. Stipples represent 0.05 kPa increments in lung pressure and correspond to stipples in VRP above. Figure 12. (lower left) (a) Female VRP for an /i/ vowel as produced by the model. Data circles correspond to averaged values for 10 females by Gramming (1988). (b) Range of lung pressure vs. F_o from which the model was computed. Stipples represent 0.05 kPa increments in lung pressure and correspond to stipples in VRP above. Figure 13. (lower right) (a) Female VRP for an /ɑ/ vowel as produced by the model. The two closed regions correspond to the two polygons shown below. (b) Range of lung pressure vs. F_o from which the model was computed. The polygons represent paths that might be taken in P_L-F_o space as pitch and loudness are changed.

Figure 11(b) are identical to those shown in Figure 9(b), and data points for the female VRP have been retained for comparison with former cases. Note that the range of intensity in Figure 11(a) is greatly reduced (by 15-20 dB in the center of the VRP) in relation to that of Figure 9(a). The top boundary is flat and the threshold boundary is pulled below 60 dB for all frequencies. Note also that the increments in SPL are more evenly spaced. Since all parameters other than $G(F_o)$ in equation (22) have been kept constant, we conclude that the vocal tract provides an average gain of 10-15 dB over the no vocal tract case. The difference is greater at the high end of the F_o range than the low end, suggesting that $G(F_o)$ indeed asymptotes to 1 as F_o approaches zero.

In Figure 12(a), the formant frequencies have been changed to correspond to a female /i/ vowel (formant frequencies 310, 2790, 3310, 3570, 4400, 5500, 6500, 7500, 8500, and 9500 Hz). Resonant Qs remained the same as for /ɑ/. Data points are again after Gramming (1988), based on an average over 10 female subjects. Note the peaking of SPL for the model at 155 Hz and 310 Hz. This is where the second harmonic and the first harmonic, respectively, line up with the first formant, giving a strong boost of output power. This ripple is apparently washed out in the averaging process in the female subject data, as in the former case of the /ɑ/ vowel.

Above 350 Hz, the maximum intensity does not rise much with F_o, even though the modeled lung pressure rises dramatically (Figure 12b). This is a direct result of the fact that F_o has exceeded F_1. The vocal tract provides little advantage over the no vocal tract case when this condition occurs. Threshold intensities of the model drift downward, in fact, while P_{th} is rising. This result is not seen in the averaged subject data of Gramming. It does agree, however, with the single subject data reported by the same author (Figure 1b). This discrepancy in the subject data is not clear at this point.

Overall, predicted SPL values are slightly lower than measured ones for the vowel /i/. At least two explanations can be offered: (1) It takes more lung pressure to produce /i/ at fundamental frequencies above F_1, or (2) the model is in error above F_1, possibly due to wrong assumptions about the spectral slope or the vocal tract transfer characteristics.

Returning now to the more commonly used /ɑ/ vowel, a mapping between the P_L-F_o domain and the SPL-F_o domain can be demonstrated with some windows, as shown in Figure 13. Mapping the parallelograms in Figure 13(b) into the VRP domain of Figure 13(a) yields some interesting results. First, note that constant (or nearly constant) P_L lines do not produce constant SPL lines as F_o is changed. Vocal tract resonance is responsible for this distortion. It would appear to be difficult for a subject to maintain constant intensity while executing a musical scale or a pitch glide. This has been experienced (informally) by subjects trying to hold the needle of an SPL meter constant while changing pitch. Intensity fluctuations on the order of ± 3 dB occur involuntarily. This is readily understandable on the basis of Figure 13(a), particularly since the patterns of fluctuation change across the F_o range (first mapping versus the second mapping).

Changing intensity at constant F_o, on the other hand, is much easier, since the vertical lines of the closed figures are not distorted. This has also been borne out (informally) by having subjects execute crescendos at constant pitch. It is generally not too difficult to hold F_o constant while SPL is increased or decreased. An exception is the low F_o region. When frequencies are less than about 100 Hz, vertical lines in the P_L-F_o space do not map into vertical lines in the SPL-F_o space. This can be demonstrated by considering an expanded view of the male VRP of Figure 8 in the 80-200 Hz region, shown in Figure 14. In this low frequency region, there is evidence of a strong F_o-P_L dependence. This dependence has been modeled quantitatively by Titze (1989a) on excised canine larynges. Although the F_o-P_L dependence was expressed in terms of vocal fold length, it is readily

verified that the expression

$$F_o = F_\infty (1 + e^{1.59 - .0246 F_\infty P_L}) \quad (34)$$

matches the data very well, where F_∞ is an asymptotic fundamental frequency (approached when P_L tends to zero). In Figure 14(b), the vertical arrays of stipples are located at F_∞, whereas in Figure 14(a), the "bended" F_o lines are calculated and plotted for each SPL. It is seen that F_o increases substantially with lung pressure, at rates of 30-100 Hz per kPa (3-10 Hz per cm H_2O), depending on the mean F_o. This may explain why it is difficult to execute a crescendo at very low pitches. In order to hold F_o constant, it is likely that vocal fold length (i.e., cricothyroid muscle activity) must be decreased as P_L is raised. There may be a limit for such a length reduction (e.g., zero cricothyroid activity). For this reason, the upper boundary of the VRP seems to fall off sharply as F_o approaches its lower limit.

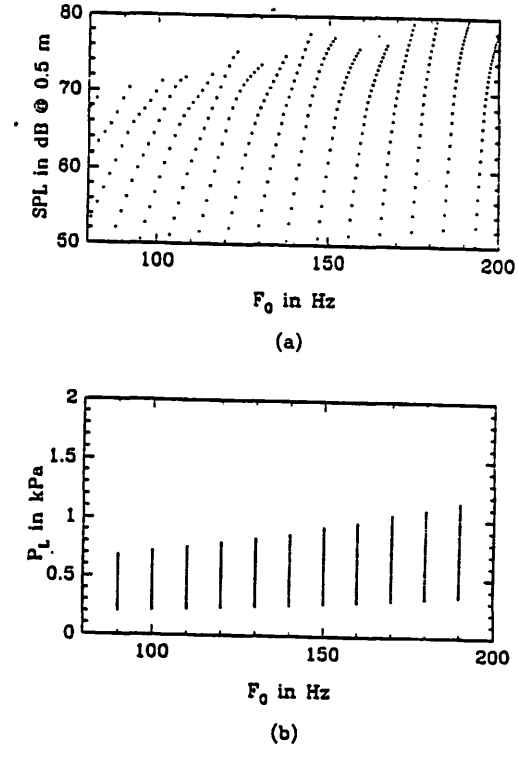

Figure 14. VRP of Figure 8 expanded in the low frequency region to show the F_o-P_L dependency.

Conclusions

Some interpretations of the Voice Range Profile have been offered by formal investigation of the co-variation of intensity (I) and fundamental frequency (F_o). The average I-F_o slope (6-10 dB/octave) can be attributed to two phenomena; (1) that the vocal tract transmits power more effectively as F_o is raised to meet F_1, and (2) that lung pressure is necessarily increased with F_o to stay above phonation threshold pressure.

The vocal tract imposes an intensity fluctuation with changing F_o that is explainable on the basis of formant tuning of selective harmonics. This creates a ripple effect at the boundaries, but the ripple may be less prominent at the lower intensity boundary because of increased formant damping (when voicing is soft, falsetto-like, or breathy).

Constraints on lung pressure limit the overall I range, but additional high and low F_o limitations on the I range may be attributed to more subtle phenomena. A restricted intensity range at low F_o may result from an inherent inability to keep the vocal fold tension low when lung pressure is raised to obtain greater amplitudes of vocal fold vibration. A restricted I range at high F_o may result from a reduced benefit obtained from the vocal tract for a spectrally rich source. The

vocal tract favors the fundamental rather than the harmonics at high F_o, making spectrally weaker productions (head voice or mixed voice) nearly as intense as chest voice (modal) productions.

Some of the empirical relations that were used to express the dependence of glottal waveform shape parameters on lung pressure are in need of more refinement. In particular, the ways in which waveform skewing, open quotient, and spectral slope vary with lung pressure and fundamental frequency need to be investigated with a greater variety of subjects. Voice Range Profiles can then be interpreted with greater confidence. Future applications of this model should be directed towards a better understanding of vocal registers, voice classification, and voice disorders.

Acknowledgement

This work was supported by Grant No. DC 00387-03 from the National Institute of Health. The author appreciates the kind assistance of Linnie Southard in manuscript preparation.

References

Ananthapadmanabha, T. (1984). "Acoustic analysis of voice source dynamics." Speech Transmis sion Laboratory. Quarterly Progress and Status Report, 3/1984, 1-52. Stockholm, Sweden: KTH.

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

Coleman, R.F., Mabis, J.H., & Hinson, J.K. (1977). "Fundamental frequency-sound pressure level profiles of adult male and female voices." Journal of Speech and Hearing Research, 20, 197-204.

Damsté, P.H. (1970). "The phonetogram." Practica Otology-Rhinology-Laryngologica, 32, 185-187.

Fant, G. (1960). *Acoustic theory of speech production.* (The Hague, The Netherlands: Mouton).

Fant, G., & Lin, Q. (1987). "Glottal voice source-vocal tract acoustic interaction." Speech Trans mission Laboratory. Quarterly Progress and Status Report, 1/1987, 13-27. Stockholm, Sweden: KTH.

Fant, G., & Lin, Q. (1988). "Frequency domain interpretation and derivation of glottal flow param eters." Speech Transmission Laboratory. Quarterly Progress and Status Report, 2-3/1988, 1-21. Stockholm, Sweden: KTH.

Finkelhor, B., Titze, I., & Durham, P. (1987). "The effect of viscosity changes in the vocal folds on the range of oscillation." Journal of Voice, 1(4), 320-325.

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

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

Gramming, P. (1988). *The phonetogram: An experimental and clinical study.* (Malmö, Sweden: Department of Otolaryngology, University of Lund).

Gramming, P. & Sundberg, J. (1989). "Spectrum factors relevant to phonetogram measurement." Journal of the Acoustical Society of America.

Holmberg, E., Hillman, R., & Perkell, J. (1988). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice." Journal of the Acoustical Society of America, 84(2), 511-529.

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

Iwata, S. (**1988**). "Aerodynamic aspects of phonation in normal and pathologic larynges." In ?, edited by O. Fujimura, (Raven Press, New York), pp. 423-431.

Kinsler, L., & Frey, A. (**1950**). *Fundamentals of acoustics.* (Wiley, New York)

Klingholz, F., & Martin, F. (**1983**). "Die quantitative Auswertung der Stimmfeldmessung." Sprache-Stimme-Gehör, 7, 106-110.

Klingholz, F., Martin, F., & Jolk, A. (**1986**). "Das Dreidimensionale Stimmfeld." Laryngologie, Rhinologie, Otologie, 65(10), 588-591.

Klingholz, F. (**1990**). *Das Stimmfeld.* (Verlag J. Peperny, München, Germany).

Komiyama, S. (**1972**). "Phonogram - A new method for evaluating voice characteristics." Otologia (Fukuoka), 18, 428-440.

Kunze, L. (**1962**). "An investigation of changes in subglottic air pressure and rate of airflow accom panying changes in fundamental frequency, intensity, vowels, and voice register in adult male speakers." Unpublished Doctoral Dissertation, University of Iowa, Iowa City, Iowa.

Ladefoged, P., & McKinney, N. (**1963**). "Loudness, sound pressure, and subglottal pressure in speech." Journal of the Acoustical Society of America, 35, 454-460.

Morse, P. (**1948**). *Vibration and sound.* New York: Acoustical Society of America.

Pabon, J.P.H., & Plomp, R. (**1988**). "Automatic phonetogram recording supplemented with acoustic voice-quality parameters." Journal of Speech and Hearing Research, 31(4), 710-722.

Peterson, G. & Barney, H. (**1952**). "Control methods used in the study of vowels." Journal of the Acoustical Society of America, 24, 175-184.

Schutte, H.K. (**1980**). *The efficiency of voice production.* (Groningen Druk, The Netherlands).

Schutte, H.K., & Seidner, W. (**1983**). "Recommendation by the Union of European Phoniatricians (UEP): Standardizing voice area measurement/phonetography." Folia Phoniatrica, 35, 286-288.

Seidner, W., Krüger, H., & Wernecke, K.D. (**1985**). "Numerische Auswertung spektraler Stimmfelder." Sprache-Stimme-Gehör, 9, 10-13.

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

Sundberg, J. (**1987**). *The science of the singing voice.* (Northern Illinois University Press, DeKalb, Illinois).

Sundberg, J., & Gauffin, J. (**1979**). "Waveform and spectrum of the glottal voice source." In *Frontiers of speech communication research: Festschrift for Gunnar Fant* , edited by B. Lindblom & S. Ohman. (Academic Press, London). pp. 301-320.

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

Titze, I.R. (**in press**). "Phonation threshold pressure: A missing link in glottal aerodynamics." Journal of the Acoustical Society of America.

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

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

Titze, I.R. (**1989a**). "On the relation between subglottal pressure and fundamental frequency in phonation. Journal of the Acoustical Society of America, 85(2), 901-906.

Titze, I.R. (**1989b**). "Physiologic and acoustic differences between male and female voices." Jour nal of the Acoustical Society of America, 85(4), 1699-1707.

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

Footnotes

[1] The term *lung pressure* is used here for brevity instead of mean subglottal pressure. The difference is negligible for any consideration here.

NCVS Status and Progress Report - 1
June 1991, 39-48

Elastic Models Of Vocal Fold Tissues

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

Abstract

Elastic properties of canine vocal fold tissue (muscle and mucosa) were obtained through a series of experiments conducted *in vitro* and were modeled mathematically. The elastic properties play a significant role in quantitative analysis of vocal-fold vibrations and theory of pitch control. Samples of vocalis muscle and mucosa were dissected and prepared from dog larynges a few minutes premortem and kept in a Krebs solution at a temperature of 37 ± 1 °C and a pH of 7.4 ± 0.05. Samples of muscle tissue and mucosa were stretched and released in a slow, sinusoidal fashion. Force and displacement of the samples were measured with a dual-servo system (ergometer). After digitization, stress-strain data for samples of muscle tissue and cover tissue were averaged. The stress-strain data were then fitted numerically by polynomial and exponential models.

Nomenclature

A, B	Exponential model parameters
A_0	Cross-sectional area of specimen at its reference state
E, E_b, E_c	Young's moduli, kPa
F	Longitudinal force
L	Sample length
L_0	Reference length
m	Sample mass
ε	Strain
ρ	Density
σ	Lagrangian stress (kPa)

Introduction

This paper will report some data sets on vocal fold tissue elasticity and present a few modeling applications for them. Although similar data have been reported previously, quantitative analysis and modeling were not performed completely in the past. Methodologies for obtaining stress-strain curves now have been refined considerably, and order-of-magnitude calculations have been replaced by more precise constitutive equations.

Elasticity of vocal fold tissues is not only a key factor in the control of fundamental frequency in phonation (Titze and Talkin, 1979; Titze, Jiang and Druker, 1988), but also a major determinant of some abnormalities in the voice, such as breathiness and hoarseness (Isshiki, Ohkawa, and Goto, 1985). Although vocal fold tissues are viscoelastic, thereby demonstrating hysteresis effects and time-dependent stress-strain (Alipour-Haghighi & Titze, 1985, 1990), purely elastic behavior is approximated at a small rate of loading, where viscous effects are negligible (Christensen, 1971). This fact is demonstrated by Muller, Milenkovic and MacLeod (1985) on perioral tissue through a frequency response analysis of passive perioral tissue due to cyclic loading. Their study showed that the perioral tissue stiffness remained constant below 10 Hz and the tissue exhibited elastic-like behavior. In a cyclic process, as in vibration, the stress-strain loop becomes repeatable and predictable after several cycles and this repeatable state is called pseudo-elasticity (Fung, 1979). Stiffness of the tissues may, therefore, be approximated sufficiently for steady phonation from the cyclic stress-strain relationships at very low frequencies (one Hz or less).

Anatomically, the vocal folds consist of a layered structure; each layer exhibits a different mechanical property (Hirano, 1977). This structure seems to be responsible for the wide range of phonation frequencies and qualities that can be produced. In this study, we have followed Hirano's suggestion of simplifying the vocal fold morphology to a two-layer system, a muscular layer (body) and a non-muscular layer (cover). The superficial cover consists of epithelium (Hirano, et al. 1982), and the vocal fold body consists of the vocalis muscle. Since the cover assumes most of the vibration near the surface (Saito et al., 1985), knowledge of its elastic properties is of primary importance. Nevertheless, the elastic properties of both layers affect vocal fold vibration (Titze and Talkin, 1979). The vocal ligament, which assumes a major portion of stiffness in human vocal folds (Van den Berg, 1959), does not exist in the canine vocal fold and was not studied here.

In the last ten years, considerable attention has been given to the measurement of vocal fold elasticity. Kakita, Hirano and Ohmaru (1981) and Hirano, Kakita, Ohmaru, and Kurita (1982) investigated mechanical properties of vocal fold tissues and reported ranges and orders of magnitude for Young's moduli and shear moduli of different layers of canine vocal fold tissue. Their procedure consisted of hanging various weights on tissue samples and measuring elongations with a microscope. Although their results provided some of the first quantitative estimates of elastic moduli, there were some methodological difficulties. The reference length was not defined in their studies, and the viability of their tissue was questionable since it was excised post mortem.

Perlman, Titze and Cooper (1984) and Perlman (1985) investigated elastic properties of similar vocal fold tissues in viable conditions. By applying a stepwise elongation to the samples and measuring the force and elongation, a secant Young's modulus for both the body and the cover was reported. The secant Young's modulus, which is the ratio of total increase in stress to total increase in strain from a reference length, would be identical to the tangent Young's modulus for linearly elastic materials. When the material is nonlinear (such as in vocal fold tissues), differences will exist between the two Young's moduli, especially when the reference lengths are not the same.

Perlman et al. (1984) used a so-called physiological length (the length where a muscle generates its maximum active force) as the reference length (zero strain). The investigators began the stepwise elongation process at this reference point and collected passive stress-strain data for the body samples. Since the physiological length is likely to be above the typical operating length of muscles used for length control (Rack and Westbury, 1969; Finocchio and Luschei, 1985), the data were limited to large strains.

Perlman and Titze (1985) investigated the elastic properties of vocal fold muscle tissue using the *in situ* length as a reference length. The method of elongation was stepwise, which caused the number of data points to be small (less than seven) in the elongation process. They reported stress-strain data for vocal fold muscle with age and breed as control parameters. They also compared their results to the conditions where physiological length was used as a reference length and concluded that the choice of *in situ* length gave less variability in their results.

Perlman and Durham (1987) investigated stress-strain data for canine vocal fold cover. They used thin (approximately 1 mm) and thick (approximately 2 mm) cover samples and reported stress-strain data with animal sex as the control parameter. Their reference length was based on one gram of force, which gave a similar reference length to this study. The method of elongation and the definition of Young's modulus remained the same, however.

The data reported by Perlman et al. (1984, 1985, 1987) from our laboratory on the vocal fold stress-strain were detailed, but not ideal for the purpose of mathematical modeling. The sparsity of data points and the lack of low-strain data forced us to repeat some of the experiments with a different methodology.

In the present study, the reference length was defined as an average of fiber lengths measured *in situ*. Muscle samples were not stimulated for more than one single twitch (to check viability). Each sample was mounted in the chamber at the average *in situ* length, which was the defined reference length. Instead of stepwise elongation, a slow cyclic stretch and release paradigm at one Hz was employed to elongate the samples and collect data. Thus, not only was a smoother curve obtained, but the tissue had a shorter overall stretch time. Another advantage of the present study was that more than one sample from the same viable larynx (e.g., the left and right vocal fold) could be used in the same experiment.

Experimental Method

To obtain elastic properties of vocal fold tissues, ten samples of vocal fold muscle and nine samples of vocal fold cover were dissected and prepared from viable, excised canine larynges in a manner described by Perlman (1985) and kept in Krebs solution aerated with 95% oxygen and 5% carbon dioxide to preserve viability. A piece of cartilage was left at both ends of each sample for mounting and application of the force. Since values of strain were expected to be dependent upon reference length, it was very important to define and measure this length in a consistent manner to reduce errors. Before dissection, lengths of each sample were measured *in situ* using a caliper with an accuracy of 0.1 mm.

The muscle samples were approximately 3-5 mm in both of the transverse dimensions (width and thickness). Due to the angle between the thyroid and arytenoid cartilage surfaces, the muscle fiber lengths were not equal; thus the sample length was defined as an average of the maximum and minimum length of the fibers measured *in situ*. The length of muscle samples ranged between 15-19 mm. The cover samples had a more uniform cross section, with an approximate 0.5-1.5 mm thick-

ness and an approximate 4-7 mm width. The length was 13-17 mm. Using a Tevdek (3-0) suture through the attached cartilages (thyroid and arytenoid), the samples were mounted in a water-jacketed chamber (Fig. 1) containing an electrolyte solution, maintained at a temperature of 37° ± 1°C and pH of 7.4 ± 0.05 during the experiment. Force and elongation of the samples were controlled and measured with a dual-servo system (Cambridge Technology, model 305 ergometer). The ergometer had force resolution of 50 mg, displacement accuracy of 0.02 mm and rise time of 6 ms. Samples were mounted vertically in the water-jacketed chamber, with the initial length set as close to the previously measured *in situ* length (ranging between 15-19 mm) as could be set without any slackness in the suture. This length was defined as the reference length. Occasional slackness in the suture, which required an additional 0.5-1.0 mm elongation, may have caused errors on the low strain values of stress.

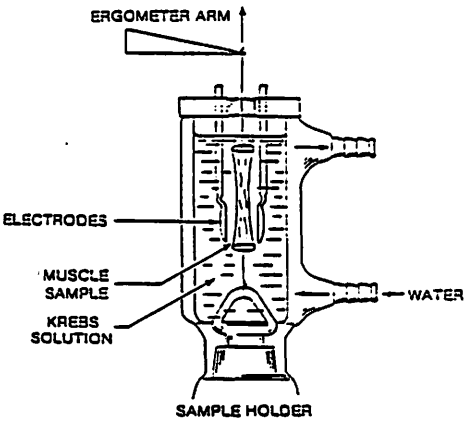

Fig. 1:Schematic of sample mounting. Warm water circulates in a jacket to keep temperature at 38 °C. Platinum electrodes are used for checking on the viability of muscle samples with field stimulation method.

Since measurement of the cross-sectional area of the samples in the chamber was impossible, an average cross-sectional area was calculated from the volume of the sample (without the attached cartilages) divided by the reference length (Kakita, et al 1981; Hirano et al, 1982). Perlman (1985) measured the densities of canine vocal fold tissues and found that they were almost constant (1.04 for the body and 1.02 for the cover). Using these densities, the volume was calculated from the measured mass, which was determined with a Mettler balance model AE 100 (0.1 mg resolution). The cross-sectional area was calculated from

$$A_0 = m / (\rho L_0) \quad (1)$$

where m is mass, ρ is density, and L_0 is reference length. The length and cross-sectional area were then used in calculations of stress and strain (from force and elongation measurements).

Force-elongation data were obtained by applying a sinusoidal signal to the ergometer, which resulted in stretch and release for each sample. The output force and position signals were sampled with a 12-bit A/D converter on a PDP 11/44 computer for digital recording of the data (at a sampling frequency of 1000 Hz per channel).

Since major length changes of vocal folds during phonation occur at about 10 Hz and below, the force-elongation data were collected in a frequency range of 0.1-10 Hz. For each sample, data were collected in descending frequency from 10 Hz to 0.1 Hz. Ten samples of muscle and nine samples of cover were used. Sutures transmitted the force of the ergometer to the sample during loading. During the return, however, there was no force transmitted and tissue shortened by its own elasticity. Thus, the force always stayed positive and only the loading portions of the data were used for mathematical analysis.

Results

Time series for force and elongation were plotted against each other to obtain force-elongation curves. These curves formed a closed loop, with the upper curve corresponding to stretch and the lower curve corresponding to release. Figure 2 shows typical force-elongation hysteresis curves

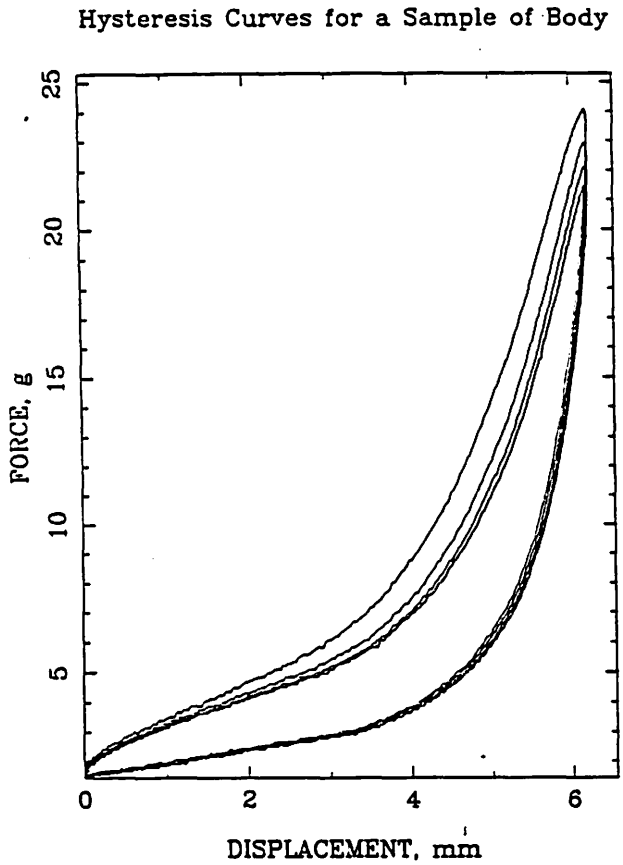

Fig. 2:Typical force-elongation hysteresis curves for a vocal fold muscle (body) sample at frequency of 0.5 Hz.

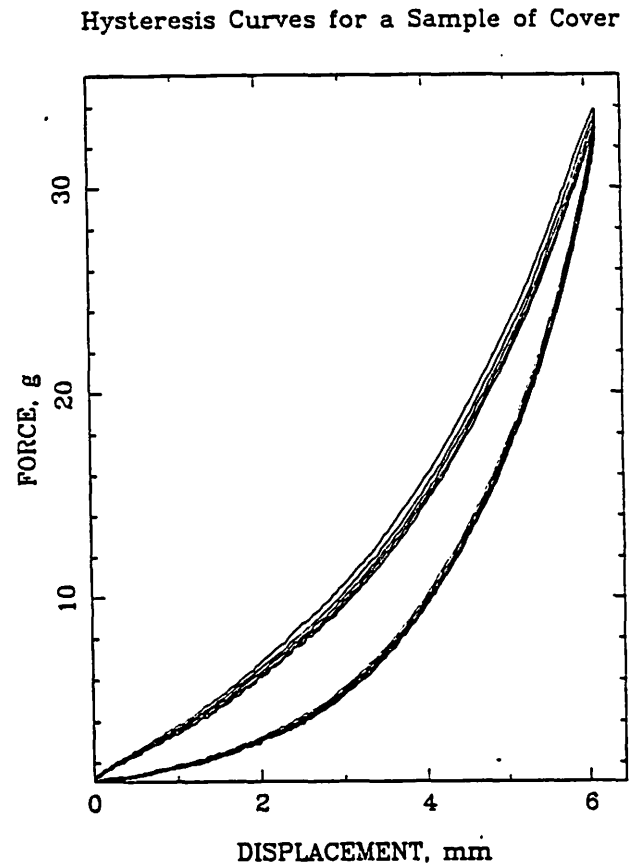

Fig. 3:Typical force-elongation hysteresis curves for a vocal fold cover sample at frequency of 0.5 Hz.

for the vocal-fold body (muscle) and Figure 3 shows similar curves for the vocal-fold cover at 0.5 Hz. Because of the relaxation properties of tissue, these force-elongation loops move downward for a few cycles at the beginning of the process. For a repeatable stress-strain relation, data were collected after five to ten cycles of repetition. Stress and strain were then obtained for the defined reference length and the calculated cross-sectional areas

$$\sigma = F / A_0 \tag{2}$$

and

$$\varepsilon = (L-L_0)/L_0 \tag{3}$$

where F is the (longitudinal) force applied to the sample, σ is the stress, L is the final sample length, and ε is the strain.

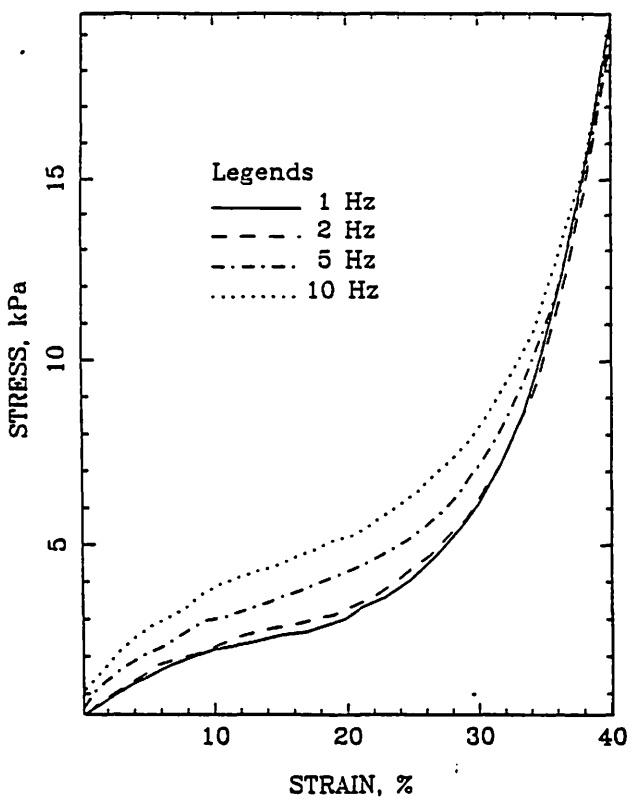

Fig. 4:Stress-strain curves of a sample of vocal fold muscle (body) stretched and released from high to low frequencies.

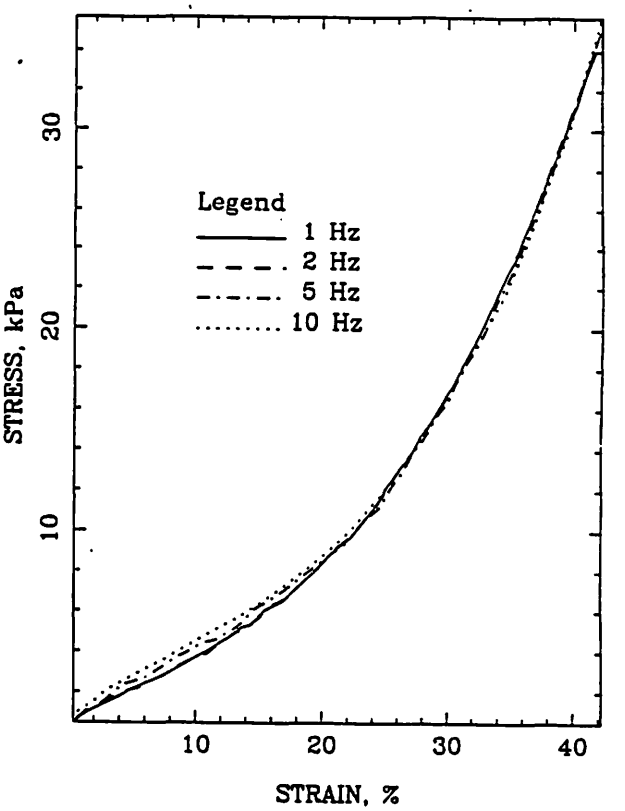

Fig. 5:Stress-strain curves of a sample of vocal fold cover tissue stretched and released from high to low frequencies.

Figure 4 shows stress-strain curves of a sample of muscle for various frequencies of stretch. Each curve represents the loading portion of hystresis curves of that frequency after few cycles (preconditioning). When the stretch-release frequency was decreased, lower values of stress were obtained, i.e. the stress-strain curves moved downward. However, below 1 Hz, there were no noticeable changes in these curves. These trends are due to viscoelastic properties of the vocalis muscle (Alipour-Haghighi & Titze, 1985). The force-elongation data for the vocal-fold cover (Fig. 5) was not affected as much by the frequency of stretch and release. This is probably due to the more elastic nature of the cover. This fact was also observed in former stress relaxation measurements of these tissues (Alipour-Haghighi & Titze, 1990), where body samples showed higher viscosities and shorter time constants than cover samples. This higher viscosity of the body samples was responsible for the greater time-dependence of the stress-strain curves.

Force-elongation data from 10 samples of vocal fold muscle and 9 samples of the vocal fold cover were averaged. These averaged stress-strain data (for 1 Hz stretch and release) are shown in Fig. 6, together with their standard deviations. The greater standard deviation for muscle tissue

could be attributed to irregularity of its shape. Nonlinear behavior of vocal fold tissue is clearly noticeable from this plot, particularly for the muscle.

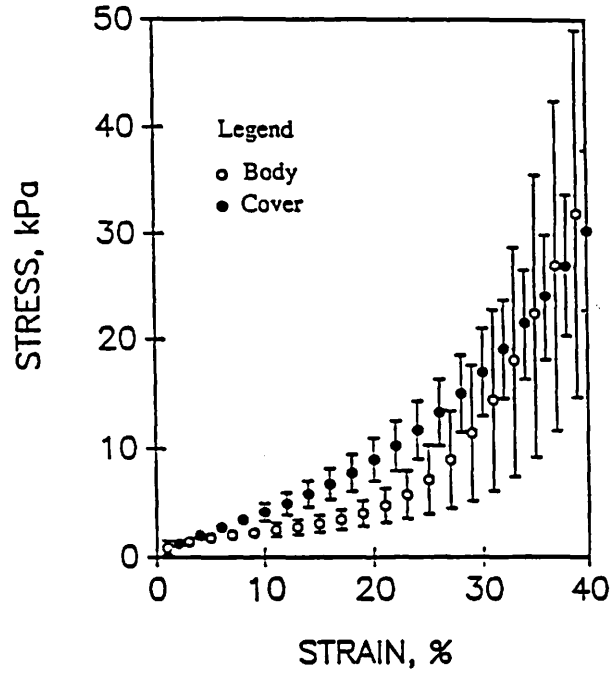

Fig. 6:Average stress-strain curves of vocal fold muscle (body) (n=10) and vocal fold cover (n=9) together with standard deviations.

Modeling

Polynomial model

The averaged data were modeled with a third order polynomial using a nonlinear least-square curve fitting algorithm. The stress-strain relation obtained for vocal-fold body tissue was

$$\sigma = 0.4 + 42.3\,\varepsilon - 341.9\,\varepsilon^2 + 1132\,\varepsilon^3, \qquad (4)$$

whereas the relation obtained for the vocal-fold cover was

$$\sigma = 0.5 + 38.3\,\varepsilon - 49.5\,\varepsilon^2 + 347.6\,\varepsilon^3 \qquad (5)$$

These expressions apply to the entire range of strain ($0 \leq \varepsilon \leq 40$ percent).

One particular case of a polynomial model is a linear model (Hook's law), with no residual stress at zero strain. This model is useful at low strains ($\varepsilon < 15\%$), where the stress-strain curves of both body and cover are fairly linear (Figure 6). Hook's law can be written

$$\sigma = E\,\varepsilon \qquad (6)$$

where E is Young's modulus for the tissue. Using a least-square optimization technique (Hamming 1973), individual data from zero to 15% strain were modeled with equation (6). Mean and standard deviations of these low strain Young's moduli for body and cover tissues are

$E_b = 20.7$ kPa $\pm$ 2.4 kPa (N=10)
$E_c = 41.9$ kPa $\pm$ 7.1 kPa (N=9).

Exponential model

While low strain portions of stress-strain curves behave nearly linearly, high strain portions appear more exponential. Exponential models have been used in the theory of pitch control (Fujisaki, Tatsumi, and Higuchi, 1981; Titze, 1989; Titze, Luschei and Hirano, 1989). A possible exponential model can be described as

$$\sigma = A(e^{B\varepsilon}-1). \qquad (7)$$

Individual data records were modeled with equation (7) using a nonlinear least-square optimization

technique (Hamming, 1973) and results were averaged. The best-fit parameters for equation (7) were

$$A = 0.89 \pm 0.55, B = 9.0 \pm 3.2 \text{ (for body)},$$

and

$$A = 7.1 \pm 2.6, \quad B = 4.2 \pm 0.73 \text{ (for cover)}.$$

The parameters for the body produces a model close to that of Titze et al. (1989), who used the equation $\sigma = 1.0e^{9.2\varepsilon}$ for mean passive stress of all combined tissues in vibration.

Discussion

Using a slow cyclic stretch and release technique, stress-strain data of canine vocal fold tissues were obtained in the range of 0-40% elongation over an approximate *in situ* length. These stress-strain data, which apply to longitudinal vocal fold tension, were modeled with polynomial and exponential functions. Elastic properties of body and cover tissues were reported graphically and through mathematical models. These elastic properties are nonlinear for both vocal fold muscle and cover tissues over most of the strain range, but some linear approximations at low strain can be made, especially for the body. At these lower strains, the cover has about twice the stiffness of the muscle, (as measured by the longitudinal Young's modulus). This may be significant in that the vocal folds operate mostly at low strains during phonation (Hollien, 1960). It has usually been thought that the cover is much less stiff than the body, even in this passive state where the muscle tissue of the body is uncontracted.

The relatively low Young's modulus in the non-activated muscle tissue of the vocal fold body can be attributed to the different proportions of collagen and elastin fibers in the layers. According to Hirano et al. (1982), there are more collagen fibers and fewer elastic fibers in the muscle than in the cover. Since the combination of collagen and elastin fibers influence the passive mechanical properties of these tissues, it might be possible to explain the difference between cover and body Young's moduli at low strain on the basis of this fiber concentration. Elastin fibers increase their stiffness more gradually at low strains (Oxlund, Manschot, and Viidik, 1988), whereas the wavy collagen fibers are first being straightened without exerting much restoring force. When larger loads are applied, however, the collagen fibers are already oriented in the direction of loading and resist with much greater stiffness (Nordin and Frankel, 1980). This would also explain the greater nonlinearity in the overall stress-strain curve of the body. The analogy between a rubber band and a cotton thread has been applied to elastin and collagen fibers, respectively.

The elastic properties of vocal fold tissues reported in this study are quantitatively different from those of Perlman et al. (1984), Perlman (1985), and Perlman and Durham (1987). The Young's moduli for body and cover reported by these investigators were 98 and 110 kPa, respectively, which are three to four times higher than our results. This discrepancy could be attributed to two major factors: definition of zero strain and definition of Young's modulus. In this study, the baseline of zero strain corresponded to a small stress required to keep the suture straight. This is a negligible strain. The earlier studies used a higher zero strain value because elongation was based on physiological length. More work is needed to identify precisely the range of elongations that are

used in phonation, both for human and for canine vocal folds, before the issue of reference length can be properly resolved.

The morphological difference between canine and human vocal folds must be mentioned again in closing. The absence of the vocal ligament in the canine is very likely to render the Young's modulus of the body lower than in the human. The ligament, which is rather firmly attached to the body, probably absorbs most of the longitudinal stress at high strains, especially when the thyroarytenoid muscle is not activated. This will leave the cover more loose to propagate a mucosal wave at high fundamental frequencies. Without the ligament, the cover would become too stiff to propagate the wave. This may be a key reason why excised canine larynges do not phonate well at high fundamental frequencies. It is also a good reason to concentrate more effort on the study of human vocal fold tissue, but given the many methodological problems to be worked out first, the order of selection has been appropriate.

Acknowledgment

The authors are indebted to Paul L. Durham and Steve Hamiel for assistance in sample preparation and data collection. The study was supported by NINCDS Grant No. NS 16320-09.

References

Alipour-Haghighi, F., and Titze, I.R. (1985). "Viscoelastic modeling of canine vocalis muscle in relaxation," J. Acoust. Soc. Am., 78, 1939-1943.

Alipour-Haghighi, F. and Titze, I.R. (1990). "Stress Relaxation in Vocal Fold Tissues." Advances in Bioengineering, BED Vol. 17, pp. 21-24, ASME, New York.

Christensen, R.M. (1971). *Theory of Viscoelasticity--An Introduction,* (Academic Press, New York).

Finocchio, V., and Luschei, E.S. (1985). "Active length-tension characteristics of naturally activated muscle: Measurement and applications."*Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control,* edited by I.R. Titze and R.C. Scherer. Denver Center for the Performing Arts (pp. 283-291).

Fujisaki, H., Tatsumi, M., and Higuchi, N. (1981). "Analysis of pitch control in singing," in *Vocal Fold Physiology*, edited by K.N. Stevens and M. Hirano (University of Tokyo Press) pp. 347-362.

Fung, Y.C. (1979). "Inversion of a Class of Nonlinear Stress Strain Relationships of Biological Soft Tissues," J. Biomechanical Engineering, 101, 23-27.

Hamming, R.W. (1973). *Numerical Methods for Scientists and Engineers* (McGraw-Hill, New York).

Hirano, M. (1977). "Structure and Vibratory Behavior of the Vocal Fold," in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and F.S. Cooper, (University of Tokyo Press, Tokyo), pp. 13-30.

Hirano, M., Kakita, Y., Ohmaru, K., Kurita, S. (1982). "Structure and mechanical properties of the vocal fold," Speech and Language, 7, 271-297.

Hollien, H. (1960). "Vocal pitch variation related to changes in vocal fold length," J. Speech Hear. Res., 3, 150-156.

Isshiki, N., Ohkawa, M. and Goto, M. (1985). "Stiffness of the Vocal Cord in Dysphonia - Its Assessment and Treatment," Acta Otolaryngol. (Stockh), Suppl. 419, pp. 167-174.

Kakita, Y. Hirano, M., and Ohmaru, K. (1981). "Physical Properties of the Vocal Fold Tissues: Measurement on Excised Larynges," in *Vocal Fold Physiology,* edited by K.N. Stevens and M. Hirano, (Univ. of Tokyo Press, Tokyo), pp. 377-397.

Muller, E.M., Milenkovic, P.H., and MacLoed, G.E. (1985). "Perioral Tissue Mechanics During Speech Production", *Mathematics and Computers in Biomedical Applications*, edited by J. Eisenfeld and C. DeLisi (North Holland), pp. 363-371.

Nordin, M. and Frankel, V.H. (1980). "Biomechanics of Collagenous Tissues," In *Basic Biomechanics of the Skeletal System* edited by V.H. Frankel and M. Nordi, (Philadelphia: Lea & Febiger), pp. 87-110.

Oxlund, H., Manschot, J., and Viidik, A. (1988). "The Role of Elastin in the Mechanical Properties of Skin," J. Biomech., 21, 213-218.

Perlman, A.L., Titze, I.R., and Cooper, D.S. (1984). "Elasticity of canine vocal fold tissue," J. Speech Hear. Res., 27, 212-219.

Perlman, A.L. and Titze, I.R. (1985). "Measurement of Viscoelastic Properties in Live Tissue," In *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control,* edited by I.R. Titze & R.C. Scherer, Denver Center for the Performing Arts, (pp. 273-281).

Perlman, A.L. (1985). *A technique for measuring the elastic properties of vocal fold tissue,* Unpublished doctoral dissertation, The University of Iowa, Iowa City, Iowa.

Perlman, A.L. and Durham, P.L. (1987). "In vitro studies of vocal fold mucosa during isometric conditions," in *Laryngeal function in phonation and respiration,* edited by T. Baer, C. Sasaki, and K. Harris, (College-Hill Press: Boston) pp. 291-303.

Rack, P.M.H., & Westbury, D.R. (1969). "The effects of length and stimulus rate on tension in the isometric cat soleus muscle," Journal of Physiology, 204, 443-460.

Saito, S., Fukuda, H., Kitahara, S., Isogai, Y., Tsuzuki, T., Muta, H., Takayama, E., Fujioka, T., Kokawa, N., and Makino, K. (1985). "Pellet tracking in the vocal fold while phonating - Experimental study using canine larynges with muscle activity," In *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control,* edited by I.R. Titze and R.C. Scherer, (Denver Center for the Performing Arts, Denver Colorado), pp. 169-182.

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

Titze, I.R., Jiang, J., and Druker, D.G. (1988). "Preliminaries to the Body- Cover Theory of Pitch Control," J. Voice, 1(4), 314-319.

Titze, I.R. (1989). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. 85(4), 1699-1707.

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

Van den Berg, J. and Tan, T.S. (1959). "Results of experiments with human larynges," Practica Oto-Rhino-Laryngologica, 21, 425-450.

NCVS Status and Progress Report - 1
June 1991, 49-60

The Establishment of Open Articulatory Postures by Deaf and Hearing Talkers

Nancy Tye-Murray, Ph.D.

Department of Otolaryngology - Head and Neck Surgery, The University of Iowa

Abstract

Previous researchers have proposed that prelingually deafened talkers do not displace the tongue body to establish vowel steady-state postures, and displace the jaw excessively. The purpose of this investigation was to evaluate the opening gesture lingual displacement patterns of three deaf and two hearing adult talkers. Cinefluorography and x-ray microbeam indicated that the deaf subjects displaced their tongue bodies during the opening gestures. However, their glossal movement trajectories were qualitatively dissimilar from those of the hearing subjects. Whereas the hearing subjects moved the tongue differently for different vowel contexts, the deaf subjects had similar trajectories for all contexts. The common trajectories suggest that some deaf talkers contract their tongue muscles such that the tongue body moves similarly for all vowels. The deaf subjects also appeared to have a less flexible tongue body during speech production than the hearing subjects. Means for quantifying and comparing the lingual behaviors of deaf and hearing talkers are considered.

The Establishment of Open Articulatory Postures by Deaf and Hearing Talkers

Several investigators have suggested that prelingually deafened talkers establish the articulatory open postures, or "vowel" steady-states, with excessive jaw displacement and minimal tongue movement (Monsen, 1976a; Osberger, 1987; Osberger & McGarr, 1982; Stevens, Nickerson, & Rollins, 1983; Tye, Zimmermann, & Kelso, 1983; Zimmerman, & Rettaliata, 1981). A large body of acoustic, perceptual and physiological data appears to support this description.

Monsen (1976a) analyzed the vowel formants of hearing-impaired adolescents. Their second formant values clustered around 1800 Hz, regardless of vowel identity (also Angelocci, Kopp & Holbrook, 1964; Martony, 1965; Monsen, 1976b; 1978; Rothman, 1976). Noting that the first

formant may reflect changes in jaw displacement whereas the second formant may reflect changes in anterior/posterior lingual displacement, Monsen proposed that the restricted range of second formant values indicates a relatively immobile tongue body. Deaf talkers may attempt to distinguish the vowels with jaw displacement. Monsen (1978) later demonstrated that a reduced second formant range is negatively correlated with overall speech intelligibility.

Phonetic transcriptions agree with the acoustic findings. The vowels produced by deaf talkers often sound schwa-like or neutralized (Hudgins & Numbers, 1942; Smith, 1975).

X-ray data indicate that deaf talkers do not differentiate the open posture steady-states with tongue position (Boone, 1966; Crouter, 1963; Kaiser, 1959; Stein, 1980). For example, Tye et al. (1983) found that two prelingually deafened talkers had more similar lingual open steady state postures for /u/ and /a/ than did two hearing controls. Such findings have been interpreted as being consistent with descriptions of an immobile tongue body and excessive jaw displacement. However, at least two studies of movement behavior conflict with this view (Stein, 1980; Tye-Murray & Folkins, 1990).

Stein (1980) studied five deaf and two hearing talkers with cinefluorography. The five deaf subjects had markedly reduced intelligibility. The speech sample included the vowels /a, u, i/ in sentence contexts. The cinefluorographic records indicated that the deaf subjects displaced their tongue dorsum as much as the hearing talkers during opening gestures (page 188). Anecdotally, Stein noted that lingual movement for both /i/ and /u/ occurred in an anterior direction (page 132).

Tye-Murray and Folkins (1990) determined whether prelingually deaf adults could correctly produce stressed and unstressed syllables across known changes in stress patterning. Three deaf subjects with poor speech intelligibility and three hearing controls spoke sets of homogeneous syllable strings with stress patterns that they could tap successfully with their finger (e.g., TAP tap tap TAP tap tap...., where TAP corresponds to a stressed syllable and tap corresponds to an unstressed syllable). Strain gauge transduction of jaw movement indicated that both deaf and hearing subjects produced different displacements and durations for the stressed and unstressed syllables. The group of deaf subjects did not produce greater jaw displacements than the group of hearing subjects.

In the present investigation the lingual trajectories, i.e., displacement in an X-Y spatial plane, of three deaf and two hearing talkers were examined using cinefluorography or microbeam. Small pellets were placed upon the tongue surface and tracked during opening speech gestures. The purpose was to determine whether the deaf subjects displaced the tongue body during opening articulatory gestures, and if so, whether displacement patterns varied with vowel context. Jaw displacement was also measured to determine how much the subjects displaced their jaws.

When studying displacement patterns of tongue pellets, investigators often separate the anterior/posterior and superior/inferior components and plot them individually as a function of time (e.g., Forrest, Abbas, & Zimmermann, 1986; Stein, 1980; Zimmermann & Rettaliata, 1981). They then analyze and report the extent of displacement in one dimension, and less often both dimensions, for the temporal interval of interest. Although this procedure permits a quantitative analysis and is appropriate for some purposes, separating the two components for analysis also presents two disadvantages. First, the measures are not referenced to the vocal tract space. Thus, two subjects may both displace the tongue body inferiorly by 4 mm when producing the vowel /a/. However, one subject may position the tongue more superiorly in the vocal tract at movement onset than the other subject, and may move it anteriorly as well as inferiorly. As such, the perceptual consequences of the two gestures may be quite different. The second disadvantage is that by separating the anterior/

posterior and superior/inferior components, we may create a qualitatively inaccurate portrayal of the movement. For example, one lingual opening gesture may have twice as much anterior and twice as much inferior displacement as another. Quantitatively, these two gestures differ. However, the movement trajectories may be qualitatively alike such that both gestures produce similar percepts.

In this investigation, both qualitative and quantitative analyses of tongue behavior were performed. Qualitative examination of gesture trajectories is consistent with perception theories suggesting that vowel identity is specified by both the opening and closing movement trajectory paths of the articulators (Browman & Goldstein, 1986; Fowler, Rubin, Remez, and Turvey, 1980; Strange, 1989).

Methods

Subjects

Three prelingually deafened male (D1, D2, D3) and two hearing female (H1, H2) adults served as subjects. H1 and H2 reported no history of hearing loss or speech problems. D1 was tested eight years before the other subjects. He produced a smaller speech sample and was studied with cinefluorography. The purpose of his original experiment differed from the present purpose. All other subjects were studied with microbeam.

D1 has no measurable hearing and does not wear a hearing aid. D2 and D3 have profound bilateral hearing losses. D2 wears hearing aids bilaterally and wore the aids during testing. His aided threshold for warble tones in soundfield, averaged for the frequencies 500, 1000 and 2000 Hz, is 55 dB HL. His unaided pure-tone threshold average in the better ear (left) is 102 dB HL. D3 wears a hearing aid infrequently because it provides minimal benefit, and had not worn it for several months prior to the experiment. He did not wear the hearing aid on the day of testing. D1 and D3 lost their hearing after contracting meningitis, at age 18 months and 22 months, respectively. D2 is congenitally hearing-impaired. D1 was educated in an oral program and communicates only with speech. D2 and D3 attended an oral preschool and elementary school, and a total communication high school. D2 and D3 indicated that they communicate orally with hearing individuals, but that they prefer total communication.

All deaf subjects recorded the Rainbow Passage (Fairbanks, 1960). Three judges who are familiar with the speech of deaf talkers used a scale ranging from I understand none of the passage (1) to I understand all of the passage (10) to rate speech intelligibility. Before presenting the recordings, the experimenter read the Rainbow Passage aloud to familiarize each judge with the stimulus material and to model a score of 10. The judges listened individually in a sound-treated booth. The recordings were presented in soundfield at a comfortable listening level. All judges assigned D1 the highest score and D3 the lowest score. On average, D1 received a score of 5.3, D2 received a score of 3.7, and D3 received a score of 1.0.

Speech Sample

The subjects read consonant-vowel-consonant syllables embedded in the carrier phrase, "That's a _____." D1 recorded the syllables boon, dude, pot, teak, map and tin. The remaining subjects read all of these syllables except map, in order to match D1's data set. They also read the following stimuli: beet, dean, pan, ban, din, bob, and tune. This speech sample included the vowels /u, a, i, I, ae/, and permitted a large set of opening gestures to be compared. All phrases were read two times each.[1] The subjects read the stimuli list once, and then repeated it a second time.

Cinefluorographic Apparatus

Cinefluorography was used to monitor the articulatory activity of D1. Kent and Moll (1972) describe the procedures in detail. Small 3.2 mm hemispherical radiopaque markers were placed on the lower lip, tongue tip and tongue dorsum. D1 spoke and counted prior to filming in order to adapt to the markers.

Microbeam Apparatus

Microbeam data were collected from D2, D3, H1, and H2 using the x-ray microbeam apparatus at the University of Wisconsin Waisman Center (Fujimura, Kiritani & Ishida, 1973; Fujimura & Abbs, 1985). Small 1-3 mm gold pellets were placed on the lower lip, tongue tip, tongue blade and tongue dorsum for tracking. The tongue tip pellet was placed on the tongue region that contacts the alveolar ridge during /t/ production. The tongue dorsum pellet was placed on the region that contacts the palate during /k/ production. The tongue blade pellet was placed midway between the tongue tip and tongue dorsum pellets. Subjects were allowed to adapt to the markers prior to the experiment by speaking. A computer-controlled, low-dosage flying spot x-ray microbeam was used to track the pellet movements.

Analysis of the Cinefluorographic Film

The cinefluorographic films were projected onto a glass tablet with a 16-mm film projector (Lafayette Instrument Model 201). Frame-by-frame tracings were made of the opening articulatory movements, beginning at the last frame in which initial consonant closure was maintained and ending at the point of maximum tongue dorsum or jaw opening, whichever was achieved last.

The tracings were analyzed with respect to a mandibular reference. This reference frame eliminates the effects of jaw displacement on tongue position. For each opening gesture, the series of frame-by-frame tracings were aligned with a template drawing of the mandible. The trajectories of the tongue pellets were then plotted during the course of the gesture, one tracing after the other (Kent & Moll, 1972).

Analysis of the Microbeam Data

The microbeam data were displayed using a SUN 68010 Computer Work Station. An X-Y Cartesian coordinate set was used to display the opening gesture tongue dorsum, tongue blade, and tongue tip pellet trajectories in the anterior-posterior dimension (X) and the superior-inferior dimension (Y) of the vocal tract. The X axis corresponded with a line drawn through the cusp of the central maxillary incisor and the maxillary molar. The intercept of the axes corresponded with the central maxillary incisor position. An opening gesture trajectory began when the primary articulator released the initial consonant closure posture and terminated when tongue dorsum or blade displacement ceased, which ever occurred last. The effect of jaw displacement was subtracted from lingual displacement before the trajectories were plotted. This was accomplished by drawing a line through reference pellets placed upon the mandible incisor and molar. For each point in time, the computer translated and rotated the other pellets to align the current jaw position with the reference line. Only the opening gestures for the syllables pot, pan, beet, boon, and tin were initially plotted. The first four syllables were chosen because the initial closure posture did not constrain glossal activity. The syllable tin was chosen because it corresponded with data collected for D1 and was the only sample with the vowel /I/.

Displacement Means and Standard Deviations

The mean inferior/superior displacements of the tongue blade, tongue dorsum, and mandible were computed for H1, H2, D2, and D3, using the absolute displacement values for every syllable in the speech sample. The displacements of the tongue pellets were averaged after subtracting the effect of jaw displacement. Due to the small number of subjects, the mean values for the two subject groups were not compared statistically (Forrest, Weismer, & Adams, 1990).

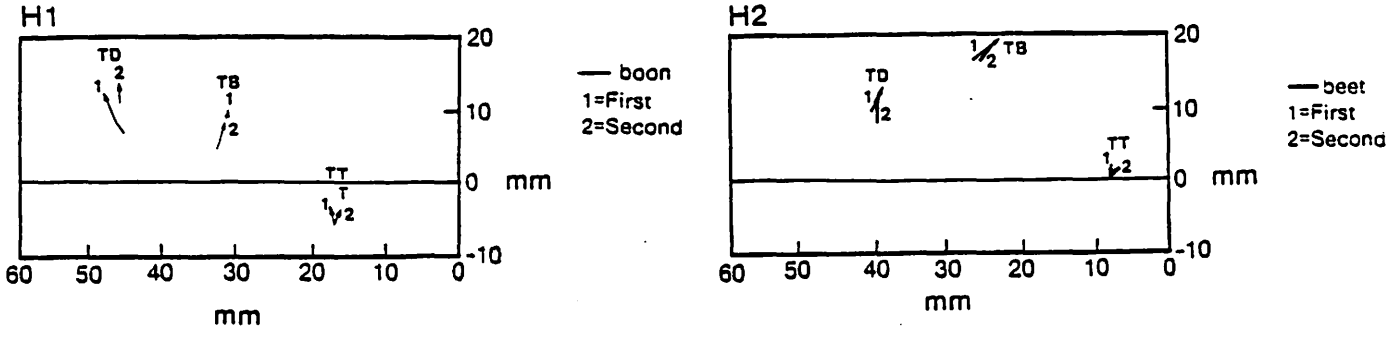

Figure 1. The lingual trajectories for the first and second productions of boon by H1 and beet by H2. Due to technical problems, only one sample of boon was available for H2. The trajectories are plotted with a mandibular reference. The beginning of an arrow corresponds to the movement onset; the point of the arrow corresponds to the movement offset. (TD = tongue dorsum; TB = tongue blade; TT = tongue tip.)

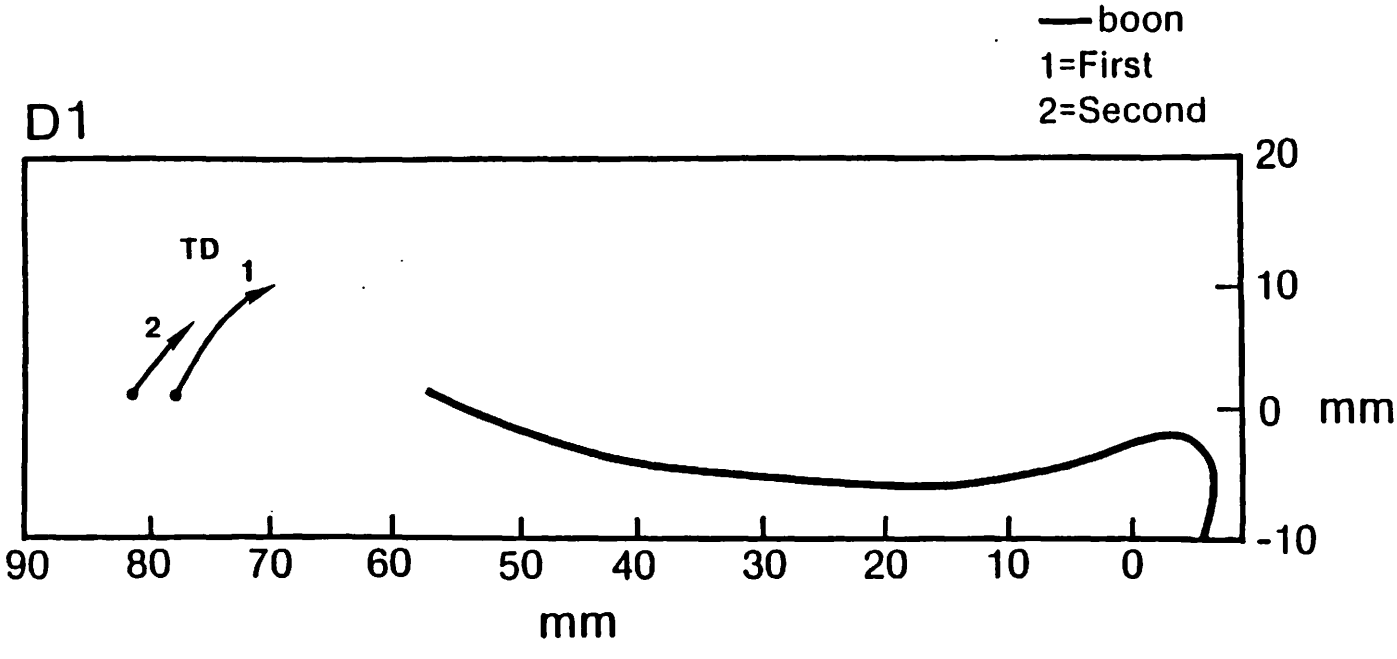

Figure 2. The lingual trajectories for the first and second productions of boon by D1. Only the tongue dorsum trajectories are displayed as tongue blade trajectories were unavailable.

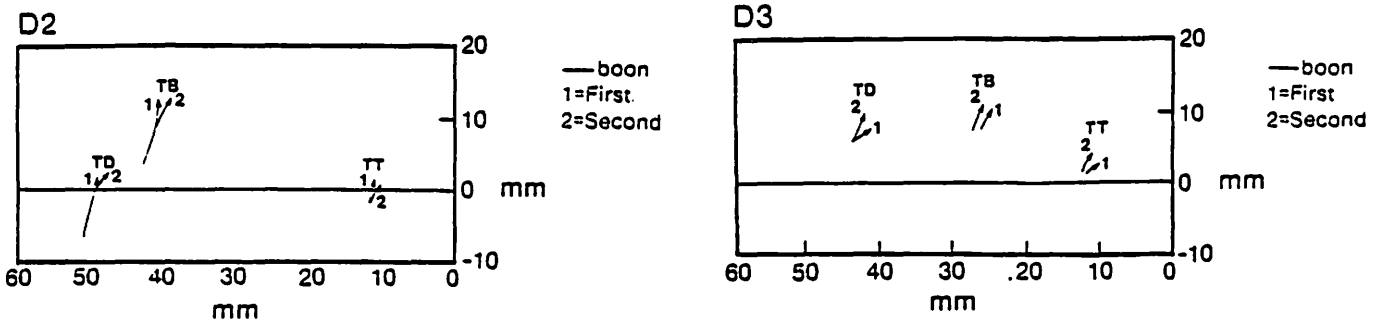

Figure 3. The lingual trajectories for the first and second productions of boon by D2 and D3.

