NCVS Status and Progress Report
Volume 4/June 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
John Folkins, Deputy Director
Julie Ostrem, Program Assistant
Julie Lemke, Secretary

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

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.
Barbara Bustillos
Patricia Benjamin, B.S.
Florence Blager, Ph.D.
Diane Bless, Ph.D.
James Brandenburg, M.D.
Larry Brown, D.M.A.
Kate Bunton, B.S.
Geron Coale, M.A.
Stefanie Countryman, M.A.
Heather Dove, M.A.
Kate Emerich, B.M.
Penny Farrell
John Folkins, Ph.D.
Charles Ford, M.D.
Steven Gray, M.D.
Judith Grayhack, Ph.D.
Chwen-Geng Guo, M.S.
Susan Hensley, M.A.

Marilyn Hetzel, Ph.D.
Margaret Hoehn, M.D.
Henry Hoffman, M.D.
Yoshiyuki Horii, Ph.D.
Joel Kahane, Ph.D.
Judith King, Ph.D.
David Kuehn, Ph.D.
Jon Lemke, Ph.D.
Julie Lemke
Erich Luschei, Ph.D.
Brian McCabe, M.D.
Deb Milbauer
Martin Milder, B.S.
Paul Milenkovic, Ph.D.
Kenneth Moll, Ph.D.
Hughlett Morris, Ph.D.
Jerald Moon, Ph.D.
John Nichols, B.A.

Lorraine Olson Ramig, Ph.D.
Julie Ostrem, B.S.
Mark Peters
Donald Robin, Ph.D.
Ronald Scherer, Ph.D.
Elaine Smith, Ph.D.
Marshall Smith, M.D.
Nancy Pearl Solomon, Ph.D.
Molly Strittmater, B.A.
Leatitia Thompson, Ph.D.
Ingo Titze, Ph.D.
Nancy Tye-Murray, Ph.D.
Vern Vail, B.S.
Katherine Verdolini, Ph.D.
William Winholtz, A.A.S.
Darrell Wong, Ph.D.
Raymond Wood, M.D.
Patricia Zebrowski, Ph.D.

Doctoral Students

Julie Barkmeier, M.S.
Todd Brennan, M.S.
David Druker, B.S.
Eileen Finnegan, M.A.

Mark Leddy, M.S.
Bruce Poburka, M.S.
Annie Ramos, M.S.
Eileen Savelkoul, M.S.

Alice Smith, M.A.
Ina Steinecke
Brad Story, B.S.
Kenneth Tom, M.A.

Postdoctoral Fellows

David Berry, Ph.D.

Young Min, M.D.

Visiting Scholars

Hanspeter Herzel, Ph.D., Germany

Meijin Nakayama, M.D., Japan

Tzu-Yu Hsiao, M.D., Taiwan

Advisory Board

Katherine Harris, Ph.D.
Minoru Hirano, M.D.

Clarence Sasaki, M.D.

Johan Sundberg, Ph.D.

Contents

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

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

Forward

It is gratifying to see that the majority of investigators in the NCVS consortium are accomplishing two major goals: (1) to interact more across research sites, and (2) to translate their research findings into the language of the consumer. To mention only two examples of the latter, Dr. Patricia Zebrowski has had multiple opportunities to explain her parent-child interaction research in stuttering to large audiences through the media. Dr. Lorraine Ramig was invited to contribute to the telerounds sponsored by the National Center for Neurogenic Communication Disorders. These opportunities are highly treasured. When one gets a chance to explain research findings the way one would to grandma at home (or to a nationwide audience), a new level of accountability for tax dollars spent is reached. I hope all of us get a turn at this.

Our trainees are a big part of the NCVS action. Several are learning how to take responsibility for first authorship of publications, and all of them are studying with more than one principal investigator. With the Center concept, we can proudly say that we have a trainer/trainee ratio that exceeds 1.0.

