NCVS Status and Progress Report
Volume 5/November 1993

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

Editorial and Distribution Information

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

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

Correspondence should be addressed as follows:

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

Primary Sponsorship

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

Other Sponsorship

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

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

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

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

Memphis State University
Department of Audiology and Speech Pathology

The University of Illinois
Department of Speech and Hearing Science

NCVS Personnel

Administration

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

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

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

Investigators, Affiliates and Support Staff

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

Doctoral Students

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

Postdoctoral Fellows

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

Visiting Scholars

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

Advisory Board

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

Contents

Forward

This 5th Status and Progress Report comes at a time when we are all thinking of the long term future of the National Center for Voice and Speech. Many things are changing right now in the health science arena. All of us are familiar, of course, with the day to day developments of President Clinton's health care package. None of us know exactly what the impact will be on the long term welfare of biomedical research, but we are all hoping that NIH funding will continue at a healthy pace. Last year the NIDCD budget was reduced from its previous year for the first time ever, which made it very difficult to obtain new grant applications. Recent announcements indicate, however, that there may be an increase in the NIDCD budget of approximately 5 percent for the coming year. This is good news, given the very tight fiscal policies that Congress is working under these days.

One thing is certain - each year we have to become more clever as researchers. The Advisory Board to NIDCD wishes us to get deeper into the molecular structure of all the organs of the human body involved in speech communication. At the same time, they wish us to understand the whole body as a system. Yet it is becoming more and more difficult to do invasive procedures, either on humans or on animals. This means that the critical data that we all need have to come from very carefully conducted experiments, those that have a high benefit to risk (or cost) ratio. On the one hand, we need large numbers of human subjects or animals to make our results statistically reliable; on the other hand, we need to conserve and protect humans and animals involved in research. This puts all of the pressure on the experimenter to obtain only those pieces of information that are absolutely vital and then to integrate the fragments in the most clever ways. I hope that our research will show that trend.

From a publication standpoint, this 5th Status and Progress Report has been arranged in a two-column format to be a little more compatible with typical journal papers. I am extremely proud of our staff here at the Iowa office for spending time and effort to make the report readable and appealing to you all. Special thanks go to Julie Lemke, Julie Ostrem, and Marty Milder who have contributed substantively to the success of the reports.

Ingo R. Titze, Director
November, 1993

Part I

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 5
November, 1993, 1-10

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

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

Abstract

The purpose of this investigation was to study the operating range of the levator veli palatini muscle for a nonspeech task (blowing) and to determine wherein that range levator activity for speech lies. Ten adult subjects without speech or velopharyngeal abnormalities participated. Levator EMG activity for speech occurred in the lower region of the total range for blowing. In two subsequent experiments involving a subset of four subjects, it was found that overall effort may have had a small effect on levator activity apart from its role in velopharyngeal closure for aerodynamic purposes. The results of the main experiment are discussed in relation to the concept of threshold of fatigue as it may influence velopharyngeal control mechanisms.

Certain muscles that are used for speech production appear to have the capacity to produce forces that far exceed those necessary or typically used for speech. Examples of such muscles include the masseter and temporalis used for jaw elevation (Folkins, 1981). It has been reported that levels of lip muscle force used for speech are only about 10-20% of maximum lip forces attainable (Barlow & Abbs, 1983). With regard to respiratory activity afforded by the powerful abdominal and chest wall muscles, humans are capable of generating airway pressures much greater than those typically used for speech. Cook, Mead, and Orzalesi (1964) found that a group of normal young adult females and males generated average maximum pulmonary pressures of 146 and 237 cm H_2O respectively. Kent, Kent, and Rosenbek (1987) reviewed additional studies reporting similar values. These values are many times greater than the approximately 6-10 cm H_2O typically achieved during speech.

Playing a wind instrument may require a great deal more intraoral air pressure than that for speech. Bouheys (1977) indicated that trumpet players often expend up to 150 cm H_2O when playing tones that are high in frequency and intensity. Bless, Ewanowski, and Dibble (1983) reported air leakage through the velopharyngeal port for 10 subjects playing wind instruments who sought help for this problem. Only three of the subjects exhibited air leakage for speech, although only after extended practice on their musical instruments. It appears that in the subjects described by Bless et al. (1983), velopharyngeal muscle activity was sufficient for the pressure demands of speech but not for the higher pressure demands associated with playing a musical instrument. It is possible that for certain individuals, such as those with a history of cleft palate, muscle strength may not even be sufficient to meet the pressure demands for speech.

In discussing maximum performance tests of speech production, Kent et al. (1987) noted that a reduced reserve capacity can impair a speaker's "flexibility" and might also render speech a taxing activity for that individual. Although the authors discussed global activities such as maximum expiratory pressures without addressing underlying anatomic and physiologic mechanisms, it is likely that considerations of prevailing level of activity in relation to reserve capacity apply to individual muscles as well as whole systems. That is, using a muscle at or near its maximum activation level for speech may be taxing for that individual muscle and may render movement of the structure, of which the muscle is a part, an unduly effortful process.

It is not known where in its total operating range each individual muscle that influences the airway functions during speech. We were interested in studying the operating range for the levator veli palatini muscle and to determine wherein that range its activity for speech lies. We were motivated by a therapy procedure that we are utilizing to strengthen the muscles of velopharyngeal closure (Kuehn, 1991).

The therapy procedure described by Kuehn (1991) involves introducing a positive air pressure into the nasal cavities using a method commonly referred to as CPAP (continuous positive airway pressure) that is used to treat patients with sleep apnea. The positive air pressure provides a resistance against which the velopharyngeal closure muscles must work. We have shown in a previous study (Kuehn, Moon, and Folkins, 1993), that levator muscle activity increases with heightened nasal cavity air pressure in subjects with and without cleft palate. However, we did not determine in that study the level of levator activity for speech in relation to its total range for either group of subjects.

Given that blowing requires an airtight velopharyngeal seal regardless of the level of intraoroal air pressure generated, we chose that task in an attempt to activate the levator muscle over its widest operating range. The data from normal subjects pertaining to levator operating ranges obtained in the current study will be used as a basis for comparison with similar information obtained from individuals with velopharyngeal abnormalities in a follow-up study.

Three experiments involving normal subjects were designed to answer the following questions:

1) What is the activity range for the levator veli palatini muscle for a nonspeech task involving blowing?

2) What is the activity range for the levator veli palatini muscle for speech in relation to the range for the blowing task?

3) What is the activity range for two control muscles, the masseter and sternocleidomastoideus, during the blowing task?

4) What is the activity range for the levator veli palatini and one control muscle, the sternocleidomastoideus, during a speech loudness task?

Method

Subjects

The subjects for Experiment 1 were five men and five women in the third, fourth, or fifth age decade. The subjects exhibited normal oral/nasal resonance balance and reported no history of speech, language, or hearing disorders. This was verified by the investigators at the time of testing. Four of the subjects, two male (S1 and S2) and two female (S3 and S4), participated in a second and third experiment in addition to Experiment 1.

Experiment 1

The tasks for Experiment 1 are summarized in Table 1 and identified more explicitly in Table 2. Both the speech and nonspeech tasks were intended to elicit a range of activity for the levator veli palatini muscle, from minimal to maximal levels. Subjects produced 10 repetitions of the speech and blowing tasks and several repetitions of the voluntary velar elevation and swallowing activities.

The subject was seated upright in a dental chair and the oral cavity was sprayed lightly with 4% lidocaine topical anesthetic. Stainless steel wire electrodes, 110 μm in diameter, were used for recording EMG activity from the levator muscle. The wires were inserted perorally into the muscle using 1/2 inch 30 gauge hypodermic needles. The needles were inserted at an angle following the course of the levator muscle, that is, in a superior, lateral, and posterior direction. The two wires for bipolar recording were placed approximately 4 mm apart and 10 mm deep into the levator muscle on the subject's right side. Placement criteria included EMG activity that was observed in association with sustained [s] production. The EMG signals were amplified using Biocommunications Electronics preamplifiers (model 301) and amplifiers (model 205).

During the blowing tasks, a segment of a polyethylene tube, 15 cm in length with a 1.77 mm inner diameter and 2.80 mm outer diameter, was inserted into the mouth to serve as a shunt in parallel with the pressure-sensing tube of the same diameters. The shunt tube enabled constant and predetermined oral pressures to be generated. Intraoral air pressure was sensed with a Honeywell Microswitch pressure transducer (model 162PC01D), amplified with a Biocommunications Electronics amplifier (model 205),

Table 1.
Summary of tasks for Experiment 1.
Tasks are identified by letter in Table 2.

SPEECH

"Say _____ again."

[m]	[sis]
[mam]	[sus]
[mim]	[pʌ]

NONSPEECH

1. Blowing: Intraoral air pressure values in cm H_2O

5	40
10	50
20	60
30	max

2. Voluntary Velar Elevation

3. Swallowing on Command

Table 2.
Key for identification of tasks in Experiment 1 for which levator veli palatini muscle activity was measured. Letters a-r represent the 18 speech tasks measured. Letters s-z represent the blowing tasks, letter A is the voluntary velar elevation task, and B is the swallowing task.

SPEECH

Peak activity for [s] in "say" in:

a) say [m] again	d) say [sis] again
b) say [mam] again	e) say [sus] again
c) say [mim] again	f) say [pʌ] again

Prevailing activity level for the utterance:

g) ... [m] ...	j) ... [sis] ...
h) ... [mam] ...	k) ... [sus] ...
i) ... [mim] ...	l) ... [pʌ] ...

Peak activity for [g] in "again" in:

m) say [m] again	p) say [sis] again
n) say [mam] again	q) say [sus] again
o) say [mim] again	r) say [pʌ] again

NONSPEECH

Blowing. Intraoral air pressures in cm H_2O

s) 5	w) 40
t) 10	x) 50
u) 20	y) 60
v) 30	z) max

A. Voluntary velar elevation on command.

B. Swallowing water from a cup.

and displayed on one channel of an oscilloscope (Tektronix model 2214).

The subject was shown the horizontal axis on the oscilloscope for which each target pressure coincided. He or she was instructed to maintain the target pressure by keeping the oscilloscope beam on the appropriate horizontal axis while blowing on the shunt tube. Target pressures were elicited in the order 10, 5, 20, 30, 40, 50, 60 cm H_2O, and the maximum pressure that the subject could generate. The maximum pressure varied across subjects and was not controlled.

After the blowing tasks, subjects produced the speech tasks listed in Tables 1 and 2. The speech samples were produced in the carrier phrase "say _____ again." Sequencing of the speech tasks was randomized across subjects. The audio signal from a dynamic microphone was amplified using a Nakamichi preamplifier and Tascam tape recorder (model 22-4).

Subjects also were asked to swallow water from a cup at various times throughout the experiment. The experiment concluded by eliciting each subject's voluntary velar elevations. Subjects were provided with audio feedback of the amplified levator veli palatini electromyographic interference patterns to assist them in voluntary velar elevations. Two (S8 and S10) of the ten subjects could not perform voluntary velar elevations even with such feedback.

Generalized Physiologic Effort

A subset of four subjects performed two additional tasks in two subsequent experiments to estimate the possible effects of overall effort on levator muscle activity apart from the functional demands on the muscle to close the velopharyngeal orifice for aerodynamic purposes. It was reasoned that overall effort, especially at higher intraoral air pressure levels, conceivably could elevate activity of muscles in the head and neck region even if those muscles were not immediately bordering the airway. For example, those muscles could provide a stabilizing force that might increase activity level linearly with increases in intraoral air pressure. Thus, we wanted to account for the possibility that increases in levator activity might be due predominantly to overall physiologic effort thereby giving an otherwise false impression of its activity in relation to velopharyngeal closure for the blowing task. Experiments 2 and 3 were conducted for that purpose.

Experiment 2

Activation levels of the masseter and sternocleidomastoideus muscles were sampled. The masseter was chosen because it can move the jaw during speech although it is not an "obligatory" muscle. That is, people can produce speech with the jaw immobilized (Lindblom, Lubker, & Gay, 1979). In a similar fashion, the sternocleidomastoideus is often regarded as an "accessory" muscle of inhalation (Hixon, 1973) and may therefore assist in respiratory activity for speech, but generally is not regarded as "obligatory" for that purpose. Moreover, neither muscle immediately borders the airway. Therefore, they were felt to be suitable as neutral muscles of the head and neck for the purpose of assessing the effects of the range of effort associated with the blowing task.

Pairs of surface electrodes (Beckman Ag-AgCl 11 mm diameter disks) were attached with adhesive collars to the skin overlying the masseter and sternocleidomastoideus muscles of each of the four subjects. Proper placement of these electrodes was assessed by having the subjects clench the teeth for masseter and to turn the head to the opposite side for sternocleidomastoideus.

The session began with two separate recordings, first with the subject clenching his or her teeth with maximum effort and then rotating the head maximally to the side opposite the sternocleidomastoideus muscle from which the recording was obtained. This provided maximum EMG activity levels for the two muscles against which the blowing activities could be compared. The subjects then

blew through the shunt tube as in Experiment 1 but only at target values in the sequence 10,30,50 cm H_2O and maximum effort. Each task was repeated 10 times. All instrumentation and data collection procedures were the same as that for Experiment 1.

Experiment 3

Experiment 3 involved the same subset of four subjects. These subjects produced vowels at three different loudness levels. Although increased loudness is associated with greater overall effort, for example in respiratory drive, it is not accompanied by greater intraoral air pressure because vowels are produced with an unoccluded oral cavity. Therefore, any increase in levator veli palatini activity with loudness would logically be attributable to some aspect or aspects of physiologic effort, but not to increased intraoral air pressure demands.

Two muscles, the levator and sternocleidomastoideus, were sampled. The latter muscle was included again as a neutral muscle as described for Experiment 2. Hooked wire electrodes were inserted into the levator veli palatini muscle on each subject's right side. Surface electrodes were affixed to the skin overlying the left sternocleidomastoideus muscle.

The subjects sustained each of the vowels [i,a,u] at a normal loudness level, louder than normal, and at their loudest level. They were instructed to produce the loudest vowels that they could, but without pain or strain. The sequence of the nine tasks (3 vowels X 3 loudness levels) was random across subjects.

To ensure that the subjects were in fact increasing their output level across loudness conditions, a sound pressure level meter (Bruel & Kjaer, 2209, set to the A scale) was used to measure the level of their vowel productions. The meter microphone was placed 2 1/2 ft from the subject's mouth. All other instrumentation and data collection procedures were the same as that for Experiments 1 and 2.

Data Analysis

EMG activity, intraoral air pressure, and audio signals were monitored on an oscilloscope (Tektronix model 5111A) and recorded on a Sony digital instrumentation recorder (model PC108M). Subsequently, EMG signals were full-wave rectified and smoothed with a 40 ms time constant. Intraoral air pressure signals also were smoothed with a 40 ms time constant. Rectified and smoothed EMG signals, smoothed pressure signals, and the audio signal were digitized at 1000 samples per second using a laboratory computer and commercially available analog-to-digital conversion software. Data then were displayed and analyzed using custom graphics and analysis routines.

Experiment 1

Figure 1 shows an example of rectified and smoothed levator veli palatini EMG activity and intraoral air pressure trace during a blowing task. A similar EMG trace was observed during voluntary elevations. For each of these tasks, average EMG activity within a 1-sec segment characterized by relatively stable EMG was chosen to represent prevailing levator EMG activity.

An example of measures obtained from a speech task is presented in Figure 2. Three measures of activation level for levator were obtained: 1) peak level for [s] in the carrier word "say," 2) prevailing level during the target utterance, and 3) peak level for [g] in the carrier word "again." For swallowing, peak activation levels were recorded.

Levator EMG activation levels were normalized within each subject. The largest peak EMG value recorded during the blowing tasks by a subject was used as a reference for that subject and was assigned a value of 100%. All other EMG values recorded for that subject were referenced to the maximum value.

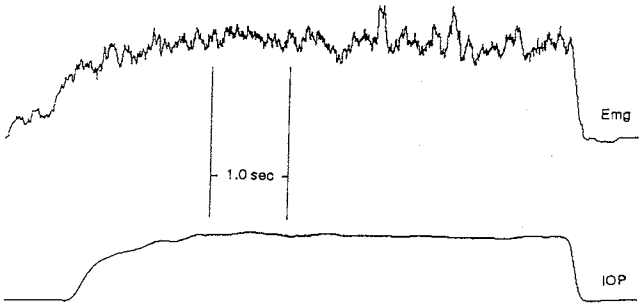


Figure 1. Rectified and smoothed EMG activity from the levator veli palatini muscle (top) and intraoral air pressure trace (bottom) for a blowing task. An interval of 1.0 sec in the mid portion of each token of the blowing tasks was sampled and analyzed as a representative measure of EMG activity for that token.

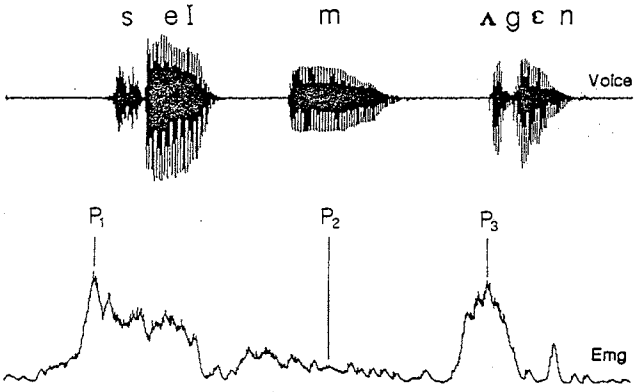


Figure 2. Audio trace (top) and rectified and smoothed EMG activity from the levator veli palatini muscle (bottom). P1 = peak EMG for [s] in the carrier word "say." P2 = prevailing EMG level during the target word in the carrier phrase. P3 = peak EMG for [g] in the carrier word "again."

A mixed model analysis of variance with one random factor (subjects) and one fixed factor (task) was used to assess the effects of the various speech and nonspeech tasks on levels of levator muscle activity. Designed a priori comparisons were made between the blowing, speech, voluntary elevation, and swallowing tasks. The mean differences among these task groups were estimated and tested against zero using an alpha level of 0.05. Finally, multiple comparisons among the eight blowing tasks, 18 speech tasks, voluntary elevation, and swallowing were performed using the Scheffe procedure.

Experiment 2

Data analysis for Experiment 2 was similar to that employed for Experiment 1. Normalization within subjects was conducted separately for each muscle (masseter and sternocleidomastoideus). Activation levels recorded during maximal effort tasks (teeth clenching for masseter and head turning for sternocleidomastoideus), were used as reference levels. As in Experiment 1, average EMG activity within a 1-sec segment characterized by relatively stable EMG activity was recorded. In instances for which no readily identifiable EMG activity could be detected, a 1-sec segment characterized by stable intraoral air pressure was chosen.

A mixed model analysis of variance with one random factor (subjects) and one fixed factor (pressure level) was used to assess the effects of pressure level on muscle activation level. Separate analyses were conducted for the masseter and sternocleidomastoideus muscles. Post hoc analyses involved Bonferroni multiple comparisons.

Experiment 3

For each vowel prolongation, average activation levels were measured for 1-sec segments characterized by relatively stable levator and sternocleidomastoideus EMG activity. Maximal activation tasks were not recorded for these muscles in this experiment. Therefore, EMG values recorded in Experiment 3 were not normalized and are reported in arbitrary units.

A mixed model analysis of variance was used to assess the effects of vowel loudness on activation levels of the levator and sternocleidomastoideus muscles. A separate analysis was conducted for each vowel. The model included vowel and loudness-within-vowel as fixed factors. Random factors included subjects, subject-by-vowel, and subject-by-loudness-within-vowel interactions. Bonferroni multiple comparisons were used to assess muscle activation differences across loudness levels within each vowel.

Results

Experiment 1

Figures 3 and 4 show 95% confidence intervals representing levator veli palatini muscle activation levels for the speech and nonspeech tasks for the 10 subjects.

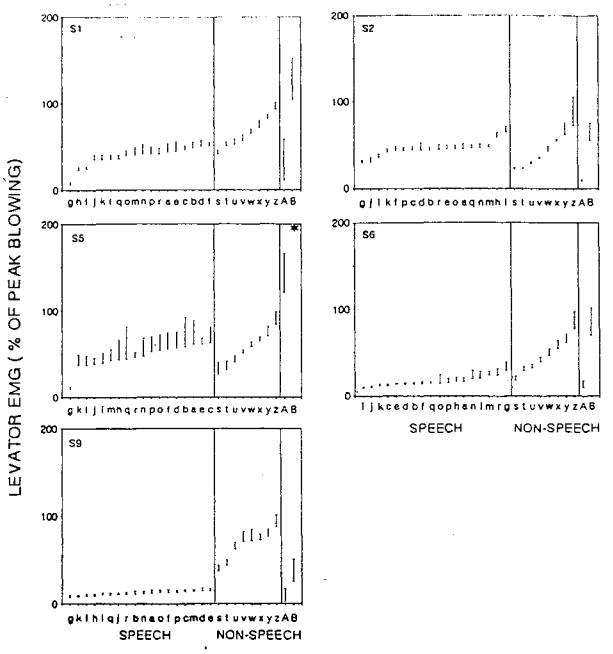


*Figure 3. Levator veli palatini EMG activity (95% confidence intervals) for speech and nonspeech tasks; male subjects. Tasks are identified by letter in Table 2. * Subject 5 activity for swallowing exceeded 200% of that for his overall peak blowing activity*

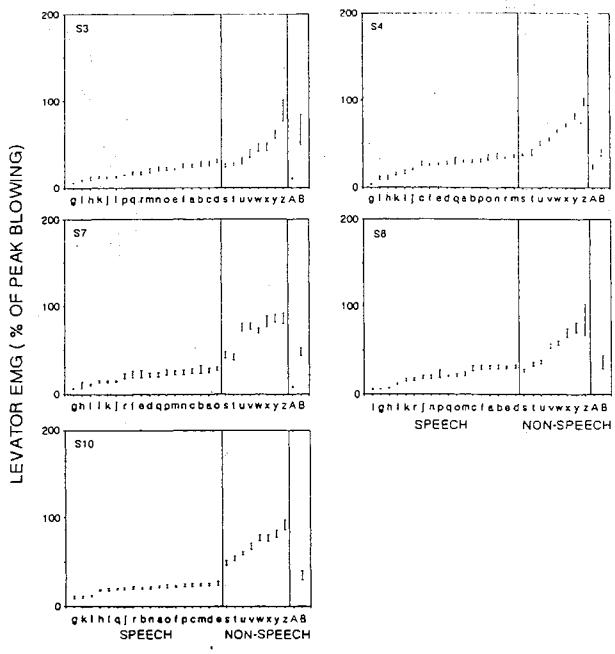


Figure 4. Levator veli palatini EMG activity (95% confidence intervals) for speech and nonspeech tasks; female subjects. Tasks are identified by letter in Table 2.

Speech tasks are identified by the letters a-r, blowing by the letters s-z, voluntary velar elevation by the letter A, and swallowing by the letter B (see Table 2).

EMG measures for the speech tasks are arranged according to increasing levels of levator EMG activity within subjects. Although the exact sequence varied across subjects, the nasal contexts are generally nearer the lower end and the stop and fricative contexts are generally at the higher end of the activity range for speech as expected.

The levator activity range across the speech tasks for each subject appears to be rather continuously variable without obvious discontinuities in the function for most subjects. Possible exceptions to this general statement might be for Subjects 1 and 5. The speech function for Subject 1 appears to be somewhat trimodal with the nasal consonant [m] (represented by the letter g in Figure 3) at the lowest level, vowels within the nasal context (letters h and i) somewhat higher, and all of the other sounds higher still. Subject 5 shows an abrupt change in levator EMG activity for the nasal consonant (letter g) versus all other speech sounds for which levator activity is higher. Although not a major focus of this study, the data for speech in Figures 3 and 4 are more consistent with a velopharyngeal mechanism that is under continuously variable control (Kent, Carney, & Severeid, 1974; Lubker, 1975) rather than binary control (Moll & Daniloff, 1971; Moll & Shriner, 1967).

For the blowing task, levator activity increased in a monotonic fashion for most subjects from the lowest intraoral air pressure generated (represented by letter s in Figures 3 and 4) to the highest intraoral air pressure (letter z). Table 3 shows the results of the analysis of variance comparing blowing tasks for the data grouped across subjects. For most comparisons, an increase in intraoral pressure was associated with a significant ($p < 0.05$) increase in levator activity. For example, levator activity was significantly greater at 20 cm H_2O compared to 5 cm H_2O, greater at 30 cm H_2O compared to 10 cm H_2O, etc.

Overall, significantly ($p < 0.0001$) greater levels of levator EMG activity were observed during blowing than during speech. Table 4 shows the results of the individual comparisons for blowing versus speech tasks. Across subjects, levator activity was significantly ($p < 0.05$) greater for blowing tasks at and above 20 cm H_2O compared to all 18 speech tasks. For the blowing tasks at 5 and 10 cm H_2O, levator activity was significantly ($p < 0.05$) greater than that for speech for 5 and 6, respectively, of the 18 speech tasks.

Table 3.
Results of analysis of variance comparing levator muscle activity for the blowing tasks. * $p < 0.05$. s-z = intraoral air pressure values in cm H_2O at 5, 10, 20, 30, 40, 50, 60, maximum respectively.

	s	t	u	v	w	x	y	z
s	---		*	*	*	*	*	*
t		---		*	*	*	*	*
u			---		*	*	*	*
v				---		*	*	*
w					---	*	*	*
x						---		*
y							---	*

Table 4.
Results of analysis of variance comparing levator muscle activity for blowing tasks to that for speech tasks across subjects. Ratios in right column indicate number of significant differences ($p < 0.05$) for each blowing task compared to the 18 speech tasks. Blowing tasks are expressed in cm H2O of intraoral air pressure. Speech tasks are identified in Table 2.

Blowing Task	Significant versus Speech Tasks
5	5/18
10	6/18
20	18/18
30	18/18
40	18/18
50	18/18
60	18/18
max	18/18

Figures 3 and 4 also show the 95% confidence intervals for levator activity associated with swallowing and voluntary velar elevation compared to that for the speech and blowing tasks. Across subjects, significantly ($p < 0.0001$) greater levels of levator EMG were observed during swallowing than during all three of the other tasks, speech, blowing, and voluntary velar elevation. However, this effect was due primarily to the much greater levels associated with swallowing for S1 and S5. Levator activity levels for swallowing were less than that for blowing for several of the other subjects. Overall, significantly ($p < 0.0001$) less levator activity was observed during voluntary velar elevation than during blowing, but there was not a significant difference between voluntary elevation and speech.

Experiment 2

Figure 5 provides information about whether effort level in the subset of four subjects had an affect on muscles in the head and neck region that do not directly border the vocal tract. The figure shows normalized mean levels of EMG activity for the masseter and sternocleidomastoideus muscles as a function of intraoral air pressures generated in the blowing tasks. The EMG levels are expressed as percentages of the maximum level of activity within each muscle as determined by maximal teeth clenching and head turning maneuvers.

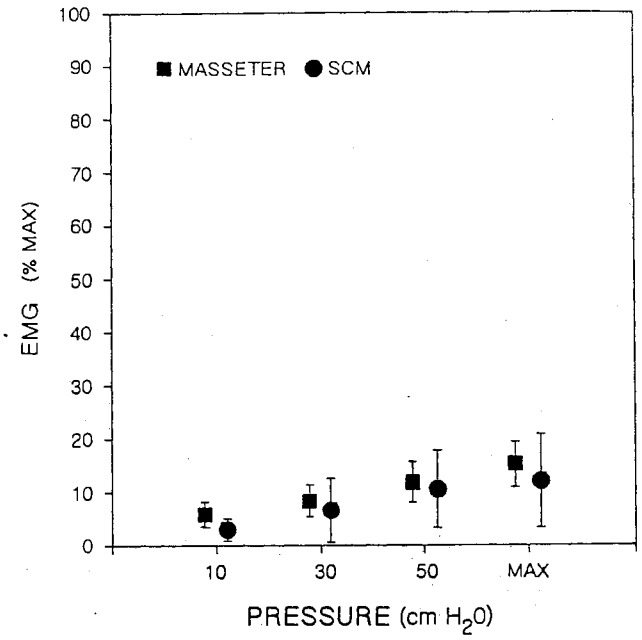


Figure 5. Mean and standard deviation values for masseter and sternocleidomastoideus (SCM) EMG activity levels versus intraoral air pressure levels.

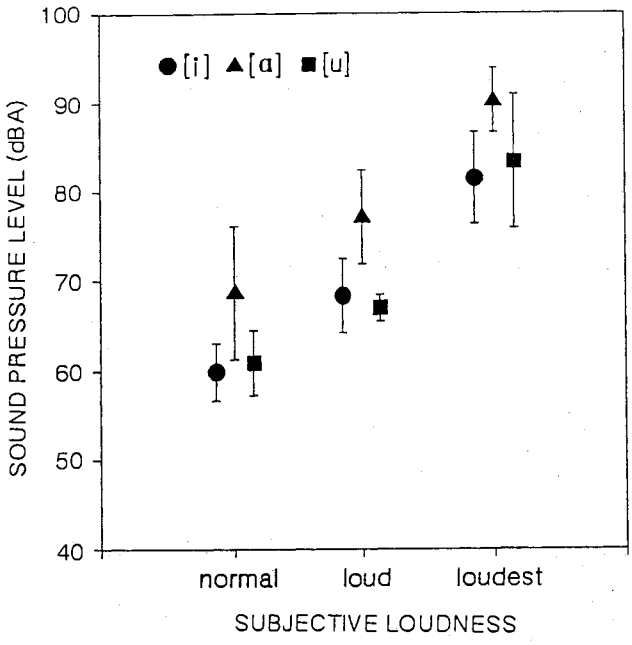


Figure 6. Mean and standard deviation values for sound pressure level in dB (A scale) versus subjective loudness levels for the three vowels [i,a,u].

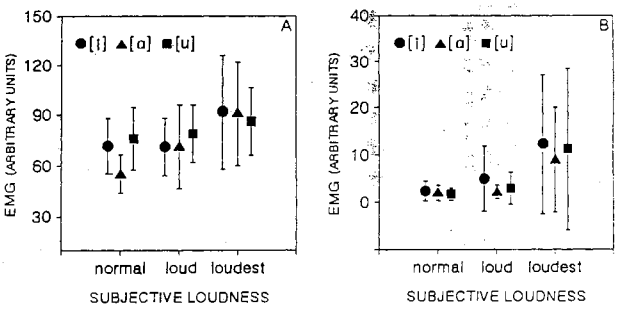


Figure 7. Mean and standard deviation values for EMG activity level in arbitrary units versus subjective loudness levels for the three vowels [i,a,u]. A. Levator veli palatini muscle. B. Sternocleidomastoideus muscle.

For the masseter muscle, comparisons for EMG activity between any two intraoral air pressure conditions were significantly ($p < 0.0001$) different from each other. For the sternocleimastoideus muscle, four of the six paired comparisons between pressure conditions were significantly ($p < 0.008$) different from each other. Sternocleidomastoideus muscle activity for comparisons between 30 and 50 cm H_2O versus maximum intraoral air pressure were not significantly different. Thus, most of the paired comparisons across pressure conditions for the two muscles were significantly different from each other suggesting a general increase in muscle activity with increased blowing effort. However, relative levels of activity remained at a low level (below 25% of maximum) for both muscles even at the highest levels of intraoral air pressures generated.

Experiment 3

Figures 6 and 7 show the results of the vowel loudness experiment for the subset of four subjects. Figure 6 shows that the subjects did increase sound pressure levels in association with their subjective increases in vowel loudness. All comparisons across loudness levels were significantly ($p < 0.0001$) different from each other.

EMG activity levels for the levator muscle are shown in Figure 7A and those for the sternocleidomastoideus muscle are shown in Figure 7B in relation to the loudness changes. Across loudness levels for the levator muscle, only the comparison involving the vowel [a] for the normal versus the loudest production reached statistical significance ($p < 0.006$). For the sternocleidomastoideus muscle, none of the comparisons across loudness levels reached statistical significance for any of the three vowels. Thus, in general, increases in loudness did not have a strong effect on activation levels of the two muscles examined.

Discussion

The primary purpose of this study was to obtain information about the range of levator veli palatini muscle activity in normal speakers and to determine wherein that

range the activity for speech lies. It was assumed that utilizing a nonspeech task such as blowing, which requires tight velopharyngeal closure, would activate the levator muscle over its widest range. Although we were primarily interested in levator activity related to velopharyngeal functioning, we wanted to account for the possibility that overall physiologic effort might influence levator activity apart from its more direct role in relation to the control of intraoral air pressure and the direction of the airstream during speech. Experiments 2 and 3 were conducted for that purpose.

The results of Experiments 2 and 3 suggest that overall effort may have had some small effect on levator activity apart from aerodynamic demands. This follows from the increases observed in the masseter and sternocleidomastoideus muscles with increases in intraoral air pressure in Experiment 2. However, the activity observed in these muscles during the blowing task was of a very low magnitude. Moreover, increases in loudness did not have a strong effect on activation levels of either the levator or the sternocleidomastoideus muscles in Experiment 3. Therefore, we conclude that the variability attributable to overall effort is minimal and that most of the variance in levator activity is related directly to its role in providing closure of the velopharyngeal port.

A major finding in this study was that levator muscle activity for speech tended to occur in the lower region of the total range for blowing. Across subjects, all levator activity levels for blowing at or above 20 cm H_2O intraoral air pressure were greater than levator activity levels observed during speech tasks. These results are interesting in view of the fact that intraoral air pressure needs for normal conversational speech are generally below 20 cm H_2O.

These results may have different explanations depending on the type of neuromuscular control acting on the velopharyngeal mechanism. For example, the increases in levator activity in the blowing task could be related to reflexive activity. In this fashion, following initial velar elevation, the levator muscle could be functioning in a largely reactive manner. This appears unlikely, however, because levator activity rose at the onset rather than following intraoral air pressure changes and remained fairly constant throughout each individual blowing maneuver (see Figure 1). Therefore, although we cannot be certain about a cause-effect relation, it appears more likely that increases in levator activity in relation to blowing with different levels of intraoral air pressure are planned by the motor mechanism and do not rely on reflexive control.

Regardless of whether reflexive control or other more automatic peripheral adjustments occurred, the levator activity levels for speech in relation to the total range for blowing suggests a relatively low effort on the part of the levator muscle during speech. This is consistent with the general concept that normal speech does not require a great deal of effort. Yet, it is possible that because speech and nonspeech tasks are qualitatively different, they draw upon different neuromuscular control mechanisms. Thus, there could be different maxima for levator activity for different tasks.

For nonrepetive, nonsustained activity such as swallowing, maximal activation of the levator muscle may be a reasonable strategy, as observed in some subjects in this study, to ensure the tightest velopharyngeal closure for each swallow to prevent nasal regurgitation. However, for repetive activity such as speech, or sustained activity such as blowing, functioning at maximum level would appear to be a poor strategy because of the likelihood of fatigue. Although the distribution of muscle fiber types in the normal human adult velopharyngeal muscles is not known, it is likely that there is a mixture of Type I and Type II fibers present and the muscles would be susceptible to fatigue owing to the Type II fibers (Johnson, Polgar, Weightman, & Appleton, 1973). Thus, to prevent fatigue, it appears parsimonious for the muscle to function nearer the lower end of its operating range for repetitive and sustained activities. Also, it is possible that the structure that is being moved, in this case the velum, may be apt to reach its intended target more consistently and in a timelier fashion than if the underlying muscle is overtaxed to the level approaching fatigue.

Mundale (1970), in a study involving hand grip strength and fatigue, found that fatigue (decrease in force generating capacity) was clearly evident with hand grip maneuvers at 20% of maximum force with periods of relaxation alternating with periods of muscle contraction. Above 20% of maximum force, the duration of each intermittent contraction was important for total endurance (resistance to fatigue) but under 20%, the duration of each contraction was less important.

Bystrom and Kilbom (1990) provided additional information about the interaction between "intensity" (force generation) and muscle contraction time. They also measured handgrip force and included EMG recording of the extensor digitorum communis as one index of handgrip fatigue. They defined "local fatigue" on the basis of combined measures of blood flow in the forearm, EMG, and subjective ratings. They found that at continuous contractions of 10%, 25%, and 40% of maximum voluntary contraction (MVC), local fatigue in the forearm was evident. Intermittent exercises at 10% MVC with 2, 5, or 10 sec of relaxation alternating with 10 sec of contraction did not lead to fatigue, nor did 5 or 10 sec of relaxation alternating with 10 sec of contraction at 25% MVC. They found the threshold of fatigue to be 16.7% MVC which was the product of the time and intensity ratio. For example, with contraction time of 7 sec and relaxation of 3 sec (7/10=.7) versus an intensity of 20% MVC, the product of

these values equals 14% (0.7 X 20%) and is below the fatigue threshold of 16.7% in their study.

Obviously, the exact numbers as stated above, if they are indeed valid, would vary depending on the muscles involved and probably many other factors as well. Robin, Goel, Somodi, & Luschei (1992) observed a difference in endurance for tongue pressure against an air-filled bulb in trumpet players and high school debaters compared to control subjects without training in those activities. Sustained tongue pressures were significantly longer at 25% and 50% of maximum pressure for the experimental subjects. Robin et al. suggested that possible exercise-related changes in the proportion of fatigue-resistant muscle fibers brought about by trumpet playing and competitive debate may have led to the observed differences in endurance times.

The notion of a threshold fatigue effect, depending on both time and force in relation to velopharyngeal control, appears worthy of exploration. A common anecdotal report by clinicians is that patients often have the capability of producing a sustained [s] or isolated words with no evidence of velopharyngeal incompetency but that the competency breaks down in connected speech, especially "casual" speech. It is possible that such individuals may not necessarily be within a fatigue state but rather have developed a pattern of velopharygeal control to avoid a fatigue state that might occur very rapidly in the presence of increased or more sustained muscle force generation. Perhaps because of weaker velopharyngeal muscles, these individuals may have a lower threshold of fatigue than individuals having normal strength and they may have developed a pattern of neuromotor control to remain below the fatigue threshold. In a recent study, Warren, Dalston, and Mayo (1993) concluded that if the velopharyngeal port is open for an inappropriately long time interval compared to normal, hypernasal speech is likely to result. Their results and conclusion fit well with the concept expressed here with regard to duration of velopharyngeal closure as affected by a fatigue threshold. That is, one way of alleviating fatigue in velopharyngeal closure muscles is to avoid excessive opposition to gravity and other forces that naturally tend to open the velopharyngeal port.

Future studies are needed to help elucidate the possible beneficial effects of resistance exercises such as CPAP therapy (Kuehn, 1991) and other therapeutic techniques designed to strengthen the velopharyngeal musculature and possibly reduce fatigue effects. In speakers exhibiting hypernasality, a reserve capacity for the levator muscle may exist above that used for speech, as shown for people with normal speech in this study. It may be possible for such individuals to tap into that reserve capacity with proper training procedures during speech. Also, speakers exhibiting hypernasality may have lower thresholds of fatigue because of weaker velopharyngeal closure muscles and other factors. It may be possible to raise the threshold of fatigue thereby utilizing a more suitable range of activity for normal speech purposes.

The current study will be extended using subjects with abnormal velopharyngeal mechanisms to determine wherein their total operating range levator activity for speech lies. We also intend to explore the notion of velopharyngeal muscle fatigue thresholds in subject groups with normal and abnormal velopharyngeal mechanisms.

Acknowledgments

This study was supported in part by PHS Research Grants DC-00976, DC-00085, and DC-01015 from the National Institute on Deafness and Other Communication Disorders and DE-10436 from the National Institute of Dental Research. We would like to acknowledge the statistical support of the Biostatistics Core of the National Center for Voice and Speech and Wendy Edwards for her assistance in data analysis and graphics.

References

Barlow, S.M. & Abbs, J.H.(1983). Force transducers for the evaluation of labial, lingual, and mandibular motor impairments. *Journal of Speech and Hearing Research,26,*616-621.

Bless, D.M., Ewanowski, S.J., & Dibble, D.(1983). Velopharyngeal valving problems related to extensive playing of wind instruments. *XIX Congress of the International Association of Logopaedics and Phoniatrics,* University of Edinburgh, Scotland, U.K.

Bouhuys, A.(1977). *The physiology of breathing* (p. 279). New York: Grune & Stratton.

Bystrom, S.E.G. & Kilbom, A.(1990). Physiological response in the forearm during and after isometric intermittent handgrip. *European Journal of Applied Physiology,60,*457-466.

Cook, C.D., Mead, J., & Orzalesi, M.M.(1964). Static volume-pressure characteristics of the respiratory system during maximal efforts. *Journal of Applied Physiology,19,*1016-1022.

Folkins, J.W.(1981). Muscle activity for jaw closing during speech. *Journal of Speech and Hearing Research,24,*601-615.

Hixon, T.J.(1973). Respiratory function in speech. In F.D. Minifie, T.J. Hixon, & F. Williams (Eds.), *Normal aspects of speech, hearing, and language* (pp. 73-125). Englewood Cliffs, NJ:Prentice-Hall, Inc.

Johnson, M.A., Polgar, J., Weightman, D., & Appleton, D.(1973). Data on the distribution of fibre types in thirty-six human muscles: an autopsy study. *Journal of the Neurological Sciences, 18,* 111-129.

Kent, R.D., Carney, P.J., & Severeid, L.R.(1974). Velar movement and timing: evaluation of a model for binary control. *Journal of Speech and Hearing Research, 17,* 470-488.

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

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

Kuehn, D.P., Moon, J.B., & Folkins, J.W.(1993). Levator veli palatini muscle activity in relation to intranasal air pressure variation. *Cleft Palate-Craniofacial Journal, 30,* 1-8.

Lindblom, B.E.F., Lubker, J.F., & Gay, T.(1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics, 7,* 147-161.

Lubker, J.F.(1975). Normal velopharyngeal function in speech. *Clinics in Plastic Surgery, 2,* 249-259.

Moll, K.L. & Daniloff, R.G.(1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America, 50,* 678-684.

Moll, K.L. & Shriner, T.H.(1967). Preliminary investigation of a new concept of velar activity during speech. *Cleft Palate Journal, 4,* 58-69.

Mundale, M.(1970). The relationship of intermittent isometric exercise to fatigue of handgrip. *Archives of Physical Medicine and Rehabilitation, 51,* 532-539.

Robin, D.A., Goel, A., Somodl, L.B., & Luschei, E.S.(1992). Tongue strength and endurance: relation to highly skilled movements. *Journal of Speech and Hearing Research, 35,* 1239-1245.

Warren, D.W., Dalston, R.M., & Mayo, R.(1993). Hypernasality in the presence of "adequate" velopharyngeal closure. *Cleft Palate-Craniofacial Journal, 30,* 150-154.

NCVS Status and Progress Report - 5
November, 1993, 11-18

Coordination of Velopharyngeal Muscle Activity During Positioning of the Soft Palate

Jerald B. Moon, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Alice E. Smith, M.A.
Department of Speech Pathology and Audiology, The University of Iowa
John W. Folkins, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Jon H. Lemke, Ph.D.
Department of Preventive Medicine, The University of Iowa
Michael Gartlan, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa

Abstract

The relative contributions of the levator veli palatini, palatoglossus, and palatopharyngeus muscles were assessed relative to a range of positions of the velopharynx during production of the vowels [ɑ] and [i] by four normal adult speakers. The results indicate that velopharyngeal positioning is determined by the relative contributions of the levator veli palatini, palatoglossus, and palatopharyngeus muscles. There was an increase in coefficients of determination (i.e. amount of closure level variability explained) when activity levels of all three muscles are included in the statistical model compared to activity in any one muscle analyzed independently. Both consistent and inconsistent relationships among activity levels in the three velopharyngeal muscles studied were observed across speaker and vowel produced.

According to Bernstein (1967), a given motor task can be performed in a variety of ways. A change in activity of one structure in a system may induce variations in the activity of other structures to accomplish a desired task. The motor control system imposes constraints on the component structures to simplify the control process. That is, function based interaction rules are established. Fowler, Ruben, Remez, and Turvey (1980) refer to such an interactive system as a coordinative structure. In speech, the components of the coordinative structure might be articulators (e.g. the lip or the jaw) or muscles, and their actions are nested within the overall goal of perceptually adequate speech output.

A number of studies have addressed motor control of the speech articulators within the perspectives of Bernstein (e.g. Folkins and Canty, 1986; Folkins, Linville, Garrett, and Brown, 1988; Gracco, 1988). However, with the exception of a theoretical discussion by Folkins (1985), control of the velopharyngeal mechanism has not been considered within this context. Within the velopharyngeal mechanism, a coordinative structure might be indicated by interactions among the velum and lateral pharyngeal walls, or interactions among activation levels of the velar muscles. On a muscular level, the levator veli palatini, palatoglossus, and palatopharyngeus might be thought of as a coordinative structure. While levator veli palatini muscle activity is associated with velar elevation gestures, Fritzell (1969) noted that the extent of velar elevation and magnitude of levator veli palatini muscle activity was not always highly correlated. Similarly, Kuehn, Folkins, and Cutting (1982) found in a study of oralized vowels and fricatives that "levels of levator muscle activity independent of other muscle activity were not directly related to velar position" (p. 30). Kuehn et al. suggested that a trading relationship might exist among the levator veli palatini, palatoglossus, and palatopharyngeus muscles in positioning the velum.

They postulated further that various combinations of activity in these three muscles might be associated with the same velar position.

Although a speaker may not vary the relative combination of velar and lateral pharyngeal wall movement during multiple repetitions of a speech task, the overall amount of closure used to produce a speech sample will often vary greatly from repetition to repetition. Supposedly, the variation in velopharyngeal opening interacts with changes in the impedance of the oral cavity to produce desired percepts of nasality (Folkins, 1985). There is also great variability in the size and duration of the bursts of electromyographic activity recorded from the velar muscles during speech. Typically, it is not possible to distinguish which aspects of this variability in any one muscle are related to variability in velar opening and which are related to interactions among muscles to produce the same movements.

Moon and Jones (1991) have shown that visual feedback can be used to teach speakers to control and vary the amount of velopharyngeal opening used to produce a speech sample. One advantage of their procedure is that one can ensure that a range of velopharyngeal openings is systematically studied. By manipulating the extent of velopharyngeal opening, one could evaluate the combinations of velopharyngeal muscle activity that may be associated with a given velopharyngeal opening size. While previous investigators have studied activation levels and ascribed roles to individual velopharyngeal muscles, there have been no systematic studies of interactions between the levator veli palatini, palatoglossus, and palatopharyngeus muscles. Such information is important to our understanding of both normal and ultimately disordered velopharyngeal function. This study investigates relative contributions of the levator veli palatini, palatoglossus, and palatopharyngeus muscles in positioning of the velum during speech production.

Methods

Subjects

Four young adults, three females and one male, served as subjects for this investigation. All were judged by the experimenters to have normal resonance balance and articulation. None reported a history of speech, language, or hearing disorders. One subject (subject A) had been trained in singing.

Phototransduction

Velopharyngeal opening and closing gestures were transduced using the phototransducer system described by Dalston (1982). A description of the specific device used in this study is provided by Moon and Jones (1991). The transducer was passed transnasally and positioned with the light emitting fiber below the velopharyngeal port and the light detecting fiber placed above the velopharyngeal port. Phototransducer output was then amplified for each subject to produce 0 volts during rest nasal breathing and a 2 volt deflection during velopharyngeal closure for [i]. Because the phototransducer cannot provide absolute velopharyngeal opening area, the range of closure was denoted as 0% (0 volts) to 100% (2 volts).

Electromyography

Following a light appliaction of topical anesthetic (4% Lidocaine), bipolar hooked wire electrodes were inserted into the levator veli palatini, palatoglossus and palatopharyngeus muscles on each subject's right side. The electrodes were constructed of 110 m stainless steel wire (Medwire 316 SS 3T), and were inserted using half-inch 30 gauge hypodermic needles. Levator veli palatini electrodes were inserted at the dimple of the elevated velum in a posterior, lateral, and superior direction, following the course of the levator muscle. Depth of insertion was approximately 10 mm. Palatoglossus and palatopharyngeus electrodes were inserted into the midportion of the anterior and posterior faucial pillars, respectively. Depth of insertion was approximately 3 mm. Spacing between all electrode pairs was approximately 5 mm. Verification of electrode placement in the levator veli palatini was made during production of [s]. Verification of placements in palatoglossus and palatopharyngeus was made during swallowing. Electrodes were repositioned or reinserted if no electromyographic signal was obtained or if the presence of artifact was observed during placement verification tasks.

Speech Tasks

Each subject was positioned to view a two channel storage oscilloscope. The target velopharyngeal closure level was displayed on one channel. The amplified and low pass filtered (30 Hz) phototransducer output signal was displayed on the second oscilloscope channel. Both the target and phototransducer signals were recorded on a digital instrumentation recorder (Sony PC108M).

Six speech conditions were employed: the vowels [ɑ] and [i] each produced at 25, 50, and 75% closure. Following the procedures developed by Moon and Jones (1991), subjects viewed the phototransducer output on the oscilloscope and attempted to match the target level throughout the duration of each 10 second trial. The subjects were instructed to phonate the vowel normally for approximately 1-2 seconds and then to open the velopharynx to the target level for the remainder of the trial. A minimum of ten attempts were recorded within each condition.

Analyses

Recorded electromyographic and phototransducer signals were digitized using a 10 KHz sampling rate, rectified, smoothed with a 40 ms time constant, and downsampled by a factor of eight. Within each trial, 625 ms

segments characterized by a stable phototransducer signal were extracted for analysis. The 625 ms time window was chosen to ensure that electromyographic activity was also relatively stable. Segments were chosen regardless of whether or not the target closure level was attained. That is, the targets were utilized only to elicit a range of closure levels and accuracy of target matching was not a measured variable in this study. The extracted segments were then analyzed to determine relative velopharyngeal closure level in the 0 to 100% range and the average corresponding activity level in each of the three velopharyngeal muscles.

Separate analyses were performed for each subject and vowel combination. The analyses included univariate and multivariate regression analyses (response surface analyses) of muscle activation level as a function of relative closure level using linear, quadratic, and interaction (in multivariate analyses) terms. Prior to analysis, the electromyographic data were normalized within each muscle for each subject. For example, the activation levels for levator veli palatini during production of the vowel [ɑ] by Subject A were converted to percentage of maximum activation using the maximum raw activation level of levator during normal production of that vowel by that subject as the reference.

Results

Univariate Analyses

Subject-specific coefficients of determination (R^2 expressed as a percentage) and mean square errors are shown in Table 1 for closure level regressed against each of the three muscles during production of [ɑ] and [i] by each subject. With three exceptions out of 24, R^2 values for the full model (linear plus quadratic terms) are statistically significant ($p < 0.0001$). However, the coefficients are low in many instances. In 18 of the 24 cases (4 subjects x 3 muscles x 2 vowels) less than 50% of the variability in observed velopharyngeal closure level can be explained by level of activity in any one muscle.