Results

Trajectory Variability

The movement trajectories for the two productions of each syllable were qualitatively similar for every subject. Examples of syllable repetitions for the hearing subjects appear in Figure 1. Examples for the deaf subjects appear in Figures 2 and 3.

Only one sample of each syllable is presented in the following figures. Only the tongue dorsum and tongue blade trajectories will be considered.

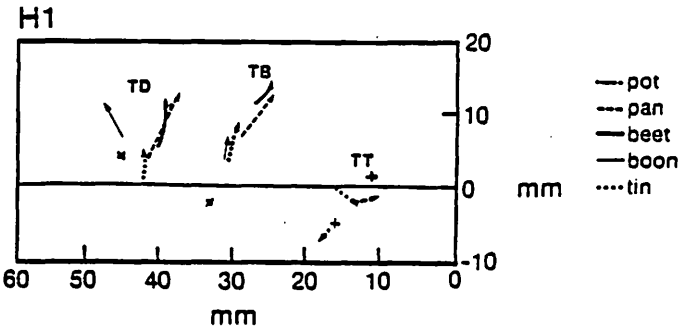

Figure 4. The lingual trajectories for the opening gestures produced by H1. Crosses indicate that no movement occurred for pot.

Hearing Subjects

The lingual trajectories for the hearing subjects appear in Figures 4 and 5. Figure 4 displays the tongue pellet trajectories for H1. The tongue dorsum marker moved anteriorly/superiorly for pan, superiorly for beet and tin, and posteriorly/superiorly for boon. The tongue blade marker moved anteriorly/superiorly for pan, tin, and beet and superiorly for boot. Both pellets were immobile for pot.

Figure 5 shows that H2's tongue dorsum marker moved superiorly for pan and beet, and posteriorly/superiorly for boon, pot and tin. The tongue blade marker moved posteriorly/ inferiorly for pot, posteriorly/ superiorly for boon, and anteriorly/superiorly for beet, tin and pan.

A comparison of Figures 4 and 5 reveals qualitative similarities between the opening gestures by H1 and H2. For both subjects, dorsal movement terminated most posteriorly for boon. The most anterior blade

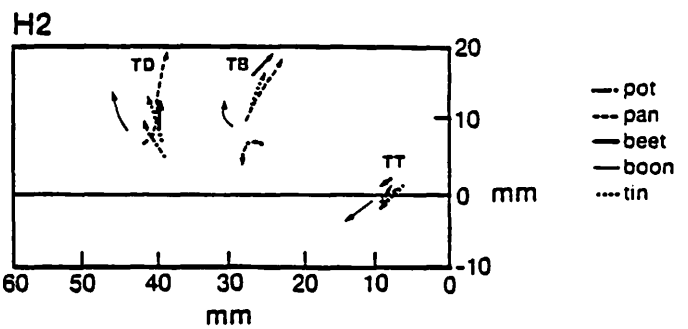

Figure 5. The lingual trajectories for the opening gestures produced by H2.

termination positions occurred for beet and ban. The most inferior blade termination positions occurred for pot, and these positions were considerably lower than the corresponding dorsal pellets. The most extensive displacements occurred for pan. The trajectories displayed in Figures 3 and 4 accord with the traditional view of /a/ as a low vowel, /ae, i, I/ as high front vowels, and /u/ as a high back vowel.

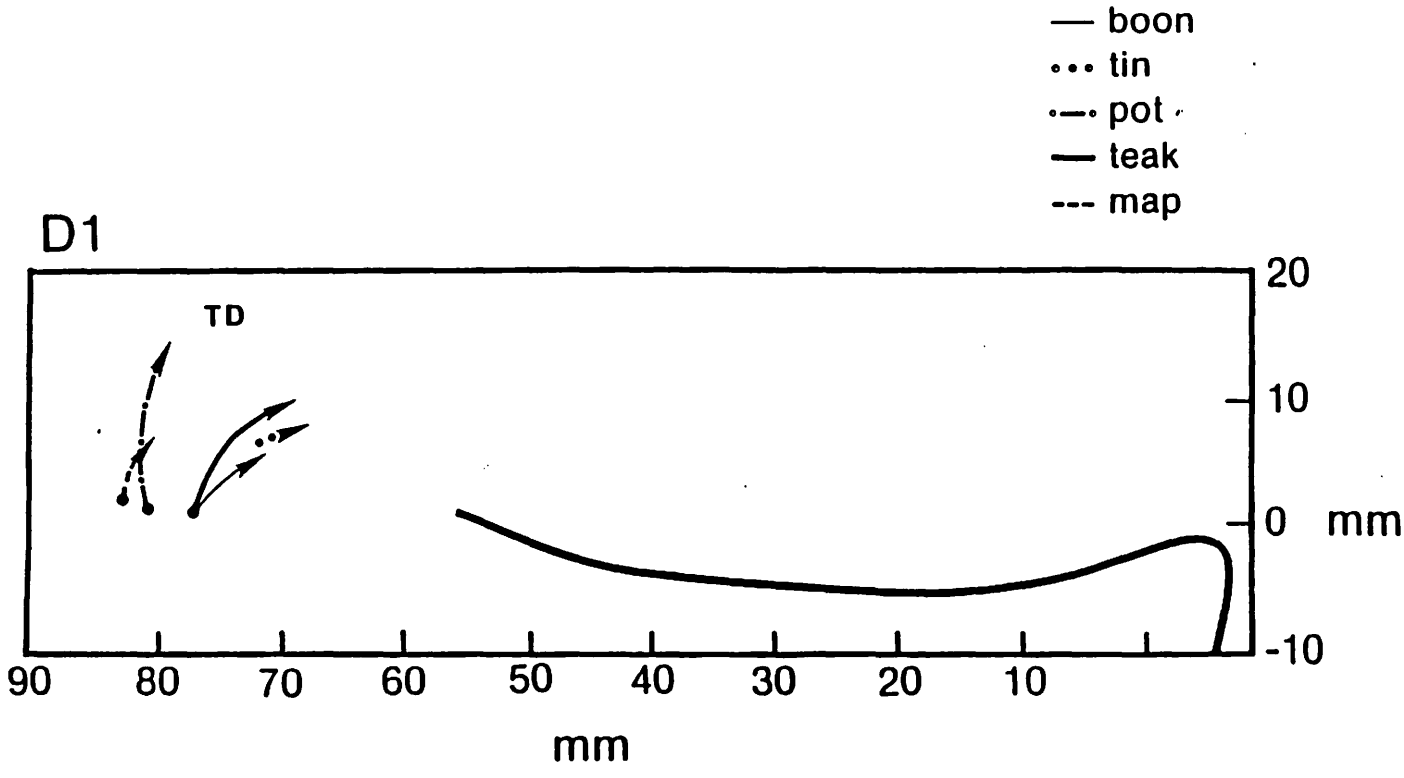

Figure 6. The lingual trajectories for the opening gestures produced by D1.

Deaf Subjects

The lingual trajectories for the deaf subjects appear in Figures 6-8. Figure 6 indicates that D1's tongue dorsum trajectories projected anteriorly/superiorly for all syllables, although the anterior displacement was small for pot. The greatest displacement occurred for the syllables teak and pot; the least occurred for map and tin.

D2's lingual trajectories appear in Figure 7. For all syllables except tin, displacement was in the anterior/superior direction. For tin, marginal posterior/superior displacement occurred. The greatest displacement was associated with

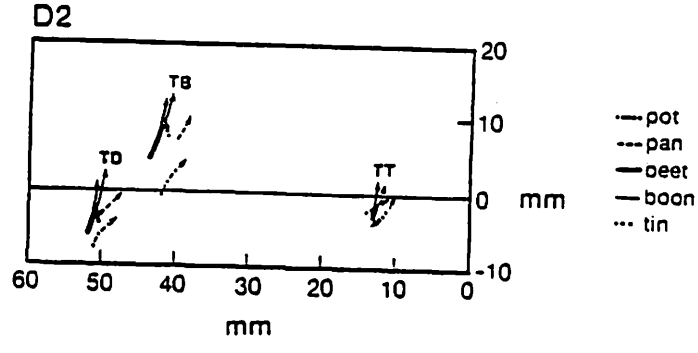

Figure 7. The lingual trajectories for the opening gestures produced by D2.

beet and boon (where the trajectories almost overlap); the least was associated with tin.

D3's lingual trajectories are displayed in Figure 8. For all syllables except tin, displacement occurred in a superior direction. For boon, beet, and pan, displacement also occurred anteriorly. The least extensive displacement was noted for the syllable pan.

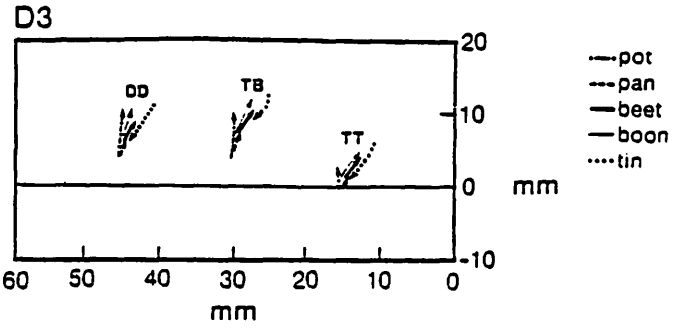

Figure 8. The lingual trajectories for the opening gestures produced by D3.

For D3's tin, the tongue dorsum and tongue blade pellets moved posteriorly/inferiorly. Since his lingual trajectories were different for this syllable, lingual motion appeared to be affected by the releasing action of the tongue tip for the initial consonant closure. To evaluate this possibility, the opening gesture trajectories for his syllables teak, dean, din, tune and dude were also examined. This examination indicated that the tongue body was immobile for tune and dude. For the remaining syllables, the tongue dorsum and tongue blade trajectories were like those noted for tin; displacement occurred in a posterior/inferior direction. A similar examination of these syllables by D2 showed that an initial alveolar closure had little effect on the trajectories toward the /u, i/ steady-state postures. Displacement was less extensive and more variable, but in the same direction as when preceded by bilabial closure.

Comparison of the Hearing and Deaf Subjects

For both hearing subjects, the tongue dorsum and tongue blade pellets did not always follow similar trajectories for a particular syllable (for H1: boon, beet, and tin; for H2: all syllables except pan). For example, H2's tongue dorsum pellet moved posteriorly/superiorly while the tongue blade marker moved posteriorly/inferiorly during the opening gesture of pot. In contrast, the tongue dorsum and tongue blade trajectories were similar for a given syllable for both D2 and D3.

For D2 and the hearing subjects, the most inferior initiation and termination positions for the dorsal and blade pellets occurred for pot; the most superior termination position for the blade pellets occurred for beet. Apart from these similarities, the differences between the hearing and deaf groups are striking. For example, the beet and boon trajectories extended across similar vocal tract regions for the deaf but not the hearing subjects. The vowel /ae/ was associated with small displacement with the one group but not the other.

Displacement Means and Standard Deviations

Table 1 presents the means and standard deviations for the opening articulatory gestures. On average, D3 did not displace the mandible more than H1 and H2, although he displaced the tongue blade and tongue dorsum less. Conversely, while D2 demonstrated greater mandibular displacement than either hearing subject, he demonstrated comparable lingual displacement. The tongue pellets of the hearing subjects displaced at least twice as much as their mandible, on average, whereas the deaf subjects demonstrated similar displacement magnitudes for both articulators.

Discussion

The three deaf subjects displaced the tongue body during opening articulatory gestures. Therefore, they cannot be characterized as not displacing the tongue body. However, the displacement trajectories were similar for all vowel contexts, with the exception of syllables with an initial alveolar closure by D3. For these syllables, the tongue displaced posteriorly/inferiorly. The common glossal trajectories suggest that some deaf talkers contract their tongue muscles such that the tongue body moves similarly for all vowels.

Some deaf talkers "front" their vowels (Crouter, 1963; McGarr & Gelfer, 1983; Stein, 1980) whereas others "back" the vowels (Boone, 1966; Geffner, 1980; Nober, 1967; Smith, 1975). Thus deaf talkers may vary in whether the stereotyped glossal muscle contractions result in an anterior (as in this investigation) or posterior vocal tract constriction.

D1 and D2 did not demonstrate an effect of initial alveolar closure upon the opening gesture lingual trajectories. This finding agrees with Stein (1980), who showed that the open steady-state postures of five deaf talkers were uninfluenced by consonant closure identity. In contrast, D3 demonstrated different lingual trajectories for syllables beginning with initial /t, d/ and syllables beginning with /b, p/.

Future investigations could attempt to quantify the kinds of qualitative differences that were noted in the lingual trajectory patterns. For example, Georgopoulos, Kalaska, and Massey (1981) developed a method for characterizing arm movement trajectories. For each arm movement pattern, they identified 19 equidistant points and defined each with X-Y coordinates. They then calculated the variability of points across repetitions of a movement gesture. The present investigation suggests that prelingually deafened talkers may employ stereotyped tongue gestures in establishing different vowel steady states. If the lingual trajectories were quantified as in Georgopoulos et al. (1980), and then examined for different vowel contexts, one might expect that the dispersion of any given point would be greater for a hearing than for a deaf talker.

The tongue dorsum and tongue blade pellets followed similar paths for D2 and D3 but usually different paths for the two hearing subjects. This finding suggests that deaf talkers may not contract and extend their tongues like hearing talkers. They may have a less flexible tongue body during speech production because they have developed aberrant principles for constraining tongue movement (Tye et al., 1983; Tye-Murray, 1987). Perhaps future investigations can document these differences quantitatively also. For example, one might divide the duration of the opening gesture into equal temporal intervals. At each interval, the distance between pellets placed upon the tongue dorsum and tongue blade could be computed. If the two pellets followed similar paths throughout the gesture, as might occur with prelingually deafened talkers, than the distance between them should show little variability as a function of time.

D3 and the hearing subjects demonstrated similar mean jaw displacements during the opening articulatory gestures (Table 1). This finding agrees with Tye-Murray and Folkins (1990). H1 and H2 appeared to displace their tongues relatively more than their mandibles, whereas D2 and D3 did not. As such, deaf talkers may not displace the jaw excessively in an absolute sense, but may employ excessive displacement relative to the extent of their lingual displacements. These behaviors, in conjunction with stereotypical tongue movement, may account for the neutral-sounding vowels produced by some deaf talkers.

Table 1
Inferior/Superior Displacement Means (Standard Deviations) for the Opening Articulatory Gestures

	Articulator (mm)		
Subject	Mandible	Tongue Blade	Tongue Dorsum
H1	2(1)	5(3)	5(3)
H2	2(2)	4(3)	7(4)
D2	4(2)	5(3)	5(3)
D3	2(1)	2(2)	3(2)

Acknowledgments

I thank Anne Smith, Randall Monsen and an anonymous reviewer for their editorial comments, and Robert Nadler and the staff of the Waisman Center for assistance in data collection. This work was supported by NHI Grants N520466, NS16373 and NS-0755-22.

References

Angelocci, A. A., Kopp, G. A., & Holbrook, A. (**1964**). "The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys." Journal of Speech and Hearing Disorders, 29, 156-170.

Boone, D. (**1966**). "Modification of the voices of deaf children." Volta Review, 68, 686-694.

Browman, C. P., & Goldstein, L. M. (**1986**). "Towards an articulatory phonology." *Phonology Yearbook,* 3, 219-252.

Crouter, L. (**1963**). *A cinefluorographic comparison of selected vowels spoken by deaf and hearing subjects*. Unpublished master's thesis, University of Kansas, Kansas City.

Fairbanks, G. (**1960**). *Voice and articulation drillbook.* (New York, Harper & Row).

Forrest, K., Abbas, P. J., & Zimmermann, G. N. (**1986**). "Effects of white noise masking and low pass filtering on speech kinematics." Journal of Speech and Hearing Research, 29, 549-562.

Forrest, K., Weismer, G., & Adams, S. (**1990**). "Statistical comparison of movement amplitudes from groupings of normal geriatric speakers." Journal of Speech and Hearing Research, 33, 386-389.

Fowler, C. A., Rubin, P. E., Remez, R. E., & Turvey, M. T. (**1980**). "Implications for speech production of a general theory of action." In *Language Production,* edited by B. Butterworth. (Academic Press, New York) pp. 373-420.

Fujimura, O., & Abbs, J. H. (**1985**). "A computer controlled x-ray microbeam system for articulatory studies." Paper presented at the meeting of the American Association of Physicists in Medicine, Seattle, WA.

Fujimura, O., Kiritani, S., & Ishida, H. (1973). "Computer controlled radiography for observation of movements of articulatory and other human organs." Computers in Biology and Medicine, 3, 371-384.

Geffner, D. (1980). "Feature characteristics of spontaneous speech production in young deaf children." Journal of Communicative Disorders, 13, 443-454.

Georgopoulos, A., Kolaska, J., & Massey, J. (1981). "Spatial trajectories and reaction times of aimed movements: Effects of practice, uncertainty and change in target location." Journal of Neurophysiology, 46, 725-743.

Hudgins, C. V., & Numbers, G. C. (1942). "An investigation of the intelligibility of the speech of the deaf." Genetic Psychology Monographs, 25, 289-392.

Kaiser, L. (1959). "Speech and handling of language in normal and abnormal Dutch school children." Folia Phoniatrica, 11, 219-229.

Kent, R., & Moll, K. (1972). "Tongue body articulation during vowel and diphthongal gestures." Folia Phoniatrica, 24, 278-300.

Martony, J. (1965). "Studies on the speech of the deaf: I." Quarterly Progress Status Report, Speech Transmission Laboratories, Stockholm: Royal Institute of Technology, 3, 16-24.

McGarr, N. S., & Gelfer, C. E. (1983). "Simultaneous measurements of vowels produced by a hearing-impaired speaker." Language and Speech, 26, 233-246.

Monsen, R. B. (1976a). "Normal and reduced phonological space: The production of English vowels by deaf adolescents." Journal of Phonetics, 4, 189-198.

Monsen, R. B. (1976b). "Second formant transitions of selected consonant-vowel combinations in the speech of deaf and normal-hearing children." Journal of Speech and Hearing Research, 19, 279-289.

Monsen, R. B. (1978). "Toward measuring how well hearing-impaired children speak." Journal of Speech and Hearing Research, 21, 197-219.

Nober, E. H. (1967). "Articulation of the deaf." Exceptional Children, 33, 611-621.

Osberger, M. J. (1987). "Training effects on vowel production by two profoundly hearing-impaired speakers." Journal of Speech and Hearing Research, 30, 241-251.

Osberger, M. J., & McGarr, N. S. (1982). "Speech production characteristics of the hearing impaired." In *Speech and Language: Advances in Basic Science and Research,* edited by N. Lass (Academic Press, New York) pp. 221-283.

Rothman, H. (1976). "A spectrographic investigation ofconsonant-vowel transitions in the speech of deaf adults." Journal of Phonetics, 4, 129-136.

Smith, C. R. (1975). "Residual hearing and speech production in deaf children." Journal of Speech and Hearing Research, 18, 795-811.

Stein, D. M. (1980). *A study of articulatory characteristics of deaf talkers,* (Doctoral dissertation, University of Iowa, 1980). Dissertation Abstracts International, 41, 1327B.

Stevens, K. N., Nickerson, R. S., & Rollins, A. M. (1983). "Suprasegmental and postural aspects of speech production and their effect on articulatory skills and intelligibility." In *Speech of the Hearing Impaired,* edited by I. Hochberg, H. Levitt, & M.J. Osberger, (University Park Press, Baltimore) pp. 35-51.

Strange, W. (1989). "Evolving theories of vowel perception." Journal of the Acoustical Society of America, 85, 2081-2087.

Tye, N., Zimmermann, G. N., & Kelso, J. A. S. (1983). "'Compensatory articulation' in hearing impaired speakers: A cinefluorographic study." Journal of Phonetics, 11, 101-115.

Tye-Murray, N. (1987). "Effects of vowel context on the articulatory closure postures of deaf speakers." Journal of Speech and Hearing Research, 30, 99-104.

Tye-Murray, N., & Folkins, J. W. (1990). "Jaw and lip movements of deaf talkers producing utterances with known stress patterns." Journal of the Acoustical Society of America, 87, 2675-2683.

Zimmermann, G. N., & Rettaliata, B. P. (1981). "Articulatory patterns of an adventitiously deaf speaker: Implications for the role of auditory information in speech production." Journal of Speech and Hearing Research, 24, 169-178.

Footnotes

[1]Due to technical problems, only one production of dean, teak and pot was collected from H1; only one production of boon and pan and no productions of bob were collected from H2; only one production of pan and dude was collected from D2; and only one production of pot and ban was collected from D3. The opening gesture lingual trajectories of bob were obtained from H1 and D3 and found to be similar to those for their pot syllables. Both syllables require an initial bilabial closure and contain /a/. The trajectories of ban were obtained for H2 and found to be similar to that of his syllable pan. Both syllables require an initial bilabial closure and contain /ae/.

NCVS Status and Progress Report - 1
June 1991, 61-86

The Efficacy of Voice Therapy for Patients with Parkinson's disease

Lorraine Olson Ramig, Ph.D., CCC-SP
Yoshiyuki Horii, Ph.D.
Department of Communication Disorders and Speech Science, University of Colorado-Boulder
and
Recording and Research Center, The Denver Center for the Performing Arts
Carolyn Mead Bonitati, M.A.
Tempe, Arizona

Abstract

This study is an investigation of the efficacy of voice therapy for patients with Parkinson's disease. Intensive voice therapy designed to target the underlying laryngeal and respiratory pathophysiologies accompanying disordered speech in patients with idiopathic Parkinson's disease was administered to 40 patients; 12 patients served as control subjects. Microphone recordings were made of all subjects before and after therapy or at one month intervals in order to measure the following variables: maximum duration sustained vowel phonation (seconds), maximum fundamental frequency range (ST), mean fundamental frequency (Hz) and fundamental frequency variability (STSD) in reading, and forced and slow vital capacities (L). In addition, subjects, spouses and speech pathologists made perceptual ratings of the subjects before and after treatment. Statistically significant differences ($p < .05$) were measured on the following variables after therapy in the treated group: maximum duration sustained phonation, maximum fundamental frequency range, mean fundamental frequency and fundamental frequency variability in reading. No statistically significant differences on these variables were measured in the control group. Corresponding changes in the perceptual variables of loudness, monotonous speech and intelligibility were measured as well. These findings support the effectiveness of intensive voice therapy for patients with Parkinson's disease and should be viewed as an initial step in documenting treatment effectiveness in this population.

Parkinson's disease is a progressive, degenerative neurological disease resulting from nigrostriatal dopamine deficiency (Hornykiewicz, 1966; Hornykiewicz and Kish, 1982). It affects approximately 1 in every 100 individuals over age 60 and 1 in every 1000 individuals under age 60 nationwide (Duvosin, 1984). At least seventy-five percent of these patients have disordered speech which limits their ability to function fully in society (Hoberman, 1958; Canter, 1961; Logeman, Fisher, Boshes and Blonsky, 1978; Oxtoby, 1981; Streifler and Hofman, 1984).

As the American population ages, the incidence of Parkinson's disease increases (Kurland, Kirtzke, Goldberg, Choi and Williams, 1973; Martilla and Rinne, 1979; Hoehn, 1983). It is not unusual for these patients to live well over ten years after initial diagnosis (Marttila and Rinne, 1979), with every Parkinson's disease patient eventually developing speech and voice problems (Selby, 1986). While neuropharmacological interventions, such as L-dopa and Sinimet, have proven effective in the management of many motor symptoms of Parkinson's disease (e.g., limb tremor and rigidity) (Klawans, 1988), the speech and voice problems of these patients are not consistently or significantly alleviated by these interventions (Leanderson, Meyerson and Persson, 1971; Hansen et al., 1984; Larson, Ramig and Scherer, 1988).

The classic description of the speech of these patients includes reduced loudness, hoarse and breathy voice, monotony of pitch, short rushes of speech and imprecise consonants (Darley, Aronson and Brown, 1969 a,b; Critchley, 1981; Aronson, 1985). These perceptual characteristics have been associated with rigidity and hypokinesia in respiratory, phonatory and articulatory musculature. For example, reduced loudness, breathy voice quality and abnormally short duration of phonation have been associated with rigidity in laryngeal musculature resulting in bowed vocal folds (Hansen, Gerratt and Ward, 1984). Reduced fundamental frequency range and monotony of pitch have been associated with rigidity in the cricothyroid laryngeal muscles, which control pitch changes (Aronson, 1985). Rigid respiratory musculature, together with reduced vocal fold adduction, have been associated with reduced maximum duration vowel phonation (Canter, 1965; Critchley, 1981). A reduced range of articulator movement, due to either acceleration (Netsell, Daniel and Celesia, 1975; Hirose et al., 1981) or rigidity (Hunker, Abbs and Barlow, 1982; Weismer, 1984) has been associated with the imprecise consonant articulation of these patients.

Despite the high incidence of Parkinson's disease and accompanying breakdown in communication, no effective method of speech therapy for these patients has been established. Efforts at traditional speech therapy have proven ineffective and speech pathologists generally remain at a loss as to how to maximally help these patients improve their communication abilities (Sarno, 1968; Allan, 1970; Greene, 1980; Aronson, 1985; Rubow and Strand, 1985; Weiner and Singer, 1989). Therefore, most patients do not receive speech therapy and their ability to communicate continues to deteriorate as their disease progresses (Morley, 1955; Logemann et al., 1978).

It was the goal of this study to evaluate a speech treatment program for patients with Parkinson's disease which focused on vocal improvement. The design of this treatment program was shaped by a number of factors: the frequent reports of disordered voice in patients with Parkinson's disease (e.g., Logemann et al. (1978) reported 89% of 200 Parkinson's disease patients had disordered voice); the clinically apparent role of disordered voice in reducing patients' communication intelligibility (Ramig, in press) and reports that intensive speech therapy focusing on phonation has been of value to patients with Parkinson's disease (Scott and Caird, 1983; Robertson and Thompson, 1984).

Method

Treatment program

Our voice treatment program for patients with Parkinson's disease was designed to impact the overall communication function of patients while generating objective data about the physiologic bases for treatment-related change. Therapy techniques were designed to improve perceptual characteristics of voice by targeting the hypothesized underlying laryngeal pathophysiology. The program was designed to achieve a balance between clinical feasibility and objective research quality quantitative documentation.

Our initial treatment program (Table 1) included five elements: (1) the phonatory disorders reported in Parkinson's disease, (2) the hypothesized underlying laryngeal and/or respiratory pathophysiology, (3) our therapy goals and basic techniques, (4) the objective acoustic and physiologic variables measured to document treatment effectiveness and (5) the perceptual variables measured to assess the impact of therapy-related change on functional communication.

For example, as the breathy, weak voices of patients with Parkinson's disease have been associated with bowed vocal folds (lack of vocal fold closure or adduction), one of our therapy goals was to increase loudness and decrease breathiness by increasing vocal fold adduction. Specific therapy techniques included those designed to maximize vocal fold closure such as pushing and lifting with phonation, techniques which have been used with patients who have vocal fold adduction problems related to laryngeal nerve paralysis (Aronson, 1985) and that first were suggested for Parkinson's disease patients by Froeschels, Kastein and Weiss (1955). To assess the impact of therapy-related changes on functional communication, the perceptual variables of loudness and overall speech intelligibility were rated. To document therapy-related changes objectively, we measured maximum duration of sustained vowel phonation. To gauge respiratory system function, we also measured slow and forced vital capacities.

The monotonous voices of Parkinson's disease patients have been associated with rigidity in the cricothyroid muscles, one of our therapy goals was to improve intonation by increasing range of motion of the cricothyroid muscle. Specific therapy techniques included those designed to maximize cricothyroid muscle function, such as phonating at highest and lowest pitch levels. To assess the impact of therapy-related changes on functional communication, the perceptual variable of intonation was rated. To document therapy-related changes objectively, we measured maximum fundamental frequency range and fundamental frequency variability in reading.

Maximum duration sustained vowel phonation and maximum fundamental range were selected for measurement and training for a number of reasons. These tasks stimulate patients to increase their phonatory effort level and to sustain an increased effort level over time. In addition, these tasks can be taught to patients at most cognitive levels, they can be analyzed simply in a clinical situation to provide on-line feedback, they can be practiced independently by patients and the methodology for quantitative analysis is feasible even in cases of dysphonic voices which are common in Parkinson's disease. Maximum duration sustained vowel phonation was selected in particular because improvement can be obtained through both increasing respiratory system support and vocal fold adduction, two physiologic variables important for increasing vocal loudness.

We designed our program to be intensive, offering daily speech treatment for a month. We hypothesized that this intensive focus would maximize patient motivation and daily carryover (Scott and Caird, 1983; Robertson and Thompson, 1984; Mayeux, Williams, Stern and Cote, 1986; Palmer, Mortimer, Webster, Bistevins and Dickson, 1986; Gauthier, Dalziel and Gauthier, 1987).

Table 1

Framework and rationale for initial program of speech therapy administered to forty patients with idiopathic Parkinson's disease; treatment philosophy is intensive therapy with a focus on phonation and immediate carryover into functional communication.

Perceptual characteristics of speech	Hypothesized laryngeal and/or respiratory pathophysiology	Therapy goals and tasks	Acoustic, physiologic variables measured	Perceptual variables measured
"Reduced loudness, breathy, weak voice" (Logemann, et al., 1978; Aronson, 1985)	Bowed vocal folds (Hansen et al., 1984), rigidity, hypokinesia in laryngeal and/or respiratory muscles; reduced adduction; reduced inspiratory, expiratory volumes (Critchley, 1981);	1) increase vocal fold adduction - isometric (pushing, lifting) with phonation increase maximum duration vowel phonation at increased intensity -think "shout" -speak over background noise	Maximum duration of sustained vowel phonation (sec)	loudness breathiness intelligibility
		2) increase respiratory support -posture -deep breath before speak -frequent breaths -phrasing of words in sentences	vital capacity (L;%)	
"Reduced pitch variability monopitch" (Logemann, et al., 1978; Aronson, 1985)	Rigidity cricothyroid muscle (Aronson, 1985)	1) increase maximum fundamental frequency range -high and low pitch scales -Sustain phonation at highest and lowest pitches	maximum range of fundamental frequency (ST)	
			variability of fundamental frequency in connected speech (STSD)	monotone intelligibility
Unsteady, hoarse, rough voice (Logemann et al., 1978)	Rigidity, hypokinesia, tremor in laryngeal and respiratory muscles (Hansen et al., 1984; Critchley, 1981)	1) increase steadiness of phonation -maximum duration tasks with constant intensity -consistent firm voice throughout sentence	improved measures of phonatory stability (coefficient of variation of frequency, coefficient of variation of amplitude, jitter, shimmer harmonics to noise ratio)	steadiness of voice hoarse rough tremorous intelligibility

All experimental subjects participated in the voice treatment program summarized in Table 1. This voice treatment program was provided within the context of a multidisciplinary treatment program designed for patients with Parkinson's disease which also involved physical therapy, exercise, occupational therapy, counseling and recreation. Within the one month period of time, each patient was involved in rehabilitation activities approximately 4-5 hours a day, including 1 hour of voice therapy. On average, the experimental patients received 12-14 hours of individual voice therapy within this one-month time frame. In addition to the activities summarized in Table 1, all patients participated in educational lectures regarding their Parkinson's disease which included information about their speech and voice characteristics. Spouse education was included when appropriate. A typical voice treatment session included a variety of maximum phonation drills and spontaneous, functional speech tasks. The same speech pathologist administered all the voice treatment.

Subjects

Fifty-two patients with idiopathic Parkinson's disease served as subjects. Their ages, sex, stages of disease (Hoehn and Yahr, 1967) medications and primary speech disorders are summarized in Tables 2 and 3 (see end of chapter). Forty of these patients served as treated experimental subjects; twelve served as untreated control subjects. Group data are summarized in Table 4.

Table 4.

Ages and numbers of subjects in the experimental and control groups classified by stage of Parkinson's disease.

		Subjects			
		Experimental Sex		Control Sex	
		M	F	M	F
Stage	I	M 65.5 (n=2) Range 64-67		M 69 (n=2) Range 59-79	64 (n=1)
	II	M 70.1 (n=9) Range 59-78	70 (n=2) 67.73	M 70.6 (n=6) Range 64-73	63 (n=1)
	III	M 70.9 (n=15) Range 59-86	72 (n=8) 59-81	M 64.5 (n=2) Range 61-68	
	III-IV	M 72 (n=1)			
	IV	M 76.6 (n=3) Range 75-80			
	V				

All experimental subjects had laryngeal examinations before treatment was administered. Bowed vocal folds were reported on pre-treatment laryngeal examinations of thirty-five of the forty experimental subjects. Tremor, dyskinesia and mucus were also reported. While patients were on a variety of medications for Parkinson's disease, these medications were not changed during the course of the study and voice data were collected at the same time post-medication throughout the study. Seventy percent of the experimental patients were residents of Arizona; thirty percent were from across the United States (e.g., California, Midwest, New York); all control patients were residents of the Denver-Boulder, Colorado area.

It should be pointed out that as reflected in stage of disease data, the experimental group of patients was somewhat more impaired than the control group. In addition, 1 of the 40 experimental subjects had been a semi-professional singer while 3 of the 12 control subjects had been semi-professional singers.

Data collection

All subjects were seated in an IAC sound-treated booth with a microphone (AKG 190E) positioned 15 cm in front of the lips. The microphone was powered by an AKG preamplifier. After amplification (DSC-240) the voice signals were recorded on to one channel of a Otari MX 5050 tape recorder. Subjects were asked to take a deep breath and sustain phonation of the vowel "ah" for as long and steadily as possible. A timer with a second hand was within view of the subject and he/she was encouraged to monitor his/her performance and sustain phonation for a longer duration each time. Encouragement was given to motivate performance after each phonation. Six to nine maximally sustained phonations were collected from each subject. In certain instances, the experimenter determined maximum performance had been reached after four or five repetitions because the subject was fatiguing and the task was terminated at that time. All experimental and control subjects were able to perform this task. Subjects were asked to read aloud the phonetically balanced "Rainbow Passage" (Fairbanks, 1960) at a comfortable rate and loudness. In certain instances because of reduced visual acuity or inability to read, subjects were unable to complete this task. Subjects were asked to generate their maximum fundamental frequency range (including falsetto and excluding vocal fry) (Hollien, Dew and Philips, 1971). The experimenter worked with each subject using both the step and gliding (glissano) methods (Reich, Frederickson, Mason and Schlauch, 1990) until both the experimenter and subject were satisfied that performance limitations had been met. This task was repeated six to nine times within each session. In certain instances, subjects were unable to complete this task because of difficulties in following directions.

Forced and slow vital capacities were measured for the experimental subjects using a spirometer (Multi-spiro AS/100). Subjects were asked to take their deepest breath and "blow out as hard and fast and long as you could" and as "long as you could " for forced and slow vital capacities respectively. Because some subjects were wheel chair-bound or had balance problems, all subjects performed this task in a seated position. This task was repeated three times and the best performance was recorded as data. In some cases because of cognitive or motor limitations subjects were unable to complete this task.

In order to probe the subjects' impressions of the treatment effectiveness as well as the impact of the treatment on subjects' daily living, perceptual rating scales were completed by the subject and his/her spouse at the time of the voice data collection. Variables such as loudness,

monotonous voice and intelligibility were rated. Because of reduced visual acuity, task confusion, inability to read or lack of spousal participation, a number of subjects and over half of the spouses were unable to complete these scales.

In addition, as a preliminary probe of the impact of voice treatment on conversational speech characteristics and intelligibility, two professional speech pathologists blindly rated pre- and post-video tapes of a subgroup of these patients while they spoke in individual interview situations. One of the speech pathologists was familiar with the experimental subjects, the other speech pathologist was completely unfamiliar with these subjects. This subgroup of patients was representative of the experimental subject group as a whole in terms of pre-treatment speech disorder severity. Visual analogue rating scales were selected for all perceptual ratings because of their increased sensitivity when compared with other forms of scaling such as equal-appearing-intervals (Kempster, 1984).

Because of potential instabilities in measures of maximum performance (Kent et al., 1987) as well as the variability associated with Parkinson's patients performance, it was a critical aspect of this study that subjects were well-trained for these tasks and that they consistently generated maximum performance. The same experimenter collected all the data reported here and was careful to elicit maximum performance as determined by her clinical decision. In addition, for the experimental subjects, clinical treatment data were collected daily on the measures of maximum duration sustained phonation and maximum fundamental frequency range using a timer and visipitch respectively, inorder to obtain a general gauge of the representativeness of the experimental data. It was expected that the daily clinical measures would serve as a general indicator of consistency in pre- to post-treatment changes across experimental and clinical settings.

All experimental data were collected before the month-long treatment program (pre-) and after the treatment program (post-) for the experimental subjects and at one month intervals for the control subjects.

Data analysis

To obtain measures of duration of maximally sustained vowel phonation, each phonation was input to a digital oscilloscope (Data Precision Model 611, Data 6000). Cursors were hand positioned to mark the monitor-displayed zero crossing preceding the first negative-going peak at the onset and the zero crossing following the final positive-going peak at the offset of each vowel. To obtain measures of maximum fundamental frequency range, each attempt at maximum high or low phonation was filtered above the predicted fundamental frequency and input to the digital oscilloscope and frequency in Hertz (Hz) was read. The maximum high and low were then converted to express the maximum range in semitones (ST). Measures of fundamental frequency during reading and fundamental frequency variability during reading (semitone standard deviation; STSD) were obtained by digitizing the reading of the "Rainbow Passage" at 10,000Hz and applying the program UFO (Horii, 1987) on a PDP 11-34 computer. Standard procedures (Boeckstyns and Backer, 1989) for analysis of visual analogue scales were used to obtain perceptual data.

Statistical analysis

Because the distributions of these acoustic and physiologic data did not satisfy the assumptions of normalacy, data were analyzed with the Wilcoxon signed-ranks non-parametric test.

Results

Reliability

Intra and interexaminer measurement reliability for repeated measures of maximum duration sustained vowel phonation and maximum fundamental frequency range are reflected in correlation coefficients ranging from .94 to .99.

Acoustic variables

Because it has been recommended that maximum performance variables (sustained vowel duration and fundamental frequency range) be studied with caution (Kent et al., 1987), these data were considered in two ways for each subject: the single maximum score for each variable and the mean of the three best scores for each variable in pre- and post-treatment conditions.

Maximum duration sustained vowel phonation (seconds). Maximum and mean duration data for the variable maximum duration sustained vowel phonation for the pre- and post- conditions are summarized in Table 5. Subjects in the experimental group evidenced a pre- to post-treatment

Table 5.

Means and standard deviations for maximum vowel duration (sec), mean maximum vowel duration (sec), maximum fundamental frequency range (ST) and mean maximum fundamental frequency range (ST) for the experimental treatment group pre- and post- treatment and a control group.

			Trial	
Measure	Subjects		pre-treatment	post-treament
Maximum vowel duration(s)	Experimental (n=40)	M	21.63	25.95
		SD	9.43	9.96
	Control (n=12)	M	23.36	23.56
		SD	6.91	6.32
Mean maximum vowel duration(s)	Experimental (n=40)	M	19.86	24.07
		SD	8.74	9.47
	Control (n=12)	M	21.89	21.86
		SD	6.24	5.48
Maximum fundamental frequency range (ST)	Experimental (n=38)	M	23.41	27.38
		SD	7.18	7.55
	Control (n=12)	M	28.16	29.12
		SD	4.17	5.70
Mean maximum fundamental frequency range (ST)	Experimental (n=38)	M	20.91	25.31
		SD	6.93	7.30
	Control (n=12)	M	26.05	26.52
		SD	4.06	5.34

median increase of 17% (ranging from -23% to 190%) and 26% (ranging from -26% to 214%) on maximum and mean maximum vowel duration respectively; subjects in the control group evidenced a median increase of 3% (ranging from -16% to 22%) and 1.5% (ranging from -17% to 23%) on these variables. Wilcoxon signed-ranks test revealed significant pre- to post-treatment differences on the variables maximum duration ($p < .05$) and mean maximum duration ($p < .05$) for the experimental group but not for the control group. Pre- to post-treatment percent change for the variable mean maximum vowel duration for the experimental and control subjects is plotted in Figure 1.

In order to assess the representativeness of these experimental data, a Wilcoxon signed-ranks test was calculated on pre- to post- treatment changes measured in the daily clinical treatment situation on the variable maximum duration sustained vowel phonation for 29 of the 40 experimental subjects. Consistent with the experimental data, there was a statistically significant difference ($p < .05$) between these pre-treatment data (mean = 17.17 sec; range = 4 to 31 sec) and post-treatment data (mean = 25.7 sec; range = 12 to 40 sec).

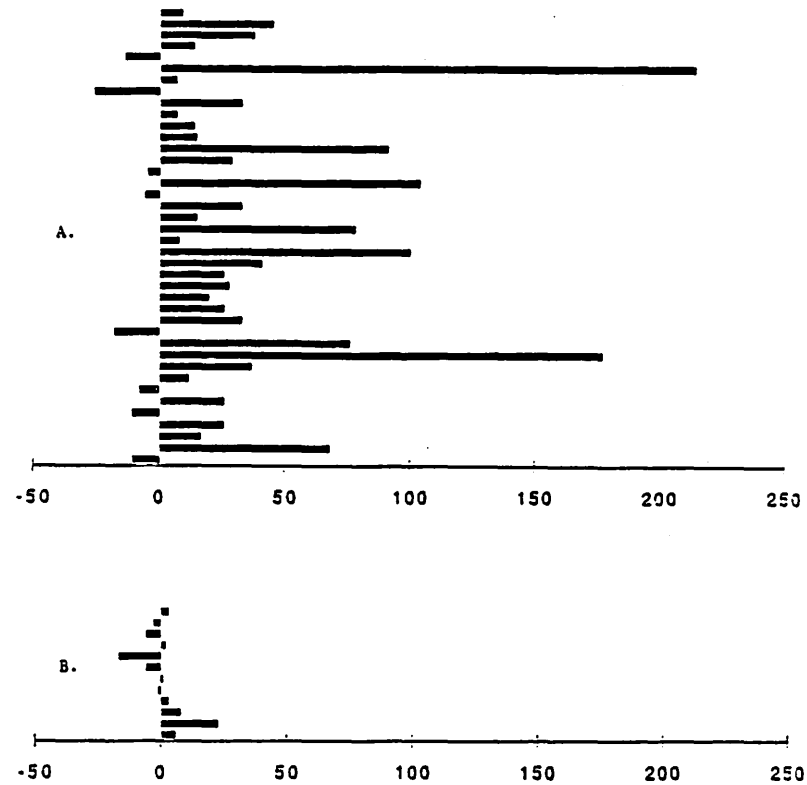

Figure 1. Percent change pre- to post- treatment on mean maximum duration sustained phonation for the experimental (A) and control (B) groups.

Maximum fundamental frequency range (ST). Maximum and mean fundamental frequency range data for the variable maximum fundamental frequency range for the pre- and post- conditions are summarized in Table 5. Subjects in the experimental group evidenced a pre- to post-treatment median increase of 12.5% (ranging from -43% to 182%) and 14% (ranging from -36% to 175%) on maximum and mean maximum fundamental frequency range respectively; subjects in the control group evidenced a median increase of 3% (ranging from -17% to 38%) and .5% (ranging from -14% to 50%) on these variables. Statistical analysis revealed significant pre- to post-treatment differences on the variables maximum range ($p < .05$) and mean maximum range ($p < .05$) for the experimental group but not for the control group. Pre- to post- percent change for the mean fundamental frequency range data for the experimental and control subjects is plotted in Figure 2.

In order to assess the representativeness of these experimental data, Wilcoxon signed-ranks analysis was carried out between pre- and post- treatment maximum phonation range data collected in the daily clinical treatment situation. Consistent with the experimental data, there was a statisti-

cally significant difference ($p < .05$) between the pre-treatment daily clinical data (mean = 28.6 ST; range = 15.1 to 47.2 ST) and post-treatment daily clinical data (mean = 34.1 ST; range = 19.2 to 53.8 ST).

Fundamental frequency variability (semitone standard deviation). Mean fundamental frequency variability data for reading of the "Rainbow Passage" for the pre- and post- conditions are summarized in Table 6. Male subjects in the experimental group (n=23) evidenced a pre- to post-treatment median increase of 12% (ranging from -29% to 80%); male subjects in the control group (n=10) evidenced a median decrease of -5.5% (ranging from -15% to 4%) on this variable. Female subjects in the experimental group (n=8) evidenced a pre- to post-treatment median increase of .5% (ranging from -31% to 31%); female subjects in the control group (n=2) evidenced a median increase of 4.5% (ranging from .01% to 9%).

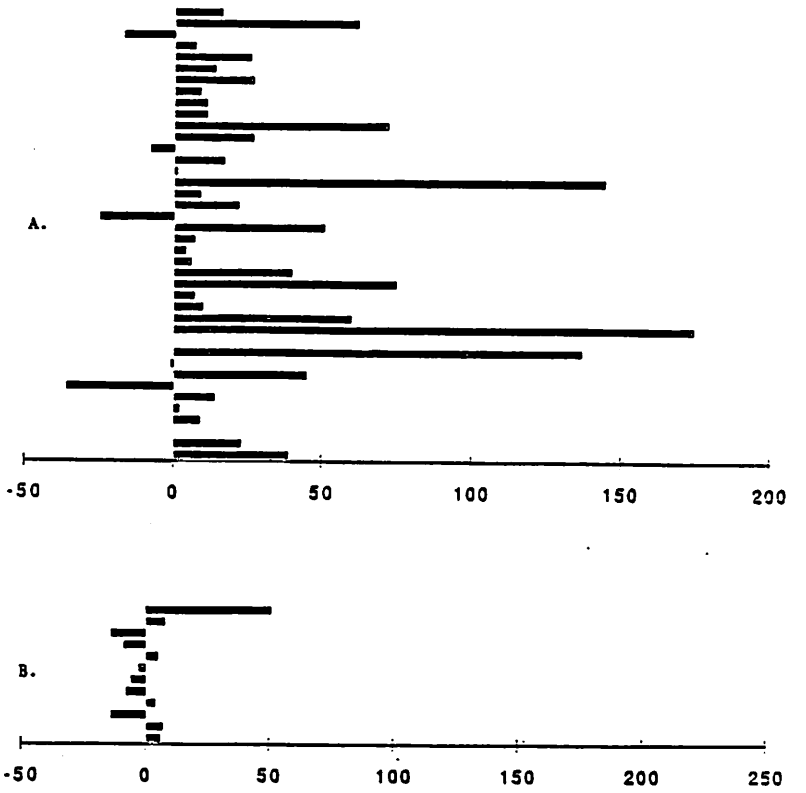

Figure 2. Percent change pre- to post- treatment on mean maximum fundamental frequency range for the experimental (A) and control (B) groups.

Table 6.

Mean fundamental frequency variation (STSD) and mean fundamental frequency (Hz) during reading from thirty-one experimental and twelve control subjects pre- and post-treatment.

Measure	Subjects	Sex		Trial: pre-treatment	Trial: post-treatment
Mean fundamental frequency variation during reading (STSD)	Experimental	M (n=23)	M	1.74	1.93
			SD	.70	.81
		F (n=8)	M	2.24	2.24
			SD	.68	.68
	Control	M (n=10)	M	1.85	1.76
			SD	.38	.46
		F (n=2)	M	2.07	2.18
			SD	.57	.71
Mean fundamental frequency during reading (Hz)	Experimental	M (n=23)	M	127.50	134.50
			SD	18.80	21.60
		F (n=8)	M	186.10	191.30
			SD	11.80	14.20
	Control	M (n=10)	M	130.40	121.50
			SD	26.30	23.10
		F (n=2)	M	196.60	207.70
			SD	5.10	2.30

Statistical analyses revealed a significant pre-to post-treatment difference on the variable fundamental frequency variability for the males in the experimental group ($p < .05$) but not for the experimental females or for any members of the control group. Pre- to post-treatment percent change on this measure for the experimental and control subjects is plotted in Figure 3.

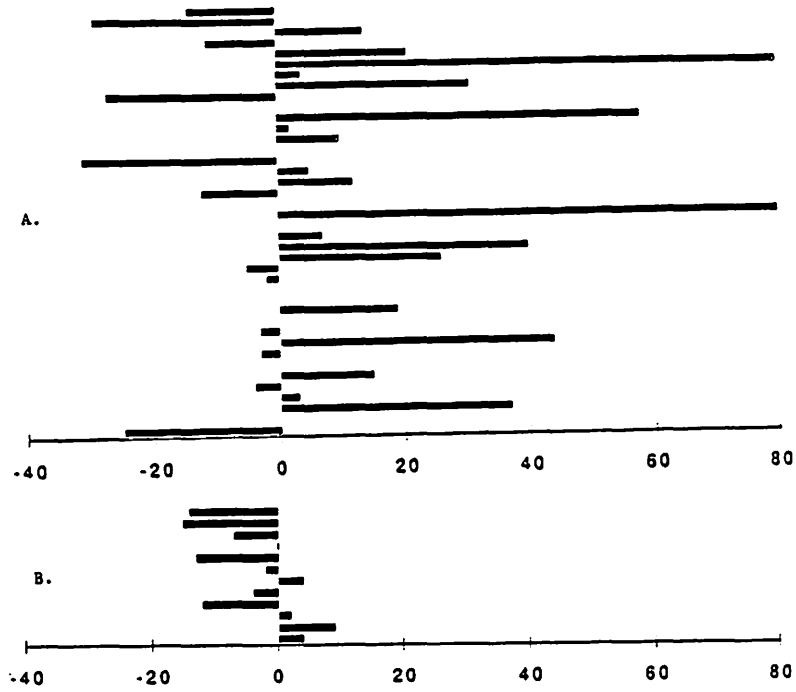

Figure 3. Percent change pre- to post- treatment on semitone standard deviation (STDS) fort he experiemtnal (A) and control (B) groups.

Mean fundamental frequency (hertz). Mean fundamental frequency data for reading of the "Rainbow Passage" for the pre- and post- conditions are summarized in Table 6. For the experimental group, median percent increase for males (n=23) was 5.0% (ranging from -18% to 26%); median percent decrease for males (n=10) in the control group was -5.5% (ranging from -29% to 15%). The median percent increase for experimental females (n=8) was 1% (ranging from -4% to 16%); median percent change for control females (n=2) was 6% (ranging from 5% to 7 %). Statistical analysis revealed a significant pre- to post-treatment difference on the variable mean fundamental frequency for the males in the experimental group ($p < .05$) but not for the experimental females or any of the control group.

Perceptual ratings

Perceptual ratings by the experimental and control subjects and speech pathologists for the pre- and post- conditions are summarized in Table 7.

Pre- to post-treatment self-ratings for loudness, monotonous voice and intelligibility were carried out by 27 of the 40 experimental subjects. The median percent improvement for self-rated loudness was 12% (ranging from -41% to 47%); this increase was statistically significant ($p < .05$). Other pre- to post- median self-ratings for the experimental subjects were 1.5% (ranging from -47% to 38%) for monotonous voice and 2% (ranging from -23 to 51%) for intelligibility; these were not statistically significant differences. The pre- to post-treatment median percent change on these variables for the twelve control subjects was 6% (ranging from -25% to 47%) for loudness, -.11.0% (ranging from -29.0% to 16.0%) for monotonous voice and 11.0% (ranging from -9.0% to 27.0%) for intelligibility. Only the pre- to post- increase in the control subjects' rating of intelligibility was statistically significant ($p < .05$).

The spouses of seventeen of the forty experimental subjects rated these variables in the pre- and post-treatment conditions. The pre- to post- treatment median percent change in the ratings of the seventeen spouses and the corresponding experimental subjects was 0% (ranging from -29% to

42%) and 13% (ranging from -41% to 38%) for loudness, 1.5% (ranging from -13% to 68%) and 1.5% (ranging from -39% to 38%) for monotonous voice and 13% (ranging from -18% to 59%) and 5% (ranging from -23% to 34%) for intelligibility respectively. The pre- to post-treatment spouses ratings of intelligiblity and the patients self-ratings of loudness were the statistically significant differences in this subset of perceptual data ($p < .05$).

Table 7.

Perceptual ratings by patients, spouses and corresponding patients, familiar and unfamiliar speech pathologists and the control group on the variables loudness, monotone and inteligibility pre- and post- treatment.

		Variable					
		Loudness		Monotone		Intelligibility	
Rater		pre	post	pre	post	pre	post
patient (self-rating) (n=27)	M	47.0	55.8	53.7	53.6	57.3	63.7
	SD	28.7	21.8	23.3	25.1	20.4	18.3
spouse (n=17)	M	45.5	53.5	60.1	69.5	49.5	64.7
	SD	22.9	22.5	24.5	21.9	23.3	19.4
corresponding patient (n=17)	M	43.4	54.0	52.0	55.6	52.0	58.6
	SD	30.5	24.6	26.3	25.6	20.2	16.6
familiar SP	M	39.2	65.9	28.3	52.9	45.2	71.2
	SD	21.7	26.0	18.5	24.3	21.3	16.9
unfamiliar SP	M	32.9	64.0	30.4	62.8	55.6	75.1
	SD	17.1	20.9	15.5	23.6	16.9	18.9
controls (n-12)	M	53.8	61.7	71.2	64.1	58.8	69.6
	SD	25.9	21.6	24.4	23.6	15.6	15.8

Two speech pathologists rated loudness, monotonous voice and intelligibility from the conversational speech of a subgroup of sixteen of these experimental patients. Because of the differences in familiarity with these patients, the ratings of the individual speech pathologists were considered separately. The median percent change in the ratings of the familiar and unfamiliar speech pathologists across sixteen subjects were respectively 26.5% (ranging from -7% to 50%) and 31.5% (ranging from -2% to 69%) for loudness, 26% (ranging from -20% to 58%) and 30% (ranging from 3% to 77%) for monotonous voice and 24% (ranging from 7% to 47%) and 20.5% (ranging from -24% to 71%) for intelligibility. All of these pre- to post-treatment ratings by the speech pathologists were statistically significantly different ($p < .05$).

Vital capacity

Forced and slow vital capacities for the pre- and post-conditions for 28 and 27 of the 30 male experimental subjects and 10 and 8 of the 10 female experimental subjects are summarized in Table 8. The median percent change for the males was -.51% (ranging from -47% to 34%) and 3% (ranging from -17% to 33%) for forced and slow vital capacities respectively. For the females, the median percent change was -1.5% (ranging from -11% to 35%) and 10% (ranging from -6% to 138%) for forced and slow vital capacities respectively. There were no statistically significant pre- to post- treatment differences on these variables.

Table 8.
Means and standard deviations for forced and slow vital capacities for the experimental treatment group pre- and post- treatment.

Measure	Sex		Trial: pre-treatment	Trial: post-treatment
Forced vital capacity (L)	M (n=28)	M	3.82	3.82
		SD	.80	.88
	F (n=10)	M	2.63	2.67
		SD	.70	.55
Slow vital capacity (L)	M (n=27)	M	3.65	3.77
		SD	.88	.77
	F (n=8)	M	2.45	2.69
		SD	.85	.71

Discussion

The data reported here support the effectiveness of an intensive program of voice therapy on the improvement of vocal function in patients with Parkinson's disease. Statistically significant differences were measured between pre- and post-treatment on the variables maximum duration sustained vowel phonation, maximum fundamental frequency range, and mean fundamental frequency and fundamental frequency variation in reading for the treated experimental group but not for the untreated control group. Improvement in perceptual measures of voice accompanied these changes in objective acoustic measures as well.

At this point one can only speculate on the underlying physiologic changes accompanying successful treatment. Given the focus and rationale for treatment, the most logical physiological changes are increases in vocal fold adduction and respiratory support during speech and increased range of motion of the cricothyroid muscle. Preliminary videoendoscopic analysis supports increased vocal fold adduction in one patient following treatment. In addition, measures of the abduction quotient (Titze, 1984) demonstrated increased vocal fold adduction after voice treatment in eight patients who were successfully treated (Ramig, Fazoli, Scherer and Bonitati, 1990). We are currently more thoroughly investigating characteristics of vocal fold adduction and respiratory support accompanying these pre- to post-treatment acoustic and perceptual findings.

Of particular interest in these data are the significant pre- to post-treatment differences for mean fundamental frequency and fundamental frequency variability during reading observed in the male experimental subjects. Unlike the other variables which had been directly trained in therapy, these variables apparently improved in these patients as a secondary effect of training. Increased mean fundamental frequency of the magnitudes observed here could be a result of increased respiratory driving pressure as well as increased vocal fold tension perhaps as a result of increased adduction. Increased intonation could be a generalization of the effect of improved maximum fundamental frequency ranges. One can speculate that as patients increased maximum fundamental frequency ranges, it became easier for them to generate greater intonation because the range of motion of the

cricothyroid muscle had been stimulated and increased. Another speculation which could explain the increase in both of these variables is that as these patients improved their speech, their attitude and affect also improved and this was reflected in the increased pitch and intonation of their voices. The explanation for these findings as well as the lack of a similar effect in the experimental female data is currently being investigated.

The perceptual data presented here were considered as a probe of the impact of therapy-related changes on functional communication. There are clearly many issues to consider in the interpretation of these perceptual data (e.g., limited sample, bias, patient and spouse problems with comprehension of the scaling procedure, placebo effect as reflected in control groups self-rating of intelligibility). However, the findings, especially those from the speech pathologists, are in general agreement with the subjective impressions of the experimenter, other professionals and caregivers who observed that many patients were easier to understand and that they participated in and initiated conversation more frequently after therapy. It is critical in future treatment efficacy work that the relationship between improved perceptual aspects of speech and acoustic and physiologic measures be well-defined in order to clarify the most efficient targets for therapy. The relationship among improved perceptual aspects of speech (especially intelligibility) and acoustic and physiologic correlates must be clarified if treatment is to be most effective and efficient. Our preliminary (Ramig, in press) and ongoing work continues to address this relationship.

This approach to speech therapy for patients with Parkinson's disease differs from previous approaches in a number of ways. The focus was on increased phonatory effort. Patients were stimulated to "speak loud" by taking deeper breaths and using more effortful adduction. This provided a relatively simple tasks for patients with almost immediate translation into improved functional communication. Parkinson's disease patients have a well-established difficulty in simultaneously executing two different movements (Beneckel, Rothwell, Dick, Day and Marsden,1986; Yanagisawa et al., 1989); this may be one explanation why previous speech therapy approaches for Parkinson's disease patients which focused on multiple levels of speech production (articulation, rate, intonation) have not been consistently successful. The tasks stimulated in this therapy program were maximum effort tasks. Patients were encouraged to push themselves to a new phonatory effort level. In fact many patients would comment that they felt as if they were "shouting" when they used the new louder speech. One could speculate that the effort level these patients used in order to increase loudness to a level of intelligible conversational speech was similar to what they had previously (pre-Parkinson's disease) generated when actually shouting. It is possible that the maximum effort treatment pushed these patients to a new effort level and the intensive (daily) treatment, stimulated this level until it became habitual. Furthermore, while respiratory and phonatory activities were the focus of therapy, it was apparent by clinical observation that articulatory precision increased in these patients post-treatment as well. These findings are in agreement with Scott and Caird (1983) who reported that phonation in particular was a prerequisite for intelligible, effective speech and as voice production improved, other aspects of speech also improved, as well as Rubow and Strand (1985) who reported a positive relationship between improved loudness and articulatory skills in one Parkinson's disease patient. These findings suggest that multiple levels of speech production can benefit from a single treatment focus on increased phonatory effort. Ongoing research will assess these speculations.

These data should be viewed as a first step in the process of evaluating the efficacy of voice treatment for patients with Parkinson's disease. They support the effectiveness of intensive voice

therapy on improving select measures of phonation in this group of patients. However, a number of critical issues must be addressed in relation to these findings. Documentation of underlying physiologic changes accompanying successful treatment is a critical element in defining treatment efficacy. The interactive roles of the respiratory and phonatory systems in treatment-related change must be understood. We are currently measuring respiratory system excursions, subglottal pressure and intraoral pressures and oral airflow in addition to acoustic, electroglottographic and perceptual measures to more clearly define physiologic changes accompanying successful treatment.

The long-term carryover of these phonatory changes must be studied. A major problem in establishing speech and voice treatment efficacy for patients with Parkinson's disease has been the reported lack of carryover. The consensus has been that while patients can make some improvements in the speech therapy room, they do not carry these improvements outside of the therapy room or maintain them for long periods of time. Allan (1970) suggested that a rapid deterioration in speech occurs in Parkinson's disease patients when traditional speech therapy is discontinued. On the other hand, Scott and Caird (1983) noted improvement up to three to four months following intensive voice therapy. In long-term studies of physical and occupational therapy programs for Parkinson's patients, up to a year of functional status maintenance including significant improvement in psychological well-being have been observed (Palmer et al.,1986; Gauthier et al., 1987). Our pilot work (Ramig, Mead and DeSanto, 1988) supports maintenance and continued improvement in speech and voice production with minimal (weekly, progressing to monthly) follow-up voice therapy. The combination of speech, voice, physical and occupational therapy data suggest that the benefits of short-term therapy may be maintained on a long-term basis. We are currently evaluating the long-term carryover of the changes reported in this research. We have developed carryover programs which involve treatment sessions provided by videotape (VCR) in th home (Ramig, Bonitati and Winholtz, 1989) and extensive roles for the caregiver and/or spouse. While conventional wisdom suggests that the changes Parkinson's disease patients make in the speech treatment room disappear when they return home, our findings (based upon the ratings and reports of spouses and caregivers) support carryover of the changes reported here outside of the treatment room. Future work will investigate more systematically the impact of increased phonatory effort on Parkinson's patients functional communication.