As in any dynamic organization, responsibility and personnel changes occur quite often. The departure of John Folkins as Deputy Director and Director of Training is bittersweet--not at all sweet for us. John has given enormous strength and wisdom to the Center and will be sorely missed. We hope that for him the move to Associate Provost is sweet. There is no doubt that The University of Iowa administration is the winner in all of this.

Other changes have occurred. Julie Ostrem and Barbara Bustillos now coordinate Continuing Education and Dissemination, respectively. This has given Dr. Diane Bless and Dr. Ronald Scherer more time to direct two important research components. I again express my deep appreciation to Ron and Diane for their leadership and continuing efforts to bring excellence to all phases of the Center.

Ingo Titze, Director
June 1993

Part I

Research papers submitted for peer review in archival journals

NCVS Status and Progress Report - 4
June 1993, 1-20

Finite Element Simulation of Glottal Flow and Pressure

Chwen-Geng Guo, M.S.
Ronald C. Scherer, Ph.D.
The Recording and Research Center, The Denver Center for the Performing Arts

Abstract

Computational studies of laryngeal aerodynamics should help clarify the relationships among configuration, airflow, surface pressure, and vocal fold movement within the larynx, and the acoustic consequences of the output glottal airflow. The penalty finite element method (Kim, 1988a, 1988b; Kim & Decker, 1989) was adopted to simulate steady air flow and air pressure through the larynx. A total of 133 conditions of different glottal configurations and inflow rates were studied. The computational results were compared to empirical data from earlier experiments. Two cases are reported, (1) constant glottal divergence (42°) but variable diameter, and (2) constant glottal diameter (0.04 cm) but variable glottal angle. For case (1), the average discrepancy for translaryngeal pressure drop between the computational results and empirical data was 6.8% for pressures between 3 and 15 cm H_2O. Flow separation occurred just downstream of the minimal glottal diameter. For case (2), the computational results for translaryngeal pressure drop differed from the empirically-derived equation predictions by an average of 89% for pressure between 3 and 13 cm H_2O. Pressure recovery in the glottis suggested that the optimal glottal diffuser angle was near 10°. Results suggest that the computational method should be sufficient to study glottal aerodynamics (assuming quasi-steady flow).

Introduction

Laryngeal aerodynamics is the study of the interrelationship among air pressures in the larynx, air flows through the larynx, and laryngeal airway geometry (Scherer, 1981, 1992; Scherer et al., 1983a). During phonation these factors change quickly with time and depend upon tissue biomechanics of the vocal folds, prephonatory glottal configuration, lung pressure, and acoustic characteristics of the respiratory tract (Titze, 1988; Rothenberg, 1981, 1983). Of primary importance in phonation is the air pressure distribution along the surface of the vocal folds. This distribution defines the external driving forces that interplay with the vocal fold surface configuration and tissue biomechanics to allow vocal fold oscillation (Titze, 1986; Scherer & Guo, 1990a). In order to adequately study laryngeal function and test theories of phonation, the pressure distributions on the vocal fold surfaces need to be known for the wide range of glottal configurations that are expected in phonation. In addition, the surface pressures are accompanied by and depend upon the airflow distributions within the laryngeal airway.

There have been relatively few studies of pressure distributions within the larynx for well defined laryngeal configurations (Scherer, 1981; Scherer & Titze, 1983; Binh & Gauffin, 1983; Gauffin et al., 1983; Scherer & Guo, 1990b). Static physical models give information on surface pressures, but are difficult to construct with adequate sampling of surface pressures. Computational procedures are needed to predict pressure distributions along all surfaces of the larynx region for any realistic configuration. Although it is conceivable that the flow field could be sampled using techniques such as hot-wire anemometry, computational techniques are also needed to predict the entire flow field within the larynx.

In recent years, computational approaches have been used by several research groups with promising results (Iijima et al., 1988, 1989, 1990, 1992; Liljencrants, 1991; Alipour & Patel, 1991). They demonstrated that computational methods were informative in the study of laryngeal flow. However, the accuracy of the computational methods was not extensively discussed. Because one can not be sure that computational methods produce relevant results, the relationship between computational results and empirical data should be established.