Table 1.
Subject-specific univariate coefficients of determination (R^2 expressed as a percentage) and mean square errors for % closure level regressed against each muscle during production of [ɑ] and [i]

	Vowel [ɑ]		Vowel [i]	
Subject	R^2	MSE	R^2	MSE
	LEVATOR			
A	30.38	430.85	22.08	96.94
B	61.47	161.80	81.19	116.58
C	46.44	540.57	27.97	505.38
D	69.87	91.39	1.59*	101.22
	PALATOGLOSSUS			
A	86.94	80.32	16.37	103.73
B	54.66	190.41	16.41	518.02
C	42.20	583.45	23.46	537.06
D	31.74	207.01	32.74	69.17
	PALATOPHARYNGEUS			
A	54.99	278.50	35.49	80.01
B	16.82	349.20	1.71*	609.16
C	0.67*	1002.60	11.69	619.62
D	40.58	180.20	34.88	66.97

* NS; $p > 0.0001$

Examples of electromyographic data associated with the best and worst coefficient of determination for the levator veli palatini during production of [i] are shown in Figure 1. These examples are from Subject B and Subject D. It is clear from the top panel that a relationship between levator activity and closure level exists. The bottom panel does not display a clearly identifiable relationship. It is also apparent from Figure 1 that subjects did not demonstrate the same range of closure levels. That is, while targets of 25, 50 and 75% closure were presented to the subjects, some (e.g. Subject D) had difficulty reaching the 25% level consistently.

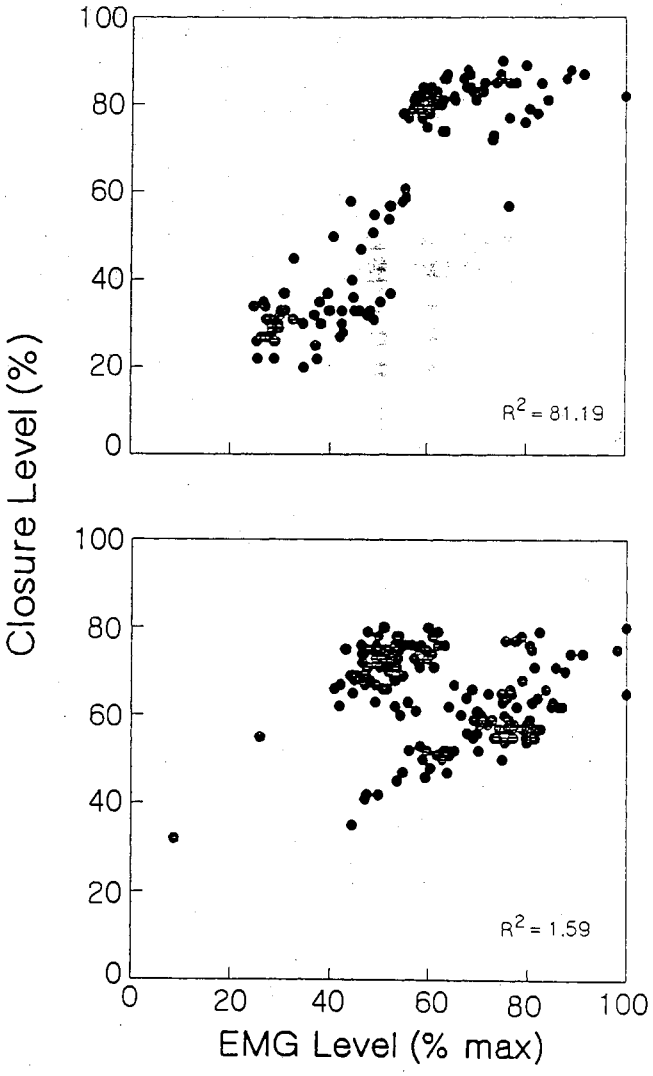


Figure 1. Data sets resulting in best (upper, Subject B) and worst (lower, Subject D) coefficients of determination for levator veli palatini during [i] production.

Multivariate Analyses

Subject-specific coefficients of determination (R^2 expressed as a percentage) for the full multivariate model are presented in Table 2. In addition, the contributions to the coefficients of determination corresponding to the partial F-tests for the set of linear effects, the set of quadratic effects, and the set of two-way interaction effects are listed. Coefficients of determination for the full model are significant ($p < 0.0001$) for each subject. The contributions of the linear terms are significant in all eight cases (2 vowels X 4 subjects). The contributions of the quadratic terms to the model are significant in seven of the eight cases. The interaction term is significant in four of the eight cases. Since each set of terms is significant in at least half of the cases, they are all retained in the final model for all subjects regardless of their statistical significance. However, it is clear from Table 2 that the linear term contributes most to the overall coefficient of determination for each subject within each vowel. The exception is the quadratic term for Subject B producing [ɑ]. On average, the quadratic and two-way interaction terms account for only 2.6% of the variability in closure level.

Table 2.

Subject-specific multivariate regressions: components of the coefficients of determination (R^2 expressed as a percentage) and mean square errors for each subject during [ɑ] and [i].

Vowel [a]

Subject	R^2 Linear	Quadratic	Interaction	Total	MSE
A	91.16	0.25*	0.30*	91.71	53.48
B	50.24	22.86	2.10	75.20	105.48
C	76.38	7.85	1.63	85.87	149.80
D	76.90	1.47	0.34*	78.71	64.72

Vowel [i]

Subject	R^2 Linear	Quadratic	Interaction	Total	MSE
A	66.31	4.72	7.10	78.13	29.25
B	78.27	3.05	2.29	83.62	107.61
C	47.58	3.10	2.05*	52.73	346.18
D	61.06	2.26	1.53*	64.84	37.73

* NS; $p > 0.05$

For each subject, the percentage of previously unexplained variability (from best muscle in univariate analysis) captured by the multivariate analysis was assessed. For [ɑ], these values are 36.5%, 35.6%, 73.6%, and 29.34% for subjects A to D, respectively. For [i], the values are 66.1%, 12.9%, 34.4%, and 46.0% for subjects A to D, respectively. Similarly, reductions in mean square error can be observed by comparing the multivariate model mean square errors in Table 2 with the univariate results in Table 1. It is evident that the multivariate mean square error values for any subject-vowel combination are consistently lower than all of the univariate mean square errors for that subject and vowel.

Using inverse distance interpolation, three dimensional mesh plots were generated from relative closure level and electromyographic muscle activity data. Figure 2 shows a three dimensional mesh plot (left) and the individual data points (right) used to derive the mesh plot. Figures 3, 4, 5 and 6 depict three dimensional mesh plots from each subject for the two vowels. Because the interactions among all three muscles and closure level require four dimensions and thus cannot be displayed on a three dimensional graph, each Figure contains the three combinations of muscle pairs.

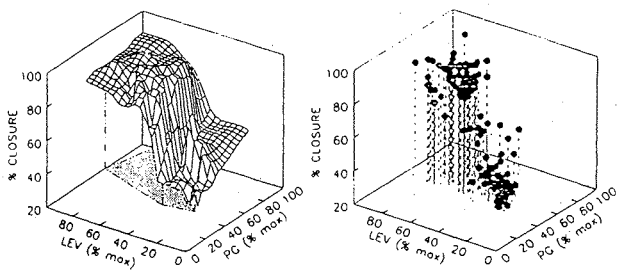


Figure 2. Three dimensional mesh plot of levator veli palatini versus palatoglossus for Subject B producing [i] (left) and individual data points (right) used to derive mesh plot.

Figures 3 through 6 clearly illustrate both similarities and differences in the nature of the influence of each muscle relative to the other muscles as a function of subject and of the vowel being produced. For example, at the top left of Figure 3, essentially unchanging levels of activity in the levator muscle are associated with decreasing levels of palatoglossus activity as closure level increases for production of [ɑ] by Subject A. The bottom graph shows decreases in both palatoglossus and palatopharyngeus with increasing closure level.

In Figure 4, a different activation pattern is observed for [ɑ] (left side) compared to the other subjects. Here, levator, palatoglossus and palatopharyngeus muscle activity all appear to increase as closure level increases. This subject (Subject B) also displays more similarity in activation patterns between the vowels [ɑ] and [i] than the other three subjects. Like Subject A, activation patterns for Subjects C (Figure 5) and D (Figure 6) also differ as a function of vowel produced.

A more systematic interpretation of muscle interactions displayed graphically in Figures 3 through 6 was accomplished using the signs of each of the parameter estimates (linear, quadratic, and two-way interaction) of the multivariate model generated for each subject and each vowel. Linear terms include LEV (levator), PG (palatoglossus), and PP (palatopharyngeus). Quadratic terms include LEV^2, PG^2, and PP^2. Two-way interaction terms include LEV*PG, LEV*PP, and PG*PP. Regardless of subject or vowel, parameter LEV is always positively related to velopharyngeal closure level. Of additional

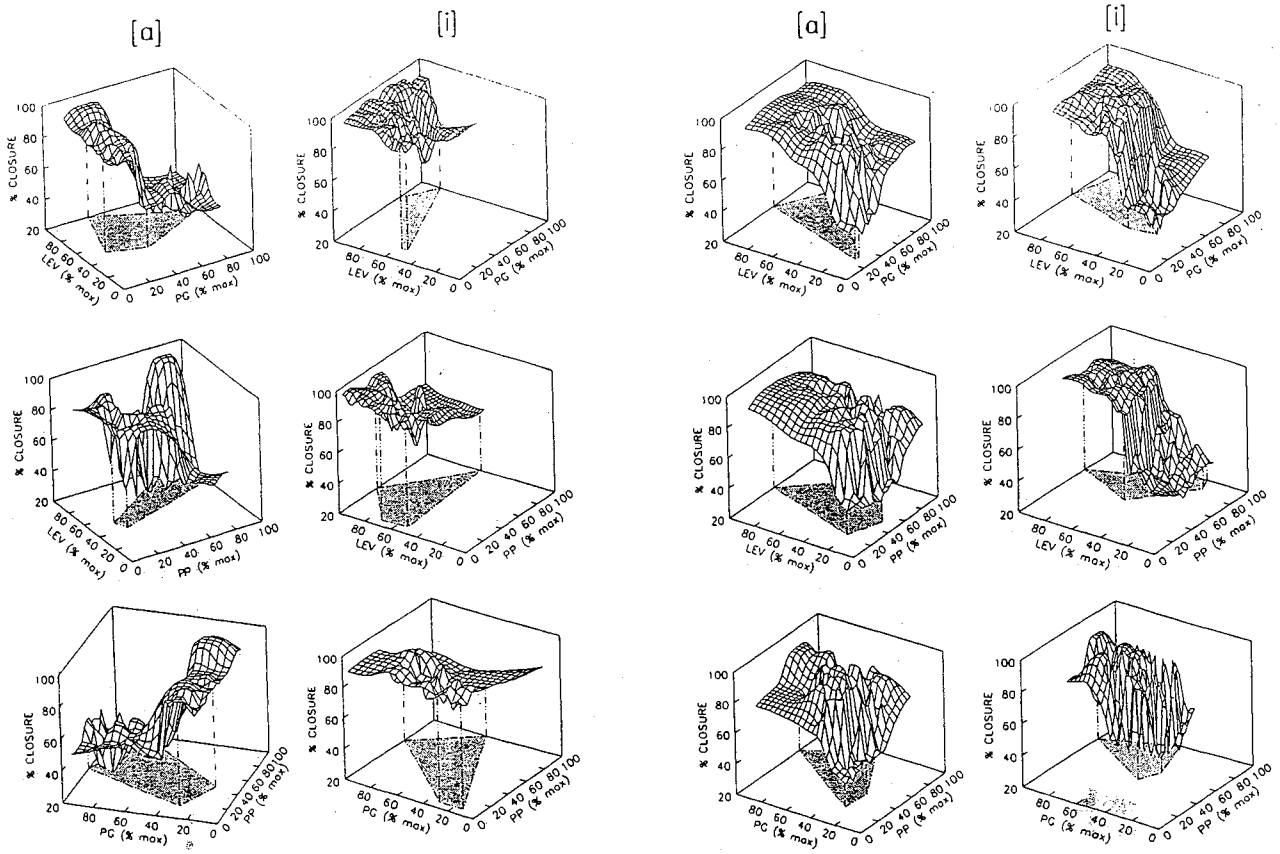


Figure 3 - left. Three dimensional mesh plots for Subject A producing the vowels [a] and [i]. Shaded areas indicate minimum and maximum recorded EMG levels for each muscle. Interpolated mesh plots extend to EMG minima and maxima throughout 0 - 100% max EMG range for each muscle. Figure 4 - right. Three dimensional mesh plots for Subject B producing the vowels [a] and [i]. Shaded areas indicate minimum and maximum recorded EMG levels for each muscle. Interpolated mesh plots extend to EMG minima and maxima throughout 0 - 100% max EMG range for each muscle.

interest is the observation that the LEV^2 parameter in the multivariate model is negatively signed during [i] for three of the four subjects and during [ɑ] by two of the four subjects. The negative sign is indicative of a curvilinear pattern of increasing levator activity with increasing closure level. This pattern is evident in the upper graphs of Figure 4.

The contributions made by the palatoglossus muscle to the multivariate model appear to be vowel specific. In addition, its contribution is more variable than that of the levator muscle. For the vowel [i], the PG parameter is always negatively signed. The negative sign suggests that palatoglossus activity decreases with increasing closure level. The pattern is less obvious for [ɑ]. Two subjects display a positive relationship between palatoglossus activity and closure level (as indicated by a positively signed PG model parameter estimate), while the other two display a negative relationship.

The PG^2 parameter appears to be more variable than LEV^2. For [ɑ], it is negatively signed for two subjects and positively signed for one subject. For [i], it is positively signed for two subjects and negatively signed for one. The negative sign is indicative of a curvilinear pattern of increasing palatoglossus activity with increasing closure level. The positive sign is indicative of a curvilinear decrease in palatoglossus activity with increasing closure level.

The contribution of palatopharyngeus muscle is also important to the multivariate model. For [i], the palatopharyngeus parameter (PP) is negatively related to closure level in all cases. For [ɑ], its relationship to closure level varies across subjects. The PP^2 parameter is negative for two subjects during [ɑ] and positive for one subject.

Discussion

The results of this study indicate that velopharyngeal positioning in space is determined by the relative contributions of the levator, palatoglossus, and palatopharyngeus muscles. This is evidenced by the dramatic increase in closure level variability explained using the multivariate model over the univariate model. These data support the notions of earlier investigators (Shelton,

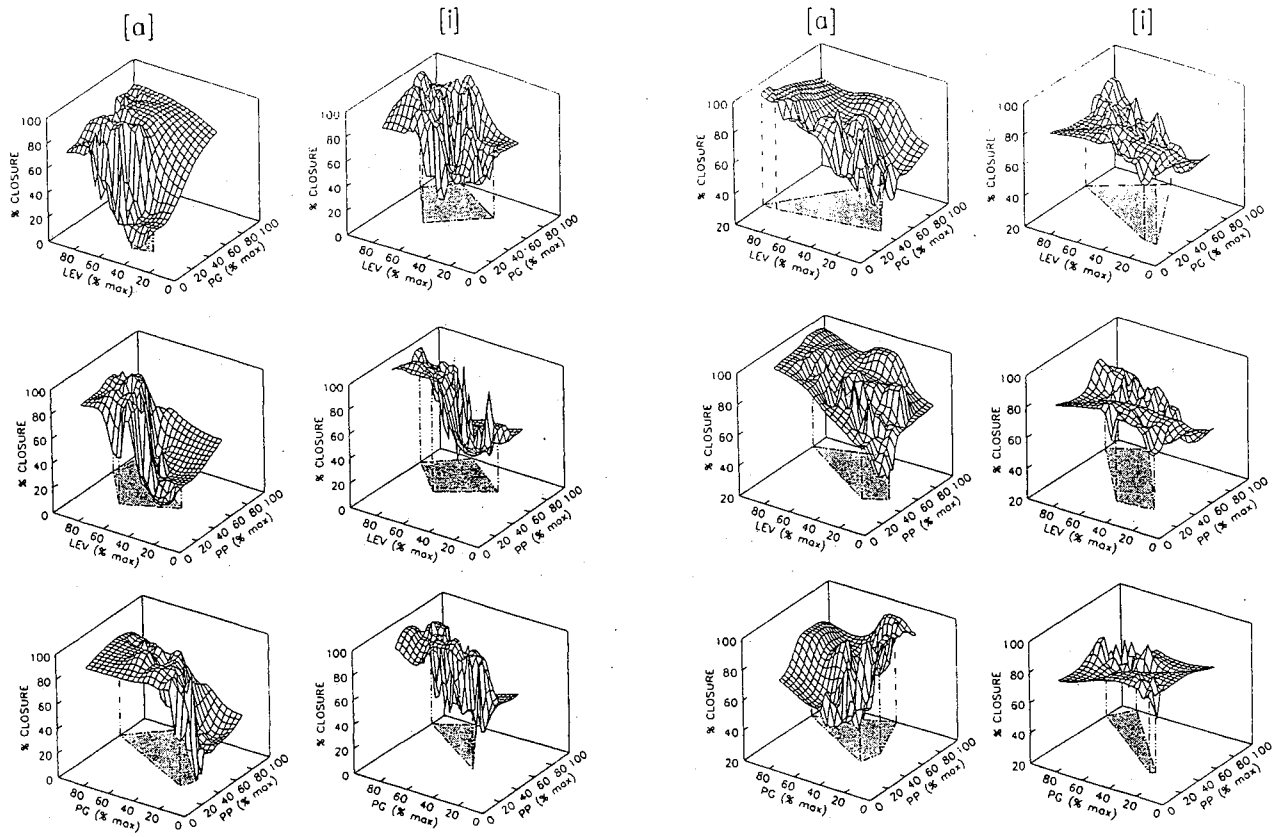


Figure 5 - left. Three dimensional mesh plots for Subject C producing the vowels [a] and [i]. Shaded areas indicate minimum and maximum recorded EMG levels for each muscle. Interpolated mesh plots extend to EMG minima and maxima throughout 0 - 100% max EMG range for each muscle. Figure 6 - right. Three dimensional mesh plots for Subject D producing the vowels [a] and [i]. Shaded areas indicate minimum and maximum recorded EMG levels for each

Harris, Sholes, and Dooley, 1970; Seaver and Kuehn, 1980; Kuehn et al., 1982) that activity levels in more than one velar muscle must be taken into account. The results of the present study cannot, however, be taken as a lack of support for early characterizations (Fritzell, 1969; Bell-Berti, 1976) of the levator muscle as the primary muscle involved in velar elevation. Based on anatomical position alone, velopharyngeal elevation must be greatly influenced by contraction of the levator veli palatini. The results of the present investigation do suggest, however, that the levator muscle performs as one element of a coordinated structure of at least three velopharyngeal muscles. Within this framework, position control of the velopharyngeal mechanism is flexible in that it allows for varying combinations of muscle activation among the constituent muscles.

It is evident from Figures 3, 4, 5 and 6 and the multivariate analysis that muscle interactions vary both across subjects and vowels. The amount of variability in each muscle's contribution to velar position is also evident from Table 1, where univariate coefficients of determination range from less than 1% to 87%. However, the multivariate analyses substantiate the importance of a coordinative structure conceptualization of velar muscle activity for positioning of the velopharyngeal mechanism. The parameter estimates of the multivariate model reveal similarities and differences across both subjects and vowels that may provide insights into the nature of the coordinative structure framework.

Regardless of subject or vowel, the linear parameter estimate for the levator muscle (LEV) was always positively related to velopharyngeal closure level. This should not be surprising given previously published data regarding the role of levator and given our knowledge of its anatomical position. That is, its anatomical position in normal speakers is conducive to velar elevation upon contraction. However, earlier investigators (Fritzell, 1969; Kuehn, Folkins, and Cutting, 1982) suggested that velar position and magnitude of levator muscle activity do not appear to be directly related. This conclusion is supported by the univariate and multivariate analyses conducted in this study.

Our results show that the contributions made by the palatoglossus muscle to the multivariate model appear to be vowel specific. In addition, its contribution is more

variable than that of the levator muscle. During [i], palatoglossus activity was negatively related to closure level for all four speakers. For [ɑ] the relationship was mixed; positive for two speakers and negative for two speakers. These observations raise questions regarding the effects of competing activity (i.e., tongue elevation and pharyngeal wall positioning) on the nature of coordinative structure interactions within the soft palate. Production of [i] requires elevation of the tongue whereas production of [ɑ] does not. Contraction of the palatoglossus muscle to produce upward movement of the back of the tongue for [i] would be expected to produce a downward pull on the soft palate. One might speculate that less palatoglossus activity occurs at higher velopharyngeal closure levels because a) the elevated tongue position can be maintained at least in part by mechanical linkage forces, and/or b) increased palatoglossus activity in association with increased levator activity might move the back of the tongue too high. During [ɑ] however, the consequences of increased palatoglossus muscle activity and its interaction with levator muscle activity are not as important. That is, more variability in tongue back positioning might be tolerated during [ɑ]. There may be support for this notion in Table 1, where univariate coefficients for palatoglossus are substantially lower for [i] for three of the four subjects. Further, three of the four subjects revealed partial F statistic values from the multivariate analysis for the palatoglossus muscle that were much lower for [i] than for [ɑ]. In other words, palatoglossus appears to divide its role during [i], and may be less of a factor for this vowel.

If the palatoglossus muscle has little influence on velar positioning during [i], one must ask whether the other muscles have more influence relative to their performance during [ɑ]. The results of both the univariate and multivariate analyses do not provide convincing evidence of this. The univariate analysis revealed a greater influence of levator during [i] for only one of the four speakers. The partial F statistic values from the multivariate analysis were greater for levator during [i] for only two of the four speakers. However, the interaction between levator and palatoglossus muscle did change as a function of the vowel produced. For [i], the LEV*PG interaction term was always positively related to closure level. During [ɑ] the relationship was again inconsistent. One might interpret the positive LEV*PG term to indicate that the levator and palatoglossus muscles were not working independently during this task. That is, for a given level of closure, the interaction suggests that greater levels of activity were observed in the these muscles than would be expected had the muscles been working separately. Finally, the multivariate coefficients of determination during [i] tended to be lower (with one exception) than those obtained for [ɑ]. That is, less of the variability in closure level could be explained by activation levels of the three muscles and their interactions during [i] than during [ɑ]. This may have been due to the effects of competing activity associated with tongue elevation, or that we have not sampled all of the activity in the three muscles. It might also indicate that activity levels in other muscles (e.g. musculus uvulae, superior constrictor, and perhaps the more transverse fibers of palatopharyngeus) play a role in velar positioning that becomes more important during certain vowels. Of course, we can only speculate on the possible influence of these additional muscles. However, Fritzell (1969) alluded to the possible influence of superior constrictor activity on velopharyngeal closure and the consequence of increased superior constrictor activity on the relative influence of the levator muscle in some subjects. While Kuehn, Folkins, and Cutting (1982) observed that superior constrictor was active for all speech sounds studied, they were unsure whether the magnitudes of activity seen were "sufficient to contribute in a substantial way to the interaction of forces for velar movement" (p. 34). Regarding musculus uvulae, Kuehn, Folkins and Linville (1988) proposed an extensor role. That is, contraction of the musculus uvulae was proposed to exert a compressional force along the top side of the velum which would tend to straighten the curved velum. Such a straightening gesture could be used to modify velar position in a manner unaffected by changes in activity levels of the levator and palatoglossus musculature. However, Kuehn et al., (1988) found musculus uvulae activity to be highly correlated with levator veli palatini.

Finally, the contribution of palatopharyngeus muscle to the multivariate model may also be important in the positioning of the velopharyngeal mechanism. Palatopharyngeus activity during [i] was always negatively related to closure level. It is of interest that the nature of palatopharyngeus involvement as a function of vowel tends to parallel that of palatoglossus even though palatopharyngeus would not be expected to be involved in tongue elevation. It is also of interest that, during the production of both [i] and [ɑ], the LEV*PP interaction term in the multivariate model was always negative. In contrast to the positive LEV*PG interaction, this interaction may be interpreted to indicate independent influences of these two muscles on closure level.

The results of this study are in partial agreement with earlier characterizations of the role of palatopharyngeus during speech production. Fritzell (1979) suggested that palatopharyngeus activity was associated with narrowing of the pharynx during [ɑ]. As was the case with palatoglossus activity during tongue elevation for [i], one might characterize narrowing of the pharynx for [ɑ] as a competing activity that would affect the interrelationships among these muscles during the production of [ɑ]. However, we observed that palatopharyngeus was always negatively related to closure level during [i], but its relationship to closure during [ɑ] was variable. It is difficult to explain this observation using the same arguments presented in the case of the palatoglossus muscle. However, it could be argued

that, for these four subjects, palatopharyngeus plays less of a role in velopharyngeal positioning than do levator or palatoglossus.

To summarize, control of position of the velopharyngeal mechanism can be explained using a coordinative structure notion of muscle interaction. Activity levels of individual muscles alone do not account for velar position as well as the combined interactive activation levels of the levator, palatoglossus, and palatopharyngeus. Multivariate models depicting the role of each muscle and their interactions in velopharyngeal positioning during the tasks studied here allow for a description of such interactions. Interactions among the levator veli palatini, palatoglossus, and palatopharyngeus and velopharyngeal closure level were observed in the data that describe the complexity of control of the velopharyngeal mechanism that was once assigned exclusively to the levator veli palatini muscle. It is clear from our data that there may be some consistent relationships between velopharyngeal muscle activity and their interactions and control of the velopharyngeal mechanism. Some of these relations may be associated with anatomic positioning (i.e. levator veli palatini muscle always positively related to velar elevation) while others may be related to the relationship between velopharyngeal movement and movement of other speech articulators (i.e. palatoglossus always negatively related to velopharyngeal positioning during [i]). It is also clear from our data that some variability in muscle activity and their interrelations also exist. Between speaker variability may be due to variations in anatomical positioning of the muscles (Kuehn and Azzam, 1978) affecting relative force levels or simply to motor system variability common to all controlled systems. Additional investigations are underway to further delineate relationships among velopharyngeal muscles during speech production.

Acknowledgement

The authors would like to thank Kice Brown for his assistance with statistical analysis, and David Kuehn for his comments during the preparation of this manuscript.

References

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

Bernstein N. The coordination and regulation of movements. New York:Pergamon Press, 1967.

Dalston R. Photodetector assessment of velopharyngeal activity. Cleft Palate Jour 1982; 19:1-8.

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

Folkins J, Canty J. Movements of the upper and lower lips during speech: interactions between lips with the jaw fixed at different positions. Jour Speech Hear Res 1986; 29:348-356.

Folkins J, Linville R, Garrett J, Brown C. Interactions in the labial musculature during speech. Jour Speech Hear Res 1988; 31:253-264.

Fowler C, Ruben P, Remez R, Turvey M. Implications for speech production of a general theory of action. In: Butterworth B, Ed.. Language Production. New York:Academic Press, 1980.

Fritzell B. The velopharyngeal muscles in speech. Acta-Otolarygol. Suppl 1969; 250:1-81.

Gracco V. Timing factors in the coordination of speech movements. Jour Neuroscience 1988; 8:4628-4639.

Iglesias A, Kuehn D, Morris H. Simultaneous assessment of pharyngeal wall and velar displacement for selected speech sounds. Jour Speech Hear Res 1980; 23:429-446.

Kuehn D, Azzam N. Anatomical characteristics of palatoglossus and the anterior faucial pillar. Cleft Palate Jour 1978; 15:349-359.

Kuehn D, Folkins J, Cutting C. Relationships between muscle activity and velar position. Cleft Palate Jour 1982; 19:25-35.

Kuehn D, Folkins J, Linville R. An electromyographic study of the musculus uvulae. Cleft Palate-Cranio Jour 1988; 25:348-355.

Lubker J. Normal velopharyngeal function in speech. Clin Plast Surg 1975; 2:249-259.

Moon J, Jones D. Motor control of velopharyngeal structures during vowel production. Cleft Palate-Cranio Jour 1991; 28:267-273.

Seaver E, Kuehn D. A cineradiographic and electromyographic investigation of velar positioning in non-nasal speech. Cleft Palate-Cranio Jour 1980; 17:216-226.

Shelton R, Harris K, Sholes G, Dooley P. Study of nonspeech voluntary palate movements by scaling and electromyographic techniques. In:Bosma J, ed. Second symposium on oral sensation and perception. Springfield:CC Thomas, 1970.

NCVS Status and Progress Report - 5
November, 1993, 19-28

Tongue Strength and Endurance in Mild to Moderate Parkinson's Disease

Nancy Pearl Solomon, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Daryl M. Lorell, M.S.
Department of Speech Pathology and Audiology, The University of Iowa
Donald A. Robin, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Robert L. Rodnitzky, M.D.
Department of Neurology, The University of Iowa Hospitals and Clinics
Erich S. Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Weakness and fatigue of the speech production system may contribute to articulatory imprecision and timing difficulties in the speech of people with Parkinson's disease. Nineteen individuals with Parkinson's disease and 19 healthy matched control subjects were tested for strength and endurance of the tongue. Tongue function was evaluated by the Iowa Oral Performance Instrument, a pressure sensing device. In addition, speech was evaluated for articulatory imprecision, overall speech defectiveness, and speech rate. Subjects with Parkinson's disease were found to have lower tongue strength but comparable tongue endurance when compared to matched control subjects. The contributions of peripheral and central processes of fatigue are discussed in light of the present findings. The subject groups did not differ significantly for overall speech defectiveness or interpause speech rate, but the speech articulation of the subjects with Parkinson's disease was perceived as less precise than that of the control subjects. Because most of the subjects in this investigation had perceptibly normal or mildly disordered speech, a potential relationship between tongue function and speech proficiency could not be examined adequately.

A common characteristic of the speech disorder of Parkinson's disease, hypokinetic dysarthria, is imprecision of articulation (Canter, 1965; Chenery, Murdoch, & Ingram, 1988; Darley, Aronson, & Brown, 1969; Ewanowski, 1964; Laszewski, 1956; Logemann, Fisher, Boshes, & Blonsky, 1978; Morrison, Rigrodsky, & Mysak, 1970; Solomon & Hixon, 1993; Tanner, 1976). Logemann, Boshes, and Fisher (1972) reported that articulatory errors were due to inadequate constriction or occlusion of the upper airway. Their analysis revealed a "sequence of articulatory degeneration" from posterior to anterior placements of articulation with disease progression.

Another aspect of speech that can be abnormal in this population is speech rate. Speech rate may be too fast (Hammen, 1990; Hammen, Yorkston, & Beukelman, 1989; Hanson & Metter, 1983; Tanner, 1976; Yorkston, Hammen, Beukelman, & Traynor, 1990), too slow (Anthony & Farquharson, 1975; Boshes, 1966; Kammermeier, 1969; Peacher, 1950), variable (Critchley, 1981; Darley et al., 1969; Ludlow & Bassich, 1983; Metter & Hanson, 1986), accelerating (Critchley, 1981; Hirose et al., 1981; Streifler & Hofman, 1984), or normal (Alp, 1988; Pitcairn, Clemie, Gray, & Pentland, 1990). Individual differences between speakers appear to be the rule (Canter, 1963).

The clinical characteristics of rigidity and bradykinesia have been hypothesized to account for the predominant articulatory and temporal abnormalities in Parkinson's disease (Hunker, Abbs, & Barlow, 1982). Increased background muscle activity in lip muscles has been reported as evidence for rigidity (Hunker et al., 1982; Leanderson, Meyerson, & Persson, 1972; Marquardt, 1973).

Rigidity often is assumed to cause hypokinesia in Parkinson's disease, but this concept has been challenged (Caliguiri, 1987). The presence of bradykinesia in the orofacial system has not been established definitively. Although selective reductions in lip and/or jaw displacement and velocity have been reported in people with Parkinson's disease during speech (Caligiuri, 1987; Conner, Abbs, Cole, & Gracco, 1989; Forrest, Weismer, & Turner, 1989; Hunker et al., 1982), the relationship between displacement amplitude and velocity has been demonstrated to be normal (Forrest et al., 1989). Although the mechanisms for movement difficulties in the orofacial system are unknown, the tongue appears to be affected for speech. Acoustic data indicate that the center frequencies of vowel formants may be abnormal (Tanner, 1976) or the extent and speed of formant transitions are reduced for speakers with Parkinson's disease (Conner, Ludlow, & Schulz, 1989; Forrest et al., 1989).

In addition to the classic motor signs of Parkinson's disease, weakness and fatigue of the speech production system may contribute to articulatory imprecision and timing difficulties. A small number of studies have found reduced tongue strength or endurance to be related to speech disorders in populations other than Parkinson's disease. Children and adults with a variety of articulation and fluency disorders (Palmer & Osborn, 1940) and adults with amyotrophic lateral sclerosis (Dworkin, 1978; Dworkin, Aronson, & Mulder, 1980) have demonstrated lower than normal tongue strength. Children with developmental apraxia of speech were found to have normal tongue strength but reduced tongue endurance (Robin, Somodi, & Luschei, 1991).

Weakness and fatigue have been recognized as common symptoms of Parkinson's disease as early as its original description by James Parkinson in 1817. However, the few objective studies of muscle strength and endurance in Parkinson's disease have provided equivocal results. Wilson (1925) provided examples of reduced strength in various muscles of a few people with parkinsonism, but indicated that the more pervasive problem is slowness of muscle contraction and relaxation, and the inability to maintain contractions. Schwab, England, and Peterson (1959) argued that weakness is not a problem in Parkinson's disease, because normal amplitudes and directions of finger movements were achieved voluntarily ("voluntary ergogram" from the first dorsal interosseous muscle) with adequate motivation, and normal movement was elicited from electrical stimulation of the same muscle ("electronic ergogram"). Again, they noted that endurance was a primary problem.

Saltin and Landis (1975) reported that maximal isometric strength of the ankle and knee flexors was similar for 6 subjects with moderate to severe Parkinson's disease and healthy control subjects. Koller and Kase (1986) also found no difference for isometric hand grip strength (maximal effort using a dynamometer, averaged over 2 trials) between 21 subjects with mild Parkinson's disease and normal subjects. However, the subjects with Parkinson's disease demonstrated significantly decreased maximum isotonic muscle strength of the wrist, arm, and knee, measured by averaging the second, third, and fourth repetitions of maximal extension/flexion movements. The authors concluded that the subjects with Parkinson's disease were weaker than the control subjects, but only for repetitive tasks. Similarly, Tzelepis, McCool, Friedman, and Hoppin (1988) found that 9 subjects with mild to moderate Parkinson's disease were not impaired for single maximum efforts but were impaired for repetitive efforts involving the respiratory system.

Contrary to these findings of normal isometric strength, Yanagawa, Shindo, and Yanagisawa (1990) reported decreased maximal strength, measured as maximal torque produced by voluntary ankle dorsiflexion, in 15 subjects with mild to moderate Parkinson's disease. However, normal torques were obtained when the common peroneal nerve was electrically stimulated. Because weakness was apparent with voluntary activation but strength was normal when the muscle was activated involuntarily, a central rather than peripheral mechanism for muscle weakness was indicated. These results indicate that muscle or joint stiffness did not contribute to demonstrated weakness.

We are aware of only one study that systematically examined endurance in Parkinson's disease. Koller and Kase (1986) defined endurance as the number of repetitions of maximum extension/flexion movements of the wrist, arm, and knee to fatigue or until only 50% of the maximum strength could be generated. They found that endurance was greater for subjects with Parkinson's disease than for control subjects. This measure of endurance is difficult to interpret because the level of force (strength) and the rate of repetitions (movement velocity) can differ between subjects. Case study reports have clearly indicated a progressive decline in muscle strength over time that is quite different than that seen in healthy subjects (Schwab et al., 1959; Wilson, 1925).

Examination of strength in the orofacial system of people with Parkinson's disease has been reported in a few case studies. Barlow and Abbs (1983) described a subject with PD who did not demonstrate tongue weakness but exhibited instability in maintaining a steady force during tongue elevation. Dworkin and Aronson (1986) reported lower than normal maximum tongue "strength," measured by calculating the area under a force curve, in one subject with Parkinson's disease. Lip weakness has also been reported in Parkinson's disease (Netsell, Daniel, & Celesia, 1975; Wood, Hughes, Hayes, & Wolfe, 1992). Netsell et al. (1975) studied muscle activity of the upper lip in 22 people with Parkinson's disease (some of whom had been treated with thalamic surgery) and reported evidence of weakness (reduced amplitude and duration of electromyographic

activity) in at least one representative subject. Wood et al. (1992) used a labial force transducer to assess maximum force generation and found weakness of the lower lip, but not the upper lip, in 10 subjects with Parkinson's disease, 8 of whom had dysarthria. We are unaware of studies that examined endurance in the orofacial system in people with Parkinson's disease.

In the present investigation, we tested strength and endurance of the tongue and hand in 19 people with mild to moderate idiopathic Parkinson's disease and 19 control subjects matched for physical characteristics. Assessing hand function was deemed informative as an indicator of general muscle functioning. Individuals with Parkinson's disease may demonstrate differential impairment of the extremities and midline structures. The extant literature in Parkinson's disease has not addressed specifically orofacial strength and endurance in relation to speech. To address possible relations between general tongue function and speech production, speech samples from the subjects were evaluated by experienced speech-language pathologists for severity of articulatory imprecision and overall speech defectiveness. In addition, speech rate was calculated from an acoustic record.

Method

Subjects

Subjects were 19 adults diagnosed with idiopathic Parkinson's disease recruited from the Movement Disorders Clinic at the University of Iowa Hospitals and Clinics.[1,2] Individual data pertaining to physical characteristics and disease severity for the subjects with Parkinson's disease are provided in Table 1. The subjects were in mild to moderate stages of Parkinson's disease as judged on a modified Hoehn and Yahr scale (1967; Fahn et al., 1987) by a neurologist on the same day as data collection. Mild disease (Stages 1 or 2) was present in 12 subjects, mild-to moderate (Stage 2.5) in 5, and moderate (Stage 3) in 2. The subjects had no neurologic or speech disorders other than those associated with Parkinson's disease. Sixteen subjects with Parkinson's disease were taking antiparkinsonism medications although none experienced clinical fluctuations in their motor signs. Unfortunately, we were not able to coordinate data collection with the drug cycle because of scheduling constraints.

Nineteen neurologically normal adults were recruited from the community to match, one-to-one, subjects with Parkinson's disease for sex and age (within 3 years).

Table 1.
Demographic information for subjects with Parkinson's disease, including stage of disease (Hoehn & Yahr, 1967), sex, age, weight, and height, and for the matched control subjects, including age, weight, and height. Subjects were matched for sex.

	PARKINSON'S DISEASE					CONTROL		
Subject	Stage	Sex	Age (yr)	Weight (kg)	Height (cm)	Age (yr)	Weight (kg)	Height (cm)
A	1	F	54	75	175	55	76	163
B	1	F	72	70	163	72	75	158
C	1	M	64	106	185	62	105	185
D	2	F	49	90	172	49	95	163
E	2	M	58	99	185	61	100	183
F	2	F	61	79	160	58	74	168
G	2	M	73	87	179	71	86	183
H	2	F	76	59	160	73	60	158
I	2	M	64	72	172	64	68	173
J	2	M	66	97	185	67	97	178
K	2	M	73	76	181	70	78	183
L	2	F	65	79	161	64	78	168
M	2.5	F	46	78	173	49	80	163
N	2.5	F	64	87	165	65	85	163
O	2.5	M	73	108	170	72	96	178
P	2.5	M	66	76	178	65	76	183
Q	2.5	F	71	55	166	72	55	175
R	3	M	55	78	185	52	80	168
S	3	M	73	97	178	74	95	174

In addition, subjects were matched as closely as possible for weight (all were within 5 kg with the exception of subject pair O) and height (within 10 cm except for subject pairs A and R). These variables have been found to correlate with strength in various skeletal muscles (Burke, Tuttle, Thompson, Janey, & Weber, 1953; Collumbine, Bibile, Wikramanayake, & Watson, 1950; Larsson & Karlsson, 1978; Petrofsky & Lind, 1975; Robin, Somodi, & Luschei, 1991). Control subjects had negative histories for neurologic, speech, or language disorders, and were not taking medications that would affect motor performance. All subjects spoke General American English as their native language.

Procedures

The Iowa Oral Performance Instrument (IOPI) was used to assess strength and endurance. The IOPI has been described in detail previously (Robin et al., 1991; Robin, Goel, Somodi, & Luschei, 1992). In brief, the IOPI measures pressure exerted upon a small, air-filled bulb, and displays the result digitally (in kPa) or by a multi-light LED display. For the tongue, a small plastic bulb is placed against the hard palate immediately posterior to the alveolar ridge, and the subject pushes against the bulb in a rostral direction with the anterior portion of the tongue dorsum.

[1] In our original presentation of these data (Lorell, Solomon, Robin, Somodi, Luschei, & Rodnitzky, 1992), 23 subjects were included. For this final analysis, 2 subjects were eliminated because English was not their first language, 1 for having a history of drug abuse, and 1 for being in a severe stage of the disease (Hoehn & Yahr, Stage 4).

[2] Two of these subjects were described previously in a preliminary report [Subjects P-L ("Mrs. S.") and P-Q ("Mrs. H."); Solomon et al., 1993].

Table 2.
Results for strength and endurance of the hand and the tongue for the subjects with Parkinson's disease and the matched control subjects.

	PARKINSON'S DISEASE				CONTROL			
Subject	Strength (kPa)		Endurance (s)		Strength (kPa)		Endurance (s)	
	Tongue	Hand	Tongue	Hand	Tongue	Hand	Tongue	Hand
A	71	117	12	42	65	119	26	75
B	65	158	37	84	76	120	24	34
C	88	168	15	28	70	196	18	18
D	70	106	30	52	63	128	53	40
E	84	186	32	56	86	224	33	50
F	64	115	30	50	73	126	37	54
G	60	118	24	127	71	151	30	53
H	53	83	11	39	78	121	17	62
I	63	191	30	12	64	115	17	70
J	72	151	24	31	80	156	15	21
K	38	139	49	23	64	157	16	35
L	70	112	45	83	53	169	43	60
M	66	158	10	30	69	189	25	94
N	30	104	4	49	62	150	42	55
O	75	182	18	35	79	201	38	45
P	57	120	39	144	57	193	25	90
Q	29	149	55	52	50	126	32	30
R	26	143	23	38	90	195	51	60
S	46	110	13	25	66	137	23	37
M	59.3	137.4	26.4	52.4	69.3	156.5	29.7	51.7
SD	18.1	31.1	14.2	34.8	10.6	34.1	11.7	21.0

For the hand, the subject grips a hand bulb which is placed in the palm of the preferred hand.

To measure maximum strength, subjects squeezed the bulb as hard as possible. The best of two trials was taken as the maximum. Hand strength was determined first, then tongue strength. Following the maximum strength maneuvers, endurance of the hand and then the tongue (one trial each) was measured. Subjects were instructed to maintain 50% of the maximum pressure as long as possible. The LED display on the IOPI was used for visual feedback and verbal encouragement was provided. Trials were timed with a stopwatch, and were terminated when the subject abruptly dropped the pressure or when 50% of the maximum pressure could not be maintained.

A speech sample was collected from all 38 subjects. Subjects described the Cookie Theft picture from the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983). Later, a segment of each speech sample was recorded onto another tape in random order. Four speech-language pathologists with 5 or more years of clinical experience (not the investigators involved in data collection) rated the speech samples for articulatory precision and overall speech defectiveness on a six-point scale (0=normal, 1=mild, 2=mild-to-moderate, 3=moderate, 4=moderate-to-severe, 5=severe). Judgements by the 4 listeners were averaged to provide a single numeric result for each speech sample. In addition, speech rate was determined by measuring the acoustic waveform with the C-Speech software program for personal computers (Milenkovic & Read, 1992). The duration of speech, with the exclusion of pauses > 250 ms, was determined. The number of syllables was divided by the duration of speech. This procedure resulted in a measure of "interpause speech rate."[3]

Statistical Analysis

A repeated measures multivariate analysis of variance with one within-subjects factor was used to analyze the strength and endurance data. Two variables were included in the analysis: structure (tongue and hand) and function (strength and endurance). The within-subjects factor was subject group (Parkinson and control). This analysis allowed for paired comparisons of matched subjects. The Wilcoxon signed-rank test was used to test for differences between paired perceptual judgements of speech (equal judgements for a pair were considered missing data; a correction was conducted for tied ranks). Speech rate between pairs of subjects was compared with a 1-sample t-test for paired data (2-tailed probability). For all analyses, a probability level of 0.05 was assigned.

Results

Measures of tongue and hand strength and endurance for each subject are provided in Table 2. A statistically significant difference between the subject groups was realized when all variables were considered [$F(2,17)=4.393$; $p=0.029$]. The difference was due to strength [$F(2,17)=4.645$; $p=0.025$], not endurance [$F(2,17)=0.359$; $p=0.704$]. A significant difference between the structures (hand and tongue) was found for both strength and endurance [$F(2,17)=121.9$; $p=0.0001$]; both were greater for the

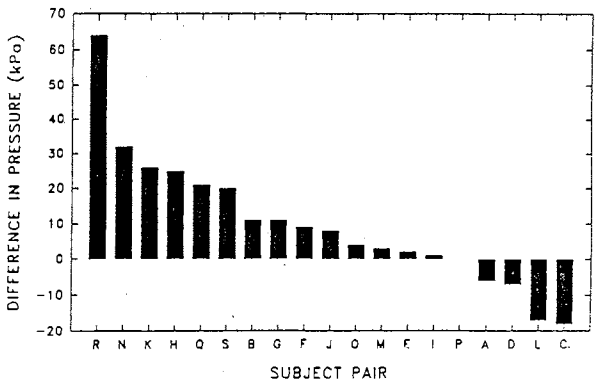


Figure 1. Differences in tongue strength (pressure, in kPa) plotted for matched pairs of subjects (Control-Parkinson). Positive differences indicate that the control subjects exhibited greater pressures than did subjects with Parkinson's disease.

[3]Measures of speech rate that exclude pauses may provide more meaningful information for the speed of speech articulation than if pauses are included (Alp, 1988; Hammen, 1990; Metter & Hanson, 1986; Till & Goff, 1986). Recent studies in Parkinson's disease have excluded pauses from measures of speech rate (Hammen, 1990; Solomon & Hixon, 1993).

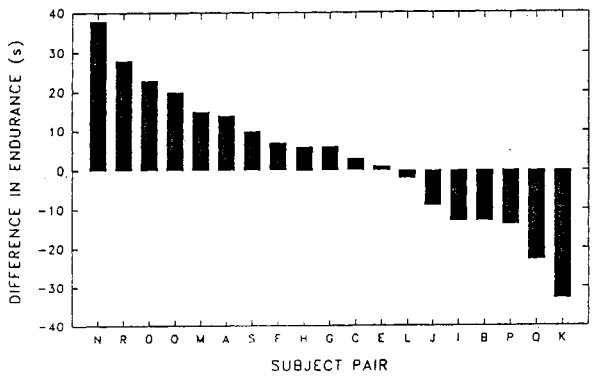


Figure 2. Differences in tongue endurance (duration, in s) plotted for matched pairs of subjects (Control-Parkinson). Positive differences indicate that the control subjects exhibited greater endurances than did subjects with Parkinson's disease.

hand. However, the interaction between subject group and structure was not significant [F(2,17)=0.524; p=0.601], indicating that the structures were not affected differentially for the subject groups.

Differences between matched pairs of subjects for tongue strength are illustrated in Figure 1, and for tongue endurance in Figure 2. In the graphic displays, the result for the subject with Parkinson's disease was subtracted from that for the matched control subject. The data were ordered in terms of the magnitude and direction of the differences. Therefore, the subject order is different for the two graphs. For tongue strength (Figure 1), it is clear that most of the data are positive (i.e., the control subject had greater tongue strength than the matched subject with Parkinson's disease). For tongue endurance (Figure 2), the data are more evenly split between positive and negative differences.

The results for the analyses of speech are provided in Table 3. Average judgements by the speech-language pathologists indicated that articulatory imprecision and overall speech defectiveness were not present (i.e., were normal) or were mild for all but 2 subjects with Parkinson's disease (Parkinson subjects R and S). Articulatory imprecision was slightly but significantly greater for the subjects with Parkinson's disease (V=35, n=15; Z=-1.65, p=0.049, after correction for tied ranks). No difference was revealed for overall speech defectiveness (V=63, n=17; Z=-0.88, p=0.188 after correction).

Inspection of Figures 1 and 2 reveals that severity of disease (Table 1) did not relate systematically with measures of strength or endurance for these subjects (Note placement of Subjects A, B, and C, the most mildly affected, and Subjects R and S, the most severely affected, on the graphs.). Similarly, severity of speech disorder

Table 3.
Results for the perceptual analysis of speech for articulatory imprecision and overall speech deficiency, and measures of interpause speech rate (syllables per second) for the subjects with Parkinson's disease and the matched control subjects.

	PARKINSON'S DISEASE			CONTROL		
Subject	Articulation	Overall Speech	Interpause Speech Rate (syl/s)	Articulation	Overall Speech	Interpause Speech Rate (syl/s)
A	0	0	3.96	0.50	0.25	4.81
B	1.25	1.50	4.49	0	0.75	4.70
C	1.0	1.50	5.15	0.25	0.75	5.11
D	0	0	4.19	0	0.25	3.91
E	0	0	4.43	0.25	0	3.83
F	0	0	5.04	1.00	0.75	5.12
G	0.75	0.50	4.25	0.25	0.75	5.31
H	0.5	0.75	4.19	0.25	1.00	3.23
I	0.75	0.25	3.79	0.75	1.00	4.60
J	0.75	0	4.30	0.75	0.25	4.53
K	1.25	1.00	4.17	0.25	0.25	3.78
L	1.00	1.50	5.62	0	0	4.79
M	0	0	4.90	..	0	3.67
N	0	0	4.97	0.25	0.50	4.18
O	0.50	0.25	4.66	0.75	0.50	4.96
P	0.50	0.50	5.16	1.25	1.25	4.19
Q	0.75	0.50	4.98	0.50	0	5.27
R	2.25	2.00	5.43	0	0	3.80
S	2.25	2.50	5.44	0	0.25	5.07
M	0.694	0.671	4.691	0.368	0.447	4.465
SD	0.710	0.777	0.543	0.376	0.396	0.626

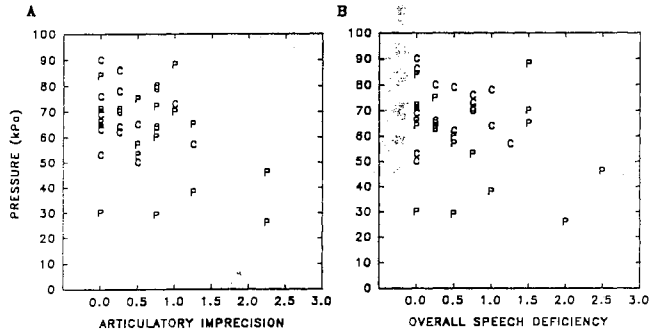


Figure 3. Tongue strength (maximal pressure) for subjects with Parkinson's disease (P) and control subjects (C) plotted against average perceptual judgements of articulatory imprecision (A) and overall speech deficiency (B); 0 = normal, 1 = mild, 2 = mild-to-moderate, 3 = moderate.