The traditional format for delivery of speech treatment services to patients with Parkinson's disease has been as an independent therapy offered two or three times a week. The data reported here were collected from patients who have participated in a daily multidisciplinary treatment program which involved various therapies (physical, occupational, exercise, counseling) in addition to voice therapy for as often as 4-5 hours a day five days a week for one month. It could be hypothesized that patients who participated in a daily multidisciplinary treatment program would maximize their chance for successful rehabilitation because of the continuous stimulation by multiple treatments. Our pilot data (Ramig et al., 1990) collected from patients who participated in the intensive voice treatment program in the absence of a multidisciplinary program support comparable improvement in post-treatment measures of voice production. We are currently evaluating the role of the multidisciplinary treatment as well as the intensiveness of treatment in the changes observed following voice therapy.

Identification of patient suitability or prognostic factors for voice treatment success is another issue of importance in studies of efficacy of voice treatment for Parkinson's disease. Motiva-

tion, dementia, stage of disease and depression are some of the patient variables which have been suggested as related to general therapy success (Duvosin, 1984; Mayeux et al., 1986; Gauthier et al., 1987). Our preliminary data support voice treatment success across various stages of disease and degrees of dementia (Bonitati, Ramig and Beck, 1989). However, we have observed that lack of motivation and depression may be variables that limit voice treatment success with Parkinson's patients (Ramig et al., 1990). We are currently systematically evaluating these possible prognostic factors in relation to short and long-term therapy success. As Tompkins, Jackson and Schultz (1990) have recently suggested, formulating prognoses is one of the primary tasks facing speech-language pathologists.

Acknowledgements:

This research was made possible by the generosity and support of the family of Mrs. Lee Silverman which was the impetus for the work in this area. Deep appreciation is extended to the patients who participated in this study for their contributions and continuous challenges. The authors gratefully acknowledge the contributions of Ronald C. Scherer, Jon H. Lemke and Carl K. Brown of the Biostat Core, Center Grant University of Iowa NIH #1 P60 DC00976, Katy H. Fazoli, Kristin K. Larson, Mary R. Mailand, Stephanie M. Countryman, Patricia A. Benjamin and Roberta Borne.

The data presented here were supported in part by OE-NIDRR grant #H133G00079. Manuscript preparation was supported in part by NIH RO1 DC01150-01. Data analysis was supported by Center Grant University of Iowa NIH#1 P60 DC00976.

Bibliography

Allan, C. M. (**1970**). "Treatment of non-fluent speech resulting from neurological disease: treatment of dysarthria," Brit. J. Dis. Comm. **5**, 3-5.

Aronson, (**1985**). *Clinical Voice Disorders*. (Thieme-Stratton, New York).

Beneckel, R., Rothwell, J. C., Dick, J. R. R., Day, B. L., and Marsden, C. D. (**1986**). "Performance of simultaneous movements in patients with Parkinson's disease," Brain **109**, 739-757.

Blonder, L. X., Gur, R., and Gur, R., (**1989**). "The effects of right and left hemiparkinsonism on prosody," Brain and Language **36**, 193-207.

Bonitati, C., Ramig, L., and Beck, J. (**1989**). "Parkinson's disease with severe dementia: effectiveness of intensive voice therapy," ASHA **31**, 118.

Boshes, B. (**1966**). "Voice changes in parkinsonism," J. Neurosurg. **24**, 286-288.

Canter, G. J. (**1965**). "Speech characteristics of patients with Parkinson's disease: II. Physiological support for speech," J. Speech Hear. Dis. **30**, 44-49.

Critchley, E. M. R. (**1981**). "Speech disorders of Parkinsonism: a review," J. Neurol. Neurosurg. Psychiatry **44**, 751-758.

Darley, F. L., Aronson, A., and Brown, J. (**1969a**). "Differential diagnostic patterns of dysarthria," J. Speech Hear. Research **12**, 246-249.

Darley, F. L., Aronson, A., and Brown, J. (**1969b**). "Clusters of deviant speech dimensions in the dysarthrias," J. Speech Hear. Research **12**, 462-496.

Duvosin, R. C. (**1984**). *Parkinson's Disease: a guide for patient and family.* (2nd ed.) (Raven Press, New York).

Fairbanks, G. (**1960**). *Voice and Articulation Drillbook.* (Harper and Brothers, New York).

Forrest, K., Weismer, G., and Turner, G. (**1989**). "Kinematic, acoustic and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults," J. Acoust. Soc. Am. **5**(6), 2608.

Froeschels, E., Kastein, S., and Weiss, D. A. (**1955**). "A method of therapy for paralytic condition of the mechanisms of phonation, respiration, and glutination," J. Speech Hear. Dis. **20**, 365-370.

Gauthier, L., Dalziel, S., and Gauthier, S. (**1987**). "The benefits of group occupational therapy for patients with Parkinson's disease," Am. J. Occupat. Ther. **41**, 360-365.

Green, H. C. L. (**1980**). *The Voice and Its Disorders.* (Pitman Medical, London).

Hansen, D. G., Gerratt, B. R., and Ward, P. H. (**1984**). "Cinegraphic observations of laryngeal function in Parkinson's disease," Laryngoscope **94**, 348-353.

Hirose, H., Kiritani, S., Ushijima, T., and Sawashima, M. (**1981**). "Patterns of dysarthric movements in patients with Parkinsonism," Folia Phoniatr. **33**, 204-215.

Hoberman, S. G. (**1958**). "Speech techniques in aphasia and Parkinsonism," J. Mich. St. Med. Soc. **57**, 1720-1723.

Hoehn, M. (**1983**). "Parkinsonism treated with levodopa: progression and mortality," J. Neuro. Trans. (Suppl.) **19**, 253-264.

Hoehn, M., and Yahr, M. (**1967**). "Parkinsonism: onset, progression, and mortality," Neurol. **17**, 427.

Hollien, H., Dew, D., and Philips, P. (**1971**). "Phonational frequency ranges of adults," J. Speech Hear. Research **14**, 755-760.

Horii, Y. (**1987**). Personal Communication.

Hornykiewicz, O. (1966). *Metabolism of brain dopamine in human parkinsonism: neurochemical and clinical aspects*, edited by E. Costar, L. Cote, and N. Yahr (Raven Press, New York).

Hunker, C. J., Abbs, J. H., and Barlow, S. M. (1982). "The relationship between Parkinsonian rigidity and hypokinesia in the orofacial system: a quantitative analysis," Neurology 32(7), 749-754.

Kempster, G. B. (1984). "A multidimensional analysis of vocal quality in two dysphonic groups," An unpublished doctoral dissertation. Northwestern University, Evanston.

Kent, R. D., Kent, J. F., and Rosenbek, J. C. (1987). "Maximum performance tests of speech production," J. Speech Hearing Disorders 52, 367-387.

Klawans, H. L. (1988). "Natural history of Parkinson's disease," A paper presented at the ninth international symposium on Parkinson's. Jerusalem. United Parkinson's Foundation Newsletter, 3:1.

Kurland, L. T., Kirtzker, L. T., Goldberg, I. D., Choiz, N. W., and Williams, G. (1973) *Epidemiology of neurologic and sense organ disorders*, edited by L. T. Kurland, J. F. Kirtze, and Goldberg (Harvard University Press, Cambridge).

Larson, K. L., Ramig, L., and Scherer, R. (1988). "Acoustic analysis of voice: the on-off effect of medication in Parkinson's disease," A paper presented at the Clinical Dysarthria Conference (San Diego).

Leanderson, R., Meyerson, B. A., and Persson, A. (1972). "Lip muscle function in parkinsonian dysarthria," Acta Otolaryngol. 74, 354-357.

Logemann, J. A., Fisher, H. B., Boshes, B., and Blonsky, E. R. (1978). "Frequency and occurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson's patients," J. Speech Hear. Disorders 42, 47-57.

Martilla, R. J., and Rinne, V. K. (1979). "Changing epidemiology of Parkinson's disease: predicted effects of levodopa treatment," Acta Neurol. Scand. 59, 80-87.

Mayeux, R., Williams, J. B. W., Stern, Y., and Cote, C. (1986). "Depression and Parkinson's disease," in *Adv. Neurol.*, edited by M.D. Yahr and K. J. Bergmann Adv. Neurol. (New York: Raven Press, 241-250).

Morley, D. E. (1955). "The rehabilitation of adults with dysarthric speech," J. Speech Hear. Dis. 20, 58-64.

Netsell, R., Daniel, B., and Celesia, G. G. (1975). "Acceleration and weakness in Parkinsonian dysarthria," J. Speech Hear. Dis. 40, 170-178.

Palmer, S. S., Mortimer, J. A., Webster, D. D., Bistevins, R., and Dickinson, O. L. (**1986**). "Exercise therapy for Parkinson's disease," Arch. Phys. Med. Rehab. **67**, 741-745.

Ramig, L. A., Fazoli, K., Scherer, R., and Bonitati, C. (**1990**). "Changes in phonation of Parkinson's disease patients following voice therapy," A paper presented at the Clinical Dysarthria Conference, San Antonio, TX.

Ramig, L. A., Mead, C., and DeSanto, L. (**1988a**). "Voice therapy and Parkinson's disease," ASHA **30**(10), 128.

Ramig, L. O., Scherer, R. C., Klasner, E. R., Titze, I. R., and Horii, Y. (**1990**). "Acoustic analysis of voice in amyotrophic lateral sclerosis: a longitudinal case study," J. Speech Hear. Dis. **55**(1), 2-14.

Ramig, L. O. (**in press**). "The role of phonation in speech intelligibility: a review and preliminary data from patients with Parkinson's disease," in *Intelligibility in Speech Disorders: Theory, Measurement and Management*, edited by R. Kent (John Benjamin, Amsterdam).

Ramig, L. O., Bonitati, C., and Winholtz, W. (**1988**). "Speech therapy: Parkinson's disease," a videotape produced by The Recording and Research Center of The Denver Center for the Performing Arts.

Reich, A. R., Frederickson, R. R., Mason, J. A., and Schlauch, R. S. (**1990**). "Methodological variables affecting phonational frequency range in adults," J. Speech Hear. Disorders **55**, 124-131.

Robertson, S., and Thompson, F. (**1984**). "Speech therapy in Parkinson's disease: a study of the efficacy and long-term effect of intensive treatment," British Journal of Disorders of Communication **19**, 213-224.

Rubow, R., and Strand, E. (**1985**). "A microcomputer-based wearable biofeedback device to improve transfer of treatment in Parkinsonian dysarthria," J. Speech Hear. Dis. **50**, 178-185.

Sarno, M. T. (**1968**). "Speech impairment in Parkinson's disease," Arch. Phys. Med. Rehab **49**, 269-275.

Scott, S., and Caird, F. L. (**1983**). "Speech therapy for Parkinson's disease," J. Neurol. Neurosurg. Psych. **46**, 140-144.

Selby, G. (**1986**). "Parkinson's disease," in *Handbook of Clinical Neurology*, **6**, edited by P.J. Winken and G. W. Bruyn (North Holland Publishing Company, Amsterdam).

Streifler, M. and Hofman, S. (**1984**). "Disorders of verbal expression in Parkinsonism," in *Adv. Neurol.* edited by R.G. Hassler and J.F. Christ, **40**, 385-393.

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

Tompkins, C. A., Jackson, S. T., and Schultz, R. (1990). "On prognostic research in adult neurogenic disorders," J. Speech Hear. Res. 33(2), 398-401.

Weiner, W. J., and Singer, C. (1989). "Parkinson's disease and nonpharmacologic treatment programs," J. Amer. Geriat. Soc. 37, 359-363.

Weismer, G. (1984). "Articulatory characteristics of parkinsonian dysarthria: segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination," in *The Dysarthrias: Physiology, Acoustics, Perception, Management*, edited by M. McNeil, J. Rosenbek, A. Aronson (College Hill Press, New York), 101-130.

Yanagisawa, N., Fujimato, S., and Tamaru, F. (1989). "Bradykinesia in Parkinson's disease: disorders of onset and execution of fast movement," Eur. Neurol (suppl.) 29, 19-28.

Table 2.

Age, sex, stage, medication and primary speech symptoms for forty experimental subjects.

Experimental Subjects	Age	Sex	Stage	Medication	Primary Speech Symptoms
1	76	M	II	Sinemet Artane	reduced loudness
2	75	M	IV	Sinemet Meladryl	breathy voice harsh voice
3	73	M	III	Sinemet Artane Symmetrel	reduced loudness hoarse voice imprecise articulation
4	81	F	III	Sinemet Isosorbide Dinitrate Cardizem	reduced loudness harsh voice palilalia
5	59	M	II	Sinemet	hoarse voice harsh voice
6	77	M	II	Artane	lowered pitch harsh voice
7	72	M	III-IV	Sinemet	reduced loudness hoarse voice breathy voice
8	70	M	III	Sinemet Cogentin Symmetrel	reduced loudness monotone
9	73	M	II	Sinemet	reduced loudness
10	59	F	III	Sinemet	reduced loudness monotone hoarse voice breathy voice vocal tremor

Table 2. (Continued)

Experimental Subjects	Age	Sex	Stage	Medication	Primary Speech Symptoms
11	73	M	III	Levadopa Sinemet Artane	hoarse voice breathy voice vocal tremor
12	76	M	III	Sinemet	reduced loudness imprecise articulation palilalia
13	77	F	III	Sinemet Deprenyl	reduced loudness monotone harsh voice imprecise articulation
14	67	M	III	Sinemet	reduced loudness hoarse voice imprecise articulation
15	75	F	III	Levadopa Sinemet	reduced loudness monotone hoarse voice harsh voice vocal tremor increased rate
16	59	M	III	Sinemet	reduced loudness harsh voice imprecise articulation palilalia increased rate
17	53	M	III	Sinemet Parlodel	reduced loudness monotone imprecise articulation
18	73	M	III	Sinemet	reduced loudness

Table 2. (Continued)

Experimental Subjects	Age	Sex	Stage	Medication	Primary Speech Symptoms
19	75	M	IV	Sinemet Artane	reduced loudness vocal tremor breathy voice imprecise articulation increased rate
20	67	M	I	Sinemet Theo-dur	hoarse voice imprecise articulation
21	72	M	II	Sinemet Artane	reduced loudness breathy voice harsh voice imprecise articulation
22	86	M	III		imprecise articulation
23	76	M	II	Sinemet	reduced loudness imprecise articulation
24	55	M	II	Sinemet Artane Deprenyl	reduced loudness harsh voice
25	74	F	III	Sinemet Sinequan	reduced loudness harsh voice imprecise articulation pitch breaks
26	67	F	II	Sinemet Artane	lowered pitch hoarse voice increased rate
27	78	M	II	Sinemet	imprecise articulation
28	69	F	III	Synthroid Sinemet Symmetrel Periactin	

Table 2. (Continued)

Experimental Subjects	Age	Sex	Stage	Medication	Primary Speech Symptoms
29	73	F	II	Sinemet Diazoxide	reduced loudness palilalia
30	65	M	II	Sinemet Artane	reduced loudness hoarse voice harsh voice imprecise articulation
31	68	F	III	Sinemet Elavil	reduced loudness imprecise articulation
32	74	M	III	Artane Sinemet Symmetrel	reduced loudness imprecise articulation
33	73	F	III	Ativan Sinemet Halcion Cardizem Sinequan	vocal tremor harsh voice imprecise articulation
34	64	M	I	Sinemet Symmetrel	reduced loudness hoarse voice breathy voice imprecise articulation
35	73	M	III	Sinemet Symmetrel	reduced loudness vocal tremor harsh voice imprecise articulation
36	67	M	III	Symmetrel Artane	reduced loudness vocal tremor breathy voice imprecise articulation

Table 2. (Continued)

Experimental Subjects	Age	Sex	Stage	Medication	Primary Speech Symptoms
37	68	M	III	Sinemet	reduced loudness vocal tremor hoarse voice breathy voice
38	80	M	IV	Sinemet	wet-hoarse voice imprecise articulation
39	75	M	III	Sinemet	imprecise articulation hyponasal
40	77	M	III	Sinemet	reduced loudness hoarse voice breathy voice harsh voice imprecise articulation

Table 3

Age, sex, stage, medication and primary speech symptoms for twelve control subjects.

Experimental Subjects	Age	Sex	Stage	Medication	Primary Speech Symptoms
1	59	M	I	Sinemet	reduced loudness
2	64	F	I	Sinemet	reduced loudness
3	70	M	II		reduced loudness vocal tremor
4	68	M	II	Sinemet Artane	reduced loudness hoarse voice harsh voice
5	61	M	III	Sinemet	reduced loudness harsh voice imprecise articulation
6	82	M	II	Sinemet	imprecise articulation
7	79	M	I	Sinemet Levadopa	reduced loudness
8	67	M	II	Sinemet	reduced loudness monotone
9	63	F	II	Sinemet	reduced loudness palilalia
10	73	M	II	Sinemet	reduced loudness harsh voice imprecise articulation
11	68	M	III		reduced loudness
12	64	M	II	Sinemet Symmetrel	hoarse voice harsh voice imprecise articulation increased rate

NCVS Status and Progress Report - 1
June 1991, 87-97

Vocal Evaluation of Patients with Recurrent Respiratory Papillomatosis

Steven D. Gray, M.D.
Heather Dove, M.A.
Steve Alder, B.S.
Department of Otolaryngology - Head and Neck Surgery, The University of Utah
Julie Barkmeier, M.S.
David Druker, B.S
Department of Speech Pathology and Audiology, The University of Iowa
Cynthia Shive, R.N.
Duane VanDemark, Ph.D.
David Jones, Ph.D.
Department of Otolaryngology - Head and Neck Surgery, The University of Iowa

Abstract

Eight patients, (six adults, 2 children), following resolution of recurrent respiratory papillomatosis, received vocal testing to determine vocal function. Vocal testing included a perceptual evaluation; a voice frequency and intensity range profile; and a questionnaire regarding vocal dysfunction in various situations. Perceptual results indicate that the voices were abnormal in roughness, strain, and breathiness. Vocal frequency range was a mean of 15.8 semitones for adults (patient range 10-20 semitones) and the intensity range was 25.1 dB (range of 14-33 dB) compared to normals of 37 semitones and 40 dB. Responses to the questionnaire revealed most patients were generally satisfied with their voice, although they did not feel it functioned normally. Laryngeal stroboscopic examination at the time of vocal testing suggests the vocal dysfunction could be due to scarring of the vocal folds.

Introduction

Recurrent respiratory papillomatosis (RRP) is the most common benign laryngeal neoplasm in children. Adults may also be afflicted, but the disease tends to be less severe. These papillomas can cause stridor, hoarseness and respiratory distress necessitating surgical intervention to maintain an adequate airway. Functional impairment of the larynx may occur as a result of the RRP or the treatment. Crockett described complications as a result of surgical intervention for RRP. Glottic scarring and anterior and posterior glottic webbing occurred in patients who had frequent, repeated surgeries.[1] In most patients, the RRP eventually goes into remission.

Treatment of severe RRP, although not controversial, is not standard with respect to timing of surgical intervention. Treatment patterns range from performing microlaryngoscopy with papilloma removal only when necessary for establishing an airway to surgically removing an asymptomatic isolated papilloma. Justification of such divergent philosophies range from trying to reduce the number of surgeries to as few as possible (former example) to trying to "cure" the disease and "prevent" it from spreading (latter example). Little attention has been given to the long-term sequelae of these treatment patterns. This study was designed to achieve the following purposes: determine voice function in patients previously afflicted with severe RRP; ascertain if voice dysfunction impaired their life in family, social, or employment situations; and, if possible, determine the cause of voice dysfunction.

Methodology

Patient Information

Eight patients treated for papillomatosis participated in this investigation. The patients ranged in age from 7 to 43 years (see Table 1). The patients had been free of disease for an average of four years.

Table 1. - RRP Patient Description

Patient	Age	Gender	No. of Surgeries	Years Since Last Surgery
1	34	M	14	4
2	35	F	2	4
3	31	M	2	5
4	43	M	1	3
5	38	M	3	4
6	43	M	>50	3
7	10	M	49	2.5
8	7	M	10	4

Patients 1-5 developed laryngeal papilloma as adults, whereas patients 6-8 had laryngeal papilloma as children. Patient 2 developed papilloma during her pregnancy and subsequently experienced surgeries described as cord stripping. Patient 6 developed papilloma before the age of 2 and, by his report, spent the first ten years of his life hospitalized with a tracheostomy, requiring

papilloma removal every week or two. His disease went into remission around age 10 and he had one recurrence at age 30. Two children, now free of disease, are also included in the voice intensity and frequency range (VIFR) and stroboscopic analysis, but were not given the questionnaire. Nor were they included in the perceptual study because they represent too small a group, even though we feel they represent a very good voice outcome.

Perceptual Evaluation

Evaluators. Ten first-year graduate students specializing in speech-language pathology and having experience in voice disorders participated as listeners. All evaluators completed training for this research by listening to a voice disorder practice tape which illustrated the perceptual qualities they would later be asked to rate. Sample ratings were performed as part of the training.

Experimental Tapes. In addition to the six patients, four normal-voiced individuals were recorded reading the rainbow passage. A lapel microphone and a digital audio recorder were used. Recordings of the speakers reading the rainbow passage were then randomized. Each voice recording was presented three times on the listener tape, resulting in a total of 30 sets being presented to the listeners. The rating form is presented in Fig 1. Each item was rated by each listener following presentation of the recorded passage.

Voice sample #:
Rater:

VOICE QUALITY PERCEPTUAL RATING SCALE

Quality	Normal	Mild	Moderate	Severe
Clear vs rough/ gravelly quality				
Easy vs strained/ forced/tight quality				
Breathiness				
Aesthenia (weakness, lack of power)				
Audible breathing				
Degree of overall abnormality				

Pitch: LOW sev mod mild N mild mod sev HIGH

Voice pereeived as male _____ female _____

Age perceived as 5-10 10-15 15-20 20-25 25-30 30-35 35-40 40-45 45-50 50-55 55-60 60-65

Figure 1

Experiment. Evaluators were seated in a quiet lab and, using headphones, listened to recordings at a comfortable loudness level. Raters were presented the experimental tape of normal and papilloma voices randomly. They were instructed to listen to each recorded passage and to rate the speaker using the perceptual rating form previously described.

Analysis. The mean rating for each item for each speaker and across all speakers was determined. Mann-Whitney statistical analysis was used to determine if qualities of voice perception were different in the papilloma group when compared to the normal group. This analysis was needed since the data was not evenly distributed.

Voice Intensity and Frequency Range. Individual intensity and frequency range profiles were obtained by seating each patient in a soundproof booth and positioning an Epson sound-level meter at a distance of 30 centimeters and at a 45-degree angle from each patient's mouth. The A weighted scale was used during loudness measurements. In addition, a Yamaha keyboard was utilized to provide the patients with target frequencies.

Patients were instructed to sustain phonation of "ah" during the measurement of loudness at the target pitch. Initially, the experimenter requested the patient to sustain phonation of "ah" at a comfortable pitch. After the experimenter matched this pitch on the keyboard, the patient was instructed to sustain phonation of the same pitch as softly as possible. This was called "threshold phonation." The experimenter perceived this threshold as being reached when the patient produced a periodic voice signal, rather than a whispered breathiness. After measuring threshold and maximum loudness at a comfortable pitch, the experimenter provided higher and then lower target pitches using both the keyboard and modeled sustained phonation. Maximum loudness and threshold (minimum loudness) were reproduced for three tokens (trials). This method of examination continued until the envelope of intensity and frequency range had been determined. The patient's total intensity and frequency range was then plotted using Sigma Plot software, and mean intensity and frequency were determined from the three trials using Mystat (by Systat, Inc.) for graphic analysis.

Questionnaire. A questionnaire was used to determine epidemiological data and to assess patient satisfaction with voice and functional utility of the voice in social and workplace situations. It was 50 questions long and included reliability questions. Only those questions that elicited responses which helped us interpret the objective data and their results will be discussed.

Results

The perceptual study demonstrates vocal dysfunction in Roughness (p=0.02), Strain (p-0.01), Breathy (p=0.01), and Overall (p=0.03) voice quality in voices of previous papilloma patients. Voice weakness and audible breathing were not statistically different than normals. Table 2 shows individual rating scores for each patient and normals. Figure 2 displays this data showing the standard deviations compared to normals. Table 3 provides the average threshold and maximum loudness measures across all frequencies as well as each patient's frequency range in semitones. For all measures, a large degree of variability in intensity and

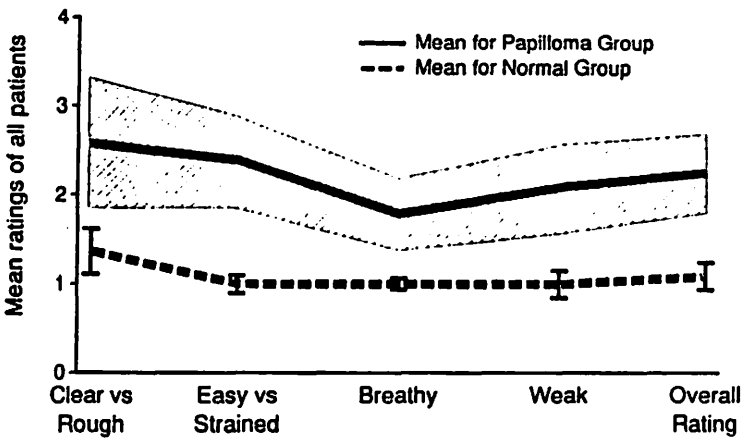

Figure 2. Perceptual evaluation of voices of papilloma patients as compared to normal voices. Broad bands indicate standard deviations for the voice quality items. Voice quality items are rated 1 if normal.

frequency range was demonstrated. The average loudness range across all patients was 25.1 dB with a range across all adult patients of 14-33 dB. The average normal voiced individual usually exhibits a 40 dB loudness range.[3,4] Examining individual measures of loudness revealed that the loudness range was moderately reduced in all patients. When measures were averaged across all of the adult patients, the average frequency range was 15.8 semitones with a range of 10-20 semitones. This average frequency range was markedly less than the average of 37 semitones found in normal voiced individuals.[2]

Table 2. - Perceptual Results

Patient	Clear vs. Rough	Easy vs. Strained	Breathy	Weak	Overall Rating
1	1.7 SD 0.4	1.13 SD 0.3	1.1 SD 0.3	1.1 SD 0.3	1.2 SD 0.4
2	2.5 SD 0.6	2.6 SD 0.6	1.6 SD 0.6	2.1 SD 0.8	2.3 SD 0.6
3	3.2 SD 0.4	3.2 SD 0.5	2.3 SD 0.7	2.7 SD 0.6	3.0 SD 0.3
4	1.5 SD 0.6	1.1 SD 0.3	1.0 SD 0.2	1.0 SD 0.0	1.2 SD 0.4
5	2.0 SD 0.3	1.7 SD 0.6	1.1 SD 0.3	1.4 SD 0.6	1.7 SD 0.6
6	3.4 SD 0.5	3.5 SD 0.5	2.9 SD 0.5	3.2 SD 0.6	3.4 SD 0.6
1-6 totals	2.6 SD 0.7	2.4 SD 1.0	1.8 SD 0.8	2.1 SD 1.0	2.3 SD 0.9
normals	1.4 SD 0.5	1.0 SD 0.2	1.0 SD 0.0	1.0 SD 0.3	1.1 SD 0.
P values	P=0.02	P=0.01	P=0.01	P=0.08	P=0.03

Perceptual results of voices of patients in remission from laryngeal papillomas compared to normal voices. Top number represents mean, lower number is standard deviation.

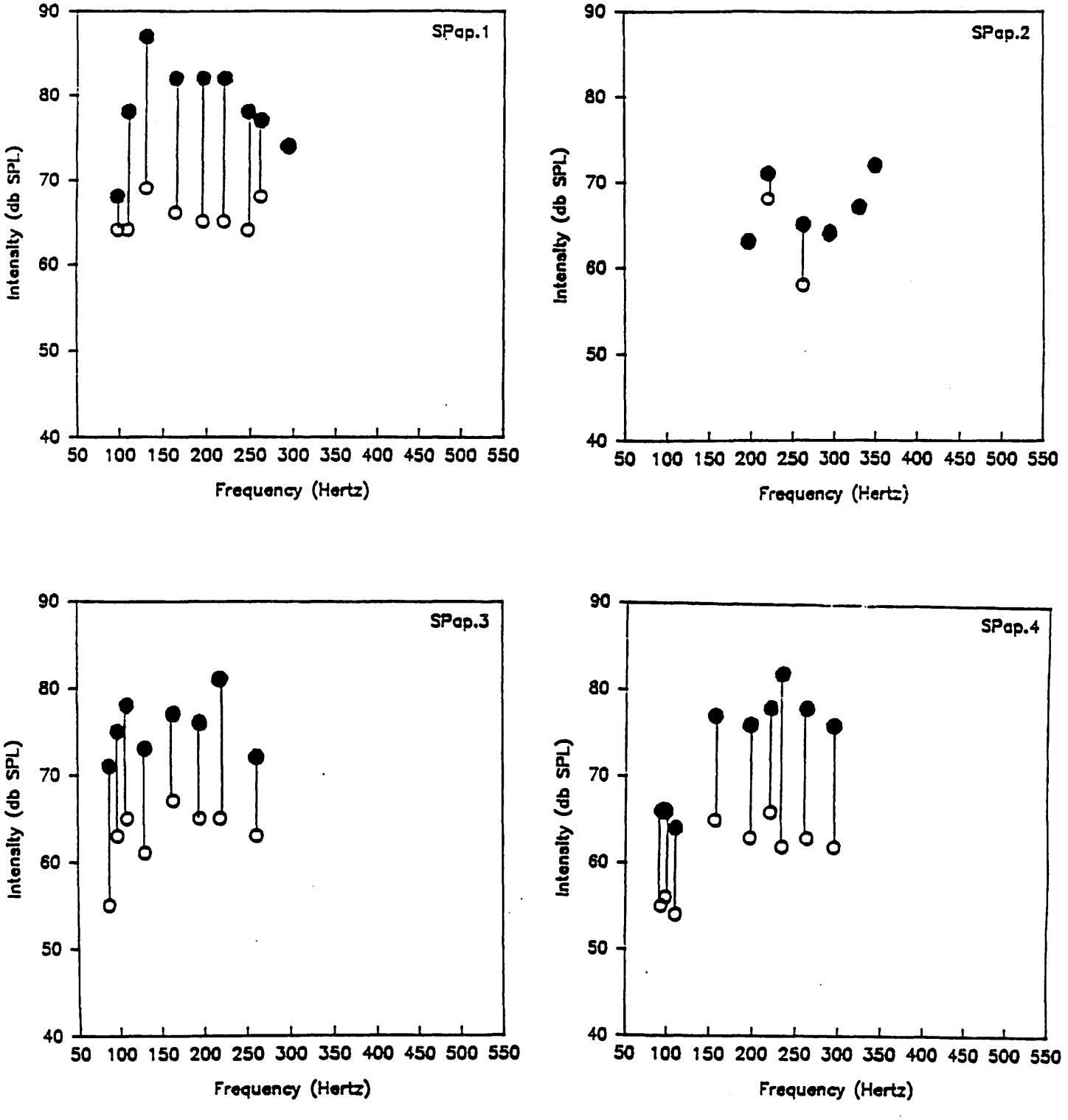

Figure 3: Voice intensity and frequency range profiles on eight patients previously inflicted with laryngeal papillomas. Subjects seven and eight are children. Subject six has fiarly wide ranges, but has very poor voice quality on perceptual ratings.

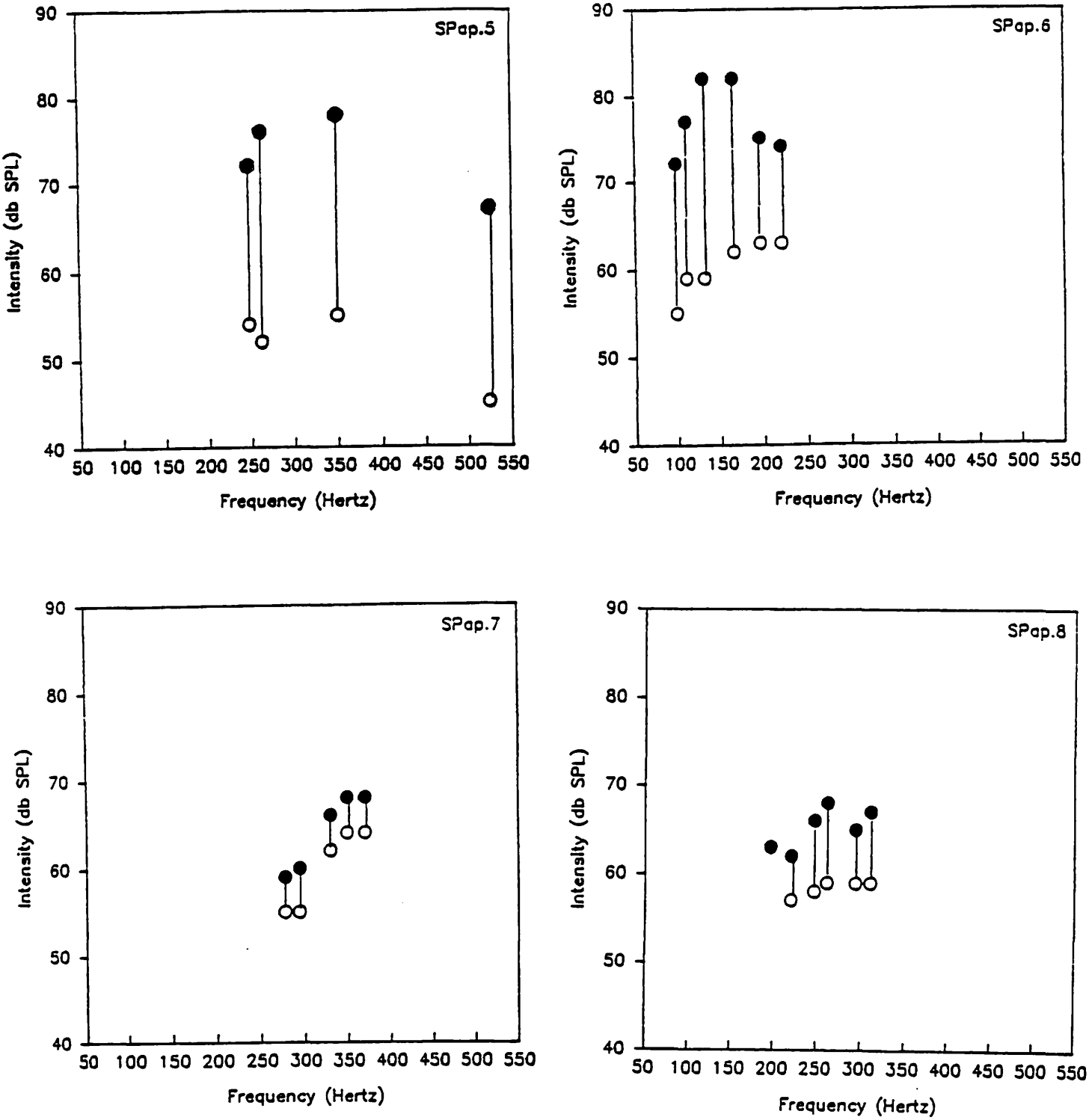

Figure 3: (Continued)

The intensity and frequency profiles of each patient can be seen in Fig 3. Patients 7 and 8 represent the two children. No established normals are available for this age group. Children are not expected to have the intensity and frequency range of adults. These tests will be interesting to follow during the next few years.

Table 3. - Voice Intensity and Frequency Ranges

Patient	Mean Intensity Range in dB*	Mean Frequency Range in Semitones
1	23 (64-87)	19
2	14 (58-72)	10
3	26 (55-81)	19
4	28 (54-82)	20
5	33 (45-78)	13
6	27 (55-82)	14
7	13 (55-68)	5
8	11 (57-68)	8

*First number in parentheses refers to threshold phonation or minimum loudness, second number refers to maximum loudness.

Mean frequency range for adult voices - 15.8 semitones
Mean intensity range for adult voices - 25.1 dB
Normal frequency range - 37 semitones
Normal intensity range - about 40 dB
Mean threshold pressures - 55 dB

Questionnaire results indicated that five of six were either happy or moderately happy with their voice. Nevertheless, four of six said their voice was not normal and all four selected the terms hoarse, harsh, breathy, and unsteady to describe these abnormal vocal qualities. Thirty-three percent said they were self-conscious about their voice when talking to others and avoided conversing in non-familiar social situations. Sixty-seven percent indicated that people made comments about their voice during conversations. Interestingly, only one patient had received speech therapy or speech guidance. No one felt that his or her voice had interfered with employment, although some indicated that a better voice might have allowed them to pursue a better or different job.

Stroboscopic examination demonstrated that some common abnormalities were non-straight leading edges (edges were bumpy or irregular), stiff or adynamic segments, and out of phase (with the opposite fold) vocal fold vibration. As one would expect, those voices that were perceptually the poorest had the most marked stroboscopic findings. Patients 2, 3, and 6 had stiff, adynamic

segments with poor mucosal waves, suggestive of severe lamina propria injury. Patient 6 also had a marked anterior vocal fold web.

The two children examined, while not included in the perceptual analysis, or the questionnaire, probably have fairly normal VIFR for their ages and seem to the authors to have perceptually fairly normal vocal quality. Neither child nor parent complained of voicing or speech difficulty. One patient (7) had a normal stroboscopic examination ; the other patient (8) could not tolerate stroboscopy but had normal-appearing vocal folds during direct laryngoscopy.

Discussion

In this preliminary study of vocal function, normal voices were different from those of patients who have had recurrent respiratory papillomatosis. The differences were noted in the perceptual voice qualities, the voice intensity and frequency ranges, and the stroboscopic examinations. Although the questionnaire indicated that most were generally satisfied with their voice, the majority indicated that they did not feel their voice functioned normally.

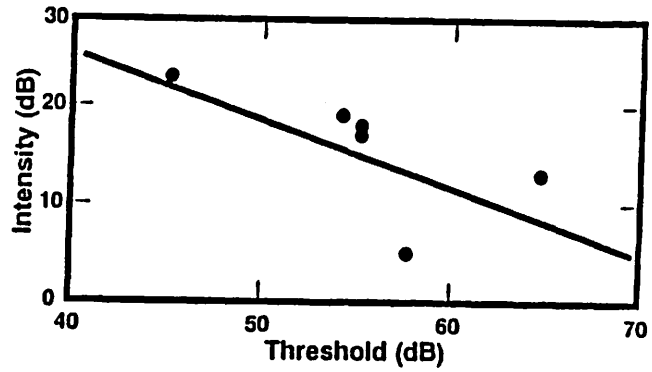

Figure 4. Threshold phonation plotted against vocal range of intensity in patients who previously had laryngeal papilloma. High threshold phonation correlates with decreased intensity range.

Figure 4 demonstrates the relationship between the threshold phonation (minimum) and the range of intensity. A higher threshold phonation correlates with decreased intensity range.[4] Factors that raise threshold phonation are not well identified. Scarring of the lamina propria has been indirectly associated with increased threshold phonation.[5]

It is noteworthy that vocal dysfunction persists despite remission of the laryngeal papilloma. Since papilloma are an epithelial disease, no permanent injury to the lamina propria should be present following remission. In this study, voice abnormalities seemed to be associated with two visible, general categories seen during laryngeal stroboscopy. Those two categories were, first, non-straight leading vocal fold edges due to possible microscopic, inactive, residual disease, or to surgical injury; and second, injury to the lamina propria, resulting in scarring and stiffening and, thus, preventing vibratory function.

Such findings raise concerns about whether the voice difficulties experienced by these patients have arisen from the disease or the treatment. Vocal fold injury due to surgery has been previously described in patients with RRP. It is now apparent that long-term voice dysfunction may be experienced as a result of the disease or the treatment. It is interesting that poor voice outcome does not seem to be directly related to the number of surgeries experienced. Patients 1, 4, 5, 7, and 8 all appeared to have more favorable voice outcomes, yet they experienced similar numbers of surgeries as a group than those with poor voice outcomes. Unfortunately, it is not possible to evaluate the surgical technique (not performed at this institution) used in some of these patients or the severity of disease, both factors that are likely to be important in voice outcome.

Surgical technique probably plays a considerable role in providing the optimal voice following remission. Phonosurgical principles should be applied to this type of surgery as they are applied

to the surgical removal of other benign lesions. Surgical tissue destruction should be limited to the epithelium, and care should be taken not to violate the deeper layers of the lamina propria. Adhering to these principles can be difficult for two reasons: sometimes the tissue planes of the vocal folds (and at times, the entire vocal fold) can be hard to distinguish and an underlying desire exists to surgically "get all of it." We sometimes assume that longer intervals between surgery are better; however, waiting too long between intervals can allow the disease to progress to a state in which the vocal folds are not recognizable during surgery, thus increasing the potential injury. Second, the desire to "get it all" may lead to overzealous surgical removal. It is becoming apparent that with juvenile recurrent respiratory papillomatosis, few patients are "cured" by surgery. Considerable doubt can be raised regarding the concept that leaving papilloma is a source of reinfection since the human papilloma virus has been isolated in many clinically non-diseased sites in these patients.[6]

Our current policy, similar to that of other centers, has been to remove the papillomas before vocal fold structures become obscure and not to attempt total removal of all papilloma in juvenile cases if doing so risks injury to the vocal folds. This policy is based on the assumption that, eventually, the disease will go into remission.

It should be mentioned that satisfaction with voice does not correlate with good perceptual ratings, good VIFR, or favorable stroboscopy. Patient 6, who had one of the poorest voice outcomes, indicated he was extremely happy with his voice. One can only wonder how this response relates to his experience with a tracheostomy for the first 8-10 years of his life. Satisfaction with voice seemed to be more related to the expectations of the individuals for their voice and the demands they placed on it than on actual voice quality.

Conclusion

Vocal outcome of patients who previously had laryngeal papillomatosis was different from normal voices. As a group perceptual ratings were significantly different from normal, although some patients were rated within normal limits. Voice intensity and frequency ranges were reduced from normal. Laryngeal stroboscopy commonly revealed vocal fold stiffening and irregular vocal fold edges in these patients. We suggest that surgical removal of papilloma be performed before the vocal folds become obscured, that overzealous removal of papilloma be avoided, and that, when possible, the surgical planes of the vocal folds be respected, thereby confining surgical injury to the epithelium. Such guidelines may provide better eventual outcome.

Acknowledgement

The work was supported, in part by NIH grant No. 7K08 DC00036-03, and in part by 1 P60 DC0097801.

Bibliography

[1]Crockett, D.M., McCabe, B.F. and Shive, C.J. (1987). "Complications of Laser Surgery For Recurrent Respiratory Papillomatosis." Anal Otol Rhinol Laryngol, 96:639-44.

[2]Coleman, R.F., Mabis, J.H., and Hinson, J.K. (1977). "Fundamental Frequency-Pressure Level Profiles of Adult Male and Female Voices." Speech Hearing Res, 20:197-204.

[3]Komiyama, S., Watanabe, H., and Ryu, S. (1984). "Phonautographic Relationship between Pitch and Intensity of the Human Voice. Folia Phoniatrica, 36:1-7.

[4]Titze, I.R. (1991). "Acoustic Interpretation of the Voice Range Profile (Phonoetogram)." Accepted for publication, Speech Hearing Res.

[5]Wexler, D.B., Jiang, J., Gray, S.D. and Titze, I.R. (1989). "Phonosurgical Studies: Fat-graft Reconstruction of Injured Canine Vocal Folds". Ann Oto Rhin Laryng. 98:668-73.

[6]Pignatari, S.S.N., Smith, E.M., Gray, S.D., Shive, C. and Turek, L.P. (**Submitted for publication**) "Detection of HPV Infection in Diseased and Nondiseased Sites of the Respiratory Tract in Recurrent Respiratory Papillomatosis Patients by DNA Hybridization."

NCVS Status and Progress Report - 1
June 1991, 99-109

Vocal Evaluation of Thyroplastic Surgery in the Treatment of Unilateral Vocal Fold Paralysis

Steven D. Gray, M.D.

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

Julie Barkmeier, M.S.

Ingo Titze, Ph.D.

David Druker, B.S.

Department of Speech Pathology and Audiology, The University of Iowa

David Jones, Ph.D.

Department of Otolaryngology - Head and Neck Surgery, The University of Iowa

Abstract

Vocal function is assessed in 15 patients who received thyroplasty type I for rehabilitation of unilateral vocal fold paralysis. Function is assessed by perceptual evaluation, voice intensity and frequency range profile, and questionnaire. Goals of the study are: to evaluate efficacy of thyroplasty in improving voice; to compare postoperative thyroplasty voices to normal voices; and to correlate objective measurements to the results of the questionnaire regarding satisfaction with voice. Perceptual evaluations were performed by randomizing normal and thyroplasty voices on a recording tape. The voices on the tape were then rated by independent, blinded, trained listeners. The perceptual qualities of pitch, intonation, and loudness were not statistically different than normals, however, voice qualities of strain, breathy, hoarse, harsh, and unsteady were different than normals. Mean frequency range and mean intensity range was moderately to severely reduced from normals with a wide variation being present in the results. The correlation between the higher threshold phonation pressures and decreased intensity ranges found in some patients is discussed.

Questionnaire results indicate that a high degree of satisfaction with the surgery is present (92%). Extreme or general satisfaction with their voice was present in 73%. The most difficult voicing is experienced at work with 25% needing to adjust their employment to accommodate their voice abilities. These results indicate that thyroplasty type I is effective in partially rehabilitating

unilateral vocal fold paralysis. Voice function is still not normal, probably in part due to the underlying disease.

Introduction

Symptomatic unilateral vocal fold paralysis (UVP) has generally been treated with teflon injection to augment the paralyzed vocal fold. Although teflon (Polytef paste, trademark Ethicon/ Polytetra-Fluorethylene) injection has been used successfully by a large number of surgeons, some dissatisfaction with certain aspects of the procedure, the material, migration, and long-term results has led to the trial and development of alternative methods to rehabilitate the unilateral paralyzed vocal fold. One of these methods, thyroplasty type I[1,2] has received considerable interest from laryngeal surgeons. This study evaluates some longer term results following thyroplasty type I surgery.

Many voice procedures have been evaluated by comparing postoperative voice to preoperative voice.[1,2,3,4,5,6,10] In contrast, our study compares vocal results of thyroplasty voices to normal voices rather than to preoperative voices. We have done this for several reasons: (1) normal voice provides standards by which other methods can be cross-compared; (2) this comparison allows identification of the areas of rehabilitated UVP which may still need improvement; and (3) since many methods, including teflon injection, collagen injection, thyroplasty surgery, and reinnervation procedures, are capable of improving the pre-operative voice condition of unilateral vocal fold paralysis, this type of comparison may lead to objective assessment of the various methods.

Interpretation of data reported in studies that compare postoperative to preoperative voice can be difficult, as the following examples indicate. May reported 76% improvement of all UVP with reinnervation procedures and up to 95% improvement in patients with no other laryngeal pathology except UVP.[4] Improvement was defined as postoperative improvement in one or more of three categories: loudness, pitch, and quality. Vocal results, did not appear to be normal, however, since he went on to state, "Further, 9 of these 19 (patients with only UVP pathology) who improved were considered to have near-normal voice postoperatively." Tucker reported the long-term results of reinnervation of UVP in 31 patients.[5] "Voice spectrograms were obtained in an attempt to arrive at some reliable assessment of the quality of voice that had been achieved. It was our conclusion that such analysis is not sufficiently precise to provide an objective measurement of voice quality and, therefore, could not be used to assess the results obtained. Therefore, the voice results obtained were graded subjectively by the surgeon, the patient and/or a speech pathologist..." Ford and Bless, in a preliminary study of collagen injection for vocal fold augmentation, found that airflow, phonation time, and intensity increased significantly, and vocal self-perception data demonstrated overall improvement.[6]

The comparison trials of preoperative voice to postoperative voice are valuable. However, by comparing postoperative voices to normal voices we hope to establish objective standards that can then be used in studies of alternative methods. This will allow us to evaluate the efficacy of thyroplasty surgery in comparison to other methods of improving voice.

The goals of this study were threefold: 1) to evaluate the efficacy of thyroplasty in improving voice; 2) to compare postoperative thyroplasty voices to normal voices; and 3) to correlate objective measurements to questionnaire results, voice intensity and frequency range profiles, perceptual quality assessments, and voice functioning in social situations and employment.

Methodology

Fifteen patients who were at least one year post-phonosurgery for correction of unilateral vocal fold paralysis participated in this study. These patients received voice intensity and frequency range profiles (VIRP), voice recording for perceptual analysis, and a questionnaire regarding voice abilities in the workplace and social situations. During the time period in which their surgeries occurred, 1986-1988, no patients were selected out for teflon or arytenoid adduction procedures. The thyroplasty technique used was similar to that described by Isshiki and Kaufman, sometimes using modifications (Isshiki) and cartilage, silastic, or gortex as implants.[1,2,7] The patients varied as to etiology of vocal cord paralysis and size of glottic gap. They ranged in age from 17 to 83 years.

Perceptual Evaluation

Evaluators. Ten speech-language pathologists with experience in voice disorders participated as listeners. All listeners completed training for this research by listening to a voice disorder practice tape which illustrated the perceptual qualities they would later be asked to rate. Sample ratings were performed as part of the training.

Experimental tapes. In addition to the fifteen patients, four normal-voiced individuals were recorded reading 10 sentences (Table 1). A lapel microphone and a four channel VHS recorder were used. Recordings of the speakers were randomized in the following manner: First, of the ten recorded sentences, five were randomly assigned to set 1 and five to set 2 for that speaker. Next, each set for each speaker was randomly ordered on a cassette tape for presentation to the experienced listeners. Of the 19 speakers, eight sets were randomly chosen for a third presentation to obtain reliability measures. Thus, eight speakers were presented three times on the listener tape, resulting in a total of 46 sets being presented to the listeners.

Table 1
Sentences Used For Perceptual Evaluation

1. Most boys like to play football.
2. Do you have a brother or sister?
3 Ted had a dog with white feet.
4. Nick's grandmother lives in the city.
5. We go swimming on a very hot day.
6. Tom has ham and eggs for breakfast.
7 Can you count to nine?
8 Do you want to take my new cap?
9. Do you know the name of my doll?
10. Jack likes cheese sandwiches for lunch.

Speaker rating forms. The rating form consisted of nine bipolar voice quality items on either end of a 7-point, equal- appearing interval scale. Each item was rated by each listener following presentation of each set of five sentences. The nine bipolar items are presented in Table 2. The first three items are rated 4 for normal--pitch, intonation, and loudness--while the rest of the

items are rated 1 for normal.

Table 2.
Voice Quality Rating Items Form

Pitch:	Low	1	2	3	4	5	6	7	Animated
Loudness:	Soft	1	2	3	4	5	6	7	Loud
	Unforced	1	2	3	4	5	6	7	Strained
	Not Breathy	1	2	3	4	5	6	7	Severely Breathy
	Not Hoarse	1	2	3	4	5	6	7	Severely Hoarse
	Not Harsh	1	2	3	4	5	6	7	Severely Harsh
	Steady Voice	1	2	3	4	5	6	7	Shaky Voice
Normal Voice (overall)		1	2	3	4	5	6	7	Severely Abnormal

Experiment. Listeners were seated in a soundproof booth 3 feet directly in front of a speaker. The sound level of speaker recordings at the position of the listener was approximately 70 dB SPL.

Listeners were presented the experimental tape of normal and thyroplasty voices randomly. They were instructed to listen to each set of five sentences and to rate the speaker on the nine bipolar items. The next sentence set was presented when the listeners indicated that they had finished rating the previous set.

Analysis. Interreliability and intrareliability measures were obtained. The mean rating for each bipolar item for each speaker and across all speakers was determined. Mann-Whitney statistical analysis was used to determine if qualities of voice perception were different in the thyroplasty group when compared to the normal group.

Voice Intensity and Frequency Range

Intensity and frequency range profiles were obtained for each patient by seating him or her in a soundproof booth and positioning an Epson sound-level meter at a distance of 30 centimeters and at a 45-degree angle from each patient's mouth. A weighted scale was used during loudness measurements. In addition, a Yamaha keyboard was utilized to provide the patients with target frequencies.

Patients were instructed to sustain phonation of "ah" during the measurement of loudness at the target pitch. Initially, the experimenter requested the patient to sustain phonation of "ah" at a comfortable pitch. After the experimenter matched this pitch on the keyboard, the patient was instructed to sustain phonation of the same pitch as softly as possible. This was called "threshold phonation." The experimenter perceived this threshold as being reached when the patient produced a periodic voice signal, rather than a whispered breathiness. After measuring threshold and maximum loudness at a comfortable pitch, the experimenter provided higher and then lower target pitches using both the keyboard and modeled sustained phonation. Maximum loudness and threshold (minimum loudness) were reproduced for three tokens (trials). This method of examination continued until the envelope of intensity and frequency range had been determined. The patient's total intensity and frequency range was then plotted using Sigma Plot software, and

mean intensity and frequency were determined from the three trials using Mystat (by Systat, Inc.) for graphic analysis.

Questionnaire

A questionnaire was used to determine epidemiological data and to assess patient satisfaction with voice and functional utility of the voice in social and workplace situations. It was 50 questions long and included reliability questions. Only those questions that elicited responses which helped us interpret the objective data and their results will be discussed.

Results

Perceptual Evaluation

The mean ratings and standard deviations for the perceptual evaluation of the thyroplasty group and normal group are presented in Figure 1. Overall, the patient population was rated as approximately normal on the pitch, intonation, and loudness items. On the remaining items, a large degree of variation in rating scores occurred across subjects. Generally, the patients were rated as slightly abnormal on strained, breathy, hoarse, and harsh items. All these items were statistically different from normals: strained - p=0.003; breathy - p=0.003; hoarse - p=0.005; harsh - p=0.002; unsteady - p=0.01; and overall - p=0.003.

Voice Intensity and Frequency Range

Table 3 includes each patient's average threshold phonation and maximum loudness measures across all frequencies, as well as each patient's frequency range in semitones. Across all patients, a large degree of variability in intensity and frequency range was demonstrated.

When frequency results were averaged across all of the patients, the average frequency range was 16 semitones. This value is less than half of the average frequency range of 37 semitones found in normal-voiced individuals.[8]

The average total intensity range (softest threshold phonation of any trial to loudest phonation of

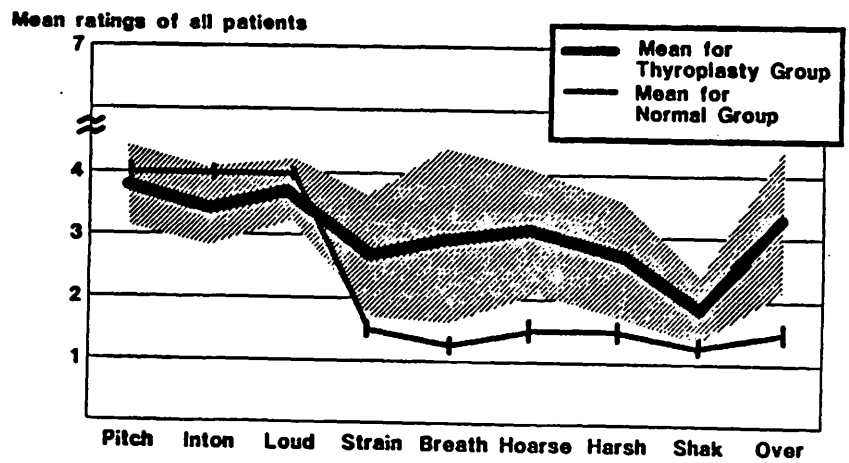

Figure 1. Perceptual evaluation of voices of thyroplasty patients as compared to normal voices. Broad bands indicate standard deviations for nine bipolar voice quality items. The first three items are rated 4 for normal, and the remaining six are rated 1 for normal.

any trial) across all patients was 22 decibels. Normal-voiced individuals usually exhibit a 30-40 decibel loudness range.[9] Mean threshold phonation was 59 dB, with a standard deviation of 6 dB. Therefore, loudness range was severely reduced in all patients. Figures of the two best and of the two worst voice intensity and frequency range profiles are included for inspection (Fig. 2). Figure 3 shows the mean total intensity range plotted against the mean threshold phonation. Note that the higher the mean threshold phonation becomes, the lower the intensity range is. This result is expected since maximum loudness is difficult to increase. Increases in subglottic pressure result in moderate loudness gains in softer phonation, while equal increases in subglottic pressure result in minimal gains in loud phonation. Because of this relationship, threshold phonation becomes a fairly important factor in determining total intensity range. Emphasis in phonosurgery should be placed on providing a result which is conducive to low threshold phonation pressures.

Table 3. Frequency and Intensity Ranges for Voices of Thyroplasty Patients*

Patient Age	Sex	Mean Frequency Range (semitones)	Mean Threshold Loudness (dB)	Mean Maximum Loudness (dB)	Time Since Surgery (yr)
50	F	9	59.5	70.8	1.5
74	M	38	60.1	70.1	2.0
41	F	7	52.2	63.4	2.5
83	F	16	66	70	2.0
17	F	5	49.3	60.3	4.0
44	M	21	65.4	75.9	1.3
64	F	24	59.3	69.4	1.5
80	F	6	64.8	70.3	2.0
53	F	9	54.9	60.9	1.0
67	F	24	58.2	65.3	2.5
74	F	21	55.2	69.8	3.0
27	M	7	66.6	73.0	5.0
37	M	17	66.6	68.9	3.5
51	F	17	63.5	72.0	3.5
*25	F	24	58.3	63.8	4.5

Mean age = 49 years (SD 19 years).
Mean frequency range = 16 semitones (SD 9 semitones).
Average norm = 37 semitones.
Mean threshold phonation = 59 dB (SD 6dB).
Mean total intensity range = 22 Db.
Average normal intensity range = 30-40 dB (see text).

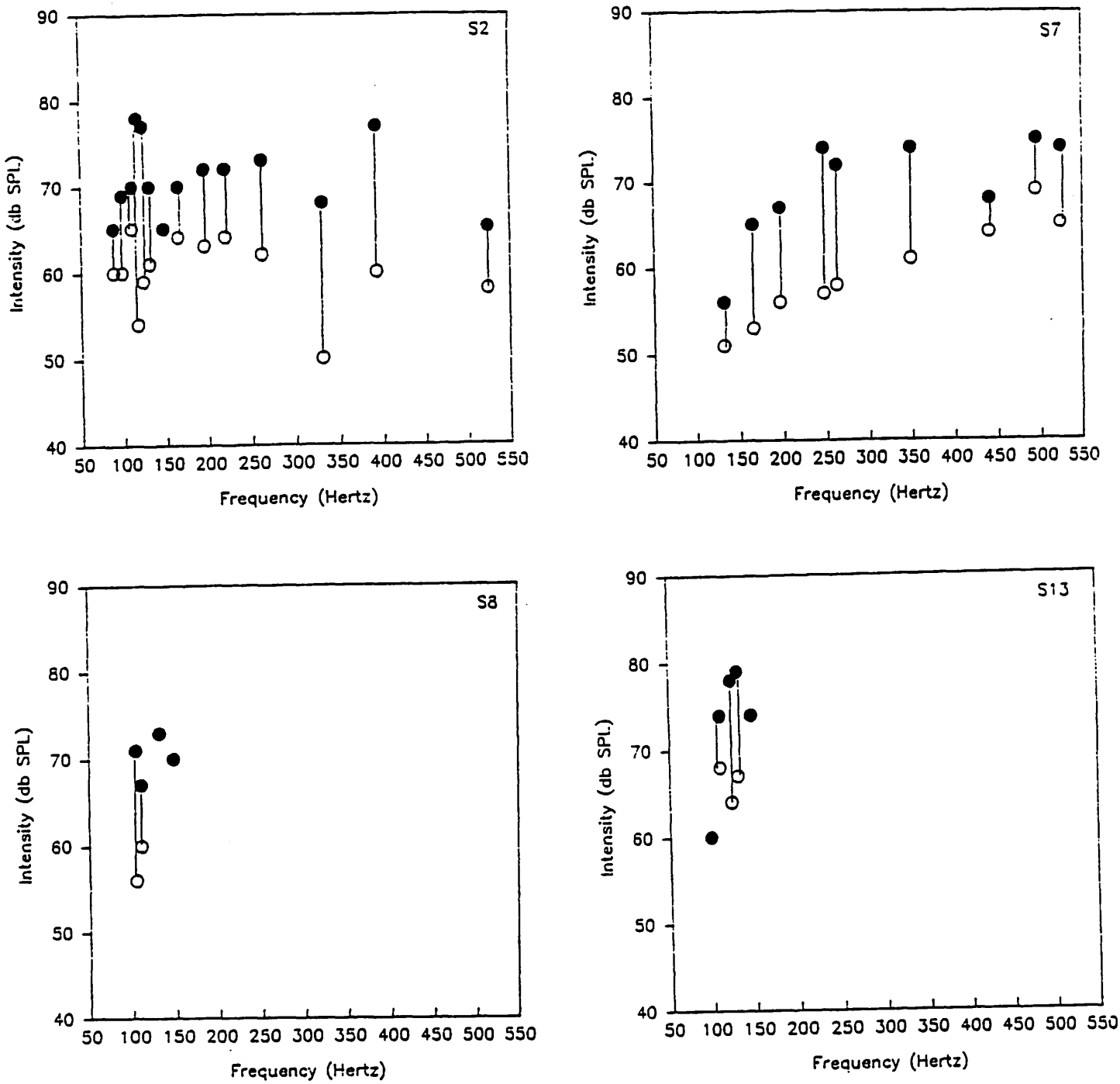

Figure 2. The two best and two worst voice intensity range profiles. Top left. Male with good frequency range and fair intensity range at its fundamental frequency of approximately 130 Hz. Top right. Female with good intensity and frequency range. Bottom left. Female 80 years old with very poor frequency and intensity range. Bottom right. Male with very poor frequency range.