In this study, a penalty finite element method (Kim, 1988a, 1988b; Kim & Decker, 1989) was adopted to analyze two-dimensional, steady glottal flow. Data will be compared to results from earlier physical model studies, and criteria for adequate computational modeling will be discussed. The computational approach used here appears to show satisfactory validity for the aerodynamic study of static flow and pressure over a wide range of realistic laryngeal configurations.

Numerous computational methods can be adopted. Owing to the complexity of laryngeal geometries, not every method can produce results with high computational efficiency and acceptable accuracy. The computational method used in this study is compatible with the irregular boundaries of the vocal folds.

The use of static flow modeling, whether physical or computational, assumes that the aerodynamics of the glottis is essentially quasi-steady. Support for this assumption has come from theoretical (Flanagan, 1958) and empirical (Mongeau et al., 1992) studies, although application of this assumption over the wide range of possible phonation conditions may be challenged. In this study, it is assumed that the resistive characteristics of the glottis are significant.[1]

[1]Physical models using static glottal configurations have been used to derive pressure-flow-geometry equations governing the aerodynamics of the glottis (Wegel, 1930; van den Berg et al., 1957; Ishizaka & Matsudaira, 1972; Scherer, 1981; Scherer et al., 1983a, 1983b; Binh & Gauffin, 1983; Gauffin et al., 1983; Scherer & Titze, 1983; Scherer & Guo, 1990b, 1991). The validity of using *static shapes* in physical models of the larynx, and similarly therefore for computational modeling using static glottal shapes, depends on the quasi-steady flow assumption. This assumption means that the pressure and flow fields within the glottis for any instant of a dynamically changing glottis are essentially the same as those existing for the static glottis of the same configuration and translaryngeal pressure. There are theoretical and empirical studies that support the assumption over a significant range of phonation. Arguments were made by Flanagan (1958, 1972) and Crystal (1966) that the ratio of glottal inertance to glottal resistance was a time constant small compared to the fundamental period for frequencies up to about 400 Hz and for higher frequencies as subglottal pressure became greater. That is, the reactive aspect of glottal impedance was small compared to the resistive aspect,

A total of 133 conditions of different glottal geometries or inflow rates were studied. These cases were divided into two groups. One group had geometries with a constant glottal angle but variable diameters, and the other group had a constant diameter with variable glottal angles. For most cases, there were parallel empirical data for the translaryngeal pressure drop or pressure profile data for comparison so that the accuracy of the computational method could be examined. Other concerns for computational studies, e.g., convergence criteria and number of iterations, are reported so that the efficiency of the computational method can be discussed.

Finite element computational methods for fluid flow require the solution of Navier-Stokes and continuity equations. The next section discusses this and other modeling aspects relative to the penalty finite element method.

Computational Modeling

The Navier-Stokes equations for two-dimensional, laminar, steady, and incompressible flow (reasonable simplifying assumptions for steady flow through the glottis), ignoring gravitational forces, can be given as

$$v_x \frac{\partial v_x}{\partial x} + v_y \frac{\partial v_x}{\partial y} + \frac{1}{\rho}\frac{\partial p}{\partial x} - \frac{\mu}{\rho}\left(\frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_x}{\partial y^2}\right) = 0 \tag{1}$$

$$v_x \frac{\partial v_y}{\partial x} + v_y \frac{\partial v_y}{\partial y} + \frac{1}{\rho}\frac{\partial p}{\partial y} - \frac{\mu}{\rho}\left(\frac{\partial^2 v_y}{\partial x^2} + \frac{\partial^2 v_y}{\partial y^2}\right) = 0 \tag{2}$$

and

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} = 0 \tag{3}$$

supporting the quasi-steady assumption. More recently, Mongeau, Coker and Kubli (1992) constructed a physical model of a dynamic orifice simulating a uniform glottis. They found that up to about150 Hz (the highest frequencies reported), the dynamic flow resistance sufficiently matched the static resistance to support the quasi-steady assumption. This held over the entire cycle except for an initial onset acoustic transient and for flow entrainment at orifice closure. Their experimental arrangement excluded a duct downstream of the orifice. The onset transient might be minimized if the addition of the inductance of a downstream duct were added, and the offset entrainment was found to be small (2% of the flowrate near closure).