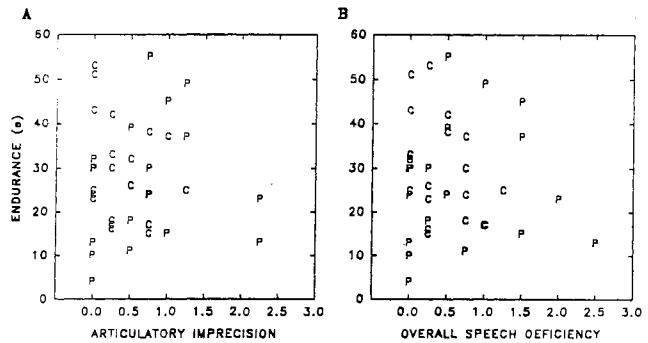


Figure 4. Tongue endurance at 50% maximal pressure for subjects with Parkinson's disease (P) and control subjects (C) plotted against average perceptual judgements of articulatory imprecision (A) and overall speech deficiency (B); 0 = normal, 1 = mild, 2 = mild-to-moderate, 3 = moderate.

(Table 3) bore no obvious relation to tongue strength (Figure 3) and endurance (Figure 4). One exception to these observations is Parkinson Subject R's; he was one of the most severely affected subjects and the differences between his data and his matched control subject's data for tongue and hand strength and endurance are among the greatest (Table 2, Figures 1 and 2).

Interpause speech rate (Table 3) did not differ between subject groups ($t(18)=1.3179$; $p=0.204$). Speech rate is plotted against tongue strength in Figure 5A and tongue endurance in Figure 5B. Graphic inspection of these plots reveals no obvious relations for these subjects.

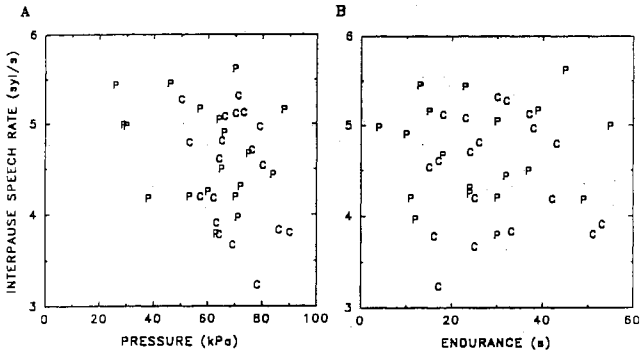


Figure 5. Interpause speech rate (syllables/s) for subjects with Parkinson's disease (P) and control subjects (C) plotted against tongue strength (maximal pressure) and tongue endurance at 50% maximal pressure.

Discussion

Subjects with mild to moderate Parkinson's disease were found on average to have less strength in the tongue and hand than 19 matched control subjects. Strength was determined by the maximum pressure exerted on an air-filled bulb. Despite the contribution of Subject Pair R to the large difference in tongue strength between subject groups (Parkinson Subject R had the lowest and Control Subject R had the greatest tongue strength of all subjects), the difference appears to be real for the subject groups in general. Previous research generally has indicated that limb isometric strength (i.e., maximal force, torque, or pressure generated during a single maximal voluntary contraction) is not reduced in Parkinson's disease (Koller & Kase, 1986; Saltin & Landis, 1975; Tzelepis et al., 1988; for an exception, see Yanagawa et al., 1990). However, weakness has been reported in the tongue (Dworkin & Aronson, 1986) and lips (Netsell et al., 1975; Wood et al., 1992).

Endurance, defined in this report as the maximum duration for maintaining 50% of the maximum pressure, did not differ systematically between the subject groups. This finding was unexpected because people with Parkinson's disease often complain of fatigue. This perception of fatigue may relate to the muscle weakness demonstrated for these subjects; muscle weakness corresponds with an increased sense of effort (Gandevia, 1982; McCloskey, 1981), a critical component of fatigue (Edwards, 1981; Enoka & Stuart, 1985). In addition, weakness may provide a possible explanation for the finding of normal endurance. That is, the target pressure during the endurance task would be lower in the subjects with Parkinson's disease than in the control subjects. Because of the lower target pressure, the task may require less effort and recruitment of more fatigue-resistant motor units than when strength is normal. Koller & Kase's (1986) finding of greater than normal endurance values in subjects with Parkinson's disease is consistent with this hypothesis.

The inability to maintain a target contraction is due to a process termed "force failure." Force failure can result from central or peripheral sites in the motor system. As described previously, Yanagawa et al. (1990) used voluntary contractions and nerve stimulation to demonstrate that muscle weakness in Parkinson's disease may be due to central and not peripheral mechanisms. Measures for the assessment of central fatigue may provide critical information for the understanding of fatigue in Parkinson's disease. We have been investigating perceptions of effort in an attempt to elucidate central fatigue in normal and neurologically disordered individuals (Solomon et al., in press; Solomon, Robin, Dorothy, & Luschei, 1992; Somodi, Robin, & Luschei, 1993). Thus far, our research suggests that neurologically normal young adults can accurately and consistently perceive various levels of effort (Somodi et al., 1993) such that low pressures are generated at low effort levels and progressively higher pressures are generated at higher effort levels. In addition, healthy young adults can maintain a constant sense of effort that corresponds with an exponential decline in pressure (Solomon et al., 1992). Based on the results of the present investigation, we examined the ability of 2 of the subjects with Parkinson's disease (Subjects P-L and P-Q) and an additional subject with parkinsonism to perform tasks related to the perception of effort (Solomon et al., in press). Each of these subjects complained of fatigue. Although the disordered subjects could maintain constant effort levels in a manner comparable to neurologically normal subjects (Solomon et al., 1992), they had difficulty producing pressures that corresponded to various effort levels (Solomon et al., in press). We are continuing to explore strength, endurance, and perceptions of effort in a study of people with moderate to severe Parkinson's disease.

The relation between tongue strength and endurance (assessed with non-speech tasks) and speech is unclear. Most of the subjects in the present investigation, including the control subjects, were judged to have normal or mildly disordered speech. In fact, judgements for overall speech defectiveness and measures of interpause speech rate did not differ between the two subject groups, and articulatory imprecision was only slightly greater for the subjects with Parkinson's disease.

The 2 subjects who were judged as having mildly to moderately impaired speech (Parkinson Subjects R and S) had lower than normal tongue strength and endurance: tongue function results were especially abnormal for Subject R. In addition, their interpause speech rates were among the fastest. This last observation contrasts with a speculation we advanced in a previous paper, that increased speech rate may correspond with better tongue endurance in Parkinson's disease (Solomon et al., in press). We based this hypothesis on the finding that "supernormal" speakers (debaters) with fast speech rates (M=414 words/min, or approximately 8.9 syllables/s) had supernormal tongue endurance (approximately 95 s at 50% of maximal pressure; Robin et al., 1992). It should be clarified that the debaters and control subjects in that study were asked to speak as fast as possible while maintaining intelligibility. In contrast, the present subjects' speech rate was determined from a task of habitual speech rate. Apparently, substantially more data are needed to assess whether a relationship between speech rate and tongue endurance holds for speech disordered populations.

Although individual subjects can be described whose data appear to support the expectation of impaired tongue function co-occurring with disordered speech, this relationship clearly cannot be addressed by the current data. Future studies of larger groups of subjects, subjects with more severely affected speech, and changes in speech and non-speech tongue function over time may clarify this issue.

Acknowledgments

This research was supported by Grant No. R03 DC01182 and Grant No. P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders. We gratefully acknowledge Lori B. Somodi, Samuel K. Seddoh, Daniel L. Keyser, and Judith K. Dobson for their assistance with this project.

References

Alp, L. A. (1988). The effects of imposed speech rate reductions on linguistic parameters and speech characteristics. Unpublished master's thesis. University of California, Santa Barbara.

Anthony, J.K.F., & Farquharson, I. M. (1975). Clinical research in speech pathology, The Journal of Laryngology and Otology, 89, 1169-1184.

Barlow, S. M., & Abbs, J. H. (1983). Force transducers for the evaluation of labial lingual and mandibular function in dysarthria. Journal of Speech and Hearing Research, 26, 616-621.

Boshes, Benjamin (1966). Voice changes in parkinsonism. Journal of Neurosurgery, 24, 286-288.

Burke, W. E., Tuttle, W. W., Thompson, C. W., Janey, C. D., & Weber, R. J. (1953). The relation of grip strength and grip-strength endurance to age. Journal of Applied Physiology, 5, 628-630.

Caligiuri, Michael P. (1987). Labial kinematics during speech in patients with parkinsonian rigidity. Brain, 110, 1033-1044.

Canter, Gerald J. (1963). Speech characteristics of patients with Parkinson's Disease: I. Intensity, pitch, and duration. Journal of Speech and Hearing Disorders, 28, 221-229.

Canter, Gerald J. (1965). Speech characteristics of patients with Parkinson's Disease: III. Articulation, Diadochokinesis, and Over-All Speech Adequacy. Journal of Speech and Hearing Disorders, 30, 217-224.

Chenery, H. J., Murdoch, B. E., & Ingram, J. C. L. (1988). Studies in Parkinson's disease: I. Language and perceptual speech analyses. The Australian Journal of Human Communication Disorders, 16, 17-29.

Collumbine, H., Bibile, S. W., Wikramanayake, T. W., & Watson, R. S. (1950). Influence of age, sex, physique and muscular development on physical fitness. Journal of Applied Physiology, 2, 488-511.

Connor, N. P., Abbs, J. H., Cole, K. J., & Gracco, V. L. (1989). Parkinsonian deficits in serial multiarticulate movements for speech. Brain, 112, 997-1009.

Conner, N. P., Ludlow, C. L., & Schulz, G. M. (1989). Stop consonant production in isolated and repeated syllables in Parkinson's disease. Neuropsychologia, 27, 829-838.

Critchley, E.M.R. (1981). Speech disorders of Parkinsonism: A review. Journal of Neurology, Neurosurgery, and Psychiatry, 44, 751-758.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. Journal of Speech and Hearing Research, 12, 246-269.

Dworkin, J. P. (1978). Protrusive lingual force and lingual diadochokinetic rates: A comparative analysis between normal and lisping speakers. Language, Speech, and Hearing Services in the Schools, 9, 8-16.

Dworkin, J. P., & Aronson, A. E. (1986). Tongue strength and alternate motion rates in normal and dysarthric subjects. Journal of Communication Disorders, 19, 115-132.

Dworkin, J. P., Aronson, A., & Mulder, D. W. (1980). Tongue strength in normal subjects and dysarthric patients with amyotrophic lateral sclerosis. Journal of Speech and Hearing Research, 23, 828-837.

Edwards, R. H. T. (1981). Human muscle function and fatigue. In R. Porter & J. Whelan, (Eds.). Human muscle fatigue: Physiological mechanisms (pp. 1-18). London: Pitman Medical.

Enoka R. M., & Stuart, D. G. (1985). The contribution of neuroscience to exercise studies. Federation Proceedings, 44, 2279-2285.

Ewanowski, S. J. (1964). Selected Motor-Speech Behavior of Patients with Parkinsonism. Dissertation Abstracts International, 24/11, 4860. (University Microfilms No. AAC6407081)

Fahn, S., Elton, R. L., and Members of the UPDRS Development Committee (1987). Unified Parkinson Disease Rating Scale. In S. Fahn, C. D. Marsden, D. B. Calne, & M. Goldsteing (Eds.), Recent developments in Parkinson's disease (Vol. 2, pp. 153-304). New York: MacMillan.

Forrest, Karen, Weismer, Gary, & Turner, Greg S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. Journal of the Acoustical Society of America, 85, 2608-2622.

Gandevia, S. C. (1982). The perception of motor commands or effort during muscular paralysis. Brain, 105, 151-159.
Goodglass, H., & Kaplan, E. (1983). The assessment of aphasia and related disorders (2nd ed.). Philadelphia: Lea & Febiger.

Hammen, V. L. (1990). The effects of speaking rate reduction in parkinsonian dysarthria. Dissertation Abstracts International, 51/09B, 4304. (University Microfilms No. AAC9104243)

Hammen, V.L., Yorkston, K.M., & Beukelman, D.R. (1989). Pausal and speech duration characteristics as a function of speaking rate in normal and parkinsonian dysarthric individuals. In K.M. Yorkston & D.R. Beukelman (Eds.), Recent Advances in Clinical Dysarthria (213-224). Boston: College-Hill.

Hanson, W. R., & Metter, E. J. (1983). DAF speech rate modification in Parkinson's disease: A report of two cases. In W. R. Berry (Ed.), Clinical dysarthria (pp.231-251). San Diego: College-Hill Press.

Hirose, H., Kiritani, S., Ushijima, T., Yoshioka, H., & Sawashima, M. (1981). Patterns of dysarthric movements in patients with parkinsonism. Folia phoniatrica, 33, 204-215.

Hoehn, M.M., & Yahr, M.D. (1967). Parkinsonism: Onset, progression, and mortality. Neurology, 17, 427-442.

Hunker, C. J., Abbs, J. H., & Barlow, S. M. (1982). The relationship between parkinsonian rigidity and hypokinesia in the orofacial system: A quantitative analysis. Neurology, 32, 755-761.

Kammermeier, M. A. (1969). A comparison of phonatory phenomena among groups of neurologically impaired speakers. Dissertation Abstracts International, 30/06, 2945. (University Microfilms No. AAC6920026)

Koller, W., & Kase, S. (1986). Muscle strength testing in Parkinson's disease. European Neurology, 25, 130-133.

Larsson, L., & Karlsson, J. (1978). Isometric and dynamic endurance as a function of age and skeletal muscle characteristics. Acta physiologica scandinavia, 104, 129-136.

Laszewski, Z. (1956). Role of the department of rehabilitation in preoperative evaluation of Parkinsonian patients. Journal of American Geriatrics Society, 4, 1280-1284.

Leanderson, R., Meyerson, A., & Persson, A. (1972). Lip muscle function in parkinsonian dysarthria. Acta otolaryngologica, 74, 350-357.

Logemann, J., Boshes, B., & Fisher, H. (1972). The steps in the degeneration of speech and voice control in Parkinson's disease. In J. Siegfried (Ed.), Parkinson's disease: Rigidity, akinesia, behavior, Vol. 2: Selected Communications on Topic (pp. 101-111). Bern Stuttgart Vienna: Hans Huber.

Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and Hearing Disorders, 43, 47-57.

Lorell, D. M., Solomon, N. P., Robin, D. A., Somodi, L. B., Rodnitzky, R., & Luschei, E.S. (1992, April). Tongue strength and endurance in individuals with Parkinson disease. Paper presented at the Conference on Motor Speech, Boulder, CO.

Ludlow, C. L., & Bassich, C. J. (1983). The results of acoustic and perceptual assessment of two types of dysarthria. In W. B. Berry (Ed.), Clinical Dysarthria (pp. 121-153). San Diego: College-Hill Press.

Marquardt, Thomas P. (1973). Characteristics of speech production in Parkinson's disease: Electromyographic, structural movement, and aerodynamic measurements. Dissertation Abstracts International, 34/08B, 4113. (University Microfilms No. AAC7402217)

McCloskey, D. I. (1981). Corollary discharges: Motor commands and perception. In V. B. Brooks, (Ed.), Handbook of physiology: Section 1, Vol. II, The nervous system: Motor control, Part 2. (pp. 1415-1447). Bethesda, MD: American Physiological Society.

Metter, E. J., & Hanson, W. R. (1986). Clinical and acoustical variability in hypokinetic dysarthria. Journal of Communication Disorders, 19, 347-366.

Milenkovic, P. H., & Read, C. (1992). CSpeech Version 4: User's Manual. Madison, WI: University of Wisconsin-Madison.

Morrison, E. B., Rigrodsky, S., & Mysak, E. D. (1970). Parkinson's disease: Speech disorder and released infantile oroneuromotor activity. Journal of Speech and Hearing Research, 13, 655-666.

Netsell, R., Daniel, B., & Celesia, G. G. (1975). Acceleration and weakness in Parkinsonian dysarthria. Journal of Speech and Hearing Disorders, 40, 170-178.

Palmer, M. F., & Osborn, C. D. (1940). A study of tongue pressures of speech defective and normal speaking individuals. Journal of Speech Disorders, 52, 133-140.

Peacher, William G. (1950). The etiology and differential diagnosis of dysarthria. Journal of Speech and Hearing Disorders, 15, 252-265.

Petrofsky, J. S., & Lind, A. R. (1975). Isometric strength, endurance, and the blood pressure and heart rate responses during isometric exercise in healthy men and women, with special reference to age and body fat content. Pflügers Archiv: European Journal of Physiology, 360, 49-61.

Pitcairn, T. K., Clemie, S., Gray, J. M., & Pentland, B. (1990). Impressions of parkinsonian patients from their recorded voices. British Journal of Disorders of Communication, 25, 85-92.

Robin, D. A., Goel, A., Somodi, L. B., & Luschei, E. S. (1992). Tongue strength and endurance: Relation to highly skilled movements. Journal of Speech and Hearing Research, 35, 1239-1245.

Robin, D. A., Somodi, L. B., & Luschei, E. S. (1991). Measurement of strength and endurance in normal and articulation disordered subjects. In C. A. Moore, K. M. Yorkston, & D. R. Beukelman (Eds.), Dysarthria and apraxia of speech: Perspectives on management (pp. 173-184). Baltimore: Paul H. Brooks.

Saltin, B., & Landis, S. (1975). Work capacity, muscle strength and SDH activity in both legs of hemiparetic patients and patients with Parkinson's disease. Scand. J. cin. Lab. Invest., 35, 531-538.

Schwab, R. S., England, A. C., & Peterson, E. (1959). Akinesia in Parkinson's disease. Neurology, 9, 65-72.

Solomon, N. P., & Hixon, T. J. (1993). Speech breathing in Parkinson's disease. Journal of Speech and Hearing Research, 36, 294-310.

Solomon, N. P., Robin, D. A., Dorothy, S. I., & Luschei, E. S. (1992, April). Sense of effort in the tongue and hand: An exponential decay model of central fatigue. Paper presented at the Conference on Motor Speech, Boulder, CO.

Solomon, N. P., Robin, D. A., Lorell, D. M., Rodnitzky, R. L., & Luschei, E. S. (in press). Tongue function testing in Parkinson's disease: Indications of fatigue. In J. Till, K. Yorkston, & D. Beukelman (Eds.), Motor speech disorders: Advances in assessment and treatment. Baltimore: Brooks.

Somodi, L. B., Robin, D. A., & Luschei, E. S. (1993). A model of "sense of effort" during maximal and submaximal contractions of the tongue. Unpublished manuscript.

Streifler, M., & Hofman, S. (1984). Disorders of verbal expression in parkinsonism (pp. 385-393). In R.G. Hassler & J.F. Christ (Eds.), Advances in Neurology (Vol. 40). New York: Raven Press.

Tanner, D. C. (1976). Spectrographic, pausimetric, and intelligibility measures in Parkinson's disease. Dissertation Abstracts International, 37/09, 4415. (University Microfilms No. AAC7705901)

Till, J. A., & Goff, A. M. (1986, November). Task variables affecting temporal structure and respiratory patterns in speech [Abstract]. Asha, 28, 102.

Tzelepis, G. E., McCool, F. D., Friedman, J. H., & Hoppin, F. G., Jr. (1988). Respiratory muscle dysfunction in Parkinson's disease. American Review of Respiratory Disease, 138, 266-271.

Wilson, S. A. (1925). The Croonian lectures on some disorders of motility and of muscle tone, with special reference to the corpus striatum. The Lancet, 209, 1-10, 53-62.

Wood, L. M., Hughes, J., Hayes, K. C., & Wolfe, D. L. (1992). Reliability of labial closure force measurement in normal subjects and patients with CNS disorders. Journal of Speech and Hearing Research, 35, 252-258.

Yanagawa, S., Shindo, M., & Yanagisawa, N. (1990). Muscular weakness in Parkinson's disease. In M. B. Streifler, A. D. Korczyn, E. Melamed, & M. B. H. Youdim (Eds.), Advances in Neurology, Vol. 53: Parkinson's disease: Anatomy, pathology, and therapy (pp. 259-269). New York: Raven Press.

Yorkston, K. M., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. Journal of Speech and Hearing Disorders, 55, 550-560.

NCVS Status and Progress Report - 5
November 1993, 29-41

Interpretation of Biomechanical Simulations of Normal and Chaotic Vocal Fold Oscillations with Empirical Eigenfunctions

David A. Berry, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Hanspeter Herzel, Ph.D.
Department of Physics, Humboldt University (Berlin, Germany)
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Katharina Krischer
Fritz Haber Institut der Max Planck Gesellschaft (Berlin, Germany)

Abstract

Empirical orthogonal eigenfunctions are extracted from biomechanical simulations of normal and chaotic vocal fold oscillations. For normal phonation, two dominant empirical eigenfunctions capture the vibration patterns of the folds and exhibit a 1:1 entrainment. The eigenfunctions show some correspondence to theoretical low-order normal modes of a simplified, three-dimensional elastic continuum, and to the normal modes of a linearized two-mass model. The eigenfunctions also facilitate a physical interpretation of energy transfer mechanisms in vocal fold dynamics. Subharmonic regimes and chaotic oscillations are observed during simulations of a lax cover, in which case at least three empirical eigenfunctions are necessary to capture the resulting vocal fold oscillations. These chaotic oscillations might be understood in terms of a desynchronization of a few of the low-order modes, and may be related to mechanisms of creaky voice or vocal fry. Furthermore, some of the empirical eigenfunctions captured during complex oscillations correspond to higher-order normal modes described in earlier theoretical work. The empirical eigenfunctions may also be useful in the design of lower-order models (valid over the range for which the empirical eigenfunctions remain more or less constant), and may help facilitate bifurcation analyses of the biomechanical simulation.

Introduction

With any model of a physical or physiological process, there is always a trade-off between simplicity and completeness. The model should be simple enough to be useful in conceptualization and prediction, but also complete enough to represent the process accurately.

This certainly applies to vocal fold models. Early one-mass and two-mass models (Flanagan & Landgraf, 1968; Ishizaka & Flanagan, 1972) were simple enough to be described in a few pages of print. They were elegant in that they helped conceptualize the interaction between airflow and tissue movement to produce self-oscillation. But there is considerable doubt that they represented the geometry and the viscoelastic properties of the vocal folds adequately for the study of voice disorders or special vocal qualities. More recent models by Titze and Talkin (1979), and Titze and Alipour (in review) have enough biomechanical detail to model the three-dimensional layered structure of vocal fold tissue, but a heavy price is paid in terms of mathematical complexity and speed of computation. Furthermore, interpreting the dynamics of such intensive descriptions of the vocal folds can be a formidable task, particularly if irregular, chaotic vibrations occur (Titze, Baken & Herzel, 1993).

One way to facilitate the physical interpretation of a vibrating structure is to calculate its principal modes of vibration. Sometimes, even complicated vibration patterns can be explained by a relatively small number of orthogonal modes.

The paper begins with an introduction to modal analysis of vocal fold tissues, followed by a brief description of the biomechanical simulation used in this investigation (Titze & Alipour, in review). Next a set of empirical eigenfunctions extracted from simulations of both normal and chaotic vocal fold oscillations is presented. The physical significance of these eigenfunctions is discussed and related to current theories of voice production and nonlinear dynamics (e.g. airflow-tissue energy transfer mechanisms and desynchronization mechanisms). The empirical eigenfunctions are also compared with the modes captured by the two-mass model (Ishizaka & Flanagan, 1972) and to modes predicted for a simplified elastic continuum (Titze & Strong, 1975). Finally, future directions of research using this procedure are discussed.

Modal Analysis of the Vocal Folds

Modal analysis is a basic technique used to analyze many vibrating structures. Traditionally, it refers to the process of determining the normal (natural) modes and frequencies of a linear (or linearized) system. It is a powerful technique because it provides a framework in which a system can be decomposed into a set of independent vibration patterns, each with a characteristic (although not necessarily unique) frequency. Experimentally, these normal modes/frequencies can be observed immediately after a system is pulse excited, or during a forced, sinusoidal excitation (provided the driving frequency coincides closely enough with one of the systems' natural frequencies). One of the major limitations of this technique is that it is only valid for linear systems. However, in practice many systems are approximately linear for small-amplitude oscillations.

Theoretical Normal Modes

The concept of normal modes and frequencies is not new to speech science. For example, formants are frequencies which correspond to the normal modes of the vocal tract. They have been discussed extensively in terms of vowel production. The same concepts have not been exploited to the same degree for an understanding of vocal fold movement, although the study of normal modes in vocal fold tissues does have its beginnings. Almost two decades ago, Titze and Strong (1975) theoretically determined normal modes of the vocal folds. By examining a single fold and treating it as a three-dimensional, elastic, compressible medium, and by assuming a rectangular parallelepiped with simple boundary conditions (anterior, posterior, and lateral boundaries fixed; medial, superior, and inferior boundaries free), normal modes were expressed in terms of elementary sines and cosines. For comparison with empirical modes to be shown later, the theoretical modes are reviewed and briefly discussed. The lateral displacements ξ of the x-modes are given by:

$$\xi(x,y,z,t) = A\exp^{i\omega_x t}\cos\frac{(2n_x-1)\pi x}{2D}\sin\frac{n_y\pi y}{L}\cos\frac{n_z\pi z}{T} \quad (1)$$

and the vertical displacements ζ of the z-modes are given by:

$$\zeta(x,y,z,t) = B\exp^{i\omega_z t}\cos\frac{(2n_x-1)\pi x}{2D}\sin\frac{n_y\pi y}{L}\cos\frac{n_z\pi z}{T} \quad (2)$$

where n_x, n_y, and n_z are integers indicating the order of the modes; L, D, and T are the length, depth and thickness of the folds, respectively; A and B are arbitrary constants; and ω_x and ω_z are the radian frequencies of vibration. Any possible y-displacements (anterior-posterior direction) are neglected. This is based on experimental evidence that the trajectories of vocal fold fleshpoints are mostly planar (Baer, 1981; Saito, Fukuda, Isogai & Ono, 1981; Saito, Fukuda, Kitahira, Isogai, Tsuzuki, Muta, Takyama, Fujika, Kokawa & Makino 1985).

In order to distinguish the modes, the order indices (n_x, n_y, and n_z) must be specified and the modes need to be identified as either x or z modes (the assumption of compressible tissue allows the decoupling of such modes). In practice, the n_x index is usually not specified because the standing wave pattern governed by n_x (the first cosine term) is assumed to be constant (e.g., the likelihood of reflections from the fixed lateral boundary is small because of high attenuation in the thyroarytenoid muscle). Thus, following nomenclature introduced previously (Titze & Strong, 1975; Titze, 1976; Titze, 1988), the modes are designated as either x-$n_y n_z$ or z-$n_y n_z$ modes. Conceptually, the n_y and n_z indices indicate how many half-wavelengths occur along the longitudinal and vertical dimensions, respectively.

A few of the lower-order modes are shown in Figure 1. Figure 1a shows a superior view (upper) and coronal view (lower) of the x-10 mode. From the superior view, the commonly observed oval glottis is displayed. From the coronal view, all the lateral tissue displacements are in phase along the vertical dimension. An x-11 mode is displayed in Figure 1b. In the coronal view, the tissue at the top of the folds is 180 degrees out of phase with the tissue at the bottom of the folds. Variations of these lowest-order x modes describe some of the most commonly observed vocal fold vibration patterns (Moore & Von Leden, 1958; Hirano, 1975). Indeed, an appropriate combination of these modes is known to be essential for self-oscillation of the folds (Titze, 1988). An x-21 mode is illustrated in Figure 1c, and a z-10 mode in Figure 1d (sagittal view on top). These modes are not as easily observed because (1) the superior aspect, which is almost always used in high-speed films and videostroboscopy, is not ideal for viewing z-modes (saggital or coronal views would be better), and (2)

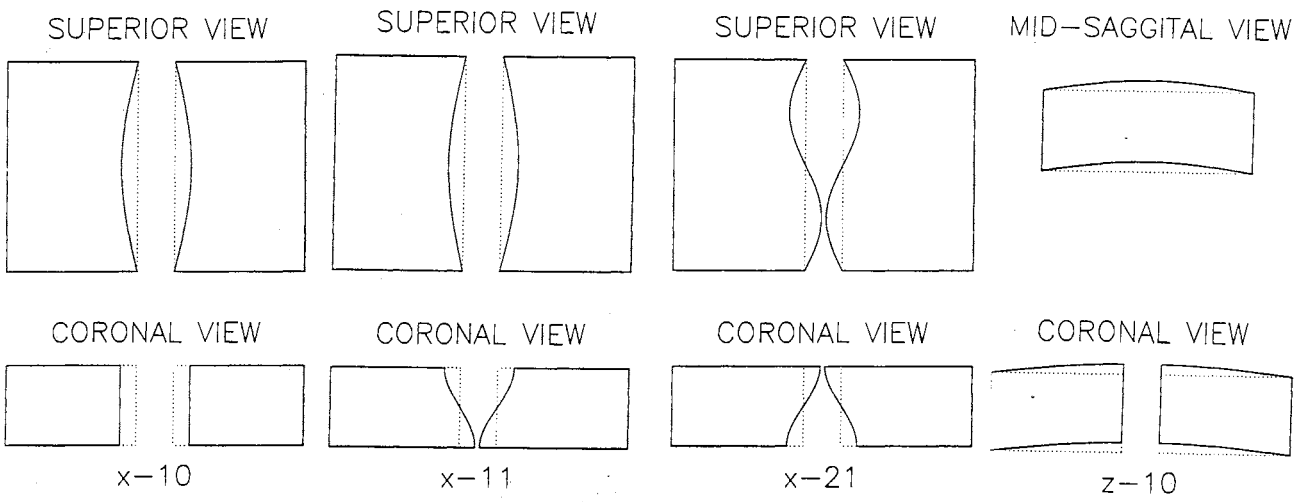


Figure 1. A few of the low-order, theoretical, normal modes are shown from superior and coronal views: (a) x-10, (b) x-11, (c) x-21. A saggital and coronal view is shown for (d) z-10. An artificial separation of left and right folds is used in order to display the true theoretical modes without deformation from collision.

higher-order modes (such as the x-21 mode) usually have smaller vibrational amplitudes and are often not excited.

Normal Modes in Low-Order Models

Not long after these normal modes were introduced, Titze (1976) claimed that (1) self-oscillation of the vocal folds consists of "approximately linear combinations of the normal modes," and (2) that "self-oscillation ... occurs at ... one of the natural frequencies of oscillation, usually the lowest." Titze demonstrated the plausibility of these concepts through an analysis of the two-mass model (Ishizaka and Flanagan, 1972). The normal modes of the two-mass model were shown to be analogous to the lowest order *x*-modes of the simplified elastic continuum; that is, the mode where the two masses are in phase is similar to the x-10 mode, and the mode where the two masses are 180 degrees out of phase is similar to the x-11 mode. The ability of the two-mass model to self-oscillate can be explained, in large measure, by the existence of these two modes, which facilitate energy transfer from the airflow to the tissue (Stevens, 1977; Broad, 1979; Titze, 1988). For "typical" Ishizaka and Flanagan (1972) parameters, the natural frequencies of the normal modes are 120 Hz and 201 Hz, respectively (Titze, 1976).

Entrainment of the Modes

Self-oscillation during normal phonation also involves a 1:1 "entrainment" of the modes. Entrainment is a phenomenon in which a nonlinear coupling of system variables causes the natural frequencies of the system to shift so as to be related by an integer ratio. For example, a 1:1 entrainment in the two-mass model means that both modes oscillate at the same frequency. Such an entrainment has been observed over a wide range of parameters in the two-mass model (Herzel, Steinecke, Mende & Wermke, 1991). As predicted by Titze (1976), this entrainment occurs at a frequency very close to the natural frequency of the lowest-order mode. Over a certain range of parameters (e.g., those corresponding to low stiffness of the upper mass and weak coupling of the masses), the breakdown or desynchronization of this 1:1 entrainment has also been observed in the two-mass model (Herzel et al., 1991). In such parameter regions, various nonlinear phenomena have been observed including subharmonic regimes, beating-like toroidal oscillations (i.e., low-frequency modulations), and chaotic motion (Herzel et al., 1991).

Subharmonics, low-frequency modulations, and chaos are also commonly observed in patients with vocal disorders (Herzel & Wendler, 1991; Baken, 1991; Herzel, Berry, Titze & Saleh, in review) and during infant cries (Mende, Herzel & Wermke 1990). Consequently, this desynchronization of the modes is believed to be an essential mechanism of many vocal disorders (Titze et al., 1993; Herzel et al., in review).

Experimental Studies of Normal Modes

To date, most of the discussion of normal modes in vocal fold tissues has been in a theoretical sense. Direct measurement of the modes has proven problematic, partially because of the small dimensions of the vocal folds (on the order of 1 cm). Traditional modal analysis in which accelerometers are used to trace trajectories at various locations on a structure would undoubtedly yield unsatisfactory results. The number of accelerometers that could be placed on the folds would be limited, the weight of accelerometers might alter the modes significantly, and the

ability to firmly attach a device to the elastic tissue of the folds would be limited.

Perhaps new optical techniques and high-speed video show the greatest promise for an adequate modal analysis of the folds to be performed. Indeed, the most common way to observe vocal fold vibrations is through high-speed films and videostroboscopy. As noted previously, many vibration patterns similar to the theoretical normal modes have been observed using these techniques. However, vibration patterns are usually a combination of several modes, so the vibration patterns observed are rarely oscillations of a single mode. Still, if these optical techniques can be used to trace trajectories of vocal fold fleshpoints, empirical methods (e.g., such as the method of empirical eigenfunctions to be described) can be used to decompose the vibration patterns into distinct, orthogonal modes of vibration.

Although little experimental work has been done to quantify the normal modes of the vocal folds, some attempts have been made to measure the resonant frequencies of the folds (Kaneko, Masuda, Shimada, Suzuki, Hayasaki & Komatsu, 1986). Kaneko et al. (1986) used ultrasonic techniques to measure these resonant frequencies. Measurements were made immediately following a mechanical pulse-excitation to the thyroid cartilage, and immediately before the patients began phonating. Kaneko et al. (1986) observed two dominant resonances of the folds below 625 Hz. Interestingly, the lower resonance corresponded well to the phonation frequency that followed.

Modes In the Biomechanical Simulation

From the start, it is acknowledged that there is no substitute for direct measurement of the modes in vocal fold tissues. However, until this becomes feasible, there are additional theoretical approaches that can be used to investigate these modes, particularly with the help of a biomechanical simulation of vocal fold movement (Titze and Alipour, in review). The simulation uses a finite element approach to the solution of viscoelastic waves in a continuum (Titze and Talkin, 1979). A series of experimental studies have been performed to quantify the elastic properties of vocal fold tissues (Alipour & Titze 1985, 1991), and more work in this area is in progress. Indeed, the development of this simulation has been an effort to integrate many independent measurements and theoretical considerations into one coherent "picture" of vocal fold vibration.

Furthermore, with the simulation it is easy to trace the trajectories of points interior to the folds. With optical methods, only surface points can be tracked (Baer, 1981). Nevertheless, direct measurement of the modes in vocal fold tissues (perhaps with x-ray pellet techniques as described by Saito et al., 1985) will be invaluable, and may be a key to future refinements of the biomechanical simulation.

Empirical Eigenfunctions

As pointed out earlier, traditional modal analysis is limited by the fact that it is only valid for linear systems. However, there are many nonlinearities associated with the vocal folds. One of these is the nonlinear stress-strain curves of vocal fold tissues. Another is the nonlinear pressure-flow relation in the glottis. A third is the nonlinearity associated with vocal fold collision. While for small transient oscillations these nonlinearities might be neglected, self-sustained oscillation depends critically on at least one nonlinear constitutive equation. Indeed, for many vocal fold configurations linear dynamics is not even approximately true.

However, the method of empirical orthogonal eigenfunctions (Lorenz, 1956) has been used for many years to extract physically meaningful structures from nonlinear systems. For example, Lumley (1967) advocated the technique as a way to extract "coherent structures" from a turbulent flow. In recent years, the method has become a popular technique in a variety of problems in fluid dynamics (Sirovich, 1987; Aubry, Guyonnet & Lima, 1991; Deane, Kevrekidis, Karniadakis & Orszag, 1991; Armbruster, Heiland, Kostelich & Nicolaenko, 1992). The method is an application of a general technique familiar to many disciplines, and has also been referred to as the singular value decomposition (Golub & Van Loan, 1983), singular spectrum analysis (Vautard, Yiou & Ghil, 1992), principal-components analysis (Zahorian & Rothenberg, 1981), principal factor analysis (Johnson & Wichern, 1982), the Karhunen-Loève expansion (Fukunaga, 1972), the proper orthogonal decomposition (Lumley, 1967), and the bi-orthogonal decomposition (Aubry, Guyonnet & Lima, 1991).

Furthermore, Breuer and Sirovich (1991) have recently shown that, for a general class of linear systems, the empirical eigenfunctions actually reduce to the linear normal modes. The ability to extract physically meaningful structures from both linear and nonlinear systems makes the method of empirical orthogonal eigenfunctions a particularly useful tool for analyzing vocal fold movement. In the case of small-amplitude vibrations for which the tissue stress-strain curves are approximately linear, the empirical eigenfunctions should be related to the normal modes of vocal fold tissues. For larger amplitude vibrations for which tissue nonlinearities become important, the eigenfunctions should appear as distortions of the normal modes, i.e., a reflection of the new nonlinear phenomena (Breuer and Sirovich, 1991).

Moreover, the statistical nature of this technique makes it well-suited for the present investigation. In a sense, the method is "blind" to all the complexities of the biomechanical simulation (e.g., nonlinearities in stress-strain curves, complex geometry of the folds, layered tissue structure, tissue incompressibility which induces a coupling between lateral and vertical modes, aerodynamic forces,

collision forces, viscous losses). Such complexities forbid an analytical solution of the modes, but present no difficulties for the method of empirical eigenfunctions.

The method of empirical orthogonal eigenfunctions differs from a traditional normal mode analysis in that it does not determine "modes" directly from the equations of motion. Rather, "modes" are determined by statistical correlations of the output variables, i.e., a covariance matrix is generated and eigenvectors are computed. The eigenvectors are orthogonal and are guaranteed to be optimal in the sense that they regenerate the output data with minimum least-square error (for any arbitrary number of eigenvectors). Unlike a normal mode analysis, the method of empirical eigenfunctions does not calculate all the possible modes of a system. Rather, it only extracts those "modes" which are excited. For the present investigation, the excited "modes" are the focus of interest, and are used as a tool for interpreting vocal fold dynamics during self-oscillation.

Procedures

Trajectories from the Simulation

Empirical eigenfunctions were calculated based on the output of a biomechanical simulation of vocal fold movement (Titze and Alipour, in review). The simulation was run as part of a complete speech synthesis system, including sub- and supraglottal systems. The biomechanical model of the folds consists of nine longitudinal layers as shown in Fig. 2, where the posterior edge of the folds is in the foreground. Anterior and posterior boundaries are fixed. Each layer consists of 32 finite elements (triangles) or 26 nodes (fleshpoints), as shown on the left side of Fig. 3. The elements which correspond to the body (or muscle) are marked "B", the elements which correspond to the ligament are marked "L", and the elements which correspond to the cover (or mucosa) are marked "C". Each of these regions possesses distinct elastic properties. Three nodes per layer are placed on a fixed lateral boundary. Thus, there are 207 nodes per fold (9 layers x 23 nodes/layer) which are free to oscillate. As in earlier investigations, lateral and vertical motions are allowed, but no movement along the anterior-posterior direction. With two degrees of freedom per node, there are 414 total degrees of freedom if left and right folds are symmetric, and 828 if asymmetric. Although the simulation is equipped to handle asymmetric folds, all the runs for this analysis employed left-right symmetry.

Nodal trajectories for parameters corresponding to normal phonation are shown on the right side of Fig. 3. The trajectories are taken from the fifth longitudinal layer (the layer mid-way between anterior-posterior boundaries), which is the layer with the most lateral movement. Qualitative similarities exist between these trajectories and fleshpoint trajectories observed experimentally (Baer, 1981; Saito et al., 1981; Saito et al., 1985). The x and z coordinates from the trajectories of each of the 207 nodes were used as the input for calculating the covariance matrix, and the resulting eigenfunctions. Although the simulation was run at a sampling rate of 20 kHz, the nodal coordinates were only saved at a rate of 5 kHz, which was found to be sufficient. Frequencies above about 1 kHz were essentially non-existent in the trajectories (measured on a power spectrum, they were at least 40 dB below the strongest frequency).

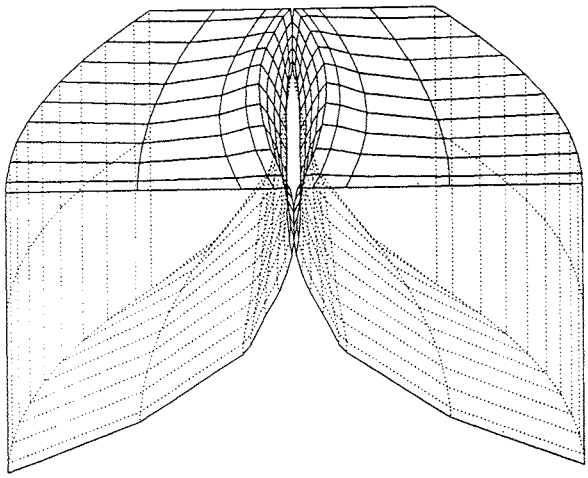


Figure 2. A view of the biomechanical simulation immediately before glottal closure is shown, with the posterior edge of the folds in the foreground. There are nine layers positioned along the anterior-posterior length.

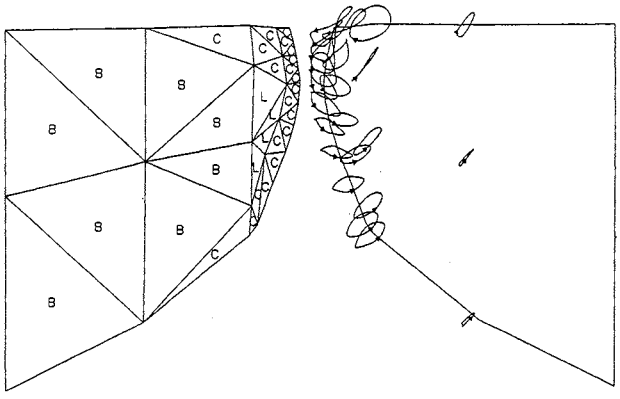


Figure 3. A coronal view of the fifth longitudinal layer. On the left side of the figure, the 32 elements/layer are displayed and distinguished as corresponding to the body ("B"), the cover ("C"), or the ligament ("L"). On the right side of the figure, trajectories of vocal fold fleshpoints are shown for parameters corresponding to normal phonation.

Calculation of Empirical Eigenfunctions

First, the nodal coordinates R_i were separated into mean and oscillatory components:

$$R_i(t) = \overline{R_i} + r_i(t), \quad i=1,2,...,414 \qquad (3)$$

where the bar denotes a mean value. The mean represents the dynamic equilibrium of the system, and the remaining oscillatory component represents the time-varying displacements about this equilibrium. A covariance matrix was generated using the time-varying displacements:

$$S_{ij} = \frac{1}{N}\sum_{k=1}^{N} r_i(t_k)\, r_j(t_k), \qquad i,j = 1,2,...,414 \qquad (4)$$

where t_k are the discrete times at which the coordinates are sampled, and N is the total number of time samples. The eigenvectors of the covariance matrix correspond to the empirical eigenfunctions. At any time t_k, the nodal displacements may be expressed as a linear combination of the empirical eigenfunctions ϕ_j (Deane et al., 1991):

$$r_i(t_k) = \sum_{j=1}^{414} a_j(t_k)\,\phi_j(i), \qquad i=1,2,...,414 \qquad (5)$$

where $\phi_j(i)$ is the i-th component of the j-th eigenfunction and $a_j(t_k)$ is the temporal coefficient of the j-th eigenfunction at time t_k. The temporal coefficients a_j may be computed by projecting the eigenfunctions ϕ_j onto the time-varying displacements:

$$a_j(t_k) = \sum_{i=1}^{414} r_i(t_k)\,\phi_j(i), \qquad j=1,2,...,414 \qquad (6)$$

The temporal coefficients themselves may be thought of as temporal eigenfunctions, and correspond to the eigenvectors of a temporal correlation matrix which may also be generated from the original data (Sirovich, 1987). Both the spatial and temporal eigenfunctions are orthogonal, and reveal distinct features of the dynamics of the system. The spatial eigenfunctions (sometimes referred to as "topos", Aubry et al., 1991) reveal topological patterns in the data and are analogous to the normal modes of linear systems. The temporal eigenfunctions (sometimes referred to as "chronos", Aubry et al., 1991) reveal information about possible entrainment of the modes, and capture the frequencies at which the modes oscillate.

Each pair of spatio-temporal eigenfunctions has a corresponding eigenvalue, which quantifies the degree to which the eigenfunctions can regenerate the nodal trajectories (in terms of variance). Often just a few eigenfunction pairs capture the essential dynamics of a system (Deane et al., 1991), which facilitates a reduction of the system as well as a physical interpretation of the dynamics.

All the covariance matrices calculated in this study were generated with one second of stationary output (5000 time frames). Initial transients and other nonstationary segments were not used in calculating the covariance matrices and resulting empirical eigenfunctions. The dominant vibration frequencies of these modes ranged between 80 and 160 Hz, so 80 to 160 cycles were used in calculating the modes.

Results

Normal Phonation

First of all, we consider the results of the analysis for typical parameters corresponding to "normal" phonation. The normalized eigenvalues computed for this simulation are shown in Table 1. The eigenvalues are shown in descending order. The far right column shows a cumulative sum of the eigenvalues. From this table, we see that the first eigenfunction explains about 72% of the variance of the nodal trajectories, and the second eigenfunction about 26% of the variance. Together the first two eigenfunctions explain approximately 98% of the variance, suggesting the dominance of just a few primary modes. These results were consistent over a wide range of elastic constants and subglottal pressures.

Table 1.
Normalized eigenvalues for modes of "normal" phonation (Young's Modulus of the cover, E_c = 2 kPa).

Mode Number	λ_i (%)	Cumulative sum of λ_i (%)
1	72.5	72.5
2	25.2	97.7
3	1.5	99.2
4	0.5	99.7

A coronal view of the first eigenfunction is shown in Fig. 4a. Frames 1 and 2 display maximum and minimum excursions of the eigenfunction, respectively (solid lines). The dotted lines show the mean coordinate values. By examining the motion of the folds near the top of the glottal air passage (e.g., see the top five medial nodes which are bolded on either side), one can note the correspondence of this eigenfunction with the x-11 mode. That is, there is a higher and lower portion of the folds which are 180 degrees out of phase. Consequently, this eigenfunction is largely

responsible for alternately shaping a divergent (frame 1) and convergent (frame 2) glottis. In addition, there is considerable vertical motion similar to the z-10 mode. This coupling of x and z modes is not surprising given the incompressibility of the tissue (Titze, 1976). Tissue incompressibility implies that the overall tissue volume does not change, so if the folds are compressed laterally, they must bulge out vertically, and vice versa.

A coronal view of the second eigenfunction is shown in Fig. 4b. Again frames 1 and 2 show maximum and minimum excursions of this eigenfunction. Again, note that near the top of the folds (the region that might approximate a rectangular parallelepiped), this eigenfunction is qualitatively similar to the x-10 mode (Fig. 1a), and is largely responsible for the net lateral movement of the folds in this region.

Fig. 5 shows the temporal coefficients associated with each of these eigenfunctions; the solid line displays the temporal coefficients for the first eigenfunction and the dotted line illustrates the time coefficients for the second eigenfunction. The temporal coefficients for both eigenfunctions are nearly sinusoidal with a sine/cosine relationship (mode 1 lags mode 2 by about 90 degrees). A simple analysis shows that the modes are synchronized in such a way that energy transfer may occur from the airflow to the tissue, enabling self-oscillation. Specifically, note that the solid line can be expressed as sin(t). A maximum in this line occurs for a divergent glottis (see Fig. 4a, frame 1), and a minimum in the solid line occurs for a convergent glottis (see Fig. 4a, frame 2). If Bernoulli's law is taken as approximately valid, then the intraglottal pressure will be relatively low for a divergent glottis and relatively high for a convergent glottis (Titze, 1988). As a first-order approximation, one might say that the intraglottal pressure is in phase with -sin(t).

The dotted line can be expressed as cos(t). A maximum in this line occurs when the folds are most opened (Fig. 4b, frame 1), and a minimum occurs when the folds are closed (Fig. 4b, frame 2). Because this mode roughly corresponds to the net lateral displacement of the tissue, a rough estimate of the net lateral velocity of the tissue is given by the time derivative of cos(t), or -sin(t). Thus, an examination of the two dominant spatio-temporal eigenfunctions of the biomechanical simulation reveals an in-phase relationship between the intraglottal pressure and the net tissue velocity, which allows energy transfer from the airflow to the tissue.

It is already well-known that this condition must be satisfied if self-oscillation of the folds is to occur in the presence of dissipation. However, the important point is that this method of analysis reduced several hundred trajectories of the biomechanical simulation to essentially two modes of vibration. With this reduction, the dynamics of a biomechanical model with many degrees of freedom could be discussed and interpreted as easily as the dynamics of a much more constrained, low-order model. The ability to reduce large amounts of data to essential dynamics will be crucial for understanding more complex output from the biomechanical simulation.

As a word or caution, it should be noted that because the biomechanical simulation was reduced to essentially two modes of vibration for parameters corresponding to normal phonation, the biomechanical model was in no way reduced to a two-mass model. Even for normal phonation, the most dominant mode of the biomechanical simulation was not simply a lower-order x-mode such as might be captured by a two-mass model, but an x-mode coupled with a z-mode. Furthermore, although the modes of a two-mass model may be qualitatively similar to

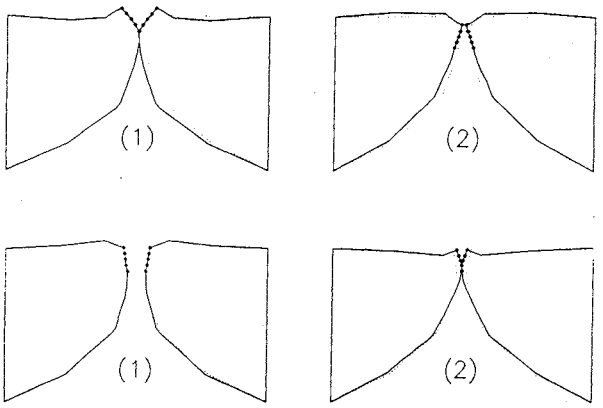


Figure 4. A coronal view of the two strongest spatial eigenfunctions for normal phonation. The first eigenfunction is shown in (a, top) and the second eigenfunction is shown in (b, bottom). In both cases, frame 1 corresponds to a maximum excursion of the eigenfunction, and frame 2 corresponds to a minimum excursion.