Questionnaire

The questionnaires indicated a high patient approval/satisfaction rating as a result of the surgery (92%). Only one person did not feel surgery improved his voice. Thus, when asked to respond to the statement "Surgery on my throat helped improve my voice," 92% responded "yes." When asked if they were happy or satisfied with their voice, 13% (2 of 15) said they were "extremely happy" with their voice, 60% (9 of 15) said they were "generally satisfied," and 27% (4 of 15) said they were "generally not satisfied" with their voice. No one expressed extreme dissatisfaction or unhappiness with his or her voice. However, 87% of the individuals indicated that their voices were still abnormal. Difficulties indicated included softness, breathiness, hoarseness, and unsteadiness of vocal quality. In addition, 48% indicated a need to frequently clear their throats, and 28% frequently experienced coughing. Tension, felt in the neck (21%), throat (36%), and shoulders (16%), was an additional complaint.

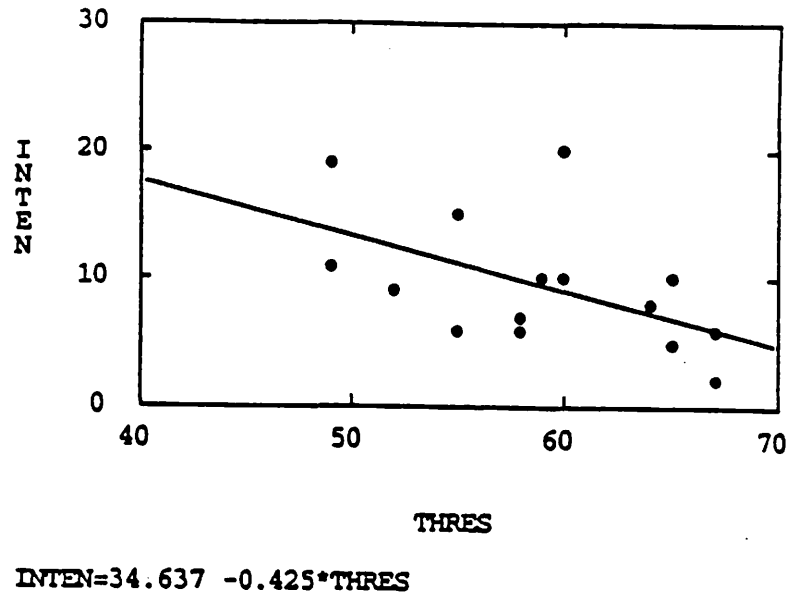

Fig. 3 Mean intensity range of thyroplasty voices plotted against threshold phonation pressure. Note that as threshold pressure increases, mean intensity range decreases. Emphasis in phonosurgery should be placed on providing a result which is conducive to low threshold phonation pressures.

Regarding voice usage, individuals reported that they generally talked about the same length of time as before surgery. Fourteen percent felt self-conscious about their voices, 14% thought other people noticed the abnormality of their voices, and 28% indicated that other people commented on the abnormality of their voices.

Individuals were frequently asked to repeat what they said (43%), and they experienced increased difficulties with voicing or vocal fatigue at the end of the day or when they were tired (25%). Another common complaint was frequently running out of air and needing to work hard to maintain voicing.

With regard to vocal performance in social situations and employment, 4 out of 15 (one nonresponder) said that their voice interfered with their work. These four people had to seek new employment or change job responsibilities to accommodate their voice dysfunction. However, 13 out of 15 (93%) indicated their voice did not interfere with family activities.

The findings from this questionnaire indicate that these patients were happy that they had achieved voicing improvement. However, most of these individuals felt they did not have normal voices, and some were experiencing difficulties related to their abnormal voices in work and social areas.

In an attempt to correlate reduction of frequency range or intensity range with dysfunctional voices in the workplace environment or social situations, we performed cross correlation by using mean frequency range and mean total intensity range with responses in the questionnaire. Surpris-

ingly, no correlation was found between satisfaction with voice, employment difficulty, and reduction in frequency range or intensity range.

Discussion

Overall, patients reported a high percentage of satisfaction (92%) with surgery and post-surgery improvement. Nevertheless, most of these patients did not feel their voices were completely normal. The most difficult voicing occurred at work with 27% requiring employment changes. It is likely that certain jobs require more extensive voice use. No attempt was made in this study to determine which occupations were abandoned and which were retained.

Even though 73% indicated extreme or general satisfaction with their voice, when their voices were compared with normal voices in a blinded, randomized fashion, the thyroplasty voices were perceptually rated as minimally to moderately abnormal. Pitch, intonation, and loudness were found to be no different from normals; but acoustic perturbational qualities, such as hoarseness or harshness, and perceptual qualities associated with aerodynamics, like straining or breathiness, were significantly different from normal. This study confirms the findings of Hammarberg, who studied teflon-injected patients, preoperatively and postoperatively, by using trained listeners to rate taped voice segments in a blinded, randomized fashion.[10] That study found that breathiness, loudness, and pitch were most likely to be corrected from the preoperative state; however, roughness, diplophonia, and grating were not significantly different from preoperative conditions. In this study we found that not only can pitch and loudness be improved from preoperative conditions, but they are essentially normal when compared to normal voices. However, other qualities, including breathiness, although improved from a preoperative state, are rated as not normal. Hammarberg pointed out that the perceptual qualities of grating and harshness are occasionally worse following teflon injection.[10] Our thyroplasty voices as a group were also rated as hoarse and harsh when compared to normals. It should be noted that most surgeons, while correcting glottic gaps under local anesthesia, focus on loudness and breathiness of the patient's voice as the endpoint. Therefore, it would be expected that these would be the preoperative qualities most likely to improve. As voice surgeons become more experienced and ears become more trained, correction of voice under local anesthesia will require listening to and monitoring harshness and threshold phonation as well.

Although further work needs to be done, various factors may be responsible for the harshness that may accompany a thyroplasty voice: over medialization, curved or non-straight medial leading vocal fold edge, or compression of the vibrating portion of the lamina propria by the medializing implant. Hammarberg indicated that over medialization with resultant grating and harshness tended to occur with correction of small glottic gaps.[10] In our study, we observed that the latter two conditions, a non-straight vocal fold edge and implant compression of the vibrating portion of the lamina propria, tended to occur with those who had a very wide glottic gap preoperatively. Because of this, we now use arytenoid adduction techniques more liberally to medialize the vocal fold in patients with large glottic gaps.

Following up on the results regarding voice difficulties at work, we were eager to see if an association existed with reduced voice intensity range or frequency range. Our enthusiasm was increased by the observation that vocal fatigue was commonly experienced by our patients during testing of voice intensity and frequency ranges. We had assumed that some employment skills (coaching, sales, teaching, etc.) required large ranges of intensity and frequency. Vocal fatigue was, in fact, a major complaint of thyroplasty patients. But, when questionnaire responses to vocal

fatigue or vocal employment difficulties were correlated with range of intensity or frequency, no statistically significant association was found. Three of the four subjects who had employment difficulties had frequency ranges of less than ten semitones, so a trend may be present. Reduction in intensity range or frequency range did not correlate with voice satisfaction, either. It is possible that the small numbers in this study will not allow us to recognize these associations. These findings suggest that patients learn to optimize voice production and modify voice behavior to within the more narrow range that they have.

Our results regarding vocal performance and employment are difficult to interpret. It appears that within the parameters studied and the superficial employment questions asked, the perceptual qualities of the voice or the voice ranges of intensity or frequency do not reflect how functional the voice is in the workplace environment. This study is unable to show that the abnormal perceptual qualities that correlate with acoustical perturbations or poor aerodynamics, or with reduced voice range or intensity, provide prognostic information regarding the functional capabilities of the rehabilitated voice. This is remarkable since most of us assume that a voice which sounds bad, performs bad. And, although this may be true, we were unable to show it in this study.

It is apparent that rehabilitation with thyroplasty, although offering significant patient satisfaction and improvement from preoperative voice, at the present time does not offer total recovery to normal voice. This is not surprising, since the underlying condition of vocal paralysis is still present. Reinnervation procedures may hold promise in this area, used alone or in combination with other techniques. No randomized studies or blinded ratings by non-biased observers are yet available regarding these techniques. This study has led us to counsel our preoperative patients in the following ways. "It is likely that voice surgery (thyroplasty) will restore your voice to normal loudness, pitch, and intonation. Nevertheless, your voice may still be harsh and grating (not normal). Voice ranges of intensity and frequency will likely be reduced (probably in large part due to the underlying disease resulting in fold paralysis)." To those that are employed we indicate that they may need to find occupations that allow them to function within their voice capabilities. The slightly harsh voice that may result from augmentation/medialization procedures may be problematic for certain professions (radio announcers, telephone talking).

As mentioned before, we now select out those with larger glottic gaps for arytenoid adduction with or without thyroplasty. To date, no non-biased, randomized analysis has been performed on patients receiving arytenoid adduction.

Conclusion

Vocal fold paralysis rehabilitated with thyroplasty type I provided a high degree of patient satisfaction; however, their voices, as a group, were rated as slightly harsh, breathy, and strained. Voice range reduction occurred in both frequency and intensity. Results compare favorably to the teflon injection technique used to rehabilitate the unilaterally paralyzed vocal fold. Comparison of results to normal voices is intended to allow easier future comparison of techniques. It appears that many techniques improve the preoperative state of UVP. Identifying the optional combinations of patient and technique will likely become an important issue as further data correlating the results of techniques to age, sex, and employment become available.

Bibliography

1. Isshiki, N., Taira, T., Kojima, H., and Shoji, K.: Recent Modifications in Thyroplasty Type I. Ann Otol Rhinol Laryngol, 98:777-779, 1989.
2. Koufman, J.R.: Laryngoplasty for vocal cord medialization: An alternative to Teflon. Laryngoscope, 96:726-31, 1986.
3. Maves, M.D., McCabe, B.F., and Gray, S.D.: Phonosurgery: Indications and pitfalls. Ann Otol Rhinol Laryngol, 98:577-580, 1989.
4. May, M. and Beery, Q.: Muscle-Nerve Pedicle Laryngeal Reinnervation. Laryngoscope, 96:1196-1200, 1986.
5. Tucker, H.M., and Rusnov, M.: Laryngeal Reinnervation for Unilateral Vocal Cord Paralysis: Long-Term Results. Ann Otol, 90:457-459, 1981.
6. Ford, C.N., and Bless, D.M.: A Preliminary Study of Injectable Collagen in Human Vocal Fold Augmentation. Otolaryngol Head Neck Surg, 94:104-112, 1986.
7. Crumley, R.L., and Izdebski, K.: Voice Quality Following Laryngeal Reinnervation by Ansa Hypoglossi Transfer. Laryngoscope, 96:611-616, 1986.
8. Coleman, R.F., Mabis, J.H., and Hinson, J.K.: Fundamental Frequency-Pressure Level Profiles of Adult Male and Female Voices. Speech Hearing Res, 20:197-204, 1977.
9. Komiyama, S., Watanabe, H., and Ryu, S.: Phonautographic Relationship between Pitch and Intensity of the Human Voice. Folia Phoniatrica, 36:1-7, 1984.
10. Hammarberg, B., Fritzell, B., and Schiratzki, H.: Teflon Injection in 16 Patients with Paralytic Dysphonia: Perceptual and Acoustic Evaluations. Jour Speech Hearing Disorders, 49:72-82, 1984.

NCVS Status and Progress Report - 1
June 1991, 111-119

Difference Limen for Frequency Modulation Extent In the Synthesized Vowel /a/

Yangho Cho, Ph.D.
Yoshiyuki Horii, Ph.D.
Department of Communication Disorders and Speech Science, University of Colorado, Boulder
and
Recording and Research Center, The Denver Center for the Performing Arts

Abstract

Difference limen for frequency modulation extent was studied using the synthesized vowel /a/ with modulation rates of 5 and 7 Hz and a base frequency of 130 Hz. Results of paired-comparisons by 50 listeners revealed that DL defined as the 75% correct response level was about .3 semitones for both modulation rates. The implications of the finding are discussed in relation to singers' ability to control vibrato rates and extents and their variability.

Introduction

One of the intriguing phonatory maneuvers routinely produced by singers is vocal vibrato. In vocal vibrato, singers not only phonate at target frequencies, but also modulate frequencies above and below the target within about 0.5 to 2 semitones at approximately 4 to 8 times/sec. Thus, vocal vibrato exemplifies a complex frequency control maneuver.

The literature is equivocal on the issue of controllability of rate and/or extent of frequency modulation in vocal vibrato (Vennard, 1967; Ruhlman, 1984; Shipp, Leanderson and Sundberg, 1979; Keidar, et al.; 1984; Campbell and Michel, 1979). A recent study by King (King, 1989; King and Horii, 1989), however, strongly indicated that the rate of frequency modulation is under voluntary control between 3 and 7 Hz while extent is not. On the average, King's 11 trained singers matched rates of 3, 5, and 7 Hz within 10% and extent of 0.5, 1, and 1.5 semitones within 60%. King's findings not only have significant implications to the theory of vocal vibrato production, but also naturally leads to another important area of investigation regarding frequency control capability.

It is reasonable to assume that one would not develop motor control unless the results of the control were perceptible. King's subjects demonstrated their ability to match the frequency modulation rate suggesting that they perceived the differences in the rate. One cannot conclude, however, that they did not perceive the differences in the extents simply because they could not match the extent. Possibly, they failed to control extent in spite of being able to hear differences in the extent.

Detectability of the presence of frequency modulation, rather than the discrimination of different degrees of frequency extent, has been investigated from an audiometric point of view (Shower and Biddulph, 1931; Moore, 1976; Moore and Glasberg, 1989). Although studies directly examining difference limen for extent of frequency modulations in vibrato have not been conducted to the authors' knowledge, there also have been a number of investigations on the perception of frequency modulations in speech-like signals (Danaher and Osberger, 1973; Mermelstein, 1978; Horst, 1989). In this context, the frequency modulation often refers to formant frequency shifts. Horst (1989), Danaher and Osberger (1973), Mermelstein (1978), for example, employed one- to five-formant synthesized vowels in attempts to find DLs for "formant transition." As such, the base signals were complex tones with multiple harmonics and the frequency modulation patterns were either "rising" or "falling" from one frequency to another, analogous to vowel formant transitions for consonant-vowel syllables. It was not obvious how the discriminability of extent of frequency modulations in vibrato (in repeated modulations) would compare with the discriminability of formant transitions (in a single modulation).

The purpose of the present study was to examine listeners' ability to discriminate differences in magnitudes of extent of 5-Hz and 7-Hz frequency modulations. Knowledge of such perceptual constraints would provide further insight into human phonatory capabilities. Primary questions asked in this investigation were: (1) What are DLs defined as the 75% response level for the extent of frequency modulations in synthesized vibrato /a/? (2) Are the extent DLs different for 5-Hz versus 7-Hz modulation rates?

Method

Auditory Stimuli

Auditory stimuli were synthesized /a/ with vocal vibrato generated by a VAX computer using the program SPEAK at The Recording and Research Center, The Denver Center for the Performing Arts. SPEAK is a speech synthesis program developed by Titze (1983).

Two sets of 12 synthesized /a/s with differing amounts of extent of frequency modulation were used in this study. All stimuli had the base (fundamental) frequency of 130 Hz. The extent of frequency modulation increased from 2% to 7.5% with a 0.5% increment. These percent values correspond to the extents of .68 semitones to 2.42 semitones with an increment of .16 semitones. The rate of frequency modulation for one set was 5 Hz and the other was 7 Hz. The duration of each stimulus was one second.

Auditory Discrimination Test Tapes

An auditory discrimination test tape was created in a paired comparison paradigm for the 5-Hz and 7-Hz tests. A pilot study indicated that comparisons of two stimuli separated by more than six steps were clearly distinguishable. Therefore, it was decided that only stimulus pairs separated by six or less steps be used for the discrimination tests. Pairing in this manner resulted in a total of 51 possible pairs for each set.

Taking into consideration a possible order effect within each pair, it was decided to include, additionally, the same 51 pairs but to reverse the order of stimuli within each pair. The final discrimination test tape for each modulation-rate condition, therefore, contained a total of 102 pairs. The orders of the 102 pairs in the 5-Hz and 7-Hz tests were randomized.

The interstimulus interval (the interval between the first and second stimulus within each pair) was 1.5 seconds, and the interpair interval (the interval between two pairs, during which the subject must respond) was 2.5 seconds. A brief 500 Hz complex tone was provided after every 5th pair in order to assist the subject in verifying which pair was being presented. The total duration of each test was approximately 22 minutes. A 5-minute recess was provided between the 5-Hz and 7-Hz tests.

In addition to each discrimination test, each tape contained (1) the 12 stimuli in the order of the least (2%) to the most (7.5%) extent, and (2) 12 practice pairs. The primary purpose of these stimuli was to let the subject listen to the whole range of extents being investigated and to familiarize himself/herself with the required task, e.g., the timing of stimulus presentation and the response format.

Subjects and Experimental Procedures

A total of 50 adults, 29 males and 21 females, served as the subjects. Their age ranged from 22 to 46 with a mean of 31 years. The subjects passed a pure-tone hearing screening test at 500, 1000 and 2000 Hz at 20 dBHL in at least one ear (ANSI, 1979). Of the 50 subjects, 9 were considered to be trained singers.

The subject was seated in a sound-treated booth. After instructions, the subject put on headphones (TDH-39 with Zwislocki cushion) and listened to the 12 stimuli (via a TEAC tape recorder model 3-XR) in the order of the least (2%) to the most (7.5%) extent. The subject was then informed that there would be a series of 12 practice pairs of stimuli he/she just heard. The subject was asked to indicate on the response sheet which of the stimuli of the pair appeared to have the greater frequency extent. It was emphasized that he/she must choose one of the two stimuli even if he/she was not sure. The 12 practice pairs were then presented. The presentation level was set at 70 dBA using a General Radio Sound Level Meter and a 6 cc-coupler under the headphones. The order of the test parts (the 5-Hz and 7-Hz tests) was counterbalanced so that half of the listeners received the 5-Hz test first and the other half received the 7-Hz test first.

Results

Reliability Measures

Reliability was investigated by employing the test-retest method. Ten subjects were selected, based on their scores of the 5-Hz test. In order to select representative subjects from the sample of subjects, their scores were ranked in descending order. Then, every 5th subject was selected for the re-test. The retest procedures were identical to the first test. Since the statistical testing results to be described later indicated that the differences between the 5-Hz and 7-Hz tests were not significant, only the 5-Hz test was adopted for retest. The interval between the first test and re-test ranged from 7 days to 10 days.

Results indicated that the average of the first scores of the ten subjects was 85.9, while that of the second was 86.3. The difference between these means was not statistically significant, as evidenced by a correlated t-test ($t = -.338$, $df=9$). The Pearson product-moment correlation coeffi-

cient, on the other hand, was .805, highly significant ($p < .05$). These two statistical results indicated that 1) there were no differences between the first and second tests, and that 2) those who scored poorly in the first test tended to score poorly in the second test and vice versa. It was concluded that the reliability of the perceptual measurements was satisfactory.

DL for the Extent of Frequency Modulation

As stated earlier, DL was defined as the 75% correct response level. As a preliminary to the DL calculation, for each subject, the total correct responses in percent at each step difference were calculated for the 5-Hz and 7-Hz tests. The results are presented in Table 1 wherein the total number of responses for each modulation test, and means and standard deviations of percent correct responses, are shown for each of the step differences for the 5-Hz and 7-Hz tests, respectively. As expected, correct responses increased as the step differences increased for both tests. For the 5-Hz test, the one-step difference yielded 66% correct responses, while two-step difference yielded 78% correct responses. DL being defined as the 75% correct response level, DL for this set of data was located between the one and two step difference (which corresponded to .5% and 1.0% frequency extent difference, respectively). DL was determined to be .85% frequency extent using interpolation.

Table 1

Means and standard deviations of percent correct responses at each of the six step differences for the 5-Hz and 7-Hz tests.

Step Difference	Total No. of Responses	5-Hz % Correct		7-Hz % Correct	
		Means	SDs	Means	SDs
1	1100	66.2	10.0	65.7	10.4
2	1000	78.3	10.7	79.9	9.3
3	900	85.3	9.6	89.3	9.6
4	800	90.0	8.7	93.3	9.4
5	700	94.2	8.2	94.5	7.9
6	600	96.7	7.7	96.2	6.5

For the 7-Hz test, the one step difference yielded 66% correct responses and the two step difference yielded 80% correct responses. The interpolation procedure yielded a DL of .81% frequency extent. These two DL values in percent were converted to DL in semitones using the conversion formula:

$$\text{SEMITONES} = (12/.301)\log_{10}(1 + \%\text{value})$$

The conversion yielded DL of .15 semitones for the 5-Hz test and .14 semitones for the 7-Hz test with an average DL of .145 semitones. It should be noted that these DL values in semitones are from the mean frequency to the lowest (or highest) frequency in the modulated signals. DL in terms of frequency extent from the smallest to the greatest is derived by doubling these DL in semitones, i.e., .29 semitones.

As seen in the table, the extent discrimination scores increased monotonically. An ANOVA test was applied to examine possible effects of the modulation rates, step differences and their interaction. The ANOVA test indicated highly significant step difference effects ($F=234.27$, $df=5, 245$, $p < .0001$). The modulation rate and its interaction with step difference, however, were nonsignificant ($F=2.48$, $df=1, 49$ and $F=1.88$, $df=5, 245$, respectively). The post hoc Newman-Keuls test, which used combined mean scores from each of the six step differences, revealed that the means of each of the six step differences were significantly different from each other except the difference between the 5-step and 6-step conditions.

Discussion

DL for the Extent of Frequency Modulation

The primary purpose of the present investigation was to find the DL for extent of frequency modulations in vibrato and to gain insight why the singers in the King (1989) study could not match extent of frequency modulation but could match rate of modulation. In particular, a question was raised whether the singers failed to match extent accurately (1) because they could not discriminate the extent, or (2) because they could discriminate but could not control extent in their phonatory attempts.

DL was found to be .30 semitones (peak-to-valley range) for the 5-Hz modulation stimuli, and .28 semitones for 7-Hz modulation stimuli. Production studies (Gemelli et al., 1954; Horii, 1989a; Luchinger and Arnold, 1965; Mason, 1965; Mason and Zemlin, 1966; Seashore, 1938; Sjostrom, 1947; Vennard, 1967; Winckel, 1953) reported .5 semitones to 2.0 semitones as a typical range of frequency extent in vocal vibrato. The present results indicate that, in ideal conditions of sustained, steady vibrato vowels, one would be able to distinguish perceptually an extent difference of about .3 semitones.

How does the DL found in the present study (.3 semitones or about 1.7% frequency extent) compare with DL for formant transition? Mermelstein (1978) found an average DL of 9 to 14% of the formant frequencies, while Danaher and Osberger (1973) reported DL of 3 to 4%. Thus, the DL found in this study appears to be much smaller than DL for formant transitions. However, the current DL value, 1.7%, is in the range reported by Horst (1989). Horst (1989) used one-formant (2 kHz) stimuli with a triangular spectral envelope and near 100 Hz fundamental frequencies. The modulation pattern was bell-shaped, that is, the fundamental and harmonic frequencies rose and fell in the shape of a smooth bell only once. He found DL for the detection of modulation to be about 0.5 to 2% and DL for the discrimination of extent of modulation to be about 1.5 to 3%.

One study of an estimate of DL for extent of frequency modulation in vibrato-like signals was conducted by Horii and Scherer (1988). They reported DL of .32 semitones, comparable to the DL values found in the present investigation. These comparable DL values are noteworthy especially when one considers major methodological differences of the two studies. The Horii and Scherer study employed (1) triangular waves instead of vowels, and (2) the method of adjustment instead of paired comparison paradigms. Comparability of the DL values in these two studies help strengthen the validity of the findings.

Another method of gauging the perceptual ability of discrimination is to calculate the number of different stimuli discriminable from the lowest value to the highest value of the variable of interest. In other words, in the present situation, how many different stimuli could be distinguished

from 0.5 semitone extent to 2.0 semitone extent with the DL of .3 semitones? Starting with the .5 semitone extent, the next distinguishable stimulus would be the one with .8 semitone extent (.5 + .3) and the next distinguishable stimulus would be the one with 1.1 semitone extent (.8 + .3). Continuing the process, it would be found that only 6 different stimuli would be distinguishable from .5 to 2.0 semitone extent.

In actual production of vocal vibrato by singers, it is not possible to produce vowels with constant frequency modulation extent. Thus, in any vibrato vowel, there would be variability in the extent of the frequency modulation within a single production. In addition, other acoustic features such as frequency modulation rates, intensity and spectra would also vary in single productions. Thus, it is reasonable to expect much greater difficulty (and therefore, greater DL) in distinguishing different degrees of frequency extents. The DL value of .3 semitones, therefore, needs to be viewed as the lower limit of DL.

It is also noteworthy that variability of frequency modulation extent in sustained vibrato vowels was about .25 semitones (Horii, 1989b). In other words, even when singers were instructed to produce a sustained vowel with vibrato as "steadily or regularly" as possible, the frequency extent varied on the average about .25 semitones. This magnitude of extent variability is again comparable to DL values. It may be speculated that the singers indeed felt their vibrato productions were "steady" and "regular" since their extent variabilities were just barely noticeable or unnoticeable.

These considerations lead to a partial answer to the question of why King's subjects could not well match the extent of synthesized vibrato vowels. According to the present results, the target extent of .5, 1.0 and 1.5 semitones used in the King study should have been distinguishable if they were presented in a paired comparison paradigm. Since they were presented singularly, the target extent of 1.0 semitone, for example, could have been heard as having an extent somewhere between .7 to 1.3 semitones. In matching the target, due to the production accuracy and variability coupled with other factors such as memory and self-monitoring capability (which cannot be better than DL), singers could theoretically produce a vibrato vowel with an off-target frequency extent by as much as .6 semitones. The singers' poor performance on extent matching, therefore, can be attributed, in part, to the discriminability of the target extents and the controllability of extents of their own productions.

Nonsignificant Modulation Rate Effects on Extent Discrimination

The present study revealed nonsignificant effects of the frequency modulation rates of the 5-Hz and 7-Hz on the extent discrimination scores. This finding is somewhat unexpected because at the rising or falling frequency within one modulation, the rate of frequency change is obviously much faster in the 7-Hz modulation than the 5-Hz modulation signals. For example, the 5% extent signal at 130 Hz had a frequency change of 13 Hz (5% above 130 Hz and 5% below 130 Hz) within one modulation. The rate of frequency rise or fall for the 5-Hz modulation signal, therefore, was 130 Hz/second (13 Hz/100 millisecond). The rate for the 7-Hz modulation signal, on the other hand, was 182 Hz/second. A small percentage change in frequency extent, therefore, was associated with much greater change in the rate of frequency change in the 7-Hz modulation signal than the 5-Hz modulation signal. The present finding indicated that the listeners did not have better discrimination of extent for the 7-Hz modulation signal in spite of the greater changes in the rate of frequency change within one modulation. Had the single modulations of the 5-Hz and the 7-Hz modulation signals been paired and presented, it is not unlikely that different discrimination scores could

have been demonstrated. Further research on the use of repetitive (like the current stimuli) versus single modulations for frequency discrimination is warranted.

Significant Step Difference Effects

Through a pilot study, it was expected that the DL would be located somewhere between .5% (one-step difference) and 3.0% (six-step difference). Thus, it was not unexpected that the discrimination scores would result in an overwhelmingly significant step difference effect in the ANOVA test. What was not foreseen was the gradual, rather than sharp, improvement of the discrimination scores as the step difference increased (Figure 1). The discrimination scores reached a plateau only after the step difference was 2.5% (five-step difference) as evidenced by the post hoc Newman-Keuls test. The gradual, statistically significant, improvement of the discrimination scores up through the five-step difference appears to indicate that 1) the extent increment of .5% was adequate, 2) the discrimination errors, although decreased, persisted even at the large step differences possibly due to relatively long signal durations for this kind of paired-comparison tasks.

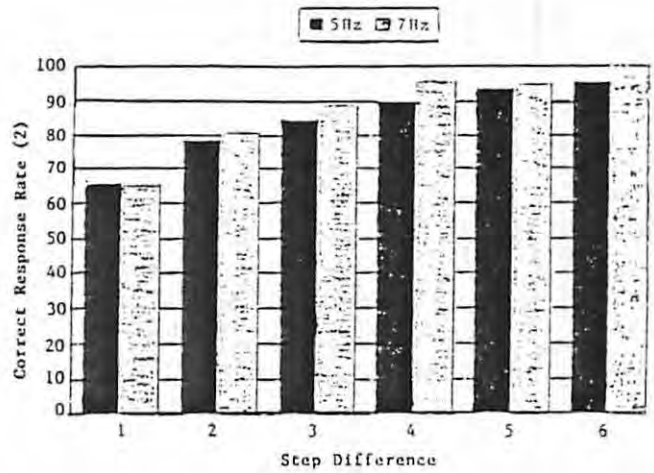

Figure 1. Percent correct responses in the 5-Hz and 7-Hz tests.

Tests of Assumptions Underlying the Research Design

There were two assumptions in the research design that could be tested post hoc. These were 1) gender differences and 2) music versus non- music majors.

The male subjects on the average scored 168.7 on combined 5-Hz and 7-Hz tests. The female subjects on the other hand scored 174.3. The results of a t-test supported the assumption that there would be no gender difference ($t=1.81$, $df=48$, $p>.05$).

The subject pool happened to include 9 music majors. Although the size of this subgroup was small and it was not in the original design, possible differences in the discrimination performances between the music and non-music majors were explored. The difference in the discrimination scores for both 5-Hz and 7-Hz tests between music major and non-music major subjects (176.85 vs. 170.12) was not significant at the .05 level ($t=.49$, $df=48$).

Conclusions

1. The DLs calculated in this study were .30 semitones for the 5-Hz tests and .28 semitones for the 7-Hz tests.
2. The effects of auditory discrimination at the 5-Hz rate of frequency modulation in vocal vibrato are not different from those at 7-Hz.
3. The effects of step differences are significant.
4. There is no effect from the interaction of modulation rates and step differences.
5. There is no gender difference in frequency extent discrimination.

6. The differences in the discrimination scores for both 5-Hz and 7-Hz test between music major and non-music major subjects are not significant.

Acknowledgments

This work was supported by grant 1 P60 DC00976-01 by The National Institutes on Deafness and Other Communication Disorders to The Denver Center for the Performing Arts through the National Center for Voice and Speech.

References

American National Standards Institute (ANSI). (**1978**). *Methods for Manual Pure-Tone Threshold Audiometry, (S3.21),* (New York: American National Standards Institute, Inc.).

Campbell, W.M. and Michel, J.F. (**1979**). "The effects of auditory masking on vocal vibrato," In *Transcripts Eighth Symp. Care Professional Voice,* edited by V. Lawrence and B. Weinberg, 1:50-55.

Danaher, E.M., and Osberger, M.J. (**1973**). "Discrimination of formant frequency transitions in synthetic vowels," Journal of Speech and Hearing Research, 439-451.

Downie, N.M. and Heath, R.W. (**1965**). *Basic Statistical Methods, (2nd ed.).* (Harper and Row, New York).

Gemelli, A., Sacerdote, G., and Bellusi, G. (**1954**). "Analisi electroacustica della voce contat," Boll. Soc. Ital. Fonet Sper., 3-8.

Horii, Y. (**1989a**). "Frequency modulation characteristics of sustained /a/ sung in vocal vibrato," Journal of Speech and Hearing Research, 32, 829-836.

Horii, Y. (**1989b**). "Acoustical analysis of vocal vibrato: a theoretical interpretation of data," Journal of Voice, 3-1:36-43.

Horii, Y., and Scherer, R.C. (**1988**). "The jnd's for rate and extent of frequency modulations," J. Acoust. Soc. Am., 83 (S1), S31(A).

Horst, J. W. (**1989**). "Detection and discrimination of frequency modulation of complex signals," Journal of the Acoustical Society of America, 85, 2022-2030.

Keidar, A., Titze, I. and Timberlake, C. (**1984**). "Vibrato characteristics of tenors singing high C's," In *Transcripts Thirteen Symp. Care Professional Voice*, edited by B. Lawrence and B. Weinberg, 105-110.

King, J.B. (**1989**). *Control of Rate and Extent of Frequency Modulation in Vocal Vibrato by Trained Singers,* Doctoral Dissertation. University of Colorado.

King, J.B. and Horii, Y. (**1989**). "Rate and extent control of vibrato by trained singers." ASHA, 31, 155.

Luchsinger, R., and Arnold, G. (**1965**). *Voice-Speech-Language.* (Wadsworth, Belmont).

Mason, R.M. (**1965**). *A Study of the Physiological Mechanisms of Vocal Vibrato,* Unpublished doctoral dissertation. University of Illinois.

Mason, R.M. and Zemlin, W.R. (**1966**). "The phenomenon of vocal vibrato," The NATS Bulletin, 22:12-17.

Mermelstein, P. (**1978**). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," Journal of Acoustical Society of America, 63:572-580.

Moore, B. C. J. (**1976**). "Comparison of frequency DL's for pure tones and modulated tones," British Journal of Audiology, 10, 17-20.

Moore, B. C. J. and Glasberg, B. R. (**1989**). "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation," Journal of the Acoustical Society of America, 86, 1722-1732.

Ruhlman, J. (**1984**). "Science Session III," In *Transcripts Thirteen Symp. Care Professional Voice*, edited by V. Lawrence and B. Weinberg, 118-129.

Seashore, C. (**1938**). *The vibrato. Studies in the Psychology of Music*, (Vol. 1), University of Iowa, Iowa City.

Shipp, T., Leanderson, R. and Sundberg, J. (**1979**). "Vocal vibrato/rate and extent of vibrato as a function of vowel, effort and frequency." In *Transcripts Eighth Symp. Care Professional V Voice*, edited by V. Lawrence and B. Weinberg, Part I:50-56.

Shower, E. G., Biddulph, R. (**1931**). "Differential pitch sensitivity of the ear," Journal of the Acoustical Society of America, 3, 275-287.

Sjostrom, L. (**1947**). "Experimental-phonetischeutersunchugen des vibratophanomes der sinstimme," 10th Nord, Otolaryngol. Kongr. Stockholm.

Titze, I. (**1983**). "Synthesis of sung vowels using a time-domain approach," Transcripts Eleventh Symp. Care of the Professional Voice, Part I: Scientific Sessions.

Vennard, W. (**1967**). *Singing: the mechanism and the Technique*. (Fisher, New York).

Winckel, F. (**1953**). "Physikalische kroterien fur objecktive stimmbeurilung." Folia Phoniatrica, 5:232.

Part II

Summaries and tutorial reports of Training, Dissemination and Continuing Education

NCVS Status and Progress Report - 1
June 1991, 121-123

Training Update

John Folkins, Ph.D., Training Coordinator

Department of Speech Pathology and Audiology, The University of Iowa

The need to attract among the brightest and best to careers in biomedical research has been well-documented. Although enrollment has increased over the past 15 years in business, computer science and engineering, 36 to 44 percent drops have been noted in the number of students embarking in research careers in biological sciences, physical sciences, and social sciences (Kenneth Green, 1985). By the year 2004, experts predict a shortfall of 8,000 Ph.D.s in the natural and physical sciences within the United States (McIntosh, 1989).

Furthermore, the need for research personnel in the field of communication sciences and disorders has been described as "critical" in the National Strategic Research Plan of NIDCD (April 1989). Training of such individuals was listed as a first priority. With the award to establish the National Center for Voice and Speech, aspiring researchers are provided funding to allow them to prepare for research careers as predoctoral students, postdoctoral trainees, or visiting scholars.

In general, the structure of the NCVS traineeships are quite flexible. All trainees are offered the opportunity to be involved in the research process and open discussion of ideas. The NCVS philosophy stresses open-minded scholarly inquiry and learning skills more than memorizing existing theories and facts related to specific topic areas.

Furthermore, NCVS personnel believe that many approaches to voice and speech, as well as other areas of communication disorders, need to be integrated with other areas of science. For example, our speech and voice research program cuts across the domains commonly associated with physiology, clinical psychology, experimental psychology, otolaryngology and music. Obviously, researchers need to be aware of these areas to be able to avoid building theories that can be challenged by insights already well established in other areas. Perhaps, more importantly, researchers with a good exposure to many knowledge bases will have more options for building theories that extend our understanding of voice and speech disorders. NCVS members are committed to meeting this end by accepting trainees with backgrounds in a variety of areas and introducing them to research in voice and speech.

All trainees will be involved in number of supervised research projects throughout training. This might occur as a combination of individual projects and projects done in collaboration with faculty or other trainees. These projects are expected to result in convention presentations and published manuscripts. Trainees are encouraged to do research projects with faculty who are not their mentors and faculty identified with a variety of disciplines.

Additionally, trainees will be required to attend the yearly NCVS conferences. Attendance and participation in weekly Proseminars for trainees at the University of Iowa is also required. Proseminar presentations are based on finished projects, proposed research, open discussions, or guest speakers from other programs. Every investigator, postdoctoral trainee, and predoctoral candidate presents at least annually.

Predoctoral trainees

Predoctoral trainees are anchored in the doctoral degree program in the Department of Speech Pathology and Audiology, located in the Wendell Johnson Speech and Hearing Center at the University of Iowa. All of the proposed training programs are expected to span a number of academic disciplines and take full advantage of all consortium sites.

Each predoctoral student is required to develop and defend a plan of research experiences and coursework that will meet the competencies required for his or her research goals. Competencies in four tool areas are required: research design, analog instrumentation, computers and mathematics.

Each doctoral student in the Department of Speech Pathology and Audiology is reviewed yearly by the faculty. The review includes student performance in research projects, coursework, and clinical activities. The outcome of the review is reported to the student by his or her advisor.

Postdoctoral trainees

To be considered as a candidate for an NCVS postdoctoral fellowship, the applicant must have either successfully completed a Ph.D., M.D. or the equivalent. The consortium allows the postdoctoral fellows to use the clinical resources of the University of Iowa, University of Wisconsin-Madison, and the University of Utah.

Like the predoctoral trainee, the postdoctoral fellow will meet with a group of five faculty members--chosen by the trainee in consultation with the faculty advisor--within the first few months of training. Prior to this meeting the trainee and advisor work out a specific statement of career goals; a description of competencies brought into the program; and a list of needed competencies, experiences and skills to prepare optimally for the stated career goals. The committee discusses these materials, makes alterations as appropriate, and approves a plan of study.

When possible, planning meetings for postdoctoral trainees who are located at non-Iowa sites will be held during the regular yearly NCVS conferences.

At the end of the trainee's research experience, he or she will meet with the same five training faculty members that were involved with the planning meeting. The training experiences will be critiqued by all members of the group. A written report summarizing the meeting will be prepared by the trainee's mentor and submitted to the training director.

Two candidates were selected as NCVS postdoctoral trainees in Summer 1991.

Linnea Peterson received her medical degree from the University of Iowa in May 1991. During her traineeship, she plans to work most directly with Henry Hoffman in the Department of Otolaryngology at the University of Iowa.

David Berry earned his Ph.D. in physics from Brigham Young University in Spring 1991. He plans to collaborate with Ingo Titze and specialize in signal analysis of voice.

Visiting scholars

Visiting scholars may train at the University of Iowa, Denver Center for Performing Arts, University of Wisconsin-Madison, or University of Utah. Training periods may vary from two

weeks to six months. Visiting scholars are established researchers for whom the training visits are expected to introduce new scientific areas, allow them to acquire new research skills or update and enhance existing research skills. Just as importantly, visiting scholars will bring new ideas and perspectives to NCVS investigators on a continuing basis.

Because the focus of the visiting scholarship is the mature scientist, planning meetings are deemed inappropriate. However, visiting scholars are expected to give presentations covering their planned research projects. Such seminars are open to all NCVS investigators and mentors and a format encouraging discussion is encouraged.

Carlin Hegeman, Ph.D., a professor at the University of Northern Iowa, will join the NCVS as a visiting scholar beginning in August. During his six-month tenure at the University of Iowa, Hegeman plans to work with John Folkins, Jerry Moon and Donald Robin on the motor control of the velum during speech.

NCVS Status and Progress Report -1
June 1991, 125-126

Dissemination Update

Ronald Scherer, Ph.D., Dissemination Coordinator

Recording and Research Center, The Denver Center for the Performing Arts

The physical and emotional well-being, and even a person's livelihood, may be threatened by a voice or speech disorder. Yet, affected individuals and the population as a whole know too little about protection, possible rehabilitation and enhancement of voice and speech.

As many as three percent of the adult population suffers from voice disorders, and as many as one in ten of the general population experiences a speech or voice disorder some time during his or her life. Estimates of voice disorders in children range between 6.1 percent and 9.1 percent. Stuttering prevails at 3.4 percent in male children and 2.0 percent in female children in the 6-11 year age group. Voice and speech disorders occur in 75 percent of all Parkinson's patients, a disease that affects 1 in 100 individuals over 60 years of age.

The main thrust of the dissemination project of the NCVS is to distribute information about (1) voice disorders--not just to professional users of the voice, but to the general public--and (2) stuttering.

Groups which may benefit most from information are the parents, relatives, teachers and care providers for young children who may be at risk to develop stuttering; and teachers, singers, lawyers, actors, ministers, salespeople, broadcasters, and other individuals who depend upon their voices in their careers. The NCVS dissemination group is especially targeting information to these individuals.

Project leaders also hope to inform the general public about the factors that effect these disorders, what can be done to prevent them, what can be done to habilitate them, what can't be done, and what is not yet known. Armed with additional knowledge, people can better understand the disorders observed in themselves and others, take appropriate preventative measures, seek professional help, and develop realistic expectations.

Stuttering

Research over the past decade has shown that beginning stutterers and children at risk for developing stuttering can be reliably differentiated from their normally fluent peers on the basis of specific characteristics of their speech disfluencies (Zebrowski, Conture, and Cudahy, 1985). In addition, recent discussions in the clinical literature in stuttering, motivated by studies of parent-child verbal interaction (Zebrowski and Conture, 1989), have provided suggestions for helping parents establish a fluency-enhancing environment for their stuttering children.

One goal of the project is to offer such suggestions to the public. While early diagnosis coupled with information-sharing and counseling with parents may not prevent the emergence of stuttering, it may prevent the development of a chronic, long-term problem.

In addition to issues related to childhood stuttering, a goal of the Center is to provide adults who stutter with information about evaluation and treatment to allow them to select the most appropriate clinician. Materials may help them evaluate their options for different approaches to therapy.

Voice Disorders

Special concern is directed to the professional voice user: salespeople, teachers, lawyers, actors, singers, ministers, broadcasters, etc. Other important consumers of information about voice care are those with disease conditions such as asthma, those who live in adverse conditions (i.e., excessively dry air) or those with nerve damage or complications from intubation. Informing these groups about preventive care, early warning signs, and steps to take at early stages, may greatly reduce the likelihood of the development of voice disorders.

To meet the goals of reaching individuals who may experience or are at risk of developing a voice or speech disorder, project members are in the process of: reviewing existing materials and make specific decisions about what to offer to the public and under what circumstances; determine areas most in need of additional materials (workshops, brochures, videotapes, pamphlets, audiotapes, etc.); produce new materials; distribute these materials to the public; and evaluate the effectiveness of the materials.

Feedback forms, distributed and collected at the end of all workshops, allow the group to evaluate their effectiveness in providing new and understandable information. Secondly, careful tabulations are kept for requests for brochures, pamphlets, etc. Finally, project members continue to follow research in progress and longitudinal studies to provide the most up-to-date information to the consumer as possible.

Continuing Education Update

Diane Bless, Ph.D., Continuing Education Coordinator

Department of Communicative Disorders and
Department of Surgery, Otolaryngology Division, University of Wisconsin-Madison

The pace of scientific advances underscores a critical need for continuing education in all areas of science. However, for several reasons, a specific need for continuing education exists for practitioners in voice and speech. There appears to be a lag between new findings in the laboratory and clinical practice. Also, consumers are entitled to expect the service they receive is a quality, state-of-the-art product. Finally, the process of keeping professionally current has become increasingly difficult for practicing clinicians.

To correct these problems, NCVS members are developing workshops and symposia, and reviewing papers, software and videotapes to help meet the continuing education needs of speech-language pathologists, voice teachers, drama coaches and otolaryngologists. These continuing education activities are designed to bridge the gap between the laboratory and clinical practice, and to provide easy access to new information; videotape modules and software can be used in the privacy of one's home or office at a time that is convenient to the busy practitioner. Software has the additional advantage of providing immediate feedback regarding knowledge of results or effectiveness of learning.

One goal of the NCVS is to make learning easily accessible. To this end, the NCVS Status and Progress Report will include review papers and lists of materials developed by NCVS staff. The CE Review section in this progress report consists of four papers written by NCVS investigators: "Aerodynamic Assessment in Voice Production" by Ronald Scherer, "Use of the Electromyogram as a Diagnostic Aid for Voice Disorders" by Erich Luschei, "The Role of Phonation in Speech Intelligibility: A Review and Preliminary Data from Patients with Parkinson's Disease" by Lorraine Ramig, and "Speech Therapy for Neurological Disorders of the Larynx" by Lorraine Ramig and Ronald Scherer. These papers are meant to further the education of practicing clinicians. They identify and discuss issues not found in texts or research papers that are important to the application of the measures to assessment and treatment of the clinical population.

NCVS Status and Progress Report -1
June 1991, 129-149

Use of the Electromyogram (EMG) as a diagnostic aid for voice disorders

Erich S. Luschei, Ph.D.

Department of Speech Pathology and Audiology, University of Iowa

Harry T. Hoffman, M.D.

Department of Otolaryngology - Head and Neck Surgery, University of Iowa

This document is intended for the participants of the Conference on the "Scientific Advances in the Assessment and Treatment of Voice Disorders: An Integrated Apporach." Please do not cite it as a publication.

Introduction

I (Erich Luschei) am a neurophysiologist who has conducted, for approximately 20 years, research concerned with how the nervous system controls movements of the eyes, mandible, pharynx, and larynx. Much of this research has involved the use of experimental animals, but a significant fraction has been with human subjects. Many of my studies have involved the recording and analysis of the electromyogram (EMG).

In several of my experiments in recent years, recording laryngeal muscle EMG has been the major investigative tool. Besides helping conduct the studies, I have served as a subject on numerous occasions, and had EMG electrodes placed in my laryngeal muscles approximately 20 times. More recently, I have begun to work closely with an otolaryngologist, Dr. Harry Hoffman, and other professionals at the University of Iowa to develop an on-going program of diagnostic electromyography of laryngeal muscles. Invasive techniques are not warranted in all patients with voice problems, certainly, but in particular cases, information gained using these techniques can provide clear answers to important questions. These techniques are complicated enough, however, that they cannot be effectively used in a casual manner. It is my strong feeling that a program dedicated to this purpose has to be approached as a major commitment of time and energy, with the

realization that it will only become truly effective as experience is gained in both the technique and interpretation of the recordings.

Although I have written most of this document, Harry Hoffman has provided critical input concerning patient procedures, and has provided novel ideas about applications and improvments of the techniques. Dr. Hoffman is not, however, to be held responsible for any of the silliness the reader will encounter below.

Recent History of Laryngeal EMG Recording

Diagnostic EMG recording in laryngeal muscles is not entirely "uncharted" water. Several Japanese otolaryngologists, in particular Dr. Minoru Hirano, have used this procedure for many years. I am fortunate to have had Dr. Hirano on the other end of the needle most of the time I have been a subject. Another productive laboratory has been that directed by Dr. Christy Ludlow at the National Institutes of Health. This laboratory has, while attempting to treat patients with complex problems involving the voice and speech, used laryngeal recording in a consistent manner, quantitatively analyzing records, and correlating measures with problems and changes in response to therapy. Much of what I know about the potential and problems of this procedure has come from careful study of Dr. Ludlow's and her colleagues' publications, and by direct consultation with her. A few of the references covering the work of these and other investigators is given at the end of this document.

Uses of Laryngeal EMG Recording: Some Examples

It is probably premature to attempt to characterize all of the voice disorders that could be better understood using laryngeal EMG recording. The "frontiers" of this endeavor will probably be pushed forward in the context of research programs, but it is my presumption that most people at this conference are more interested in what might be learned on the basis of our current knowledge. Two examples may be illustrative in this regard.

Case 1: An elderly woman who is slowly recovering from Guillain Barré syndrome has regained appreciable strength in the limbs, but her vocal folds remain adducted to the degree that she must continue to use a tracheostomy to breath. Are her laryngeal muscles still paralyzed, or are they active but unable to move the vocal folds because the long period of immobility has resulted in a greatly reduced range of motion of the arytenoids? The absence of any EMG activity in the thyroarytenoid muscle during maneuvers that ordinarily produce adduction would suggest continued paralysis; clear evidence of EMG activity would suggest the latter conclusion. But if you don't see any EMG activity, how do you know you've placed the electrode in the TA? Maybe the electrode is in non-muscular tissue. More on this case later.

Case 2: A patient has stridor on inspiration, so much so that she finds it difficult to maintain normal ventilation even when at rest. Her vocal folds remain adducted while breathing. This could be because of paralysis of the PCA, which ordinarily is active on inspiration. Or this condition could result from unusual activity of the adductors during inspiration. Ideally, records from both PCA and TA would be best, but PCA is difficult to record from directly. TA, on the other hand, is relatively easy. Absence of TA activity on inspiration would make PCA paralysis probable. On the other hand, high levels of activity in TA during inspiration would suggest an organic problem with the

coordination of the respiratory-laryngeal system (which presumably involves only brainstem neurons), or the patient has a "functional problem."

Traumatic injury to the larynx can produce immobility of a vocal fold either by arytenoid subluxation or damage to the recurrent laryngeal nerve. Hoffman et al. (1991) have used a combination of diagnostic procedures, including intraoperative EMG recording, to make this diagnosis in several patients.

Another potential area of application is in the injection of botulinum toxin in the laryngeal muscles. Some physicians who perform this procedure find it useful to record EMG activity from the tip of an insulated injection needle to verify that the tip is in a laryngeal muscle rather than in non-muscular tissue or even in the airway (Blitzer et al., 1988; Miller et al., 1987). Once the instrumentation for EMG recording is in place, such recording can very easily be added to the injection procedure.

Monitoring of muscle activity during certain surgical procedures may be used to warn of imminent damage to motor nerves. Electromyographic recording of facial muscles is routinely used in some hospitals to prevent damage to the facial nerve during surgery on the middle and inner ear. It seems very possible that a similar approach, monitoring the activity of laryngeal muscles, could prevent accidental damage to the recurrent laryngeal nerve during explorations and surgery in the neck. It must be realized, however, that such monitoring has be done by individuals who understand the source of the EMG, and can distinguish EMG from the extraneous signals ("artifacts" to a neurophysiologist) that can plague the technique. To quote a scrub nurse in a university hospital: "Yeah, we have a machine that monitors the facial muscles, but it gives a warning so often that mostly we turn it off or ignore it." In at least one other major hospital setting known to the presenter, such monitoring is done by an electromyographer. That is an excellent situation, but my feeling is that non-specialists could be very reliable at detecting abnormal EMG activity if they were trained to do so. One of my purposes in this document is to try to provide some of that training.

Basic Instrumentation

Amplifiers

To detect and study the activity of laryngeal muscles requires one or more differential amplifiers which have a selection of gains between about 1000 to 50,000 (60 dB to 94 dB) between 120 Hz and 5 kHz. These characteristics are easily obtained with modern electronics. Some EMG specialists might like the lower cutoff frequency somewhat lower, but being familiar with the clinical situation, in which the 60-Hz noise problem cannot always be eliminated at the source, I've come to the conclusion that attenuating the response in the neighborhood of 60 Hz is a sensible (if cowardly) way out. One could pay for a lot more high frequency response, but there is virtually no EMG information above 5 kHz.

A very important feature of the amplifier is its "common mode rejection ratio" (CMRR). When used with the proper electrode configuration, an amplifier with a high CMRR is the most effective way to prevent amplification of the 60 cycle electrical fields which are the bane of all who would dare to try and record the EMG outside of an electrically-shielded facility.

There are several companies who manufacture amplifiers which are well-suited to recording EMG activity. All of them have specifications which are very close to being "ideal". One feature that I feel is necessary for recording with human subjects is not, however, available on all brands. I

strongly feel that human subjects should be electrically isolated from Earth ground during these procedures. This requires the use of an isolation stage. I will discuss this issue in more detail in a "Break-out" session later in the conference. The cost of a suitable amplifier could be marvelously variable. For those to whom money is not an object, there is hospital equipment manufactured specifically for diagnostic electromyography labs such as those found in major neurological clinics. But one doesn't have to pay so much. One of the finest amplifier systems I have examined is manufactured by a small company in Madison, Wisconsin (Biocommunications Electronics, Inc), and its cost is very reasonable. For the "do-it-your-selfer," a schematic of the single-channel portable amplifier which we have been using will be provided gratis to anyone who is interested. Cost of the electronic components is about $100. Note that while this amplifier would meet all requirements that I know of for electrical safety, the mechanical layout and fabrication techniques can be important: whoever puts the system together should be familiar with fabrication of medical electronics.

Note that all differential amplifiers have three leads. One input is usually called "ground" although I URGE you to NOT think of it as a "ground." It is better thought of as a reference voltage which is treated as a "Zero" with respect to the other two leads, which are the inputs to the amplifier. This "reference" lead should never be connected to Earth ground. It is attached to the subject's skin over a non-muscular regions. The forehead works well for this purpose. The signal which is amplified by the amplifier is the difference in the voltage between the other two leads. These two leads are usually attached to a bipolar electrode, which has two exposed metallic surfaces located close together. When these two exposed surfaces are in a muscle, a difference in voltages between them can only be produced by electrical events close to the electrode; i.e. differential recording with a bipolar electrode selects events close by, and "rejects" far-field events such as muscle activity in remote muscles and the infernal 60-cycle field.

Monitors of the EMG

Audio monitors: A type of monitor that has traditionally been used with the EMG is an audio amplifier connected to a loudspeaker. This type of monitor has several advantages. It is inexpensive, you don't have to watch it, and the human can, with some experience, use his or her "ears" to perform an acoustic analysis that can provide critical on-line information. For example, the sound of muscle activity that is close to an intramuscular electrode sounds "crisp" and "cracking" because it has mainly high-frequency components. EMG from muscles located at greater distance has more of a "rumble" to it, like "surf." This is because the tissue that intervenes between the EMG source and the electrodes acts like a low-pass filter (DeLuca, 1979). Single motor unit "spikes", which are produced by the near-simultaneous action potentials in the muscle cells innervated by the same motoneuron, produce a "crack" or "pop" which has a unitary sound. The ear can quickly detect the pattern and rate of firing of a single motor unit. In fact, a subject who can control his or her muscles can quickly learn to turn a unit on and off by listening to its discharges.

The limitation of an audio monitor is that it provides no permanent record, and the amplitude of the EMG cannot be determined with any precision. It also cannot reveal the presence of some problems such as high frequency interference (as from a near-by video monitor) and large low-frequency movement artifacts. Another problem with audio monitors is that they produce a terrible 60 cycle din when the electrode is removed or a wire becomes detached. A very good idea is to have an audio monitor with a handy volume control, or an assistant who can anticipate the "blast" and turn the volume down before it happens.

Oscilloscopes: I personally would never approach the challenge of using laryngeal EMG to study a subject or patient without an oscilloscope. It can be used to reveal the many types of artifacts that interfere with EMG recording, and by knowing the gain of the amplifier which is being used, it can be used to estimate the actual amplitude of the signals. Waveforms which can't be discriminated by the audio monitor carry potentially important information. If an oscilloscope were to be purchased for this use, then a "low-end" digital oscilloscope, costing about $1000, would be quite adequate. A digital oscilloscope can "freeze" a picture "on-line", and thus the EMG signal can be studied in order to make a decision during the procedure. Older analog oscilloscopes are still quite useful, however, for verification of the signal's basic characteristics.

Recorders: It will probably be very important for any person or group of people making decisions based upon EMG recording to be able to directly compare records from different patients. There should be, therefore, some means of recording these data. Recording systems can be VERY expensive, but if you will promise NEVER to tell anyone, I will reveal a deep dark secret. A decent stereo cassette tape-recorder from your local commercial audio store will do just fine for recording EMG. You only have to choose a recorder that allows you to set the record level, i.e. it ought to have a knob or slide lever that sets a VU meter or equivalent display. In many instances, you will only need one channel for the EMG. You can use the other to record the voice of the patient or as a "chatter" channel. In order to avoid "clipping" the signals, however, you will need to calibrate the record level meter by putting in a 1 Khz sinewave at some set amplitude (within the range of the "line-in" of the tape recorder), and determining what record level setting produces a 0-dB level. Leave this setting on the record-level control, and just choose an amplifier gain that produces an "average" amplitude the same as that used for the test sinewave.

Of course if one needs more channels and/or DC capabilities (if perhaps you were also recording intraoral pressure), then "instrumentation" recorders, which come in many forms--all expensive--are needed. In a few years, however, it is likely that the cost of research-grade digital tape recorders will plummet from their current atmospheric realm.

Note that a recorder is basically useless unless one has the ability to analyze the records. For this purpose, an oscilloscope or access to a computer with an analog-to-digital converter is necessary. I will discuss various approaches to quantitative study of EMG records in a later section.

Electrodes

Concentric needle electrodes: There are basically two types of bipolar electrodes that are used for intramuscular recording: concentric needle electrodes (which are rigid), and "hooked wire" electrodes (which are made of fine wire and are flexible once the insertion needle has been removed).

In most EMG diagnostic clinics, such as those found in departments of neurology, concentric needle electrodes are used almost exclusively. This is because the waveform and firing patterns of single motor units are of primary interest in diagnosing most neurological diseases that affect muscle activity. These electrodes can be advanced and withdrawn by the clinician to isolate particular units. When used in muscles directly under the skin during isometric contraction, in which there is relatively little actual movement of the muscle tissue, these electrodes are relatively painless to the patient, and do little damage to the muscle. When used to record from laryngeal muscles, however, this type of electrode has serious problems. First of all, "gross" EMG, as distinct from single motor unit activity, is useful for determining if and how much a muscle can be activated. Concentric

bipolar electrodes are very selective for single motor units, so one might see little, if any, evidence of activity in a normal muscle, even at fairly high levels of activation.

A more serious problem, in my view, is that movement of a muscle with a rigid needle in it hurts like the dickens. I've experienced this several times. Pain from this source may have large effects on the motoneuron pool which one is attempting to study, as well as hurt the patient. If the larynx remained very still, this probably wouldn't be a significant factor, but large, vigorous, involuntary laryngeal movements can occur even with a very cooperative patient as a result of the stimuli that produce swallowing or coughing. Significant intramuscular damage could be caused by a large laryngeal movement with a rigid needle in it. This is because the larynx moves relative to the skin. The place where the needle penetrates the skin acts as a point of angular rotation as the larynx moves up and down. One only need imagine what goes on inside a ripe tomato if you stick a hypodermic needle halfway into the tomato and then move the hub of the needle up and down. The laryngeal muscle are relative small, and they contain important sensory and motor nerve branches which could be damaged by the moving edges of the needle.

If an investigator topically anesthetized the laryngeal mucosa to prevent coughing, and used suction to remove saliva to prevent swallowing, rigid needles would not be a serious problem. However, several of the maneuvers that one usually uses to determine which laryngeal muscle is being recorded involve large laryngeal movements, in particular swallowing and pitch changes. Another factor to consider is that concentric bipolar electrodes are quite expensive, and they are now used as a disposable item because they cannot be sterilized with gas or steam.

Had you not guessed by this time, I will admit that I much prefer the use of bipolar hooked-wire electrodes for recording from laryngeal muscles.

Bipolar Hooked-wire electrodes: Hooked wire electrodes have been used for laryngeal EMG recording for many years (Hirano and Ohala, '69). They are made with very fine, flexible, insulated wires. For recording gross EMG (as distinct of the attempt to record single motor units), about one millimeter of insulation is removed from each wire. These two wires are placed in the lumen of a suitable hypodermic needle so that the bared portions extend just beyond the tip, and then the ends of the wire are bent over to form a "hook." When the needle is placed into the muscle and then withdrawn, the hooks catch in the muscle and the wires are left behind in the muscle. Because of the flexibility of the wires, the muscle may make large movements without causing pain. Most people find these electrodes quite comfortable. They may be left in place for long periods of time during swallowing, speech, and singing.

While there are several good reasons for using hook-wire electrodes, they have the major disadvantage of not permitting adjustments of the electrode position once the hypodermic needle has been withdrawn. Obviously the electrode cannot be pushed further into the muscle. One could imagine (naively) that one could gently tug on the wires and pull them along the route of the wires through the muscle. To do this, however, the hooks have to be straightened out, and once this is done, the wires do not generally remain in the muscle. For all practical purposes, one gets one attempt per needle. Most people who use these electrodes generally have a stack of extra electrode assemblies in order to try a second or third time to get the wires into the desired muscle.