The theoretical and experimental support for the quasi-steady assumption of laryngeal flow is insufficient at this time, however, to declare it to be valid for all cases of glottal configuration, translaryngeal pressure, and phonatory frequency. This is especially true for the diverging glottal shapes, where flow separation is likely for sufficiently large divergence angles. Davies, McGowan and Shadle (1993) point out that, in general, flow separation can be sensitive to acceleration effects and therefore the flow may not be quasi-steady when separation exists. The assumption is made in this paper, however, that the resistive aspect of glottal impedance is most likely significant and should be studied independently. Further modifications may entail corrections or recasting of the glottal resistance results.

where v_x, v_y are velocity components in the x and y direction, respectively, ρ is the air density, p is the pressure, and μ is the molecular viscosity of air.

In the penalty method, the continuity equation (3) is expressed as

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} = -\frac{1}{\lambda}p \tag{4}$$

where λ is the penalty number. As the penalty number approaches infinity, the conservation of mass is satisfied. The value of the penalty number can affect the rate of convergence of the solution (Engelman et al., 1982). Based on the work of others, in this study a penalty number of $\mu/\rho \times 10^{10}$ was adopted (Kim, 1988b; Kim & Decker, 1989).[2]

The finite element equations were constructed by using the Galerkin method (Fletcher, 1984). The velocities were interpolated using biquadratic shape functions, and the pressure was interpolated using linear shape functions defined on a triangular element which is contained inside the quadratic element (see Kim & Decker, 1989, for details). This pressure interpolation function can yield rapidly convergent solutions with high accuracy (Kim, 1988a, 1988b; Kim & Decker, 1989). A 9-node velocity and 3-node pressure flow element was used throughout the whole flow domain.

The assembled global equations were solved by a frontal solver (Irons, 1970) and the direct iteration method (Taylor & Hughes, 1980). For each iteration the solutions were updated using an under-relaxation method to ensure a stable iteration (Anderson et al., 1984). The iterations of the assembled global equations were executed until the error norms became smaller than the convergence criterion E. The error norm e_i was defined as:

$$e_i = \underset{j=1,N}{Max} \left(\left| \frac{a_{i,j}^n - a_{i,j}^{n-1}}{a_{i,\max}^n} \right| \right) \tag{5}$$

where the subscript $(i = v_x, v_y, or\ p)$ denotes each component of the flow variables, n is the iteration level, $a_{i,j}$ denotes the nodal value of the i-th flow variable at the j-th node, $a^n_{i,max}$ denotes the maximum value of the i-th flow variable in the n-th iteration level, and N is the total number of nodes. The convergence criterion E is a preset number and should be small enough to guarantee acceptable accuracy. E was not a constant in the computations of this study. That is, occasionally (ref. to Tables 1 and 2) E was slightly increased, and successful convergence was accepted if the pressure and flow results appeared highly reasonable.

[2]The use of kinematic viscosity in the penalty number is somewhat arbitrary, although the penalty number itself needs to be large (personal communication with S.W. Kim).

Table 1

Computational/empirical comparisons for case 1:constant diverging glottal angle of 42° and variable minimal glottal diameters.