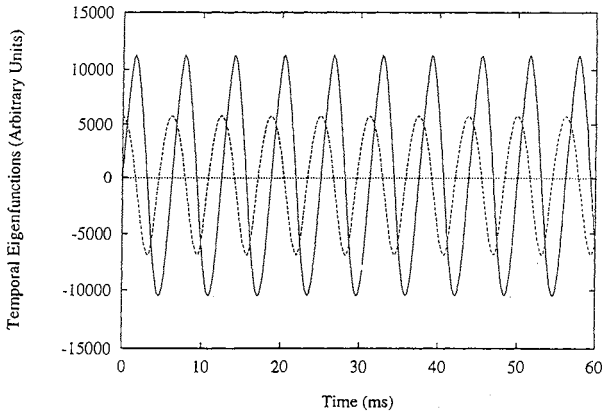


Figure 5. The two dominant temporal eigenfunctions for parameters corresponding to normal phonation. The solid line corresponds to eigenfunction 1 and the dotted line corresponds to eigenfunction 2.

the lower-order modes of a simplified elastic continuum, it is questionable whether two bar-shaped masses can adequately capture the smoothly varying shape of the glottis. The discontinuities introduced by such gross spatial discretization would likely have an adverse effect on synthesis.

Moreover, the biomechanical simulation has hundreds of degrees of freedom which allow it to be excited into many modes of vibration not possible for the two-mass model. The fact that just a few of the lower-order modes are excited for a range of parameters corresponding to normal phonation is to be expected and might even be viewed as one validation of the biomechanical simulation. For other parameter configurations, additional modes are excited in the simulation. The study of these modes may yield additional insights into vocal fold dynamics, and may have relevance for an understanding of voice disorders.

Chaotic Oscillations

Because the vocal folds are nonlinear systems with many degrees of freedom, bifurcations and chaos should be expected for certain parameter configurations (Glass & Mackey, 1988; Titze et al., 1993). Indeed, bifurcations and chaos appear in some of the simplest models of vocal fold vibration (Awrejcewicz, 1990; Herzel et al., 1991). They are also observed in more complex models (Wong, Ito, Cox & Titze, 1991), in models which incorporate left-right asymmetries (Ishizaka & Isshiki, 1976; Wong et al., 1991; Smith, Berke, Gerratt & Kreiman, 1992), and most recently in finite element simulations of the folds (Titze et al., 1993). Furthermore, an acoustical analysis of many types of rough voice (e.g., creaky voice, vocal fry, vocal disorders and newborn infant cries) reveals an intimate relationship between voice mechanics and bifurcations and chaos (Herzel & Wendler, 1991; Herzel et al., in review, Mende et al., 1990).

Previously, it has been suggested that the origin of chaos in vocal fold vibrations may be the desynchronization of a few of the oscillatory modes (Titze et al., 1993; Herzel et al., in review). Specifically, creaky voice or vocal fry is thought to be induced by a lax cover, which could lead to the desynchronization of a few of the low-order x and z modes (Herzel et al., in review).

To investigate this hypothesis, the lax cover was simulated by decreasing the transverse Young's modulus of the cover (E_c) in the biomechanical simulation. Starting at 2 kPa (which is the value of E_c used in the simulation of "normal" phonation), E_c was gradually lowered and the resulting acoustic output was observed. No unusual behavior was noticed until E_c reached values below 0.6 kPa, at which point the signal became irregular, and perceptually rough. At 0.4 kPa, the signal became regular again, but with a doubling of the original period (an "octave jump"), which appeared as alternating high and low amplitudes in the acoustic output. Such phenomena (e.g., irregular oscillations, low frequencies, and alternating high/low amplitudes) are characteristic of the acoustic output of creaky voice (Hollien & Michel, 1968). Listening to the acoustic output also gave the perception of creaky voice.

A spectral bifurcation diagram (e.g., Lauterborn, 1986) is shown in Fig. 6, where E_c is slowly varied from 0.35 to 0.65 kPa. From left to right, one views transitions from a subharmonic regime to chaos to the periodic regime characteristic of normal phonation. This figure shows striking similarities to spectrograms of newborn cries (Mende et al., 1990) and to acoustic cavitation experiments (Lauterborn, 1986). A more complete bifurcation analysis of this region will be treated in a forthcoming paper.

For the present investigation, empirical eigenfunctions were determined at $E_c = 0.4$ kPa and $E_c = 0.5$ kPa. Table 2 shows the eigenvalues for both parameter configurations. At $E_c = 0.4$ kPa, four eigenfunctions are needed to describe the nodal trajectories in as much detail

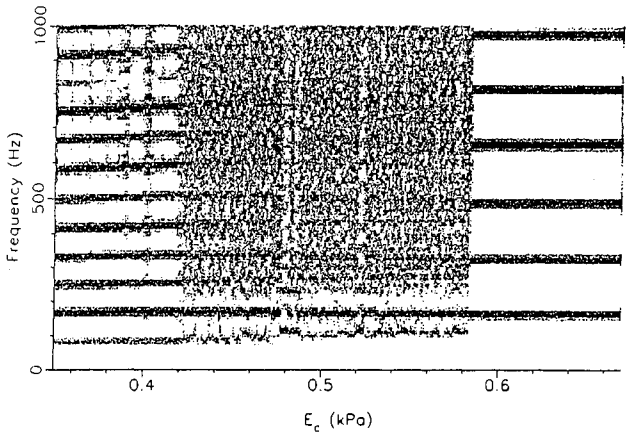


Figure 6. A spectral bifurcation diagram as E_c is slowly varied from 0.35 to 0.65 kPa, in increments of 0.01 kPa every 400 ms. Transitions from a subharmonic regime to chaos to periodic motion are displayed.

Table 2.
Normalized eigenvalues for modes of E_c = 0.4 kPa and 0.5 kPa.

	E_c = 0.4 kPa		E_c = 0.5 kPa	
Mode Number	λ_i (%)	Cumulative sum of λ_i (%)	λ_i (%)	Cumulative sum of λ_i (%)
1	43.9	43.9	45.6	45.6
2	30.9	74.8	27.0	72.6
3	16.1	91.0	12.5	85.1
4	7.1	98.1	5.2	90.3
5	0.7	98.8	3.1	93.4
6	0.5	99.3	1.6	95.0

(in terms of variance) as the first two eigenfunctions at E_c = 2 kPa (see Table 1). At E_c = 0.5 kPa, additional eigenfunctions are needed. However, even for the complicated, nonperiodic behavior at E_c = 0.5 kPa, relatively few eigenfunctions are needed to capture most of the variance of vocal fold dynamics. Out of 414 possible eigenfunctions, only six are needed to describe the motion in considerable detail.

Moreover, the first three spatial eigenfunctions at E_c = 0.5 kPa are essentially equivalent to the first three spatial eigenfunctions at E_c = 0.4 kPa (approximately 90% agreement). Table 3 shows the dot product of the first three eigenfunctions of the two parameter configurations. Furthermore, the three eigenfunctions of E_c = 0.5 kPa also explain over 90% of the variance for higher values of E_c corresponding to normal phonation. Thus, during the entire cascade, three low-order "modes" explain most of the variance.

The fact that this simulation of partial differential equations (PDE's) can be projected onto just a few eigenfunctions is reminiscent of the findings of Saltzman (1962) and Lorenz (1963) in relation to Bernard convection. In their studies, it was found that close to the onset of convection there were only a few dominant modes, which led to the derivation of the celebrated Lorenz equations (1963). While Lorenz employed a trigonometric expansion, in this study empirical eigenfunctions might be appropriate to reduce the original PDE's to a small set of ordinary differential equations (ODE's). Such reductions may be useful for the design of lower-order models (which can nevertheless simulate various vocal qualities), and may help facilitate bifurcation analyses over specific parameter regions of the model (Deane et al., 1991).

The essential difference in the system at E_c = 0.4 kPa and E_c = 0.5 kPa is revealed by the temporal eigenfunctions, as illustrated in Fig. 7. The temporal eigenfunctions of E_c = 0.4 kPa are nearly periodic, while the temporal eigenfunctions of E_c = 0.5 kPa are nonperiodic.

Table 3.
Dot product of first three modes of Ec = 0.4 kPa and 0.5 kPa

Mode Number	$< \phi_{i,EC=0.4} \mid \phi_{i,EC=0.5} >$
1	0.913
2	0.883
3	0.901

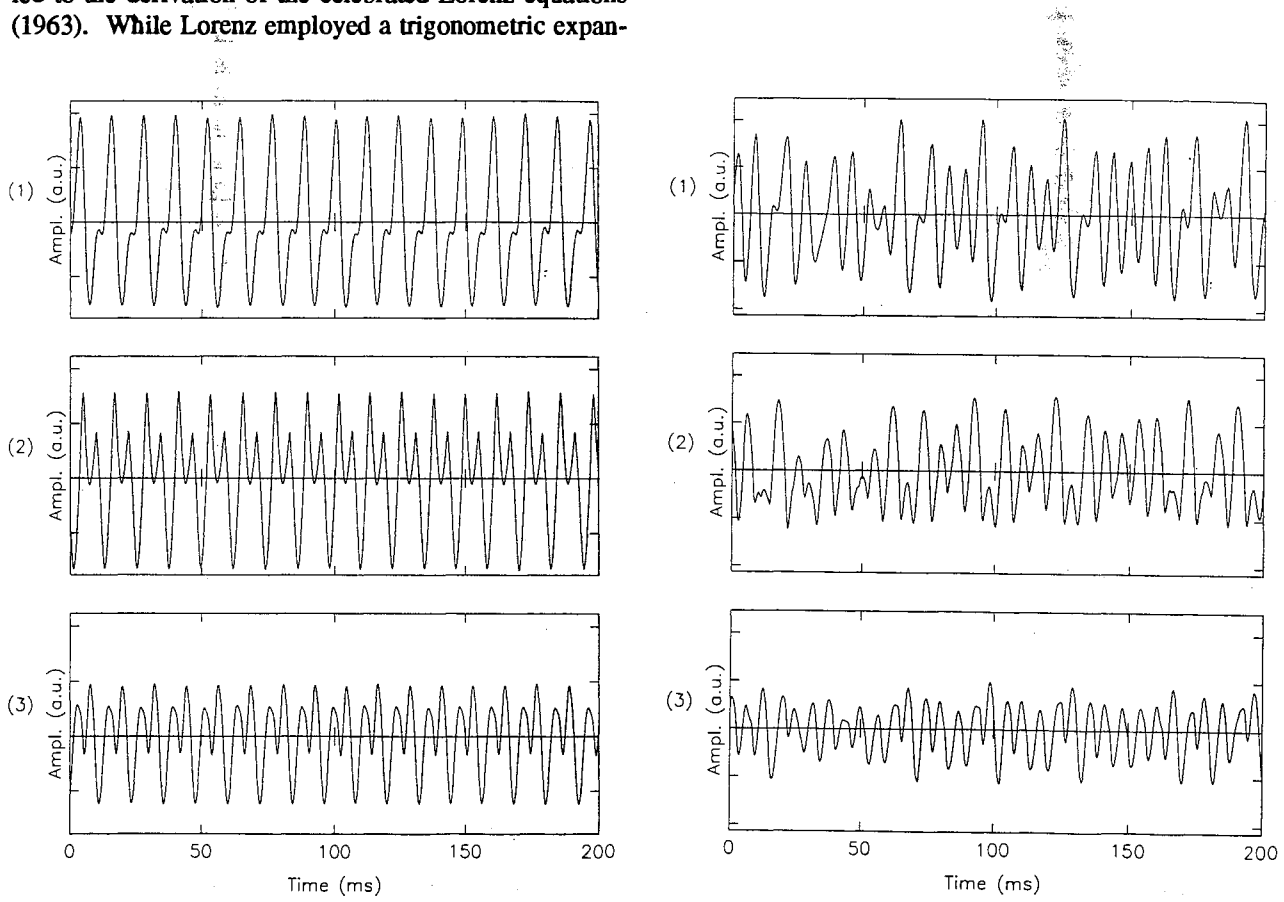


Figure 7. The first four temporal eigenfunctions for E_c = 0.4 kPa (a - left) and E_c = 0.5 kPa (b - right).

Phase portrait reconstructions of the attractors for $E_c = 0.4$ kPa and $E_c = 0.5$ kPa were generated using the first three temporal eigenfunctions of each configuration. The temporal eigenfunctions of $E_c = 0.4$ kPa results in the attractor of Fig. 8a, which shows weakly-modulated periodic motion. The nonperiodic temporal eigenfunctions of $E_c = 0.5$ kPa yield the chaotic attractor shown in Fig. 8b. These results lend support to our earlier claims that the origin of chaos in vocal fold vibrations is the desynchronization of a few of the low-order modes.

The first three spatial eigenfunctions of $E_c = 0.5$ kPa are shown in Fig. 9. In this case, it may not be possible to claim a definite relationship between the empirical eigenfunctions and the normal modes of the simplified folds. Indeed, many factors (e.g., tissue incompressibility, complex geometry, nonlinearities) may cause significant deformations in the modes of vibrations. Nevertheless, the eigenfunctions appear to be manifestations of simple, low-order modes. For example, the first eigenfunction (Fig. 9a) shows some resemblance to a z-10 mode, the second eigenfunction to an x-11 mode, and the third eigenfunction to an x-10 mode.

In addition, the sixth eigenfunction for $E_c = 0.5$ kPa is analogous to a higher-order normal mode (e.g., the x-21 mode). Although not as commonly observed, these higher-order modes have been viewed occasionally with high-speed cinematography (Rubin & Hirt, 1960). Fig. 10 shows a superior view of this eigenfunction. This eigenfunction did not appear, at least as clearly, in the more stable oscillations corresponding to $E_c = 0.4$ kPa and $E_c = 2$ kPa. This may be related to the fact that this is an unstable eigenfunction, and is thus usually only excited during more unstable, nonperiodic vibrations. Even in the complex oscillations from which this eigenfunction was extracted, the higher-order eigenfunction was so weak that it could not be visually detected in the overall vibration pattern.

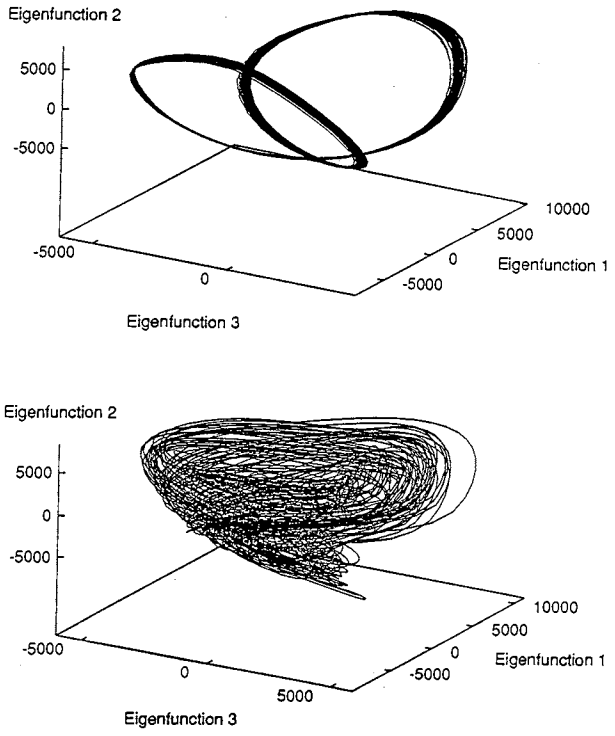


Figure 8. Phase portrait reconstructions of attractors for $E_c = 0.4$ kPa (a, top) and $E_c = 0.5$ kPa (b, bottom) using the first three temporal eigenfunctions from each. Weakly-modulated periodic motion is shown in (a) and a chaotic attractor in (b).

Discussion

As is well-known from high-speed films, stroboscopy, and sophisticated models, vocal fold vibrations exhibit complex three-dimensional patterns. However, normal phonation produces fairly periodic acoustic output. These observations may be explained by the fact that only a few modes are excited and all the modes are entrained. This concept has been substantiated through examining the empirical eigenfunctions of a biomechanical simulation of vocal fold vibrations during self-oscillation. Even though hundreds of degrees of freedom exist, two eigenfunctions explain 98% of the variance of the nodal trajectories. By viewing the high-order simulation as a superposition of just two dominant eigenfunctions, an interpretation of the mechanism of self-oscillation of the folds is facilitated. These eigenfunctions are qualitatively similar to x-11 and x-10 modes (although the x-11 mode is also coupled with a z-10 mode which is partially a result of tissue incompressibility).

The technique of empirical eigenfunctions is also useful for describing irregular oscillations related to rough voice. Changing parameters of the biomechanical model leads to subharmonic regimes and chaos. However, despite complex motion, a relatively small number of eigenfunctions

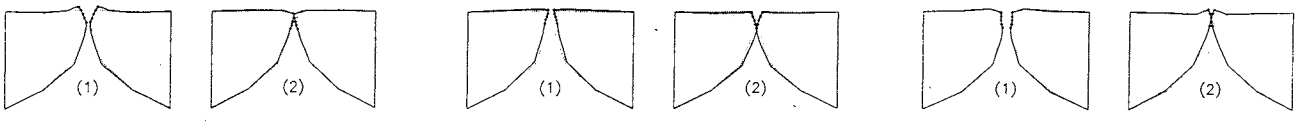


Figure 9. Coronal views of the first three spatial eigenfunctions at $E_c = 0.5$ kPa. Eigenfunction 1 shows some resemblance to a z-10 mode, eigenfunction 2 to an x-11 mode, and eigenfunction 3 to an x-10 mode. Coronal views are shown as in Figure 4.

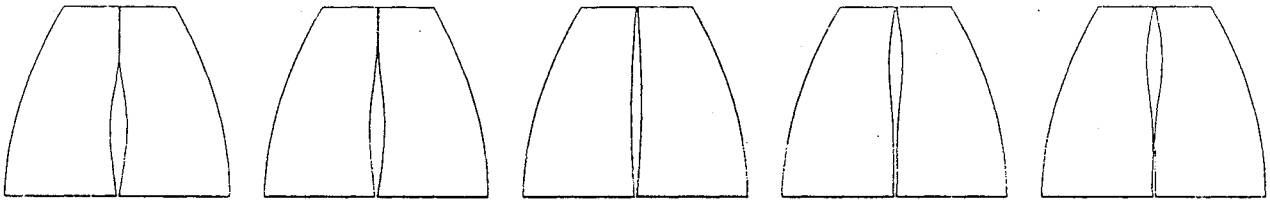


Figure 10. A superior view of the sixth spatial eigenfunction for E_c = 0.5 kPa. This eigenfunction is similar to an x-21 mode. To illustrate this eigenfunction, a series of 5 sequential snapshots are shown from left to right.

captures the essential dynamics of the folds. In addition, some of the more subtle dynamics captured by "weaker" eigenfunctions correspond with higher-order normal modes.

Although the oscillation patterns in the simulation changed substantially with decreasing stiffness of the cover, the spatial "modes" remained more or less the same. Consequently, it is shown that the appearance of chaos in vocal fold oscillations may be understood in terms of a desynchronization of a few of the low-order "modes".

Furthermore, such eigenfunctions may be useful for designing lower-order models capable of simulating specific vocal disorders, and for performing bifurcation analyses. The technique can also be useful for evaluating and refining biomechanical simulations of the folds, and for assessing the impact of various design parameters on modal shapes. In particular, one might start from the conditions of the simplified folds for which analytic solutions of the normal modes are known (Titze & Strong, 1975). Then the deformation of the normal modes caused by nonlinearities, complex geometry, and tissue incompressibility may be observed systematically and independently as one complexity at a time is added to the system. The technique can also be used to determine eigenfunctions from empirical data obtained directly from vocal fold tissues, provided such data can be obtained. In general, the method of empirical eigenfunctions enhances the study of vocal fold dynamics by allowing the principal modes of vibration to be extracted during self-oscillation, despite inherent nonlinearities.

Acknowledgments

This work was supported by Grant No. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. The authors also thank I.G. Kevrekidis for introducing them to the method of empirical eigenfunctions.

References

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

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

Armbruster, D., R. Heiland, E.J. Kostelich and B. Nicolaenko (1992). Phase-space analysis of bursting behavior in Kolmogorov flow. Physica D 58, 392-401.

Aubry, N., R. Guyonnet and R. Lima (1991). Spatiotemporal analysis of complex signals: theory and applications. J. Stat. Phys. 64, 683-739.

Awrejcewicz, J. (1990). Bifurcation portrait of the human vocal cord oscillations. Journal of Sound and Vibration 136, 151-156.

Baer, T. (1981). Investigation of the phonatory mechanism. ASHA Reports (11), 38-46.

Baken, R.J. (1991). Géométrie fractale et évaluation de la voix: Application préliminaire á la dysphonie. Bull. d'Audiophonogie. Ann. Sc. Univ. Franche-Comté 7(5-6), 731-749.

Breuer, K.S. and L. Sirovich (1991). The use of the Karhunen-Loève procedure for calculation of linear eigenfunctions, Journal of Computational Physics 96, 277-296.

Broad, D. (1979). The new theories of vocal fold vibration. In *Speech and Language: Advances in Basic Research and Practice*, N. Lass (ed.), New York, NY: Academic Press.

Deane, A.E., I.G. Kevrekidis, G.E. Karniadakis and S.A. Orszag (1991). Low-dimensional models for complex geometry flows: application to grooved channels and circular cylinders, Phys. Fluids A 3(10), 2337-2354.

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

Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*, New York, NY: Academic Press.

Glass, L. and M.C. Mackey (1988). *From Clocks to Chaos*. Princeton, NJ: Princeton University.

Golub, G.H. and C.F. Van Loan (1983). *Matrix Computations*. Baltimore, MD: John Hopkins University Press.

Herzel, H., I. Steinecke, W. Mende and K. Wermke (1991). Chaos and bifurcations during voiced speech. In *Complexity, Chaos, and Biological Evolution*, E. Mosekilde (ed.), New York, NY: Plenum Press, 41-50.

Herzel, H. and J. Wendler (1991). Evidence of chaos in phonatory samples. Proceedings of EUROSPEECH, Genova.

Herzel, H., D.A. Berry, I.R. Titze and M. Saleh (in review). Analysis of vocal disorders with methods from nonlinear dynamics. J. Speech. Hear. Res.

Hirano, M. (1975). Phonosurgery: Basic and Clinical Investigations. Official Rep. 76th Annual Convention Oto-Rhino-Laryngo. Soc., Kurume, Japan.

Hollien, H. and J. Michel (1968). Vocal fry as a phonational register. J. Speech. Hear. Res. 11, 600-604.

Ishizaka, K. and J.L. Flanagan (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords, Bell Syst. Tech. J. 51, 1233-1267.

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

Johnson, R.A. and D.W. Wichern (1982). *Applied Multivariate Statistical Analysis*, Englewood Cliffs, NJ: Prentice-Hall.

Kaneko, T., T. Masuda, A. Shimada, H. Suzuki, K. Hayasaki, K. Komatsu (1986). Resonance characteristics of the human vocal folds in vivo and in vitro by an impulse excitation. In *Laryngeal Function in Phonation and Respiration*, T. Baer, C. Sasaki, K. Harris (eds.), 349-377. Boston, MA: Little, Brown and Company.

Lauterborn, W. (1986). Acoustic turbulence. In *Frontiers in Physical Acoustics*, D. Sette (ed.), Amsterdam, North-Holland: Italian Physical Society, 123-144.

Lorenz, E.N. (1956). Technical Report 1, Statistical Forecasting Program, Department of Meteorology, Massachusetts Institute of Technology.

Lorenz, E.N. (1963). Deterministic nonperiodic flow. J. Atmos. Sci. (20), 130-141.

Lumley, J.L. (1967). The structure of inhomogeneous turbulent flows. In *Atmosphere Turbulence and Radio Wave Propagation*, A.M. Yaglom and V.I. Tatarski (eds.), Moscow: Nauka, 166-178.

Mende, W., H. Herzel and K. Wermke (1990). Bifurcations and chaos in newborn cries. Phys. Letters A (145), 418-424.

Moore, G. and H. Von Leden (1958). Dynamic variations of the vibratory pattern in the normal larynx. Folia Phoniatrica (10), 205-238.

Rubin, H.J. and C.C. Hirt (1960). The falsetto: a high-speed cinematographic study. Laryngoscope 70, 1305-1324.

Saito, S., H. Fukuda, Y. Isogai and H. Ono (1981). X-ray stroboscopy. In *Vocal Fold Physiology*, K. Stevens (ed.), 95-106. Tokyo: University of Tokyo Press.

Saito, S., H. Fukuda, S. Kitahira, Y. Isogai, T. Tsuzuki, H. Muta, E. Takyama, T. Fujika, N. Kokawa and K. Makino (1985). Pellet tracking in the vocal fold while phonating: experimental study using canine larynges with muscle activity. In *Vocal Fold Physiology*, I.R. Titze and R.C. Scherer (eds.), Denver Center for the Performing Arts: Denver, CO.

Saltzman, B. (1962). Finite amplitude free convection as an initial value problem. J. Atmos. Sci. 19, 329-341.

Sirovich, L. (1987). Turbulence and the dynamics of coherent structures, Parts I-III, Quarterly of Applied Mathematics XLV, 561-590.

Smith, M.E., G.S. Berke, B.R. Gerratt and J. Kreiman (1992). Laryngeal Paralyses: Theoretical considerations and effects on laryngeal vibration. J. Speech. Hear. Res. 35(3), 545-554.

Stevens, K.N. (1977). Physics of laryngeal behaviour and larynx modes. Phonetica 34, 264-279.

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

Titze, I.R. (1976). On the mechanics of vocal-fold vibration, J. Acoust. Soc. Am. 60(6), 1366-1380.

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

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

Titze, I.R., R. Baken and H. Herzel (1993). Evidence of chaos in vocal fold vibration. In *Vocal Fold Physiology: Frontiers in Basic Science*, I. Titze (ed). San Diego, CA: Singular Publishing Group, Inc.

Titze, I.R. and F. Alipour-Haghighi (in review). Three source models for voice synthesis, J. Acoust. Soc. Am.

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

Vautard, R., P. Yiou and M. Ghil (1992). Singular-spectrum analysis: a toolkit for short, noisy chaotic signals. Physica D 58, 95-126.

Zahorian, S.A. and M. Rothenberg (1981). Principal-components analysis for low-redundancy encoding of speech spectra. J. Acoust. Soc. Am. 69(3), 832-845.

NCVS Status and Progress Report - 5
November, 1993, 43-47

The Effect of Subglottal Pressure on Fundamental Frequency of the Canine Larynx With Active Muscle Tensions

Tzu-Yu Hsiao, M.D.
Departments of Otolaryngology and Physiology,College of Medicine, National Taiwan University
Nancy Pearl Solomon, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Erich S. Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Kang Liu, M.D.
Department of Speech Pathology and Audiology, The University of Iowa
Tsu-Ching Fu, Ph.D.
Department of Physiology, College of Medicine, National Taiwan University
Mow-Ming Hsu, M.D.
Department of Otolaryngology, College of Medicine, National Taiwan University

Abstract

The relation between subglottic pressure and fundamental frequency of vocal fold vibration was studied by means of evoked phonation in an in vivo canine model. The evoked-phonation model involved electrical stimulation of the midbrain that resulted in consistent responses by respiratory and laryngeal musculature accompanied by phonation. The dynamic stiffness properties of the vocal folds, especially the "cover," were investigated by delivering various amounts of air pressure to the larynx from an opening in the trachea. Fundamental frequency of vocal fold vibration increased linearly with subglottic pressure. The slopes ranged from 22.4 to 118.7 Hz/kPa in 7 animals. The results indicated that the dependence of fundamental frequency on subglottic pressure is a passive mechanical phenomenon.

Introduction

The vocal fold has a unique multilayered structure. Owing to this characteristic structure, the concept of a cover-body complex has become a prevalent hypothesis in explaining vocal fold function [1]. Modes of vocal fold vibration are determined by a combination of the mechanical characteristics of the cover and the body of the vocal folds [2]. In canines, vocal fold oscillation during phonation is confined primarily to the cover of the vocal folds [1,3,4]. The cover includes a surface layer of mucus, the stratified squamous epithelium, and Reinke's space also known as the superficial layer of lamina propria.

For fundamental frequency (F_0) control, the coordinated action of the cricothyroid and thyroarytenoid muscles, as well as other intrinsic and extrinsic laryngeal muscles, results in the effective stiffness of the cover and body of the vocal folds. In addition, subglottic pressure (P_s) influences the F_0 of vocal fold vibration to some degree [5-9]. The influence of P_s on F_0 can be included in the effective stiffness of the vocal folds [9]. As P_s increases, the amplitude of lateral excursions of the vocal folds will increase; consequently, vocal fold tension increases dynamically. Theoretically, if the coordinated actions of the intrinsic laryngeal muscles can be held constant, the relation between P_s and the F_0 of vocal fold vibration should be linear [9].

In this experiment, we used an in vivo canine model of evoked phonation by electrical stimulation of the midbrain. Such stimulation elicits activation of laryngeal and respiratory muscles that results in a phonatory response [10]. The response is repeatable and robust over many trials of stimulation. Known air pressures were applied to the larynx through an opening in the trachea. Under these conditions, the relation between P_s and F_0 was investigated in order to determine whether the results of previous studies of the cover-body hypothesis, conducted in the absence of active muscle tensions, would be the same as the results obtained with naturally coordinated activity of the laryngeal muscles.

Method

The study was conducted with 7 hound-like mongrel dogs, weighing approximately 20 kg. The animals were anesthetized to surgical levels (absence of corneal and deep pain reflexes) with pentobarbital (IV, 25mg/kg). The adequacy of the anesthesia was checked frequently, and additional pentobarbital was given as needed. At the conclusion of the experiment, animals were euthanized with a large dose of pentobarbital.

The animals were positioned for the evoked phonation experiment in a stereotaxic apparatus. A schematic representation of the stereotaxic frame and electrode carrier, with representation of the various measurements that were made, is provided in Fig. 1. The animal's head was fixed by using earbars inserted firmly in the external auditory meati, infraorbital ridge holders which were in the plane of the earbars, and a maxillary brace. The cortex was exposed by cutting a 1.5 cm diameter hole through the parietal bone and lacerating the underlying dura; this opening was centered 10 mm anterior and 5 mm lateral to ear bar zero. At these stereotaxic coordinates, a coaxial bipolar electrode (Rhodes NE-100) was inserted vertically into the brain to an initial depth of 20 cm dorsal to ear bar zero.

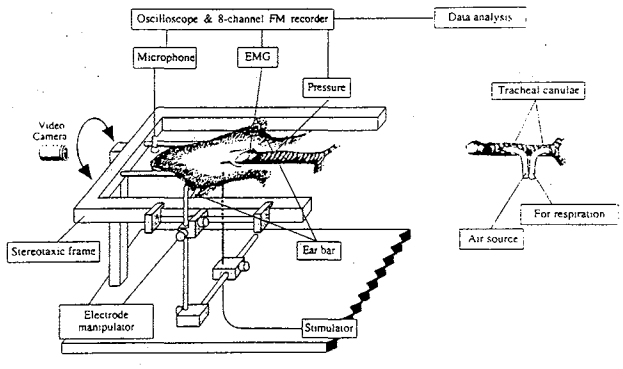


Figure 1. Schematic representation of the experimental set-up used.

Before beginning exploration of the midbrain with electrical stimulation for the purpose of eliciting phonation, the animal was rotated from a prone to a supine position to allow access to the ventral neck. This is because rotation of the animal can produce a change in the position of the brain, and previously established responses to stimulation may be lost.

A midline skin incision of the ventral neck was made and the larynx and trachea were identified. Bipolar stainless steel hooked-wire electrodes (75 um diameter) were inserted into cricothyroid (CT), thyroarytenoid (TA), lateral cricoarytenoid (LCA), and posterior cricoarytenoid (PCA) muscles for electromyographic (EMG) recording. The locations of the electrodes were verified by dissecting the related muscles after the experiment. Two low-pressure cuffed cannulae were inserted in the trachea, one between the third and fourth and one between the sixth and seventh tracheal rings. The more caudally located cannula was used for the animal's natural respiration. The more rostral cannula was connected to an air source that supplied warmed and humidified air to the larynx (Concha-Therm II). The pressure in the trachea, controlled with a pressure regulating valve (Fairchild Model 10), was measured with a pressure transducer (Micro Switch 143PC03G).

With the animal remaining in a supine position, electrical stimulation to sites within the midbrain was delivered until vocalization was elicited. Cites from 5 to 20 mm dorsal to ear bar zero, at 1 mm intervals, were stimulated. If acceptable phonation was not obtained with stimulation at any depth with the initial anterior-posterior and medial-lateral coordinates, the electrode was withdrawn and additional vertical tracks were made at 1 mm deviations from previously attempted locations. Electrical stimuli consisted of a 2-3 s train of 0.2 ms pulses at a rate of 200 Hz. An electrode placement was considered acceptable for the experiment if low-pitched phonation with minimal cricothyroid muscle activity (Fig. 2) was elicited consistantly at low current levels (0.3 - 0.5 ma). Once selected, the site for electrical stimulation of the midbrain was constant throughout the experiment.

With concurrent applications of electrical stimulation to the midbrain and air pressure to the larynx, phonation could be sustained. During the evoked response it was necessary to occlude partially the outlet of the tracheal cannula connected to the lungs. If left open, the lack of the respiratory resistance ordinarily provided by the closed glottis caused the lungs to deflate rapidly, producing a series of rapid brief phonations associated with interruptions of the activity of the laryngeal muscles. This effect has been described previously [11-13] and is known to be a reflex mediated by afferents in the vagus nerve. The air pressure delivered to the larynx (P_s) during evoked phonation ranged from the lowest to the highest pressures that could sustain phonation in the modal, or chest, register (about 0.3 to 4

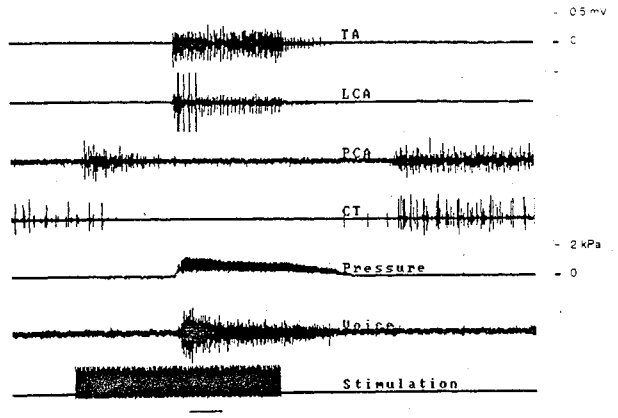


Figure 2. A typical response to a 2-s electrical stimulation (bottom trace) from four intrinsic laryngeal muscles (thyroarytenoid (TA), lateral cricoarytenoid (LCA), posterior cricoarytenoid (PCA), and cricothyroid (CT)), subglottic pressure (provided by the dog's own lungs), and voice in dog #5 is illustrated. The electrode placement in the midbrain was 11 mm anterior, 5 mm lateral, and 14 mm dorsal to earbar zero. Note that TA and LCA activity is present and PCA and CT activity is absent during phonation. The bar at the bottom of the figure represents 300 ms. During this time, mean P_s is 1.2 kPa, and mean F_0 is 135 Hz.

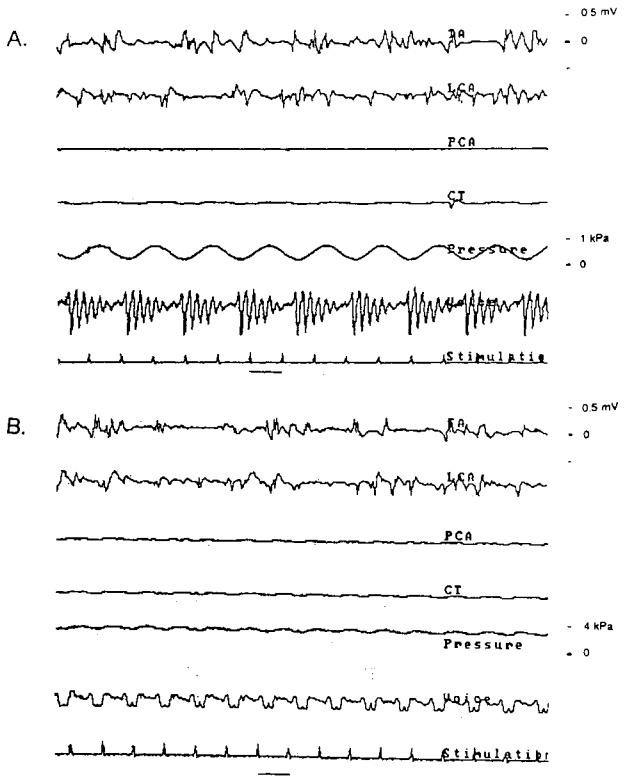


Figure 3. Subglottic pressure (P_s) is applied to the larynx at a low level (A: P_s = 0.5 kPa) and a high level (B: P_s = 2.7 kPa) during midbrain stimulation. Thyroarytenoid (TA) and lateral cricoarytenoid (LCA) activity appear similar for the two conditions. Voice fundamental frequency (F_0) is 109 Hz in A and 186 Hz in B. The time bar is 5 ms.

kPa). Pressure was increased in increments of about 0.1 kPa in an ascending series of phonations, and then, after a period of rest, a descending series was recorded. The P_s was approximately constant during each midbrain stimulation. Analysis was, however, based upon the directly recorded pressure in the trachea. The voice was recorded by an electret microphone placed just outside the animal's mouth.

Data were recorded on an 8-channel FM data tape recorder (Hewlett Packard 3968A, bandpass DC-2.5kHz). Recorded signals were a midbrain-stimulation marker, EMG from laryngeal muscles (including TA, LCA, PCA, CT), subglottic pressure, and voice. The recorded data were subsequently processed with an IBM PC data acquisition and processing program (Codas DATAQ). F_0 was calculated from the voice signal using an FFT analysis function. Mean P_s was obtained by averaging the pressure signal during phonation over a period of approximately 0.5 s, beginning about 0.5 s after the onset of phonation. At least 16 simultaneous measures of P_s and F_0 were obtained from each animal. The data pairs were plotted and analyzed statistically using a graphics and analysis program for the PC (Sigmaplot 4.0).

Results

Consistent low-pitched phonation was elicited from each animal during midbrain stimulation. Laryngeal activity during phonation was characterized by increased EMG activity of the TA and LCA muscles and suppression of CT and PCA muscle activity (Fig. 2). The level of activation of the TA and LCA remained the same when large differences of pressure were delivered to the larynx (Fig. 3).

The F_0-P_s data pairs for 3 animals (Numbers 2, 4, and 7) are plotted in Figure 4. The solid lines represent the linear regressions for each data set. For these 3 animals and the other 4 animals (see Table 1), the data points fell close to their respective regression lines. This result is reflected in the R^2 values, which were close to 1 for all of the experiments (Table 1). Thus, F_0 was demonstrated to be linearly related to P_s.

The change in F_0 with P_s (the slope of the regression lines) ranged from 22.4 Hz/kPa (or 2.3 Hz/cm H_2O) to 118.7 Hz/kPa (or 12.1 Hz/cm H_2O). The extreme values for slope are illustrated in Fig 4 (dogs 2 and 7), along with one having an intermediate value (dog 4). P_s across all experiments ranged from 0.32 kPa to 4.08 kPa. The data from the experiments with the smallest and the largest ranges of P_s are include in Figure 4 (dogs 2 and 4 respectively). F_0 for all elicited phonations ranged from 81 Hz to 255 Hz, and were perceived to be in modal register.

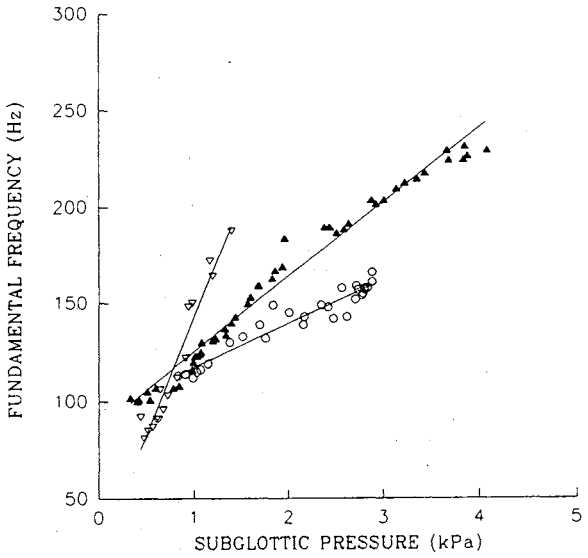


Figure 4. Voice fundamental frequency is plotted against subglottic pressure for dogs 2 (open triangles), 4 (closed triangles), and 7 (open circles). The solid lines are regression lines for each data set.

Table 1.
Summary of the Relation Between P_s and F_o

Dog No.	Range of P_s (kPa)	Range of F_o (Hz)	Pairs of data	Slope of reg. line (Hz/kPa)	R^2
1	0.5 - 3.7	143 - 239	41	31.9	.985
2	0.5 - 1.4	81 - 188	16	118.7	.972
3	0.6 - 2.6	122 - 255	50	66.0	.976
4	0.3 - 4.1	102 - 230	48	38.8	.989
5	0.9 - 3.7	105 - 208	25	34.4	.945
6	0.4 - 3.0	117 - 196	24	31.4	.975
7	0.8 - 2.9	113 - 166	30	22.4	.960

Discussion

It is well documented that vocal fold vibration is a mechanical phenomenon[14]. The mechanical parameters include subglottic aerodynamic power and the mechanical characteristics of the vocal folds' cover and body. The purpose of the present experiment was to examine the mechanical characteristics of the vocal fold cover during phonation with normal active muscle tensions. A modified in vivo model was used in which the midbrain of anesthetized dogs was stimulated electrically to evoke phonation[10]. Consistent phonations with similar patterns of muscular activity (increased activity of the TA and LCA muscles and decreased activity of the CT and PCA muscles) were elicited over many repetitions. By using repeated evoked phonation, minor differences in the patterns of intrinsic laryngeal muscle activity could be disregarded and we could examine various aspects of the mechanical characteristics of phonation.

From a model advanced by Titze[9], it can be inferred that F_0 increases linearly with P_s. The mechanism for this relationship is that an increase in P_s will increase the lateral excursion of vocal fold displacement (the amplitude of the vibration). Because the amplitude to length ratio is not small, vocal fold tension will increase dynamically, resulting in a rise in F_0 with P_s. The present results confirmed this inference. The data illustrated a clear linear relationship between F_0 and P_s. Thus, the mechanical nature of vocal fold vibration was supported.

Similar results have been reported by Moore and Berke[8]. They used an in vivo canine model with recurrent laryngeal nerve stimulation to elicit vocal fold adduction, and also found that F_0 increased linearly with P_s. Interestingly, the range of P_s for which this linear relationship was found ranged from approximately .3 to 4 kPa (3 to 40 cm H_2O) with midbrain stimulation in the present study and 4 to 12 kPa (40 to 120 cm H_2O) with recurrent laryngeal nerve stimulation by Moore and Berke. Therefore, it appears that this linear relationship can be generalized to a wide range of pressures.

The degree to which F_0 changed with P_s across animals was quite variable, ranging from 22.4 to 118.7 Hz/kPa. This variability has been attributed to different pre-phonatory vocal fold length conditions[9]. It was not possible to directly evaluate Titze's hypothesis about the effect of pre-phonatory vocal fold length in the present experiment because the direct view of the vocal folds was not adequate. This is an obvious area in which the experiment may be improved.

References

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

2.Fujimura O. Body-cover theory of the vocal fold and its phonetic implications. In: Stevens K, Hirano M, eds. Vocal fold physiology. Tokyo: University of Tokyo Press, 1981:271-288.

3.Saito S, Fukuda H, Kitahara S, et al. Pellet tracking in the vocal fold while phonating -- Experimental study using canine larynges with muscle activity. In: Titze IR, Scherer RC, eds. Vocal fold physiology. Denver: The Denver Center for the Performing Arts Inc, 1983:169-182.

4.Fukuda H, Saito S, Kitahara S, et al. Vocal fold vibration in excised larynges view with an X-ray stroboscope and an

ultra-high-speed camera. In: Bless DM, Abbs JH, eds. Vocal fold physiology. San Diego: College-Hill Press, 1983:238-252.

5.van den Berg J. Sub-glottal pressure and vibrations of the vocal folds. Folia Phoniatr 1957;9:65-71.

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

7.Rothenberg M, Mahshie J. Induced transglottal pressure variations during voicing. J Phon 1986;14:365-371.

8.Moore DM, Berke GS. The effect of laryngeal nerve stimulation on phonation: A glottographic study using an in vivo canine model. J Acoust Soc Am 1988;83:705-715.

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

10.Solomon NP, Grayhack JP, Liu K, Luschei ES. Patterns of laryngeal and respiratory activation during elicited phonation [Abstract]. ASHA 1992;34:175.

11.Testerman RL. Modulation of laryngeal activity by pulmonary changes during vocalization in cats. Exp Neurol 1970;29:281-297.

12.Garrett JD, Luschei ES. Subglottic pressure modulation during evoked phonation in the anesthetized cat. In: Baer T, Sasaki C, Harris KS, eds. Laryngeal function in phonation and respiration. Boston: College-Hill Press, 1987:139-153.

13.Davis PJ, Zhang SP, Bandler R. Pulmonary and upper airway afferent influences on the motor pattern of vocalization evoked by excitation of the midbrain periacqueductal gray of the cat. Brain Research 1993;607:61-80.

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

NCVS Status and Progress Report - 5
November, 1993, 49-52

Miniature Head Mount Microphone for Acoustic Analysis

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

Abstract

Acoustic analysis of voice sometimes requires a constant mouth to microphone distance. In this study a miniature head mount condenser microphone was compared to a larger, professional grade condenser microphone typically mounted on a stand. Long term and short term amplitude and frequency perturbation measures of human phonation were made for comparison. The results indicate that for this type of analysis only small differences exist between the two microphones. This suggests that errors associated with variable source to microphone distance can be reduced without losing baseline quality in transducing voice signals for analysis.

Introduction

In a previous study of the effects of microphone type and placement on voice perturbation measures (Titze and Winholtz, in press), it was found a mouth to microphone distance of a few centimeters was desirable for high precision perturbation analysis. At close distances, however, a small change in mouth orientation can cause a significant change in amplitude and phase of the microphone output. If small children or patients with head tremor (or other uncontrollable movements) are subjects for recording, meaningful data collection and analysis may be difficult.

To minimize error due to motion artifact, the obvious solution is to use a head mounted microphone that maintains a constant distance of a few centimeters from the mouth. However, the aerodynamic artifacts for near-to-mouth diameters have not been documented well for many vocal tasks. The solution proposed here may therefore not be applicable for all vocal tasks.

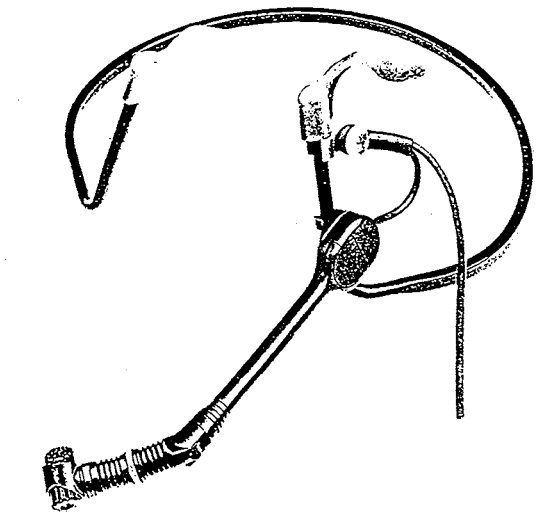
This study will focus on comparison between a miniature head mounted microphone and a larger high quality microphone mounted on a stand. Amplitude and frequency perturbation measures on human sustained vowel phonation will be used in the comparison.

Description

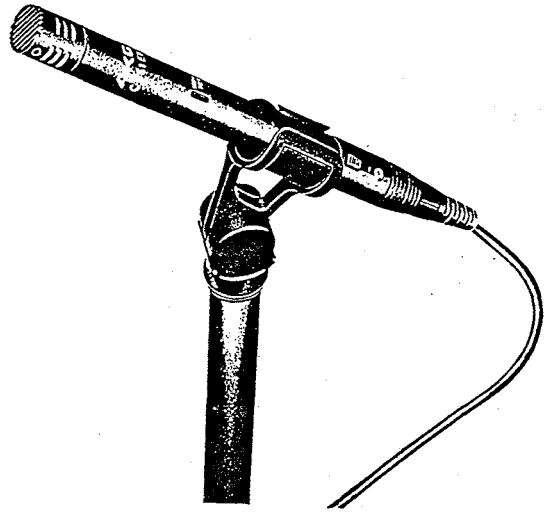
The two microphones chosen for this study were (1) an AKG C410 (Fig. 1a - next page) miniature cardioid condenser head mount and (2) an AKG C451EB (CK-22 capsule) (Fig. 1b - next page) standard mount omnidirectional condenser. Figure 2 (next page) shows the manufacturers published frequency response curves for the two microphones (Fig. 1a for the 410 and Fig. 1b for the 451). It is apparent from the curves that the low frequency response of the 410 begins to rolloff near 150Hz, whereas no roll-off is seem for the 451. Sensitivity was measured by placing the microphones 4cm in front of a loud speaker producing a 200Hz sinewave at 80dB sound pressure level (SPL) and subtracting a preamplifier gain of 60dB. The 410 had a sensitivity of -58.6dB and the 451 had a sensitivity of -49.6dB, indicating a 9dB greater sensitivity for the 451.

Method

Twenty normal subjects, 10 male and 10 female, were asked to sustain the vowel [ɑ] at two pitches, low (100Hz for males, 200Hz for females) and medium (200Hz for males, 400Hz for females) at a comfortable loudness level. The subjects were seated in a chair with a headrest and instructed to keep their head position constant. To test the low end of the fundamental frequency range, two of the male subjects were also asked to phonate at 75Hz, as well



(a)



(b)

Figure 1. Diagram of the microphones (a) AKG C410 miniature head mount microphone, (b) AKG C451 EB standard mount microphone.

as to produce vibrato at the low and medium pitches. In addition, one male subject was asked to sustain the vowel while moving his head in one of two manners: drift, a slow (.5Hz) movement from the axis of the mouth to approximately 30° off axis; and wobble, a slightly faster (1Hz) periodic movement from 0° to 30° on either side of the mouth axis. This task was performed for low and medium F_o. Drift was to approximate what a normal subject might do (particularly a child), and wobble was to approximate what a subject with a head tremor might do.