Choosing the wire for these electrodes is a compromise between using wires that are too ductile and wires that are too stiff. Very ductile wires can be inserted with very fine needles, and cause no sensation in the muscle. But their hooks are very weak, and they often do not "hook" into the muscle, and get pulled out with the needle. On the other hand, their weak hooks probably do minimal damage when these wires have to pulled out after the recording is over. I have tried 0.001

inch diameter stainless steel, and while I loved having it IN the muscle, it generally didn't stay there. In our current efforts, we have been using 0.003 inch diameter stainless steel wire. It generally gets and holds the "hook," and produces only a slight "funny feeling" in the muscle. It probably damages some muscle cells when the hook is pulled out (see Fig. 1 in later section), but there is no pain associated with this event, so the damage is probably small. Stainless steel wires larger than 0.003 inch diameter are too stiff in my experience. They produce a sensation that borders on discomfort, and they hurt when they are pulled out. A list of some of the vendors who sell wires suitable for these electrodes is given at the end of the document.

Once the wire is chosen, one has to select a needle size. We use a 1.5 inch 25 gauge hypodermic for a pair of 0.003 inch diameter wires which have an insulated overall diameter of 0.004 inches. In some cases, one can choose a needle that will accept the wires, but the fit is so close that the wires often "hang up" when the needle is withdrawn, so the wires get pulled out. Make sure the wires have enough room in the needle to accommodate slight bends in the wire.

When placing the electrode in a package for sterilization, it is important to use a relatively large package so the wires can be loosely looped in a way that allows them to be straightened out without forming kinks when the package is opened. Once formed, these kinks are impossible to remove: one might as well discard the electrode.

Stripping the insulation from ends of the wires that are to be connected to the amplifier may be done before or after the electrode is assembled. Connecting the amplifier to these wires can be more troublesome than one might suppose. If one uses very long pieces of fine wire, then these leads may be taken directly to the inputs to the amplifier and put under a screw-type terminal (gold-plated surfaces are a VERY good idea). However, these long wires have a tendency to become tangled and form kinks. Shorter fine-wire leads work better. One may take light flexible copper wires ending in a spring-loaded clip from the amplifier to the fine-wire electrodes. Whatever type of clip one uses, keep in mind that the very small size of the wire requires a close tolerance fit of the "jaws," and that an oxide coating on either the wires or the clip's surfaces can cause electrical "noises" that can interfere with recording. Test clips with gold-plated surfaces are, for this reason, much to be preferred (for example Pomona Model 4233 "Micrograbber").

Injection Needle Electrode: As noted earlier, it is helpful to be able to record muscle activity from the tips of the needle used to inject botulinum toxin. A special needle may be purchased for this purpose. It has an insulating lacquer over the length of the shaft, with approximately one millimeter bared at the tip. One of the inputs to the amplifier should be clipped to the hub of the injection needle during insertion. The other amplifier lead should be attached by a good skin electrode on the forehead next to the reference electrode. You will be recording differentially between the forehead and the injection needle. However, one may assume there is little muscle activity under the forehead electrode, so the EMG which is seen will be from the needle tip. One might suppose you could simply attached the unused amplifier input directly to the amplifier reference and save the trouble of attaching another skin electrode. You could do this, but it will greatly exacerbate the 60-cycle noise problem if any such noise is present. This is because shorting one of the input electrodes to the reference input ruins the common-mode rejection ability of the differential amplifier. When using the injection needle electrode, keep in mind that it is capable of "seeing" muscle activity at a greater distance than bipolar electrodes that are close together. Nevertheless, activity close to the needle tip will sound "crisp" whereas activity at a distance will have the duller, rumbling, sound. In particular, if single motor units are observed it is certain the injection needle tip is in muscle.

Preoperative Evaluation and Preparation of the Patient

Examination and Consent: The head and neck of the patient should be examined by a physician, including examination of the larynx using either indirect or fiberoptic laryngoscopy. Preoperative documentation of vocal fold movements and voice may be obtained at this time if so desired. Alternatively, video and audio recordings may be obtained during EMG recording, recognizing that there are potential problems of patient stress associated with the insertion of laryngeal EMG electrodes.

One of the primary goals of the examination is to discover whether there are contraindications to the procedure. Contraindications would include:

1. A bleeding disorder (coagulopathy, coumadin, aspirin.)

2. Altered anatomy precluding percutaneous placement of needle electrodes (infections or previous surgeries on the neck or laynx.)

3. Allergies to any of the drugs to be used, in particular the local anesthetic used at the electrode insertion site.

4. The inability of the patient to cooperate (young children, individuals with psychiatric disorders, or those with severely compromised health.)

Consent should be obtained from the patient. The following is an example of the way we discuss the procedure with the patient:

> "We would like to see how some of the muscles in your voicebox are working. To do this, we need to insert a needle with the electrodes inside of it through the skin on the front of your neck. We will deaden the skin where the needle is inserted. Insertion of the needle generally does not cause discomfort, but the needle sometimes makes you feel like you need to cough. However, please do not cough or swallow if you can help it while we are inserting the needle. After a little while, the feeling that you need to cough will go away.
>
> We have done this procedure many times without any problems, but there are some possible complications such as bleeding, infection, reaction to the anesthetic, or damage to structures in the region. These are rare, but they could affect your voice, swallowing, or breathing. Quite a few patients report that they feel like they have a minor sore throat for a day or two after this procedure."

If the plan during the recording is to have the patient produce singing-like phonation, it is a good idea to demonstrate the maneuvers to be used prior to the actual recording session. The patient is less stressed then, and one has more opportunity to do a little voice training. Have the practice until he or she seems comfortable with the task (unless, of course, such practice would be contraindicated by their voice disorder).

The patient generally is not sedated. Oxygen by a nasal cannula should be available. Equipment for and personnel experienced in airway management should be readily available ("crash cart", 14 gauge angiocath, and trach set.)

Preparation of the patient: The patient should be placed in a comfortable supine position with the head supported during the electrode insertions. The back may have to be elevated, however, if the larynx is to be observed with an endoscope during the insertion to see the needle enter the vocal fold (see below). Whatever position is used, the neck will need to be in the neutral position. Topical sprays (we use an aerosol of neosynephrine/pontocaine mixture) can be used if nasoendoscopy is to used, but no topical anesthetic is used on the larynx. Palpate the neck to identify the midline, cricoid cartilage, lower border of the thyroid cartilage, thyroid notch, and hyoid bone. Such identification can be difficult in obese patients. Inject 0.5 cc of 1% lidocaine with 1:100,000 epinephrine superficially in a small wheal over the midline cricothyroid ligament (for TA recording) and 1 cm inferiorly over the lower border of the cricoid (for CT recording). While these steps are being taken, another member of the team can be placing the reference electrode on the forehead.

If a tracheostomy is present, it is necessary to remove it for the laryngeal EMG testing, therefore these procedures can only be carried on those patients who can tolerate short term removal of the tracheostomy tube. A nasal speculum can be placed into the tracheostomy site to maintain the airway during testing.

Placing the Electrode in the Desired Muscle

Thyroarytenoid muscle: Pierce the skin in the midline with the electrode directed superolaterally through the cricothyroid ligament, and insert the needle to a depth of 1.5 to 3 cm depending on thickness of the overlying tissues, angle of entry, and size of the larynx. The electrode should enter the TA by a submucosal route without entering the airway. If the route is too superficial, the electrode may enter the sternohyoid or cricothyroid muscle. If the route is too deep or too medial, the electrode may enter the airway. In this event, the patient may feel a severe need to cough or swallow. One way to monitor this process is to record from the electrode as the needle is inserted. Movements of the electrode will cause some movement artifact, but one should still be able to hear muscle potentials when the tip is in a muscle. If the electrodes enter the airway, there will a "blast" of 60-cycle noise from the audio monitor because the electrodes will be "open" when in air rather than tissue.

Cricothyroid muscle: Pierce the skin in the midline and direct the needle posterolaterally along the long axis of the pars obliqua, aiming at the lower surface of the thyroid cartilage posterior to the tuberculum. The needle should not penetrate the cricothyroid ligament. If the route is too superficial, the electrode will likely end up in the sternohyoid muscle, whereas if it is too deep it will probably be in the lateral cricoarytenoid muscle.

Verifying the electrode location: Assuming that the electrode has been placed without incident, one faces the Big Mystery. In what muscle, if any, is the electrode located? When the larynx being observed is known a priori to be functionally normal, one can, by assuming that all normal people use the muscles of the larynx in the same way, perform certain verification gestures to deduce which muscle the electrode is in. One thing that happens fairly often with a hooked-wire electrode is that it "follows" the needle out, but then gets hung up just under the skin. Thus a good verification gesture is to have the patient push downward with their chin against your hand. This

activates strap muscles, and if you hear or see "crisp" EMG, then the electrode is definitely not usable. TA will exhibit a brief, vigorous, burst of activity during a swallow. CT generally will not activate greatly for a swallow, but will activate strongly for phonation at the subject's highest pitch, but will exhibit very little activity during phonation at a low pitch. TA activity can also go up during phonation at high pitch, so this "criterion" is certainly not absolute. However, the anatomy is such that one is unlikely to hit TA when aiming for CT, and vice versa. To my mind, much of the logic behind complex verification procedures is circular, and once you start to work with patients in whom the activation of these muscles may be abnormal, there is little basis for being very dogmatic about the result of these "verifications".

The posterior cricoarytenoid and interarytenoid muscles are very unlikely unless one intends to record from them (by using a transoral approach). Distinguishing between TA and LCA on the basis of their behavior is apt to be a real problem. Some authorities feel confident about telling them apart by their behavior, but most records I've seen in print in which both TA and LCA have been studied show two muscles whose activity differs only in small details.

Respiratory activity is usually, but not always, seen in laryngeal muscles. The usual pattern is for the adductors to be more active during exhalation. The amount of activity is usually very small during quiet breathing, compared to the level of activity seen with phonation and other laryngeal maneuvers, however, and I have seen good records from normal people in which there was no evidence of respiratory activity in TA or CT. On the other hand, vigorous activity of TA during inspiration, which was seen with the patient in Case 2 (above), is highly unusual. Generally speaking, the presence or absence of respiratory activity would not be definitive as a sign of which muscle the electrode is in (unless the needle went all the way through the lumen of the larynx and ended up in the PCA...oops!)

An alternative verification procedure that is useful for TA is to have an endoscopic or nasoendoscopic view of the glottis during the needle insertion. The movement of the needle into the vocal fold can usually be seen. Additionally, if the person making the insertion gently moves the needle shaft, this will be seen to produce the expected movement of the vocal fold. Such a procedure can be, however, counterproductive it if makes the patient cough or gag while the needle is being inserted. Another problem I have noted in this procedure is that the endoscope seems to act a conduit for carrying a large 60 cycle electrical field near the electrodes. There may be a way of thwarting this effect from an endoscope, but I don't know what it is as of yet.

Post-operative procedures: After observations have been made, the electrodes may be removed by gently tugging them until the hooks bend. If the wires are grasped at their emergence from the skin, then you can determine how deep the electrodes were in the tissue. With some experience, this information may become useful in understanding the results that are obtained.

The patient should be observed for a minimum of 30 minutes following the procedure. If there is any question of laryngeal injury, indirect or fiberoptic endoscopy should be performed to confirm an adequate airway and document the degree of injury. If there are no complications, however, the patient may be released after the waiting period, and told that they may eat whenever they want to.

Recognizing and Getting Rid of Artifacts (Noise)

The 60-cycle Beast: As most people know, the power supplied to homes, hospitals and labs by the Electric Company comes out of the wall (in the U.S.) as a 60 Hz sinewave at a root-mean-square value of about 115 volts. Most people think that this is so because electrical engineers found that electrical power could be distributed MUCH more efficiently from generating plants to users spread all over the countryside if the voltage in the distribution lines was VERY high, i.e. >100,000 volts. But this required stepping-up the voltage coming from the generators, and stepping-down the voltage when it was delivered to the houses etc. Can you imagine 100,000 volts in your toaster? Twelve microseconds until the toast (and you) are a nice "toast" color. Stepping voltages up and down can be done easily and efficiently if you use transformers, but transformers demand that the voltages used produce alternating current. DC was doomed from the start. Sixty Hz was chosen because that is lowest frequency which is above the critical fusion frequency of a human under conditions of bright light (so we don't see our lights flickering even though a photocell will show you that they are flickering at 60 Hz). The above explanation is a ruse, however. The REAL reason is that Thomas Edison had a pathological hatred of neurophysiologists, and he was advised that alternating current in the power system, rather than DC, would create problems for neurophysiologists for generations to come. And it has.

In dealing with 60-cycle interference, there are three important steps to take:

1. Minimize the electrode impedance, including the impedance of the reference electrode.
2. Remove the sources of the interference.
3. Shield against the interference, and/or selectively filter it out of the amplification process.

Reducing Source Impedance

Other factors being the same, the amount of 60-cycle problem one encounters can be related to the "source" resistance of the electrodes, which in this case is the sum of the resistances from each electrode to the reference electrode. Consider the extreme: suppose you directly connect both input electrodes directly to the reference electrode with a piece of wire. You could put that "electrode" configuration in the middle of a generator and you wouldn't have a problem. At the opposite extreme, suppose the electrodes are "open," i.e. there is nothing but air between the electrodes and the reference electrode. The source impedance is now the input impedance of the amplifier, which will probably be something like 10^{12} ohms. In this condition, very small 60-cycle fields will be able to create comparatively large voltages.

There is really little that one can do about the resistance of the electrode contact with the body fluids. One has a small area of bared stainless steel wire on the end of each wire, and that it that. The thing that one CAN do something about is the resistance between the reference electrode and skin of the patient. Applying a skin electrode is probably familiar to most readers. Let me emphasize, however, what you were probably told at the time you got your first lesson; short of being cruel, scrub the skin vigorously. You can make the skin "pink up" without leaving a scab, and it will, in the end, be to the patient's benefit if you are able to get good noise-free recordings. Commercial preparations are available to aid this process. At the extreme, one can use acne preparations, which are roughly equivalent to liquid sandpaper. I would hesitate to use them on anyone except a well-paid volunteer. A preparation which I do use, however, that is a "mild abrasive" version of an acne scrub, is called Omni-Gel. It was recommended to me by an audiologist

who specializes in ABR recording, and its use significantly reduced the 60-cycle level of the laryngeal EMG recordings in my lab.

Appreciation of the effect of having a high source impedance will help you understand a frequent problem in EMG recording: the mysterious appearance of high level of 60-cycle noise where it was absent earlier in the recording, or where exactly the same conditions were used in a previous recording session. The best guess is that either a connection is "open" somewhere in the circuit, you have an air bubble between the gel in the skin electrode and the metallic part of the electrode, or perhaps the electrode tip has been pushed into the lumen of the larynx.

Eliminating the Source of the 60-cycle Field

Interference from 60-cycle sources comes from two types of "fields." The easy problem (usually) to solve is caused by inductive fields from transformers or wires carrying large AC currents which are physically close to the electrodes. You cannot "shield" against this type of field, but physically moving the source of the magnetic field away from the recording site is usually sufficient to solve the problem. Moving just a few feet sometimes helps a great deal. I used to have a significant problem with 60-cycle pick-up when I was testing amplifiers on my electronic workbench in my lab. I was always somewhat disappointed in their performance in this regard. Then one day I went searching for the source of the field. I unplugged every piece of equipment in the lab, and turned out all the lights. Yet the 60-cycle problem was still there. Then I took the leads of the amplifier and moved them like an antenna, seeing where I could move them that would make the 60 cycle signal become bigger or smaller. I soon noted that moving the amplifier and leads toward the ceiling made the signal become very small, whereas moving it toward the floor greatly increased the signal. I then moved the bench laterally about 4 feet, and the problem simply went away. I had accidently placed the workbench directly over a large electrical "main" which was located in a conduit in the concrete of the floor. Morale: there is always a "source" of 60-cycle noise. If you can find it and turn it off or move it further away, you can get rid of the problem.

The other type of 60-cycle field is electrostatic. Any two conductors separated by an insulator form a capacitor. If a capacitor has a significant capacitance (a number that indicates how much charge can be stored at a given voltage difference between the conductors), then an AC voltage impressed upon one of the conductors can "couple" some fraction of that AC voltage to the other conductor. Large currents do not have to flow to cause this type of interference. Equipment which is plugged into the wall socket, but which is not turned on, can still cause problems. The powercord to the equipment, and the wires running internally to the power switch can form a conductor which has a peak voltage of +/- 160 volts with respect to ground. Any other conductor in the area, in particular the leads to the amplifier, can have a large AC voltage coupled to it if the resistance to ground is large (see above re source impedance). Therefore, one strategy of finding the source of a 60-cycle problem is to systematically unplug any unnecessary equipment in the vicinity of the recording site. If the equipment has to be plugged in, then sometimes moving the offending equipment to another location is helpful.

Florescent light can be a terrible problem. They all have a "ballast" in them, which increases the voltage across the florescent tube to about 750 volts. Turning off the lights close to the recording site usually gets rid of the problem, but if one has to use high amplifier gains, then one may have to turn off all the florescent lights. Incandescent lights, which use only 115 volts, cause far less of a problem, so overhead surgical lights are often tolerable. I would add that sometimes florescent lights

are not the source of the problem. You can test this simply by turning off the lights to see if the problem goes away.

One factor involved in getting rid of 60-cycle noise which I know about empirically, but don't really understand, is the effect of having large ungrounded metal structures of any type close to the patient or amplifier. For example, we have found that to get rid of the 60-cycle noise problem in the minor surgery room we have been using recently for our laryngeal recording we had to ground the metal frame holding the patient's bed. It is important to note that the patient was not in contact with this frame, so we were NOT grounding the subject. I've seen similar effects from grounding the metal frame of a table being used as a workbench. I would be dishonest, however, if I did not also add that I have on occasions noted that ungrounding a large metal object near the recording site decreased the 60 cycle problem! This comment brings me to my summary of what to do to get rid of 60-cycle noise: behave logically to identify the source of the field and them remove it. When you have done all you can with that strategy, try anything that would seem attractive to an imaginative monkey.

Shielding and Filtering

After the source of the 60-cycle noise has been minimized, then one has two additional strategies: shielding and filtering. If one is working in a normal environment, having an electrically-shielded room to do the recording is generally not practical. I should note parenthetically that I have been amused to see several special electrically-shielded rooms outfitted with florescent lights inside the room! Some people shield the leads going from the subject to the amplifier, but this makes the leads rather stiff and much heavier than is desirable. A much more common strategy that accomplishes the same thing is to use a small input stage on the amplifier (a small box) that is held close to the patient. Thus the leads can be kept fairly short. If these leads are only a few feet long, then shielding them in a cable with a braided "shield" really doesn't help much. Shielded input leads can also be a major source of movement artifact (discussed later), so I recommend short unshielded leads to the amplifier.

I have steadfastly opposed 60-cycle "notch filters" for most of my professional career. My feeling was that they discourage the attempt to find the source of the problem and get rid of it. Now that I have been confronted with the challenge of recording in an environment that I can't control to my liking, I have to admit that I have designed the instrumentation system which is being built for our continued laryngeal recording experiments with 60-cycle notch filters. They will help by reducing the 60-cycle component of the noise ten-fold, compared to what it would be without the filter. They will also produce corresponding amplitude and phase distortions of the biological signals in this frequency range, but that should not be a problem for diagnostic EMG recording. Notch filters will not do as much as you might suppose at getting rid of "60-cycle" noise, however. The simple reason is that the actual waveform which is usually recorded is not a simple 60-Hz sinewave. It can have very strange peaks and lumps, and they are all represented in the signal as harmonics of 60 Hz, and these harmonics just waltz right through the filter. In the end, you do your best, and dream of setting up a research station for high gain neurophysiological recording in the Gobi desert; a station powered entirely by batteries.

Besides 60-cycle noise, one can encounter other forms of extraneous noise signals. Three that are relevant to the applications being considered here are 1) high frequency fields from TV or computer monitors, 2) electrocautery, and 3) electrical devices such as motors that have "brushes." Interference from a TV monitor will be seen on an oscilloscope as a broad band of "fuzz" around the

recorded signal. You can't hear it on the audio monitor, because most audio amplifiers can't reproduce it and/or most of the "seniors" can't hear that high (about 16 kHz). Generally it isn't a serious problem, but if it is too large, it can "block" the input stages of the amplifier and thus affect amplification of the EMG signal. The easiest solution is to move the TV monitor as far away from the patient as possible. A few feet can help a lot. Electrocautery can produce sparks, and a spark creates a burst of electrical "white noise," and those components in the bandpass of the amplifier can produce an enormous "spike" in the recording. Where this observation is applicable, I suspect, is in any attempt to make an automatic EMG monitor during a surgical procedure. This automatic monitor has got to be "smart" enough to disregard electrical noises created by any equipment used in the surgery.

Intrinsic artifacts (noises): The main type of artifact one has to recognize is related to movement of the electrodes in the muscle, or less frequently, movements of the electrode leads between the patient and the amplifier inputs. In most cases, these electrode movements cause large low-frequency excursions of the recording baseline. If the amplifier is designed correctly, however, it will not amplify signals below the low-frequency cut-off, and these signals won't be too large. In some amplifiers, however, the input stage may be DC-coupled, and have enough gain that these large movement artifacts actually cause the input stages to "block." Because the low-pass filtering may take place after the input stage, one may not see the large excursions of the baseline. All one will see is that the output periodically goes dead-silent for awhile, and then seems to recover spontaneously after a few seconds. Let me add a historical note for the electronic buffs.

> The circuit that could reduce the gain of the input stage to low frequencies without putting a capacitor-resistor circuit in the input to the amplifier (which basically destroys the CMRR and input impedance of the amplifier) escaped me and most other people for many years. The solution depends upon the judicious placement of <u>one</u> capacitor in an otherwise conventional circuit. I swiped this design from a circuit designed by Rick Konipaki (Biocommunication Electronics, Inc.). I don't know if Rick invented it, but he has my thanks in any case. This design allows one to put a fair gain in the input stage before the isolator, keep the inherent CMRR of the instrumentation amplifier, and never have to see the amplifier system "block." Wonderful!

The source of the large low-frequency signal produced by movement of metal electrodes in biological tissues is, I think, due to physical disturbance of the electrochemical system at the interface of a metallic conductor in a salt solution. In short, there is a surprisingly large DC junction potential at each electrode. In effect, each electrode acts like a miniature "battery." The polarity of each junction potential is such that most of the effect cancels. However, tiny physical differences can create a "standing" DC voltage at the inputs to the amplifier. To see this phenomenon for yourself, take a bipolar electrode and place it in physiological saline (or simpler, place it in the saliva in your mouth) and connect the leads to a modern digital voltmeter, which has a high input impedance. You will see, I predict, a DC voltage of 1-15 millivolts which will jump abruptly if you jiggle the wires. These jumps, which may be many millivolts in amplitude, are the source of the problem.

Sometimes one sees a lovely sinusoidal signal at the frequency of phonation when recording from the larynx. Don't suppose you are observing phase-locking of the EMG. This is another version of the movement artifact discussed above. Another version is even trickier, however.

Sometimes one sees brief "spikes" associated with the start of phonation or with laryngeal maneuvers such as swallowing. They look somewhat like large single motor units, except for two things. They are usually monophasic (go in only one direction from the baseline), and while they may occur several times in each burst, they don't exhibit a repetitive pattern at a respectable rate, i.e. something on the order of 10-50 spikes/second. My guess is that these spikes are caused by the bared portions of the wires actually touching one another internally in the muscle. This doesn't happen very often, and generally goes away after the wires have wiggled into a more comfortable position.

Even with bipolar hooked-wire electrodes having large tip exposures it is fairly common to observe single motor unit activity in laryngeal muscle. One has to separate these potentials from artifacts that can produce spike-like potentials in the record. Real single motor unit potentials are distinguished by several features: 1) they are almost always biphasic, 2) they usually fire at least several times in a row at rather regular intervals, 3) they generally correlate with some aspect of motor activity, 4) they do not correspond to external events like someone switching a piece of electronic equipment on and off, 5) they disappear just about the time you decide they ARE single motor units and you start to make a record of them.

Interpreting EMG Waveforms

Two obviously abnormal waveforms: At this stage of experience, I can't be very authoritative about what is or is not an abnormal EMG. But at the extremes, one doesn't have to be much of a genius to get the idea. The first case is one in which no EMG signals are present during maneuvers that normally produce muscle activation. This is a condition we observed in Case 1, mentioned in the beginning of this document. In this case, one must conclude (assuming a cooperative patient) that the muscle is paralyzed or that the electrode is not in muscle tissue at all. Certain details of this recording session suggest to me that the electrode was in fact in muscle, hence the conclusion is that the patient's laryngeal muscles were still paralyzed. The main observation was that when the electrode wires were gently tugged there was, shortly after the movement artifact, the unmistakable sounds of a biological cell dying, with its attendant "injury potential." When an excitable cell, such as a neuron or muscle cell, is mechanically injured, it is locally depolarized and this usually causes a period of action potentials, which start initially at a high rate and then taper off to silence. In the case of a neuron, the sound through an audio monitor is exactly like that which we associate with a person going off the roof of a ten-story building. It generally curls the toes of the initiate the first time it happens. What I heard from the EMG monitor was similar, except the rates were much lower. The potentials also sounded "tinky" compared to a normal single motor unit. Looking at the oscilloscope, I caught a glance of a few of these potentials on the screen, and they appeared monophasic. Fortunately, we had a recorder going continuously, so I was able to go back later and study these records.

> The record shown in Figure 1 (next page) illustrates one of these responses to moving the electrode wires in this silent muscle.

The record also illustrates the minimum 60-cycle interference we could achieve until we grounded the metal frame supporting the patient's bed. Such grounding completely eliminated this problem.

The record starts off with a movement artifact. After the artifact abates, one can see a series of very brief monophasic spikes that occur at gradually increasing intervals. The arrow points out

one of these spikes that is shown at a faster time base on the right of the slow record. These monophasic spikes were very brief. A single motor unit from one of my normal laryngeal recordings is shown for comparison below. The monophasic nature and brevity of the patient's laryngeal muscle spikes suggest that it is different in important ways from a normal single motor unit record. I am pretty confident these were action potentials from a single muscle cell which had been injured by moving the end of the electrode. The monophasic waveform is caused by the fact that the action potentials were excited close the electrode, and travelled away from it, disappearing into the "distance." A biphasic waveform is produced by action potentials approaching, passing, and then going away from a pair of electrodes (which is what happens usually in a normal muscle). The brevity is accounted for by the fact that only a single muscle cell was producing the potential. In this case, one sees an action potential which is not "smeared" in time. The normal single motor unit response is really a summation of many action potentials in many muscle cells which are firing almost at the same time. Due, however, to slightly different conduction distances to the various muscle cells in a single motor unit, the summated potential is considerably longer that the potential in a single muscle cell, i.e. it is "smeared" in time.

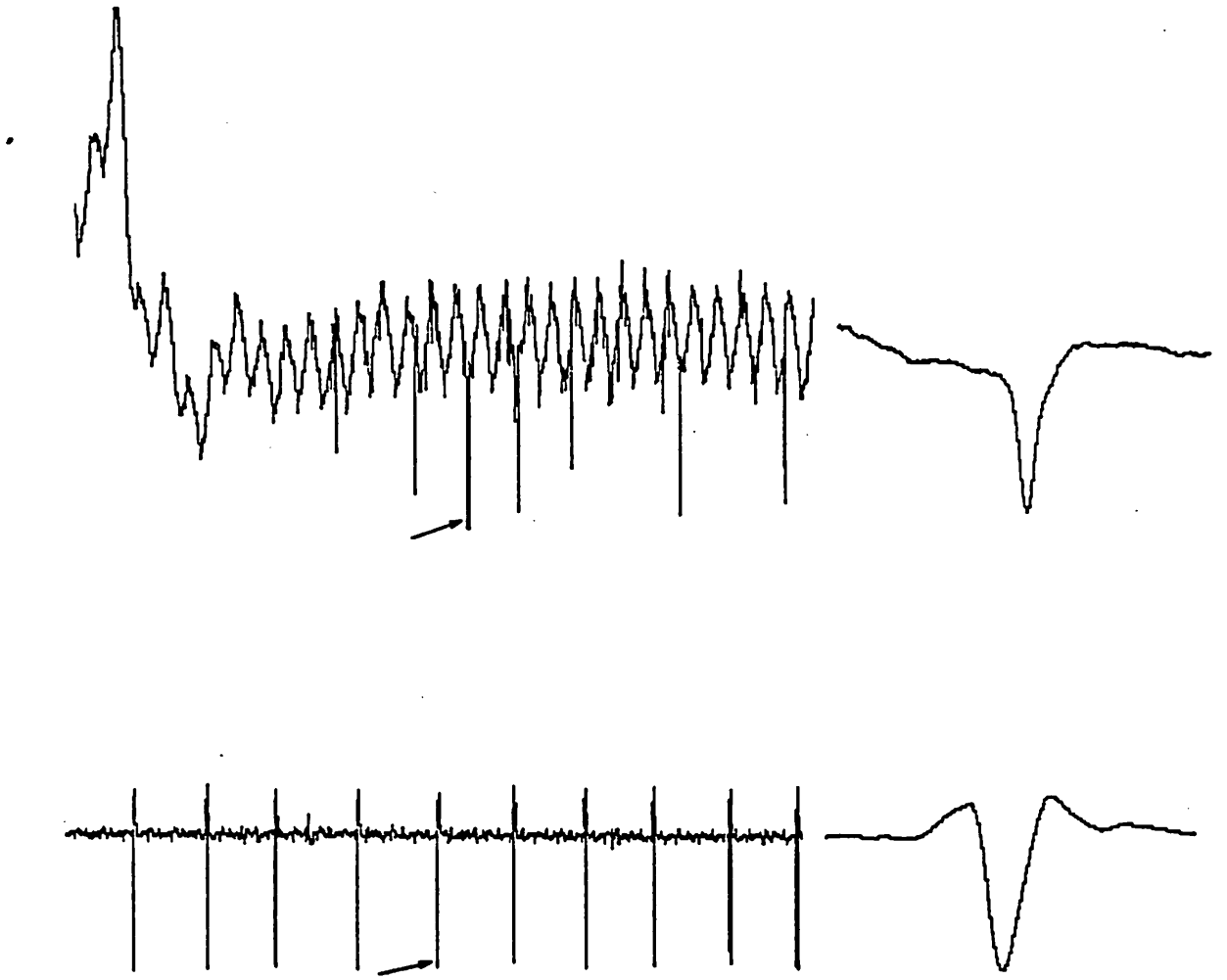

Figure 1. Top: EMG records of the thyroarytenoid muscle "unit" observed in the patient with Guillain-Barre syndrome (Case 1). Bottom: Single motor unit discharges observed in the TA muscle of a normal subject. Low speed records (left) are 500 ms long. High speed records (right) of the unit waveform indicated by the arrow are 4 msec long. Amplitude of the unit in the top records is about 125 μV; that in the bottom record is about 450 μV.

A second type of obvious abnormality is that seen in a muscle which has been reinnervated by a relatively small number of motor axons. This is a situation I have read about, but had not had the opportunity to see until we placed the tip of a Botox injection needle in the immobile vocal fold of a spasmodic dysphonia patient who had the RLN on that side surgically denervated several years previously. What we saw was a "forest" of enormous, fat, single motor units on the oscilloscope. Such waveforms were initially reported in the muscles of polio patients who survived the disease and had a few surviving motoneurons take control of whole muscles. Surviving motor axon terminals, and also motor axon terminals reinnervating a muscle, have the capacity to "sprout" and produce functional motor endplates on other muscle cells that have don't have contact with another motor axon. Ultimately, a whole muscle may be comprised of a few "monster" motor units. The large number of muscle cells in each unit makes their waveforms larger than normal, and the greater temporal dispersion of action potentials increases the duration of the single motor unit potential.

I would point out that while these observations are my own interpretations of these particular records, they don't represent novel observations or interpretations. These types of potentials have been studied by electromyographers for many years. I have described them here to illustrate that attention to some of the details of these recordings can be helpful in establishing the diagnosis. These observations also illustrate the fact that interpretation of the EMG may be significantly aided by an understanding of the basic anatomy and physiology of muscle.

Quantitative measures of EMG activity: In many cases, it will not be obvious whether the waveforms or the degrees of muscle activity observed in the EMG record are normal or abnormal. It would be marvelous if there were well-established measures of EMG activity which had been "normed" to records from a large number of individual having normal voices and laryngeal function. I hope that some day such standards exist, but one should keep in mind that currently we are on very "soft" ground if we attempt to look at an EMG record and assert that it is either normal or abnormal. One of the things I have learned, as have other scientists studying the EMG activity associated with many types of movements in normal humans (and animals), that the "normal" pattern of muscle activation during a movement is a statistical statement. There is actually a very large range of patterns within a group of functionally normal individuals. The situation is even more complicated in the case of the larynx, where one has only a general idea of which muscle the electrode is in. Then one has to consider the possibility that different parts of the same (anatomically-defined) muscle may be used differently by the nervous system. I do not mean to suggest that the situation is hopeless. What I DO hope to do is discourage such unadorned and undocumented statements as "the laryngeal muscle activity of all the spasmodic dysphonia patients was abnormal." The quoted statement is only a slightly modified version of a sentence I have seen published in a professional journal.

One of things that would facilitate the development of standards for interpreting laryngeal EMG records would be the use of the same measures and at least some common behavioral tasks by the various laboratories and clinics involved in the enterprise. Let us suppose, as an example, that we have EMG records from TA and CT during a swallow, quiet breathing, sustained /i/ at a comfortable level, and an ascending pitch phonation. One might also include speech samples of standard phrases or sentences. How are these records going to be converted to numerical values? Analyzing the EMG during sustained tasks is relatively simple if one simply wants to know the "average" level of the EMG. I say "simple" because I am assuming a computer with an A/D converter is available. In this case, one digitizes an appropriate portion of the record, takes the absolute value of the numbers (which is what "rectification" does), sums all the values in the array,

and then divides the sum by the number of samples in the array. The result is the average value of the EMG during the period. This number then has to be converted to a "real" number, however. One has to consider the gain of the amplifier used during recording, the gain of the recording process (usually 1.0), and the scale of the A/D converter. These numbers will allow you to convert the average EMG to the units of volts (which will usually be in the range to 10-500 V). An alternative procedure is to create the root-mean-square (RMS) value of the EMG. The main difference between the "mean" and RMS measure is that the latter gives more "weight" to the large amplitude peaks in the record (De Luca, '79). I personally think the "Mean" is just fine, and most people understand it better. If one is attempting to quantify EMG during quiet breathing or ascending pitches, then it doesn't make sense to take the mean of the whole record. But it would be useful to calculate the mean value over portions of the record (during peak inspiration and exhalation, and at the lowest and highest pitch)

Many people want to make comparisons between values from different muscles and/or different people. If one compares these mean values expressed in voltage, then the usual result is variability which totally destroys the "power" of any statistical tests that are used. To solve this problem, people "norm" the voltages in various ways. One can either norm all values to the highest value or the lowest value seen. There is nothing wrong with norming data in either way, but if the results are presented in only this form, then they cannot be directly compared. If however, the values are also reported in voltages, then they can be directly compared, and at least potentially be "pooled" to help describe laryngeal muscle activity in various classes of individuals.

I fully realize that the average EMG levels, expressed in voltage, can be quite variable. The signal depends on electrode configuration and location within a muscle. Filter characteristics can also be a factor. For example, the presence of low frequency signals can have a large effect on the mean values. Nevertheless, these values will have to be reported in a recognized unit (voltage) if a standard is ever to evolve. The variability will always probably be high enough that it will be unlike that we can say we know that the "TA activity in this spasmodic dysphonia patient is abnormally high." It seems possible, however, that we might be able to conclude that "EMG levels are in the highest 20% of values seen in normal speakers." EMG values can add information to a diagnostic process. By themselves, it is unlikely that they can ever be definitive.

It would be useful to be able to quantify dynamic changes in EMG records. Consider swallows, for example. These events can be described in terms of duration from start to finish, peak values achieved, and the average value during the event. Quantifying EMG responses during repetition of simple utterances is possible (see Hirose, 1987), but the results are usually non-numerical. Nevertheless, knowledge of the general pattern of activation of a laryngeal muscle during the production of a standard phrase could be useful. Such a record can be created by using a moving average to smooth a record of the absolute value of the EMG. Figure 2 illustrates such a record from the TA muscle of two normal speakers and two severe stutterers reciting "We were away a year ago." Can you pick out the stutterers?

Another quantitative measure that can be obtained on at least some records pertains to the behavior of single motor units. I am currently analyzing records we have from previous experiments that contain identifiable single motor units. One can characterize motor unit behavior in terms of firing rate and firing rate variability. The motor unit waveforms can be quantified in terms of amplitude, duration, and polarity. One can only speculate at this time whether this will help diagnostically, but it would appear to have that potential.

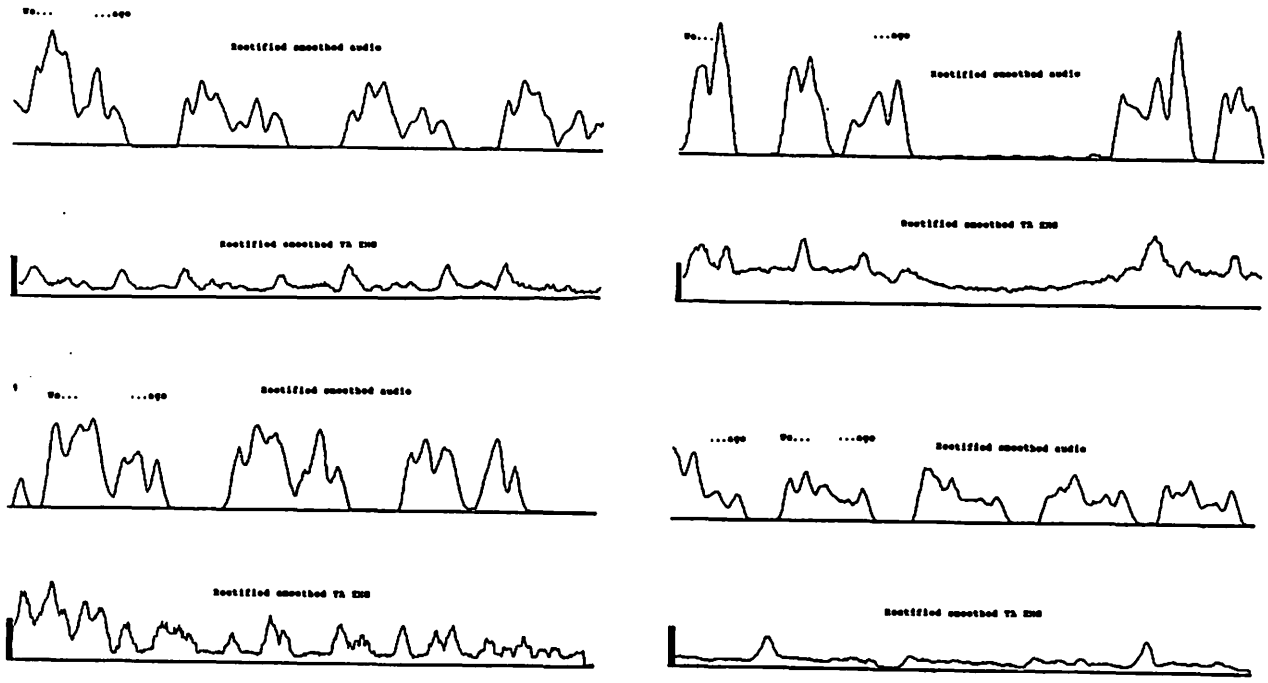

Figure 2. Records of rectified smoothed audio and TA EMG during repeated productions of the sentence "We were away a year ago" by four different speakers, two of whom are severe stutterers. The other two subjects are normal speakers. The smoothing process was exactly the same for all records: one pass with a 10 point moving average. The baseline represents 0 volts; the calibration bar for all TA records represents 100 μV. The amplitude of the audio records is relative. The records are all 10 seconds long.

References

Blitzer, A., Brin, M.F., Fahn, S., and Lovelace, R.E. Localized injections of botulinum toxin for the treatment of focal laryngeal dystonia (Spastic dysphonia). Laryngoscope, 98, 193-197, 1988.

De Luca, C.J. Physiology and mathematics of myoelectric signals. IEEE Transactions on Biomedical Engineering, Vol. BME-26, 313-325, 1979.

Hirano, M, and Ohala, J. The use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. Journal of Speech and Hearing Research, 12, 362-372, 1969.

Hirano, M. The laryngeal muscles in singing. In Hirano, M., Kirchner, J.A., and D.M. Bless (eds) Neurolaryngology, College-Hill Press, Boston, 209-230, 1987.
Hirose, H. Laryngeal articulatory adjustments in terms of EMG. In Hirano, M., Kirchner, J.A., and D.M. Bless (eds) Neurolaryngology, College-Hill Press, Boston, 200-208, 1987.

Ludlow, C.L., Baker, M., Naunton, R.F., and Hallet, M. Intrinsic laryngeal muscle activation in spasmodic dysphonia. In Benecke, R., Conrad, B., and Marsden, C.D. Motor Disturbances I, Academic Press, New York, 119-130, 1987.

Ludlow, C.L., and Fujita, M. The coordination of intrinsic laryngeal muscle activation during phonatory and non-phonatory tasks. In J.D. Newman (ed) The Physiological Control of Mammalian Vocalization, Plenum Publishing Corp., New York, N.Y., 109-118, 1988.

Ludlow, C.L., Schultz, G., and Naunton, R.F. The effects of diazepam on intrinsic laryngeal muscle activation during respiration and speech. Journal of Voice, 2, 70-77, 1988.

Ludlow, C.L., Sedory, S.E., and Fujita, M. Correlations between intrinsic laryngeal msucles during different speech gestures. In Kiritani, S., Hirose, H., and Fujisaki, H. Festschrift for Osamu Fujimura, Ohmsha Publishers, Tokyo, Japan, in press.

Miller, R.H., Woodson, Gayle, E., and Jankovic, J. Botulinum toxin injection of the vocal fold for spasmodic dysphonia. Acta Otolaryngology and Head and Neck Surgery, 113, 603-605, 1987.

Shipp, T., Izdebski, K., Reed, C., and Morrissey, P. Intrinsic laryngeal muscle activity in a spastic dysphonia patient. Journal of Speech and Hearing Research, 50, 54-59, 1985.

Sources for Wire for Hooked-Wire Electrodes

1 Medwire, 121 S. Columbus Ave., Mount Vernon. N.Y. (914) 664 5300. This company has supplied most of the wire that I have personally used. The wire I usually get from them is "316 SS 3T" (0.003" stainless steel wire, plastic insulated with an overall diameter of 0.0045").

2. Cooner Sales Company, 9186 Independence Ave., Chatsworth, California (213) 882-8311. This company has a wide variety of wires that could be used in electrode manufacture.

3. A-M Systems, Inc., 11627-A Airport Road, Everett, Washington (800) 426-1306. This is another source of the 0.003" stainless steel wire, except that they provide it insulated with teflon rather than plastic. I've never personally used the teflon-insulated wire. It might be a little harder to strip the insulaton that the Medwire (but I'm guessing).

4. California Fine Wire Co., 338 S. 4th St., P.O.Box 556, Grove City, California 93433. This company supplies a large selection of small "exotic" wires, and will manufacture wire to specifications. Dr. Anne Smith (Purdue University) has extensively used a wire from this company (Product #304, Stainless Steel/Heavy Poly nylon coating, 0.002" wire, 0.0027 maximum outside diameter) and found that it is stiff enough to hold a "hook," yet will accomodate a pair of wires in a fine needle without binding. It comes in units of 1000' at about $80/unit.

5. Surprising as it seems, there is ONE tool that "outstrips" all competitors when it comes to stripping the insulation from these fine wires without damaging the wire itself. There is only one source of this tool that I know of. The tool is called the "Fine Cutting Broad Blade, cat# D-TWES-7 in the catalog of: Small Parts Inc., 6891 N.E. Third Ave., PO Box 381966, Miami, Florida, 33238 (305) 751 0856.

Also note that this company sells tweezers for microscopic work (D-TWEX-3) for $17.00. That sounds like a lot but it is far cheaper than exactly the same "forceps" obtained from a surgical supply or biological supply. One way or the other, a pair of fine forceps is essential for making hooked-wire electrodes. When you get them, you might as well buy two because you will eventually drop one of them on their tips, and they are basically worthless after that.

NCVS Status and Progress Report - 1
June 1991, 151-166

Aerodynamic Assessment In Voice Production

Ronald C. Scherer, Ph.D.

Research and Recording Center, The Denver Center for the Performing Arts

Introduction

The purpose of this paper is to discuss selected aerodynamic techniques and measures relevant to clinical voice research and assessment.

In speech, aerodynamics usually refers to the average air pressures and airflows that are produced as part of the peripheral mechanics of the respiratory, laryngeal and supralaryngeal airways. Aerodynamics in general is a branch of mechanics that deals with reactions such as pressures, forces, resistances, and movements caused by relative motion between air and solid boundaries (Considine, 1976). Aerodynamics is related to acoustics through the fluid mechanics of air motion (Allen, 1982). In this paper, emphasis will be placed on the inextricable relationships among airflows, air pressures, vocal fold vibration, glottal dynamics, and the output acoustic signal.

There are a number of purposes for measuring laryngeal aerodynamics. One purpose is to describe laryngeal function quantitatively. Quantitative aerodynamic measures for voice tend to follow a dichotomy between gross and fine characteristics (as do acoustic measures, ref. e.g., Scherer et al., 1988). Measures of gross characteristics are taken over a few hundred milliseconds or more. Examples are average subglottal pressure, average airflow, calculated flow resistance (subglottal pressure divided by airflow), subglottal aerodynamic power (subglottal pressure times flow), acoustic efficiency (output power divided by subglottal aerodynamic power), sound pressure level, and other measures or derived variables dealing with average or integrated values. On the other hand, fine measures of aerodynamic characteristics deal with short term calculations on the order of tens of milliseconds or less, and often deal with aspects of pressures and flows within individual phonatory cycles. These measures deal with characteristics of the glottal flow such as peak values, slopes at specific times, relative delays, dynamic (time varying) subglottal pressure and vocal tract pressures, and intraglottal pressures. They also are associated with specific vibratory motions of the vocal folds as well as spectral details.

A second purpose for measuring laryngeal aerodynamics is to determine laryngeal abnormality. Gross aerodynamic measures are well established in clinical assessment, whereas fine measures only recently have been adopted.

A third purpose for measuring laryngeal aerodynamics is to help in establishing standard procedures in voice recording and analysis, and in determining valid interpretations of those analyses. The lack of voice recording and analysis standards for clinical purposes is of vital concern, and aerodynamic measurement must take a prominent role along with acoustic, glottal kinematic, and other measures.

Fourthly, an important purpose for measuring laryngeal aerodynamics is to obtain new and relevant data to enhance theories and models of phonation. Testing, revising and extending phonatory theories will bring functional bases into clearer focus for clinical relevance.

Clinically, aerodynamic measures have been used primarily to monitor change of the voice and to attempt to evaluate degrees of dysphonia. According to Hirano (1981), the most "popular" measures have been the mean airflow rate and the maximum phonation time, which appear to show significant change between pre- and post-treatment of vocal fold paralysis and polyps. Hirano also points out that glottal acoustic efficiency appears to be significantly reduced in a number of patient groups following treatment.

Average Aerodynamic Measures

The following discussion summarizes some of the common methods for obtaining average aerodynamic measures and variables related to phonation. Informative references for this material can be found in Baken (1987) and Hirano (1981, 1989).

The spirometer has been used for many years to measure flow volumes (Webster, 1978). It is an "expandable compartment consisting of a movable, statically counterbalanced, rigid chamber or bell, stationary base, and a dynamic seal between them" (Webster, 1978, p. 454). As Rau and Beckett (1984) pointed out, spirometers like the Collins P-900 can be used with prolonged phonation to obtain vital capacity, phonation volume and phonation time, allowing for the estimation of average flow and the phonation quotient (vital capacity divided by phonation time), but less expensive and portable spirometers (e.g., the Propper Compact Spirometer and the Ventilation Monitor) can also be as effective clinically. If a clinician can obtain only maximum phonation time, an estimate of vital capacity based on height and age should be considered (Bless & Hirano, 1982).

More commonly, however, the pneumotachograph, a gas volume flowmeter, is the choice for recording average flow rate during voice and speech production. The device consists of a metal housing for a resistance element (usually wire mesh or tubules) through which flow passes. An electrical pressure transducer is used to measure the pressure drop across the resistance. The pressure drop is linear with respect to the airflow in the operating range, and has sufficient accuracy, sensitivity and frequency response for clinical and research average flow applications. These devices are rugged and can be used for many years while maintaining calibration. The pressure transducers that are used with them need to be checked for calibration regularly.

Hot wire anemometry has a wide frequency response and therefore can be used to record both the mean flow and fast changing aspects of the flow. The necessary linearization of the

circuitry typically has been dealt with electronically when the items are purchased. Wire filaments within the devices are fragile, so these devices are not expected to be as rugged as the pneumotachographs. An advantage of these devices has been to measure an AC/DC ratio of flow (alternating amplitude of the flow divided by the average flow) that may have clinical relevance (Isshiki, 1981; Kitajima, 1985; Wilson & Starr, 1985; Woo, 1986). However, the AC aspect is dependent upon the resonance structure of the vocal tract, and the accuracy of hot wire flowmeters at low flows is poor (Isshiki, 1981, 1985).

Body plethysmography is a method whereby the patient or subject sits within a box in order to measure respiratory events. Boyle's law inversely relating pressure and gas volume is the primary principle behind body plethysmography operation. For speech purposes, Hixon (1972) gives an excellent review (see also Warren, 1976). The configuration enclosing the entire body can be used to measure lung volume, airway resistance, and speech alveolar pressure and airflow. A configuration with the head exposed above the box allows undistorted simultaneous acoustic measurements. Problems of pressure fluctuation due to temperature and humidity changes are controllable (e.g., Schlegelmich, undated). The essential advantages of body plethysmography are lack of discomfort for the subject, no risks to the subject, no placement of devices within the subject, and no necessary assistance by a physician (Hixon, 1972). However, there appear to be inconveniences of calibration, subject placement, and perhaps physical space for the box.

Devices such as magnetometers (Hixon et al., 1973; Hixon & Putnam, 1983; cf. Reich & McHenry, 1990), mercury strain gauges (Baken, 1977; Cavallo & Baken, 1985), and inductive plethysmography (e.g. Respitrace and Respigraph; Bless et al., 1982) are non-invasive and do not impede the speech and voice mechanisms. These techniques place devices on or around the upper thoracic wall and the abdomen so that motion of the rib cage and abdomen can be measured fairly easily if only in the sense of timing for expansion or contraction of the two regions (an important issue for disordered patients). These devices can be calibrated to measure airflow rate. The mercury strain gauges appear to be inferior to the other two techniques with respect to limits of volume measurement, flow monitoring, and slippage artifacts (Bless et al., 1982). For all these techniques, calibration must be performed with each subject, postural artifact can be significant, and estimation of instantaneous airflow for speaking can be inaccurate, problems not present using a pneumotachograph connected to the upper airway.

It is important to point out that different groups have adopted different techniques to measure airflow rate, vital capacity, and air volume with apparently equal satisfaction. Familiarity with a particular technique appears to reduce the severity of controllable disadvantages.

Airflow rate in normal subjects can cover a large range. A critical range would appear to be approximately 40 to 200 cc/s (Hirano, 1981; cf. Kitajima, 1985, and Wilson & Starr, 1985, who suggest wider ranges), and when used alone, is not often a clinically sensitive measure. The mean flow rate appears to be relatively high in many cases of recurrent laryngeal nerve paralysis and large tumors (Hirano, 1981), and the mean flow rate can be useful in following treatment of the larynx if significant changes are made in glottal competency. The primary worth of the mean flow rate, however, is probably in conjunction with subglottal pressure in measures of glottal resistance and subglottal aerodynamic power. Glottal resistance (subglottal pressure divided by airflow), rather than flow rate alone, is more

informative regarding glottal competence, and subglottal power (subglottal pressure times airflow) specifies the power that can be used and dissipated by the larynx and upper airway during speech.

The discussion now will turn to techniques related to average subglottal pressure.

The U-tube manometer technique has the patient or subject expire into a tube that is attached to one side of the manometer, allowing the column of fluid to displace vertically. The displacement between the menisci at the tops of the two columns corresponds to the measure of the airway pressure produced.

A simple clinical procedure to measure relatively low levels of respiratory pressure is the configuration by Hixon et al. (1982) whereby the patient blows into a straw that is positioned within a glass of water. The depth of the straw below the surface of the water indicates the approximate amount of pressure necessary for a bubble of air to exit the straw within the water.

The U-tube manometer and the straw in the glass techniques are simple and allow relatively accurate measures of clinically relevant respiratory pressures. These techniques, however, obviously do not measure subglottal pressure during phonation, but do indicate some level of potential for the creation of respiratory pressures.

Strain gauge devices, variable inductance devices, and other electrical pressure transducers have been used for many years for accurate measurement of physiological and speech pressures. They are used with pneumotachographs to measure airflow, with tracheal puncture procedures to measure subglottal pressure directly, and with oral pressure techniques to estimate subglottal pressure indirectly. Calibration can often stay within 1 to 3%. They can be durable, and they have frequency responses which cover the average range of expected pressure fluctuations (up to about 100 Hz; Edmonds et al., 1971). Pressure transducers such as these must be kept in constant calibration.

The measurement of subglottal pressure using a tracheal puncture procedure (e.g., Isshiki, 1964) is a direct way to measure subglottal pressure. The procedure is not without risk because of possible injury to the skin and subdermal air injection. The tube that is placed through the puncture hole within the trachea must be fairly rigid, relatively short, and placed perpendicular to the flow stream so that only the static normal pressure is measured.

The esophageal balloon technique has been used to approximate subglottal pressure during phonation (van den Berg, 1956; Lieberman, 1968; Cavagna & Margaria, 1968; Bouhuys et al., 1968; Leanderson et al., 1987, used a Gaeltec miniature pressure transducer within the esophagus). A balloon is fed into the upper part of the esophagus. There is no puncture to the skin. The measure of subglottal pressure is indirect and requires a correction for lung volume. There can also be insertion problems with respect to specific location and calibration.

Currently in use in some research settings are miniature pressure transducers, such as the Millar device (Mikro-Tip(R) Catheter Pressure Transducer, Millar Instruments, Inc.), in which sensitive pressure transducers are placed along a line that can be fed through the nose and pharynx to the larynx region (Koike & Perkins, 1968; Kitzing & Lofqvist, 1975; Koike, 1981; Koike et al., 1983; Cranen & Boves, 1985; Iwata, 1988; Miller & Schutte, 1990). This kind of device can be situated between the arytenoid cartilages and allows simultaneous sub- and supraglottal pressure recordings. The transducers are small and sensitive with a wide frequency response (0 to 20,000 Hz approximately) so that either average or broad-band pressure signals can be recorded. The larynx must be anesthetized for this procedure. The transducers can be difficult to cali-

brate, and have a potentially strong temperature sensitivity. Normal laryngeal function is influenced if the device gets close to the membranous portion of the vocal folds.

The technique of strong interest at the moment for both clinical and research purposes is the estimation of subglottal pressure from measurements of oral pressure (Rothenberg, 1973; Smitheran & Hixon, 1981; Lofqvist et al., 1982; Shipp, 1973; cf. the interruption method, Sawashima et al., 1986). In this procedure, a CVC string such as /pip/, /paep/ or /baep/ is repeatedly produced during a single exhalation by the subject or patient in a legatto (smooth) manner. When the lips close for the bilabial consonant, the glottis opens in order to produce the voiceless consonant. The pressure below the vocal folds during the vowel production is communicated to the oral cavity as the air flows from the trachea. The estimate of subglottal pressure is taken from a measure of the oral pressure during the consonant closure. This particular technique is indirect but relatively accurate (Lofqvist et al., 1982, found an average difference from direct measures of subglottal pressure of 0.85mm H_2O), is relatively simple, and if airflow is also measured (during the vowel portion) an estimate of glottal resistance can be obtained. The task can be used with patients who can completely occlude the oral airway so that airflow from the trachea quickly creates pressure equilibration. If the syllable rate is too slow, there is the risk of respiratory pumping during the consonant closures, giving oral pressures larger than the subglottal pressure during the vowel. If the syllable rate is too fast, there may not be enough time for the oral pressure to build up to the value of the subglottal pressure. Smitheran and Hixon proposed a rate of 1.5 syllables per second.

There are a number of systems that can be used to measure simultaneous average airflow and estimated average subglottal pressure. A fairly complete system appears to be the Phonatory Function Analyzer PS-77H (Nagashima Medical Instruments Co.; Sawashima & Aoki, 1982; Iwata, 1988; Wilson & Starr, 1985). It measures airflow, subglottal pressure, and sound pressure level, makes a number of calculations, and offers a number of displays. The complete system, including hardware and software, is expensive. The system described by Dr. Stephen Barlow given elsewhere in these proceedings (also Barlow et al., 1989) appears to be quite powerful in clinical and research assessment for aerodynamics in voice production. The Glottal Enterprises flow system (Rothenberg, 1973) can be modified to obtain average airflow for the same purposes. The PERCI-PC Version 2.0 (Microtronics Corp.) is configured to allow the measurement of oral flow and oral pressure for subglottal pressure estimation. The Kay Elemetrics Corporation has recently offered a device (Aerphone II, Model 6800) that allows the measurement of airflow and estimated subglottal pressure.

What is the relationship between these measures of average airflows and average air pressures and clinical needs? Changes of pressures and flows can be examined over time for a patient undergoing phonatory change or for post-treatment checks. For example, phonosurgery often changes the adductory nature of the larynx, and relatively large changes in airflows and air pressures and derived measures of glottal resistances and subglottal power may occur. Also, for example, regaining steadiness of the voice due to Botox treatment in patients with spasmodic dysphonia can be monitored by using the patient's average airflow during sustained vowels.

These average values of pressure and flow are related to basic concepts of phonation relevant to the clinic (Titze, 1986, 1988; Scherer, 1991). The concept of the subglottal pressure threshold (Titze, 1988) that allows the vocal folds to just begin oscillation is important from a clinical point of

view. Lower threshold pressures would correspond to less expiratory work to phonate, thus allowing phonation (and speaking) to feel easier.

The fundamental frequency of the voice depends primarily on the force per cross-sectional area of the vocal fold tissue, and the vocal fold length (Titze, 1988). In addition, the subglottal pressure plays a significant role in the change of fundamental frequency especially for lower frequencies. As the subglottal pressure increases, the lateral extension of the vocal folds in each of the vibratory cycles is increased, which increases the effective length and tension of the vocal folds (Titze, 1989a). This raises the pitch. This frequency control through subglottal pressure change may contribute strongly to intonation control during speaking.

Many studies have indicated that intensity increases as subglottal pressure increases (e.g., Isshiki, 1964; Ladefoged & McKinney, 1963; Iwata, 1988). Titze and Sundberg (1990) show that glottal geometry (glottal adduction, vocal fold medial convergence, vocal fold medial shape and the vertical phasing of the vocal folds) also may play a significant role in intensity control. The cyclic airflow just above the glottis is the signal affected by the subglottal pressure and intracycle glottal movements, and is the signal by which acoustic intensity is obtained (glottal flow is also dependent on the vocal tract acoustic characteristics, see below). It is shown later in this paper and also in the paper in these proceedings by Dr. Childers that the glottal flow waveform characteristics are relevant to intensity and spectral aspects.

If changes due to treatment of the voice are relatively subtle but meaningful, as is the case for many nonsurgical voice problems of quality, comfort, ease and endurance, average values of aerodynamic aspects of the voice may be insufficient to reflect those changes. There is a large variability for airflow across patients as well as within a particular patient (Bless & Hirano, 1982), and clinically relevant changes may be buried in this variability. Also, important measures like glottal flow resistance may not correspond well to clinical voice quality judgments (Holmberg & Leanderson, 1983). The detection of phonatory changes and accurate mapping to diagnostic categories and voice qualities may require finer analysis. Measures of subtle voice change, smaller measurement error, and a strong connection to phonatory theory are required.

Discussion now will center around the measurement of finer aspects of the aerodynamics of voice production.

Fine Measures of Laryngeal Function

The vocal folds constitute an oscillatory system that is essentially flow-induced. This means that, through the interaction of airflow and the vocal fold surfaces, there is an "asymmetric" intraglottal driving force during the vibratory cycle. Titze (1986, 1988) has shown that the pressure within the glottis is more positive when the vocal folds are moving laterally than when the vocal folds are moving medially. This difference in intraglottal pressure, more positive for the out-going vocal folds and less positive or negative for the in-going vocal folds, constitute the asymmetric driving forces necessary to maintain phonation (Titze, 1986, 1988). These external forces must overcome the damping characteristics of the vocal folds in order to maintain phonation. The asymmetric driving forces in the glottis are primarily due to the subglottal pressure, inertance of the air in the vocal tract, and the changing glottal shapes during vocal fold motion. If the dynamic glottal shape is abnormal due to abnormal vocal fold surface contours, the asymmetric air pressure forces may be insufficient for vocal fold oscillation (Scherer & Guo, 1990). The intraglottal forces are important, but there currently is no direct measurement of them due to vibration disruption by

intraglottal transducers. Progress in this area is ongoing, however (Scherer & Titze, 1982; Scherer et al., 1985; Reed et al., 1990).