The experiment was conducted in an IAC isolation booth, 3.2m deep by 3.5m wide by 2.4m high. Ambient SPL of the booth, measured with a B&K 2230 SPL meter on the

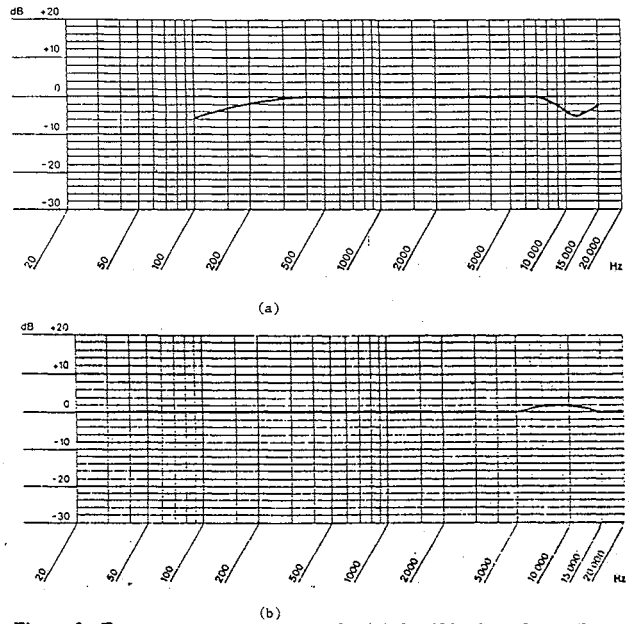


Figure 2. Frequency response curves for (a) the 410 microphone, (b) the 451 microphone.

linear weighing scale (20 Hz to 20 KHz), was 53dB near the center of the booth. The microphones were positioned at a 4cm distance, one at +90° to the mouth axis and one at -90° to the mouth axis. The two microphone signals were preamplified (ATI M-1000) and recorded simultaneously with a Panasonic SV-3700 DAT recorder. Later the signals were played out of the DAT recorder, high pass filtered at 60 Hz (24dB/oct), amplified by Tektronix AM502 amplifiers and sampled by a DSC-200 16 bit A/D converter at 20K samples/second. A two second segment near the middle of each subjects phonation was digitized.

A least squares waveform matching algorithm (Milenkovic, 1987; Titze and Liang, in press) was used to analyze the two second segments of the digitized signals. This algorithm provides the highest precision in F_o extraction for small perturbations (Titze and Liang, in press). The perturbation measures included CV (coefficient of variation, applicable to slow deviations from the mean amplitude and F_o) and PF1 (the mean rectified value of the first order perturbation function, applicable to more rapid cycle-to-cycle variations). PF1 is commonly called jitter when applied to an F_o contour and shimmer when applied to an amplitude contour. A correlation of these measures for the different vocal tasks was used to compare the microphones.

Results

Since the recordings with the two microphones were simultaneous, individual tokens could be compared with a scatter plot. This is shown in Figure 3 for the amplitude perturbation measures and Figure 4 for the frequency perturbation measures. Tokens for the twenty normal subjects are represented. The X axis represents

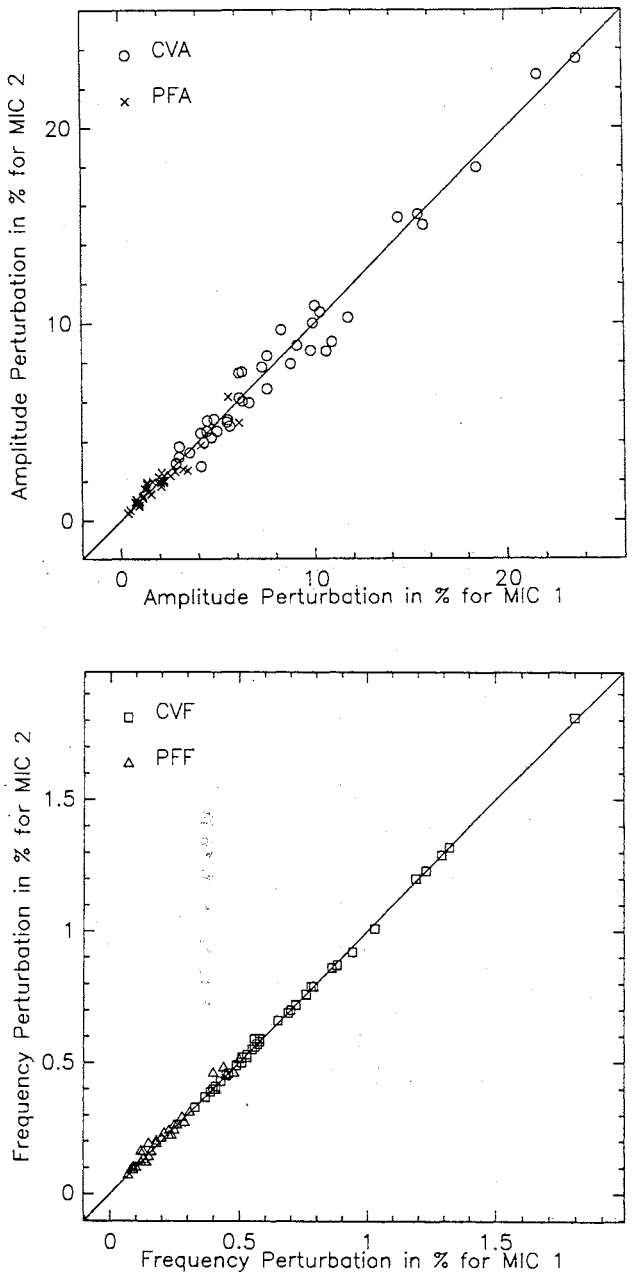


Figure 3 (top). Scatter plot of microphone 1 (the 410) versus microphone 2 (the 451) for amplitude perturbation analysis (CVA and PFA) using ten male subjects and ten female subjects at two F_os. Figure 4 (bottom). Scatter plot similar to Figure 3 for frequency perturbation measures.

measures for microphone 1 and the Y axis for microphone 2.

The plots indicate that, for the type of amplitude and frequency perturbation analysis used here, the microphones gave similar results over two octaves of F_o. Although the points do not all fall on the diagonal, an inconsistency is not apparent between the microphones. One would expect to see a trend of one microphone producing consistently lower or higher numbers than the other, particularly in the amplitude measures, if a major discrepancy existed between the microphones.

Table 1.
Comparison of perturbation analysis for simultaneous recording of two microphones for subjects producing very low F_o, vibrato, and head motion.

SUBJECT	F_o	MIC	CVA	PFA	CVF	PFF
			LOW F_o			
M6	75Hz	1	8.73	4.75	0.96	0.55
		2	8.78	4.74	0.99	0.55
M11	75Hz	1	7.02	3.64	1.35	0.41
		2	6.55	3.22	1.35	0.41
			VIBRATO			
M2	100Hz	1	9.64	2.00	2.36	0.63
		2	10.56	2.48	2.34	0.64
	200Hz	1	3.92	1.10	1.36	0.21
		2	2.57	0.93	1.36	0.22
M6	100Hz	1	16.40	1.42	0.84	0.22
		2	14.77	1.22	0.84	0.23
	200Hz	1	2.79	0.70	1.09	0.15
		2	3.19	1.40	1.09	0.15
			HEAD MOTION - DRIFT			
M12	100Hz	1	3.77	2.27	0.35	0.20
		2	9.48	1.34	0.36	0.21
	200Hz	1	5.20	1.53	0.54	0.18
		2	6.30	1.42	0.54	0.14
			HEAD MOTION - WOBBLE			
M12	100Hz	1	6.10	2.87	0.49	0.27
		2	20.74	2.03	0.49	0.26
	200Hz	1	9.44	1.86	0.55	0.56
		2	16.00	1.11	0.49	0.47

Table 1 shows results aimed at testing low F_o, low frequency modulation in the voice, and a moving source. With the higher rolloff of the 410 frequency response, perturbation measures from phonation with F_os below this rolloff would be expected to produce higher measures for the 410, especially in the case of low F_o with vibrato. The data did not indicate a compelling difference. However, there were several interesting observations with the head motion task. As expected, the long term amplitude measures (CVA) were affected the most. For subject MI2, CVA measures for wobble were two to three times higher with the stationary microphone than with the constant distance microphone. There were no strong effects for fundamental frequency measures, however.

Table 2 (following page) presents the correlation data for the various vocal tasks. Over two octaves of F_o the correlations were all greater than 0.85 when no motion artifacts were present. Most of the correlations were above 0.95. No statistical difference was found between the microphones at the $p < .05$ level. For the head motion tasks, however, the correlations were all below .85 ($p < .05$), signifying a difference between the recordings.

Table 2.
Correlation of the two microphones perturbation analysis data for normal subjects, very low F_o, vibrato, and head motion.

GROUP	F_o	N	CVA	PFA	CVF	PFF
Normal male	100Hz	10	.953	.995	.999	.997
	200Hz	10	.951	.961	1.000	.984
Normal female	200Hz	10	.987	.851	1.000	.983
	400Hz	10	.985	.870	1.000	.997
Very Low F_o/ Vibrato	ALL	6	.979	.963	1.000	1.000
Motion	ALL	4	.518	.769	.859	-.309

Conclusion

The higher rolloff of low frequency response for the head mounted microphone was expected to produce a degradation of perturbation measures at low F_o (below the flat response of the microphone), especially in the presence of vibrato. Frequency modulation, with its concomitant amplitude modulation, was expected to affect the long-term amplitude measures. However, the coefficient of variation of amplitude (CVA) measures did not indicate a significant difference. A small variability in short term measures may be partially explained by individual differences in microphone phase distortion of the acoustic signal and by differences in microphone sensitivity (Titze and Winholtz, in press).

Simple tests with head motion indicated that movement artifact can inflate long term amplitude measures by two to three times. Since the microphones gave similar perturbation measures when no movement was present, the head mounted microphone showed a clear advantage by maintaining a constant source to microphone distance.

Some additional comments are in order. Because the head mounted microphone moves with the head, the microphone cable must have strain relief to eliminate motion noise that can be conducted through the cable. Also, the electrical output of the microphone is unbalanced from the microphone to the XLR connector. This is unfortunate because it is susceptible to high levels of electromagnetic interference, such as that radiating from a stroboscopic light source. Some further experimentation may be needed to determine if the microphone can be used in various different clinical environments.

It is important to note that this study used only sustained vowel phonation tasks. The effects of aerodynamic artifacts at close source to microphone distances also awaits further study. The head mounted microphone may not be the ideal choice for all vocal tasks. However, based upon the correlation data for a limited set of tasks performed here, it appears that the miniature head mounted microphone can become a standard for voice perturbation analysis.

Acknowledgment

This study was supported by a grant from the National Institutes of Health, grant No. R01 DC00387-04.

References

Milenkovic, P. (1987). Least mean square measures of voice perturbation. Journal of Speech and Hearing Research, 30, 529-538.

Titze, I.R., & Liang, H. (in press). Comparison of F_o extraction methods for high precision voice perturbation measurements. Journal of Speech and Hearing Research.

Titze, I.R., & Winholtz, W.S. (in press). The effect of microphone type and placement on voice perturbation measurements. Journal of Speech and Hearing Research.

NCVS Status and Progress Report - 5
November, 1993, 53-59

Comparison Between Clinician-Assisted and Fully Automated Procedures for Obtaining A Voice Range Profile

Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
and
The Recording and Research Center, The Denver Center for the Performing Arts
Darrell Wong, Ph.D.
The Recording and Research Center, The Denver Center for the Performing Arts
Martin A. Milder, B.S.
Department of Speech Pathology and Audiology, The University of Iowa
Susan R. Hensley, M.A.
The Recording and Research Center, The Denver Center for the Performing Arts
Lorraine O. Ramig, Ph.D.
Department of Communication Disorders and Speech Science, University of Colorado-Boulder
and
The Recording and Research Center, The Denver Center for the Performing Arts
Neil B. Pinto, M.E.
The Recording and Research Center, The Denver Center for the Performing Arts

Abstract

A comparison was made between two methods of obtaining a Voice Range Profile. One method was traditional, involving a clinician who gave instructions, motivated the subject to achieve the greatest intensity range, and determined when the goal was achieved. The second method was completely automated, involving the use of a video tape for instruction and a computer for elicitation and evaluation. Results indicated that there is no obvious preference for the use of either method, but some differences are noted.

Introduction

The Voice Range Profile (VRP), also called the phonetogram (Damsté, 1970), is used as a clinical tool to establish a vocalist's range of intensity from the lowest to the highest fundamental frequencies (Coleman, Mabis, & Hinson, 1977; Schutte, & Seidner, 1983; Gramming, 1988; Klingholz, 1990). Typically, the VRP is obtained by prompting the vocalist with a series of pitches from a keyboard instrument, or a pitch pipe, and requesting the softest and loudest productions possible (Coleman, 1993). The clinician determines whether the effort was at the right pitch. Motivating the vocalist to give the best performance (their absolute softest and absolute loudest) takes considerable time and effort from the vocologist (clinician or teacher). It is not unusual to spend a half-hour to get a satisfactory VRP.

Given this investment in time, and the possibility of varied results due to human interaction, it seems logical to ask if the task could be economized and standardized by full automation. Each vocalist would receive exactly the same instructions and coaching (by video tape), and the frequency-intensity information would be gathered and plotted by computer (Pabon & Plomp, 1988). The present investigation deals with the trade-offs between the loss of information and the gain in economy when the VRP procedure is fully automated.

Methods

The study was conducted in an IAC recording booth (11'5" x 10'5" x 8') at the Recording and Research Center of the Denver Center for the Performing Arts. Two separate protocols were established for obtaining a VRP, and all subjects followed both protocols.

Subjects

A total of 20 subjects were involved in the study. These consisted of 10 males ranging in age from 30 to 42 years, with a mean age of 37.4 years, and 10 females ranging in age from 24 to 48 years, with a mean age of 32.9 years. As determined by a questionnaire, the subjects were free of any history of laryngeal pathology and each subject reported that he/she was in good vocal health at the time of the study. A laryngeal examination was not conducted for economic reasons. None of the subjects had had formal voice training and all were unfamiliar with the experiment. The subjects were divided into two groups. Five females and five males did Protocol 1 followed by Protocol 2, while the other groups of five females and five males did the protocols in reverse order.

Protocol 1

A certified speech language pathologist (SLP-CCC) explained the procedure for obtaining a VRP to each subject. After the procedure was clarified, each subject was asked to produce a sustained /a/ vowel at a comfortable pitch and intensity. The purpose of this was to determine a starting pitch which would presumably fall within a comfortable range for each subject. (Invariably, the pitch volunteered by each subject was within a few whole-tones of his/her lowest sustainable frequency in modal register.) The clinician then used a small keyboard instrument to prompt the subject with successive target pitches.

In order to facilitate comparison of the clinician-assisted VRP with the automated program, the chosen target pitches corresponded roughly to the center of each "frequency bin" contained in the computerized program. The target pitches determined in this manner consisted of a whole-tone scale spanning the subject's entire range and containing the pitches c, d, e, f#, g# and a# in each successive octave.

For the most part, a fairly regular pattern of pitch presentation was followed from subject to subject. That is, each subject was asked to produce the target pitches in descending order from his or her starting pitch down to his or her lowest sustainable tone, then in ascending order from the starting pitch to the highest tone he or she could produce. The best of three efforts at each pitch and intensity level throughout the subject's range was recorded. For some subjects, a shift into a higher vocal register was facilitated by allowing them at some point to skip from modal register

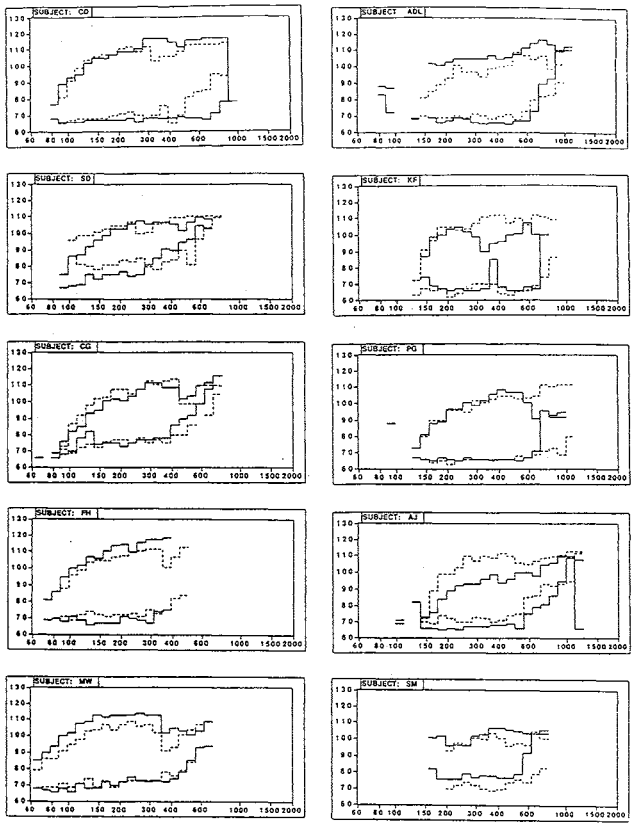


Fig. 1(a) Fig. 1(b)

Figure 1. Comparison of automated (solid lines) and clinician-assisted (dashed lines) Voice Range Profiles. (a) 5 normal male subjects and (b) 5 normal female subjects. The automated procedure was administered first in both cases.

up to a tone at or near the top of the vocal frequency range and working down from there. (This was especially helpful in eliciting falsetto productions from the male subjects.) In some cases, after reviewing the overall profile, the clinician had the subject repeat his or her efforts in a particular pitch region if she had reason to feel the best performance for those tones was not captured on the first trial.

The subjects were encouraged to produce their least and greatest SPL efforts at each target pitch without regard to musical quality of the tone, but, at the same time, without causing discomfort. When appropriate, modeling was provided by the clinician. Acceptable productions were those judged to be "stable" and at least 1-2s in duration. The sound level of the phonations were measured with a calibrated Bruel and Kjaer Type 2230 sound level meter at a mouth-to-microphone distance of approximately 8 cm. The meter settings were RMS, fast, and linear frequency weighting (20-20 kHz). The clinician determined by ear whether a particular pitch was adequate or should be revisited.

Protocol 2

Standardized instructions were given using a 5 minute video tape (narrative shown in the Appendix). The tape gave a brief explanation of the purpose of the VRP and how it would be elicited. The subject was encouraged to do as well as possible without straining the voice and was also informed that phonation in any tone and loudness order was acceptable. The tape featured a person working interactively with the computer to obtain the VRP. Following the video tape, a head mounted microphone (AKG C410), located 8 cm from the mouth at an angle of 45 degrees from the center of the lips, was placed on the subject. The subject was reminded to use only the /a/ vowel, to refrain from adjusting the microphone, and to give the best effort without straining the voice. Any questions regarding the concept or the procedure were answered at this time, as long as they did not pertain to performance expectations. The voice signal was acquired directly by computer and Fo and intensity were extracted and displayed in real-time. The voice signal was also recorded on DAT for archival purposes.

The Fo and intensity extraction algorithms were implemented on a MacII FX computer using LabView, a graphical block-diagram oriented digital signal processing system. The system consisted of display and programming software as well as hardware digitizing and processing boards (TMS320C30 based).

The microphone signal was digitized at 22 kHz and segments of 512 points were analyzed with the DSP board. For each segment, the cepstrum was calculated (Noll, 1967) and the fundamental frequency was extracted by finding the largest peak in the cepstral signal subsequent to the initial onset transient (Hess, 1983).

While the pitch was extracted, the intensity was obtained from the variance of the microphone signal (average squared difference from the mean) to eliminate low frequency drift. For display, the intensity was converted to decibels. Once both Fo and intensity were obtained, they were compared against the noise threshold and not used if below the noise threshold (noise was below 65 dB in the recording booth). The intensity-frequency data pair was then displayed as a black "X" on the screen and added to a list of data pairs kept for the last 1.5 seconds. The list was then examined to see if it met a steadiness criterion. The criterion was defined as: a minimum of two on the list, less than 4 dB SPL variation, less than 1.5 seconds time duration elapsed, and a maximum of 0.09 log units in frequency variation ($\log(F_{max})-\log(F_{min})<0.09$). This criterion was a compromise between stability and subject frustration. Once the criterion was met, the average intensity and Fo for the list was calculated. If this point was above the current maximum or below the current minimum for a frequency bin (indicated by a blue and a red line, respectively), a replacement line was drawn.

The automated program allowed the subject to select any pitch at any time, and subjects could return and improve their effort any time during the 30 minute period.

Results

Figure 1 (previous page) shows overlays of VRPs obtained with the two procedures on ten normal subjects. The automated procedure (solid lines) was administered first and the clinician-assisted procedure (dashed lines) was administered second. Males are on the left panel (Figure 1a) and females are on the right (Figure 1b). Each subject is identified by letters in the upper left corner of each plot. Fundamental frequency is plotted in equal logarithmic intervals and intensity is plotted in dB.

Figure 2 is similar to Figure 1 except that the two procedures are reversed and different subjects were used. Thus, qualitatively the results are the same, but individual differences are evident.

Tables 1 and 2 (next page) contain data extracted from the figures. The area of the VRP envelope is measured

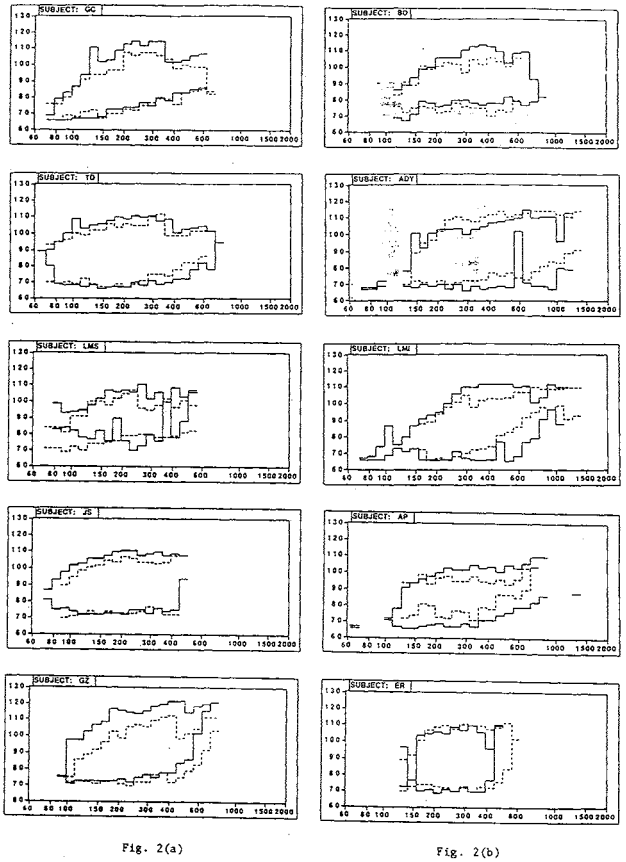


Figure 2. Comparison of automated (solid lines) and clinician-assisted (dashed lines) Voice Range Profiles. (a) 5 normal male subjects and (b) 5 normal female subjects. The automated procedure was administered second in both cases.

Table 1.
Comparison of computer versus clinician elicited VRPs. Numbers given in parentheses discount bins recorded due to subharmonics.

SUBJECT	AREA (dB x BIN)		NUMBER OF BINS	
Clinician	Automated	Clinician	Automated	
ADL	577	423	20(18)*	17
ADY	659	649	24(21)	21
AJ	434	478	21(20)	19
AP	523	260	21(20)	17
BD	458	398	18	16
CD	836	654	22	20
CG	368	442	22(21)	18
ER	333	424	12	14
FH	558	472	15	15
GC	520	417	20	19
GZ	580	447	20	18
JS	505	426	17	15
KF	424	636	15	18
LM	347	361	17	18
LMI	613	413	24(20)	19
MW	671	502	22	20
PG	458	607	19(18)	18
SD	335	347	18	18
SM	282	326	14	13
TD	684	545	22	19
MEAN	508.3	461.4	19.2(18.5)	17.6
SD	142.6	109.6	3.3(2.8)	2.1
MIN	282	260	12	13
MAX	836	654	24 (22)	20

as a cumulative sum of dB range per bin. Each bin is a unit wide, and is centered about a tone on the musical scale. For those bins in which only one loudness level was achieved, the range for that bin was set to 1 dB. The areas ranged from 260 to 654 dB bins for the clinician-obtained scores and 282 to 836 dB bins for the computer obtained scores. The frequency ranged from 12 to 24 bins.

The maximum loudness achieved was 122 dB using the automated procedure and 114 dB using the clinician assisted procedure. The minima were 65 and 62 dB, respectively. The maximum range for each VRP at a single tone varied from 30 to 52 dB for the automated protocol and 27 to 48 dB for the clinician-assisted protocol. Since these ranges are relative, they are more useful than the absolute maximum and minimum loudness. It should be noted, however, that the tone or bin at which the maximum range occurred varied from subject to subject and also from one protocol to the other.

A three factor repeated measures ANOVA was performed using gender, protocol order and protocol type as factors. The area, measured in dB bins, was used as a score. The analysis indicated that gender was not highly significant ((16,1), f=0.37 with significance 0.552). Protocol order was also not highly significant ((16,1), f=0.06 with significance 0.803).

Protocol type was significant at the 10% level, but not at 5% ((16,1), f=3.13 with significance 0.096). The cross effects of these factors were all not significant.

A scattergram of clinician-obtained vs computer-obtained VRP area scores is shown in Fig. 3. The Pearson bivariate, two tailed correlation between these two methods yields a correlation coefficient of 0.5286 and significance of p = 0.017, indicating moderate correlation between the protocols.

A subjective, visual comparison of the computer versus clinician obtained VRPs presented in Figures 1 and 2 reveals similarities between the profiles obtained by the two protocols for most of the subjects. Since the two protocols were recorded on different days, it is possible that the absolute positioning of the VRP varied according to the distance between the lips and the recording transducer. It

Table 2.
Maximum and minimum dB level attained and frequency at which maximum dB range was attained.

SUBJECT	MAX dB		MIN dB		MAX RANGE @ TONE	
	Auto.	Clin.	Auto.	Clin.	Auto.	Clin.
ADL	117	109	65	68	45, 589	41, 525
ADY	115	114	66	69	44, 661	40, 263
AJ	110	113	65	69	42, 1177	41, 417
AP	110	108	66	71	38, 331	27, 331
BD	114	111	67	70	37, 372	36, 417
CD	117	114	66	65	50, 661	44, 295
CG	116	114	66	69	35, 295	38, 295
ER	111	112	69	70	40, 331	40, 331
FH	119	113	66	69	52, 295	40, 263
GC	115	108	66	66	44, 132	35, 209
GZ	122	113	71	70	44, 186	41, 417
JS	111	108	72	70	39, 209	36, 417
KF	107	112	66	62	39, 589	48, 417
LM	111	107	70	69	38, 263	28, 235
LMI	112	110	65	66	47, 525	35, 295
MW	114	109	66	68	45, 148	36, 235
PG	109	112	65	63	43, 417	43, 935
SD	110	111	67	78	33, 263	30, 525
SM	107	105	76	69	30, 417	35, 372
TD	111	112	66	67	43, 186	39, 166
MEAN	112.9	110.8	67.3	68.4		
SD	4.0	2.6	2.9	3.3		
MIN			65	62		
MAX	122	114				

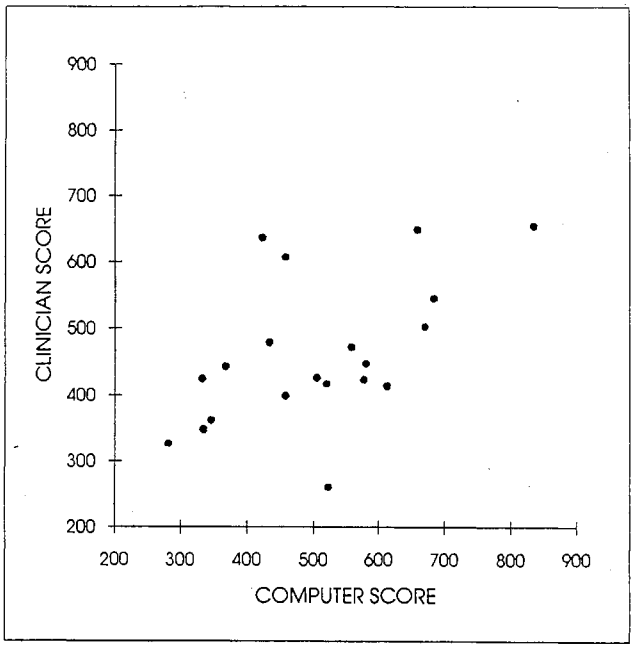


Figure 3. A scattergram of VRP area scores obtained for each subject.

is also possible that there were significant day to day differences in the performance of the subjects, as noted by Coleman (1993). These differences can appear on the VRP figure as vertically shifted boundaries. The area calculation for each VRP is independent of vertical positioning, however. If one discounts this vertical shift, the similarities between the two profiles become more apparent. This is particularly evident in the profiles of subjects CG, FH, and MW in Figure 1 and TD, JS, and ER in Figure 2. Of particular interest is the unusual "square" profile of subject ER, which is faithfully duplicated in the automated session.

Frequency (bin) range between the two VRPs typically differed by one or two bins at either end. Some notable outliers in the lower frequency bins are evident in the profiles of CG, PG, and AJ in Figure 1 and ADY, LMI, AP, and ADL in Figure 2. These do not reflect actual pitch range differences but appear to be related to the detection of subharmonics, which are discussed more fully below.

Comparison of the profiles suggests that for some subjects, clinician involvement appeared to be of great benefit. For example, the computer-elicited profile for subject PG in Figure 1 reveals a markedly reduced performance in the higher frequencies as compared to the clinician-assisted profile. A review of the audiotape revealed that, for the most part, there was a reduced overall effort by the subject during the automated session, both in terms of dynamic range and length of phonation. There appeared to be a lack of motivation for producing the higher tones.

The computer-elicited profile of KF reveals a depressed area on the high SPL tracing through this subject's mid-frequency range. Intuitively, one might suspect that this depressed area of performance could be related to difficulty on the part of the subject in negotiating a shift from a heavy, chest-type register to a lighter register. A review of the audiotape from this subject's computer-automated session suggested, however that KF did indeed have some difficulty producing high SPLs without the assistance of a clinician in this transitional region. One also might suspect that the "gap" on the low SPL tracing in this region would be typical of reduced dynamic control in the region of register shifts; in this particular instance audiotape review suggests that this gap was likely due to difficulty with phonation stability, i.e., the computer rejected those attempts as unstable.

In addition to occasional problems with phonation stability (which seems to be more prevalent with higher frequency productions in general, and with lower SPLs in the modal register), gaps in the contour of the automated VRP also occur when a subject, for one reason or another, does not revisit a certain tone in an effort to better his or her performance. This was the case with subject LMS and accounts for the lower trace peaks at about 185 Hz and 520 Hz. The gap at 400 Hz was due to pitch stability problems.

A number of the subjects demonstrated improved overall performance in terms of SPL range in the automated sessions and several commented on the benefit of visual feedback on their performance, particularly with regard to self-motivation. Subject AP is a subject whose SPL range in the automated session showed marked improvement over the clinician-assisted session. It is interesting to note that, although this subject had significant difficulties with pitch matching in the clinician-assisted session, the automated VRP reveals a fairly smooth profile. Undoubtedly AP benefitted considerably from visual feedback in the automated session, and the lack of a requirement to match particular pitches.

One source of variation between the automated VRPs and the clinician-elicited VRPs can be attributed to the fact that occasionally the computer extracts and records a "subharmonic" rather than the perceived fundamental frequency. One might argue that this is an accurate acoustic representation of what is actually being produced and therefore a valid measure in the VRP. On the other hand, it is easy to see how a VRP that demonstrates frequency ranges below 150 Hz in female subjects could be very misleading to a clinician. If these measures reflect subharmonic activity one octave below the actual pitch target or perceived fundamental frequency for that phonation, they should at least be labeled as a different vocal quality.

Many times these subharmonics appear as obvious outliers on the low frequency end of the VRP as can be seen in the profiles for CG, PG, and AJ in Figure 1 and ADY and AP in Figure 2. These readings reflect the computer response to subharmonics produced during the production of tones in the modal register. In some cases the presence of subharmonics in the lower bins is less easily identified. Subject LMI in Figure 2 appears to have a continuous profile extending well below what would be expected from a female subject. Review of the audiotape revealed that although her lower range was quite impressive, the recording of bins below 100 Hz were actually subharmonics produced during production of tones one octave above. This is the case with the lower bins recorded in subject ADL as well.

The recording of subharmonic activity can present more subtle variations elsewhere in the profile, since production of very high frequencies occasionally results in subharmonic recordings that might impact on the shape of the mid-portion of the VRP. It is felt that this is not a major problem since the energy produced by these subharmonics tends to fall within the profile envelope and thus does not replace SPL values recorded at the actual target pitch. The fairly close agreement between overall VRP shapes for these subjects would seem to confirm this.

Conclusions

Results of this study suggest that, for normal subjects, a full Voice Range Profile can be obtained with a

fully-automated protocol. This may save a clinician considerable time and effort. By using fixed, objective criteria for determining steady-state phonation, clinician variability in acceptance criteria is eliminated. The computer considers voice stability insofar as it is measurable in pitch variability, intensity variability and timing. Musicality is not a criterion.

The visual feedback offered by the automated procedure can be a helpful aid to both the clinician and the subject, since there are sometimes difficulties in eliciting vocal utterances (because of pitch-matching and pitch-perception problems).

For those clinicians who have trouble identifying pitch, this tool relieves them of this burden. The subjects also benefit since they are not required to match specific pitches, but may choose pitches randomly throughout the frequency range. There are some subjects who might prefer presentation of a pitch target or would benefit from a reminder of a pitch which needs to be revisited. The program could be modified to work with a keyboard so that any tone desired could be presented through an earphone.

Review of the audiotapes from the automated sessions coupled with subjective analysis of the profile displays suggest that for some subjects the automated VRPs may underestimate the high pitch ranges (due to tracking difficulties) and overestimate the lower pitch ranges (due to the recording of subharmonics). These problems could be controlled by including a brief period of clinician intervention at the end of an automated session to confirm the validity of the lower bins in the VRP and to fill in the gaps at the upper end (and elsewhere, if necessary). The introduction of the clinician would make it difficult to maintain consistency of presentation to the subject, however. The alternative is to leave it to the subject to interpret their effort.

The question of how to deal with the recording of subharmonics in the automated profile bears further investigation. Perhaps future modifications to the computer algorithm might include a means of reducing sensitivity to subharmonics. When the cepstral technique is used for pitch tracking, the largest peak in the cepstrum is identified as the fundamental (excluding the transient part at the beginning of the time line). When subharmonics are present, the largest peak in the cepstrum may appear at the period of the subharmonic, indicating that this peak is a better candidate for the fundamental. A possible solution might be to provide a keyboard for the subjects to prompt themselves. The computer could identify the tone and narrow its expected pitch period range so that it rejected fundamental frequencies more than half an octave away. The subject would then be expected to phonate close to the keyboard tone. This would restrict the subject by requiring pitch matching (although the order of the pitches is still determined by the subject). This problem illustrates the need for a better definition of fundamental frequency when the voice strays from being a steady state phonation to displaying diplophonic, rough, or even chaotic characteristics.

The differences in VRP areas between the protocols for each subject suggests that the VRP obtained is influenced by the protocol (but only at the 10% significance level). It would be useful to study whether for a particular protocol, a subject can achieve a consistent VRP over time.

One might wonder how much of the increased VRP area demonstrated in some of the computer-obtained profiles was a result of greater performance on the high SPL side. This does bring up a two-fold concern: (1) Does one run the risk of some over-zealous subjects straining their voices in the situation where clinicians are not present to monitor vocal behavior, and (2) Are the high SPLs underestimated in the clinician-assisted sessions because of clinician bias? The clinician may have had a preconceived notion of what is safe as a maximum SPL level, based on effort perceived from the subject. For this reason, it is possible that a clinician might not push for the loudest productions. The problems of self-inflicted vocal abuse in an automated procedure and over-guarding in a clinician-assisted procedure need further addressing.

Acknowledgement

The work was supported by Grant No. R01 DC 00387-06 from the National Institute on Deafness and Other Communication Disorders.

References

Coleman, R.F., Mabis, J.H., and 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.

Coleman, R.F. (1993). Sources of variation in phonetograms. *Journal of Voice, 7*(1), 1-14.

Damsté, H. (1970). The phonetogram. *Practica-Oto-Rhino-Laryngologica, 32,* 185-187.

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

Hess, W. (1983). *Pitch determination of speech signals: Algorithms and Devices.* Heidelberg, Germany: Springer-Verlag.

Klingholz, F, (1990). *Das Stimmfeld* [Voice area]. Munich: Verlag J. Peperny.

Noll, A.M. (1967). Cepstrum pitch determination. *Journal of the Acoustical Society of America, 41,* 293-309.

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

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

Appendix

Narrative on Video Tape*

Thank you for participating in this experiment at the Recording and Research Center.

On the computer screen behind me is a voice range profile program. The purpose of the program is to measure your loudest and softest efforts from low pitch to high pitch to determine your vocal range.

Please watch this video tape in its entirety before using the program. On the horizontal axis, the program will display your pitch, from lowest to highest. On the vertical axis, the program will display your loudness, from soft to loud. There is a range switch in the lower right-hand corner of the screen. To control the switch, use the mouse to move the curser on the switch. Press the button to change the switch from the "M" position to the "F" position. The "M" position is for pitches less than 500 Hz, or this portion of the screen. The "F" position is for pitches greater than 500 Hz, or this portion of the screen.

When you make a sound an "X" will appear on the screen, indicating the current pitch and loudness levels. When a sound is held for approximately one second, a blue and a red bar will appear on the screen. [Demonstrates] "Ah." The blue bar indicates the loudest effort at that pitch, and the red bar indicates the softest effort at that pitch.

To move the blue bar up, make a louder sound at the same pitch. [Demonstrates] "Ah." To move the red bar down, make a softer sound at the same pitch. [Demonstrates] "Ah."

To increase your voice range profile, repeat the same process at different pitches. If the "X" appears between the bars, they will not move. They will only move when you are louder than the blue bar or softer than the red bar. [Demonstrates] "Ah."

Here's an example of how to begin building your voice profile. [Demonstrates] "Ah." "Ah." "Ah." "Ah." "Ah."

On the screen you see an example of a completed voice range profile.

You have 30 minutes to do this task. Take your time. Remember, you can always go back to a particular pitch and try for a louder or softer sound.

The purpose of the experiment is to establish your loudest and softest sounds at every pitch. During the experiment strive for maximum effort, but avoid discomfort in your voice. When you are through, use the phone and dial the extension provided.

Thanks again for participating in this experiment.

*The tape is available by writing to: The Recording and Research Center, 1245 Champa Street, Denver, CO 80204.

NCVS Status and Progress Report - 5
November, 1993, 61-67

Perceived Pauses and Durational Characteristics of Oral Reading and Impromptu Speech

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

Abstract

Speaking-time ratios (STR), utterance and pause durations of oral reading and impromptu speech were investigated for twenty young adults (ten males and ten females). The reading material was the first paragraph of the Rainbow passage. The impromptu speech was elicited by asking the subject to describe a picture. Results showed that oral reading was associated with greater STRs and mean utterance durations and smaller pause durations than the impromptu speech. No gender differences were found for any of the three measures. The average STR value was .87 for oral reading and .69 for impromptu speech. The STR value of .87 for oral reading was much greater than the value of .70 reported by a normative study employing 120 young adults. A potential cause of such disparity and clinical significance of STR measurements was discussed.

Suprasegmental durational measures of human speech reflect important underlying cognitive and physiologic processes. The suprasegmental durational measures include total speaking time, utterance and pause durations, voiced speech segment durations (also called phonation time) and voiceless speech segment durations. The total speaking time is the sum of the utterance and pause durations. The utterance duration is sometimes called articulation time, and is the sum of voiced and voiceless segment durations. Given the number of syllables or words in the utterances, the rate of speech can be expressed in terms of words/minute or syllables/minute by dividing the number by the speaking time. The rate can be calculated with or without pause durations included in the speaking time.

From the primary duration measures, some ratio measures can be derived. The speaking-time ratio is the ratio of the sum of utterance durations to the total speaking time. The phonation-time ratio is calculated as the ratio of the sum of voiced segment durations to the sum of utterance durations (excluding pauses) or to the total speaking time (including pauses). The articulation-phonation time ratio is the ratio of the sum of utterance durations to the sum of phonation times.

Durational measurements have been useful in describing speech characteristics, for example, of dramatic reading (Fairbanks and Hoaglin, 1941), of superior esophageal speakers (Horii, 1983b), of a dysphonic patient (Watanabe et al., 1987), of stutterers (Horii & Ramig, 1987) and of dysarthric patients (Till & Alp, 1991). Analysis of pause frequency and durations of spontaneous speech, furthermore, has received increasing attention as a means of examining latencies of underlying cognitive processes (see for example, review papers by Goldman-Eisler, 1972; O'Connell & Kowal, 1981; Rochester, 1973).

There is, however, a paucity of normative data. The limited normative data in the literature, furthermore, are often difficult to compare and are nondefinitive regarding possible age and gender differences in durational characteristics of connected speech (Hartman & Danhauer, 1976; Oyer & Deal, 1985; Walker, 1988). These data are difficult to evaluate due to confounding factors of speech tasks (e.g., oral reading, recitations from memory, spontaneous speech, conversational speech), differences of the

reading texts, the amount of speech samples, and variations in definitions of specific measures. Other factors influence durational measures in connected speech as well. These factors include speaking conditions (loud versus soft), emotional states, type of audience (children versus adults, for example) and subject matters.

Speech tasks (oral reading versus impromptu speech, for example) affect not only the magnitude but also the interpretations of durational measurements. In oral reading, breath grouping is normally determined by the linguistic structure of utterances. Readers pause at linguistically appropriate boundaries (e.g., phrase and sentence boundaries) primarily to replenish the air reservoir in the lungs. In impromptu speech, on the other hand, the speakers focus not on the oral delivery but on the on-going formulation of ideas and transformation into linguistically appropriate strings of utterances. Obviously, the subject matter of the impromptu speech profoundly affects utterance and pause durations.

The paucity and difficulty of comparisons of the limited normative data are also attributable to instrumental difficulties and differences. Durational measures of connected speech inherently require relatively long speech samples. Traditional analog methods such as oscillographic and spectrographic analyses and the use of graphic level recorders are not well suited for the analysis of such large samples and are time-consuming, laborious and costly. Recent advancement of technology including digital methods and software, however, promises an increasing amount of durational studies of connected speech (Horii, 1983a; O'Connell & Kowal, 1981; Ruder & Jensen, 1970; Till & Alp, 1991; Walker, 1988; Watanabe et al., 1987).

Prior to applications of durational analyses to speech produced by individuals with various disorders, further accumulation of normative data is warranted. The purpose of the present investigation was to examine STR of oral reading of a standard text and impromptu speech by young adult males and females. The oral reading task was chosen because it provided tighter control of speaking conditions, linguistic content and structure, and size of speech samples. The impromptu speech condition was included in the study to obtain preliminary STR values for such speaking condition and to compare with the oral reading tasks. The oral reading task was of primary interest for the comparative purposes with equivocal STR values reported in the literature. STR values ranging from .53 to .91 for oral reading tasks can be found in the literature (Fairbanks & Hoaglin, 1941; Horii, 1983c; Walker, 1988).

Method

Subjects

A total of twenty young adults, ten females and ten males, served as subjects. They ranged in age from 18 to 27 years with the mean age of 23.9 years. The mean ages of the male and female groups were identical (23.9 years). They were all native American English speakers and all spoke the same dialect of American English usually characterized as General American. They were free from speech-, language-, reading- and hearing-problems. All subjects were considered to be untrained speakers.

Speech Materials and Recording Procedures

For the purpose of comparison with the Walker study (1988), the first paragraph of the Rainbow passage (Fairbanks, 1960) was selected as the oral reading text. The passage has been frequently used (especially the first paragraph) for speech research as well as in speech clinics. The first paragraph consisted of 98 words and took on the average 30 seconds to read. Impromptu speech was elicited by asking the subject to describe a picture. For the impromptu speech, no attempt was made to replicate the Walker study. The order of the task was counterbalanced so that the half of the subjects started with oral reading while the other half started with impromptu speech.

Each subject was seated in a sound-treated booth. A condensor microphone (Sony ECM50) was placed approximately 15 cm from the subject's lips. The voices were recorded at a transport speed of 7.5 ips on an AMPEX magnetic tape recorder located in an adjacent room.

Perceptual Analysis of Pause

Five female graduate students were randomly selected from a research method course in the Department of Communication Disorders and Speech Science, University of Colorado to provide perceptual judgments as to location of pauses. The students were native speakers of English, and were free from problems in speech, language, hearing and reading. Each student was seated in a quiet room, listened to oral reading played via a loudspeaker (Ampex 622) and was instructed to place slash (/) marks on a response form at "pauses" detected in the oral reading. The response form had the first paragraph of the Rainbow passage typed in double spacing. "Pause" was not defined in the instructions and its definition was left to the students. Each student was allowed to listen to the recordings as many times as necessary. The results of pause identification were used to determine appropriate values of input parameters to an automatic durational analyzer (described below) and to verify agreement in numbers and locations of pauses between the perceptual and automatic analyses.

Durational Analysis Procedures

The recorded voices were played back, fullwave rectified and smoothed by an RC lowpass filter (Dobkin, 1969). The rectifying and smoothing extracted the intensity envelope in real time. This intensity envelope was digitized by a 12-bit analog-to-digital converter at a rate of 1000 times per second, and stored on a 386 microcomputer disk

using CSRE (Canadian Speech Research Environment) software.

The digital intensity envelope was subsequently analyzed for utterance and pause durations by software developed by Horii (1983a). Given a specification of the maximum amplitude threshold for pause, the minimum duration of pause, and the minimum duration of utterance, the program identified utterances and pauses and printed out means and standard deviations of utterance and pause durations, number of utterances and speaking-time ratio.

Results

Perceptual Pause Analysis

Results of the pause identifications yielded unanimous agreements among the five listeners for 18 of the 20 readings. For the remaining two readings, two of the five listeners detected an additional pause. For each of the twenty readings, the number and locations of perceived pauses were noted. For those two readings with nonunanimous results, the results of the majority (3 of 5 listeners) were employed.

Reliability and Validity of Measurements

The reliability and validity of the automatic duration analyzer have been investigated and found satisfactory (Horii, 1983a). For example, when square pulses of 1 and 10 pulses/sec were submitted to the analyzer, results were within 2 milliseconds with an average difference of 0.3 milliseconds from oscillographic hand measurements. Because the recording quality affects the reliability and validity, additional tests were conducted for the voice samples under investigation. In particular, values to be used for the program input parameters (the maximum amplitude threshold for pause, the minimum pause duration, and the minimum utterance duration) were carefully explored before the final analysis.

Eventually, the following values were deemed appropriate: the maximum amplitude threshold of .2 volts (about 34 dB below the peak amplitudes, about 10 volts, of the intensity envelope), the minimum pause duration of 100 ms and the minimum utterance duration of 200 ms. These parameter values produced number and locations of pauses identical to the perceptual results for each of the twenty readings.

As an example, Figure 1 illustrates an intensity envelope with cursors positioned at the beginning (A) and end of the second utterance (B). The graphics in the middle row show an expanded waveform around the cursors (A and B), while the graphics at the bottom show waveforms further expanded at the cursors. The CSRE display shows that the duration between the two cursors is 3630 ms. Pauses are indicated by arrows (hand-drawn by the investigator after perceptual verification). A computer printout from the

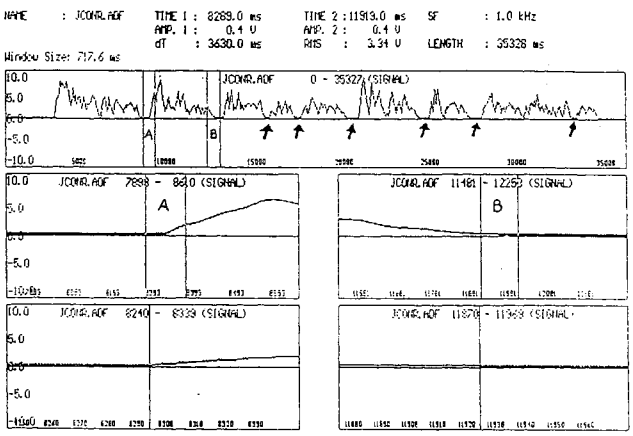


Figure 1. A CSRE display of the intensity envelope with cursors positioned at the beginning and end of the second utterance (A-B). Arrows indicate pauses detected by the automatic duration analyzer and verified perceptually.

automatic duration analyzer for the same intensity envelope listed the duration of the second utterance, in particular, to be 3629 ms. Overall, results of the automatic analyzer and manual cursor positioning demonstrated a Pearson product-moment correlation of .999 with the average absolute measurement difference of less than 3 ms. The number and locations of pauses were identical to the listeners' responses.

Table 1.

Individual and group results for oral reading. Variables were mean utterance durations (U), utterance duration standard deviations (Usd), mean pause durations (P), pause duration standard deviations (Psd) (all in milliseconds) and speaking-time ratio (STR).

	U	Usd	P	Psd	STR
Male Subjects					
1	3092	1319	406	197	.90
2	2917	1172	368	114	.89
3	3083	980	419	170	.89
4	2121	701	512	248	.82
5	2069	897	568	276	.80
6	2044	1117	475	180	.82
7	2261	1065	310	133	.89
8	2506	584	492	166	.85
9	3237	1500	372	244	.91
10	2348	851	495	285	.84
Mean	2568		442		.86
S. D.	445		75		.04
Female Subjects					
1	2498	1114	627	318	.81
2	3021	1315	360	168	.90
3	2630	1417	427	213	.87
4	2422	1333	332	148	.89
5	3772	1263	528	73	.89
6	2514	952	387	170	.88
7	2871	1641	297	127	.91
8	3484	1170	419	73	.91
9	2321	623	321	98	.89
10	2578	1020	571	209	.83
Mean	2811		427		.88
S. D.	457		107		.03

Table 2.
Individual and group results for impromptu speech. Variables were mean utterance durations (U), utterance duration standard deviations (Usd), mean pause durations (P), pause duration standard deviations (Psd) (all in milliseconds) and speaking-time ratio (STR).

	U	Usd	P	Psd	STR
Male Subjects					
1	2018	1093	1490	1759	.58
2	1324	661	1023	816	.57
3	1395	902	721	472	.66
4	1443	866	947	597	.61
5	1813	1041	898	1094	.67
6	2527	1753	645	676	.80
7	2078	1519	888	819	.71
8	1841	1077	781	530	.71
9	1787	807	842	602	.68
10	2235	1441	643	489	.78
Mean	1846		888		.68
S. D.	366		233		.07
Famale Subjects					
1	1542	841	1950	1532	.45
2	1966	1054	721	457	.74
3	1825	949	800	511	.70
4	1959	1048	637	376	.76
5	1832	1255	1266	1358	.60
6	1788	836	1207	1159	.60
7	2004	895	970	719	.70
8	1818	1460	735	544	.72
9	2804	1510	518	285	.85
10	2118	975	506	268	.81
Mean	1966		931		.69
S. D.	316		420		.11

Table 3.
Means and standard deviations (in parentheses) of the mean utterance durations (U), mean pause durations (P), both in milliseconds, and speaking-time ratio (STR) for the oral reading and impromptu speech by the twenty young adults.

Variables	Oral Reading	Impromptu Speech
U	2689 (467)	1906 (347)
P	434 (93)	909 (341)
STR	.87 (.04)	.69 (.09)

Measurement Results

Tables 1 (previous page) and 2 show the results of the durational analysis for the oral reading and spontaneous speech, respectively. Tables show individual and group results for mean utterance duration (U), utterance duration standard deviation (Usd), mean pause duration (P), pause duration standard deviation (Psd) and speaking-time ratio. All durational measures are presented in milliseconds.

Two-factor (gender x task) analysis of variance with repeated measures revealed that the gender differences and the gender-task interactions were nonsignificant at the .05 level for utterance and pause durations and for speaking-time ratio. The task differences, however, were significant at the .001 level for each of the three variables (Utterance duration: $F1,18=25.45$. Pause duration: $F1,18=39.12$. Speaking-time ratio: $F1,18=63.66$).

Because the male and female differences were nonsignificant, the measurement results were combined, and Table 3 (following page) shows means and standard deviations (in parentheses) of the mean utterance and pause durations in milliseconds and speaking-time ratio. As seen in the table, the oral reading task produced longer mean utterance (2689 ms), shorter pause (434 ms) and greater speaking-time ratio (.87) than the impromptu speech (1906 ms, 909 ms and .69, respectively). The greater speaking-time ratio was a necessary consequence of the increased utterance duration and decreased pause duration in oral reading.

Discussion

General Findings

The current investigation revealed that the oral reading was associated with greater mean utterance durations and smaller pause durations compared to impromptu (picture description) speech. The findings were not surprising because of the nature of the tasks. When individual data were examined, only one subject had shorter mean utterance and only two subjects yielded longer mean pause durations in oral reading than impromptu speech. There was, however, no exception regarding greater speaking-time ratios for oral reading than impromptu speech. On the average, the mean utterance and pause durations and speaking-time ratio of oral readings were approximately 141%, 48% and 126%, respectively, of the values for the impromptu speech.

Walker (1988) reported an average speaking-time ratio of .70 for 120 young adults reading the same text, i.e., the first paragraph of the Rainbow passage. The present investigation yielded the speaking-time ratio ranging from .80 to .91 with a mean of .87 for the twenty young adults. The current finding is consistent with the findings of earlier studies (Horii, 1983b; Horii and Ramig, 1987) that reported average speaking-time ratios of about .85 for the same text by adult subjects.

The average STR value of .87 found in the present study is much greater than the values reported by Fairbanks and Hoaglin (1941) ranging from .53 to .71. In their study, however, the subjects were six amateur actors and were asked to read a 27-word passage in five different emotional states, i.e., contempt, anger, fear, grief and indifference. The passage was "There is no other answer. You've asked me that question a thousand times, and my reply has always been the same. It always will be the same" (p.85). With such a dramatic reading, there could be more and longer pauses in the reading yielding low STR values.

For the impromptu speech, the average STR value of .69 found in this study was difficult to compare with findings of other studies mainly because of the differences

in the manner of eliciting impromptu speech. Walker (1988) reported an average STR of .62 while Till and Alp (1991) reported an average of .78 when their subjects were asked to engage in conversational monologue with a listener. As stated earlier, the task (picture descriptions, for example) and the subject matter would have considerably affected the utterance, pause durations and frequencies.

The Effect of the Maximum Amplitude Threshold for Pause

As mentioned earlier, the automatic analysis program uses the maximum amplitude threshold to define a pause. To assess the effect of the threshold on the durational measures, STR in particular, two readings (one from the male and one from the female group) were randomly selected and submitted to the durational analyzer with various values of the amplitude threshold.

Results are summarized in Figure 2 where the abscissa is the amplitude threshold in volts and the ordinate is the resulting STR values. The x's represent results for the male subject (#9) and the o's for the female subject (#3). The amplitude thresholds between the two vertical dotted lines yielded the number and locations of pauses identical to those perceptually identified by the majority of the five listeners.

As seen in the figure, the effect of the amplitude threshold was systematic as expected. As the threshold increased, the STR decreased. When the threshold was set too low, such as .1 volts in the figure, the entire envelope was above the threshold, and thus no pause was detected, resulting in an STR of 1.0 for the male subject. For the female subject, the same .1 volts threshold produced only two pauses and the STR was an inflated .97. On the other hand, when the threshold was set too high (above .4 volts for these examples), the STR decreased, with more pauses detected than the perceptual results. Although not shown in the figure, it should be obvious that considerable changes occurred to the calculated means and standard deviations of utterance and pause durations when the number of pauses (and utterances) changed for the same reading, such as between the thresholds of .4 and .6 volts. The figure shows, however, that STR values were less affected by the change in the number of pauses and utterances.

Consideration of the effect of the amplitude threshold dictates that the best threshold is the lowest threshold that yielded the same number and locations of pauses as the perceptual results. This may not be always possible for some voice recordings with poor signal-to-noise ratios. Furthermore, certain types of speech, e.g., speech produced by stutterers and dysarthrias, may not yield near-unanimous perceptual identification of pauses. Indeed, the term "pause" itself can become a source of controversy both in terms of definition and measurement.

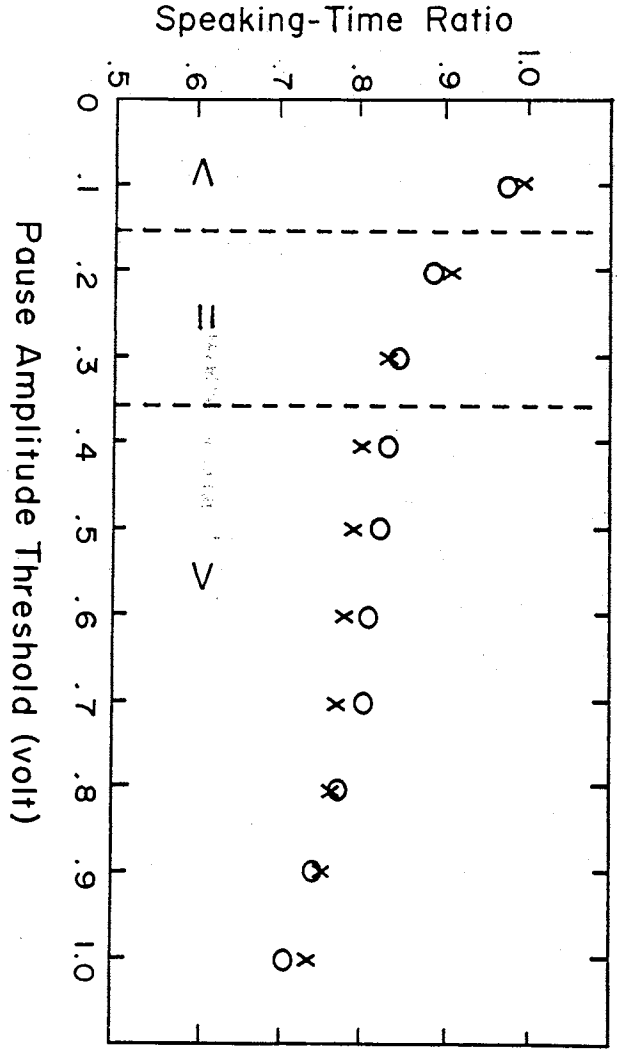


Figure 2. STR as a function of the maximum amplitude threshold of pause for randomly selected readings of a male subject #9 (x's) and a female subject #3 (o's). The threshold values between the dotted lines yielded the number and locations of pauses identical between the automatic analyzer and perceptual results.

Issues of the Definitions of Pause

The automatic duration analyzer defined the pause as segments of the intensity envelope with the maximum amplitude of .2 volts and with the minimum duration of 100 ms. In essence, a "pause" was a "silent gap" in the acoustic signals. Fortunately, for the oral reading of the particular text used in the study, the "silent gaps" were nearly identical to the listeners' intuitive definition of "pause". The only exceptions which occurred were between "look" and "but" in the sentence "People look, but no one ever finds it." For these instances, there were no "silent gaps" in the acoustic signals. As perceptual studies demonstrated, "pause" can be perceived (via vowel prolongations prior to a phrase

boundary, for example) without silent gaps in the acoustic signals. It should also be noted that the automatic duration analyzer treats so-called "filled pauses" or hesitation pauses with overt vocalizations such as "uh ..." or "humm ..." as utterances. In oral reading, such filled pauses rarely occur, and indeed none occurred in the present reading samples.

Consideration of these issues surrounding "pause" suggests that the use of agreement between the automatic analyzer and perceptual results may not always be an appropriate criterion for determining the thresholds of pauses. STR values were found, however, to be quite insensitive to variations of amplitude thresholds in contrast to mean utterance and pause durations. Further experiences in durational analyses will hopefully allow determination of the amplitude threshold without rigorous perceptual experiments.

Significance of STR Measurements

An STR value of .85, for example, means that 85% of the total speaking time is spent for actual speaking and 15% for pauses. As stated earlier, such information is useful in delineating general characteristics of speech produced by individuals with various speech disorders. More specific merits of STR measurements, however, lie in the fact that, in oral reading tasks at least, STR serves as a reasonable estimate of efficacy of the speech production mechanisms as reflected in temporal patterns of respiration. In an investigation of respiratory airflow associated with oral reading, Horii and Cooke (1978) found exhalatory durations of their eight adult normal-speaking subjects to be about 87% of the total respiratory cycle (exhalation plus inhalation). Other investigators reported similar values ranging from 80% to 90% of speech breathing of normal-speaking subjects (Itoh, 1975; Itoh & Horii, 1985; Itoh et al., 1982; Till & Alp, 1991).

When the exhalatory air is not efficiently used in connected speech, utterance durations become shorter because the air reservoir is expended more quickly, resulting in smaller percentages of exhalatory durations. The percentage becomes smaller yet if the speaker attempts to compensate for the quicker consumption of air reservoir by deeper and longer inhalations. Such inefficient usage of air reservoir can occur at the laryngeal level and/or supraglottic articulatory levels due to various structural and neurological problems. Lower than normal STR values, therefore, would indicate less efficient respiratory functions for speech. Prior to or concurrent with investigations of various speech disorders, further normative data must be accumulated because the number of the present investigation, i.e., 20, was admittedly small, and the age group was limited to young adults. The effects of the types of reading texts, the manner of eliciting impromptu speech, and the sample sizes warrant further investigation.

Acknowledgments

This work is supported in part by an NIH grant 1P60DC00976 and RO1NS24409.

References

Dobkin, R. C. (1969). Precision AC/DC converters. National Semiconductor Corporation, LB8, 1-2.

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

Fairbanks, G. & Hoaglin, L. W. (1941). An experimental study of the durational characteristics of the voice during the expression of emotion. Speech Monographs, 8, 85-90.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. Language and Speech, 15, 103-113.

Hartman, D. & Danhauer, J. (1976). Perceptual features of speech for males in four perceived age decades. Journal of Acoustical Society of America, 59, 713-715.

Horii, Y. (1983a). An automatic analysis method of utterance and pause lengths and frequencies. Behavior Research Methods & Instrumentation, 15, 449-452.

Horii, Y. (1983b). Utterance and pause lengths and frequencies in oral readings by superior esophageal speakers and by normal speakers. Computers and Biomedical Research, 16, 446-453.

Horii, Y. (1983c). Some acoustic characteristics of oral reading by ten- to twelve-year-old children. Journal of Communication Disorders, 16, 257-267.

Horii, Y. & Cooke, P. A. (1978). Some airflow, volume, and duration characteristics of oral reading. Journal of Speech and Hearing Research, 21, 470-481.

Horii, Y. & Ramig, P. R. (1987). Pause and utterance durations and fundamental frequency characteristics of repeated oral readings by stutterers and nonstutterers. Journal of Fluency Disorders, 12, 257-270.

Itoh, M. (1975). Respiratory timing, volume and airflow characteristics of the deaf individuals during speech. Unpublished dissertation. Purdue University.

Itoh, M. & Horii, Y. (1985). Airflow, volume, and durational characteristics of oral reading by the hearing-impaired. Journal of Communication Disorders, 18, 393-407.

Itoh, M., Horii, Y., Daniloff, R., & Binnie, C. A. (1982). Selected aerodynamic characteristics of deaf individuals during various speech and nonspeech tasks. Folia phoniatrica, 34, 191-209.

O'Connell, D. C. & Kowal, S. (1981). Pausology. In W. Sedlow & S. Sedlow (Eds.), Computer uses in the study of languages (Vol. 3): Cognitive approaches. The Hague: Mouton.

Oyer, H. H. & Deal, L. V. (1985). Temporal aspects of speech and the aging process. Folia phoniatrica, 37, 109-112.

Rochester, S. S. (1973). The significance of pauses in spontaneous speech. Journal of Psycholinguistic Research, 2, 51-81.

Ruder, K. F. & Jensen, P. J. (1970). An apparatus and procedure for the perceptual study of speech pauses. Journal of Experimental Analysis of Behavior, 14, 287-289.

Till, J. A. & Alp, L. A. (1991). Aerodynamic and temporal measures of continuous speech in dysarthric speakers. In Moore, C. A., Yorkston, K. M., & Beukelman, D. R. (Eds.), Dysarthria and Apraxia of Speech: Perspectives on Management. Baltimore: Paul H. Brookes Publishing Co.

Walker, V. G. (1988). Durational characteristics of young adults during speaking and reading tasks. Folia phoniatrica, 40, 12- 20.

Watanabe, H., Shin, T., Oda, M., Fukaura, J. & Komiyama, S. (1987). Measurement of total actual speaking time in a patient with spastic dysphonia. Folia phoniatrica, 39, 65-70.

NCVS Status and Progress Report - 5
Novmeber 1993, 69-72

Arytenoid Separation for Impaired Pediatric Vocal Fold Mobility

Steven D. Gray, M.D.
Division of Otolaryngology, University of Utah
Steven M. Kelly, M.D.
Division of Otolaryngology, University of Utah
Heather Dove, M.A., CCC/SP
Division of Otolaryngology, University of Utah

Abstract

Impaired vocal fold motion may result from cricoarytenoid joint fixation, bilateral vocal fold paralysis, or interarytenoid scarring. Traditional surgical techniques have focused on lateralization or resecting the arytenoid for airway improvement. This paper discussed three cases of bilateral reduced vocal fold motion of neurogenic etiology treated with posterior cricoid grafting to cause a wider resting position of the vocal folds and arytenoids. Airway improvement occurred in all. Voice results have been encouraging. Advantages of this procedures are: symmetrical vocal folds, no vocal fold or joint scarring, larynx remains a candidate for electrical pacing when that becomes available. Acoustic and aerodynamic voice results presented. Results should be considered preliminary.

Introduction

Impaired vocal fold mobility may result from a variety of disorders including cricoarytenoid joint fixation, neurologic injury resulting in vocal fold paralysis or paresis, and interarytenoid scarring. As has been recognized by other authors, distinguishing between these disorders is sometimes difficult even at the time of direct laryngoscopy.[1] Distinction is important, however, as the best treatment modality for each situation may differ. Bilateral vocal fold paralysis in the pediatric population, for example, is often transient. Rosin et al. recently reported that 3 of 19 patients (16%) requiring tracheotomy for bilateral vocal fold paralysis had spontaneous resolution allowing decannulation. Six others were decannulated following ventriculoperitoneal shunts for hydrocephalus. Another seven, however, could not be decannulated at the time of their study.[2]

Tracheotomy for airway compromise from impaired vocal fold mobility has been the gold standard of treatment. Because of the care and problems associated with tracheotomies, many surgeons in the past have described various alternative techniques to improve the glottic airway. These techniques have focused on adequacy of the airway to allow decannulation while avoiding aspiration and maintaining some voice. Various authors have described arytenoidectomy or partial arytenoidectomy for bilateral impaired vocal fold mobility, mostly in adults.[3,4] Only a few cases of arytenoidectomy in children have been reported.[5,6] Dennis and Kashima have described success with endoscopic C02 laser partial cordectomy in adults with vocal fold impairment[7] Kashima also discussed his results with C02 laser transverse cordotomy for bilateral vocal fold impairment[1] Multiple methods of lateral fixation of the vocal fold have had varying success, again mostly in adults.[8,9] Tucker has described his success with laryngeal reinnervation.[10]

Interarytenoid division with long-term stenting has been described for impaired vocal fold mobility from posterior glottic stenosis. Goodwin et al. reported six adult patients, five of whom were tracheotomy dependent, treated with posterior glottic scar excision through a midline thyrotomy and long-term stenting without grafts. All were successfully decannulated with normal vocal fold mobility in three and improved mobility in the others. Two patients retained normal voices and the remaining patients had improved voices, as judged subjectively.[11]

There have also been reports of children with abnormal vocal fold mobility secondary to interarytenoid scarring and posterior glottic stenosis who have been treated with posterior cricoidotomy and cartilage grafting. Zalzal reported 11 children treated for posterior laryngeal stenosis by this method, five of whom had associated vocal fold impairment. All five were decannulated and had return of normal vocal fold mobility. Voice was evaluated subjectively only, with all voices being "husky or hoarse" and two of the five being worse than preoperatively.[12] In another paper, Zalzal et al. reported vocal quality results in 16 children decannulated following laryngeal reconstruction, including five patients treated with posterior cricoidotomy and cartilage grafting for abnormal vocal fold mobility associated with stenosis. Voice evaluation consisted of analysis of quality, pitch, volume, resonance, speaking rate, intelligibility, and overall voice severity as judged by speech pathologists. Two of the five were judged with severe, one of the five with moderate, and two of the five with mild overall voice disorder. Most had either prominent breathiness, hoarseness, or low volume.[13] Cotton reported 61 children who underwent posterior glottic cartilage grafting for laryngotracheal stenosis. Forty-seven of these children also had some degree of bilateral vocal fold impairment with 20% being fixed bilaterally. Preoperative and postoperative voice analysis was restricted to subjective analysis by parents.[14]

This paper reviews the cases of three children with bilateral impaired vocal fold mobility from presumed neurogenic etiology treated with an extended posterior cricoid split and cartilage grafting. In these three patients there was no evidence of posterior glottic stenosis or interarytenoid scarring as a cause of the vocal fold immobility. The goal was to accomplish arytenoid separation and a larger posterior glottis, while maintaining laryngeal symmetry, mobile arytenoids, and healthy membranous vocal folds.

Materials and Methods

Subjects

These patients, ranging in age from 23 months to 7 years underwent posterior cricoid split with costal cartilage grafting for bilateral vocal fold impaired mobility. A summary of each patient is found in Table 1.

Preoperative Evaluation

Thorough history and physical examinations were performed on all patients. Flexible laryngoscopy and, in some instances, flexible bronchoscopy were used for dynamic airway evaluation. Rigid laryngoscopy and bronchoscopy under general anesthesia were performed and videotaped. Patients were selected that had no symptoms of aspiration preoperatively and who demonstrated adductory motion on laryngoscopy. Although no electromyography was performed, clinically these children demonstrated poor or no fold abduction. They did show adduction, indicating ability to close the glottis.

Table 1.

	Patient 1	Patient 2	Patient 3
age	7 years	3 1/2 years	23 months
etiology of vocal fold impairment	closed head injury, trans herniation	bilateral recurrent nerve injury following bilateral cricopharyngeal myotomies	congenital
associated problems	none	laryngomalacia, GE reflux, cricoesophageal stricture	BPD
vocal fold mobility	bilateral abductor paralysis minimal adduction	bilateral vocal fold paralysis	bilateral abductor paaralysis slight adduction
laryngeal stenosis/ posterior glottic scarring/ aytenoid fixation	none	none	grade II subglottic stenosis
tracheotomy dependent/ duration	no - severely limited physical activity secondary to stridor	yes - 3 years, 4 months	yes - 19 months
graft type	posterior costal cartilage	posterior costal cartilage	anterior and poster ior costal cartilage
stenting duration	11 days	30 days	28 days
time to decannulation after surgery	13 days	90 days	49 days

Surgical Approach

The surgical approach to the posterior cricoidotomy with cartilage graft is similar to that previously described by Cotton and Zalzal.[12,14] The posterior cricoidotomy is performed and carried superiorly slightly into the interarytenoid muscle. If there is interarytenoid scarring, the scar tissue is divided but not excised. The uninvolved interarytenoid muscle is left intact. Some mild arytenoid separation will occur by dividing the posterior cricoid; greater separation occurs with division of the fibrous network immediately superior to the cricoid. Further separation may be accomplished with progressive (inferior to superior) division of the interarytenoid muscle. However, we do not recommend complete or extensive division of the inter-arytenoid muscle since adequate arytenoid separation occurs without it and the interarytenoid muscle is a likely key adductor whose function will ideally be preserved. Interarytenoid muscle function is probably an important factor in preventing postoperative aspiration. Surgical injury to this muscle should be minimized. In cases of impaired mobility due to interarytenoid scarring, muscle division may be required.

It is not necessary to place part of the cartilage graft into the interarytenoid area. This superior positioning of the graft may in fact impair arytenoid adduction, leading to aspiration. Therefore, the cartilage is sutured to the cricoid without superior cartilage extension. We leave a perichondrial flap extending superiorly to cover the exposed inferior portion of the interarytenoid muscle, although this may not be necessary.

A laryngeal stent is placed from above the level of the stoma to a few millimeters above the false folds. The

stent we now use for this procedure is an endotracheal (ET) tube which has been halfway crimped horizontally with a towel clamp and then autoclaved for 30-45 seconds so that the ET tube maintains the crimped shape. (Photos 1A-D; see center-bound photographic plate). (Although this technique was taught to us by Dr. Rodney Lusk in 1985, I have heard that others have independently used this technique as well.) This creates a stent with a V shape anteriorly to fit in the anterior commissure and a rounded, broad shape posteriorly to fit in the posterior glottis. The stent above and below the crimp is round. We have switched to this crimped stent for this particular procedure since we suspect that rounded stents have occasionally caused some membranous fold compression atrophy and abduction of the arytenoids leading to some dysphonia and breathiness though this has not been confirmed by comparison studies. Preoperative and postoperative laryngeal configurations show that the posterior glottis is open, but the membranous folds remain more medialized (Photos 2A and B; see photo plate). The postoperative configuration is due to two factors: arytenoids in a wider resting separated state, and arytenoids which are still medially rotated in a resting or semi-adducted position. The stent is secured with 2-0 prolene sutured through the cricoid and tied over the strap muscles subcutaneously and left for approximately two - four weeks.

Voice Analysis

Voice recordings were made on all three patients allowing evaluation of mean habitual fundamental frequency, jitter, and overall quality. Patients #1 and #2 were cooperative enough to allow limited aerodynamic evaluation and more detailed voice analysis including maximum phonation time, DC airflow, mean habitual intensity, and intensity range. Patient #3 was too young to undergo additional voice or aerodynamic evaluation.

Results

Average follow up time was 23 months. All three patients were successfully decannulated with time to decannulation ranging from 13 days to three months postoperatively. Decannulation was delayed beyond the immediate stent removal period in one patient (#2) because of postoperative granulation tissue. This patient also had persistent preoperative supraglottic collapse and arytenoid prolapse from laryngomalacia, which further delayed her decannulation. All patients continued to have abnormal vocal fold mobility postoperatively. All patients are able to participate in normal play activity with adequate airway function according to their parents. Audible breathing is still present in the patient with laryngomalacia. No patient has had difficulty with aspiration.

Voice results are listed in Table 1. Perceptual voice quality was judged by our speech pathologist on a seven point scale with one being normal and seven aphonic. Patient #1 was rated as normal and the others were given a rating of two because of mild breathiness. All patients had age appropriate mean fundamental frequency. Mean jitter was normal in patients #1 and #2 but increased in patient #3. Patients #1 and #2 underwent additional aerodynamic and voice analysis. The AC:DC airflow ratio was normal in both patients. Signal:noise ratio was normal in patient #2 and just slightly low in patient #1. DC airflow was within normal limits (using adult values) for patient #2 but probably represents a high flow rate for children. DC flow was elevated for patient #1. This indicates a high flow rate through the glottis during speech. Nevertheless, both patients' ability to entrain the flow into vocal pulses (AC:DC ratio) was adequate enough (ratio was 1.6 and 3.4) that they have no or minimal breathiness and normal intensity.

Discussion

Bilateral vocal fold immobility has been a difficult problem to manage clinically. The gold standard of treatment at most institutions has been tracheotomy. Many techniques have been described including arytenoidectomy, vocal fold lateralization, partial cordectomy/cordotomy and laryngeal reinnervation to avoid tracheotomy or to allow decannulation. The objective of treatment not only focuses on airway adequacy but also on the avoidance of aspiration and on the preservation or improvement of voice. In the pediatric population, some patients are found who do not have complete bilateral vocal fold paralysis, but rather demonstrate impaired abduction of both vocal folds. These have no aspiration, adequate voice, but experience significant airway obstruction.

Our clinical experience indicates that adduction generally seems to be a stronger component than abduction of the vocal folds. In unilateral paralysis it is quite common to see over- adduction by the opposite fold. When some adduction is present preoperatively, separating the arytenoids (instead of arytenoidectomy) capitalizes on this adductor motion to provide glottic closure. This may explain in part why patients in this small series did not experience aspiration postoperatively.

Posterior cricoidotomy with cartilage grafting and stenting for treatment of posterior glottic and subglottic stenosis has had good success.[12,14] Posterior cricoidotomy with cartilage grafting has been indicated for posterior glottic and/or sub-glottic stenosis and total glottic or subglottic obstruction.[12] We propose expansion of these indications to include selected patients with bilateral impaired vocal fold mobility secondary to abductor paralysis or paresis with some residual adductory motion.

Selection criteria for this procedure would include some residual adduction of the vocal folds, presence of laryngeal sensation and intact swallowing function as aspi-

ration can be a serious complication if careful selection of patients is not done. Certainly, if there is evidence of aspiration preoperatively, any arytenoid separation would not be indicated. We are encouraged by the airway and voice results obtained by using this procedure. Our subject numbers are few and the results are considered preliminary. No comparison with other methods was performed. Although this procedure is not recommended for all patients who have bilateral vocal fold paralysis or cricoarytenoid joint fixation, we feel that it does offer advantages to those who have residual adductor vocal fold motion. This procedure is not felt to be more advantageous compared to standard arytenoidectomy or lateralization procedures in patients with no vocal fold mobility.

Table 2.
Advantages/Disadvantages of Arytenoid Separation Compared to Arytenoidectomy

Advantages:
- Preserves any residual vocal fold motion
- No membranous vocal fold scarring
- Treats concurrent interarytenoid scar or posterior subglottic stenosis
- Laryngeal symmetry maintained
- Preserves vocal structures for future rehabilitation developments, i.e. laryngeal pacing or reinnervation
Disadvantages:
- Tracheotomy required
- Open procedure - longer operating time, possibility of anterior commissure misalignment
- Laryngeal stenting required
- Donor site morbidity

Acknowledgment

This work was supported by Public Health Service #P60-DC00976.

References

1. Kashima HK. Bilateral vocal fold motion impairment: Pathophysiology and management by transverse cordotomy. *Ann Otol Rhinol Laryngol.* 1991;100:717-721.
2. Rosin DF, Handler SD, Potsic WP, Wetmore RF, Tom LWC. Vocal cord paralysis in children. *Laryngoscope.* 1990;100:1174-1179.
3. Ossoff RH, Duncavage JA, Shapshay SM, Krespi YP, Sisson GA Sr. Endoscopic laser arytenoidectomy revisited. *Ann Otol Rhinol Laryngol.* 1990;99:764-771.
4. Crumley RL. Endoscopic laser medial arytenoidectomy for airway management in bilateral laryngeal paralysis. *Ann Otol Rhinol Laryngol.* 1993;102:81-84.
5. Choi SS, Cotton RT. Arytenoidectomy in children. Presented at the American Broncho-Esophagological Association Meeting. Palm Beach, FL, 1990.
6. Cohen SR. Arytenoidectomy in Children. *Laryngoscope.* 1993;83:1293-1299.
7. Dennis DP, Kashima H. Carbon dioxide laser posterior cordectomy for treatment of bilateral vocal cord paralysis. *Ann Otol Rhinol Laryngol.* 1989;98:930-934.
8. Geterud A, Ejnell H, Stenborg R, Bake B. Long-term results with a simple surgical treatment of bilateral vocal cord paralysis. *Laryngoscope.* 1990;100;1005-1008.
9. Amedee RG, Mann WJ. A functional approach to lateral fixation in bilateral abductor cord paralysis. *Otolaryngol Head Neck Surg.* 1989;100:542-545.
10. Tucker HM. Long-term results of nerve-muscle pedicle reinnervation for laryngeal paralysis. *Ann Otol Rhinol Laryngol.* 1989;98:674-6.
11. Goodwin WJ, Isaacson G, Kirchner JC, Sasaki CT. Vocal cord mobilization by posterior laryngoplasty. *Laryngoscope.* 1988;98:846-848.
12. Zalzal GH. Rib cartilage grafts for the treatment of posterior glottic and subglottic stenosis in children. *Ann Otol Rhinol Laryngol.* 1988;97:506-511.
13. Zalzal GH, Loomis SR, Derkay CS, Murray SL, Thomsen J. Vocal quality of decannulated children following laryngeal reconstruction. *Laryngoscope.* 1991;101:425-429.
14. Cotton RT. The problem of pediatric laryngotracheal stenosis: A clinical and experimental study on the efficacy of autogenous cartilaginous grafts placed between the vertically divided halves of the posterior lamina of the cricoid cartilage. *Laryngoscope.* 1991;101:12;pt 2;suppl 56;1-34.

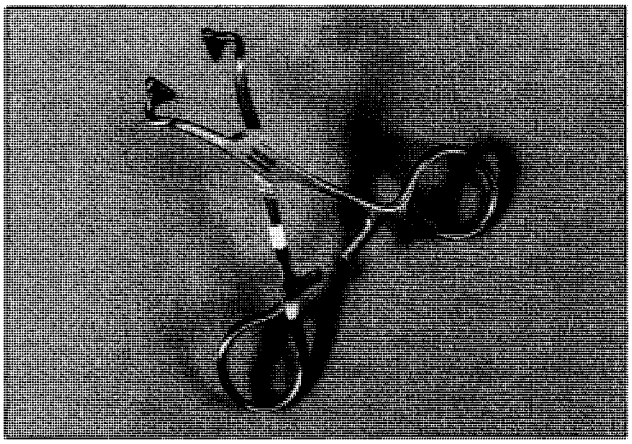


Photo 1A) Edna non-perforating towel clamp.

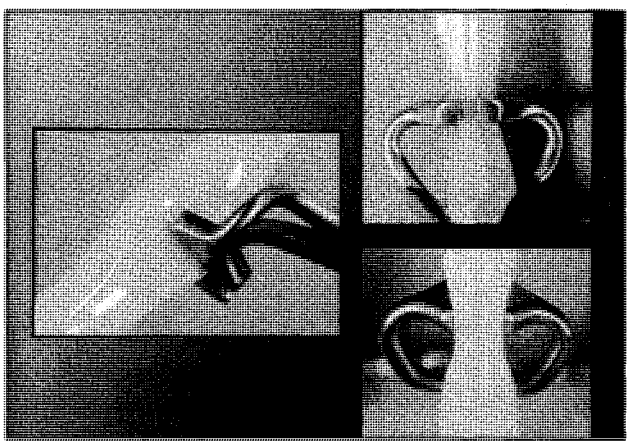


Photo 1B) Application of clamp to curved portion of endotracheal tube, clamp is applied to half of endotracheal tube so that anterior half is tightly together and posterior half is rounded. Stent and clamp are then placed into autoclave for 45 seconds. Clamp is then removed.

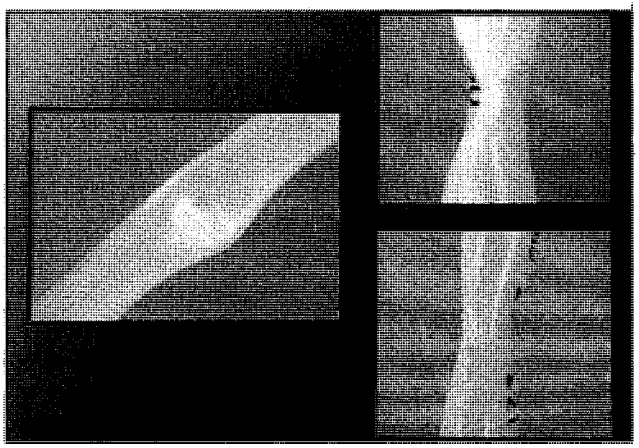


Photo 1C) Photos of stent after autoclave and clamp removed.

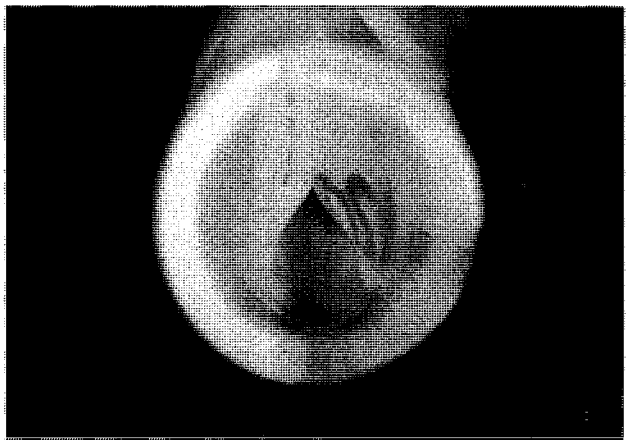


Photo 1D) Notice that the internal view of the stent shows a tear drop configuration to the part that will be positioned at the glottis. Stent will be placed so that membranous cords fit into tightly clamped area while rounded portion fits into posterior glottis.

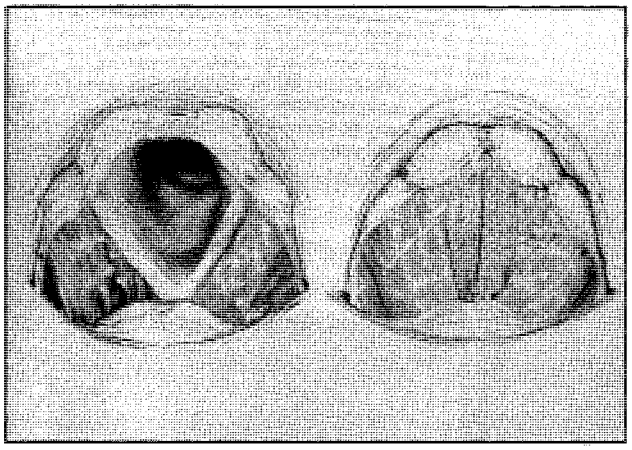


Photo 2A) Digitized images captured and imported from videotape of normal abduction and adduction of the human larynx. Notice that during normal abduction the vocal processes rotate outward and the entire arytenoids slightly separate from each other. Both actions provide a wide glottic airway.

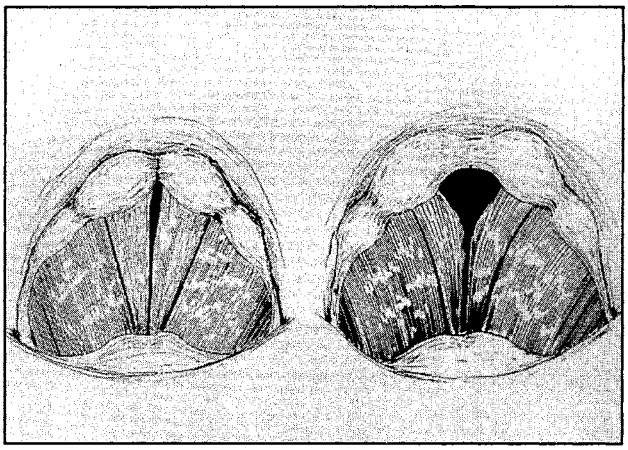


Photo 2B) Digitized images captured from videotape of the larynx before and after arytenoid separation surgery as described in the paper. Note that the body of the arytenoids are further separated but the arytenoids are still in a relatively adducted position. If adductory motion is still somewhat intact and the interarytenoid muscle functions then the arytenoid bodies will still approximate during adduction.

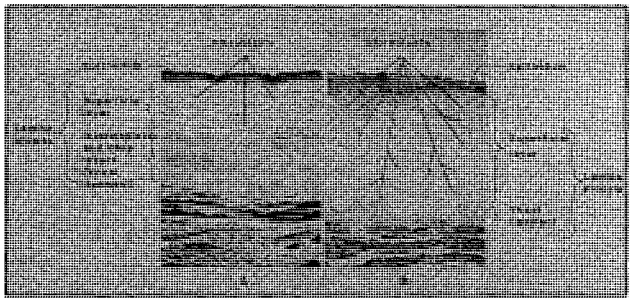


Photo 3) Histological picture demonstrating the differences in density of fibroblasts in the lamina propria of the vocal fold mucosa of an adult (photo A, at left) and a 4-year-old child (photo B, at right). Hematoxylin-Eosin stain. (From Hirano M, Kurita S, Nakashima T: Growth, development, and aging of the human vocal folds, in Bless, Diane M; Abbs, James H (eds.): Vocal Fold Physiology: Contemporary Research and Clinical Issues, *San Diego, California, College-Hill Press, 1983, page 36. Used by permission).*

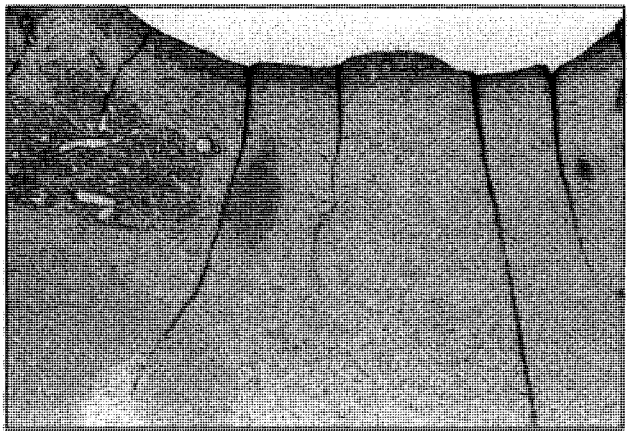


Photo 4) Goat larynx specimen after 3 month placement of Cotton/Lorenz stent in larynx. Section through vocal process of arytenoid shows erosion of cartilage. (Hematoxylin-Eosin stain, 30X).

Part II

Tutorial reports and summaries of Dissemination of Information, Continuing Education and Training

NCVS Status and Progress Report - 5
November, 1993, 73-82

Examination of the Laryngeal Adduction Measure EGGW

Ronald Scherer, Ph.D.
The Recording and Research Center, The Denver Center for the Performing Arts
Vernon Vail, B.S.
The Recording and Research Center, The Denver Center for the Performing Arts
Bruce Rockwell, M.D.
Department of Radiology, Mercy Hospital, Pittsburgh

Introduction

Adduction of the larynx at the level of the vocal folds and arytenoid cartilages is a primary peripheral and mechanical control variable in phonation. Laryngeal qualities from breathy to constricted phonation are dependent on glottal adduction (e.g., Scherer, Gould, Titze, Meyers and Sataloff, 1988b). A noninvasive measure of laryngeal adduction is therefore of importance and interest for both theoretical and applied purposes.

The electroglottograph is a noninvasive instrument that provides a signal related to glottal kinematics (ref. Baken, 1987 and 1992, Colton & Conture, 1990, and Orlikoff, 1991, for a relatively thorough review of principles, history, pitfalls, and relationships to laryngeal function). Values of the electroglottograph (EGG) waveform function correspond strongly to the amount of contact area between the two vocal folds, but not in ways completely understood or straightforward (Childers & Krishnamurthy, 1985; Childers, Alsaka, Hicks & Moore, 1987; Anastaplo & Karnell, 1988; Scherer, Druker & Titze, 1988a; Childers, Hicks, Moore, Eskenazi & Lalwani, 1990; Titze, 1990). The shape of the EGG waveform may be related to specific configurations and motions of the vocal folds relevant to normal, abnormal, and trained voices (e.g., Fourcin, 1974; Titze, 1984, 1989; Dejonckere & Lebacq, 1985; Childers, Alsaka, Hicks & Moore, 1986; Baken, 1987; Scherer & Titze, 1987; Gerratt, Hanson & Berke, 1987; Painter, 1988; Motta, Cesari, Iengo & Motta, 1990; Brown & Scherer, 1992).

This study examines a simple measure of the waveform of the electroglottograph, called the EGGW measure, to determine its relationship to other measures of adduction, including a direct measure of the gap between the vocal processes of the arytenoid cartilages. The results will show that EGGW appears to be a significant measure of adduction, at least for the limited number of subjects and phonatory conditions reported in this study.

Definition of the Measure EGGW

Figure 1 illustrates an electroglottographic waveform during normal phonation and the definition of the simple measure EGGW (the W stands for width). At the 25% level (Orlikoff, 1991; cf. Rothenberg & Mahshie, 1988, who used 35%, and Higgins & Saxman, 1993, who used 40%), the distance A on Fig. 1 corresponds to an approximate time of glottal closure, and distance B to an approximate time of glottal opening. The ratio of A to A+B is an estimate of the portion of the cycle the glottis is closed. EGGW is equal to A divided by A+B, and could be called a glottal closed quotient. EGGW is obtained for each cycle

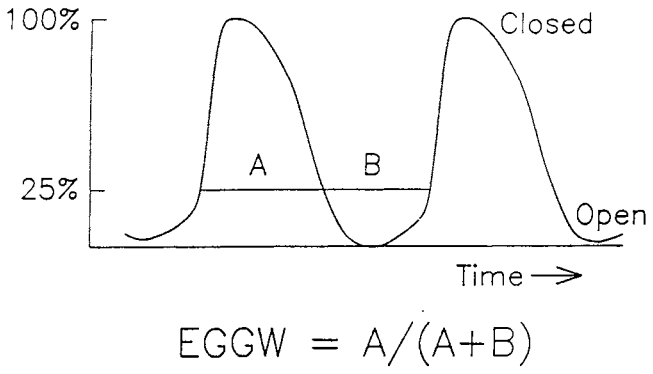


Figure 1. Definition of the EGGW measure taken at the 25% height location on the electroglottograph waveform. The upper portion of the waveform corresponds to maximum glottal closure (or maximum glottal contact area), and the lowest portion of the waveform to maximum glottal opening (or minimum glottal contact area).

of phonation. The type of electroglottograph used throughout this study was the Research Laryngograph produced by Dale Teaney (1987).

Comparison Measures

Titze (1984) suggested the Abduction Quotient, Qa, a ratio of the width of the glottis essentially between the vocal processes to twice the amplitude of motion of a vocal fold. Qa is one of a number of measures given by Titze (1984) obtained from a theoretical approach to the mechanics of motion of the vocal folds. The Qa measure was obtained from the software analysis and synthesis program GLIMPES. Qa tends to decrease as vocal quality changes from breathy to normal to pressed (Scherer et al., 1988b). Values of Qa above approximately 0.5 are associated with hypoadduction, and below -1.0 with hyperadduction (Scherer et al., 1988b).

The derivative of the EGG signal may give prominent positive and negative peaks that can be used to approximate the glottal open quotient (e.g., Childers et al., 1990). As Figure 2 helps to illustrate, the positive and negative peaks of the EGG derivative refer to locations near glottal closure and opening, respectively, during which the EGG signal changes (increases and decreases, respectively) the fastest. The distance B of Figure 2 divided by A+B is a quotient, designated Qodegg, and is an approximation to the glottal open quotient.

If the glottis is viewed with stroboscopy and recorded onto video tape, the glottal open quotient can be estimated by noting the number of frames the glottis appears to be partly to most fully open to the total number of frames for the entire glottal phonatory cycle, giving Qostrb.

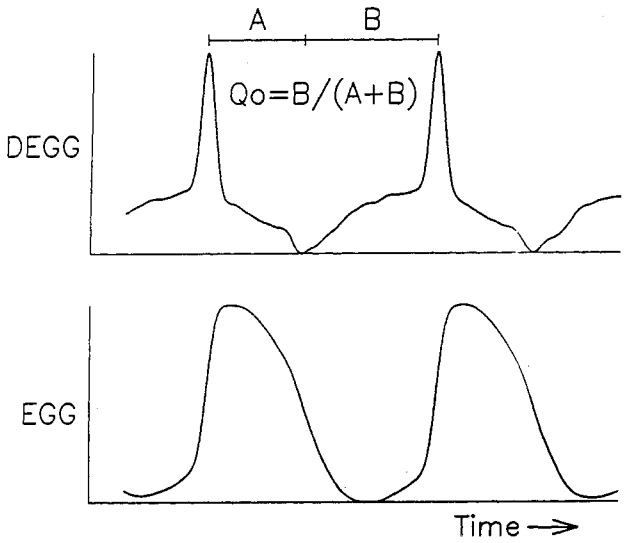
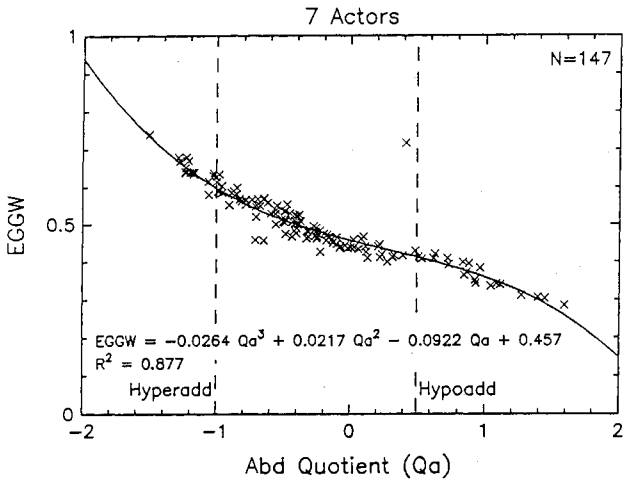


Figure 2. Definition of the glottal open quotient obtained by using the differentiated EGG waveform. The upper trace is thedifferentiation of the lower EGG waveform. Markers on the upper trace are taken at the maximum and minimum values of the differentiated EGG waveform.

Relationship Between EGGW and Qa

Seven normal adult community actors (4 males and 3 females, age range of 23-35 years, and no reported history of nontransient vocal problems), were asked to produce three prolonged /a/ vowels in a steady manner at comfortable pitch and loudness levels (and equal effort levels) for each of the vocal qualities breathy, normal and pressed (or constricted). The middle of each EGG recording was digitized for one second by a 16 (effective 15) bit analog to digital system (Digital Sound Corporation 200 Audio Data Conversion System) at 20,000 samples per second, and stored in a VAX 11/750 computer. Analyses were performed on three consecutive cycles near the



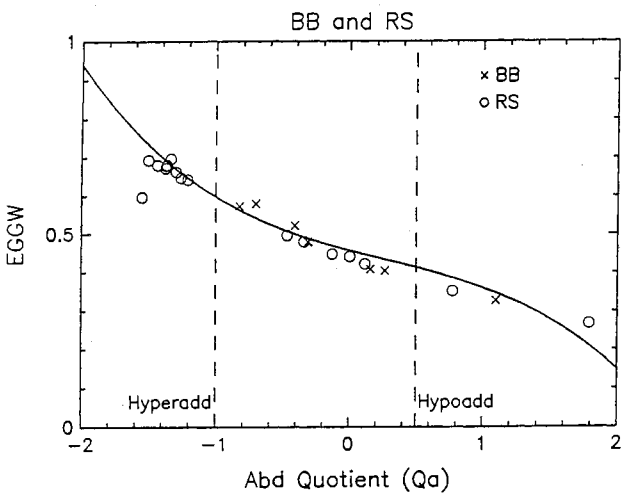


Figure 3 (top). Relationship between the abduction quotient (Titze, 1984) and the EGGW measure taken at the 25% height of the EGG waveform for 7 community actors with normal voices. The subjects prolonged the vowel /a/ over a wide range of intended voice qualities from very breathy to very pressed or constricted. The vertical dashed lines mark expected regions of hyperadduction and hypoadduction determined from Scherer et al., 1988b. Figure 4 (bottom). Relationship between the abduction quotient and the EGGW measure for two normal male subjects, BB and RS, as a replication of the study shown in Fig. 3.

beginning, middle and end of each utterance. About 25% of the (Qa only) data were discarded due to the inability of GLIMPES to successfully run (Scherer et al., 1988b; also ref. Scherer & Titze, 1987).

Figure 3 shows the data comparing EGGW and Qa fit to a cubic equation with an R^2 of 0.877. This suggests a reasonably strong relationship between these two variables. On a retest using a professional tenor (BB) and a non-professional bass-baritone (RS), again using comfortable pitch for sustained /a/ but over a wide range of adduction intentions, the relationship shown in Figure 3 was supported as shown in Figure 4 (average difference of 11.9%, sd = 18.6%, between the data and the cubic fit of Figure 3).

The vertical dashed lines on Fig. 3 and Fig. 4 represent approximate markers for perceived hyperadduction and hypoadduction (Scherer et al., 1988b). The corresponding range between these conditions (the "normal" range) involves values of EGGW between 0.4 and 0.6. This would suggest that for normal larynxes, values of EGGW below 0.4 may correspond to the perceptual label of hypoadduction and above 0.6 to hyperadduction.

Relationship Among EGGW25, EGGW50, EGGW75 and Qa

The measure EGGW discussed above (and throughout this report) was taken at the 25% amplitude location from the baseline of the EGG waveform. The measure can be taken at any reasonable height location, however. Using the data associated with Figure 3, Table 1 shows the correlations among EGGW taken at the 25, 50 and 75% height levels, and Qa. The correlations are reasonably high (minimum = 0.711). EGGW at the 50% level may be the measure of choice when there appears to be too much noise or waveform distortion on the lower portions (open glottis region) of the EGG waveform. A reasonable correspondence between the EGGW50 and EGGW25 values is $y = 1.067x + 0.081$, where x is the EGGW50 value and y is the associated predicted EGGW25 value. The highest correlation is between the measures EGGW50 and EGGW75, suggesting that these measures are essentially redundant. The table also indicates that EGGW25 is the most highly correlated measure with Qa (r = -0.912) of the three EGGW measures.

Table 1.
Pearson product moment correlation values for EGGW and Qa measures. N for Qa correlations is 147, and 189 for the other correlations.