A miniature pressure transducer like the Millar device can be used to measure the fine structure of pressure signals in the vocal tract. Figure 1 (after Fant, 1983; cf. Titze, 1986) shows both the subglottal and the supraglottal pressure recorded during human phonation, as well as the translaryngeal pressure (the difference between the subglottal pressure and the supraglottal pressure) and the "glottal area" function as depicted from photoglottography. The subglottal pressure measurement indicates a specific peak of pressure which occurs at the time of glottal closure. This is followed within the cycle by a fluctuating pressure that corresponds primarily to the influence of the subglottal resonances. The supraglottal pressure has a prominent minimum pressure, which is the negative pressure rarefaction in the supraglottal tract right above the vocal folds at the time of vocal fold closure. The damped oscillations of pressure during the rest of that cycle correspond primarily to the resonances of the vocal tract above the glottis. The fine structure thus deals with pressure magnitudes, pressure waveforms due to resonance effects, timing of pressure changes, and the relationship to dynamic configurational changes of the glottis. These details should change with lung pressure, glottal adduction, tissue morphology, and motion of the vocal folds. For example, the faster the glottal closing and the greater the lung pressure, the larger will be the negative peak rarefaction pressure above the folds and the "water hammer" positive peak pressure subglottally; if the folds close longitudinally, or have a swelling of the tissue, these peak pressures would be hypothesized to be lower and broader because of the relatively longer closing time. The wider the abduction, the greater is the potential for cross-leakage of resonance effects between the sub- and supraglottal regions causing a mixture oscillation pressure structure. Asymmetric vocal fold vibration should also affect the peak magnitudes and resonance structure, depending on vocal fold movement and abruptness of closure. These ideas suggest that the fine structure of the pressure signals should relate diagnostically to adduction, vocal fold morphology and vocal fold movement. The intraglottal pressures could be estimated

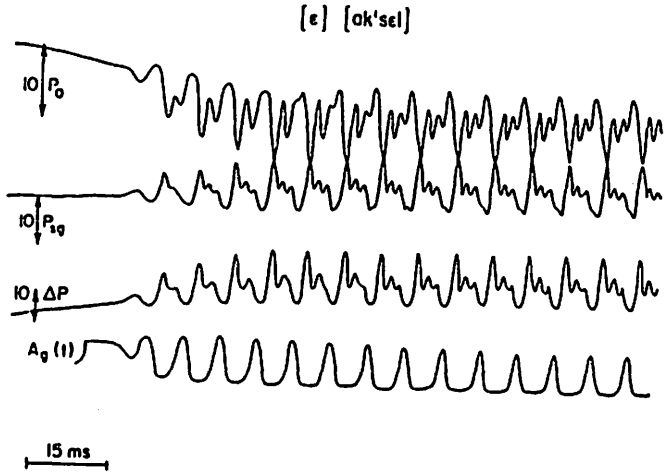

Figure 1. Laryngeal pressures and glottal area from transillumination for a stressed vowel. Top trace: supraglottal pressure. Second trace: subglottal pressure. Third trace: transglottal pressure. Bottom trace: glottal area. After Fant, 1983.

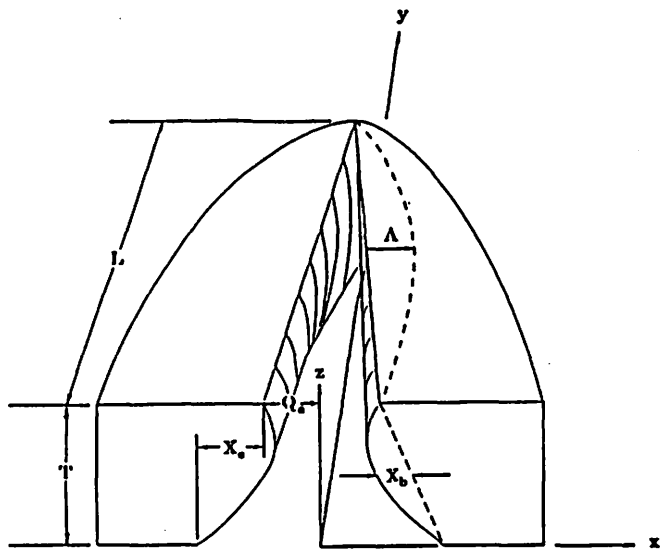

Figure 2. Schematic of glottal geometric parameters (Titze, 1989b). See text.

from these kinds of dynamic pressure changes if there were dynamic configurational measures of the vocal folds.

It is conceivable that early clinical problems in vocal fold morphology (such as early nodules, cysts, granulomas or cancer) would result in changes of pressure signals that create irregularities requiring fine display and measurement. Figure 2 indicates some of the specific laryngeal parameters that could be considered in explaining dynamic and configurational aspects of phonation. This diagram, borrowed from Titze (1989b), indicates the amplitude of motion of the vocal folds A, adduction of the vocal folds Q_a, a convergence factor X_c, and a shape or bulging factor X_b. In addition, parameters dealing with large and small structural changes on the surface of the vocal folds as well as longitudinal symmetries in vibration could be considered.

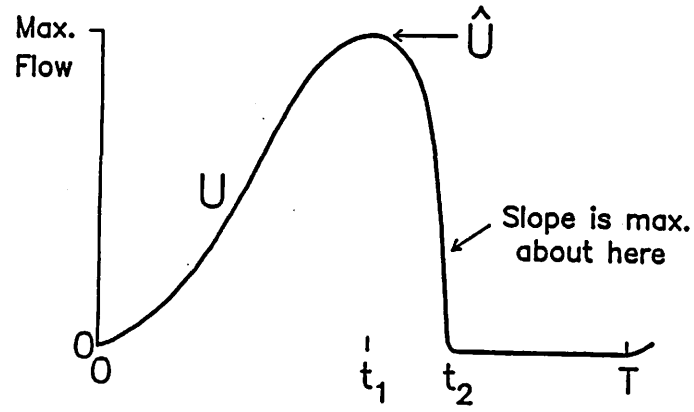

Figure 3. Schematic glottal volume velocity waveform, U. Maximum flow occurs at time T_1, and miniumum flow occurs at T_2. T is the period of the flow cycle.

Figure 3 shows a schematic of a single cycle of the volume velocity flow (integrated across the glottis, FN1) that exits the glottis during one cycle of phonation. When the vocal folds open, the flow begins to come out but is retarded in its exit because of the inertance of the air in the vocal tract (Rothenberg, 1981a). The flow reaches a peak value just after the glottis reaches its maximum area, and further delay in the airflow continues until the flow is shut off when the vocal folds return to the midline. From the time noted by T_2 to the end of the period T there is a relatively flat baseline indicating no flow or a constant level of DC flow that might occur. This is a typical shape of the volume flow waveform. Figure 4 shows a real volume flow waveform, as well as its derivative and a corresponding electroglottograph signal. The signals were obtained from a male producing a low pitch during the /pae pae pae/ sequence in an experiment on

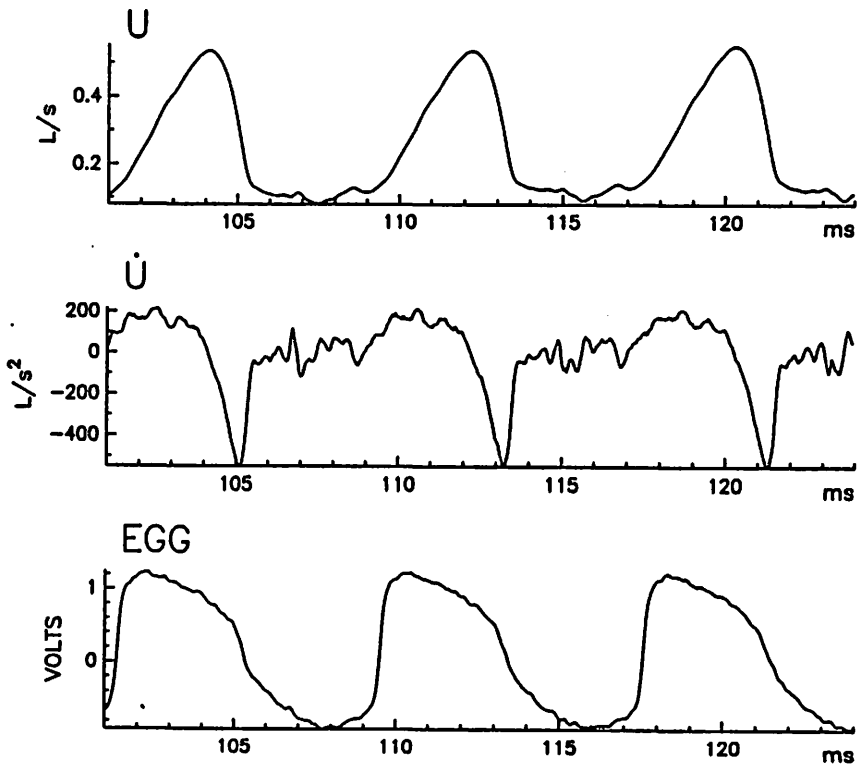

Figure 4. Inverse filtered glottal volume velocity U for a male subject phonating at a low pitch. EGG is the electroglottographic signal, and U(dot) is the derivative of the volume velocity waveform.

loudness control (Sundberg et al., 1990;Scherer et al., 1989; Titze & Sundberg, 1990).

Sundberg and Gauffin (1979, and Gauffin & Sundberg, 1989) have shown that the peak value of the volume velocity waveform (i.e., the difference between the maximum flow and the baseline flow) corresponds fairly strongly to the acoustic level of the fundamental frequency at higher levels of flow. In addition, and perhaps more importantly, the maximum slope of the change in the volume velocity during the closing of the glottis (the peak negative derivative of the flow) is strongly related to the sound pressure level of the voice (also, Scherer et al., 1989; Sundberg et al., 1990), and is considered to be the main excitation of the vocal tract (e.g., Fant, 1979). Thus the ratio of the peak flow to the peak derivative flow has relevance not only to the fine structure of the volume velocity cycle but also to the acoustic structure and quality of the resulting sound (the fundamental frequency level compared to the overall sound pressure level).

The inverse filtered glottal volume velocity flow can be obtained by using the output (microphone) pressure or the output oral flow. In addition, the reflectionless tube has been used in the past to determine the inverse filtered flow. Microphone techniques, discussed by Dr. Childers in his paper of this conference, have the primary limitation of not including the DC (or bias) levels of the airflow. The inverse filtered microphone signal results in the differentiated glottal flow signal because of the included acoustic radiation from the lips. This is a useful signal if only to yield the peak negative derivative value. Integration is necessary if the volume velocity waveform is desired. Compared to inverse filtering the oral flow, microphone techniques currently include a wider frequency range within which to obtain the glottal flow (e.g., Milenkovic, 1986, whose method is now automated into his C-Speech software). The reflectionless tube method (Sondhi, 1975; Monsen & Engebretson, 1977; Hillman & Weinberg, 1981) appears less convenient and flexible, especially regarding the correction of the volume flow cycle in the closed phase.

The Glottal Enterprises system for inverse filtering oral flow to estimate glottal volume flow (Rothenberg, 1973) uses a circumferentially vented pneumotach mask that is placed over the face. Because of meshed holes near the plane of the lips, there is relatively little acoustic loading by the mask to affect the airflow and air pressures exiting the mouth. The airflow during phonation crosses screens in the mask to create an overall dynamic pressure drop that is detected by pressure transducers. The pressure drop is linearly related to the airflow from the mouth. The frequency response for transducing the airflow is at least 1000 Hz (Rothenberg, 1977; Badin et al., 1990). The output airflow has variations in value during each cycle that correspond to the vocal tract resonance effects. With the Glottal Enterprises system, the operator manually changes the frequency and bandwidth of the first two resonances which are electronically used to inverse filter the flow signal, that is, remove the first two resonances. This inverse filtering procedure results in a signal that corresponds to the instantaneous (spatially integrated) volume flow at the glottal level. The technique is fairly easy, static calibration is not difficult, and a side pressure tube allows the estimation of subglottal pressure through measurement of the oral pressure. The acoustic structure of the voice is validly transduced through the mask to about 1000 Hz (Badin et al.,1990).

During the closed portion of the glottal cycle, the flow may be absent. Alternatively, the flow may be minimal and approximately constant through the posterior glottis. The inverse filtering process should reduce the flow to a relatively flat portion corresponding to glottal closure if there is evidence (e.g., via the electroglottograph, stroboscopy, or otherwise) that there is membranous glottal closure.

There are some cautions when using the Glottal Enterprises flow system. Vocal tract resonance effects (Rothenberg, 1981b; Fant & Ananthapadmanabha, 1982; Hillman & Weinberg,

1981) appear as ripples on the left flow-increasing side of the waveform, and during the closed or minimal flow portion. The former may be part of the flow pattern due to incomplete damping of the vocal tract acoustic effects prior to cycle onset, whereas the latter usually would need to be removed given zero or constant flow from the glottis. It is sometimes difficult to remove the resonances of the vocal tract sufficiently to feel confident that one obtains a valid inverse filtered waveform signal. The technique should be more valid at lower pitches when glottal closure can take up a significant portion of the period, and resonance effects would be damped out before the onset of the subsequent cycle. The technique is more difficult when the fundamental frequency is high or near the first formant. In addition, because the technique often does not yield sharp corners, it is sometimes difficult to measure with confidence the open quotient and speed quotient. It is recognized, however, that real flow may not have sharp corners due to gradual opening and closing of the glottis with the corresponding gradual turn-on and turn-off of the flow. A question has also been raised (refer to Titze's accompanying paper in this proceedings) that there may be a temperature effect using the Glottal Enterprises system due to breathing into the pressure tubes over a prolonged period. Also, the mask must be placed firmly against the face to prevent air leaks.

The inverse filtered flow at this time is clinically useful and should be incorporated as a routine procedure as its technology and accuracy continue to be studied. Measures of peak flow, peak derivative flow, bias offset flow, speed quotient and open quotient are highly relevant to voice quality, vibratory motion of the vocal folds, and glottal adduction. The Boston group, utilizing the efficient MITSYN (Henke, 1989) program to interactively obtain the inverse filtered flow waveform (Perkell et al., 1990), as well as other measures, have been applying the inverse filter technique to normals (Holmberg et al., 1988, 1989) and to patient populations (Hillman et al., 1989, 1990), especially in studies dealing with laryngeal hyperfunction and vocal fold lesions. These studies are helping to establish normative data collections.

It is further emphasized that the aforementioned fine detail of pressures throughout the vocal tract and of the volume velocity flow through the glottis must be accompanied by measures of the dynamic configuration of the glottis and vocal fold motion during vibration. Some information can be obtained from the electroglottographic signal, from the photoglottographic signal, and fiberoptic techniques such as stroboscopy. But other techniques, not yet known (GLIMPES, Titze, 1984, is one analysis/synthesis approach), must be developed to track subtle changes of glottal shape and other motion aspects of the vocal folds. This would allow a relatively complete picture of the interdependence of pressures, flows, and dynamic geometry that is needed to enhance the assessment of pathological, normal,and exceptional phonatory behavior.

Summary

Clinical assessment of the aerodynamics of voice production can be viewed on two levels. The gross level consists of average values of subglottal pressure (Ps), average airflow (U), glottal resistance (Ps/U), and subglottal power (PsU). These measures are relatively easy to obtain, change with relatively gross changes of laryngeal function or morphology, and should be standardized and routinely obtained in the clinic. The fine level of aerodynamic assessment consists of detailed, instantaneous cyclic measures of subglottal and supraglottal pressure, translaryngeal pressure, inferred intraglottal pressure, and the glottal volume velocity signal obtained through inverse filtering the output oral flow or pressure signal. Peak pressure and flow values, cyclic pressure and flow

waveform patterns, the peak flow derivative, timing of cyclic pressure and flow events, the bias flow offset, and the open quotient are important measures that will be meaningful for the characterization and diagnostics of subtle change in voice quality, intensity, vocal fold morphology, and neuromuscular events. This level of measurement should prove to differentiate significant changes in voice production for both the patient and the person in vocal training. Although significant data have been obtained for patients, normals and performers, normative data and standards for recording and measurement of these aerodynamic aspects are necessary.

We need further modeling and empirical research regarding the control of fine aerodynamic events, that is, the interdependence of dynamic pressure, flow, vocal fold movement, glottal configuration, laryngeal morphology, and acoustic structure. We also need to develop automatic extraction of meaningful measures from the fine and gross aerodynamic signals related to pathology, normal function, and elite performance. We need to develop normative data for the fine aerodynamic structure measures for various phonatory conditions including lesions and diseases. We need the prediction of the dynamic configuration and vocal fold movement with help from the fine structure of pressures and airflows around the larynx in addition to measures from noninvasive recordings related to glottal motion such as the electroglottograph and photoglottograph signals. Understanding the anomalies as well as consistent correlates to vocal pathologies may lie in the fine structure of aerodynamics, and this is where the future lies in more powerful diagnostic and follow-up decisions in function and phonatory improvement.

Affordable, standardized and easily used techniques need further commercial development for clinical examination and extraction of aerodynamic measures of phonation. These measures should consist of both the gross and fine structure of phonatory aerodynamcis, as well as the acoustic, configurational and vibratory aspects. Basic research will say much about the relevance of the fine structure to normal and pathological function, and commercial interests should track these developments as closely as possible.

Acknowledgements

Thanks are extended to Dr. Kimberly Fisher for reviewing an earlier version of this manuscript. Support has been provided by grant number 1 P60 DC00976-01 by the National Institute on Deafness and Other Communication Disorders to The Denver Center for the Performing Arts through The National Center for Voice and Speech.

Footnotes

FN1. Although at any instant in time there is a certain volume of air that exits the larynx, that volume of air is the addition (called the integration) of the air particle velocities across the glottis. The air particle velocities are not the same at each location along the glottis (nicely shown by Berke et al., 1989, for the canine). Tracking particle velocities at different glottal locations would be powerful diagnostically, but rather difficult to implement. Indeed, tracking particle velocities across the vocal tract at various locations would help clarify the aeroacoustic interdependence of airflow, air pressure, impedance and surface boundaries (ref. McGowan, 1988).

Bibliography

Allen, J.E. (1982). *Aerodynamics, The Science of Air in Motion, Second Edition,* (McGraw-Hill Book Company Limited, New York).

Badin, P., Hertegard, S., and Karlsson, I. (1990). "Notes on the Rothenberg mask," STL-QPSR 1/1990.

Baken, R.J. (1977). "Estimation of lung volume change from hemicircumferences," J. Speech Hear Res. 20, 808-812.

Baken, R.J. (1987). *Clinical Measurement of Speech and Voice,* (Little, Brown and Company, Boston).

Barlow, S.M., Suing, G., Grossman, A., Bodmer, P. and Colbert, R. (1989). "A high-speed data acquisition and protocol control system for vocal tract physiology," J Voice 3/4, 283-293.

Berke, G.S., Moore, D.M., Monkewitz, P.A., Hanson, D.G. and Gerratt, B.R. (1989). "A preliminary study of particle velocity during phonation in an in vivo canine model," J Voice 3/4, 306-313.

Bless, D.M. and Hirano, M. (1982). "Estimation of glottal airflow and MPT as a clinical tool," Paper, American Speech-Language-Hearing Association Convention, Toronto, November.

Bless, D.M., Hunker, C.J. and Weismer, G. (1982). "Comparison of non-invasive methods to obtain chest-wall displacement and aerodynamic measures during speech," In *Transcripts of the Tenth Symposium Care of the Professional Voice, Part I: Instrumentation in Voice Research,* edited by V. Lawrence, (The Voice Foundation, New York), 43-51.

Bouhuys, A., Mead, J., Proctor, D.F. and Stevens, K.N. (1968). "Pressure-flow events during singing," Annals of the New York Academy of Sciences, 155/1, 165-176.

Cavagna, G.A. and Margaria, R. (1968). "Airflow rates and efficiency changes during phonation," Annals of the New York Academy of Sciences, 155/1, 152-164.

Cavallo, S.A. and Baken, R.J. (1985). "Prephonatory laryngeal and chest wall dynamics," J. Speech Hear Res. 28, 79-87.

Considine, D.M. (1976). *Van Nostrand's Scientific Encyclopedia, Fifth Edition,* (Van Nostrand Reinhold Company, New York).

Cranen, B. and Boves, L. (1985). "A set-up for testing the validity of the two mass model of the vocal folds," In *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control,* edited by I.R. Titze and R.C. Scherer, (The Denver Center for the Performing Arts, Denver), 500-513.

Edmonds, T.D., Lilly, D.J. and Hardy, J.C. (1971). "Dynamic characteristics of air-pressure measuring systems used in speech research," J. Acoust. Soc. Am. 50, 1051-1057.

Fant, G. (1979). "Glottal source and excitation analysis," STL-QPSR/1, 85-107.

Fant, G. (1983). "Preliminaries to analysis of the human voice source," STL-QPSR 4/1982, 1-27.

Fant, G. and Ananthapadmanabha, T.V. (1982). "Truncation and superposition," STL-QPSR 2-3/1982, 1-17.

Gauffin, J. and Sundberg, J. (1989). "Spectral correlates of glottal voice source waveform characteristics," J. Speech Hear Res. 32, 556-565.

Henke, W.L. (1989). *MITSYN Languages, Language Reference Manual,* V7.0, (WLH, 133 Bright Rd., Belmont, MA, 02178).

Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M. and Vaughan, C. (1989). "Objective assessment of vocal hyperfunction: an experimental framework and initial results," J. Speech Hear Res. 32, 373-92.

Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M. and Vaughan, C. (**1990**). "Phonatory function associated with hyperfunctionally related vocal fold lesions," J Voice 4/1, 52-63.

Hillman, R.E. and Weinberg, B. (**1981**). "Estimation of glottal volume velocity waveform properties: A review and study of some methodological assumptions," In *Speech and Lan guage, Advances in Basic Research and Practice, Volume 6*, edited by N.J. Lass (Academic Press, New York), 411-473.

Hirano, M. (**1981**). *Clinical Examination of Voice.*, (Springer-Verlag, New York).

Hirano, M. (**1989**). "Objective evaluation of the human voice: clinical aspects," Folia Phoniatrica, 41/2-3, 89-144.

Hixon, T.J. (**1972**). "Some new techniques for measuring the biomechanical events of speech production: One laboratory's experiences," In *Proceedings of the Conference Orofacial Function: Clinical Research in Dentistry and Speech Pathology, ASHA Reports 7*, (American Speech and Hearing Association, Washington), 68-103.

Hixon, T.J., Goldman, M.D. and Mead, J. (**1973**). "Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung," J. Speech Hear Res. 16, 78-115.

Hixon, T.J., Hawley, J.L. and Wilson, K.J. (**1982**). "An around-the-house device for the clinical determination of respiratory driving pressures: a note on making simple even simpler," J. Speech Hear Dis. 47, 413-415.

Hixon, T.J. and Putnam, A.H.B. (**1983**). "Voice abnormalities in relation to respiratory kinematics," Seminars in Speech and Language 5/4, 217-231.

Holmberg, E.B., Hillman, R.E. and Perkell, J.S. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice," J. Acoust. Soc. Am. 84, 511-529.

Holmberg, E.B., Hillman, R.E. and Perkell, J.S. (**1989**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal and high pitch," J Voice, 3/4, 294-305.

Holmberg, E. and Leanderson, R. (**1983**). "Laryngeal aerodynamics and voice quality," In *Transcripts of the Eleventh Symposium Care of the Professional Voice, Part II, Medical/Surgical Sessions: Papers*, edited by V. Lawrence, (The Voice Foundation, New York), 124-129.

Isshiki, N. (**1964**). "Regulating mechanisms of vocal intensity variation," J. Speech Hear Res. 7, 17-29.

Isshiki, N. (**1981**). "Vocal efficiency index," In *Vocal Fold Physiology*, edited by K.N. Stevens & M. Hirano, (Tokyo University Press, Tokyo), 193-207.

Isshiki, N. (**1985**). "Clinical significance of a vocal efficiency index," In *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I.R. Titze & R.C. Scherer, (The Denver Center for the Performing Arts, Denver), 230-238.

Iwata, S. (**1988**). "Aerodynamic aspects for phonation in normal and pathologic larynges," In *Vocal Physiology: Voice Production, Mechanisms and Functions.*, edited by O. Fujimura, (Raven Press, Ltd., New York), 423-431.

Kitajima, K. (**1985**). "Airflow study of pathologic larynges using a hot wire flowmeter," Ann. Otol. Rhinol. Laryngol. 94, 195-197.

Kitzing, P. and Lofqvist, A. (**1975**). "Subglottal and oral air pressures during phonation. Preliminary investigation using a miniature transducer system," Med. Biol. Eng. 13, 644-648.

Koike, Y. (1981). "Sub- and supraglottal pressure variation during phonation," In *Vocal Fold Physiology*, edited by K.N. Stevens and M. Hirano (eds.), (Tokyo University Press, Tokyo), 181-192.

Koike, Y., Imaizumi, S., Kitano, Y., Kawasaki, H., and Hirano, M. (1983). "Glottal area time function and supraglottal pressure variation," In *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D.M. Bless and J.H. Abbs, (College-Hill Press, San Diego), 300-306.

Koike, Y. and Perkins, W. (1968). "Application of a miniaturized pressure transducer for experimental research," Folia Phoniatrica 20, 360-368.

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

Leanderson, R., Sundberg, J. and von Euler, C. (1987). "Breathing muscle activity and subglottal pressure dynamics in singing and speech," J Voice 1/3, 258-261.

Lieberman, P. (1968). "Direct comparison of subglottal and esophageal pressure during speech," J. Acoust. Soc. Am. 43, 1157-1164.

Lofqvist, A., Carlborg, B. and Kitzing, P. (1982). "Initial validation of an indirect measure of subglottal pressure during vowels," J. Acoust. Soc. Am. 72/2, 633-635.

McGowan, R.S. (1988). "An aeroacoustic approach to phonation," J. Acoust. Soc. Am. 83, 696-704.

Milenkovic, P. (1986). "Glottal inverse filtering by joint estimation of an AR system with a linear input model," IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-34/1, 28-42.

Miller, D.G. and Schutte, H.K. (1990). "Formant tuning in a professional baritone," J Voice 4/3, 231-237.

Monsen, R. and Engebretson, A. (1977). "Study of variation in the male and female glottal wave," J. Acoust. Soc. Am. 62, 981-993.

Perkell, J.S., Holmberg, E.B. and Hillman, R.E. (1991). "A system for signal processing and data extraction from aerodynamic, acoustic and electroglottographic signals in the study of voice production," J. Acoust. Soc. Am. (in press).

Rau, D. and Beckett, R.L. (1984). "Aerodynamic assessment of vocal function using hand-held spirometers," J. Speech Hear Dis. 49, 183-188.

Reed, C.G., Doherty, E.T. and Shipp, T. (1990). "Piezoelectric transducers for the direct measurement of medial forces in the glottis," J. Acoust. Soc. Am. 88(S1), S151(A).

Reich, A.R. and McHenry, M.A. (1990). "Estimating respiratory volumes from rib cage and abdominal displacements during ventilatory and speech activities," J. Speech Hear Res. 33, 467-475.

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

Rothenberg, M. (1977). "Measurements of airflow in speech," J. Speech Hear Res. 20, 155-176.

Rothenberg, M. (1981a). "Acoustic interaction between the glottal source and the vocal tract," In *Vocal Fold Physiology*, edited by K.N. Stevens and M. Hirano, (Tokyo University Press, Tokyo), 305-323.

Rothenberg, M. (1981b). "An interactive model for the voice source," STL-QPSR 4/1981, 1-7.

Sawashima, M. and Aoki, Y. (1982). "Phonation analyzer, simultaneous measurements of pitch, intensity and air flow during phonation," Medical Instruments 52, 342-345.

Sawashima, M., Honda, K., Niimi, S. and Hirose, H. (**1986**). "Some clinical data on aerodynamic examination using the airway interuption method," Ann. Bull. RILP 20, 217-224.

Scherer, R.C. (**1991**). "Physiology of phonation: A review of basic mechanics," In *Phonosurgery: A Symposium on Assessment and Surgical Management of Voice Disorders,* edited by D. Bless and C. Ford, [Raven Press, New York (forthcoming)].

Scherer, R.C., Cooper, D., Alipour-Haghighi, F. and Titze, I.R. (**1985**). "Contact pressure between the vocal processes of an excised bovine larynx," In *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control,* edited by I.R. Titze & R.C. Scherer (The Denver Center for the Performing Arts, Denver), 292-303.

Scherer, R., Gould, W.J., Titze, I., Meyers, A. and Sataloff, R. (**1988**). "Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis," J Voice 2, 230-244.

Scherer, R.C. and Guo, C.G. (**1990**). "Effects of vocal fold radii on pressure distributions in the glottis," J. Acoust. Soc. Am. 88(S1), S150(A).

Scherer, R.C., Sundberg, J. and Titze, I.R. (**1989**). "Laryngeal adduction related to characteristics of the flow glottogram," J. Acoust. Soc. Am. 85(S1), S129(A).

Scherer, R.C. and Titze, I.R. (**1982**). "Vocal fold contact stress during phonation," J. Acous. Soc. Am. 71(S1), S55(A).

Schlegelmilch, R.M. (**undated**). *The physics of body plethysmography.* (Erich Jaeger GmbH & Co.KG, Wurzburg).

Shipp, T. (**1973**). "Intraoral air pressure and lip occlusion in midvocalic stop consonant production," J Phonetics 1, 167-170.

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

Sondhi, M.M. (**1975**). "Measurement of the glottal waveform," J. Acoust. Soc. Am. 57, 228-232.

Sundberg, J. and Gauffin, J. (**1979**). "Waveform and spectrum of the glottal voice source," In *Frontiers of Speech Communication Research,* edited by B. Lindblom & S. Ohman, (Academic Press, London), 301-320.

Sundberg, J., Scherer, R., and Titze, I. (**1990**). "Loudness regulation in male singers," Paper presented at the Nineteenth Symposium Care of the Professional Voice, Philadelphia, June.

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

Titze, I.R. (**1986**). "Mean intraglottal pressure in vocal fold oscillation," J Phonetics 14, 359-364.

Titze, I.R. (**1988**). "The physics of small-amplitude oscillation of the vocal fold." J. Acoust. Soc. Am. 83/4, 1536-1552.

Titze, I.R. (**1989a**). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. 85/2, 901-906.

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

Titze, I.R. and Sundberg, J. (**1990**). "Acoustic power in speech and singing," Submitted for publication.

van den Berg, Jw. (**1956**). "Direct and indirect determination of the mean subglottal pressure," Folia Phoniatrica 8, 1-24.

Warren, D.W. (1976). "Aerodynamics of speech production," In *Contempory Issues in Experimental Phonetics,* edited by N.J. Lass, (Academic Press, New York), 105-137.

Webster, J.G. (1978). *Medical Instrumentation; Application and Design,* (Houghton Mifflin Company, Boston.)

Wilson, F.B. and Starr, C.D. (1985). "Use of The Phonation Analyzer as a clinical tool," J. Speech Hear Dis. 50, 351-356.

Woo, P. (1986). "Phonatory volume velocity recording by use of hot film anemometry and signal analysis," Otolaryngolgy - Head and Neck Surgery, 95/3, Part 1, 312-318.

NCVS Status and Progress Report - 1
June 1991, 167-190

Speech Therapy for Neurological Disorders of the Larynx

Lorraine Olson Ramig, Ph.D., CCC-SP
University of Colorado-Boulder
Denver Center for the Performing Arts
The Lee Silverman Center for Parkinson's (Scottsdale, Arizona)

Ronald C. Scherer, Ph.D.
Denver Center for the Performing Arts
University of Colorado-Boulder

Overview

Neurological disorders of the larynx reduce speech intelligibility (1,2). Speech therapy therefore is an important part of overall patient management. Whether speech therapy occurs in conjunction with medical (surgical, pharmacological) treatment or as a separate treatment strategy, it facilitates maximum improvement in speech intelligibility. The speech pathologist, neurologist and otolaryngologist function as a team in order to provide comprehensive diagnosis and management of patients with neurological disorders of the larynx.

The speech therapy approaches presented in this chapter are organized by laryngeal pathophysiology rather than specific neurological disorders. The traditional approach has attempted to relate neurological diagnoses with certain voice characteristics and consequently certain treatment approaches (1,2). We assume that a direct relationship exists between the laryngeal pathophysiology (biomechanical conditions of the oscillator) and the resulting voice characteristics, regardless of the specific diagnosis. This approach accommodates the numerous sources of variation in voice characteristics accompanying neurological disorders, such as compensatory behaviors, multiple neural pathologies and neuropharmacological effects, and facilitates application of principles of normal laryngeal function to treatment.

When laryngeal dysfunction is a primary contributor to reduced speech intelligibility, voice therapy is designed to maximize improvement in speech intelligibility by changing or compensating for underlying laryngeal pathophysiology. Laryngeal dysfunctions in patients with neurological disorders include problems in adducting the vocal folds (hypoadduction, hyperadduction) (3), producing a stable voice (phonatory instability) (4,5) and coordinating movements (phonatory incoordination) (3). These dysfunctions may reduce speech intelligibility by affecting the perceptual characteristics of pitch, loudness, quality, intonation and voice- voiceless contrasts (Figure 3).

When laryngeal function is affected due to a neurological disorder, other components of the speech mechanism frequently are affected as well (2,6). Therefore, it is important to assess laryngeal function and apply treatment strategies within the framework of the entire speech mechanism. One such framework includes nine functional components which are primarily responsible for producing speech: diaphragm, abdomen, rib cage, larynx, velopharynx, posterior tongue, lips and jaw (7) (Figure 1). Using a sampling of perceptual, acoustic, aerodynamic and physiologic measures, the speech pathologist evaluates the individual and interactive contribution of all components to a reduction in speech intelligibility and makes hypotheses about the pathophysiology underlying the disordered speech (8) (Figure 2). This information, used in combination with otolaryngological and neurological findings, will result in a program of speech therapy that will maximize treatment outcomes for each patient.

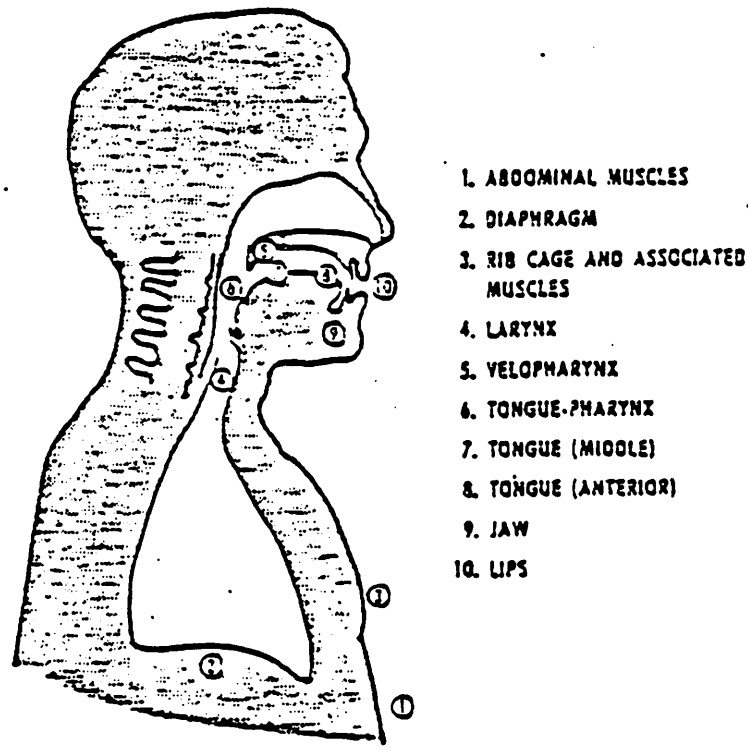

Figure 1. Functional components of the speaking mechanism, showing areas where the airstream may be valved. Adapted from Netsell (176, 177).

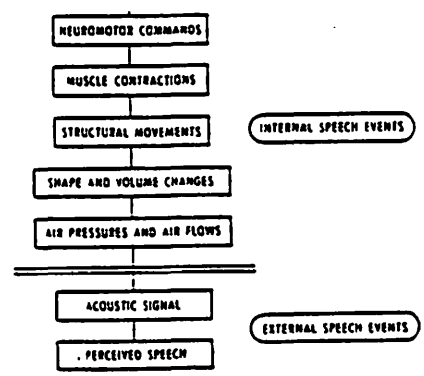

Figure 2. Components and levels and events involved in different neuropathologies of speech and language. Adapted from Kent.

Problems with Vocal Fold Adduction

Adduction and Voice Production

Phonation occurs when oscillations of the vocal folds create a pulsing airflow (acoustic signal). Oscillations of the vocal folds are created when sufficient biophysical conditions are met for laryngeal configuration (vocal fold size, shape, approximation), aerodynamics (translaryngeal and vocal fold surface pressures and airflow), tensions (active and passive) on the vocal fold tissues, and biomaterial properties (elasticities, viscosity, density). Discussions of these conditions for phonation are presented in numerous modelling papers (9,10, 11, 12, 13).

If values for these conditions are inappropriate , phonation will not occur. For example, if the vocal folds are not sufficiently adducted (hypoadducted), the airflow required to set them into motion to produce an acoustic signal may be too large for the respiratory system to create. Alternately, the adduction may be so great (hyperadducted) that the required air pressure to initially separate the vocal folds and maintain phonation is so large that again the respiratory system is inadequate to meet the need.

Adduction is a primary configurational variable of the larynx. The closeness of the vocal folds is determined by the closeness of the vocal processes of the arytenoid cartilages, and by the medial configuration of the vocal folds. Full adduction of the vocal processes may bring the vocal folds together, but bowing may prevent full longitudinal contact, thus causing a possible flow "leak" during phonation.

The quality of the laryngeal tone (and corresponding spectrum) is highly dependent upon the degree of adduction. The voice quality series breathy - normal - pressed is associated with greater adduction as seen by fiberoptic video recordings (14) and corresponds to upward tilting of the glottal flow source spectrum (greater energy in higher harmonics) and specific glottal flow waveform changes (15). At approximately the same subglottal pressure, the breathy and pressed phonations will give less sound pressure (or less loudness) than normal degrees of adduction (15). Measurement of adductory pressures in the larynx would be highly informative for clinical concerns, but studies of this nature are difficult (16).

Adduction may be related to vocal pitch. Phonatory pitch is thought to be strongly related to the ratio of effective stiffness to effective mass (17). The effective stiffness of the vocal folds is a function of their tension (especially the tension of the epithelial layer), and thus pitch will rise as the tension rises with vocal fold elongation (due primarily to cricothyroid muscle contraction or lateral stretch with increased subglottal pressure)(17). If the effective mass in vibration decreases because of medial compression, the pitch also may rise (18,19).

Measures of adduction have been related to the electroglottograph (EGG) signal under the assumption that the EGG signal is proportional to contact area (20,21). Adduction measures that may be useful clinically, such as the Abduction Quotient (22), Adduction measure (23) and EGGW (24) appear highly promising, but need extensive validation.

Hypoadduction

Certain neurological disorders are accompanied by inadequate vocal fold adduction or hypoadduction. The particular type and extent of hypoadduction may be associated with the site and extent of the related neurological damage (2). Hypoadduction may accompany a variety of neurological disorders, but is typically associated with lower motor neuron (flaccid) involvement which is characterized by paresis (weakness) or paralysis (immobility), atrophy and fatigue (2).

Some patients have no vocal fold adduction during speech and are unable to produce phonation voluntarily (25). Closed head injury with brain stem contusion (26,27) and brain stem CVA are common etiologies for this form of hypoadduction. Other patients have reduced adduction and only partially adduct their vocal folds during speech. This may be unilateral or bilateral and result from damage to (or a disease of) the laryngeal adductor muscles (lateral cricoarytenoid and interarytenoids) (2,5), to the nerve that innervates them (recurrent laryngeal nerve branch of the vagus) (28), or to a disease of brain stem motor neurons, such as flaccid forms of amyotrophic lateral sclerosis (2,29). Other patients may adduct their vocal folds posteriorly, but they do not close medially and are said to be bowed. This form of hypoadduction has been associated with bilateral superior laryngeal nerve paralysis (2,30,31) and Parkinson's disease (32). Progressive hypoadduction involving restriction of adductor-abductor movements, associated with fatigue accompanying continuous talking has been observed in patients with myasthenia gravis (31).

The major effect of hypoadduction on speech intelligibility is reduced loudness. This is frequently accompanied by a breathy, hoarse voice quality and in some cases diplophonia (the perception of two pitches). Aerodynamic and acoustic data may show high airflow, short or absent closed phase, low signal to noise ratio, reduced intensity range and short maximum duration (6,33). In certain cases of hypoadduction, patients may be unable to produce adequate voice-voiceless contrasts.

Therapy for hypoadduction-laryngeal

The primary focus of voice therapy for patients without voluntary phonation is to obtain regular voicing (25). This may be accomplished by attempting to elicit reflexive phonatory behavior (i.e., laughing, coughing) on a repetitive basis with simultaneous stimulation of subglottal pressure using externally applied abdominal pressure (25). Sapir and Aronson reported that digital manipulation of the larynx and coughing were used successfully in eliciting phonation from two patients with whispered phonation accompanying closed head injury (34).

For patients with reduced adduction accompanying muscle or nerve damage, the primary goal is to increase loudness and reduce breathy, hoarse voice quality by increasing vocal fold adduction. Procedures to accomplish this are probably the most well-referenced of all techniques in the area of neurological voice disorders (2, 35, 36, 37). The techniques include pushing, pulling and lifting while phonating (38). For example, a patient pushes the palms of his/her hands together or pulls up on the arms of his/her chair while producing an "ah" vowel. The goal is to maximize adduction by "reinforcing the sphincter action of the laryngeal muscles engaged in phonation" (38, p.365). When done as a systematic exercise, progressing from vowels to syllables to words to phrases, pushing helps to strengthen the muscles of adduction (39). Recently these pushing exercises have been used to improve adduction and increase loudness in patients with bowed vocal folds accompanying Parkinson's disease (40, 41,42). Other adduction enhancing techniques include: hard glottal attack, turning the head to one side or the other (to increase tension on the paralyzed fold), digital manipulation of the thyroid cartilage (to more closely approximate the vocal folds) (2) and speaking at a higher pitch (to make use of the adductory function of the cricothyroid muscle) (37).

An important variable in treatment success is a patient's ability to monitor improved adduction or loudness (43). This has been facilitated through various forms of biofeedback, such as visual feedback of vocal fold adduction through videoendoscopy (44,45) or intensity feedback through a voice light or visipitch (25,40,41).

Therapy for hypoadduction-support systems

In order to facilitate the goal of increased loudness and improved quality, the respiratory system is often a focus of treatment in cases of hypoadduction. The goal of respiratory treatment is to achieve a consistent subglottal pressure during speech that is produced with minimal fatigue and appropriate breath group lengths (46).

The first step may be the stabilization of posture (3, 35, 47, 48, 49). This may involve the use of support devices such as neck braces, girdling, expiratory boards or paddles (50) that are implemented in collaboration with the physical therapist and physician. For some patients, working on respiratory support from the supine position is most effective (25, 51).

Once the patient's postural support is maximized, he/she may then be instructed to increase levels of subglottal pressure. One of the most frequently targeted goals is "five cm of water pressure maintained for five seconds " (52, p.329). Various procedures assist the patient to systematically increase the amounts of air pressure he/she can produce and the duration (53). The hypothesis is that this training improves the strength and coordination of the respiratory muscles so that the patient will be able to produce sound at a constant loudness and quality for longer periods (25). Other techniques to strengthen respiratory muscles include exercises against a resistive load, tracking of a visible analogue of respiratory behavior and controlled exhalation (3, 52, 53, 54, 55, 56).

In order to train improved coordination of respiration and phonation, various techniques such as maximum duration vowel phonation (35, 40, 41, 42, 57) and phonation with simultaneous respiratory (e.g., respitrace) and vocal (e.g., visipitch) feedback have been suggested (25). During speech tasks, patients have been encouraged to take breaths more often, produce fewer syllables on each exhalation (43, 58) and to initiate phonation at the beginning of exhalation (25). The goal is to assist the patient in organizing his/her linguistic output into units that are manageable by his/her peripheral speech mechanism (43) and that afford him/her the greatest gains in speech intelligibility.

The patient with hypoadduction may also be encouraged to maximize oral resonance in order to increase loudness and quality. The patient would be encouraged to enlarge the oral cavity by dropping his/her jaw , openning his/her mouth and lowering his/her tongue while speaking. This posture facilitates maximum transfer of acoustic sound energy and production of a louder voice (37,59). In order to compensate for reduced loudness, patients may be encouraged to slow their rate and overarticulate. These techniques may improve overall speech intelligibility by allowing the patient additional time to reach the articulatory targets in order to produce the most precise articulation within his/her capabilities (59, 60).

Patients with vocal fold hypoadduction also may have weak or paralyzed velopharyngeal and articulatory mechanisms. In the case of velopharyngeal insufficiency, surgical or prosthetic intervention may be necessary in order to provide adequate physiological support after which speech therapy is initiated (43). Management of the velopharyngeal and articulatory mechanisms are summarized in various works (3, 25, 61, 62).

Speech therapy and medical management in combination for hypoadduction

In certain cases of hypoadduction, a combination of medical management and speech therapy will offer maximum results. After a patient has achieved the best voice possible within the physiological limits of his system through speech therapy, medical intervention designed to maximize vocal fold adduction may be necessary (63, 64, 65, 66, 67, 68, 69, 70, 71). After such surgery, speech therapy will teach the patient how to achieve maximum vocal efficiency within

the new range of function afforded by his/her modified laryngeal mechanism (72, 73, 74).

In other cases, such as in myasthenia gravis, the primary improvement in voice production may occur with medical management (thymectomy or anticholinesterase drugs) (75) and exercises to increase vocal fold adduction should not be attempted (59, 76).

Before medical management or speech therapy, or when neither provide the necessary gains in speech intelligibility, augmentative forms of communication (discussed below) may be useful.

Hyperadduction

Certain neurological disorders result in excess vocal fold adduction or hyperadduction. In some cases, the ventricular (false) vocal folds may hyperadduct as well (2, 77). The particular type and extent of hyperadduction may be associated with the site and extent of the related neurological damage. Hyperadduction most frequently occurs in cases of upper motor neuron system disorders characterized by spasticity and hypertonicity and extrapyramidal system diseases accompanied by abnormal involuntary movements (for example, tics, chorea, dystonia) which may be focal or generalized (2). Hyperadduction may also occur in cases of long term use of psychotropic and antiparkinson drugs and is then considered a symptom of tardive dyskinesia (78, 79).

In some patients, hyperadduction is so extreme and continuous that they are unable to generate phonation. Severe cases of adductory laryngospasm associated with adductor spastic dysphonia is an example of this. Other patients may have moderate hyperadduction which results in strained-strangled, harsh voice quality with excessively low pitch and reduced loudness (2) such as in cases of bilateral stroke (pseudobulbar palsy) and spastic cerebral palsy. Still other patients have a mild, continuous hyperadduction resulting in pitch breaks and a pressed, strained, harsh quality, such as in cases of spastic amyotrophic lateral sclerosis (2). When random periods or "bursts " of hyperadduction or hypoadduction occur, they are frequently associated with extrapyramidal hyperkinesias such as Huntington's disease, dystonia and myoclonus. The resulting periods of adductory or abductory laryngeal spasm may generate random, intermittent periods of spastic or breathy voice arrest respectively (2, 4, 80).

It should be pointed out that hyperadduction may be compensatory. For example, a patient may have weak respiratory support or velopharyngeal closure and hyperadduct in order to manage the air stream for adequate loudness (56). The harsh, high-pitched voice with reduced loudness and pitch variations observed in patients with spinal forms of multiple sclerosis has been associated with the extreme neck and laryngeal tension and upper thoracic and clavicular breathing adopted by these patients in an attempt to produce phrases of normal length (3).

The major effect of hyperadduction on speech intelligibility is a quality disorder ranging from a pressed, strained harsh voice to the inability to produce phonation. Aerodynamic and acoustic data reveal decreased mean air flow rate (81) and reduced fundamental frequency and fundamental frequency range (82). Smitheran and Hixon reported extremely high measures of laryngeal resistance for a speaker who had strain-strangled voice quality following multiple, bilateral cerebrovascular accidents (83).

Therapy for hyperadduction-laryngeal

The primary focus of voice therapy for patients with hyperadduction is to decrease the pressed strained voice by reducing vocal fold hyperadduction. Procedures to accomplish this include those designed to relax laryngeal musculature and facilitate easy voice onset. These techniques

frequently begin with progressive whole body relaxation (84, 85, 86, 87) and then focus on relaxing laryngeal musculature. Techniques of biofeedback enhanced relaxation previously used on facial musculature (88, 89) have been applied to laryngeal musculature (90, 91). The specific form of feedback may be EMG or visual (videoendoscopy). For example, surface electrodes may be attached to a patient's extrinsic laryngeal muscles and he/she uses auditory or visual feedback of the muscle tension to guide his/her relaxation. A progressive hierarchy is established from quiet breathing through vowels, syllables, words, sentences to conversation, with the patient continuously targeting low levels of extrinsic laryngeal muscle tension. Another approach for relaxing laryngeal muscle tension and thereby hyperadduction includes manual massage of extrinsic laryngeal muscles (2, 3).

Some approaches such as the chewing approach (92), the yawn-sigh approach (93), chanting (94) and the delayed auditory feedback approach (59) are based on the hypothesis that when phonation is produced in the context of these reflex-like (2) or continuous phonation responses, it will be more relaxed and less hyperadducted. This relaxed voice production is then shaped into more natural sounding conversational speech. The breathy sigh can be shaped into a relaxed vowel, then to single syllable words beginning with "h", followed by open mouth vowels and a nasal consonant or continuant, followed by short duration nasal humming (95).

Improved voice quality has been reported in strain-strangled voice when a speaker raised his pitch, rotated his head backward and initiated utterances from a high lung volume; these behaviors were associated with decreased air way resistance (83). It was suggested that improved voice quality was the result of passive abduction of the vocal folds brought about by the tracheal tug associated with a lower diaphragm position at a higher lung volume level (83).

Therapy for hyperadduction-support systems

In order to facilitate the goal of improved voice quality, the respiratory system is often a focus of treatment in patients with hyperadduction. The goal of respiratory treatment is to achieve consistent, steady air flow with relaxed respiratory musculature (95).

The first step may be stabilization of posture. Supporting the abdominal musculature with an elastic band or selective positioning in a reclining wheel chair has helped some patients with spastic dysarthria produce better air flow with less effort and reduced strain-strangled phonation (95).

Once the patient's postural support is maximized, he/she may then be instructed in procedures of relaxed abdominal breathing (59, 95) in order to provide the greatest respiratory support with the minimum muscle tension throughout the speech mechanism. These activities may be combined with the progressive relaxation procedures referenced earlier (84, 85, 86, 87). Patients are encouraged to keep their phonatory productions short and relaxed (95).

In order to remove the laryngeal focus and encourage reduced hyperadduction, some clinicians (73) encourage "placement" of the vocal resonance in the frontal nasal area, "tone focus in the mask area". A modest improvement in articulatory precision without overflow of tension into the oral or laryngeal/respiratory musculature also has been suggested to improve overall speech intelligibility (95).

Speech therapy and medical management in combination for hyperadduction

In some patients with hyperadduction, a combination of speech therapy and medical management will afford the best vocal results. For example, adductor spastic dysphonia is a disorder of hyperadduction that has a reputation of being especially resistant to speech therapy (2, 94). Because adductor spastic dysphonia may be the result of many different disorders (a musculoskeletal tension disorder, a conversion reaction, a neurological disorder of organic (essential) tremor or dystonia), it may require many types of therapy (2). The following approaches have been suggested (2) : 1) symptomatic voice therapy including musculoskeletal tension reduction, establishing natural pitch, "tone focus in the mask area", and abdominal breathing (73) when the spastic dysphonia is due to excess muscular tension ; 2) psychotherapy when the spastic dysphonia is due to conversion reaction or to depression or anxiety and; 3) medical management such as recurrent laryngeal nerve section (96) and more recently botulinum toxin injection (97) when the disorder is associated with a neurological disorder.

Many of the techniques used to reduce laryngeal hyperadduction (e.g., relaxation, singing therapy, stuttering therapy techniques, hierarchical analysis building from periods of normal voice, chanting, chewing, glottal fry, placing the voice, speaking on inhalation and yawn-sigh) have been applied to adductor spastic dysphonia patients with varied success (94, 98).

When medical intervention for hyperadduction is performed (96,97) speech therapy procedures designed to maximize vocal efficiency are recommended (99). For example, if the post surgical voice is breathy, techniques to elevate the pitch and increase glottal attack have been suggested (37, 99).

Similarly, certain other forms of hyperadduction accompanying extrapyramidal disorders (e.g. Huntington's disease) are resistant to speech therapy. The primary treatment approaches for individuals with such hyperkinesias involve medical (pharmacological and surgical intervention), rather than behavioral management (100). Because the speech symptoms are so closely related to the underlying movement disorder, any improvement in speech symptoms is dependent on modification of the severity of the movement disorder (100). The primary speech therapy for these patients may involve instructions about managing verbal interactions such as: 1) maintainence of eye contact; 2) asking listeners to inform the patient when he/she has not been understood; 3) repeating a word when an abnormal movement has interfered and; 4) introducing the topic of conversation (99).

For certain of these patients, it may be appropriate to consider augmentative forms of communication (see below).

Problems with Phonatory Stability

Phonatory stability and voice production

Laryngeal vibrations are quasi-periodic, meaning that the period of time between the same phase marker of adjacent phonatory cycles is not constant across a series of cycles. Biomechanical asymmetries (such as unilateral tissue changes) may give rise to significant vocal fold motion asymmetries (101, 102). Neuromuscular innervation abnormalities may give rise to laryngeal muscle perturbations (103, 104, 105), abnormal glottal adduction behavior, airflow disturbances from abnormal respiratory muscle function, and vocal tract tissue oscillations. Mucus accumulation on the vocal folds can give rise to transients that sound like sharp "bubbling". Turbulent airflow through the glottis is an aperiodic noise source that causes the perception of breathiness.

The regularity of phonation can be measured or inferred from visual recordings (e.g., from high speed photography (81, 106, 107)), or from signals related to vocal fold movement such as the electroglottograph signal(108, 109) and the photoglottograph signal (110). Typically, however, perturbation measures have been from an accelerometer or a microphone (airborne acoustic) signal. For the latter, perturbation etiology is concerned with factors that can affect the glottal airflow cycles and the effects on the acoustic signal from vocal tract oscillations (14).

Cycle-to-cycle phonatory perturbations (or instabilities) can be categorized into at least two categories, short term and long term (14). Short term perturbations refer to cycle-to-cycle measures such as jitter (perturbation of frequency or period) and shimmer (perturbation of amplitude). Phonation periods typically range between 4 and 11 msec, and jitter values, average differences between adjacent cycles, usually are between 0.15 to 4% of this range. Shimmer values also are small, typically 0.5 to 5% of the cycle amplitudes. A manifestation like tremor, however, is a relatively slow perturbation (5 to 12 fluctuations per second), and detection calls for a long term measure such as the coefficient of variation (the standard deviation divided by the mean) (14), correlation measures (111) and other detection techniques (112).

Increased values of vocal perturbation measures may be associated with voice quality changes. For example, the harmonics-to-noise ratio, a combination measure of waveshape, period and amplitude change from cycle-to-cycle, has been associated with measures of roughness (113). It has been reported that the perception of roughness increased as synthesized jitter and shimmer increased over a wide range of values (114, 115, 116). However, for values of jitter and shimmer within normal limits for tokens that were synthesized or produced by humans, roughness judgments were not correlated with perturbation values (116).

The definition of jitter is important in data interpretation. If the average absolute time difference between consecutive periods is used, males tend to have greater jitter than females, but not signficantly so (109, 117). If this jitter is divided by the average period (the inverse of the fundamental frequency), then females appear to have greater jitter than males (109).

It may be possible to screen carcinoma patients successfully with perturbation testing, but not necessarily patients with morphological or tension changes of the vocal folds (117). Perturbation may not correlate well with narrow age ranges (117) or when physiological condition is balanced (118, cf. 119). Perturbation measures can depend on physiological condition (118), sex (120) and vowel (119, 120, 121).

Increased phonatory instability

Certain neurological disorders are accompanied by increased phonatory instability. The particular type, extent and regularity of instability may be associated with the site and extent of the related neurological damage (5). Long-term fluctuations and short-term changes can occur as well as random or continuous use of alternative modes of voicing such as ventricular phonation, glottal fry or diplophonia (2). These forms of instability may occur singlely or in combination and may be related to the problems of adduction discussed previously.

One type of long-term phonatory instability is vocal tremor. Vocal tremor has been associated with oscillations in the adductory-abductory system, the cricothyroid muscles or the ventricular folds (122). Frequently tremor may occur simultaneously in respiratory, laryngeal and articulatory musculature (1). The neurological origin of tremor is thought to be a central tremor generator located in the cerebellum or thalamus or a peripheral stretch reflex (123). Vocal tremor has been observed in many neurological diseases such as Parkinson's disease (2, 5), post-

encephalitic Parkinsonism (124), essential tremor (2, 125), ataxia (126) and spinal muscular atrophy (125). The major effect of long-term instabilities on speech intelligibility is perceived "tremor", shakiness or quivering in the voice. Acoustic data reveal frequency and amplitude oscillations ranging from 5 to 12 Hz (2, 125).

Short-term phonatory instabilities have been related to cycle-to-cycle irregularities in the adductor-abductor system or to irregularities in vocal fold elasticity (22). While the neural bases of short-term instabilities are not well understood, they have been associated with variations in single motor unit activity (103, 127). The major effect of short-term instabilities on speech intelligibility is perceived "hoarse, rough" voice quality (113, 114, 115, 116). Acoustic data show abnormally high measures of cycle-to-cycle differences in amplitude (shimmer) and time (jitter) in neurological diseases such as myotonic muscular dystrophy (5), Parkinson's disease (5, cf 128), amyotrophic lateral sclerosis (29).

Ventricular (false) fold phonation, glottal fry phonation and diplophonia are forms of phonatory instability accompanying a number of neurological disorders of the larynx such as Huntington's disease (4), adductor spastic dysphonia (2) and unilateral nerve paralysis (2). Ventricular phonation, which may develop as a result of excess muscular tension or as a compensatory form phonation (2), is hoarse and low-pitched, with restricted pitch and loudness ranges. Glottal fry, which may be produced by the true or ventricular folds or both in combination, may be produced under conditions of extreme glottal resistance (tension) (129) or with flaccid (relaxed) vocal folds approximated tightly with very little air flow and very little subglottal pressure (130). It is characterized by a very low pitch and a "creak or fry -like" quality. Diplophonia, which is the simultaneous production of two pitches, may occur in cases of unilateral vocal fold paralysis when each vocal fold apparently vibrates at a different frequency or in cases where both the ventricular folds and the true folds vibrate (2, 37, 131). All of these instabilities disrupt the quality of voice and reduce speech intelligibility.

Therapy for phonatory instability

The major focus of therapy for patients with phonatory instability is to reduce the unsteady, hoarse, rough voice quality by targeting steady, clear phonation. Patients are encouraged to maximize respiratory and laryngeal coordination, as discussed previously, in order to sustain steady voicing with consistently good quality. Treatments discussed earlier to promote more efficient vocal fold adduction have been reported to have positive effects on phonatory stability as well (59). For example, improved voice quality and reduced acoustic measures of phonatory instability (jitter and shimmer) have been measured in patients with Parkinson's disease after therapy designed to promote vocal steadiness and increased adduction (40, 41, 42). Speaking on inhalation has been suggested as a technique to facilitate true vocal fold vibration in patients using ventricular phonation (132). In some patients, excess saliva may interfere with stability of voice production and reminders to swallow or drink water before speaking may be helpful (133).

Enhanced auditory and visual feedback (e.g., language master, tape recorder, visipitch) is frequently needed to assist the patient in monitoring his or her voice quality. Visual feedback (videoendoscopy) has been useful in elimination of ventricular fold phonation (45, 134).

Speech therapy and medical intervention in combination

In certain cases of phonatory instability, the optimum voice is obtained through a combination of speech therapy and medical management. For example, in cases of short-term

instabilities, pharmacological treatment may provide the necessary changes in underlying laryngeal physiology to generate a more stable voice with the assistance of speech therapy. In some patients with Parkinson's disease, changes in phonatory stability appear to be related to neuropharmacological treatment (2, 135, 136).

While enhanced respiratory and phonatory interaction may positively affect the secondary (compensatory) behaviors accompanying vocal tremor, speech therapy techniques have not been effective in reducing the primary symptom of vocal tremor (2). Medical intervention (surgical or pharmacological) may offer some relief (137, 138). However, bilateral ventrolateral thalamic surgery for relief of tremor may result in significantly reduced loudness (139).

Problems with Phonatory Coordination

Phonatory Coordination And Voice Production

At one level, measures of vocal instabilities of prolonged, steady vowel phonations should be sensitive to basic neuromuscular and structural abnormalities, and thus should continue to be pursued in the attempt to aid diagnostic, therapy and functional insights. On another level, it is realized that speech is a dynamic complex of physiologic and physical change, resulting in an extremely precise, highly coordinated system, within which laryngeal function plays an essential role. At any instant in time, the intended speech output requires relative levels and change of pitch, loudness, quality, subglottal pressure and air flow, as well as relative position and position change of laryngeal cartilages and other articulators, such as the tongue, lips, velum and jaw. The proper loudness and quality of voiced sound highly depends upon the ability to produce the correct laryngeal volume velocity waveform, that is, the series of air pulses exiting the glottis during speech (140). Regulation of the vowel intensity and the efficiency with which voice sound is produced is dependent upon the fundamental frequency, subglottal pressure and the degree of adduction and amplitude of vibration of the vocal folds (141). The correct production of consonants, vowels and their interdependencies require the correct timing of abduction and adduction of the vocal folds, with ongoing dynamic changes of vocal fold tension and subglottal pressure to create appropriate prosody (17, 142, 143, 144, 145, 146, 147 148).

Problems with Prosody

Many neurological disorders are accompanied by disordered prosody (1, 25), which includes problems with stress patterning, intonation and rate-rhythm (149, 150). The vocal control prerequisites for prosody include the following: loudness variation, adequate duration of phonation, appropriate pitch level, pitch variation and acceptable voice quality (149). The particular type and extent of the disordered prosody may be associated with the site and extent of the related neurological disorder (151, 152). For example, reduced prosody or "aprosody" has been reported in patients with Parkinson's disease and right hemisphere damage (151) and has been characterized by reduced loudness, monopitch, diminished stress contrasts and rate abnormalities (1). In contrast, disordered prosody or "dysprosody" was observed in patients with ataxic dysarthria and apraxia and was characterized by excessive sweeps of fundamental frequency, scanning, staccato, dissociated and segregated patterns (1, 151). Because it is clear that disordered prosody reduces speech intelligibility in many patients (25, 153), speech therapy procedures have been designed to improve prosodic disorders.

Speech therapy for disordered prosody

For patients with reduced prosody or "aprosody" frequently seen in Parkinson's disease and right hemisphere damage, the goal is to heighten the relationship between the meaning and production of an utterance (25). If a patient understands the meaning of the utterence yet cannot signal stress, treatment should involve identifying the components of stress that the speaker can control (25). Increased fundamental frequency variation in reading was measured in a group of Parkinson's disease patients following one month of therapy directed toward stimulating the vocal production prerequisites underlying prosody (40, 41). These observations were accompanied by increases in maximum duration of phonation, maximum fundamental frequency range, phonatory stability and intelligibility (40, 41). For Parkinson's patients with excessive speaking rates, rate control also may be an effective way to improve prosody (154). Gesturing, such as touching a place on a pacing board with the production of each syllable, has been suggested for modifying rate and stress (155). This "gestural accompaniment" may allow a speaker to slow rate and use appropriate stress in order to improve intelligibility (156).

For patients with disordered or excessive prosody frequently seen in ataxic patients (126), it has been suggested that treatment generally involves reducing excessive fundamental frequency and intensity variations and increased use of duration adjustments such as vowel prolongations and pausing (43, 157). Use of breath groups (25, 158), contrastive stress drills (4, 43), contrastive intonation contour drills (159), loudness manipulation (43) and various forms of visual feedback (e.g. with the use of the visipitch) have been suggested as useful therapy approaches.

Problems with the voice-voiceless contrast

The inability to produce the voice-voiceless contrast is a characteristic of certain neurological disorders of the larynx. In this case, voiceless phonemes are voiced, unlike problems with hypoadduction in which the inability to generate sufficient voicing may result in voiced phonemes being perceived as voiceless. Extended consonantal voicing has been referred to as the laryngealization of stops (160) and the "fused pattern" of continuous voicing (151). This reduced laryngeal devoicing gesture has been observed in patients with Parkinson's disease (161) and spastic dysarthria (95). Exaggerated deterioration of the laryngeal tissues, disordered central drive to the posterior cricoarytenoid muscle or an interaction of these factors may explain the reduced voiceless interval duration in Parkinson's patients (161).

Therapy for voice -voiceless contrast

It has been suggested that in order to teach patients to produce perceptually different voice/ voiceless pairs, aspects other than the voicing should be exaggerated (25). For example, patients have successfully been taught to use variations in aspiration (25) as well as modifications in vowel duration preceding voice/voiceless pairs (3). These techniques are frequently taught within the framework of an intelligibility drill in which the patient practices minimal contrast word pairs (25).

Augmentative Communication

When speech therapy and/or medical management are unable to offer the patient a level of communication intelligibility adequate for his/her needs, an augmentative communication system should be considered (162). If the patient has only laryngeal pathology primarily affecting

loudness, then an amplification system (163, 164) or artifical larynx (165) may be useful. When other components of the speech mechanism are affected so that intelligible articulation is impossible, more sophisticated forms of augmentative communication systems may be appropriate (166). These systems range from alphabet and word boards to computer based systems (25). Selection of the appropriate system for each patient is based upon his/her cognitive, language and motor skill abilities (25, 167, 168) which actually may improve with use of an augmentative system (162). Issues related to augmentative systems are summarized in a number of works (166, 169).

Other Considerations

Because the underlying physiologic breakdowns in patients with neurological disorders are often severe and sometimes degenerative, the most realistic therapy goal may be "compensated communication intelligibility" (3, 43), rather than returning the speaker to normal communication (2). In the case of degenerative diseases, the therapy goals may change as the disease progresses (6), thus short and long-term goals and objectives should be consistent with the prognosis for the disorder (170), with the intent of maximizing communication effectiveness. Age, educational and vocational status, home environment and communicative needs should be considered in the establishment of therapy goals (170).

Voice therapy for patients with neurological diseases should be intensive, vigilant and coordinated with other forms of treatment (6). Intensive therapy, with a direct focus on underlying laryngeal pathophysiology, offered maximum gains in speech intelligibility to a group of patients with Parkinson's disease (40, 41). While there are limited data on the efficacy of voice therapy for patients with neurological disease, it has been suggested that the following variables influence treatment outcome in dysarthria: neurologic status and history, age, "automatic" adjustments, personality, intelligence and support systems (171). Dysarthric speakers can improve communication if physiologic support for speech is enhanced or if patients can learn to make better use of their residual support (3). In addition, successful communication for these patients may be facilitated by "environmental management" procedures which reduce background noise, for example, and encourage active listener roles (42, 43, 100,172).

Today our approaches to speech therapy for neurological disorders of the larynx are based upon systematic application of principles of normal laryngeal function (11, 14) and generalization of techniques used in the treatment of patients with functional voice disorders (2, 37) and dysarthrias (3, 25, 156). Research which advances our knowledge of the relationship between voice characteristics and underlying laryngeal pathophysiologies will contribute to our future treatment success.

It has been suggested that the laryngeal system can be regarded as a microcosm of the entire speech mechanism (156), and therefore the larynx may reflect fine motor central impairments prior to other speech components (25). Because voice disorders have been reported as initial symptoms of various neurological disorders, such as Parkinson's disease (2, 173), phonatory analysis should be considered for its contribution to neurological diagnosis.

Conclusions

The ability to generate maximally intelligible speech affords great psychological and communicative benefits to the patient with a neurological disorder of the larynx. The combined

services of the speech pathologist, neurologist and otolaryngologist can provide the patient with optimal speech intelligibility. Because the speech pathologist understands the production and rehabilitation of communication skills, he/she makes an effective and necessary contribution to diagnosis and treatment of patients with neurological disorders of the larynx.

Acknowledgement

We gratefully acknowledge support from grant NS24409-02 from the National Institutes of Health.

Figure 3

Laryngeal Disorder	Examples of Associated Neurologic Disorder	Perceptual Characteristics	Therapy Goals and Techniques
Hypoadduction	laryngeal nerve paralysis; Parkinson's disease	reduced loudness breathy, hoarse quality	increase loudness; increase adduction and respiratory support
Hyperadduction	spasticity; extrapyramidal disorders	pressed, harsh strain-strangled quality	reduce strained quality; relax laryngeal and respiratory musculature
Phonatory Instability	most neurological disorders of the larynx	tremorous, rough hoarse quality, pitch breaks, fry	increase steady, clear phonation; maximize respiratory and laryngeal coordination
Phonatory Incoordination			
prosody	Parkinson's disease; ataxia	reduced melody of speech; excessive melody of speech	stimulate improved melody of speech; rate control
voice/ voiceless contrast	Parkinson's disease	continuous voicing	emphsize other aspects contrasts, e.g., aspiration

Figure 4: Case example

Mr. J, a 60 year old male, was diagnosed with Parkinson's disease three years ago. One of his first complaints was the inability to project his voice. Speech assessment revealed that reduced loudness and a breathy, hoarse voice quality were the characteristics that reduced his speech intelligibility most significantly. His intonation was characterized as flat. Acoustic data supported these perceptual observations. Laryngological examination revealed moderately bowed vocal folds. An intensive (daily) program of speech therapy was initiated which focused on increasing vocal loudness, stability and intonation. Techniques included for example, maximizing vocal fold adduction through pushing exercises, maximizing respiratory and phonatory coordination and phonatory stability by practicing steady vowel phonation sustained for maximum duration and stress contrast drills. At the end of one month, Mr. J.'s speech intelligibility improved; his voice was louder, steadier and had greater intonation. Both he and his family noticed that he was easier to understand. These observations were supported by acoustic measures. Post-treatment laryngeal examination revealed increased vocal fold adduction. It was recommended that Mr. J. continue in weekly follow-up therapy for one month. He was encouraged to engage in daily voice practice at home.

References

(1) Darley FL, Aronson AE, and Brown JR. : Motor Speech Disorders, Philadelphia, W.B. Saunders, 1975.

(2) Aronson AE : Clinical Voice Disorders, ed. 2. New York, Theime, 1985.

(3) Rosenbek JC, LaPointe LL : The Dysarthrias: Description, Diagnosis, and Treatment, in Johns DF (ed): Clinical Management of Neurogenic Communicative Disorders. ed.2. Boston, Little, Brown and Company, 1985, pp 97-152.

(4) Ramig LA: Acoustic analysis of phonation in patients with Huntington's disease: preliminary report. Annals of Otology, Rhinology and Laryngology 1986;95(3):288-293.

(5) Ramig LA, Scherer RC, Titze, IR and Ringel SP: Acoustic characteristics of voice of patients with neurological disease: a preliminary report. Annals of Otology, Rhinology and Laryngology 1988;97(2):164-172.

(6)Bless DM: Voice Disorders in the Adult: Treatment. in Yoder DE and Kent RD (eds): Decision Making in Speech-Language Pathology. Philiadelphia, Decker, 1988, pp 140-143.

(7) Netsell R: Speech Physiology. in Minifie FD, Hixon TJ and Williams F (eds): Normal Aspects of Speech, Hearing, and Language. Englewoodcliffs, N.J., Prentice-Hall, 1973.

(8) Kent RD: Study of vocal tract characteristics in the dysarthrias. Presented to the Veterans Administration Workshop on Motor Speech Disorders. Madison, Wis., 1976.

(9) Ishizaka KL and Flanagan JL: Synthesis of voiced soundsb from a two mass model of the vocal cords. Bell System Technical Journal 1972; 51: 1233-1268.

(10) Titze IR: On the mechanics of vocal fold vibration. Journal of the Acoustical Society of America 1976;60/6:1366-1380.

(11) Titze IR: Biomechanics and Distributed-mass Models of Vocal Fold Vibrations. in Stevens KN and Hirano M (eds): Vocal Fold Physiology. Tokyo, University of Tokyo Press, 1981, pp 245-264.

(12) Alipour-Haghighi F and Titze IR: Simulation of Particle Trajectories of Vocal Fold Tissue during Phonation. In Titze IR and Scherer RC (eds): Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control. Denver, The Denver Center for the Performing Arts, 1983, pp 183-190.

(13) Perlman AL and Titze IR: Measurement of Viscoelastic Properties in Live Tissue. in Titze IR and Scherer RC (eds): Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control. Denver, The Denver Center for the Performing Arts, 1983, pp 183-190.
(14) Scherer RC, Gould WJ, Titze IR, Meyers AD and Sataloff RT: Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. Journal of Voice 1988;2/3, 230-244.
(15) Sundberg J: The Science of the Singing Voice. DeKalb, Northern Illinois University Press, 1987.
(16) Scherer RC, Cooper D, Alipour-Haghighi F and Titze IR: Contact Pressure Between the Vocal Processes of an Excised Bovine Larynx. in Titze IR and Scherer RC (eds): Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control. Denver, The Denver Center for the Performing Arts, 1983, pp 292-303.
(17) Titze IR and Durham PL: Passive Mechanisms Influencing Fundamental Frequency Control in Baer T, Sasaki C and Harris K (eds): Laryngeal Function in Phonation and Respiration SanDiego, College-Hill Press, 1987, pp 304-319.
(18) Honda K: Variability Analysis of Laryngeal Muscle Activities. in Titze IR and Scherer RC (eds): Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control, Denver, The Denver Center for the Performing Arts, 1983, pp 127-137.
(19) Vennard W: Singing, The Mechanism and The Technique, Boston, Carl Fisher, 1967.
(20) Scherer RC Druker DG and Titze IR: Electroglottography and Direct Measurement of Vocal Fold Contact Area. in Fujimura O (ed): Vocal Physiology: Voice Production, Mechanics and Function, New York, Raven, 1988, pp 279-291.
(21) Childers DG and Krishnamurthy AK: A critical review of electroglottography. CRC Critical Review of Biomedical Engineering 1985; 12, 131-161.
(22)Titze IR: Parameterization of the glottal area, glottal flow, and vocal fold contact area. Journal of the Acoustical Society of America 1984; 75, 570-580.
(23) Rothenberg M and Mahstie JJ: Monitoring vocal fold abduction through vocal fold contact area. Journal of Speech and Hearing Research 1988; 31, 338-351.
(24) Scherer RC, Vail V and Rockwell B: Measures of laryngeal adduction. Journal of the Acoustical Society of America 1988; (Suppl 1) 84, 581.
(25) Yorkston KM, Beukelman DR and Bell KR: Clinical Management of Dysarthric Speakers, Boston, College-Hill, 1988, pp 258.
(26) Von Cramon D: Traumatic mutism and the subsequent reorganization of speech functions. Neuropsychologia 1981; 19: 801-805.
(27) Vogel M and Von Cramon D: Dysphonia after traumatic midbrain damage: a follow-up study. Folia Phoniatrica. 1982; 34:150-159.
(28) Tucker HM: Vocal cord paralysis-1979 : etiology and management. Laryngoscope 1980; 90:585-590.
(29) Ramig L Olson, Scherer RC, Burton E, Titze IR and Horii Y: Acoustic analysis of voice in amyotrophic lateral sclerosis: a longitudinal case study. Journal of Speech and Hearing Disorders (in press).
(30) Ward PH, Berci G and Calcaterra TC: Superior laryngeal nerve paralysis: an often overlooked entity. Trans. Am. Acad. Ophthalmol. Otolaryngol. 1977; 84:78-89.
(31) Neiman RF, Mountjoy JR and Allen EL: Myasthenia gravis focal to the larynx. Arch Otolaryngol 1975;101:569-570.

(32)Hanson DG, Gerratt BR and Ward PH: Cinegraphic observations of laryngeal function in Parkinson's disease. Laryngoscope 1984;94:348-353.
(33) Hirano M, Koike Y and von Leden H: Maximum phonation time and air usage during phonation. Clinical study. Folia Phoniat. 1968;20:185-201.
(34) Sapir S and Aronson AE: Aphonia after closed head injury: aetiologic considerations. British Journal of Communication Disorders 1985;20:289-296.
(35) Murphy AT: Functional Voice Disorders, ed. 2. Englewood Cliffs, Prentice Hall, Inc., 1965, pp16-20.
(36) Cooper M: Modern Techniques of Vocal Rehabilitation. Springfield, Ill, Thomas, 1977.
(37)Boone DR and McFarlane SC: The Voice and VoiceTherapy, ed.4. Englewood Cliffs, Prentice Hall, Inc.,1988.
(38) Froeschels E, Kastein S and Weiss D : A method of therapy for paralytic conditions of the mechanisms of phonation respiration and glutination. Journal of Speech and Hearing Disorders 1955; 20/4 :365-370.
(39) Stemple JC: Clinical Voice Pathology: Theory and Management, Columbus, Charles E. Merrill, 1984, pp 135-137.
(40) Ramig LA Mead CL Scherer RC Horii Y Larson K and Kohler D: Voice Therapy and Parkinson's disease: a longitudinal study of efficacy. A paper presented at the Clinical Dysarthria Conference, San Diego, 1988.
(41) Ramig LA Mead CL and DeSanto L: Voice therapy and Parkinson's disease. Journal of the American Speech Hearing Association 1988;30/10:128.
(42) Ramig LA Mead CL and Winholtz W: Speech therapy: Parkinson's disease. A video tape of speech treatment for patients with Parkinson's disease. Produced by the Recording and Research Center of the Denver Center for the Performing Arts and The Lee Silverman Center for Parkinson's, 1988.
(43) Linebaugh CW: Treatment of Flaccid Dysarthria. in Perkins WH (ed): Current Therapy of Communication Disorders: Dysarthria and Apraxia, New York, Thieme, 1983, pp 59-67.
(44) McFarlane SC and Lavorato AS: The use of video endoscopy in the evaluation and treatment of dysphonia. Communicative Disorders 1984; 9:117-126.
(45) Bastian R: Laryngeal image feedback for voice disorder patients. Journal of Voice 1987; 1/3: 279-282.
(46) Netsell R and Daniel B: Dysarthria in adults: physiologic approach to rehabilitation. Archives of Physical Medicine and Rehabilitation. 1979, 60:502-508.
(47) Collins M, Rosenbek J and Donahue E: The effects of posture on speech in ataxic dysarthria. Journal of the American Speech Hearing Association 1982; 24:767.
(48) Putnam A and Hixon T: Respiratory kinematics in speakers with motor neuron disease. in McNeil M, Rosenbek J and Aronson A (eds): The Dysarthrias, SanDiego: College Hill Press, 1985.
(49) Porter PB, Wurth B and Stowers S: Seating and Positioning for Communication, in Yoder DE and Kent RD (eds): Decision Making in Speech-Language Pathology. Philadelphia, Decker, 1988, pp 186-187.
(50) Hixon TJ: Respiratory-laryngeal evaluation. Presented to the Veterans Administration Workshop on Motor Speech Disorders. Madison, Wis.,1975.
(51) Hardy J: Cerebral Palsy. Englewood Cliffs, NJ, Prentice Hall, 1983.
(52) Netsell R and Hixon TJ: A noninvasive method for clinically estimating subglottal air pressure. Journal of Speech and Hearing Disorders 1978;43:326-330.

(53) Hixon TJ, Hawley JL and Wilson JL: An around-the-house device for the clinical determination of respiratory driving pressure: a note on making simle even simpler. Journal of Speech and Hearing Disorders 1982; 47:413-415.
(54)Hixon TJ: Speech Breathing Kinematics and Mechanism Interferences Therefrom. in Grillner S, Lindblom B and Lubker J (eds.): Speech Motor Control. New York: Pergamon, 1982.
(55) Rubow R: Role of Feedback Reinforcement, and Compliance on Training and Transfer in Biofeedback-Based Rehabilitation of Motor Speech Disorders. McNeil MR, Rosenbek JC, Aronson AE, (eds) The Dysarthrias : physiology, acoustics, perception and management. SanDiego:College-Hill Press, 1984; pp 207-230.
(56) Putnam AHB: Respiratory Dysfunction:Management, in Yoder DE and Kent RD (eds): Decision Making in Speech-Language Pathology. Philadelphia, Decker, 1988, pp128-131.
(57)Kent RD, Kent JF and Rosenbek JC: Maximum performance tests of speech production. Journal of Speech and Hearing Disorders 1987; 52: 367-387.
(58)Rubow RT and Swift E: Microcomputer-based wearable biofeedback device to improve treatment carry-over in parkinsonian dysarthria. Journal of Speech and Hearing Disorders 1985;50:178-185.
(59) Prater RJ and Swift RW: Manual of Voice Therapy, Boston, Little Brown and Company, 1984.
(60) Yorkston KM and Beukelman DR: Ataxic dysarthria: treatment sequences based on intelligibility and prosodic considerations. Journal of Speech and Hearing Disorders 1981; 46:398-404.
(61) Netsell R: Velopharyngeal Dysfunction, in Yoder DE and Kent RD (eds): Decision Making in Speech-Language Pathology. Philadelphia, Decker,1988, pp 150-151.
(62) Johns DF: Surgical and Prosthetic Management of Neurogenic Velopharyngeal Incompentency in Dysarthria, in Johns DF (ed): Clinical Management of Neurogenic Communicative Disorders, Boston, Little, Brown and Company, 1985.
(63) Isshiki N, Tanada M and Sawada M: Arytenoid adduction for unilateral vocal cord paralysis. Archives of Otolaryngology 1978;104:555-558.
(64) Tucker HM: Reinnervation of the unilaterally paralyzed larynx. Annals Oto-Rhino-Laryngo 1977; 86:789-794.
(65) Dedo DD and Dedo HH: Vocal cord paralysis. Otolaryngology, Volume 3 Head and Neck. Philadelphia, W.B. Saunders, 1980.
(66) Woodson GE and Miller RH: The timing of surgical intervention of vocal cord paralysis. Otolaryngology Head and Neck Surgery 1981;89:264-267.
(67) Tucker HM and Rusnov M: Laryngeal reinnervation for unilateral vocal cord paralysis: long-term results. Annals of Otology, Rhinology and Laryngology 1981; 90:457-459.
(68) Weber RS: Clinical restoration of voice function after loss of the vagus nerve. Head and Neck Surgery 1985; 7: 448-457.
(69) Facs MM and Beery Q: Muscle-nerve pedicle laryngeal reinnervation. Laryngoscope 1986; 96:1196-1200.
(70) Koufman, TA: Laryngoplasty for vocal cord medialization: an alternative to Teflon. Laryngoscope 1986; 96/7;726-731.
(71) Ford CN, Martin DW, Warner TF: Injectable collagen in laryngeal rehabilitation. Laryngoscope 1984; 94:513-518.

(72) Arnold GE: Vocal rehabilitation of paralytic dysphonia. Archives of Otolaryngology 1962; 76/363:88-83.
(73) Cooper M and Cooper MH (eds): Approaches to Vocal Rehabilitation, Springfield, Illinois: Charles C Thomas, 1977.
(74) Bull T and Cook J: Speech Therapy and ENT Surgery. Oxford, Blackwell Scientific Publications, 1976.
(75) Rontal M Rontal E Leuchter W and Rolnick M: Voice spectrography in the evaluation of myasthenia gravis of the larynx. Annals of Otology, Rhinology and Laryngology 1978; 87:722-728.
(76) Walshe, TM: Disease of Nerve and Muscle. in Samuels MA (ed) Manual of Neurologic Therapeutics. Boston, Little and Brown, 1982.
(77) Davis P, Boone D, Carroll RL, Darvenzia P and Harrison G: Adductor spastic dysphonia: heterogeneity of physiological and phonatory characteristics. Archives of Otolaryngology 1987; 97:179-185.
(78) Portnoy RA: Hyperkinetic dysarthria as an early indicator of impending tardive dyskinesia. Journal of Speech and Hearing Disorders 1979; 44:214-219.
(79) Gerratt BR: Formant frequency fluctuation as an index of motor steadiness in the vocal tract. Journal of Speech and Hearing Research 1983; 26: 297-304.
(80) Shipp T, Mueller P and Zwitman D: Intermittent abductory dysphonia. Journal of Speech and Hearing Disorders 1980; 45: 281.
(81)Von Leden, H: Objective measures of laryngeal function and phonation. Ann NY Acad Sci 1968;155:56-67.
(82) Kammermeier MA: A Comparison of Phonatory Phenomena Among Groups of Neurologically Impaired Speakers. Doctoral Dissertation. Minneapolis, University of Minnesota, 1969.
(83) Smitheran J and Hixon T: A clinical method for estimating laryngeal airway resistance during vowel production. Journal of Speech and Hearing Research 1981; 46: 138-146.
(84) Jacobson E: Progressive Relaxation, ed. 2. Chicago, University of Chicago Press, 1938.
(85) Jacobson E: Anxiety and Tension Control, A Physiologic Approach. Philadelphia, J.B. Lippincott Co., 1964.
(86) Jacobson E: You Must Relax, ed.5. New York, McGraw Hill Book Co., 1976.
(87) McClosky DG: General Techniques and Specific Procedures for Certain Voice Problems. in Cooper M and Cooper MH (eds): Approaches to Vocal Rehabilitation. Springfield, Ill., Charles C. Thomas, 1977, pp 138-152.
(88) Rubow RT, Rosenbek JC, Collins MJ and Celesia GG: Reduction of hemifacial spasm and dysarthria following EMG feedback. Journal of Speech and Hearing Disorders 1984; 49:26-33.
(89)Netsell R and Cleeland CS: Modification of lip hypertonia in dysarthria using EMG feedback. Journal of Speech and Hearing Disorders 1973; 38: 131-140.
(90) Prosek RA, Montogomery AA, Walden BE and Schwartz DM: EMG biofeedback in the treatment of hyperfunctional voice disorders. Journal of Speech and Hearing Disorders 1978;53:282-294.
(91) Stemple J, Weiler E, Whitehead W and Komray R: Electromyographic biofeedback training with patients exhibiting a hyperfunctional voice disorder. Laryngoscope 1980;90: 471-475.
(92)Froeschels E: Chewing method as therapy. Arch Otolaryngol 1952; 56: 427-434.

(93)Pershall KE and Boone DR: A video-endoscopic and computerized tomographic study of hypopharyngeal and supraglottic activity during assorted vocal tasks. Transcripts of the Fourteenth Symposium: Care of the Professional Voice, New York, The Voice Foundation, 1986. (94)Boone D: The Voice and Voice Therapy, ed.3.Englewood Cliffs, NJ, Prentice Hall, 1983.
(95)Aten JL: Treatment of Spastic Dysarthria . in Perkins WH (ed): Current Therapy of Communication Disorders: Dysarthria and Apraxia, New York, Thieme, 1983, pp 69-78 .
(96)Dedo HH: Recurrent laryngeal nerve section for spastic dysphonia. Annals of Otology, Rhinology and Laryngology 1979;85:451-459.
(97)Blitzer A, Brin M, Fahn S and Lovelace R: Localized injections of botulinum toxin for the treatment of focal laryngeal dystonia (spastic dysphonia). Laryngoscope 1988;98:193-197.
(98)Freeman FJ, Cannito MP and Finitzo-Hieber T: Getting to know spasmodic dsyphonia patients. Journal of the Texas Speech and Hearing Association 1985:14-19.
(99)Dedo H and Shipp T: Spastic Dysphonia, Houston, College-Hill Press, 1980.
(100)Beukelman D : Treatment of Hyperkinetic Dysarthria, in Perkins WH (ed): Current Therapy of Communication Disorders: Dysarthria and Apraxia, New York, Thieme, 1983, pp 101-104 .
(101) Ward PM, Hanson DG, Garrett B and Berke GS: Current and future horizons in laryngeal and voice research. Annals of Otology, Rhinology and Laryngology 1989;98: 145-152.
(102)Isshiki N, Tanabe M, Ishizaka K and Broad D: Clinical significance of asymmetrical vocal cord tension. Annals of Otology, Rhinology and Laryngology 1977;86:58-66.
(103) Baer T: Investigation of the Phonatory Mechanism. in Ludlow CL and Hart MO (eds): Proceedings of the Conference on the Assessment of Vocal Pathology ; AHSA Report 11. Rockville, American Speech-Language-Hearing Association, 1981, pp 38-47.
(104) Larson CR, Kempster GB, Kistler MK (eds): Changes in voice fundamental frequency following discharge of single motor units in cricothyroid and thyroarytenoid muscles. Journal of Speech and Hearing Research 1987; 30:552-558.
(105) Titze IR : Neurologic sources of aperiodicity in phonation. J Acoustical of America (in submission).
(106)Moore P and VonLeden H: Dynamic variations of the vibratory patterns in the normal larynx. Folia Phoniatrica 1958; 10:205-238.
(107) VonLeden H, Moore P and Timke R: Laryngeal vibrations:measurements of the glottal wave. Part III. The pathologic larynx. Archives of Otolaryngology 1960;71/44:16-35.
(108) Horiguchi S, Haji T , Baer T and Gould WJ: Comparison of Electroglottographic and Acoustic Waveform Perturbation Measures. in Baer T, Sasaki C, and Harris K (eds): Laryngeal Function in Phonation and Respiration. Boston, College-Hill Publication, Little, Brown and Company, 1987, pp 509-518.
(109) Orlikoff RF and Baken RJ: Vocal jitter at different fundamental frequencies. A paper presented at the Seventeenth Symposium: Care of the Professional Voice, New York City, 1988
(110) Garrett BR, Hanson DG and Berke GS: Glottographic Measures of Laryngeal Function in Individuals with Abnormal Motor Control. in Baer T, Sasaki C, and Harris K (eds): Laryngeal Function in Phonation and Respiration. Boston, College-Hill Publication, Little, Brown and Company, 1987, pp 521-532.
(111) Koike Y: Vowel amplitude modulations in patients with laryngeal diseases. Journal of the Acoustical Society of America 1969; 45:839-844.
(112) Pinto NB and Titze IR: Unification of perturbation measures in speech signals. Journal of the Acoustical Society of America, in submission.

(113) Yumoto E, Gould WJ and Baer T: Harmonics-to-noise ratio as an index of the degree of hoarseness. Journal of the Acoustical Society of America 1982; 71/6:1544-1550.
(114) Wendahl R: Laryngeal analog synthesis of harsh voice quality. Folia Phoniatrica 1963; 15: 241-250.
(115) Wendahl R: Laryngeal analog synthesis of jitter and shimmer: auditory parameters of harshness. Folia Phoniatrica 1966; 18: 98-108.
(116) Heiberger V and Horii Y: Jitter and Shimmer in Sustained Phonation, in Lass N (ed) Speech and Language Advances in Basic Research and Practice, New York, Academic Press, 1982, pp 299-332.
(117) Ludlow CL, Bassich CJ, Connor, NP, Coulter DC and Lee YJ: The Validity of Using Phonatory Jitter and Shimmer to Detect Laryngeal Pathology. in Baer T, Sasaki C, and Harris K (eds): Laryngeal Function in Phonation and Respiration. Boston, College-Hill Publication, Little, Brown and Company, 1987, pp 492-508.
(118) Ramig LA and Ringel RL: The effects of physiological aging on selected acoustic characteristics of voice. Journal of Speech and Hearing Research 1983; 26/1:22-30.
(119) Wilcox K and Horii Y: Age and changes in vocal jitter. Journal of Gerontology 1980;35:194-198.
(120) Linville SE: Intraspeaker variability in fundamental frequency stability : an age-related phenomenon? Journal of the Acoustical Society of America 1988; 83/2:741-745.
(121) Horii Y: Jitter and shimmer differences among sustained vowel phonations. Journal of Speech and Hearing Research 1982; 25: 12-14.
(122) Ward PH, Hansen DG and Berci G: Observations on central neurologic etiology for laryngeal dysfunction. Annals of Otology Rhinology and Laryngology 1981;90:430-441.
(123) Desmedt J: Physiological tremor, pathological tremor and gradation of muscle force. Progress in Clinical Neurophysiology. Basel:S.Karger, 1979.
(124) Ward PH, Hanson DG and Berci G: Photographic studies of the larynx in central laryngeal paresis and paralysis. Acta Otolaryngol 1981; 91:353-367.
(125) Ramig LA and Shipp T: Comparative measures of vocal tremor and vocal vibrato. The Journal of Voice 1987; 1/2:162-167.
(126) Yorkston KM and Beukelman DR: Ataxic dysarthria: treatment sequences based on intelligibility and prosodic considerations. Journal of Speech and Hearing Disorders 1981;46:398-404.
(127) Titze I: Neurologic sources of aperiodicity in phonation. J Acoustical Society of America (in submission).
(128) Ludlow C, Coulter D and Gentges F: The Differential Sensitivity of Frequency Perturbation to Laryngeal Neoplasms and Neuropathologies. in Bless, DM and Abbs, JH (eds) Vocal Fold Physiology: Contemporary Research and Clinical Issues, SanDiego, College Hill Press, 1983, 381-392.
(129) Case J: Neurogenic Voice Disorders. in Case J : Voice Disorders, Rockville, Aspen,1984.
(130) Zemlin, WR: Speech and Hearing Science: Anatomy and Physiology ed.2. Englewood Cliffs, NJ: Prentice Hall, 1981.
(131) Garrett BR, Precoda K and Hansen DG: Diplophonia: features in the time domain. Journal of the American Speech Language Hearing Association 1987; 29: 91A.
(132) Greene MCL: The Voice and Its Disorders, ed.4. Philadelphia:J.B.Lippincott Company, 1980.
(133) Bonitati CL: Personal communication.

(134) D'antonio L, Lotz W, Chait D, and Netsell R: Perceptual-physiologic approach to evaluation and treatment of dysphonia. Ann Otol Rhinol Laryngol 1987;96:187-190.
(135) Larson KK, Ramig LA and Scherer RC: Acoustic analysis of voice: the on-off effect of medication in Parkinson's disease. A paper presented at the Clinical Dysarthria Conference, San Diego, 1988.
(136) Schley WS, Fenton EF and Niimi S: Vocal symptoms in Parkinson's disease treated with levodopa: a case report. Annals of Otology Rhinology and Laryngology 1982; 91:119.
(137) Asperheim MK and and Eisenhauer LA: The Pharmacologic Basis of Patient Care, ed.3. Philadelphia, Saunders, 1977.
(138)Tolosa ES and Lowenson RB: Essential tremor: treatment with propranolol. Neurology 1975; 25: 1975.
(139) Cooper I: Involuntary Movement Disorders, New York, Harper and Row,1969.
(140) Sundberg J and Gauffin J: Waveform and Spectrum of the Glottal Voice Source. in Lindblom B and Ohman S (eds): Frontiers of Speech Communication Research: Festschriff for Gunnar Fant. New York, Academic Press, 1979, pp 301-320.
(141) Titze IR: Regulation of Vocal Power and Efficiency by Subglottal Pressure and Glottal Width, in Fujimura, O (ed): Vocal Physiology: Voice Production, Mechanisms and Functions. New York, Raven,1988, pp 347-356.
(142) Sawashima M and Hirose H: Laryngeal Gesture in Speech Production, in MacNeilage P (ed): The Production of Speech. New York, Springer-Verlag, 1983, pp 11-38.
(143) Hirose H and Niimi S: The Relationship Between Glottal Openning and the Transglottal Pressure Differences During Consonant Production, in Baer T, Sasaki C, and Harris K (eds): Laryngeal Function in Phonation and Respiration. Boston, College-Hill Publication, Little, Brown and Company, 1987, pp 381-390.
(144) Lofquist A and McCarr NS: Laryngeal Dynamics in Voiceless Consonant Production, in Baer T, Sasaki C, and Harris K (eds): Laryngeal Function in Phonation and Respiration. Boston, College-Hill Publication, Little, Brown and Company, 1987, pp 403-421.
(145) Collier R: Fo Declination: The Control of its Setting, Resetting and Slope, in Baer T, Sasaki C, and Harris K (eds): Laryngeal Function in Phonation and Respiration. Boston, College-Hill Publication, Little, Brown and Company, 1987, pp 403-421.
(146) Gilfer CE, Harris KS and Baer T: Controlled Variables in Sentence Intonation, in Baer T, Sasaki C, and Harris K (eds): Laryngeal Function in Phonation and Respiration. Boston, College-Hill Publication, Little, Brown and Company, 1987, pp 381-390.
(147) Hirose H, Kiritani S and Imagawa H: High-speed Digital Image Analysis of Laryngeal Behavior in Running Speech, in Fujimura, O (ed): Vocal Physiology: Voice Production, Mechanisms and Functions. New York, Raven,1988 pp 347-356.
(148) Fujisaki H : A Note on the Physiological and Physical Basis for the Phrase and Accent Components in the Voice Fundamental Frequency Contour, in Fujimura, O (ed): Vocal Physiology: Voice Production, Mechanisms and Functions. New York, Raven, 1988 pp 347-356.
(149) Kent RD: The Dysarthric or Apraxic Client, in Yoder DE and Kent RD (eds): Decision Making in Speech-Language Pathology. Philadelphia, Decker, 1988, pp 156-157.
(149)Gibbon D: Perspectives of Intonation Analysis. Forum linguisticus, Band 9. Bern:Herbert Lang., 1976.
(150) Kent RD and Rosenbek J: Prosodic disturbance and neurogenic lesions. Brain and Language 1982; 15:259-291.

(151) Monrad-Krohn GH: Dysprosody or altered "melody of language". Brain 1947;70:405-415.
(153)Darwin CJ: On the dynamic use of prosody in speech perception. Status Report of Speech Research. New York :Haskins Laboratories SR-42/43, 1975.
(154) Yorkston KM: Prosody in the Adult, in Yoder DE and Kent RD (eds): Decision Making in Speech-Language Pathology. Philadelphia, Decker, 1988, pp 146-147.
(155) Helm N: Management of palilalia with a pacing board. Journal of Speech and Hearing Disorders 1979;44:350-353.
(156) Netsell R and Rosenbek J: Treating the Dysarthrias, in Darby J (ed) Speech and Language Evaluation in Neurology: Adult Disorders, Orlando, Grune and Stratton, 1985: pp 363-392.
(157) Yorkston KM, Beukelman DR, Minifie FD and Sapir S, Assessment of stress patterning. in McNeil MR, Rosenbek JC, Aronson AE, (eds) The Dysarthrias : physiology, acoustics, perception and management. SanDiego:College-Hill Press, 1984, pp 131-162.
(158) Lieberman P: Intonation, Perception and Language, Cambridge, Ma, The MIT Press, 1967.
(159) Murry T: Treatment of Ataxia Dyasrthria, in Perkins WH (ed): Current Therapy of Communication Disorders: Dysarthria and Apraxia, New York, Thieme, 1983, pp 79-90.
(160) Lehiste I: Some acoustic characteristics of dysarthric speech. Bibliotheca Phonetica, Fasc.2.Basel:Karger, 1965.
(161) Weismer G: Articulatory Characteristics of Parkinsonian Dysarthria: Segmental and Phrase-Level Timing, Spirantization, and Glottal-Supraglottal Coordination, in McNeil MR, Rosenbek JC, Aronson AE, (eds) The Dysarthrias : physiology, acoustics, perception and management. San Diego, College-Hill Press, 1984, pp 101-130 .
(162) Silverman F: Communication Aids for the Speechless. Englewood Cliffs, N.J.:Prentice Hall, 1980.
(163) Green M and Watson BW: The value of speech amplification in Parkinson's disease patients. Folia Phoniatr. 1968; 20: 250.
(164) Rubow RT and Swift E: Microcomputer-based wearable biofeedback device to improve treatment carry-over in parkinsonian dysarthria. Journal of Speech and Hearing Disorders 1985; 50 : 178-185.
(165) Blom E: Artificial larynx: past and present. in Salmon S and Goldstein L (eds) Artificial Larynx Handbook. New York, Greene and Stratton, 1978.
(166) Beukelnam DR and Garrett K (1988). Augmentative Communication for adults with aquired severe communication disorders. Augmentative and Alternative Communication. 4, 104-121.
(167) Beukelman D, Yorkston K and Dowden P: Communication Augmentation: A Case Book of Clinical Management, San Diego, College-Hill Press, 1985.
(168) Vanderheiden and LLoyd, 1986
(169) Berstein LE : The Vocally Impaired : Clinical Practice and Research. Philadelphia, Grune and Stratton, 1988.
(170) Kent RD: The Dysarthric or Apraxic Client,in Yoder DE and Kent RD (eds): Decision Making in Speech-Language Pathology. Philadelphia, Decker,1988, pp 156-57.
(171) Netsell R: Physiologic studies of dysarthria and their relevance to treatment. in Rosenbek JC (ed): Current Views of Dysarthria: Nature, Assessment and Treatment, Seminars in Speech and Language. New York,Thieme-Stratton,1984, pp 279-291.
(172) Berry WR and Sanders SB: in Berry WR (ed): Clinical Dysarthria, SanDiego, College Hill Press, 1983, pp 57-68.

(173) Logemann JA, Fisher HB, Boshes B and Blonsky ER: Frequency and co-occurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and Hearing Disorders 1978;43:47-57.

NCVS Status and Progress Report - 1
June 1991, 191-221

The Role of Phonation in Speech Intelligibility: A Review and Preliminary Data From Patients with Parkinson's Disease

Lorraine Olson Ramig, Ph.D., CCC-SP
University of Colorado-Boulder
Denver Center for the Performing Arts

Introduction

The role of phonation in the production of intelligible speech should be considered at various levels. The larynx should be studied for its role, in combination with the respiratory system, as the aerodynamic and acoustic source for speech production, as a generator of suprasegmental features, and as an articulator. Source characteristics of loudness and quality are dependent upon the respiratory system and vocal fold adductory and oscillatory systems. Suprasegmental features of pitch and prosody are dependent upon simultaneous modulations of source and articulatory characteristics. Laryngeal articulatory characteristics are dependent upon adductory gestures coordinated with sub- and supraglottal events to achieve segmental contrasts.

The vocal folds valve the respiratory air flow to generate the acoustic voice source. If vocal fold adduction (medial compression) or subglottal air pressure are insufficient, vocal loudness or intensity may be significantly reduced or intraoral air pressure may be inadequate for supraglottal valving. If the oscillations of the vocal folds are irregular or occur inconsistently, the voice source may be transient and voice quality may sound disordered. Respiratory, laryngeal and articulatory function must be coordinated to generate appropriate suprasegmental variations in pitch, loudness, duration and quality; disturbances of any of these variables may alter utterance meaning. If vocal fold adductory/abductory activity is inadequate, or not coordinated with sub and supraglottal events, the voice/voiceless articulatory contrast may be disturbed. All of these phonatory disruptions have the potential to reduce speech intelligibility.

A number of studies have reported a relationship between these phonatory disruptions and speech intelligibility (e.g. McGarr and Osberger, 1977; Parkhurst and Levitt, 1978; Monsen, 1983). The suprasegmental and laryngeal articulatory variables of pitch, prosody and voicing distinctions, together with their acoustic correlates, have been studied in individuals with hearing impairment,

alaryngeal voice and motor speech disorders, and have been related to reductions in intelligibility in those populations (e.g., McGarr and Osberger, 1978; Yorkston and Beukelman, 1981). However, it is the relationship between segmental variables and speech intelligibility that has been given the greatest attention in the literature (e.g., Tikofsky, 1970; Tikofsky and Tikofsky, 1964; Darley, Aronson and Brown, 1969 a,b; Smith, 1972; Yorkston and Beukelman, 1981; Weismer, Kent, Hodge and Martin, 1988; Samar and Metz, 1988; Freyman and Nerbonne, 1989). In fact, one comes away from a review of the disordered speech literature with the impression that the role of phonatory characteristics in intelligible speech production is relatively limited when compared with segmental contributions. Freyman and Nerbonne (1989) recently concluded that a significant percentage of the information contributing to the intelligibility of speech is conveyed by consonants. Carney (1986) concluded that of three "speaker-oriented" variables (segmental production, suprasegmental production and hearing loss), segmental production appeared to correlate most strongly with speech intelligibility. Lessac (1967) stated that it is consonants that make the spoken word intelligible.

A closer look at this research reveals that in most of the studies of disordered speech and intelligibility, the significant support roles of the laryngeal and respiratory systems as the aerodynamic and acoustic source for speech production have been simply assumed. Generally, intelligibility has been evaluated from samples with reasonable audibility, suggesting that an adequate source was used for speech production. Furthermore, in a number of cases, samples have been equated for audibility (Freyman and Nerbonne, 1989) and consonant identification has been the only task for the listener (e.g., Owens and Schubert, 1977; Resnick, Dubno, Hoffnung and Levitt, 1975). In certain cases, the origin of segmental breakdowns may in fact be source-related (inadequate intraoral air pressure for supraglottal valving; Pickett,1956; Arkebauer, Hixon and Hardy, 1967), yet the primary focus has been on the segmental manifestation. This issue has been addressed in part when Stevens, Nickerson and Rollins (1983) suggested the need to consider the contribution of the 'postural aspects' of speech production to intelligibility. They suggest that certain aspects of segmental and suprasegmental characteristics depend upon the overall posture that is assumed by the laryngeal and respiratory structures in preparation for speech production. For example, they point out that breathy voice quality or reduced amplitude between syllables, such as observed in speech of the deaf, result from vocal fold positioning, i.e., too wide a glottal opening or the insertion of glottal closure between syllables (Hudgins, 1937; Bernstein, Rollins and Stevens, 1978). Stevens et al. (1983) assume that posture includes not only some static state of readiness for speaking, but also the range over which the configurations or states of various structures are allowed to vary. It should be noted that a consistent reduction in intelligibility has been observed in speakers with clearly inadequate source characteristics such as severely limited respiratory support or vocal fold adduction (hypoadduction or hyperadduction) accompanying, for example, neurological disease or alaryngeal voice (Kalb and Carpenter, 1981; Simpson, Till and Goff, 1988).

The critical role of source characteristics in intelligible speech production is highlighted further when one reviews the literature on speech enhancement. Various acoustic characteristics have been studied when normal and disordered speakers have attempted to improve their intelligibility or when instrumental methods have been used to enhance the understandibility of speech (Tolhurst, 1957; Pichney, 1981; Pichney, Durlach and Briada, 1985; 1986; Spector, Subtelny, Whitehead and Wirz, 1979). Speech enhancement may occur after treatment (Till and Toye, 1988), when speech is produced with background noise (Lane and Tranel, 1971; Summers, Pisoni,

Bernacki, Pedlow and Stokes, 1988; Sundberg, Ternstrom, Perkins and Gramming, 1988), or in a performance mode (Ternstrom and Sundberg 1986), or when instrumental procedures are used to modify a signal to enhance its understandability (Lim, Oppenheim and Braida, 1978; Allen, Strong and Palmer, 1981; Montgomery and Edge, 1988). The speech enhancement literature suggests that the laryngeal source characteristics of intensity and spectral balance are among the phonatory variables that are consistently modified when a speaker attempts to improve intelligibility.

The following section will review phonatory source, suprasegmental and articulatory contributions to speech intelligibility as presented in the hearing impaired, alaryngeal, motor disorders and speech enhancement literature.

The relationship between intelligibility and phonatory characteristics in disordered and enhanced speech

It is the literature on hearing-impaired and deaf speakers that has addressed most explicitly the relationship between characteristics of phonation and speech intelligibility (e.g. McGarr and Osberger, 1978). The other primary sources for data on this issue come from the literature on alaryngeal voice, motor speech disorders, and speech enhancement. It is difficult to find any association between speech intelligibility and phonation in the laryngeal-based voice disorders literature, although the professional/performing voice literature suggests various modifications in voice characteristics for improvement in intelligibility (Berry, 1973; Turner,1977).

Contributions of the SOURCE functions of the larynx to speech intelligibility

Loudness

The role of loudness in speech intelligibility has been well established. The early work of Draegert (1951), Moore (1946), Curtis (1946) and Pickett (1956) concluded that greater intelligibility accompanies speech signals of greater intensity or force. Pickett (1956) reported drastic deterioration of intelligibility with low and high extremes of vocal effort. As loudness is raised in any voice, there is increased energy in the high part of the spectrum (spectral tilt decreases), which has been associated with improvements in the clarity of speech (Ternstrom and Sundberg, 1986). One technique used by performers which enables them to be heard over loud background noise is the "singer's formant". Performers use vocal tract modifications to shape the source characteristics for increased intensity which results in a spectrum envelop peak near 3KHz (Sundberg, 1977; Cleveland and Sundberg, 1983; Sundberg and Ternstrom, 1986).

A number of disorders of phonation result in the inability to generate adequate loudness. Breakdowns occur for functional reasons (e.g., voice disorders in the hearing impaired and deaf) and organic causes (e.g., laryngeal removal or neurolaryngeal disorders). The underlying physiologic basis may be that these individuals cannot create a sufficient air pressure difference to generate an acoustic signal of adequate amplitude due to insufficient respiratory support or inadequate or excessive adduction (hypoadduction or hyperadduction). For example, Bernstein, Rollins and Stevens (1978) reported that the tendency of deaf speakers to insert glottal closure between syllables and words creates a reduction in sound amplitude between two syllables and thus causes reduced amplitude and reduced intelligibility. Levitt, Smith and Stromberg (1974) reported intermittent phonation resulting in spasmodic and excess variation in loudness which correlated with

reduced intelligibility. Smith (1975) found that errors involving poor phonatory control (intermittent phonation, spasmodic variations of pitch and loudness, and excessive variability of intonation) were highly correlated with reductions in intelligibility.

Reduced loudness has been well documented in alaryngeal speakers using esophageal speech (e.g. Drummond, 1965; Hyman, 1955; Diedrich, 1968) and has been associated with reductions in speech intelligibility observed in that patient population (Kalb and Carpenter, 1981). Recently, the tracheoesophageal shunt procedure has allowed these patients to use respiratory air rather than esophageal air to power voice production. Data support consistent increases in intensity as well as intelligibility with the use of this shunt (Robbins, 1984; Robbins, Fisher, Blom and Singer, 1984; Tardy-Mitzell, Andrews and Bowman, 1985; Pinzola and Cain, 1988; Doyle, Danhauer and Reed, 1988). These findings can be interpreted as support for the importance of the aerodynamic source in the intelligibility of tracheoesophageal voice production .

Reports indicate that reduced loudness is a deterrent to oral communication in patients with motor speech disorders (e.g. Rosenbek and LaPointe, 1985; Aronson, 1985; Simpson, Till and Goff, 1988). Respiratory as well as laryngeal adductory insufficiencies have been suggested as physiologic bases for these loudness reductions (Simpson, Till and Goff, 1988). For example, reduced intensity, monoloudness, weak overall effort and reduced intraoral pressures have been reported in parkinsonian speech (Canter, 1963, 1965; Mueller, 1971; Kent and Rosenbek, 1982; Ludlow and Bassich, 1983). Therapy techniques have focused on improving oral communication by enhancing respiratory support for speech and vocal fold adduction (Rosenbek and LaPointe, 1985; Ramig, Fazoli, Scherer and Bonitati, 1990). In addition, amplification systems are frequently recommended for motor disordered patients with significantly reduced loudness (Greene and Watson, 1968; Rubow and Strand, 1985) and have been associated with increases in intelligibility.

Increases in loudness or intensity are frequently reported when speakers attempt to improve their intelligibility in the presence of background noise (Hanley and Steer, 1949; Draegert, 1951; Tolhurst, 1954; 1955; Lane et al, 1970; Summers, Pisoni, Bernacki, Pedlow and Stokes,1988; Sundberg, Ternstrom, Perkins and Gramming, 1988), when asked to speak more clearly (Pichney, Durlach and Braida, 1986) when asked to stress or emphasize (Lieberman,1960; Klatt,1975; Cooper, Eady and Mueller,1985), or when asked to speak in low-redundancy contexts (Hunnicutt, 1985; 1987). For example, Garber, Siegel and Pick (1980) reported that in low-pass filtering conditions, subjects increased their intelligibility by increasing their vocal intensity rather than by any articulatory changes. Summers et al. (1988) reported that speakers modify both the prosodic and segmental acoustic-phonetic properties of their speech when they talk in noise. They speak louder and slower, raise pitch and introduce changes in the short-term power spectrum of voiced segments. Similar changes have been measured when subjects have been asked to stress or speak clearly. For example, Pichney, Durlach and Braida (1986) noted substantial differences (5-6 dB) in intensities of stop consonants and higher fundamental frequency values in acoustic analyses of "clear" versus conversational speech. In shouted speech, increases in fundamental frequency, vowel duration, the frequency of format one and a reduction in spectral tilt have been reported (Rostolland and Parant, 1974; Rostolland, 1982 a,b). However, because the magnitude of these changes is much greater in shouted speech and extreme articulations accompany shouting, speech is less intelligible. Thus, Rostolland (1982 a,b) concluded that clear speech cannot be obtained by merely boosting one's overall vocal output.

ment is the consonant to vowel (C/V) ratio, that is the ratio of the power of a consonant to that of the nearest vowel in the same syllable (Hecker, 1974; House, Williams, Hecker and Kryter, 1965; Montogomery and Edge,1988). For example, Hecker (1974) increased the C/V ratio of a low intelligibility talker and produced a 3.75% improvement in intelligibility. Ono, Okasaki, Nakai and Harasaki (1982) and Montogomery and Edge (1988) using hearing impaired listeners and Gordon-Salant (1986; 1987) with both normal and hearing impaired listeners reported improvements in intelligibility for monosyllables as a result of increases in C/V ratio. When Montogomery and Edge(1988) increased consonant intensity while holding the vowel constant at low presentation levels, they found that increasing amplitude produced significant improvements in intelligibility over unprocessed speech. He concluded that the 10% to 12% improvement produced by manipulating consonant amplitude was a demonstration of the influence that even modest increases in consonant intensity can exert on intelligibility and the importance of the C/V ratio in speech recognition. Recently, Freyman and Nerbonne (1989) reported that the degree to which variations in speech intelligibility could be explained by variations in C/V ratio was found to be quite different for different consonants. For example, for /s,*(sh), *(ch)/ the C/V ratio accounted for a great deal of the variation in intelligibility; for voiceless stops, this was not the case. They suggested that this difference may be due to "the relative inaudibility of stops when speech-to-noise ratios are poor" (p.32) and support the recommendation by Turner and Robb (1987) that the audibility of consonants should be considered when analyzing speech recognition data. They concluded that when stimuli are calibrated according to vowel intensity, there is no evidence that the ratio between consonant and vowels is important for intelligibility.

In summary, it appears that generation of an adequate acoustic and aerodynamic source is a critical factor in intelligible speech production. Without sufficient vocal loudness or aerodynamic pressures, supraglottal valving for segmental production would be limited and intelligibility reduced.

Quality

Voice quality has been reported to be a source of information about physical, psychological and social characteristics of the speaker as well as playing a "vital semiotic role" in spoken interaction (Laver, 1968; Laver and Trudgill, 1979). Kohler and Dommelen (1987) suggested that the different overall human voice qualities "tense", "neutral" and "modal" may have prosodic effects on sound perception. Pittam (1987) reported that the voice qualities of "breathy", "creaky", "nasal", "tense" and "whispery" all have been reported to function communicatively (Addington, 1968; Esling, 1978; Laver, 1980; Trudgill, 1974; Scherer, 1979).

Various approaches have been used to study vocal quality acoustically and perceptually. Numerous attempts have been made to relate qualities such as "hoarseness" and "harshness" to various short-term (cycle-to-cycle) (Scherer et al, 1988) acoustic measures including jitter, shimmer and harmonics-to-noise ratio (e.g. Deal and Emanuel, 1978; Takahashi and Koike, 1975; Yumoto, Gould and Baer, 1982). In contrast, Laver (1980) and Laver, Wirz, Mackenzie and Hiller (1981) applied principles of phonetic analysis to vocal quality (similar to the concept of an articulatory setting) and suggested that the vocal tract adopts long-term muscular adjustments that underlie and act as contraints on short-term articulations. This idea appears compatible with the 'postural aspects' concept of Stevens et al (1983) discussed earlier. Based upon Laver's suggestion, Pittman (1987a,b) emphasized that voice quality is a long-term phenomenon, present all the time a person is

He used long-term spectral analysis to discriminate among the voice qualities 'breathy', 'creaky', 'tense' and 'whispery'.

Disorders of voice quality have been reported in deaf and hard of hearing individuals. Spector, Subtelny, Whitehead and Wirz (1979) reported that eleven percent of deaf individuals had voices that were harsh and excessively tense. They associated this tension with an abrupt initiation of voicing, faulty modulation of the air stream for consonant production and/or inefficient control of air expenditure and suggested that these characteristics may adversely affect vocal pitch, pitch control, loudness and speech intelligibility. McGarr and Osberger(1978) reported that children who could not sustain phonation and whose speech contained numerous pitch breaks were judged to have unintelligible speech. McGarr and Osberger (1978) concluded that there is a relationship between poor phonatory control, on the one hand, and hearing level and intelligibility, on the other. Monsen (1983) states that in the deaf, voice quality and speech intelligibility are intricately intertwined; with the relationship between voice quality and intelligibility in the deaf being quite high.

When voice quality disorders are less severe, the effect on intelligibility may be reduced. Samar and Metz (1988) reported that breathiness did not degrade the message intelligibility of speech produced by hearing impaired speakers under normal listening conditions. They interpreted this to be related to a write down procedure of intelligibility assessment in which listeners report their linguistic perceptual experience, regardless of their perceptions of vocal aberrations such as breathiness. Monsen (1983) says rarely is the voice quality of a normal hearing speaker deviant enough to impair intelligibility under good listening conditions; he suggests that in a normal hearing talker, peculiar voice quality may be considered apart from how well he is understood.

Disordered voice quality has been associated with reduced intelligibility and acceptability in alaryngeal speakers. Doyle, Danhauer and Reed (1988) associated low intelligibility in an esophageal speaker with rough voice quality. Niebor, Graaf and Schutte (1988) concluded that the quality of voice was better for patients with the tracheoesophageal puncture than esophageal speech. This improvement in quality has been associated with the addition of pulmonary air which improves listeners' perception of smoothness and acceptability (Pindzola and Cain, 1988). Relaxation of the pharyngoesophageal segment (pseudoglottis) by a myotomy has been associated with improved voice quality in tracheoesophageal speakers (Singer and Blom, 1981; Chodosh, Giancarlo and Goldstein, 1984). Weinberg and Bennett (1973) concluded that because of its more normal voice quality, the Toyko larynx was more "acceptable" than superior esophageal speech, Western Electric Reed and the Bell Electrolarynx.

Disordered voice quality has been reported in speakers with motor speech disorders such as flaccid dysarthria, pseudobulbar palsy and hypokinetic (Parkinsonian) dysarthria (Aronson, 1985). Breathiness, hoarseness, harshness, strained-strangled voice, wet and gurgly quality and tremor have been reported in these populations (Aronson, 1985; Boone and McFarland, 1988). Despite the clinical impression that these disordered qualities reduce speech intelligibility, this relationship has not been experimentally documented.

Disordered voice quality has been a focus of treatment techniques in the areas of deaf speech and motor speech disorders as well as the disordered voice literature. However, only in the area of deaf speech has an attempt been made to relate treatment related changes in voice quality to speech intelligibility. Spector et al. (1979) developed a program to reduce harsh/tense voice quality in adult deaf speakers. Perceptual judgments of pre- and post-training recordings revealed significant improvements in vocal tension, pitch register, pitch control, articulation and speech intelligibil-

ity. Boone and McFarland (1988) suggested that voice therapy can often improve the intelligibility of dysarthric speakers and various techniques for modification of disordered vocal quality in dysarthric speakers have been reviewed by Prater and Swift (1984) and Ramig and Scherer (1989). However it appears that few of these techniques have been studied explicitly for their effect on speech intelligibility.

In summary, it appears that vocal quality disorders reflecting a clear breakdown in phonatory source adequacy (such as excessive hypoadduction or hyperadduction) may have detrimental effects on speech intelligibility, such as the inability to produce voicing. The relationship between less severe disorders of voice quality and speech intelligibility is unclear at this time.

Contributions of SUPRASEGMENTAL functions of the larynx to speech intelligibility

Pitch

Vocal pitch has been defined as the perceptual correlate of the vibratory frequency of the vocal folds and has been studied perceptually and acoustically (fundamental frequency) in many populations. The relationship between pitch and speech intelligibility has been addressed primarily in the deaf and hard of hearing literature.

The overall pitch of the deaf and hard of hearing has been reported to be deviant (Green, 1956; Angelocci, Kopp and Holbrook, 1964; Martony, 1968; Bush, 1979; McGarr, Youdelman and Head, unpublished). Youdelman, MacEachron and McGarr (1989) report that inappropriate average pitch may sound unnatural and detract from the speaker's message and give conflicting cues about age and gender. Stevens et al (1983) suggested that improper adjustment of vocal fold posture can also result in a fundamental frequency contour that is overly sensitive to tongue position.

Monsen (1978), McGarr and Osberger (1978) and Samar and Metz (1988) conclude that there is no simple relationship between mean fundamental frequency and intelligibility in the deaf and hard of hearing. Sudden breaks in voicing and other evidences of inadequate control of phonation appear to be more highly correlated with poor intelligibility than inappropriate average pitch levels (McGarr, 1977; Levitt, et al, 1974).

The pitch of voice in esophageal speakers has been reported to be low (Weinberg and Bennett, 1972). Niebor et al. (1988) reported that the pitch of esophageal speakers plays an important role in assessment of voice quality. Shipp (1967) reported that one factor related to higher speech acceptability ratings was relatively higher mean fundamental frequency. The work of Robbins et al. (1984a; 1984b) suggested different fundamental frequencies for esophageal and tracheoesophageal speakers but did not address the relationship between fundamental frequency and intelligibility in these populations.

While disordered pitch has been reported in speakers with motor speech disorders (Aronson, 1985), its relationship with speech intelligibility in these populations has not been studied. For example, the pitch (fundamental frequency) of voices of patients with Parkinson's disease has been reported to be both excessively low (Darley, Aronson and Brown, 1975) and excessively high (Canter, 1963; Kammermeier,1969), but has not been identified as a factor which effects the intelligibility of these speakers.

Modification of pitch has been one goal of voice therapy throughout the years (Boone and McFarland, 1988). However, therapy-related pitch change has been associated with improvements in overall communication effectiveness rather than speech intelligibility. In relationship to speech

enhancement, Pichnney et al. (1986) found that a wider range of fundamental frequency is used in clear speech, with a slight bias toward higher fundamental frequencies.

In summary, it appears that the relationship between pitch or average speaking fundamental frequency and a speaker's intelligibility is unestablished at this time. The potentially distracting influence of a disordered pitch may be a contributing factor to intelligibility reductions.

Prosody (stress and intonation)

Communicative functions of speech prosody include conveying emotional tone (e.g., Scherer, 1986), conveying linguistic distinctions (e.g., Cooper and Sorenson, 1981), signaling meaning (Yorkston, 1988), and assigning stress (e.g., Cooper, Eady and Mueller,1985). Price and Levitt (1983) suggest that listeners may use suprasegmental information to assign an initial syntactic structure before decoding the rest of the information. Weismer (1990) has recently suggested that if this prosodic information is deviant, it may affect the listener's strategy for decoding the message. Kent (1988) suggests the following "vocal control prerequisites" for prosody: loudness variation, adequate duration of phonation, appropriate pitch level, pitch variation and acceptable voice quality.