	EGGW50	EGGW75	Qa
EGGW25	0.820	0.711	0.912
EGGW50		0.966	-0.854
EGGW75			-0.715

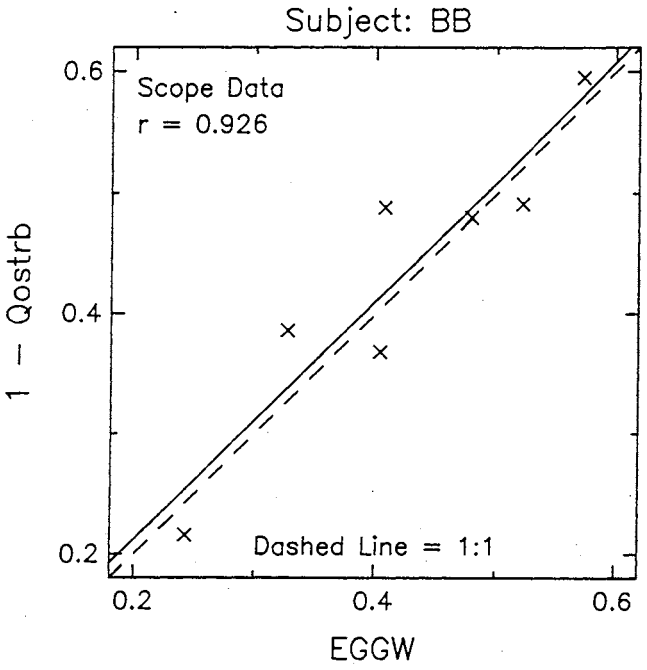


Figure 5. Relationship between the EGGW measure and the glottal open quotient obtained by counting frames of stroboscopic images of the larynx of subject BB.

Relationship Between EGGW and Qostrb

A Wolf stroboscopic system was used with subject BB to determine the open quotient of the larynx by counting the number of frames during which the glottis was partly to fully open, and the number of frames for the entire cycle. EGG was recorded simultaneously. A comfortable pitch and loudness was used. Figure 5 shows the data for corresponding Qostrob and EGGW measures (the quantity 1-Qostrb, the equivalent to a closed quotient, is used in the figure). The figure suggests that, for a wide range of adduction, video frame counting and the EGGW measure not only are strongly related (r=0.93), but the values of 1-Qostrb and EGGW at the 25% level give nearly the same value if the corresponding linear fit line is considered. The relation between EGGW and 1-Qostrb is given by 1 - Qostrb = 0.999EGGW + 0.015 in Fig. 5.

Relationship Between EGGW and Qodegg

EGGW obtained from the non-differentiated EGG signal, and the open quotient from the differentiated EGG signal, were obtained for both subjects BB and RS. The results are shown in Figure 6. The figure indicates that the EGGW measure yielded values larger than did 1-Qodegg, but with a strong relationship (r=0.982 for the two subjects combined). There appears to be a greater difference between values for EGGW and Qodegg as adduction increases. EGGW is related to 1-Qodegg in Fig. 6 by the linear equation 1-Qodegg = 0.902EGGW - 0.0139.

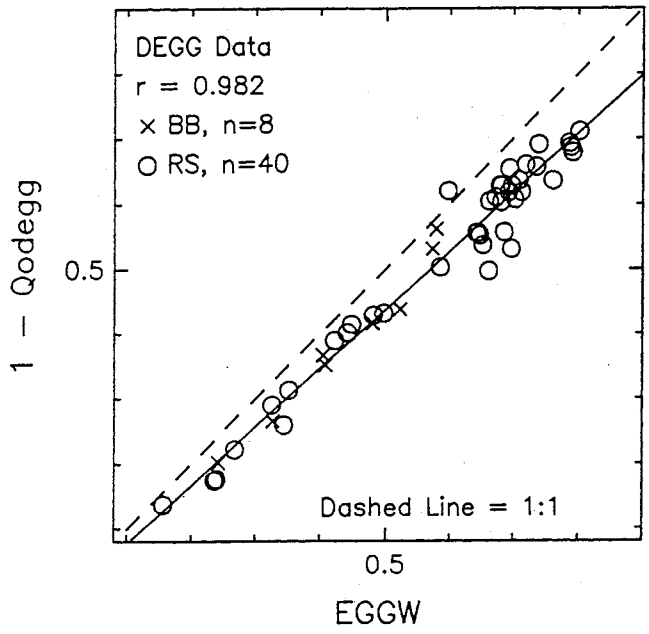


Figure 6. Relationship between the EGGW measure and the glottal open quotient obtained from the differentiated EGG waveform for subjects BB and RS.

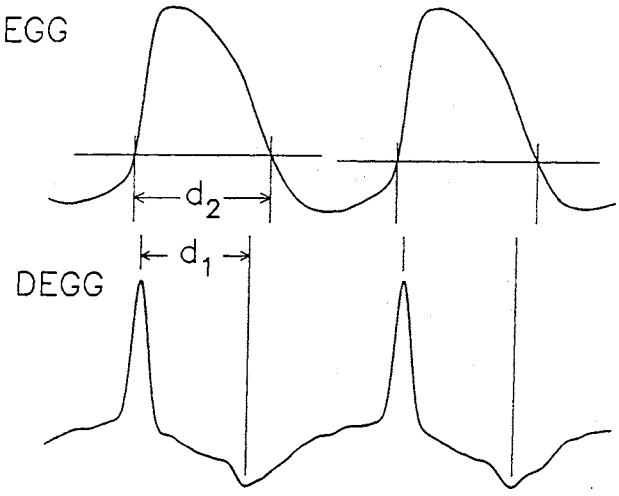


Figure 7. Relationship between the markings for EGGW on the EGG waveform, and the markings for the closed quotient on the derivative of the EGG waveform. The markings on the EGG derivative (giving distance d1) are closer together than are those for the 25% cut for EGGW (giving distance d2), showing the reason for the larger EGGW measures of glottal closed quotient in Fig. 6.

Figure 7 helps to explain the discrepancy between EGGW and Qodegg. The peaks of the derivative of the EGG waveform tend to fall in a narrower range (d1) than the 25% height markings (d2) for the EGGW measure. As adduction decreases, the discrepancy may decrease, as suggested by Fig. 6.

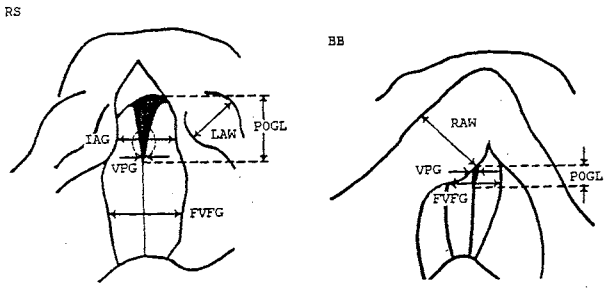


Figure 8. Tracings of laryngeal images for subjects RS and BB indicating the various measures. VPG = vocal process gap; IAG = interarytenoid gap; LAW = oblique width of the left cuneiform cartilage of subject RS; RAW = oblique width of the right cuneiform cartilage of subject BB; POGL = posterior glottal length; FVFG = false vocal fold gap.

Relationship Between EGGW and Glottal Distance Measures

The larynx of subject RS was viewed using a Wolf rigid laryngoscopic system and recorded onto video tape during a wide range of adduction conditions while producing an open hypopharynx vowel at comfortable pitch and loudness. Under visual observation and video recording, a length of cleaned soldering wire with a turned tip was passed through the vocal tract airway. The end was placed on top of the left cuneiform cartilage. This permitted the estimation of the superior oblique width of the left cuneiform cartilage (LAW as shown in Fig. 8), and thus also an approximation of the width of the gap between the vocal processes (VPG). The vocal process gap and the width of the oblique diameter of the left cuneiform cartilage were measured directly on the video monitor. Measurements were made by two people, and the calculated maximum error expected for measurements of the vocal process gap in cm was +/- 13.4%. This error was calculated using measurement variabilities for the VPG monitor measurement, the LAW monitor measurement, the estimate of the actual LAW measure determined from the width of the tip of the soldering iron (including the variability for the solder width measurement), and the height discrepancy between the level of the true vocal folds and the top of the cuneiform cartilage (assumed to be 1 cm). Figure 8 is a tracing and schematic of the glottal measures for subjects RS and BB, respectively.

Figure 9 shows the relationship between EGGW and the vocal process gap VPG for subject RS. The figure strongly suggests a reduction in the space between the vocal processes (greater adduction) as EGGW increases. The data suggest that the vocal processes touch when EGGW is between 0.60 and 0.65. For the nonlinear relationship shown in Figure 9, $VPG\,(cm) = 1.205EGGW^2 - 1.571EGGW + 0.511$, for 0.2 EGGW 0.65. The data suggest, for example, that for a vocal process gap of 0.1 cm, EGGW

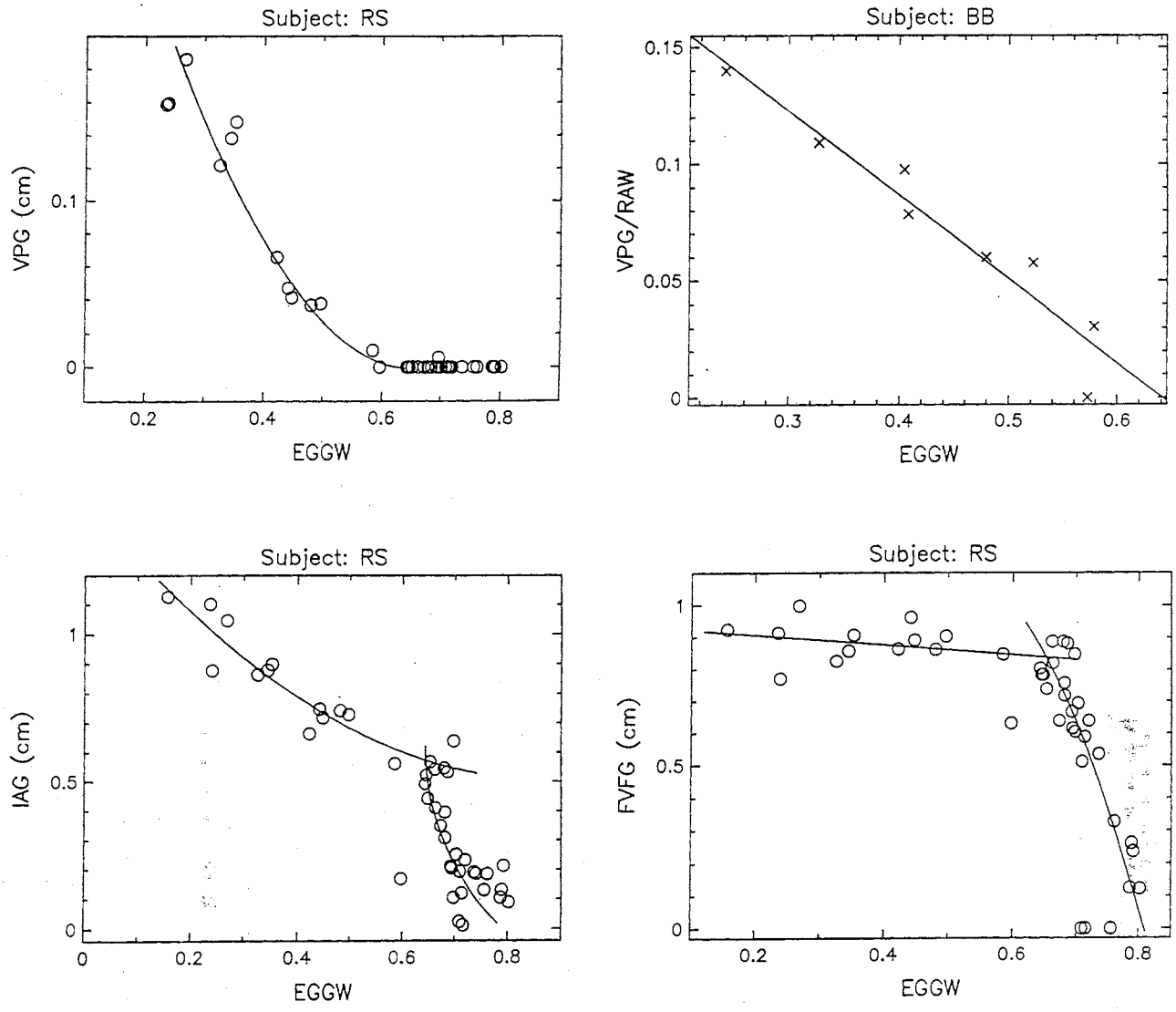


Figure 9 (upper left). Relationship between the vocal process gap VPG and EGGW for subject RS. Figure 10 (upper right). Relationship between VPG/RAW and EGGW for subject BB. Figure 11 (lower left). Relationship between the interarytenoid gap IAG and EGGW for subject RS. Figure 12 (lower right). Relationship between the false vocal fold gap FVFG and EGGW for subject RS.

equals approximately 0.36 for this subject. It is also noted that a value of EGGW = 0.6, the value near which the vocal processes touch, corresponds to the perceptual boundary of hyperadduction as discussed above for Fig. 3.

A similar experiment was performed with subject BB, although without absolute measures of the vocal process gap. The larynx of subject BB was video taped with the Wolf system and laryngeal images seen on the video monitor were copied to a Tektronix 4632 hard copy unit. The VPG measure was made at the visually consistent region where the viewed right cuneiform border intersected the vocal process border (ref. Fig. 8). The value of the VPG was normalized by the oblique diameter of the right cuneiform cartilage (VPG/RAW). Actual gap values were not obtained.

Figure 10 illustrates that the relationship between EGGW and VPG/RAW for subject BB appears linear. Again the data suggest that the vocal processes touch when the EGGW value is between 0.60 and 0.65 (the best fit line suggests 0.64, whereas there is one data point near VPG=0 at approximately 0.575).

After the vocal processes touch, greater adductory forces can approximate the arytenoid cartilages further. For subject RS, the medial boundaries of the arytenoids (cartilagenous glottis) were viewable. The interarytenoidal gap (IAG, ref. Fig. 8) was approximated by measuring the video monitor distance between the bilateral supero-medial arytenoid cartilage eminences, and normalized by the left oblique cuneiform diameter. Figure 11 shows the IAG measure (maximum measurement error of +- 11.2%) versus

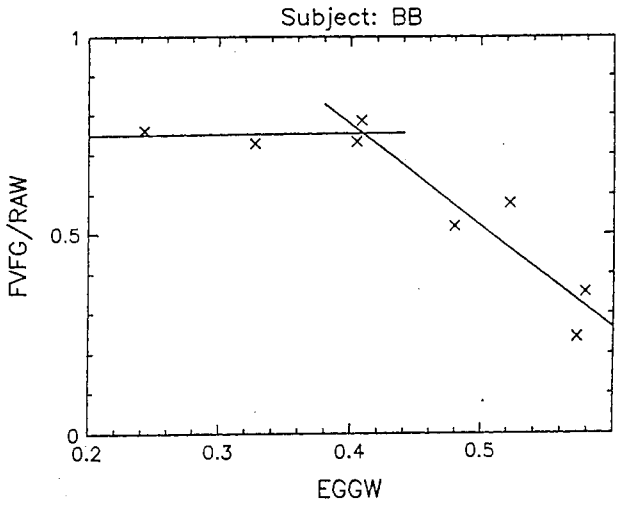


Figure 13. Relationship between FVFG/RAW and EGGW for subject BB.

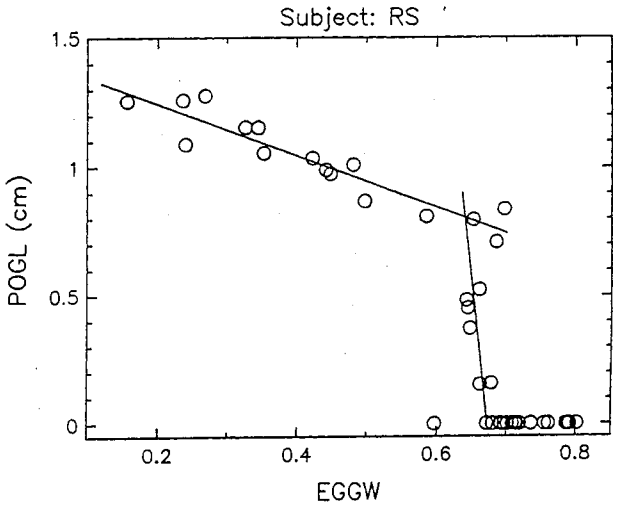


Figure 14. Relationship between the posterior glottal length POGL versus EGGW for subject RS.

the EGGW measure for a wide range of adduction. This figure strongly suggests a change in the relationship near a value of EGGW equal to 0.65, the approximate value corresponding to a vocal process gap of zero. With greater adduction, IAG decreases rapidly as EGGW increases slowly. Near the lowest values of IAG, the scatter of EGGW is relatively high. The data suggest that values of EGGW greater than 0.65 correspond to forceful adduction of the arytenoid cartilages. Smaller IAG values and larger EGGW values greater than 0.65 suggest effective compression at the vocal processes and "closer" vocal folds. The scatter of EGGW data corresponding to values of IAG below about 0.3 cm suggests adjustments of the thyroarytenoid muscles, interarytenoid muscles, and perhaps subglottal pressure resulting in a variety of widths of the closed glottis portion of the EGG waveform. It is noted that the reported distances for the IAG measure may be unique to subject RS due to

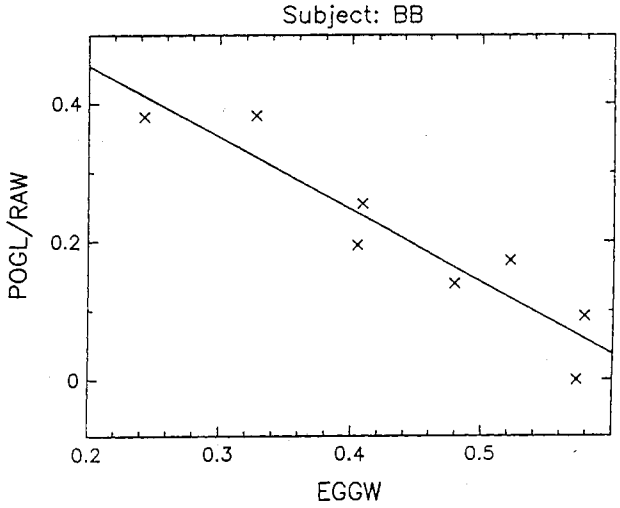


Figure 15. Relationship between POGL/RAW versus EGGW for subject BB.

individual differences of the structure of the arytenoid cartilages and adductory function across individuals.

Data shown in Fig. 11 (and Fig. 6) suggest that the total range of expected EGGW values may be 0.15 to 0.80.

Also examined were the distance changes between the ventricular folds (FVFG, the false vocal fold gap) and the anterior-posterior distance of the viewable cartilaginous glottis (POGL, the posterior glottal length) as defined in Fig. 8. Figure 12 shows the data for FVFG versus EGGW for subject RS (the estimated maximum error for the FVFG measures for RS was +-11.1%). Although there is some scatter of data, there was apparently little change in the distance between the medial edges of the ventricular folds until the EGGW values of adduction reached approximately 0.65, consistent with the IAG measure, beyond which there was a sharp change in the measure corresponding to the inferred hyperadduction. Figure 13 shows the corresponding measure FVFG/RAW for subject BB. Here the data suggest that the distance between the false vocal folds begins to decrease at a value of EGGW of approximately 0.41, a smaller number than for subject RS. A decrease in the FVFG measure may suggest greater contraction of the superior portions of the thyroarytenoid muscle lateral to the ventricular folds.

Data for the posterior glottal length POGL (ref. Fig. 8) for subject RS is given in Figure 14 (estimated maximum error for POGL data was +- 10.8%). POGL values linearly decreased as adduction (EGGW) increased until (once again) about the 0.65 value, beyond which the values dropped sharply. The distance decreased as a result of the bunching of the soft tissue on the posterior wall and greater posterior touch of the medial arytenoid surfaces at and posterior to the vocal processes. The POGL/RAW values versus EGGW for subject BB, Figure 15, show a

similar trend as for RS, that is, a relatively linear relationship of decrease of the posterior glottal length with increasing adduction over the same range of EGGW values.

Relationship to Theoretical Adduction Measures

Under the useful assumption that vocal fold tissue moves in a sinusoidal manner or that glottal area can be modelled by a truncated sinusoid, both Titze (1988) and Rothenberg and Mahshie (1988) describe the abduction quotient Qa and abduction measure D, respectively, with respect to a diagram similar to Figure 16. Tissue movement or glottal area is represented by the sinusoidal waveform, tissue contact by the baseline value zero, the distance of the vocal process of one arytenoid from the midline by W/2, and the amplitude of motion of the vocal fold by A.

Relative to Fig. 16, Titze's abduction quotient is given by $Qa = W/(2A) = -\cos(\pi Qo)$, where $Qo = To/T$, To is the time the glottis is open, and T is the period of the cycle. Rearranging this statement yields $1 - Qo = 1 - (1/\pi)\cos^{-1}(-Qa)$ [Equation 1]. Figure 17 shows this nonlinear relationship between Qa and 1-Qo. Rothenberg and Mahshie (1988) define their abduction measure $D = (1/2)(1 - \cos(\pi Qo))$ which also equals $(1/2)(1+Qa)$. Using the first expression for D with appropriate substitution of Qa leads to Equation 1. A more direct comparison of D and Qa is $1-D = (1/2)(1-Qa)$, which is also graphed in Fig. 17. It is shown that abduction quotient Qa values obtained by applying GLIMPES (Titze, 1984) to EGG recordings from humans (ref. Fig. 3) range beyond the theoretically expected; theoretical values of Qa range from -1 to +1, whereas human data values are permitted to range from about -1.5 to +1.5 as shown in Fig. 3 [later application of Qa by Titze (1990) show wider ranges of Qa than in the 1984 paper]. The form of the theoretical and actual data curves is not dissimilar in shape, however.

An abduction quotient of Qa=0 would imply that the vocal processes just touch. Figures 18a and 18b (see next page) for the data for subjects RS and BB, respectively, suggest that the vocal processes were still separated when GLIMPES gave a value of Qa=0. For subject RS, the vocal process gap was approximately 0.04 to 0.06 cm when Qa=0.

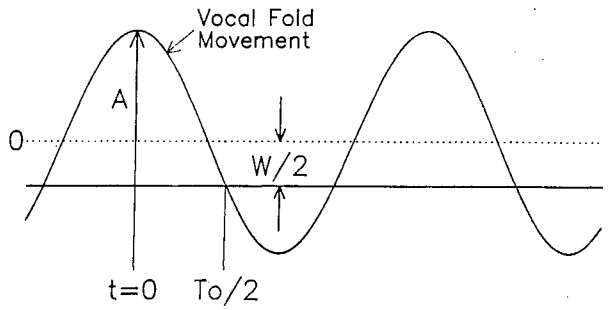


Figure 16. Sinusoidal representation of vocal fold movement. A is the amplitude of motion of one vocal fold. The dotted zero line represents the medial glottal closure location. W/2 represents half of the prephonatory glottal width at the vocal processes. To is the time the glottis is open. The figure is after Titze (1988) and Rothenberg & Mahshie (1988).

Discussion and Conclusions

A reliable and straightforward measure of glottal adduction is required for the clinical and training need of evaluating and establishing adequate phonatory sound within a wide variety of communication requirements, and for the determination of the most efficient glottal configuration from an acoustic and physiological basis (e.g., Titze, 1988; also ref. Scherer, 1991). This study examined the simple glottal adduction measure, EGGW. It is derived from the electroglottographic waveform by taking a ratio of distances (or times) obtained by an intersection line through the signal waveform at the 25% height location.

EGGW was shown to be strongly related (via a cubic equation) to Titze's (1984) abduction quotient Qa (which had been related to visual judgments of adduction in Scherer et al., 1988b). EGGW also was shown to be strongly related to (and nearly equal to) measures of the glottal closed quotient (that is, one minus the value of the glottal open quotient, 1-Qo) using frame counting from stroboscopic views, although data were not extensive. Values of EGGW were greater than 1-Qo obtained by the EGG derivative method, but in a consistent manner. Since the EGG derivative method is troublesome for EGG waveforms without clear derivative peaks, EGGW may be a more reliable method.

The most significant result of this study may be the relationship between EGGW and the actual distance be-

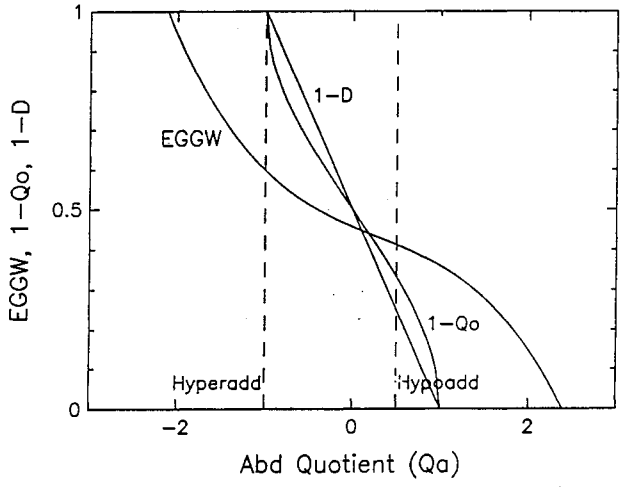


Figure 17. Relationship between human subject data and theoretically derived functions. The function between the abduction quotient Qa and EGGW is empirical (Fig. 3). The curves relating 1-Qo (one minus the open quotient) and 1-D (one minus the abduction measure) with Qa are theoretically derived from Titze (1988) and Rothenberg & Mahshie (1988), respectively.

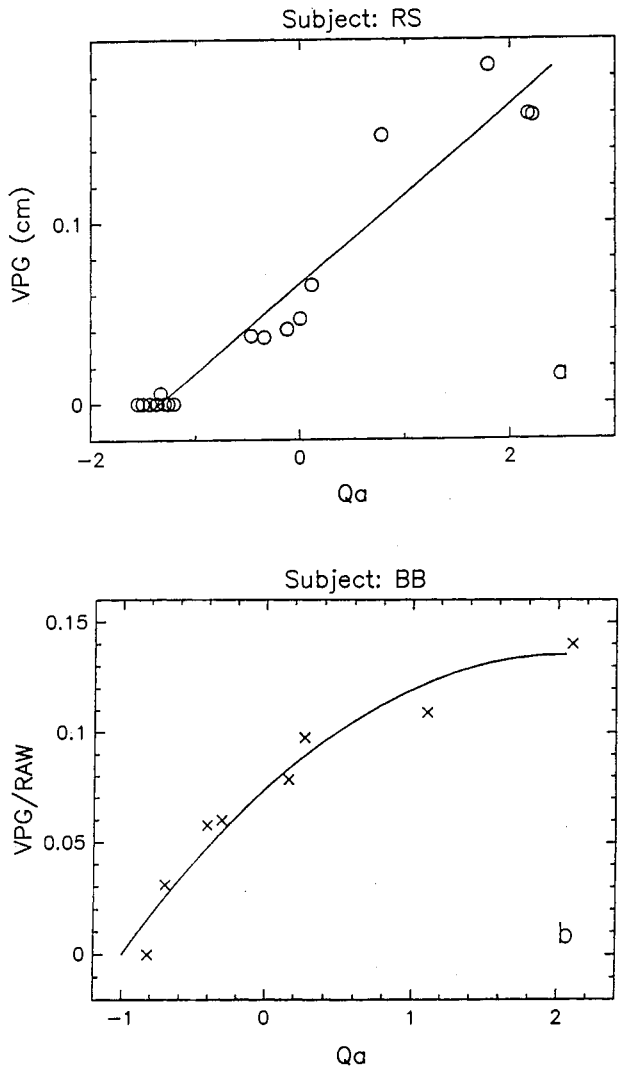


Figure 18a (top). Relationship between VPG and Qa for subject RS; b (bottom). relationship between VPG/RAW and Qa for subject BB.

tween the vocal processes of the arytenoid cartilages. The results suggested that EGGW monotonically increased as the vocal process gap decreased, at least for the comfortable pitch and loudness instruction used for the two normal subjects studied. EGGW reached a value between approximately 0.60 and 0.65 when the vocal processes just touched. EGGW tended to increase as adduction was increased after the vocal processes touched, suggesting additional compression of the vocal processes of the arytenoid cartilages. This study suggests that EGGW eventually may be useful in inferring actual glottal adduction distances in subjects or patients.

Other measures of tissue approximation, such as the interarytenoid gap, the distance between the ventricular folds, and the length of the open posterior glottis, also appear to be viable measures of glottal adduction. The degree of closure of the posterior glottis is important because it may relate to the degree of hyperadduction (as suggested here), dynamic stability of arytenoid movement and interarytenoid pressures (Scherer, Cooper, Alipour-Haghighi & Titze, 1985), and aeroacoustic influence on the glottal volume velociy signal affecting vocal tract excitation (Cranen & Schroeter, 1992). The length of the open posterior glottis is not easily seen in many people because of the "overhang" of the cuneiform and corniculate cartilages. The distance between the ventricular folds may be a relevant measure, especially when it begins to decrease during phonation, suggesting the inclusion of additional muscle forces.

In addition, this study suggests that the range of values for the EGGW measure for normal phonation (neither hypoadducted nor hyperadducted) is between about 0.4 and 0.6. Although this conclusion may be drawn from this study for normal speakers, it may not hold (for example) for classically trained male operatic voices during singing where full glottal closure might be the normal expectation (e.g., Scherer & Titze, 1987; also cf. Howard, Lindsey & Allen, 1990) and phonation would not be labelled as hyperadduction with the connotation of abnormal function.

The primary caution of this study is that EGGW is expected to be a function of vocal fold length (decreasing with greater length as the vertical glottal depth decreases), subglottal air pressure (increasing with greater pressure as larger collision forces and greater contact area are expected; Orlikoff, 1991, demonstrated a significant increase in the EGGW measure with intensity increase; also ref. Kempster, Preston, Mack & Larson, 1987, and Dromey, Stathopoulos & Sapienza, 1992), any vocal fold abnormality (e.g., increasing with edema, decreasing with bowing; ref. e.g., Kitzing, 1990; application to neurological disease problems should be feasible, e.g., Countryman & Ramig, 1993; Ramig, Scherer, Winholtz, Benjamin, Lane & Countryman, 1992), larynx height (if a lowered larynx tends to lengthen the vertical glottis dimension), and vocal tract distortions (in the sense that simultaneous tilting of the head or protrusion of the mandible, as was performed in this study with subjects BB and RS for laryngeal visualization, may place the glottis in an atypical posture). The relationship among EGGW, independent variables of phonation (vocal fold length, subglottal pressure, and arytenoid adduction), oscillatory dependencies on normal biomechanical changes of the vocal fold (e.g., degree of contraction of the vocalis muscle), and vocal fold abnormalities, needs to be mapped out. This study was performed at comfortable pitch and loudness levels only. It is expected that EGGW should be useful as a glottal adduction measure for comfortable ranges of pitch and loudness for a subject over time. Comparison of EGGW values across subjects should be made carefully. Obtaining valid EGG recordings is an obvious prerequisite (e.g., Colton & Conture, 1990; Houben, Buekers & Kingma, 1992).

Acknowledgements

Data analysis and manuscript preparation support has been provided by grant number 1 P60 DC00976 by the National Institute on Deafness and Other Communication Disorders to the The Denver Center for the Performing Arts through The University of Iowa. The authors greatly appreciate the analysis and graphics help given by Chwen-Geng Guo. An earlier version of this paper was presented at the 116th Meeting of the Acoustical Society of America, Honolulu, November, 1988. Thanks are extended to Dr. David Kuehn for a discussion of a part of this paper.

Bibliography

Anastaplo, S., & Karnell, M.P. (1988). Synchronized videostroboscopic and electroglottographic examination of glottal opening. *Journal of the Acoustical Society of America, 83/5,* 1883-90.

Baken, R.J. (1987). *Clinical measurement of speech and voice*. Boston: College-Hill Press.

Baken, R.J. (1992). Electroglottography. *Journal of Voice, 6/2,* 98-110.

Brown, L.R., & Scherer, R.C. (1992). Laryngeal adduction in trillo. *Journal of Voice, 6/1,* 27-35.

Childers, D.G., Alsaka, Y.A., Hicks, D.M., & Moore, G.P. (1987). Vocal fold vibrations: An EGG model. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration,* (pp. 181-202), Boston: Little, Brown and Company, Inc.

Childers, D.G., Alsaka, Y.A., Hicks, D.M., & Moore, G.P. (1986). Vocal fold vibrations in dysphonia: Model vs. measurement. *Journal of Phonetics 14/3/4,* 429-434.

Childers, D.G., Hicks, D.M., Moore, G.P., Eskenazi, L. & Lalwani, A.L. (1990). Electroglottography and vocal fold physiology. *Journal of Speech and Hearing Research, 33,* 245-254.

Childers, D.G., & Krishnamurthy, A.K. (1985). A critical review of electroglottography. *CRC Critical Review Biomedical Engineering, 12,* 131-161.

Colton, R.H., & Conture, E.G. (1990). Problems and pitfalls of electroglottography. *Journal of Voice, 4/1,* 10-24.

Countryman, S., & Ramig, L.O. (in press). Effects of intensive voice therapy on speech deficits associated with bilateral thalamotomy in Parkinson's disease: A case study. *Journal of Medical Speech Pathology.*

Cranen, B. and Schroeter, J. (1992). Modeling a leaky glottis. *Journal of the Acoustical Society of America, 91/4/2,* 2420(A).

Dejonckere, P.H., & Lebacq, J. (1985). Electroglottography and vocal nodules: An attempt to quantify the shape of the signal. *Folia Phoniatrica, 37,* 195-200.

Dromey, C., Stathopoulos, E.T., & Sapienza, C.M. (1992). Glottal airflow and electroglottographic measures of vocal function at multiple intensities. *Journal of Voice, 6/1,* 44-54.

Fourcin, A.J. (1974). Laryngographic examination of vocal fold vibration. In B. Wyke (Ed.), *Ventilatory and phonatory control systems* (pp. 315-333), New York: Oxford University Press.

Gerratt, B.R., Hanson, D.G., & Berke, G.S. (1987). Glottographic measures of laryngeal function in individuals with abnormal motor control. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 521-532). Boston: Little, Brown and Company, Inc.

Higgins, M.B., & Saxman, J.H. (1993). Inverse-filtered airflow and EGG measures for sustained vowels and syllables. *Journal of Voice, 7/1,* 47-53.

Houben, G.B., Buekers, R., & Kingma, H. (1992). Characterization of the electroglottographic waveform: a primary study to investigate vocal fold functioning. *Folia Phoniatrica, 44,* 269-281.

Howard, D.M., Lindsey, G.A., & Allen, B. (1990). Toward the quantification of vocal efficiency. *Journal of Voice, 4/3,* 205-212.

Kempster, G., Preston, J., Mack, R., & Larson, C. (1987). Preliminary investigation relating laryngeal msucle activity to changes in EGG waveforms. In T. Baer, C. Sasaki, and K. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp.339-348). Boston: Little, Brown and Company. Inc.

Kitzing, P. (1990). Clinical applications of electroglottography. *Journal of Voice, 4/3,* 238-249.

Motta, G., Cesari, U., Iengo, M., & Motta Jr., G. (1990). Clinical application of electroglottography. *Folia Phoniatrica, 42,* 111-117.

Orlikoff, R.F. (1991). Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects. *Journal of Speech and Hearing Research, 34*, 1066-1072.

Painter, C. (1988). Electroglottogram waveform types. *Archives of Oto-Rhino-Laryngology, 245*, 116-121.

Ramig, L.O., Scherer, R.C., Winholtz, W., Benjamin, P., Lane, K. and Countryman, S. (1992). Voice treatment and Parkinson's disease: Impact on respiratory, laryngeal and articulatory function. Paper presented to the Conference on Motor Speech Disorders, Boulder, April.

Rothenberg, M., & Mahshie, J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research, 31*, 338-351.

Scherer, R.C. (1991). Physiology of phonation: A review of basic mechanics. In C.N. Ford, & D.M. Bless(Eds.), *Phonosurgery: Assessment and surgical management of voice disorders* (pp. 77-93). New York: Raven Press.

Scherer, R.C., Cooper, D., Alipour-Haghighi, F., & Titze, I.R. (1985). Contact pressure between the vocal processes of an excised bovine larynx. In I. R. Titze & R.C. Scherer (Eds.), *Vocal fold physiology: Biomechanics, acoustics and phonatory control* (pp. 292-303). Denver, Colo.: The Denver Center for the Performing Arts.

Scherer, R.C., Druker, D.G., & Titze, I.R. (1988a). Electroglottography and direct measurement of vocal fold contact area. In O. Fujimura (Ed.), *Vocal physiology: Voice production, mechanisms and functions* (pp. 279-291), New York: Raven Press, Ltd.

Scherer, R.C., Gould, W.J., Titze, I.R., Meyers, A.D., & Sataloff, R.T. (1988b). Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. *Journal of Voice, 2/3*, 230-244.

Scherer, R.C., & Titze, I.R. (1987). The abduction quotient related to vocal quality. *Journal of Voice, 1/3*, 246-251.

Teaney, D.T. (1987). Fabre-Wheatstone electroglottograph - a precision R-F plethysmograph. Proceedings of the Ninth Annual Conference of the IEEE Engineering in Medicine and Biology Society, Boston, November, 13-16.

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

Titze, I.R. (1990). Interpretation of the electroglottographic signal. *Journal of Voice 4/1*, 1-9.

Titze, I.R. (1984). Parameterization of the glottal area, glottal flow, and vocal fold contact area. *Journal of the Acoustical Society of America, 75*, 570-80.

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

NCVS Status and Progress Report - 5
November, 1993, 83-90

Speech Therapy for Patients with Parkinson's Disease

Lorraine Olson Ramig, Ph.D., CCC-SP
Department of Communication Disorders and Speech Science, The University of Colorado-Boulder
and
The Recording and Research Center, The Denver Center for the Performing Arts

Introduction

At least 75% of patients with Parkinson's disease have disordered speech and voice, with every Parkinson's disease patient developing speech and voice disorders as the disease progresses (1, 2). Reduced vocal loudness may be one of the first signs of Parkinson's disease (3) and is a classic speech symptom together with monotone, imprecise articulation, hoarse and breathy voice, vocal tremor and short rushes of speech (3,4). These characteristics have been associated with rigidity, hypokinesia and tremor in the muscles of the speech mechanism (Figure 1). For example, reduced loudness has been related to rigidity in laryngeal musculature and bowed vocal folds (5). Decreased range of tongue, lip and jaw movement due to rigidity has been associated with imprecise articulation (6).

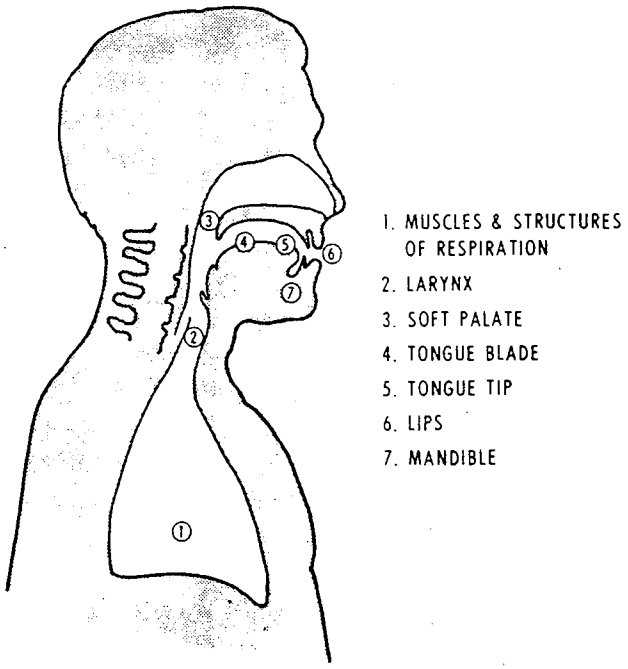


Figure 1. Functional components of the speaking mechanism, showing areas where the airstream may be valved. Adapted from Netsell.

While previous approaches to speech treatment for patients with Parkinson's disease have had limited effectiveness (7, 8, 9), this chapter will present a new method of intensive voice treatment with well-documented short and long-term efficacy (10, 11, 12). The rationale, experimental documentation of efficacy, key treatment elements and considerations for implementation of this approach will be discussed.

Traditional Approaches to Speech Therapy for Patients with Parkinson's Disease

The typical patient with Parkinson's disease who seeks or is referred for speech therapy has a moderate to severe speech disorder and complains of reduced speech intelligibility. He and his family express frustration at not being able to communicate effectively and frequently report that the patient withdraws from conversations and many social situations because people cannot understand him.

The traditional speech treatment approach for patients with Parkinson's disease has involved therapy which is delivered once or twice a week in an outpatient clinic. Typically the focus of treatment has been on improvement of articulatory precision, reduction of rate, and enhancement of intonation (13, 14). Patients, their families and speech clinicians will report some degree of speech improvement during the course of therapy, but carryover or maintenance of the treatment-related changes once therapy is discontinued is generally disappointing. Consequently the ability to communicate deteriorates in many Parkinson's Disease patients as their disease progresses (15, 16). This communication impairment limits the Parkinson's disease patients' full participation in society.

Intensive Voice Treatment for Parkinson's Disease

In 1987, Ramig and Mead developed a treatment program for patients with Parkinson's disease which focused on voice therapy rather than speech therapy. Their approach was shaped by a number of factors: the high incidence of disordered voice in Parkinson's disease (e.g., Logemann (16) reported 89% of 200 Parkinson's disease patients had disordered voice), the apparent role of reduced loudness in reducing patients' communication intelligibility, and reports that intensive speech therapy focusing on phonation has been of value to patients with Parkinson's disease (17, 18). Ramig and Mead designed a treatment program to improve perceptual characteristics of voice and functional communication by targeting the underlying laryngeal pathophysiology associated with the voice disorder (19). This approach is summarized in Table 1. For example, the breathy, weak voices of patients with Parkinson's disease have been associated with glottal incompetence (e.g., bowed vocal folds (5), anterior vocal fold gaps). A primary therapy goal is to increase loudness and decrease breathiness by increasing vocal fold adduction. The monotonous voices of Parkinson's disease patients have been associated with rigidity in the cricothyroid muscles. A second goal of therapy is to improve intonation by increasing cricothyroid muscle activity. The hoarse voices of Parkinson's disease patients have been associated with vocal fold vibratory instability. A third goal of therapy is to improve voice quality by increasing stability of vocal fold vibration.

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 increased phonatory effort 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 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

The specific tasks used in treatment were designed to address these goals through enhanced phonatory effort. This treatment program has come to be known as The Lee Silverman Voice Treatment for Parkinson's Disease (LSVT), named for the Center in which it was developed in Scottsdale, Arizona.

Documentation of Voice Treatment Efficacy

Given the limited efficacy of previous methods of speech therapy for Parkinson's Disease patients, an essential component in the development of The Lee Silverman Voice Treatment for Parkinson's Disease was to objectively quantify pre- to post-treatment improvement as well as long-term maintenance of treatment-related changes. Statistically significant increases on the variables maximum duration of phonation, maximum phonation range, and fundamental frequency in reading have been documented following this intensive voice treatment (10). The magnitudes of these pre- to post-treatment differences were significantly greater than those measured in an untreated control group (20). Three to six month follow-up data support maintenance of these post-treatment increases (12). Corresponding improvements in perceptual aspects of speech production, e.g., intelligibility and loudness, have been reported pre-, post- and follow-up treatment as well (11).

Ongoing efficacy studies are addressing the underlying physiologic bases for improved communication in patients following voice treatment (21). Measurements are being made of rib cage and abdomen kinematics, vocal fold closure, intensity, subglottal air pressure and glottal air flow as well as speech intelligibility before and after intensive voice treatment. This approach to the documentation of efficacy is summarized in Figure 2. Findings to date support statistically significant post-treatment increases in vocal fold adduction (quantified from videolaryngostroboscopy) (22) and sound pressure level (quantified from the acoustic signal) and suggest that increased vocal fold adduction is a key element in treatment success.

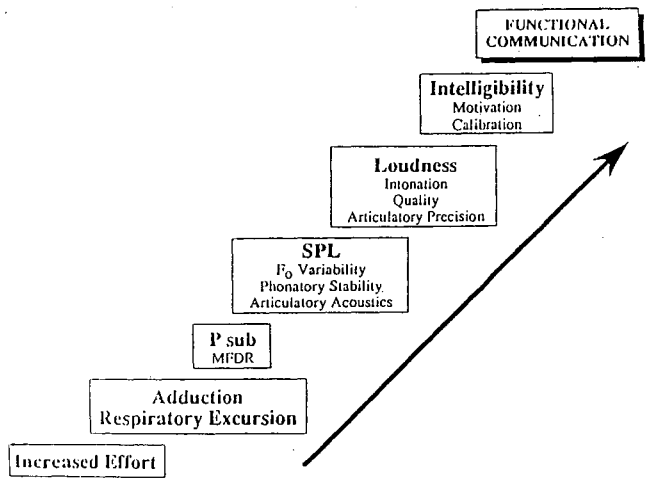


Figure 2. Model for assessment of voice treatment efficacy for patients with Parkinson's disease. Treatment stimulates increased effort. Corresponding measures are made of kinematic (respiratory excursions, vocal fold adduction), aerodynamic (subglottal pressure, maximum flow declination rate), acoustic (sound pressure level, fundamental frequency variability, phonatory stability [jitter, shimmer], articulatory, acoustics [spirantization]), perceptual (loudness, intonation, quality and articulation), intelligibility and functional communication changes following treatment.

Key Elements of Voice Treatment for Parkinson's Disease

The Lee Silverman Voice Treatment for Parkinson's Disease differs from previous approaches to speech therapy in a number of ways. The treatment focuses on voice production, is intensive (four times a week for one month), and it requires that patients be habituated to a high effort level during speech production.

The singular focus of treatment is on increasing vocal effort in order to enhance vocal fold closure and loudness. While Parkinson's disease patients do have disordered articulation and rate, the consistent focus for all sixteen sessions of treatment is vocal effort. It has been observed that even in a mild patient, a consistent vocal effort focus during all sixteen sessions is necessary to reach the habitual use of the louder voice. When the focus remains on increased phonatory effort, vocal loudness is increased and this effort generalizes to enhanced articulation as well. Both vocal and articulatory contributions to enhanced speech intelligibility have been documented following treatment which focuses on vocal effort alone (23).

The singular focus on increased vocal effort makes an immediate impact on vocal loudness and speech intelligibility with a relatively simple task- "speak loud" or "shout". After the first session of treatment, patients are often able to use a louder voice in simple greetings such as, "Hi Honey!" In many cases, this provides the initial positive feedback patients and families need to enhance their motivation and confidence to focus intently on treatment.

The primary goal of treatment is to elicit a louder voice with good quality. This is accomplished through adduction exercises which may include "pushing" exercises (24) and loud phonation. Once the louder voice is established, respiratory and laryngeal coordination at increased loudness levels is practiced. Exercises such as maximum duration vowel phonation targeting duration, constant loudness and steadiness are practiced ten to fifteen times per session with the patient being constantly urged to "go longer, louder and steadier". Fundamental frequency range is another maximum phonatory effort task that is

practiced six to ten times per session. Both of these maximum effort tasks are carried out during each of the sixteen sessions and on-line duration and frequency range data are collected on each patient's performance. These daily clinical data are useful for both patient and clinician reinforcement as well as for documentation of treatment efficacy for insurance reimbursement.

Patients are encouraged to use the same effort level they use during sustained phonation when they speak. This focus on increased phonatory effort offers a simple target for speech production tasks with the focus being "speak loud" or "shout". Parkinson's disease patients have a well-established difficulty in simultaneously executing two different movements (25, 26) and many experience impaired cognitive function (27, 28); 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. In the LSVT treatment approach, the simple focus of increased loudness in speech production is targeted hierarchically from words, phrases, sentences and conversation on a week-by-week basis. The goal for dismissal from treatment is adequate loudness during 85% of spontaneous utterances or targeted task during treatment sessions and reports of 70% carryover outside of treatment. In most cases, this goal can be reached by the fourth week of treatment.

Additional keys to treatment success include: high intensity treatment and calibration. The treatment style is highly energetic and intensive both on the part of the patient as well as the clinician. The clinician serves as an energetic motivator, constantly urging the patient to sustain phonation "longer, louder, higher and lower". Parkinson's disease patients using increased effort and emotion are able to override bradykinesia and improve task performance (29-31). It is speculated that due to phonatory treatment tasks that stimulate an increased effort level, patients are able to override their speech mechanism bradykinesia and improve phonatory and speaking performance. It is therefore essential that patients use this high effort level throughout the majority of each treatment session. Thus, individual, daily treatment is necessary for at least the first sixteen sessions of treatment. This enables the clinician to monitor closely the patient's effort level and continuously motivate the patient to increase and maintain effort levels. If the patient does not achieve these effort levels during treatment, he will surely not achieve them on his own outside of treatment. In order to habituate this high effort level outside of the treatment room, homework is assigned from the first day of treatment.

Habituation and calibration of the patient to the new phonatory effort level is another key component to treatment success. Initially, when using the louder voice, patients complain that they feel like they are "shouting or talking too loud". The clinician should view this as a positive sign because it indicates that the patient is using a higher effort level. The next critical phase is to teach the patient that this effort level is desirable and is not "too loud". This calibration phase is essential for successful treatment. If the patient is not comfortable with his louder voice he will not use it habitually. Figure 3 is frequently used to explain to patients that the level of effort they now must use to produce speech with normal loudness is comparable to the level they used pre-Parkinson's disease when they talked loudly or shouted (32). Other activities that are helpful in this calibration phase include feedback with a tape recorder,

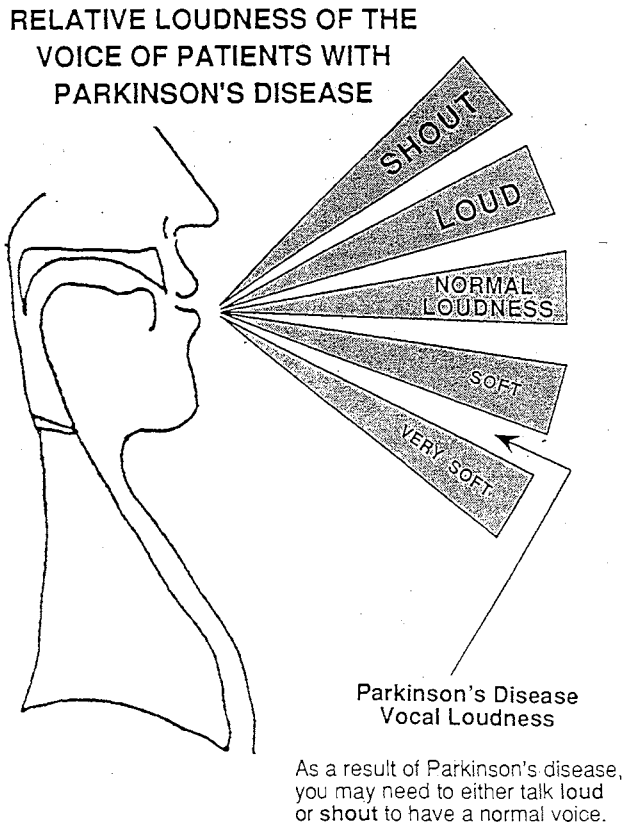


Figure 3. Loudness model used with Parkinson's disease patients to demonstrate level of vocal loudness relative to vocal effort. Adapted from Carolyn Mead Bonitati (1987).

activities in self-monitoring and group therapy (after the initial sixteen sessions). We have found it useful to provide patients with objective feedback of their intensity levels through instruments such as the Voice Light (33).