Prosodic disorders have been frequently reported in speech of the hard of hearing and deaf and related to reductions in speech intelligibility. Metz, Samar, Schiavetti, Sitler and Whitehead (1985) have suggested independent and primary roles for segmental and prosodic speech characteristics in determining intelligibility in severely to profoundly hearing impaired speakers; this has been supported by Parkhurst and Levitt (1978) and Stromberg and Levitt (1979). Stark and Levitt (1974), Gold (1975) and McGarr (1976) have shown that deaf children have difficulty in producing such features as stress, pausal juncture, and intonation. Flat monotonous speech, lacking in pitch contours has been reported in deaf speakers (Haycock,1933; Greene,1956; Hood,1966; Monsen, 1979). They suggest that the effect of these errors on speech intelligibility is significant. For example, if the speakers' voice is at the top of the frequency range, there are limitations for increases in pitch to indicate stress or to produce the rise in pitch required for some question forms (McGarr, Youldeman, Head, unpublished). Levitt, Smith and Stromberg (1974) reported that children with the same frequency of segmental errors had speech intelligibility scores differing by as much as 30%. Parkhurst and Levitt (1978) reported that excessive variations in pitch may reduce intelligibility. Breaks in pitch was one of the prosodic errors they reported to show a significant negative regression with intelligibility. McGarr and Osberger (1978) report that a grossly deviant pitch pattern may be sufficiently distracting in a communication situation to have indirect effects on intelligibility.

Prosodic features have been studied in alaryngeal speakers. Intonational contrast to signal phrase intent (Gandour and Weinberg, 1983), juncture in distinguishing ambiguous word pairs (Scarpino and Weinberg, 1981) and lexical stress to distinguish noun verb pairs (Gandour, Weinberg and Garzone, 1983) have been reported as normal or near normal in esophageal speakers. McHenry, Reich and Minifie (1982) reported that while the primary syllable stress to emphasize one word in a sentence was not within normal limits in esophageal speakers, it was high. Doyle, Danhauer and Reed (1988) and Pindzola and Cain (1988) report that the increased pulmonary support in tracheoesophageal speakers allows their prosodic feature production and intonation to closely resemble normal talkers (Robbins, 1984; Robbins et al., 1984; Shipp, 1967; Williams and Watson,1985). Weinberg (1986) reports that tracheoesophageal speakers could control fundamental

frequency, duration, and intensity to mark suprasegmental (contrastive stress, intonation) contrasts. While these findings support near normal prosodic characteristics in esophageal and tracheoesophageal speakers, the relationship between these characteristics and speech intelligibility has not been investigated.

Breakdowns in prosodic characteristics have been reported in individuals with a variety of neurological disorders and have been associated with reductions in speech intelligibility (Ansel, 1987). Right brain damaged patients exhibit a disturbance of emotional and linguistic prosody (Shapiro and Danly, 1985). Left brain damaged patients demonstrate aprosodia (Ross, Anderson and Morgan-Fisher, 1989), involvement in the control of lexical and syntactic prosody (Behrens, 1985) and linguistic use of prosody in tones (Gandour and Dararananda, 1983). Ryalls (1986) and Ryalls, Joanette and Feldman (1987) suggest a whole brain basis for prosody and Kloude, Robin, Graff-Radford, Cooper (1988) observed prosodic impairment following callosal damage. Problems with prosody also have been observed to accompany Broca's aphasia (Danly and Shapiro, 1982), cerebellar ataxia (Kent and Rosenbek, 1982) and Parkinson's disease (Kent and Rosenbek, 1982; Blonder, Gur and Gur, 1989). Voices of Parkinson and ataxic patients have been described as having reduced pitch variability and monoloudness (Canter, 1963;1965; Kent and Rosenbek, 1982).

Prosody has often been a focus of treatment with motor disordered patients (Yorkston, Beukelman, Minifie and Sapir, 1984) and has been associated with changes in intelligibility in these patients. Yorkston and Beukelman (1981) suggest that in cases where prosodic patterns are often markedly abnormal and contribute to bizarreness of dysarthric speakers, forcing marginally intelligible speakers into specific stress patterns tends to increase intelligibility and reduce bizarreness. In contrast, Boothroyd et al. (1975) and Nickerson, Kalikow and Stevens (1976) reported improvement in suprasegmentals (lower fundamental frequency, consistent fundamental frequency contours and reduced pauses) post-treatment in deaf and hard of hearing children without corresponding improvement in intelligibility. They conclude that improvement in suprasegmental attributes alone is not sufficient to result in an immediate gain in intelligibility, although it may be accompanied by improvement in the naturalness or overall quality of the speech.

Hunicutt (1985; 1987) studied the prosodic correlates of peak dB level of word pairs, their durations, fundamental frequency maximum, range, excursion and contour complexity for words in high-redundancy versus low-redundancy contexts. The correlation between any one of these factors and intelligibility was low, however the differences were in the expected direction. She concluded "If prosody indeed correlates with intelligibility, it correlates as a whole, being expressed in various combinations of higher intensity, longer durations and more lively fundamental frequency contour."

In summary, it appears that disordered prosody can effect speech intelligibility. This may occur because the listener does not receive sufficent information to decode the message or because the disordered prosody prevents the listener from using her typical decoding strategy. Improved prosody may suggest that the speaker is better able to coordinate laryngeal and sub- and supraglottal events to generate the necessary variations in loudness, duration, pitch and quality which contribute to improved intelligibility.

Contributions of ARTICULATORY functions of the larynx to speech intelligibility

The voice-voiceless contrast is the primary articulatory function of the larynx that has been studied in relation to speech intelligibility (Smith, 1975; Yorkston and Beukelman, 1988) although

recently, the initial glottal versus null (/hat/ versus /at/) contrast has been considered (Kent, Weismer and Kim, 1990; Kent, Weismer, Kent and Rosenbek,1990). Both contrasts require coordination of laryngeal adductory/abductory events together with sub- and supraglottal events in order to generate sufficient acoustic cues for the perception of the phonetic target. Breakdowns in production of the voice-voiceless contrast have been observed in the deaf and hearing impaired as well as in alaryngeal and motor disordered speakers.

One common error in deaf speech is confusion of the voiced-voiceless distinction (Smith, 1975; Mc Garr and Osberger, 1978). Observations have included voiced for voiceless errors (Smith, 1975) as well as voiceless for voiced errors (Markides, 1970). Metz et al. (1985) reported cognate pair voice onset time differences and mean sentence duration strongly predicted speech intelligibility based on readings of isolated word and contextual speech material. Monsen (1978) found that three speech characteristics: the voice onset time difference between /t/ and /d/, the second formant difference between /i/ and /*/ and a rating of the spectrographic quality of liquids and nasals emerged as the best predictors of speech intelligibility. He pointed out that correct production of these few phonemes per se is not the source of intelligibility but rather that the presence of these features suggests that a given speaker has achieved a certain level of articulatory skill. For example, he suggests that the speaker is able to coordinate the timing of the release of the plosive burst at the lips with the onset of glottal pulsation and to vary that timing for different target phonemes. He suggests that the inability to control the temporal coordination of laryngeal and supralaryngeal events may be responsible for problems with phonemic contrasts.

Alaryngeal speakers' most frequent perceptual confusion is the voice/voiceless contrast (Doyle and Danhauer, 1986); voiceless stops are often perceived as voiced. This has been related to the non-adductor/abductor nature of the postlaryngectomy voicing mechanism and the reduced air supply available to the esophageal speaker. Weinberg, Horii, Blom and Singer (1982) have hypothesized that the esophageal voicing source is optimized using the trachesesophageal shunt. Doyle and Danhauer (1986) and Gomyo and Doyle, (in press) reported that tracheoesophageal speakers achieved the production of "voicing" cues with greater effectiveness than esophageal speakers and suggested that the addition of pulmonary air supply may result in changes in VOT and vowel duration and the generation of high pressures and flows needed for fricatives and plosives and may be related to improved speech intelligibility. Robbins, Christensen and Kempster (1986) suggest also that trachesophageal speakers' ability to generate greater and more sustained air pressures and flows likely results in slower pharyngoesophageal segment vibratory decay in contrast to the rapid decay in esophageal speech. They suggest that changes in the initiation or delay of phonation onset-offset as well as other temporal aspects such as vowel duration may result in perceptually salient changes in the speech signal.

Problems with laryngeal articulatory functions have been reported in the motor disordered population as well. In cases with inadequate vocal fold adduction (hypoadduction), voiced phonemes may be perceived as voiceless. This has been observed in cases of vocal fold paralysis. In other cases, extended consonantal voicing or continuous voicing (continuation of vocal fold vibration into voiceless stop closures) has been observed in Parkinson patients (Weismer, 1984) and patients with spastic dysarthria (Freeman, Cannito and Finitzo-Hieber, 1985). Farmer (1980) found frequent voiced for voiceless substitutions and VOTs which were longer and more variable in spastic and athetoid cerebral palsy speakers. She suggests that one important factor underlying the poor intelligibility of cerebral palsy speakers is the distortion caused by increased word and interword durations. Yorkston and Beukelman (1988) reported that imprecise phonatory control

may prevent certain speakers from producing the voice/voiceless distinctions that are important in intelligibility. Kent et al. (1990a;1990b) reported voiced-voiceless (e.g., bat-pat) as well as glottal and null (e.g., hat-at) confusions in patients with Parkinson's disease as well as amyotrophic lateral sclerosis and associated this with reduced intelligibility in these patient populations.

Articulatory contrasts have been the focus of therapy designed to improve speech intelligibility. Yorkston and Beukelman (1988) suggest contrastive drills marking voiced-voiceless consonant distinctions for dysarthric patients to improve intelligibility. Till and Toye (1988) reported that patient subjects modified their VOTs after specfic feedback of a voice/voiceless intelligibility confusion and not after general feedback of communication failure. In relation to enhanced speech, Pichney et al. (1986) reported that VOT for unvoiced plosives increased substantially in clear speech. Chen (1980) observed this also in speech spoken clearly and conversationally in a carrier phrase.

In summary, it appears that there is a relationship between the articulatory functions of the larynx and speech intelligibility. Successful generation of the voice-voiceless contrast may offer the listener additional information to decode speech accurately. In addition, it may reflect the speaker's higher level of coordination of laryngeal events with sub- and supraglottal events which facilitate increased intelligibility across the board.

Summary

A common theme in the literature related to disordered phonation and speech enhancement seems to be that the disruption or the enhancement of the source aspects of laryngeal sound production underlie major changes in speech intelligibility. Once a speaker achieves a threshold of "adequacy of phonatory source" both acoustically and aerodynamically, the contribution of suprasegmental or articulatory phonatory characteristics to speech intelligibility may exist, but be less significant. Future research should assess the contribution of laryngeal function to intelligible speech production in an hierarchical way so that "phonatory source adequacy" is studied together with laryngeal contributions on suprasegmental and articulatory levels. It is critical to study systematically phonatory source contributions to intelligibility of segmental productions. One initial approach would be to employ speech synthesis to study the effects on speech intelligibility when the laryngeal and respiratory source characteristics are systematically varied while articulatory configurations remain constant.

Kent (1988;120) has stated that "Because intelligibility is the essential feature of speech communication, the assessment of intelligibility is an issue of fundamental clinical importance". The virtual absence of the concept of speech intelligibility in the voice disorders literature together with the limited discussion of the contribution of phonatory characteristics to intelligibility in all but the deaf and hearing impaired literature suggests that the role of phonation has not been sufficiently studied for its role in clinical assessment of speech intelligibility. The vast majority of the disordered voice literature relates characteristics of phonation not to reduced intelligibility but to reduced "effectiveness", "acceptability" (Shipp, 1967; Weinberg and Bennett, 1973; Pindzola and Cain, 1988) or increased "distraction" or negative perception (Moran, LaBarge and Haynes,1988; Ruscello, Lass and Podbesek, 1988). Speech intelligibility has been defined as "the degree to which the speaker's intended message is recovered by the listener" (Kent, Weismer, Kent and Rosenbek, 1989; 483). At present, the contribution of phonatory characteristics to that recovery is not established. While the deaf and hearing impaired literature has begun to document the role of

disordered phonation in speech intelligibility reductions, it is critical that disordered phonation of all etiologies be assessed for their impact on speech intelligibility. Such information would allow a common metric for determining the impact of speech breakdowns on communication functioning across client populations and offer a unifying perspective for the study of phonation and speech intelligibility. Improvement following voice therapy should be assessed for its impact on speech intelligibility as well; unless treatment can be shown to positively impact an individual's overall intelligibility, its social validity may be in question (Kent, 1988).

The following study makes an initial attempt to integrate ratings of speech intelligibility with acoustic, kinematic and perceptual data on vocal function before and after voice therapy for patients with Parkinson's disease. The study investigates the relationships among perceptual ratings and acoustic and kinematic measures of source and suprasegmental characteristics such as loudness, quality and intonation and measures of vocal fold adduction and oscillatory steadiness in patients whose vocal fold adduction improved after voice therapy. The relationships among these variables and perceptual ratings are studied to gain insight into which, if any, phonatory variables were related to improvements in speech intelligibility after therapy.

The Relationship Between Improved Vocal Fold Adduction and Speech Intelligibility in Patients with Parkinson's Disease.

Disordered communication is a problem experienced frequently by patients with Parkinson's disease (e.g. Logemann, Fisher, Boshes and Blonsky, 1978). This problem primarily has been related to disorders of articulation (e.g. Weismer, 1984) and rate (e.g. Canter, 1969). Traditional speech therapy for these patients (e.g. overarticulate, increase articulatory precision, slow rate) generally has been unsuccessful (Allan, 1970; Green, 1980; Sarno, 1968). While disordered phonation has been reported in 89% of a group of patients with Parkinson's disease (Logeman, Fisher, Boshes and Blonsky, 1978), only recently have researchers and clinicians begun to focus treatment efforts on improving the phonatory abilities of these patients (Scott and Caird,1983; Robertson and Thompson, 1984; and Ramig, Mead, Scherer, Horii, Larson and Kohler, 1988a; Ramig, Fazoli, Scherer and Bonitati, 1990). Ramig et al. (1988a) identified a "phonatory inadequacy" in this patient population and designed a treatment program to focus on increasing basic respiratory/laryngeal source characteristics. They reported significantly improved vocal abilities in patients with Parkinson's disease following intensive voice therapy which focused primarily on increasing vocal fold adduction and improving use of respiratory support for speech. It has not been established however, whether these improvements in vocal function translated into improved perceptual characteristics of speech including increased speech intelligibility in this patient population. It was the purpose of the study reported here to investigate the relationship between vocal function and perceptual measures of speech including speech intelligibility in a group of patients with Parkinson's disease who experienced maximum improvement in vocal function following intensive voice therapy.

Methods

Subjects - Seven neuropharmacologically stable patients with idiopathic Parkinson's disease were selected from a group of two-hundred patients who participated in a four-week intensive program of speech therapy. All subjects were male, ranging in age from 63 to 77 years. Three were in Stage III and four were in Stage IV of Parkinson's disease (Hoehn and Yahr, 1969). These

seven patients were selected for this study because they had attained maximum improvement in vocal function following therapy as defined by their acoustic, electroglottographic and videolaryngoscopic data (Ramig, Fazoli, Scherer and Bonitati, 1990). All seven patients demonstrated increased vocal fold adduction and phonatory stability as evidenced by a comparison of the pre- and post-treatment electroglottographic measures of the abduction quotient (Titze, 1984) and EGGW (Scherer and Vail, 1988), perceptual ratings of videolaryngoscopic examinations and acoustic measures of shimmer and the coefficient of variation of amplitude (Ramig, Fazoli, Scherer and Bonitati, 1990). All patients had participated in four-weeks of intensive (three to four times a week) voice therapy. This therapy was carried out in the context of a multidisciplinary treatment program which included physical therapy, occupational therapy, nutrition, counseling and exercise. Specific voice therapy procedures and underlying rationale have been discussed previously (Ramig et al, 1988 a,b,c) and are summarized in Table 1.

Table 1.

Framework and rationale for initial program of speech therapy administered to forty patients with idiopathic Parkinson's disease; treament philosophy is intensive therapy with a focus on phonation and immediate carryover into functional communication (Ramig et al., 1988a,b,c)

Perceptual characteristics of speech	Hypothesized laryngeal and/or respiratory pathophysiology	Therapy goals and tasks	Acoustic, physiologic variables measured	Perceptual variables measured
"Reduced loudness, breathy, weak voice" (Logemann, et al., 1978; Aronson, 1985)	Bowed vocal folds (Hansen et al., 1984), rigidity, hypokinesia in laryngeal and/or respiratory muscles; reduced adduction; reduced inspiratory, expiratory volumes (Critchley, 1981)	1) increase vocal fold adduction - isometric (pushing, lifting) with phonation increase maximum duration vowel phonation at increased intensity -think "shout" -speak over background noise	Maximum duration of sustained vowel phonation (sec)	loudness breathiness intelligibility
		2) increase respiratory support -posture -deep breath before speaking -frequent breaths -phrasing of words in sentences	vital capacity (L;%)	
"Reduced pitch variability monopitch" (Logemann, et al., 1978; Aronson, 1985)	Rigidity cricothyroid muscle (Aronson, 1985)	3) increase maximum fundamental frequency range -high and low pitch scales -sustain phonation at highest and lowest pitches	maximum range of fundamental frequency (ST)	
		4) increase fundamental frequency variation in connected speech -word emphasis -intonation in questions	variability of fundamental frequency in connected speech (STSD)	monotone intelligibility
Unsteady, hoarse, rough voice (Logemann, et al., 1978)	Rigidity, hypokinesia tremor in laryngeal and respiratory muscles (Hansen et al, 1984; Critchley, 1981)	5) increase steadiness of phonation -maximum duration tasks with constant intensity -consistent firm voice throughout sentence	improved measures of phonatory stability (coefficient of variation of frequency, coefficient of variation of amplitude, jitter, shimmer, harmonics to noise ratio)	steadiness of voice hoarse rough tremorous intelligibility

Procedures - A variety of procedures were carried out to evaluate changes in patients' speech intelligibility following therapy. Each patient participated in a video taping session which sampled approximately five minutes of conversational speech before (pre) and after (post) therapy. All data were collected at the same time post-medication for each patient. Video tape recordings

were made using a JVC Highband Saticon color video camera Model GX-S700 with a built in stereo microphone on to a JVC VHS video cassette recorder Model BR-1600U in a quiet environment. The microphone to patient distance was five feet and remained constant throughout all recordings. Given the well-established performance variability of Parkinson patients in a test situation (e.g. Weismer, 1984), conversational speech was selected inorder to maximize the likelihood of sampling functional communication abilities of these patients and minimize the effects of reading ability and visual acuity (Giolas and Epstein, 1963; Connolly, 1986). These pre- and post-video recordings were presented by a Panasonic NV8500 VHS recorder and two Panasonic CT110 nine inch color monitors in random order through earphones at constant and comfortable loudness to three speech pathologists familiar with rating the speech of Parkinson's disease patients. Netsell (1984) has suggested that perceptual ratings of speech should include at least three raters. These speech pathologists rated these samples on a visual analogue scale of speech characteristics (Schiffman, Reynolds and Young, 1981; Kempster, 1984) and on two scales of speech intelligibility: the Comunication Profile developed at the National Technical Institute for the Deaf (NTID) (Johnson,1975; McGarr and Osberger, 1978) and a modification of this profile. The following speech characteristics were rated on the visual analogue scale: loudness, shakiness of voice, scratchiness of voice, monotone, slurring, mumbling, speaking so that others can understand, participating in conversations and starting conversations. The visual analogue scale was designed to be a clinically feasible tool for both patients and professionals, which targets perceptual features related to treatment goals for these patients. Schiavetti, Metz and Sitler (1981) have demonstrated that because of nonlinearities, the NTID scale is not able to discriminate among individuals over the entire intelligibility range. We modified the NTID profile based on the findings of our pilot work which indicated that the scale in its original form was not sensitive to the type of improved communication our patients experienced after therapy. We presented both the original and modified version of this scale to listeners. Twenty percent of all ratings were repeated within task for an assessment of intratask reliability. In the clinical situation, visual analogue rating scales and NTID scales were completed by patients, caregivers, two speech pathologists and interdisciplinary team members pre- and post-treatment. The three scales are presented in Table 2.

Table 2.

Perceptual scales used in this study: 5 point NTID Intelligibility Rating and 6 point NTID Intelligibility Rating.

6 Point Modification of NTID Intelligibility Rating

1. Speech cannot be understood.
2. Speech is very difficult to understand - only isolated words or phrases are intelligible.
3. Speech is difficult to understand; however, the gist of the content can be understood.
4. Speech is intelligible with the exception of a few words or phrases.
5. Speech is intelligible, but the listener must work hard.
6. Speech is completely intelligible.

5 Point NTID Intelligibility Rating

1. Speech cannot be understood.
2. Speech is very difficult to understand - only isolated words or phrases are intelligible.
3. Speech is difficult to understand; however, the gist of the content can be understood.
4. Speech is intelligible with the exception of a few words or phrases.
5. Speech is completely intelligible.

Results

To assess intratask reliability, correlation coefficients were calculated for the within task repeated ratings of the perceptual variables on the visual analogue scale and the five and six point NTID scales by the three speech pathologists. For the visual analogue scale, these correlations ranged from .91 for loudness ratings to .98 for understandability or intelligibility ratings; coefficients for repeated ratings of the other perceptual variables fell between these two values. For the five and six point NTID scale of intelligibility, correlation coefficients were .92 and .97 respectively for repeated ratings. To assess the relationships among intelligibility assessed on the visual analogue scale and the NTID scales, correlation coefficients were calculated among these ratings. Intelligibility as assessed on the visual analogue scale correlated .69 and .70 with intelligibility ratings assigned on the five and six point NTID scales respectively. The relationship between pre- and post-therapy improvement in intelligibility (expressed in percent) as measured on the visual analogue scale and the five and six point NTID scales were reflected in correlation coefficients of .66 and .82 respectively. The average post-therapy intelligibility increase across all patients on the visual analogue scale was 21%. For the five and six point NTID scales it was 16% and 25.3% respectively.

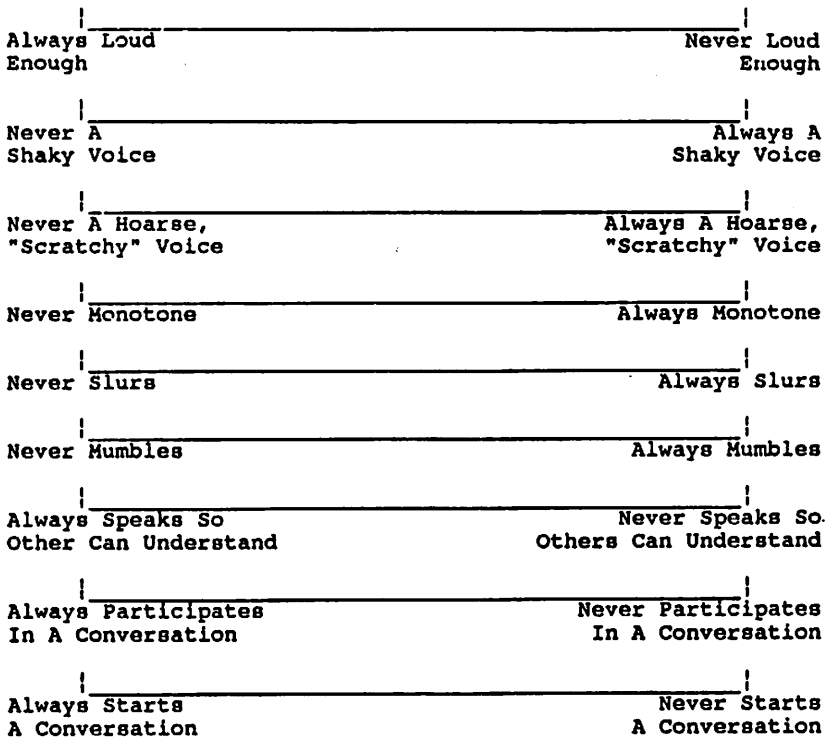

Table 2 (continued)
The third perceptual scale used in this study: Visual Analog Scale

To evaluate the impact of improved vocal function (increased adduction and stability) previously documented on perceptual measures, listener ratings were compared between the pre- and post-therapy conditions. There were statistically significant differences in listener ratings for all perceptual variables between the pre- and post-therapy conditions as assessed by paired t-tests; this included loudness, understandibility (intelligibility), scratchiness, shaky voice, monotone, mumbling and slurring from the visual analogue scale and intelligibility from the NTID scales. These data are summarized in Table 3.

To evaluate the relationship between these perceptual ratings and kinematic (e.g., abduction quotient, eggw) and acoustic (coefficient of variation of amplitude, shimmer) data, correlation coefficients were calculated between changes in perceptual and kinematic and acoustic measures of vocal function from the pre- to post-treatment conditions. The only statistically significant relationships measured by Pearson product moment correlation coefficients were between the abduction

quotient and ratings of both loudness and scratchiness; patients who increased vocal fold adduction most significantly were also rated as having the greatest increases in vocal loudness and reductions in scratchiness. There were no significant correlations with these measures and ratings of speech intelligibility.

To evaluate the impact of these perceptual changes on ratings of intelligibility pre- and post-therapy, correlation coefficients were calculated between changes in the perceptual measures and intelligibility as rated on the visual analogue scale. Ratings of monotone and shaky voice correlated with intelligibility .88 and .74 respectively; the other correlation coefficients ranged from .67 for loudness, .64 for slurs, .63 for mumbles and .32 for scratchiness. To evaluate the contribution of a combination of these various perceptual variables to overall ratings of intelligibility, a multiple regression analysis was carried out with intelligibility ratings as the criterion and the other perceptual variables as the predictors. The combination of monotone and shaky voice (intercorrelation of .53) resulted in a correlation coefficient of .94, suggesting that this combination of perceptual varaibles was able to account for 88% of the variance in speech intelligibility ratings for this group of speakers.

Perceptual variable		Therapy condition Pre	Post
Loudness	m	53.3	87.4
	SD	19.1	9.4
Shakiness	m	71.3	85.0
	SD	9.2	7.6
Scratchiness	m	59.4	87.1
	SD	25.4	8.9
Monotone	m	53.6	84.3
	SD	20.6	12.1
Slurring	m	62.1	86.6
	SD	17.7	8.7
Mumbling	m	57.9	85.7
	SD	20.7	9.6
Understandability	m	70.4	91.6
	SD	14.1	7.7
Intelligibility (NTID 5-point)	m	4.07	4.62
	SD	.40	.49
Intelligibility (NTID 6-point)	m	4.55	5.45
	SD	.73	.60

Table 3: Group means and standard deviations of perceptual ratings from the visual analogue scale and five and six point NTID intelligibility scales pre- and post- therapy for seven patients with Parkinson's disease.

Discussion

These findings demonstrate that the improvement in vocal function documented by acoustic and kinematic measures in Parkinson's patients following intensive voice therapy translated to improvement in perceptual measures as well. These Parkinson's patients improved in perceptual measures of vocal function (loudness, monotone, shaky and scratchy voice) after intensive voice therapy. In addition, it appears that the increased vocal effort also positively affected articulatory variables as evidenced by improved ratings of slurring and mumbling. Overall, ratings of intelligibility were improved in the post-therapy condition and were most closely related to ratings of reduced monotony and shakiness of speech. The finding that monotony and shaky voice were related most closely to ratings of intelligibility supports the importance of phonatory characteristics

in the intelligibility of these speakers. The patients' increased vocal effort appeared to improve overall communication performance because of improved acoustic and aerodynamic sources and perhaps because of generalization of this increased effort to articulatory performance as well.

It should be pointed out that the acoustic and kinematic data were analyzed from sustained vowel phonation and the perceptual ratings were made from conversational speech. Therefore direct relationships among acoustic, kinematic and perceptual variables can not be verified. However, these data support that the improvements measured from sustained vowels such as increased vocal fold adduction and amplitude steadiness appeared to be carried over into conversational speech and were reflected in ratings of increased loudness and reduced shakiness and scratchiness after therapy.

The relationship between improved ratings of monotony of speech and intelligibility can be interpreted in at least two ways. Perhaps the improved intonation provided the listeners with additional information to enhance speech intelligibility. On the other hand, perhaps as Price and Levitt (1983), Kent et al. (1989) and Weismer (1990) have suggested, the improved intonation patterns were consistent with listeners' perceptual set or tolerance to allow them to use familiar strategies to decode the message with greater ease and thus enhance intelligibility.

The patients in this study all had reduced vocal fold adduction associated with bowed vocal folds in the pre-therapy condition. While the voice-voiceless contrast is the primary measurement of adductory breakdown on segmental measures of speech intelligibility, our pilot work using voice-voiceless contrast word pairs with these patients suggests that most of them were able to make this contrast with ease. These same patients were rated as having reduced intelligibility and loudness and required additional effort on the part of the listener in order to be understood. One interpretation may be that in a particular contrastive word pair task the Parkinson's patients demonstrated their typical performance variablity by using maximum effort to achieve maximum performance which is not carried over to spontaneous speech. Another interpretation is that because of its segmental nature, the voice-voiceless contrast does not reflect overall breakdowns in adduction which reduce loudness across an entire utterance. Patients may be able to achieve sufficient respiratory, laryngeal and supraglottal coordination to generate the voicing contrast for a discrete unit of time, but are unable to achieve and maintain the adductory and respiratory forces necessary to generate adequate loudness across an entire utterance.

It should be noted that in many patients with Parkinson's disease, especially early in the disease process, reduced volume or loudness may be the primary complaint. While this loudness may not reduce intelligibility as documented by current procedures, it may require the "listener to work hard". The ability to document this additional effort on the part of the listener may prove to be very important. Our preliminary observations suggest that early intervention at this time with Parkinson's disease patients may be significant in terms of maintaining communication skills. If patients can learn techniques to maintain vocal function at this early point in the disease when their motivation and cognitive skills are frequently intact, it is likely that they will achieve greater success over time.

This study begins to address the interactions among acoustic, physiologic and perceptual measures of phonatory function and speech intelligibility. Future research should continue to address these relationships.

Acknowledgement

Preparation of this chapter was supported in part by NIH grant #8 RO1 DC00387-03. This research was supported in part by funds from the Lee Silverman Center for Parkinson's, Scottsdale, Arizona, the Axe-Houghton Foundation and the Parkinson's Association of the Rockies. Appreciation is expressed to Carolyn Bonitati, Katy Fazoli, Ron Scherer, Roberta Borne, Carlos Fazoli and Ken Tagawa for assistance during various parts of this project.

References

Addington, D.W. 1968. "The relationship of selected vocal characteristics to personality perception". Speech Monographs 35.492-503.

Allan, C.M. 1970. "Treatment of non-fluent speech resulting from neurological disease: treatment of dysarthria". British Journal of Disorders of Communication 5.3-5.

Allen, D.R., Strong, W.T., and Palmer, E.P. 1981. "Experiments on the intelligibility of low frequency speech codes". Journal of the Acoustical Society of America 70.1248-1255.

Andrews, G., Platt, L.J. and Young, M. 1977. "Factors affecting the intelligibility of cerebral plasied speech to the average listener". Folia Phoniatrica 29.292-301.

Angelocci, A., Kopp. G., Holbrook, A. 1964. "The vowel formants of deaf and normal hearing eleven to fourteen year old boys". Journal of Speech and Hearing Disorders 29. 156-170.

Ansel, B.A. 1985. "Acoustic predictors of speech intelligibility in cerebral palsied dysarthric adults". Unpublished doctoral dissertation, University of Wisconsin-Madison, Madison, WI.

Aronson, A.E. 1985. Clinical Voice Disorders. New York:Thieme.

Arkebauer, H., Hixon, T. and Hardy, J.1967. "Peak intraoral air pressures during speech". Journal of Speech and Hearing Research 10.196-208.

Behrens, S.J. 1985. "The perception of stress and the lateralization of prosody". Brain and Language 26. 332-348.

Bellaire, K., Yorkston, K. and Beukelman, D. 1986. "Modification of breath patterning to increase naturalness of a mildly dysarthric speaker". Journal of Communication Disorders 19.271-280.

Bennett, S. and Weinberg, B. 1973. "Acceptability ratings of normal, esophageal, and artificial larynx speech". Journal of Speech and Hearing Research 16.608-615 .

Bernstein, J., Rollins, A.M. and Stevens, K.N. 1978. "Word and syllable concatenation in the speech of the deaf". A paper presented at the meeting of the A.G. Bell Association for the Deaf, St.Louis, June.

Berry, C. 1973. Voice and the Actor. London:George G. Harrap and Company, Ltd.

Blonder, L.X., Gur, R and Gur, R.1989. "The effects of right and left hemiparkinsonism on prosody". Brain and Language. 36.193-207 .

Boone, D.R. and McFarlane, S.C.1988. The voice and voice therapy. Englewood Cliffs:Prentice Hall.

Boothroyd, A., Archambault, P., Adams, R.E. and Storm, R.D. 1975. "Use of a computer- based system of speech analysis and display in a remedial speech program for deaf children". The Volta Review 77.178-193.

Bush, M. 1979. "Articulatory proficiency and Fo control by profoundly deaf speakers". A paper presented at the meeting of the American Speech-Language-Hearing Association, November, Atlanta.

Canter, G. 1965. "Speech characteristics of patients with Parkinson's disease: III. Articulation". Journal of Speech and Hearing Disorders. 30.217-24.

Canter, G. 1963. "Speech characteristics of patients with Parkinson's disease I. Intensity, pitch and duration". Journal of Speech and Hearing Disorders 28.221-8.

Carney, A. E. 1986. "Understanding speech intelligibility in the hearing impaired". Topics in Language Disorders 6(3).47-59.

Chodosh, P., Giancarlo, H., Goldstein, J. 1984. "Pharyngeal myotomy for voice rehabilitation post laryngectomee". Laryngoscope 94.52.

Cleveland, T. and Sundberg, J. 1983. "Acoustic analysis of three male voices of different quality". Proceedings of Stockholm Music Acoustics Conference ed. by A. Askenfelt, S. Felicetti, E. Jansson, and J. Sundberg, 143-156. Stockholm:The Royal Swedish Academy of Music.

Connolly, J.H. 1986. "Intelligibility: a linguistic view". British Journal of Disorders of Communication 21.371-376.

Cooper, W.E., Eady, S.J.and Mueller, P.R. 1985. "Acoustical aspects of contrastive stress in question-answer contexts". Journal of the Acoustical Society of America 77. 2142-2156.

Cooper, W.E. and Sorenson, J.M. 1981. Fundamental frequency in sentence production. New York:Springer-Verlag.

Critchley, E.M.R. 1981. "Speech disorders of parkinsonism: a review". Journal of Neurology, Neurosurgery and Psychiatry. 44.751-758.

Curtis, J.F. 1946. "Intelligibility related to microphone position". Speech Monographs 13.8-12.

Danly, M, and Shapiro, B. 1982. "Speech prosody in Broca's aphasia". Brain and Language 16.171-190.

Darley, F., Aronson, A. and Brown, J. 1969a. "Clusters of deviant speech dimensions in the dysarthrias". Journal of Speech and Hearing Research 12.462-469.

Darley, F.L., Aronson, A. and Brown, J. 1969b. "Differential diagnostic patterns of dysarthria". Journal of Speech and Hearing Research 12.246-269.

Darley, F.L., Aronson, A. and Brown, J. 1975. Motor Speech Disorders Philadelphia: W.B.Sanders.

Deal, R.E. and Emanuel, R.W. 1978. "Some waveform and spectral features of vowel roughness" Journal of Speech and Hearing Research 21.250-264.

Diedrich, W.M. 1968. "The mechanism of esophageal speech". Sound Production in Man. New York: Annals of the New York Academy of Sciences.155.303-317 .

Doyle, P.C. and Danhauer, J.L. 1986. "Consonant intelligibility of alaryngeal talkers: pilot data". Human Communuication Canada 10.4 .21-28.

Doyle, P.C., Danhauer, J.L. and Reed, C.G. 1988. "Listeners' perceptions of consonants produced by esophageal and tracheoesophageal talkers". Journal of Speech and Hearing Disorders. 53.400-407.

Draegert, G.L. "Relationship between voice variables and speech intelligibility in high level noise". Speech Monographs

Drummond, S. 1965."The effects of environmental noise on pseudovoice after laryngectomee" . Journal of Laryngology 79.193-202 .

Esling, J.H. 1978. "Voice quality in Edinburgh: a sociolinguistic and phonetic study". An unpublished doctoral dissertation. Edinburgh University .

Farmer, A. 1980. "Voice onset time production in cerebral palsied speakers". Folia Phoniatrica. 32.267-73.

Freeman,F. Cannito, M.P. and Finitzo-Heiber, T. 1985. " Classification of spasmodic dysphonia by perceptual-acoustic-visual means". Spasmodic dysphonia:the state of the art ed. by G.Gates, 5-18. New York:The Voice Foundation.

Freyman, R.L. and Nerbonne, G.P.1989. "The importance of consonant-vowel intensity ratio in the intelligibility of voiceless consonants". Journal of Speech and Hearing Research. 32.524-535.

Gandour, J. and Dararananda, R. 1983. "Identification of tonal contrasts in Thai aphasic patients". Brain and Language18.98-114.

Gandour, J. and Weinberg, B. 1983. "Perception of intonational contrasts in alaryngeal speech". Journal of Speech and Hearing Research 26.142-148.

Gandour, J., Weinberg, B. and Garzone, B. 1983. "Perception of lexical stress in alaryngeal speech". Journal of Speech and Hearing Research 26.418-424.

Garber, S.R., Siegel, G.M. and Pick, H.L. 1980. "The effects of feedback filtering on speaker intelligibility". Journal of Communication Disorders 13 .289-294.

Giolas, T.G. and Epstein, A. 1963. "Comparative intelligibility of word lists and continuous discourse". Journal of Speech and Hearing Research, 6. 349-358.

Gold,T. 1975. "Perception and production of prosodic features by hearing impaired children". A paper presented at the Convention of the American Speech and Hearing Association, Washington, D.C., November.

Gomyo, Y. and Doyle, P.C. Forthcoming. "Perception of stop consonants produced by esophageal and tracheoesophageal speakers". To appear in Journal of Otolaryngology.

Gordon-Salant, S. 1986. "Recognition of natural and time-intensity altered CVs by young and elderly subjects with normal-hearing". Journal of the Acoustical Society of America 80.1599-1607.

Gordon-Salant, S. 1987. "Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects". Journal of the Acoustical Society of America 81.1199-1202.

Greene, D.S. 1956. "Fundamental frequency of the speech of profoundly deaf individuals". An unpublished doctoral dissertation, Purdue University, West Lafayette, IN.

Greene, M.C.L. 1980. The Voice and Its Disorders. London:Pitman Medical.

Greene, M.C.L. and Watson, B.W. 1968. "The value of speech amplification in Parkinson's disease patients". Folia Phoniatricia. 20.250.

Hanley ,T.D. and Steer, M.D. 1949. "Effect of level of distracting noise upon speaking rate, duration and intensity". Journal of Speech and Hearing Disorders 14. 363-368.

Hansen, D.G., Gerratt, B.R. and Ward, P.H. 1984. "Cinegraphic observations of laryngeal function in Parkinson's disease". Laryngoscope 94.348-353.

Haycock, G.S. 1933. The Teaching of Speech. Stoke-on-Trent, England:Hill and Ainsworth.

Hecker, M.H.L. 1974. " A study of the relationships between consonant and vowel ratios and speaker intelligibility". Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Hoehn, M. and Yahr, M. 1969. "Parkinsonism onset, progression and mortality". Neurology 17.427.

Hood, R.B. 1966. "Some physical concomitants of the perception of speech rhythm of the deaf". Unpublished doctoral dissertation, Stanford University, Stanford, CA.

House, A.S., Williams, C.E., Hecker, M.H.L. and Kryter, K. 1965. "Articulation testing methods:consonantal differentiation with a closed response set". Journal of the Acoustical Society of America 37, 158-166.

Hunnicutt, S. 1987. "Acoustic correlates of redundancy and intelligibility". STL-QPSR 2-3.7-14.

Hunnicut, S. 1985. "Intelligibility versus redundancy-conditions of dependency". Language and Speech 28.47-56.

Hyman, M., 1955. "An experimental study of artificial larynx and esophageal speech". Journal of Speech and Hearing Disorders 20. 291-299.

Johnson, D.D. 1975. "Communication characteristics of NTID students". Journal of Academic Rehabilitative Audiology 8(1,2).17-32.

Kalb, M.B. and Carpenter, M.A. 1981. "Individual speaker influence on relative intelligibility of esophageal speech and artificial larynx speech". Journal of Speech and Hearing Disorders 46.77-80.

Kammermeier, M.A. 1969. "A comparison of phonatory phenomena among groups of neurologically impaired speakers". An unpublished doctoral dissertation. University of Minnesota, Minneapolis.

Kempster, G.B., 1984. "A multidimensional analysis of vocal quality in two dysphonic groups". An unpublished doctoral dissertation. Northwestern University,Evanston.

Kent, R.D., Weismer, G. and Kim, Hyang-Hee 1990a. "Segmental and suprasegmental laryngeal dysfunction in neurological disease: amyotrophic lateral sclerosis, Parkinson's disease, stroke and cerebellar disease". A paper presented at the conference Neurological Disorders of Laryngeal Function. Los Angeles, CA.

Kent, R.D., Weismer, G. and Rosenbek, J.C. 1990b."Acoustic-phonetic dimensions of intelligibility impairment in dysarthria". A paper presented at the conference Clinical Dysarthria. San Antonio, TX.

Kent, R.D. 1988 "Speech intelligibility". Decision Making in Speech-Language Pathology ed. by D.E. Yoder and R.D. Kent 140-143. Philadelphia: Decker.

Kent, R and Rosenbek, J. 1982. "Prosodic disturbance and neurologic lesion". Brain and Language 15. 259-291.

Kent, R.D., Weismer,G. Kent, J.F. and Rosenbek, J.C. 1990. "Toward phonetic intelligibility testing in dysarthria". Journal of Speech and Hearing Disorders 54.4.482-499.

Klatt, D.H. 1975. "Vowel lengthening is syntactically determined in connected discourse". Journal of Phonetics 3.129-140.

Klouda, G.V., Robin, D.A., Graff-Radford, N.R. and Cooper, W. 1988. "The role of callosal connections in speech prosody". Brain and Language 35. 154-171.

Kohler and Dommelen 1987.

Lane , H. and Tranel, B.1970. "The Lombard signs and the role of hearing in speech". Journal of Speech and Hearing Research 14. 677-709.

Laver, J. 1968. "Voice quality and indexical information". British Journal of Disorders of Communication 3.43-54.

Laver, J. and Trudgill, P. 1979. "Phonetic and linguistic markers in speech". Social Markers in Speech ed. by K.R. Scherer and H. Giles, 1-32. Cambridge:Cambridge University Press.

Laver, J. 1980. The phonetic description of voice quality. Cambridge: Cambridge University Press.

Laver, J., Wirz, S., Mackenzie, J., and Hiller, S. 1981. A perceptual protocol for the analysis of vocal profiles. Work in Progress (University of Edinburgh, Department of Linguistics), 14. 139-155.

Lessac, A. 1967. The use and training of the human voice. New York; Drama Book Publishers.

Levitt, H., Smith, C.R., Stromberg, H. 1974. "Acoustic, articulatory and perceptual characteristics of the speech of deaf children". Speech Communication Seminar, Stockholm.

Lieberman, P. 1960. "Some acoustic correlates of word stress in American English". Journal of the Acoustical Society of America 32.451-454.

Lim, J.S., Oppenheim, A.V., and Braida, L.D. 1978. "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition". IEEE Transactions on Acoustics, Speech and Signal Processing. 26.354-358.

Logemann, J.A. Fisher,H.B., Boshes,B. and Blonsky, E.R. 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. and Bassich, C.J. 1983. "The results of acoustic and perceptual assessment of two types of dysarthria". Clinical Dysarthria ed. W.R. Berry, 121-153. San Diego: College-Hill Press.

McGarr, N.S. 1978. "The differences between experienced and inexperienced listeners in understanding the speech of the deaf". Unpublished doctoral dissertation, City University of New York.

McGarr, N.S. 1976 "The production and reception of prosodic features". A paper presented at the annual meeting of the A.G. Bell Association, Boston, Mass, June.

McGarr, N.S. and Osberger, M.J. 1978. "Pitch deviancy and intelligibility of deaf speech". Journal of Communication Disorders 11 .

McGarr, N.S., Youdelman, K. and Head, J. (in press) Curriculum to remediate the monotone voice.

McHenry, M., Reich, A. and Minifie, F. 1982. "Acoustical characteristics of intended syllabic stress in excellent esophageal speakers". Journal of Speech and Hearing Research 25.564-573.

Markides, A. 1970. "The speech of deaf and partially hearing children with special reference to factors affecting intelligibility". British Journal of Disorders of Communication. 5. 126-140.

Martony, J. 1968. "On correction of voice pitch level for severely hard of hearing subjects". American Annals of the Deaf 113.195-202.

Metz, D.E., Samar, V.J., Schiavetti, N., Sitler, R. and Whitehead, R.L. 1985. "Acoustic dimensions of hearing-impaired speakers' intelligibility". Journal of Speech and Hearing Research 28.345-355.

Monsen, R.B. 1978. "Toward measuring how well hearing impaired children speak". Journal of Speech and Hearing Research 21.197-219.

Monsen, R.B. 1979. "Acoustic qualities of phonation in young hearing-impaired children". Journal of Speech and Hearing Research 22.270-288.

Monsen, R.B. 1983. "The oral speech intelligibility of hearing-impaired talkers". Journal of Speech and Hearing Disorders 48.286-296.

Montgomery, A.A. and Edge, R.A. 1988. "Evaluation of two speech enhancement techniques to improve intelligibility for hearing-impaired adults". Journal of Speech and Hearing Research 31. 386-393.

Moore, P. 1946 "Intelligibility related to routinized messages". Speech Monographs 13. 46-49.

Moran, M.J., LaBarge, J.M. and Haynes, W.O. "Effect of voice quality on adult's perceptions of Down's syndrome children". Folia Phoniatricia 40.157-161.

Mueller, P. 1971. "Parkinson's disease: Motor-speech behavior in a selected group of patients". Folia Phoniatrica 23.333-346.

Netsell, R., Lotz, W. and Shaughnessy, A. 1984."Laryngeal aerodynamics associated with selected voice disorders". American Journal of Otolaryngology. 5(6). 397-403.

Nieboer, G.L.J., deGraaf, T. and Schutte, H.K. 1988. "Esophageal voice quality judgements by means of the semantic differential". Journal of Phonetics 16.417-436.

Nickerson, R.S., Kalikow, D.N. and Stevens, K.N. 1976. "Computer-aided speech training for the deaf". Journal of Speech and Hearing Disorders 41.120-132.

Nickerson, R.S. 1975. "Characteristics of the speech of deaf persons". The Volta Review. 77.342-362.

Nickerson, R.S., Kalilow, D.N. and Stevens, K.N. 1976. "Computer-aided speech training for the deaf". Journal of Speech and Hearing Disorders 41.120-132.

Ono, H., Okasaki, T., Naki, S., and Harasaki, H. 1982. "Identification of an emphasized consonant of a monosyllable in hearing-impaired subjects and its application to a hearing aid". Journal of the Acoustical Society of America 71.S58.

Osberberger, M.J. and McGarr, N.S. 19 "Speech Production Characteristics of the Hearing Impaired". Speech and Language, Advances in Basic Research and Practice 8. 221-283.

Owens, E. and Schubert, E.D. 1977. "Development of the California Consonant Test". Journal of Speech and Hearing Research. 20.463-474.

Parkhurst, B. and Levitt, H. 1978. "The effect of selected prosodic errors on the intelligibility of deaf speech". Journal of Communication Disorders. 11.249-256.

Pichney, M.A. 1981. "Speaking clearly for the hard of hearing". Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Pichney, M.A., Durlach, N.I. and Braida, L.D. 1985 "Speaking clearly for the hard of hearing: I:Intelligibility differences between clear and conversational speech". Journal of Speech and Hearing Research 28.96-103.

Pichney, M.A., Durlach, N.I. and Braida, L.D. 1986 "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech". Journal of Speech and Hearing Research 29.434-446.

Pickett, J.M. 1956."Effects of vocal force on the intelligibility of speech sounds". Journal of the Acoustical Society of America. 28. 5.902-905.

Pindzola, R.H. and Cain, B.H. 1988."Acceptability ratings of tracheoesophageal speech". Laryngoscope 98.394-397.

Pittman, J. 1987. "Discrimination of five voice qualities and prediction to perceptual ratings". Phonetica 44.38-49.

Pittman, J. 1987. "The long-term spectral measurement of voice quality as a social and personality marker: a review". Language and Speech 30 Part I. 1-12.

Prater, R.J. and Swift, R.W. 1984. Manual of Voice Therapy. Little, Brown Co.: Boston.

Price, P.J. and Levitt, A.G. 1983. "The relative roles of syntax and prosody in the perception of the /s/-/c/ distinction". Language and Speech 26, 3.

Ramig, L.A., Mead, C.L., Scherer, R.C., Horii, Y., Larson, K. and Kohler, D.1988a. "Voice therapy and Parkinson's disease:a longitudinal study of efficacy". A paper presented at the Clinical Dysarthria Conference, San Diego.

Ramig, L.A., Mead. C.L. and DeSanto, L. 1988b. "Voice therapy and Parkinson's disease. American Speech and Hearing Association 30/10:128.

Ramig, L.A., Mead, C.L. and Winholtz, W. 1988c. "Speech therapy:Parkinson's disease". Produced by the Recording and Research Center of the Denver Center for the Performing Arts.

Ramig, L.A., Fazoli, K., Scherer R. and Bonitati, C. 1990. "Changes in phonation of Parkinson's disease patients following voice therapy". A paper presented at The Clinical Dysarthria Conference. San Antonio, TX.

Ramig, L.A. and Scherer, R. 1989. "Speech therapy for neurological disorders of the larynx". Neurological Disorders of the Larynx ed. by A. Blitzer, C.Sasaki, S. Fahn, M.Brin and K.Harris New York:Thieme.

Resnick,S.B., Dubno, J.R., Huffnung, S. and Levitt, H. 1975. "Phoneme errors on a nonsense syllable test". Journal of the Acoustical Society of America 58, S114.

Robbins, J. 1984."Acoustic differentiation of laryngeal, esophageal, and tracheoesophageal speech". Journal of Speech and Hearing Research 27.577-585.

Robbins, J., Christensen, J. and Kempster, G. 1986. "Characteristics of speech production after tracheoesophageal puncture: voice onset time and vowel duration". Journal of Speech and Hearing Research 29.499-504.

Robbins, J.,Fisher, H.B., Blom, E.D., and Singer, M.I. 1984. "A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production". Journal of Speech and Hearing Disorders 49.202-210.

Robertson, S. and Thompson, F. 1984. "Speech therapy in Parkinson's disease: a study of the efficacy and long-term effect of intensive treatment". British Journal of Disorders of Communication 19.213-224.

Rosenbek, J.C. and LaPointe, L.L. 1985. "The Dysarthrias: description, diagnosis and treatment". Clinical Management of Neurogenic Communicative Disorders. ed. by D. Johns, 97-152. Boston: Little, Brown and Co.

Ross, E.D., Anderson, B. and Morgan-Fisher, A, 1989 "Crossed aprosodia in strongly dextral patients". Archives of Neurology 46 (2).206-209.

Rostolland,D. and Parant, C. 1974. "Physical analysis of shouted voice". A paper presented at the Eighth International Congress on Acoustics, London.

Rostolland, D. 1982 a. "Acoustic features and shouted voice". Acustica 50.118-125.

Rostolland, D. 1982b. "Phonetic structure of shouted voice". Acustica 51, 80-89.

Rubow, RT and Swift, E. 1985. "Microcomputer-based wearable biofeedback device to improve treatment carry-over in parkinsonian dysarthria". Journal of Speech and Hearing Disorders 50.178-185.

Ruscello, D.M., Lass, N.J. and Podbesek, J. 1988. "Listeners' perceptions of normal and voice-disordered children". Folia Phoniatrica 40.290-296.

Ryalls, John 1986. "What constitutes a primary disturbance of speech prosody? A reply to Shapiro and Danly". Brain and Language 29.183-187 .

Ryalls, J., Joanette, Y. and Feldman, L. 1987. "An acoustic comparison of normal and right-hemisphere-damaged speech prosody". Cortex 23.685-694.

Samar, V.J. and Metz, D.E. 1988. "Criterion validity of speech intelligibility rating-scale procedure for the hearing-impaired population". Journal of Speech and Hearing Research 31.307-316.

Sarno, M. 1968. "Speech impairment in Parkinson's disease". Archives of Physical Medicine and Rehabilitation 49.269-275.

Scarpino, J. and Weinberg, B. 1981. "Junctural contrasts in esophageal and normal speech". Journal of Speech and Hearing Research 24.120-126.

Scherer, K.R. 1986. "Vocal affect expression: a review and a model for future research". Psychological Bulletin 99.143-165.

Scherer, K.R. 1979. "Personality markers in speech". Social Markers in Speech ed. by K.R. Scherer and Giles, Cambridge:Cambridge University Press.

Scherer, R.C., Gould, W.J., Titze, I.R., Meyers, A. and Sataloff, R. 1988. "Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis". Journal of Voice 2.230-244.

Scherer, R.C. and Vail, V.J. 1988. "Measures of laryngeal adduction". Journal of the Acoustical Society of America 84 (S1), S81(A).

Schiavetti, N., Metz, D.E. and Sitler, R.W. 1981. "Construct validity of direct magnitude estimation and interval scaling of speech intelligibility: evidence from a study of the hearing impaired". Journal of Speech and Hearing Research 24.441-445.

Scott, S. and Caird, F. 1983. "Speech therapy for Parkinson's disease". Journal of Neurology, Neurosurgery and Psychiatry 46.140-144.

Schiffman, S., Reynolds, M.L. and Young, F.W. 1981. Introduction to Multidimensional Scaling: Theory, Methods and Applications. New York:Academic Press.

Scuri, D. 1935. "Restirazione e fonazione nei sordomute", In Nickerson, R.S. 1975. "Characteristics of the speech of deaf persons". The Volta Review 342-362.

Shapiro, B. and Danly, M.1985. "The role of the right hemisphere in the control of speech prosody in propositional and affective contexts". Brain and Language 25.19-36.

Shipp, T. 1967. "Frequency, duration, and perceptual measures in relation to, judgments of alaryngeal speech acceptability". Journal of Speech and Hearing Research 10.417-427.

Simpson, M.B., Till, J.A. and Goff, A.M. 1988. "Long-term treatment of severe dysarthria: a case study". Journal of Speech and Hearing Disorders 53.433-440 .

Singer, M.I. and Blom E.D. 1981. "Selective myotomy for voice restoration after total laryngectomee". Archives of Otolaryngology 107:670.

Smith, C.R. "Residual hearing and speech production in deaf children". Unpublished doctoral dissertation, City University of New York, 1972.

Smith, C.R. 1975."Residual hearing and speech production in deaf children". Journal of Speech and Hearing Research 18.795-811.

Spector, P.B., Subtelney, J.D., Whitehead,R.L. and Wirz, S.L. 1979. "Description and evaluation of a training program to reduce vocal tension in adult deaf speakers". The Volta Review 81.2.81-90.

Stark, R.E. and Levitt, H. 1974. "Prosodic feature perception and production in deaf children". Journal of the Acoustical Society of America 55, A.

Stevens, K.N., Nickerson, R.S. and Rollins, A.M. 1983. "Suprasegmental and postural aspects of speech production and their effect on articulatory skills and intelligibility". Speech of the Hearing Impaired: Research, Training and Personnel Preparation ed. by I. Hochberg, H.Levitt, M.J. Osberger, M.J., 35-51. Baltimore:University Park Press.

Stevens, K.N., R.S. Nickerson, and A.M. Rollins 1978. "On describing the suprasegmental properties of the speech of deaf children". Bolt Beranek and Newman Report No. 3955, November.

Summers, W.V., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I. and Stokes, M.A. 1988. "Effects of noise on speech production: acoustic and perceptual analyses". Journal of the Acoustical Society of America. 84/3.917-928.

Stromberg, H. and Levitt, H. 1979. "Multiple linear regression analysis of errors of deaf speeech". A paper presented at the Acoustical Society of America, Cambridge, MA.

Sundberg, J. 1977. "The acoustics of the singing voice". Scientific American 234.82-91.

Sundberg, J., Ternstrom, S., Perkins, W. and Gramming, P. 1988. "Long-term average spectrum analysis of phonatory effects of noise and filtered auditory feedback". Journal of Phonetics 16.203-219.

Takahashi, H. and Koike, Y. 1975. "Some perceptual dimensions and acoustic correlates of pathologic voices". Acta Oto-Laryngologica Supplement 338.1-24.

Tardy-Mitzell, S., Andrews, M.L., and Bowman, S. 1985. "Acceptability and intelligibility of tracheoesophageal speech". Archives of Otolaryngology. 111.213-215.

Ternstrom, S. and Sundberg, J. 1984. Acoustical aspects of choir singing. Acoustics for choir and orchestra ed. by S. Ternstrom, 12-22. Stockholm:Royal Swedish Academy of Music.

Tikofsky, R.S. 1970. "A revised list for the estimation of dysarthric single word intelligibility". Journal of Speech and Hearing Research 13. 59-64.

Tikofsky, R.S. and Tikofsky, R.P. 1964. "Intelligibility measures of dysarthric speech". Journal of Speech and Hearing Research 7.325-333.

Till, J.A. and Toye, A.R. 1988." Acoustic phonetic effects of two types of verbal feedback in dysarthric subjects". Journal of Speech and Hearing Disorders 53. 449-458.

Titze, I.R. 1984. "Parameterization of glottal area, glottal flow, and vocal fold contact area". Journal of the Acoustical Society of America 75.570-580.

Tolhurst, G.C. 1949 "Audibility of the voiceless consonants as a function of intensity". Journal of Speech and Hearing Disorders 14. 210-215.

Tolhurst, G.C. 1955. "The effect on intelligibility scores of specific instructions regarding talking" (USAM Report #NM001 064 01 35). Pensacola, FL: Naval Air Station.

Tolhurst, G.C. 1957. "Effects of duration and articulation changes on intelligibility, word reception, and listener preference". Journal of Speech and Hearing Disorders 22. 328.

Trudgill, P. 1974. The social differentiation of English in Norwich. Cambridge University Press: Cambridge.

Trudeau, M.D. , Fox, R.A. and Fornataro, L.M., 1988. "Use of the tracheostoma valve in the marking of contrastive stress and sentence intonation". Journal of Communication Disorders 21. 21-31.

Turner, James Clifford. 1977. Voice and Speech in the Theatre. (Third Edition, revised by Malcolm Morrison) London:Pittmen, Ltd..

Turner, C.W. and Robb, M.P. 1987. "Audibility and recognition of stop consonants in normal and hearing-impaired subjects". Journal of the Acoustical Society of America 81.1566-1573.

Voelker, C. 1938. "An experimental study of the comparative rate of utterance of deaf and normal hearing speakers". American Annals of the Deaf 83.274-284.

Voiers, W.D. 1964. "Perceptual bases of speaker identity". Journal of the Acoustical Society of America 36 .1065-1073.

Weinberg, B. 1988. "Speech management in the alaryngeal patient". Decision Making in Speech-Language Pathology ed. by D.E.Yoder and R.D.Kent, 148-149. Philadelphia. Decker.

Weinberg. B. and Bennett, S. 1973. "Acceptability ratings of normal, esophageal, and artificial larynx speech". Journal of Speech and Hearing Research 16.608-615.

Weinberg, B., Horii, Y., Blom, E. and Singer, M. 1982. "Airway resistance during esophageal phonation". Journal of Speech and Hearing Disorders 47.194-199.

Weismer, G. 1984. "Articulatory characteristics of Parkinsonian dysarthria". The Dysarthrias: Physiology-Acoustic-Perception-Management. ed. by M.R. McNeil, J.C. Rosenbek and A. Aronson, 101-130. San Diego:College Hill Press.

Weismer, G. 1990. "Keynote Address" A paper presented at the Clinical Dysarthria Conference. January, San Antonio.

Weismer, G., Kent, R., Hodge, M. and Martin, R. 1989. "The acoustic signature for intelligibility test words". Journal of the Acoustical Society of America.

Williams, S.E. and Watson, J.B. 1985. "Differences in speaking proficiencies in three laryngectomee groups". Archives of Otolaryngology. 111.216-219.

Yorkston, K. 1988. "Prosody in the adult". Decision Making in Speech-Language Pathology. ed. by D.E. Yoder and R.D. Kent, 140-143. Philadelphia: Decker.

Yorkston, K.M. and Beukelman, D.R. 1978. "A comparison of techniques for measuring intelligibility of dysarthric speech". Journal of Communication Disorders 11.499-512.

Yorkston, K.M. and Beukelman, D.R. 1981. "Ataxic dysarthria: treatment sequences based on intelligibility and prosodic considerations" Journal of Speech and Hearing Disorders 46.398-404.

Yorkston, K.M. and Beukelman, D.R. 1981. Assessment of Intelligibility of Dysarthric Speech. C.C. Publications: Tigard, OR.

Yorkston. K.M., Beukelman, D.R. and Bell, K.R. 1988. Clinical Management of Dysarthric Speakers. Little, Brown and Company: Boston.

Yorkston, K, Beukelman, D., Minifie, F. and Sapir, S. 1984. "Assessment of stress patterning". The dysarthrias: physiology, acoustics, perception, management. ed. by M. McNeil, J. Rosenbek and A. Aronson, 131-162. SanDiego: College-Hill Press.

Yumoto, E., Gould, W. and Baer, T. 1982. "Harmonics-to-noise ratio as an index of the degree of hoarseness". Journal of Acoustical Society of America 71. 1544-1550.

Youdelman,K., MacEachron, M. and McGarr, N.S. 1989. "Using visual and tactile sensory aids to remediate monotone voices in hearing-impaired speakers". The Volta Review 91(4).197-208.