Critical Considerations

Because the majority of Parkinson's disease patients have disorders of articulation and rate, a speech clinician not trained in the LSVT will be tempted to spend

therapy time treating these disorders. This will diffuse the focus from increased phonatory effort and reduce the magnitude of the treatment effect. In order to achieve habituation of the louder voice, it essential to keep the focus on phonation throughout all sixteen session of treatment. It has been documented that increased phonatory effort generalizes to improved articulatory precision without diffusing treatment focus to articulation (23).

Pushing the Parkinson's disease patient to a higher phonatory effort level may be challenging for the clinician. The style of the treatment is positive, energetic and consistently high effort. The clinician has the role of "infusing the patient with enthusiasm." Given the reduced affect and low-energy style of many Parkinson's disease patients, clinicians may find this to be a challenge. However, it has been reported that even on days when patients report medication problems or fatigue, if the clinician takes a directive, energetic approach, the results are positive. Patients learn that even when they don't feel their best, they can still produce intelligible speech. In order to keep the patient at a high effort level throughout the therapy session, the clinician must closely monitor the patient's output and continuously encourage this increased effort. This may be awkward for the clinician and she may not push the patient consistently. If the patient is not pushed to high phonatory effort levels during 90% of each therapy session, the resulting improvement in speech and voice will not be maximal.

The clinician may be concerned that increased phonatory effort may be abusive to voice production. Recent data document significantly improved glottal competence post-treatment with no significant supraglottal hyperadduction (22). All patients must receive an otolaryngological examination before initiating this treatment to rule out any contraindications and post-treatment laryngeal examinations are useful to document treatment-related improvements in glottal competence.

The clinician cannot underestimate the significance of calibration and habituation of the patient to this new phonatory effort level. The patient may use the louder voice in the treatment room, but if he is not completely comfortable with it, he will not use it in functional communication. Calibration and habituation begin with the first treatment session and continue daily throughout all sessions.

Follow-up Recommendations

Once the patient has demonstrated the ability to speak with increased loudness 85% of the time in spontaneous conversational speech or during tasks in the treatment room, and reports 70% carryover outside of the treatment room, therapy can be terminated. This goal requires no less than sixteen sessions of individual treatment. We recommend to the patient and his family that he continue to practice his voice exercises for 10-15 minutes at least three times a week. Many patients have found a video tape of home exercises useful (34).

Long-Term Maintenance and Follow-up Treatment

Research data has documented a clear maintenance of treatment effects up to six to twelve months without additional treatment. After six months, the maintenance varies depending upon the patient. We recommend that all patients be re-evaluated six months post-treatment. Given the progressive nature of Parkinson's disease, treatment targets may need to be modified. If the voice has deteriorated six to twelve months after treatment, the most common observation is that the patient has "fallen out of calibration" or has forgotten the level of phonatory effort required for adequate loudness. Frequently, two or three voice treatment sessions with encouragement to do homework (at least three maximum "ahs" at the beginning of the day to get calibrated for the day) will get the patient back on track.

Early Voice Treatment

Early in the course of Parkinson's disease, patients may have a monotonous voice which is reduced in volume. If voice treatment is initiated before speech intelligibility is reduced, patients may be able to develop vocal habits which will allow them to maintain communication. Improved and maintained speech intelligibility may enhance and sustain employability for Parkinson's disease patients in the workforce.

Prognostic Factors

A number of factors which increase the likelihood that patients will improve with The Lee Silverman Voice Treatment have been identified. This does not eliminate the possibility of improvement for patients who do not have these characteristics. It may mean that therapy will be more challenging.

Patients with idiopathic Parkinson's disease and the classic hypoadducted voice (bowed vocal folds, anterior glottal gap) respond well to treatment. The efficacy of intensive voice treatment on forms of Parkinsonism as well as Parkinsonian patients with hyperadducted voices is being investigated. Motivated patients who feel speech is important are good candidates for the LSVT. However, many patients who were not motivated at the beginning of treatment become motivated when they learn that they are able to improve their voices. Patients who are active communicators are good candidates; however patients who had withdrawn from communication prior to treatment report that after therapy they talk more because they feel

more confident. Patients who are stimulable to generate a louder voice are also very good candidates. While patients with adequate cognition respond well to treatment, positive post-treatment results have been observed in patients who were mildly or moderately demented (38).

Patients with Atypical Parkinson's Disease

Parkinsonian patients with atypical Parkinson's disease have been treated with the LSVT and improvements have been documented on a case-by-case basis. Pre- to post-treatment results from a patient having had a bilateral thalamotomy (34), documented improvements in phonatory stability but limited long-term carryover. Patients with laryngeal hyperadduction resulting from either physical pathology or secondary compensatory behavior, respond well to treatment when the focus is directed toward increased respiratory effort.

Voice Treatment Drug Treatment Interaction

While neuropharmacological interventions have proven effective in the management of many motor symptoms of Parkinson's disease, the speech and voice problems of these patients are not consistently or significantly alleviated by these interventions (5,36,37). In fact, in some cases, drug-related dyskinesias effect respiratory, laryngeal and oral motor musculature and are severely detrimental to speech production. Since it cannot be assumed that neuropharmological interventions will enhance speech production, it is important that patients receive speech therapy inorder to maximize their communication intelligibility. Research on the interaction between neuropharmacological treatment and The Lee Silverman Voice Treatment is ongoing.

Swallowing

Swallowing disorders (dysphagia) have been identified in all phases of the swallow (oral preparatory, oral, pharyngeal and esophageal) in patients with Parkinson's Disease (39). While there is a tendency for swallowing disorders to increase as the disease progresses, a one to one correlation has not been established. Additionally, it has been documented that patients with Parkinson's disease frequently aspirate but are unaware of their swallowing difficulties or aspiration (40). Therefore, early diagnosis through videofluoroscopy is necessary to document the presence of a swallowing disorder. This evaluation, usually done in collaboration with a speech pathologist and radiologist, provides objective data on the status of the swallowing mechanism. If a disorder exists, the speech pathologist can teach the patient compensatory techniques inorder to prevent aspiration or other swallowing difficulties and monitor the patient changes. Maximum phonatory effort tasks used to increase vocal fold adduction and improve voice production may be useful in reducing aspiration. Clinical reports document improved swallowing and less choking following the LSVT.

Summary

The majority of patients with Parkinson's disease can benefit from speech therapy designed to maintain and increase their vocal loudness. The Lee Silverman Voice Treatment for Parkinson's disease has experimentally documented short- and long-term effectiveness. The ability to communicate plays an important role in the self-concept and well-being of an individual. Therefore speech treatment can play a key role in enhancing the quality of life of a patient with Parkinson's disease.

Acknowledgement

This work was supported in part by grant from NIH R01 DC01150-03, The National Center for Voice and Speech P60 DC00976-03, The Hearst Foundation and The Family of Lee Silverman. The author gratefully acknowledges this support as well as the contributions of Carolyn Mead Bonitati, Stefanie Countryman, Annette Arnone Pawlas and Brian Edwards.

References

1. Oxtoby, M. Parkinson's disease patients and their social needs. London: Parkinson's Disease Society, 1982.

2. Streifler, M. and Hofman, S. Disorders of verbal expression in Parkinsonism. In Hassler, R.G. and Christ, J.F. (Eds.) Advances in Neurology, 1984;40:385-393.

3. Aronson, A.E. Clinical voice disorders. New York: Thieme Stratton, 1985.

4. Darley, F.L., Aronson, A. and Brown, J. Differential diagnostic patterns of dysarthria. Journal Speech Hearing, Research, 1969;12:246-269.

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

6. Weismer, G. Articulatory characteristics of parkinsonian dysarthria: segmental and phrase-level timing, spirantization and glottal-supraglottal coordination. In McNeil, M., Rosenbek, J., Aronson, A. (eds.) The Dysarthrias: physiology, acoustics, perception, management. College Hill Press: San Diego, 1984;101-130.

7. Sarno, M.T. Speech impairment in Parkinson's disease. Arch. Phys. Med. Rehab., 1968;49:269-275.

8. Allan, C.M. Treatment of non-fluent speech resulting from neurological disease: Treatment of dysarthria. Brit. J. D. Comm., 1970;5:3-5.

9. Weiner, W.J. and Singer, C. Parkinson's disease and nonpharma-cologic treatment programs. J. Amer. Geriat. Soc., 1989;37:359-363.

10. Ramig, L., Horii, Y. and Bonitati, C. The efficacy of voice therapy for patients with Parkinson's disease. National Center for Voice and Speech: Status and Progress Report, 1992, Vol. 1, pp. 61-86.

11. Ramig, L. The role of phonation in speech intelligibility: preliminary data from patients with Parkinson's disease. In Kent, R. (ed.) Intelligibility in speech disorders: theory, measurement and management.John Benjamin:Amsterdam, 1992, pp.

12. Ramig, L., Bonitati, C., Lemke, J. and Horii, Y. (in submission) Voice therapy for patients with Parkinson's disease: development of a treatment approach. Journal of Medical Speech Pathology.

13. Hoberman, S.G. Speech techniques in aphasia and Parkinsonism.Journal of the Michigan State Medical Society, 1958;57:1720- 1723.

14. LeDorze, G., Dionne, L., Ryalls, J., Julien, M. and Ouellet, L. The effects of speech and language therapy for a case of dysarthria associated with Parkinson's disease. European Journal of Disorders of Communication, 1992;27:313:324.

15. Morley, D.E. The rehabilitation of adults with dysarthric speech. Journal Speech Hearing Disorders, 1955;20:58-64.

16. Logemann, J.A., Fisher, H.B., Boshes, B. and Blonsky, E.R. Frequency and concurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and Hearing Disorders, 1978;42:47-57.

17. Scott, S. and Caird, F.L. Speech therapy for Parkinson's disease. Journal of Neurology, Neurosurgery and Psychiatry, 1968;46:140-144.

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

19. Ramig, L. and Scherer, R. Speech therapy for neurological disorders of the larynx. In Blitzer, A., Sasaki, C., Fahn, S., Brin, M. and Harris, K. (eds.) Neurological disorders of the larynx. New York City: Thieme Medical Publishers, Inc., 1992, pp. 248-278.

20. King, J., Ramig, L., Lemke, J. and Horii, Y. (in submission) Parkinson's disease: longitudinal changes in acoustic parameters of phonation. Journal of Medical Speech Pathology.

21. Ramig, L. Voice therapy for Parkinson's disease: It's worth the effort. A video tape produced by The National Center for Neurogenic Communication Disorders; University of Arizona-Tucson, 1993.

22. Smith, M., Ramig, L., Dromey, C., Darby, K. and Samandari, R.(in submission) Vocal dysfunction in Parkinson's disease: Videolaryngostroboscopic findings Pre- and Post-Voice treatment. Journal of Voice.

23. Ramig, L.O. The role of phonation in speech intelligibility: A review and preliminary data from patients with Parkinson's disease. Intelligibility in Speech disorders: Theory, Measurement and Management. John Benjamin: Amsterdam, 1992.

24. Froeschels, E., Kastein, S. and Weiss, D.A. A method of therapy for paralytic conditions of the mechanism of phonation, respiration and glutination. Journal of Speech and Hearing Disorders 1955;20:365-370.

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

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

27. Loranger, A. W., Goodell, H., and Lee, J.E. Levadopa treatment of Parkinson's syndrome. Archives of General Psychiatry, 1972;26:163-168.

28. Pollock, M., and Hornabrook, R. N. The prevalence, natural history and dementia of Parkinson's disease. Brain, 1966;89: 429-448.

29. England, A.C. and Schwab, R. S. The management of Parkinson's disease. AMA Archives of Internal Medicine, 1959;104:439-468.

30. Hallet, M., and Khosbin, S. A psychological mechanism of bradykinesia. Brain, 1980;103(2):301-314.

31. McDowell, F. H., Lee, J. E., and Sweet, R. D. Extrapyramidal disease. In A.B. Baker and R.J. Joynt (Eds.) Clinical Neurology, Philadelphia: Harper and Row, 1986;24-26.

32. Mead, Carolyn. Personal communication. 1988.

33. Wintronix, P.O. Box 384, Eastlake, CO 80614.

34. Ramig, L., Bonitati, C., and Winholtz, W. Speech therapy: Parkinson's disease; A video tape produced by The recording and research center of the Denver center for the performing arts.

35. Countryman, S. and Ramig, L. (in submission) The efficacy of voice therapy for a Parkinsonian patient having a bilateral thalamectomy. Journal of Medical Speech Pathology.

36. Leanderson, R., Meyerson, B. A. and Persson, A. Lip muscle function in Parkinsonian dysarthria. Acta Otorylaryngol, 1972;74:271-278.

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

38. Logemann, J.A. Dysphagia in movement disorders. Advances in Neurology, 1988;49:307-316.

39. Mead, C., Ramig, L. and Beck, J. Parkinson's disease with severe dementia: effectiveness of intensive voice therapy. ASHA, 31:10-118.

40. Robbins, J., Logemann, J. and Kirshner, H.S. Swallowing and speech production in Parkinson's disease. Annals of Neurology 1986;19:283-287.

NCVS Status and Progress Report - 5
November 1993, 91-98

Laryngeal Framework Surgery in Children

Marshall E. Smith, M.D.
Department of Pediatric Otolaryngology, The Children's Hospital
Department of Otolaryngology/Head and Neck Surgery, University of Colorado Health Sciences Center
Steven D. Gray, M.D.
Division of Otolaryngology/Head and Neck Surgery, University of Utah

The purpose of this chapter is to review some recent developments in surgical treatment of diseases of the larynx in children. Recently, a collection of surgical techniques known as laryngeal framework surgery has gained attention. Though these techniques have not been widely used in pediatric otolaryngology it is appropriate to examine them in the context of their potential and early attempts at clinical application to this field.

The pediatric otolaryngologist is called on to manage a variety of laryngeal disorders. These may affect one or more of laryngeal functions; swallowing, airway protection, respiration, and phonation. An underlying theme for this review is that surgical treatment of the adult larynx in one context may be applied to the pediatric larynx in another. In the adult, surgical treatment with laryngeal framework surgery is generally targeted to correct phonatory dysfunction. This may not necessarily be the sole focus of treatment in children, where airway and swallowing concerns often supercede voice considerations. For example, treatment of glottal insufficiency with laryngeal framework surgical techniques may well apply to children in addressing swallowing dysfunction while also sparing the voice.

A second theme of this review suggests that the well described techniques of airway reconstruction for laryngotracheal stenosis in children may also be viewed as surgery of the laryngeal framework. As will be discussed, laryngotracheal reconstructive surgery often affects laryngeal biomechanics and the voice, though the procedures were primarily designed for airway restoration.

Some background on the principles of laryngeal framework surgery is appropriate. In 1974 Isshiki coined the term "thyroplasty" and systematized a collection of surgical techniques that alter position, length, and stiffness of the vocal folds and change the voice through procedures on the external larynx.[1] These techniques have also become known as laryngeal framework surgery.[2,3] A concept underlying these procedures is avoidance of surgical trauma to the vocal fold mucosa. The vocal fold mucosa is essential to normal voice production. It consists of a pliable "cover" of epithelium and superfical lamina propria that drapes over a stiff "body" composed of deep lamina propria (vocal ligament) and vocalis muscle.[4] The complex interaction between these structures allows for propagation of a mucosal traveling wave. The periodic interruption of air flow from the lungs by the closing and opening of the glottis due to the traveling mucosal waves is the essence of voice production. Delicate adjustments in stiffness, length, and position of the vocal folds allow the production of a range of vocal pitches, loudness changes, and registers. Laryngeal framework surgery manipulates these laryngeal biomechanical variables to change glottal configuration and mucosal wave propagation to alter the voice.[3]

Though voice alteration has been the original focus of laryngeal framework surgery, this report attempts to broaden this view in pediatric otolaryngology to include treatment of laryngeal airway and swallowing dysfunction, in the context of also restoring or preserving voice. Important background for this discussion includes pediatric laryngeal anatomy. The developmental anatomy of the larynx and vocal folds as it relates to laryngeal framework surgery will first be reviewed, as well as potential implications for what any intervention in childhood may yield for the adult.

Anatomy

The elegant studies of Hirano et al, provide much insight into the development of the phonatory larynx. This section reviews this research, and highlights the differences in pediatric laryngeal anatomy relevant to laryngeal framework surgery. In 1981, at the second Vocal Fold Physiology Conference, Hirano et al presented a study on the growth, development, and aging of the human vocal folds.[5] They studied 88 normal larynges in patients whose ages ranged from a few hours after birth to 69 years old. Several gross anatomic and histologic variables were studied. The length of the entire vocal fold was measured, and the length of the membranous portion (anterior fold) and cartilaginous portion (posterior fold including vocal process and arytenoid).

It was found that, up to the age of 10 years, the length of the vocal fold did not vary much between males and females. At ten years of age the length of the membranous portion of the vocal fold gradually increases for males, up to 20 years of age.

At age 10, the membranous fold is 6 to 8 mm long in males and females. In females the vocal fold will increase in length to 8.5 to 12 mm by age 20. In males the length will more than double, to 14.5 to 18 mm by age 20. The length increase was not dramatic during adolescence, but gradual (Figure 1). The study by Kahane on morphology of the prepubertal and pubertal larynx also documented the changes in vocal fold length with puberty.[6] Measuring the entire vocal fold length (membranous and cartilaginous portions) before and after puberty, the average increase in female vocal folds was 4.2 mm and males was 10.9 mm, over twice as much.

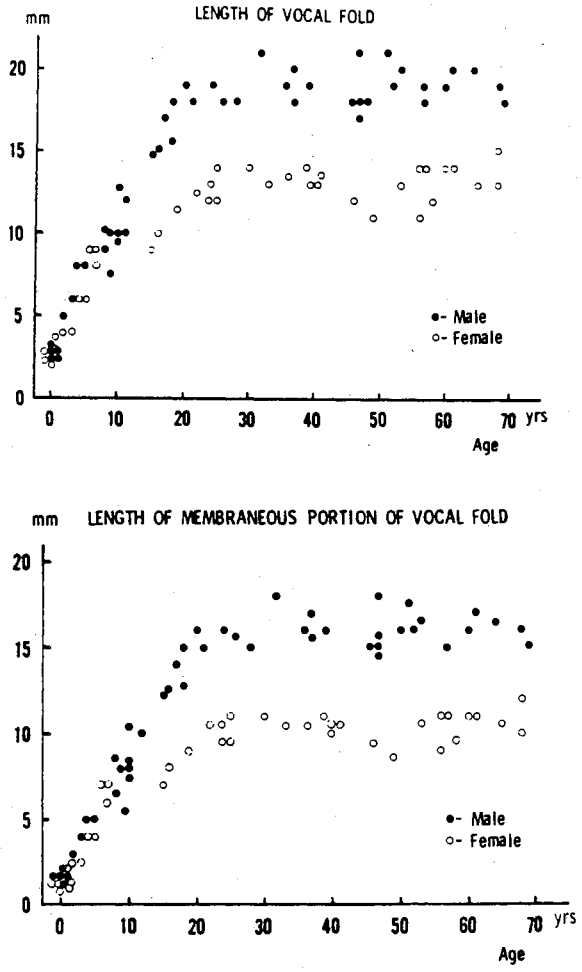


Figure 1. A (top): Length of the entire vocal fold measured in mm for 48 males and 40 females ranging from a few hours to 60 years of age. B (bottom): Length of membranous portion of the vocal fold in mm for 48 males and 40 females ranging from a few hours to 69 years of age. From Hirano M, Kurita S, Nakashima T: Growth, development, and aging of the human vocal folds, in Bless DM, Abbs JH (eds.): Vocal Fold Physiology: Contemporary Research and Clinical Issues, *San Diego, College-Hill Press, 1983, pg 26. Used by permission.*

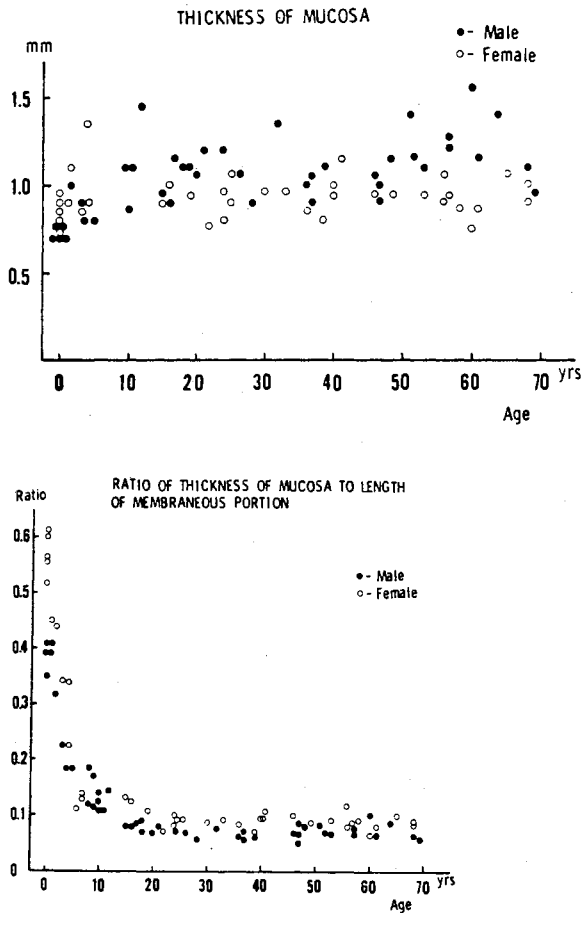


Figure 2. A (top): Thickness of the mucosa at the midpoint of the membranous portion for 48 males and 40 females, ranging from a few hours to 69 years of age. B (bottom): Ratio of thickness of mucosa to length of membranous portion for 48 males and 40 females, ranging from a few hours to 69 years of age. From Hirano M, Kurita S, Nakashima T: Growth, development, and aging of the human vocal folds, in Bless DM, Abbs JH (eds.): Vocal Fold Physiology: Contemporary Research and Clinical Issues, *San Diego, College-Hill Press, 1983, pg 26. Used by permission.*

The cartilaginous portion of the vocal fold also grows with age. It increases from about 1.25 mm in newborns to 3 mm in males and 2.5 mm in females. If a ratio is made of the length of the membranous portion to that of the cartilaginous portion of the vocal fold, the ratio is about 1.5 in newborns (see Figure 2). It increases to about 4.0 in adult females and 5.5 in adult males.[5] This ratio has value in understanding the functions of the larynx in children. In children, a larger portion of the glottis comprises the posterior glottis. This has been termed by Hirano the "respiratory glottis".[7] Indeed, respiratory and protective functions of the larynx play a larger role than phonation in infants and children. The membranous portion of the vocal folds is more susceptible to edema than adults, yet because the membranous folds (the anterior or "phonatory glottis") comprises a smaller percentage of the entire glottal area these obstructive effects are minimized, serving as a relative protection.

Hirano also studied the histology of the vocal folds in the developing larynx. He has reported extensively on the layered structure of the vocal fold.[4] It is described as having five distinct layers. The first two layers, the vocal fold epithelium and superficial layer of the lamina propria (also known as Reinke's space) comprise the "cover". Underneath this is found the intermediate and deep layers of the lamina propria (or vocal ligament) and the thyroarytenoid or vocalis muscle. The deeper layers are called the "body". The complex stiffness interaction between the cover and body facilitates phonation through its range of pitches, loudness, and registers. This layered structure goes through extensive maturational changes.[5] Up to four years of age the intermediate and deep layers of the lamina propria are not differentiated. After four years an immature vocal ligament is observed. There is in childhood a much more extensive density of fibroblasts throughout the lamina propria than in the adult (see Photo 3; center-bound photographic plate). With growth, elastic fibers of the intermediate layer develop and fibroblasts decrease while collagen fibers of the vocal ligament form. By 16 years of age the layered structure of the adult is observed. The high density of fibroblasts in the submucosa of pediatric vocal folds implies that they may be prone to scar formation from surgical trauma.

External Laryngeal Developmental Anatomy

In addition to the development of endolaryngeal structures, a study of the growth and development of the external laryngeal cartilages is helpful to understand the effects of growth on the larynx and surgical interventions to these structures. The most extensive study of this topic was done by Klock. In two reports, he published reports of extensive measurements on the anatomic dimensions of the larynx in infancy and childhood.[8,9] It was found that, in general, the growth of the overall dimensions of the larynx is linearly related (directly proportional) to crown-heel length (somatic height). Laryngeal growth is thus related to age only as overall body growth relates; that is, a sigmoidal curve with acceleration between birth and three years, then deceleration, then rapid growth phase at puberty, especially in males. Kahane also documented the changes in external laryngeal anatomy resulting from puberty.[6] Significant regional growth localized to the anterior aspect of the thyroid cartilage was measured in laryngeal specimens of pubertal males, whereas other external laryngeal measurements showed less dramatic differences between pubertal males and females.

A key principle in laryngeal framework surgery is the relationship between external laryngeal structures and the vocal folds internally. Through numerous anatomic studies Isshiki found that the position of the anterior commissure of the vocal fold is, in general, half way between the thyroid notch and the inferior border of the thyroid lamina.[1,3,10] The vocal fold is in a horizontal plane at this level. Another study by Meiteles et al generally confirmed this relationship.[11] In dissection of 10 female and 8 male cadaver larynges the anterior commissure was in all cases at or slightly above the midpoint between the thyroid notch and inferior thyroid lamina. The posterior end of the vocal fold showed some variation, in 47% of specimens it sloped downward posteriorly at the oblique line (the line on the thyroid ala joining the superior and inferior tubercles). In medialization laryngoplasty surgery, window placement at or below the inferior third of the oblique line was recommended to avoid medialization of the false folds.

Isaacson published a similar study in pediatric larynges.[12] Ten specimens, ages 10 days to 16 years were studied. The relationship of the level of the anterior commissure and the vocal fold relative to the external thyroid cartilage landmarks was consistent throughout childhood, and the same as reported by Isshiki. Thus, though the size of external laryngeal dimensions changes with growth, the relationship to endolaryngeal structures is maintained. It should also be recognized that a variety of asymmetries are present in larynges of all ages. In a recent study, Hirano et al found no directional preponderance in laryngeal asymmetries of ten newborns.[13] In adults, however, several trends were present. With age, the right thyroid lamina tends to tilt laterally, and the left thyroid lamina tends to tilt medially. Also the right cricothyroid joint is located more laterally, posteriorly, and inferiorly than the left joint. The thyroid cartilage is tilted to the right relative to the cricoid cartilage. They noted that all adult specimens were of right handed individuals, and speculated that differences in handedness may be related to the laryngeal asymmetries measured. Also of interest, despite these external asymmetries, they found that the level of the vocal folds remained relatively the same, to maintain symmetric vibration.

The effects on laryngeal growth from laryngeal framework surgery are unknown. However, results extracted from prior work in laryngotracheal reconstruction may be applicable. In general, laryngeal growth appears to maintain, despite surgical intervention. These studies have been recently reviewed by Cotton.[14] Laryngofissure had no effect on laryngeal growth in 5-week-old dogs.[15] Nasal septal cartilage and mucosal autografts were shown to increase the circumference of the subglottic space in young dogs, without affecting later laryngeal growth.[16] Cotton's work in laryngeal anterior and posterior autogenous auricular grafts in rabbits demonstrated cartilage graft viability at both sites, with the posterior cricoid grafts faring better.[14] Growth with viable new cartilage formation would be expected.

Laryngotracheal Reconstruction as Laryngeal Framework Surgery

Surgical techniques for repair of glottic and subglottic stenosis have made many advances.[17] These procedures, generally known as laryngotracheal reconstruction (LTR), have particular applicability to the pediatric population where laryngotracheal stenosis is a prevalent clinical problem. The goal of these procedures has been to restore the airway. However it can be readily appreciated that laryngeal surgery designed to address one aspect of laryngeal function may necessarily affect other functions of the larynx, eg. phonation and swallowing.[18]

Techniques of LTR may alter the position and anatomy of the vocal folds and endolaryngeal structures through external surgical manipulation of their support, the thyroid cartilage and arytenoids. In this way these procedures may be also viewed as laryngeal framework surgery. Examples of this may be seen in the commonly employed techniques of LTR. Cartilage grafts, usually from autogenous rib, are popularly used in LTR.[14,17] These may be placed in the anterior and/or posterior cricoid region. Anterior cartilage grafts alter the laryngeal framework by immobilizing the action of the cricothyroid muscles, which lengthen and tense the vocal folds. Highly placed anterior cartilage grafts may disrupt the anterior commissure and splay the vocal folds apart. Posterior cricoid cartilage grafts widen the posterior commissure. This separates the arytenoids and affects the ability of the vocal folds to approximate at the vocal processes. Posterior cricoid grafts may also impair action of the interarytenoid muscle. These effects on glottal closure may be more pronounced in the pediatric larynx since it has a relatively larger posterior glottis.

The effects of cartilage graft LTR procedures on pediatric voice have been recently studied. Several reports have described voice problems in these patients.[19-24] Smith et al reported on eight patients with voice problems following LTR.[19] The voices were frequently rough, low pitched, or breathy. Two patients exhibited reverse or inhalatory phonation. Fiberoptic laryngoscopy and laryngostroboscopy demonstrated supraglottal phonation in three, glottal incompetence in two, arytenoid fixation in two, anterior commissure blunting or widening in three, vertical asymmetry of the vocal folds in two, and vocal fold scarring (ie. absent mucosal wave) in three. Most patients exhibited more than one abnormal finding. Though the study group was not controlled and was likely representative of LTR patients with more severe voice problems, it is comparable to other reports,[23-25] and is indicative of the consequences for voice that may result from LTR. This can be a significant long term disability for a child.

As Zalzal has pointed out, the voice in children who are treated for laryngotracheal stenosis may be affected by both the underlying disease process and the surgical treatment designed to correct the problem.[23] Stenosis at the level of the free margin of the vocal folds and scar of the superficial lamina propria will inhibit vocal fold vibration and is difficult to reconstruct. It is also apparent that surgery designed to enlarge the laryngeal airway may adversely affect phonation, which requires glottal closure. Trends from recent studies indicate some additional factors that appear increase risk for a poor postoperative voice result in children who undergo LTR. These include the use of posterior cricoid cartilage grafts, combined use of anterior and posterior grafts, the long term placement of endolaryngeal stents, and multiple LTR procedures.[14,18,23,24]

Posterior glottic and subglottic stenosis can be successfully treated with posterior cricoid cartilage grafts.[14,21] This technique can also be used for treatment of impaired vocal fold mobility, such as bilateral vocal fold paralysis.[26] In the series reported by Zalzal, twelve patients were treated for posterior laryngeal stenosis with posterior cricoid cartilage grafts. The patient's voice quality was assessed by a household member who spent the most time with the patient before and after surgery (subjective perceptual assessment).[21] Of the eight patients with preoperative normal voice quality, only two had a postoperative normal voice quality with six patients reported as having hoarse or husky voice quality. In another series reported by Zalzal et al, sixteen patients had voice quality formally assessed.[23] Only four of the nine patients who received posterior grafts had breathiness, yet these four were the only patients with a breathy postoperative voice. Smith et al reported on fifteen patients that underwent "single stage" LTR (no tracheotomy tube employed or removal of the tracheotomy tube at initial surgery).[18] Of the twelve patients who were successfully extubated, at informal voice quality assessment three to six months postoperatively, seven had normal voice, four had moderate dysphonia, and one had severe dysphonia. All five of these patients had both anterior and posterior cricoid cartilage grafts placed. For three of these five, the surgery was a revision procedure.

The use of endolaryngeal stents to secure cartilage grafts in place has been well described. However, these stents appear to injure the voice, especially when used long term. In Cotton's large series of 61 patients that underwent posterior cricoid graft LTR, the duration of stenting was found to be correlated with postoperative voice assessment in that better voice results occurred when the duration of stenting was 12 weeks or less.[14] Maddalozzo and Holinger[27] reported in a series of 20 children that underwent LTR that hoarseness was not an infrequent problem in those that required stenting. In the report of Zalzal et al all sixteen patients had a stent placed, fifteen had aberrant voice quality.[23] The authors failed to find a correlation, however, between stenting duration and postoperative voice quality.

Several animal studies have examined the effects of stent/intubation on the larynx, with implications for voice problems. In a goat animal study of the effect of long-term endolaryngeal stents on the larynx, disruption of laryngeal mucosa and underlying tunica elastica, particularly in the posterior glottis, was observed in preparations that underwent endolaryngeal stent placement for 3 months.[28] Squamous metaplasia of the posterior glottic mucosa was seen, as well as erosion of the vocal process of the arytenoid (see Photo 4 on center-bound photographic plate). Epithelial hyperplasia and fibrous proliferation in the submucosa anterior to the vocal process were observed. Leonard et al studied the effect of 7-day intubation in small dogs.[29] Larynges harvested 5 weeks after extubation exhibited epithelial disruption, hypertrophy of the epithelial layer, and proliferation of subepithelial connective tissue. These changes, although mainly in the posterior glottis, also were observed in the membranous fold anterior to the vocal process. It would be expected that in the infant and pediatric larynx endotracheal tubes and stents would tend to contact more of the membranous folds and anterior glottis. This has potential implications for injury to the membranous folds. Because of the abundant and diffuse distribution of fibroblasts throughout the superficial and deep lamina propria, the membranous folds ("phonatory glottis") of the pediatric larynx may be more susceptible to voice injury from surgical trauma, stents or intubation.

A summary of suggestions for minimizing or preventing phonation problems in laryngeal framework cartilage graft surgery for pediatric laryngotracheal stenosis is given in Table 1.

Pediatric Applications of Laryngeal Framework Surgery

The approach of Isshiki to voice problems through external laryngeal framework surgery and his surgical techniques have been published extensively, including in a previous volume of *Advances*.[2] The most common and frequently reported of Isshiki's procedures is the Type I thyroplasty, also commonly known as medialization laryngoplasty.[30,31] Another laryngeal framework medialization procedure is the arytenoid adduction laryngoplasty.[32-34] The other thyroplasty types described that have been used for vocal pitch change, or spasmodic dysphonia are not generally applicable to children. The medialization laryngoplasty and arytenoid adduction procedures, however, do have promise in the treatment of selected pediatric patients. As compared with adult patients, a different approach is necessary in considering 1) indications for surgery, 2) surgical plan, and 3) anatomic differences in the pediatric larynx that influence surgical technique.

For children who are candidates for medialization laryngoplasty, aspiration may often be the major symptom

Table 1.
Laryngotracheal reconstruction procedures: potential adverse effects on voice and techniques to avoid and minimize them.
(From Smith ME, Mortelliti AJ, Cotton RT, et al: Phonation and swallowing considerations in pediatric laryngotracheal reconstruction. *Ann Otol Rhinol Laryngol* 1992; 101:731-738. Used by permission)

Laryngotracheal Reconstruction Procedures	*Potential Adverse Effects on Phonation*	*Techniques*
Anterior laryngotracheal split and/or graft	Anterior commissure disruption; vocal fold vertical asymmetry; cricothyroid muscle dysfunction; supraglottic collapse	Avoid complete laryngofissure, if possible; avoid graft placement in anterior commissure; perform exact alignment of anterior commissure
Posterior laryngotracheal split and/or graft	Increased glottic gap; impaired arytenoid adduction; arytenoid subluxation	Avoid excessive graft width; use gentle retraction of hemicricoid
Stents	Scarring of vocal fold mucosa; impaired arytenoid mobility	Minimize stenting time; use single-stage laryngotracheal reconstruction, if possible; stent below vocal folds, if possible

rather than voice. Glottal insufficiency may be due to unilateral or bilateral vocal fold paralysis or paresis. Bilateral vocal fold paralysis in adults is usually due to peripheral nerve injury, but may have central etiology in head injured or stroke patients. In infants, this problem usually has a central etiology. There may be associated cortical dysfunction and developmental delay. Because of central neurologic dysfunction, swallowing may frequently be affected and chronic aspiration present. The presence of a tracheotomy to secure the airway will not prevent aspiration, and recurrent pneumonia may result. Vocal fold medialization, by injection or exteral augmentation, has been described as treatment for aspiration by several authors.[35,36] However, the success of various techniques employed has not been systematically investigated. Medialization laryngoplasty provides an option for improving glottal competence to minimize aspiration, while attempting to avoid procedures such as laryngotracheal separation that would render the patient totally aphonic.

Medialization laryngoplasty for the treatment of swallowing and aspiration problems in selected pediatric patients has been employed (Cotton RT, personal communication). A case reported in the literature by Isaacson[12] is illustrative in this regard and will be reviewed in detail. He described a case of a 14 year old patient that was neurologically impaired from a severe closed head injury at age 4. The patient had been decannulated from a tracheotomy at age 9. A unilateral vocal fold paralysis had been treated with a Teflon® injection, yet development of stridor after the injection resulted in tracheotomy replacement. An arytenoidectomy had then been performed, but the patient developed aspiration. The boy remained dependent on gastrostomy tube feedings and took nothing by mouth. After referral to Isaacson's institution, the Children's Hospital of Pittsburgh, an operation was devised which modified the standard technique of medialization laryngoplasty. The window in the thyroid ala was designed so that the depressed cartilage window would fill in the soft tissue and provide bulk to the region of the absent arytenoid. A custom silastic block was created to secure this window. Following surgery the patient's swallowing improved to the point where he could swallow solids and semi-solids, but aspirated with liquids. Six months later the posterior glottal gap was noted to be slightly larger. An additional Teflon® injection to this region of the reconstructed arytenoid improved the patient's swallowing so that liquids could be swallowed without aspiration. Eventually, the patient no longer required the gastrostomy tube, but remained tracheotomy dependent.

Arytenoid adduction laryngoplasty also accomplishes vocal fold medialization, through suture fixation of the muscular process of the arytenoid to adduct the vocal process. While medialization laryngoplasty affects closure of the anterior (membranous) fold, arytenoid adduction is best suited to close the posterior glottis.[32,37] This also accomplishes medialization and lengthening of the vocal fold as well as appropriate alignment for phonation in vocal folds on different vertical levels. Since the pediatric larynx has a larger proportion of cartilaginous (posterior glottis) to membranous fold versus the adult, laryngeal procedures which are designed to close the posterior glottis may have pediatric application. It would appear that the design and effect of an arytenoid adduction procedure is ideally suited for the pediatric larynx. Preliminary experience suggests that it has, as expected, worked well to improve swallowing but that voice improvement has not been as pronounced and dramatic as that seen in adults. This is probably related to the fact that even though the arytenoid is adducted, the larger posterior pediatric glottis (relative to the anterior glottis) still remains partially open during phonation resulting in some persistence of a breathy voice.

Case Report: A 14-year-old female was suffering from recurrent aspiration pneumonias and a near inability to orally feed due to aspiration. She had a left vocal fold paralysis due to a neuroblastoma. The patient had undergone chemotherapy and was now in remission. Due to her recurrent aspiration pneumonias and her very poor oral intake, she was cachectic and emaciated. Although she was over five feet tall, her body weight was 86 pounds. Laryngeal examination showed a wide glottic aperture with a divergent (triangular) glottic configuration. Based on this assessment a medialization thyroplasty would not be sufficient in closing the posterior glottis.[32,33] Therefore, a left arytenoid adduction was performed. Following the procedure, aspiration was resolved, the teenager was able to gain weight, and a gastrostomy tube was avoided. The disappointing part of this case is that postoperatively the voice was perceptually breathy, even though measures of glottal aerodynamics showed considerable improvement. This case illustrates that arytenoid adduction alone may not be adequate enough by itself to correct a breathy voice associated with unilateral vocal fold paralysis in the pediatric patient. It does demonstrate the effectiveness of arytenoid adduction laryngoplasty in treatment of aspiration problems.

Case Report: A 15-year-old with an appropriate body weight and height for age, had been treated for nodular sclerosing Hodgkin's disease. She had failed therapy and then underwent a bone marrow transplant. During the course of therapy, she experienced a left vocal fold paralysis and a very breathy, weak voice. She had previously been quite active in her high school activities and felt that the breathy voice was more disabling than the rest of her disease. After waiting nine months, she underwent an arytenoid adduction laryngoplasty. Clinically, her voice was improved and objective voice aerodynamic measures corroborated this result. Over the following year as she

continued to grow, she experienced more vocal improvement. This case is used to illustrate two things. First, an arytenoid adduction can be used to treat a breathy voice in the older teenager as in adult disorders and second, continued growth through puberty is likely to improve voice results. This is attributed to change in the ratio of posterior glottis to anterior glottis, described by Hirano et al.[5,7] This ratio declines rapidly through the first ten years of life, and also declines further through the second decade. With change in this ratio, more glottal air flow during phonation would be directed through the membranous glottis to increase glottal vibration.

These cases demonstrates the utility of laryngeal framework surgery in being adaptable to treat problems affecting phonation and swallowing in the pediatric patient. Best results have been obtained in the teenage population as opposed to those under twelve years. The long term results of these procedures await further experience. Results will likely be influenced by the natural deterioration of function in children with central neurological problems. A role may be found for such procedures that ameliorate symptoms and improve laryngeal function in these children.

Surgical Technical Points

Medialization laryngoplasty is usually performed under local anesthesia in adults. This has not been our experience in children, and general anesthesia has been used. These children frequently have concomitant airway and swallowing problems that have necessitated the need for a tracheotomy. However, selected patients may be candidates for local anesthesia and light sedation; especially if they are older children who can cooperate for intraoperative voice assessment during positioning of the implant.

The anatomy of the pediatric larynx presents several differences that the surgeon is aware of when performing medialization laryngoplasty or other laryngeal surgery. The cartilage of the larynx is soft and not calcified. The angle between the thyroid ala is wide and the midline less distinctly palpable. The thyroid notch must be carefully identified. The notch may be obscured by an overriding hyoid bone, since in infants and children the larynx has not fully descended in the neck.

For thyroplastic medialization laryngoplasty, anatomical reference points regarding the level of the vocal fold have been described above. The size of the window for medialization and implant placement is proportionally smaller.[30] It is unnecessary to use a burr, saw, or drill in cutting the cartilage window. A Beaver® knife and otologic instruments, such as a canal wall knife and House elevator ("Gimmick"), may work well for these purposes. Because of the soft cartilage care must be taken in securing the implant in place.

Arytenoid adduction in the pediatric population is a more tedious procedure than thyroplastic medialization. The smaller size of the pediatric larynx makes this procedure technically challenging. The muscular process of the arytenoid is not very prominent and it is easy to misplace the suture for the muscular process too superiorly. This results in prolapse or tilting of the arytenoid cartilage anteriorly. To make sure the desired effect of arytenoid adduction is achieved, the larynx must be examined while the arytenoid is being adducted. This is best done by having the vocal folds visualized on the monitor while the surgeon is applying tension to the suture which is adducting the arytenoid. Furthermore, once the suture is tightened and the arytenoid is adducted, it is preferred to not re-intubate the patient. This requires cooperation between the anesthesiologist and the surgeon. The airway can be managed with a mask or with negative pressure ventilation. A very small endotracheal tube may be acceptable. We have not had good success in performing this procedure under local anesthesia in a pediatric patient.

Conclusions

This report has reviewed basic laryngeal investigations relevant to pediatric laryngeal framework surgery. Some clinical data has been reviewed, especially with regard to the influence on voice in children who undergo cartilage graft laryngotracheal reconstruction, which may be regarded as laryngeal framework surgery. Finally, examples of the application of medialization laryngoplasty and arytenoid adduction laryngoplasty in the pediatric age group are presented and issues discussed regarding the potential use of these surgical techniques in children. The definitive role in children of the array of phonosurgical techniques known as laryngeal framework surgery requires further clinical experience.

References

1.Isshiki N, Morita H, Okamura H, Hiramoto M: Thyroplasty as a new phono-surgical technique. *Acta Otolaryngol (Stockh)* 1974; 78:451-457.

2.Isshiki N: Laryngeal framework surgery. *Adv Otolaryngol Head Neck Surg* 1991; 5:37-56.

3.Isshiki N: *Phonosurgery: Theory and practice*. Tokyo, Springer, 1989, pp 1-233.

4.Hirano M, Kakita Y: Cover-body theory of vocal fold vibration, in Daniloff RG (ed): *Speech Science*, San Diego, College-Hill Press, 1985, pp 1-46.

5.Hirano M, Kurita S, Nakashima T: Growth, development, and aging of the human vocal folds, in Bless DM, Abbs JH (eds.): *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, San Diego, College-Hill Press, 1983, pgs. 22-43.

6.Kahane JC: A morphological study of the human prepubertal and pubertal larynx. *Am J Anat* 1978; 151:11-20.
7.Hirano M, Kurita S, Kiyokawa K, et al: Posterior glottis: morphological study in excised human larynges. *Ann Otol Rhinol Laryngol* 1986; 95:576-581.
8.Klock L, Beckwith JB: Appendix: dimensions of the human larynx during infancy and childhood, in Bosma JF (ed.): *Anatomy of the Infant Head*. Baltimore, Johns Hopkins University Press, 1986.
9.Klock LE: The growth of the human larynx from birth to adolescence. thesis, Univ Washington, Seattle, 1968.
10.Isshiki N, Okamura H, Ishikawa T: Thyroplasty type I (lateral compression) for dysphonia due to vocal cord paralysis or atrophy. *Acta Otolaryngol (Stockh)* 1975: 80:456-473.
11.Meiteles LZ, Lin P, Wenk EJ: An anatomic study of the external laryngeal framework with surgical implications. *Otolaryngol Head Neck Surg* 1992; 106:235-240.
12.Isaacson G: Extraluminal arytenoid reconstruction: laryngeal framework surgery applied to a pediatric problem. *Ann Otol Rhinol Laryngol* 1990; 99:616-620.
13.Hirano M, Kurita S, Yukizane K, et al: Asymmetry of the laryngeal framework: a morphologic study of cadaver larynges. *Ann Otol Rhinol Laryngol* 1989; 98:135-140.
14.Cotton RT: The problem of pediatric laryngotracheal stenosis: a clinical and experimental study on the efficacy of autogenous cartilaginous grafts placed between the vertically divided halves of the posterior lamina of the cricoid cartilage. *Laryngoscope* 1991: 101 (suppl 56).
15.Calcaterra TC, McClure R, Ward PH: Effect of laryngofissure on the developing canine larynx. *Ann Otol Rhinol Laryngol* 1974; 83:810-813.
16.Thomas GK, Marsden J: Subglottic enlargement using cartilage-mucosa autograft. *Arch Otolaryngol* 1975; 101:689-692.
17.Cotton RT: The management and prevention of subglottic stenosis in infants and children. *Adv Otolaryngol Head Neck Surg* 1987; 1:241-260.
18.Smith ME, Mortelliti AJ, Cotton RT, Myer CM: Phonation and swallowing considerations in pediatric laryngotracheal reconstruction. *Ann Otol Rhinol Laryngol* 1992; 101:731-738.
19.Smith ME, Marsh JH, Cotton RT, Myer CM: Voice problems after pediatric laryngotracheal reconstruction: videolaryngostroboscopic, acoustic, and perceptual assessment. *Int J Ped Otorhinolaryngol* 1993; 25:173-181.
20.Sell D, MacCurtain F. Speech and language development in children with acquired subglottic stenosis. *J Laryngol Otol* [Suppl] 1988 (suppl 17):35-38.
21.Zalzal GH: Rib cartilage grafts for the treatment of posterior glottic and subglottic stenosis in children. *Ann Otol Rhinol Laryngol* 1988; 97:506-511.
22.Zalzal GH: Use of stents in laryngotracheal reconstruction in children: indications, technical considerations, and complications. *Laryngoscope* 1988; 98:849-854.
23.Zalzal GH, Loomis SR, Derkay CS, et al: Voice quality of decannulated children following laryngeal reconstruction. *Laryngoscope* 1991; 101:425-429.
24.Zalzal GH, Loomis SR, Fischer M: Laryngeal reconstruction in children: assessment of vocal quality. *Arch Otolaryngol Head Neck Surg* 1993; 119:504-507.
25.Kearns DB, Grace AR, Parsons DS, et al: Laryngeal characteristics after reconstruction for subglottic stenosis: a clinical correlation. *J Laryngol Otol* [Suppl] 1988 (suppl 17):39.
26.Gray SD, Dove H, Kelly S: Posterior glottic surgery for impaired vocal fold mobility. Presented at the eighth annual meeting of the American Society of Pediatric Otolaryngology, Los Angeles, CA, April 20-22, 1993.
27.Maddalozzo J, Holinger LD: Laryngotracheal reconstruction for subglottic stenosis in children. *Ann Otol Rhinol Laryngol* 1987; 96:665-669.
28.Smith ME, Gray SD, O'Connor DM, Cotton RT: Endolaryngeal trauma associated with long-term endolaryngeal stents (abstract). *Otolaryngol Head Neck Surg* 1991; 105:206.
29.Leonard R, Senders C, Charpied G: Effects of long-term intubation on vocal fold mucosa in dogs. *J Voice* 1992; 6:86-93.
30.Koufman JA: Laryngoplasty for vocal cord medialization: an alternative to Teflon. *Laryngoscope* 1986; 96:726-731.
31.Maves MD, McCabe BF, Gray S: Phonosurgery: indications and pitfalls. *Ann Otol Rhinol Laryngol* 1989; 98:577-580.
32.Isshiki N, Tanabe M, Sawada M: Arytenoid adduction for unilateral vocal cord paralysis. *Arch Otolaryngol* 1978; 104:555-558.
33.Koufman, Isaacson G: Laryngoplastic phonosurgery. *Otolaryngol Clin North Am* 1991; 24:1151-1177.
34.Netterville JL, Stone RE, Luken ES, et al: Silastic medialization and arytenoid adduction: the Vanderbilt experience, a review of 116 phonosurgical procedures. *Ann Otol Rhinol Laryngol* 1993; 102:413-424.
35.Blitzer A, Krespi MD, Oppenheimer RW, et al: Surgical management of aspiration. *Otolaryngol Clin North Am* 1988; 21:743-750.
36.Eibling DE, Bacon G, Snyderman CH: Surgical management of chronic aspiration. *Adv Otolaryngol Head Neck Surg* 1992; 5:93-113.
37.Woodson GE, Murry T: Glottic configuration after arytenoid adduction. Presented at the annual meeting of the Triologic Society, Los Angeles, CA, April 1993.

Acknowledgement

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