Reynolds Number	Convergence Criterion Vel.	Convergence Criterion Pre.	No. of Iteration	Computational Translaryngeal Pressure Coefficient	Empirical Translaryngeal Pressure Coefficient	Percentage Difference %	Computational Dimensional Flow cm^3/s	Computational Dimensional Pre. cm H_2O
Diameter = 0.0614 cm								
^ 100	.0005	.0005	30	1.4596	NA	-	9.46	0.010
^ 200	.0005	.0005	38	1.2689	NA	-	18.92	0.040
^ 300	.0005	.0005	38	1.2070	1.28	5.70	28.38	0.090
^ 400	.0005	.0005	38	1.1752	1.58	*25.62	37.84	0.160
^ 500	.0005	.0005	38	1.1542	1.10	4.93	47.30	0.250
^ 584	.001	.001	38	1.1437	1.11	3.04	55.25	0.342
^ 700	.001	.001	38	1.1207	1.10	1.88	66.22	0.486
^ 800	.001	.001	38	1.1468	1.15	0.28	75.68	0.654
700	.001	.001	38	1.1319	1.10	2.90	66.22	0.491
800	.001	.001	38	1.1242	1.15	2.24	75.68	0.641
834	.001	.001	38	1.1218	1.14	1.60	78.90	0.697
900	.001	.001	39	1.1179	1.13	1.07	85.14	0.812
1000	.001	.001	38	1.1131	1.15	3.21	94.61	1.002
1094	.001	.001	N	1.1098	1.05	5.70	103.50	1.199
1200	.02	.001	72	1.1066	1.10	0.60	113.53	1.443
1300	.04	.003	38	1.1038	1.12	1.45	122.99	1.693
1353	.04	.003	45	1.1032	1.11	0.61	128.00	1.834
1400	.04	.003	56	1.1011	1.11	0.80	132.45	1.964
1500	.04	.003	75	1.0990	1.11	0.99	141.91	2.255
1631	.04	.003	41	1.0966	1.11	1.21	154.30	2.666
1907	.04	.003	36	1.0973	1.12	2.03	180.41	3.644
2162	.04	.003	36	1.0934	1.11	1.50	204.54	4.684
2510	.04	.003	35	1.0879	1.15	5.40	237.46	6.313
#2748	.04	.003	N	-	-	-	-	-
Diameter = 0.0412 cm								
1	.01	.001	20	98.9117	NA	-	0.09	0.000
10	.01	.001	19	9.9460	NA	-	0.93	0.002
50	.01	.001	19	2.2532	NA	-	4.65	0.011
200	.01	.001	18	1.3220	NA	-	18.62	0.086
400	.01	.001	39	1.2041	NA	-	37.24	0.345
612	.02	.002	37	1.1560	1.17	1.20	56.97	0.807
828	.03	.003	35	1.1286	1.18	4.36	77.08	1.477
1092	.03	.003	36	1.1081	1.16	4.47	101.65	2.570
1344	.03	.003	35	1.0943	1.13	3.16	125.11	3.892
1616	.04	.004	34	1.0830	1.16	6.64	150.43	6.090
1834	.04	.004	36	1.0766	1.16	7.19	170.73	7.248
2081	.04	.004	49	1.0699	1.17	8.56	193.72	9.331
Diameter = 0.0202 cm								
1	.01	.001	19	132.8772	NA	-	0.09	0.001
5	.01	.001	18	26.5988	NA	-	0.46	0.006
30	.01	.001	19	4.5434	NA	-	2.75	0.031
80	.01	.001	18	2.0285	NA	-	7.32	0.111
195	.01	.001	18	1.4706	1.58	6.92	17.85	0.329
317	.01	.001	17	1.3334	1.53	12.85	29.01	0.871
629	.01	.001	40	1.1998	1.36	11.78	57.56	3.427
845	.02	.002	36	1.1571	1.29	10.30	77.33	6.186
1047	.03	.003	34	1.1309	1.24	8.80	95.82	9.497
1190	.03	.003	35	1.1169	1.24	9.93	108.90	12.268
1438	.03	.003	35	1.0981	1.21	9.25	131.60	17.914
1647	.03	.003	34	1.0860	1.21	10.25	150.73	23.500
1908	.03	.003	34	1.0740	1.20	10.50	174.61	31.538
2135	.04	.004	33	1.0654	1.20	11.22	195.38	39.488
2362	.04	.004	34	1.0581	1.21	12.55	216.16	48.332
#2497	.04	.004	N	-	-	-	-	-

Average of Percentage differences: 5.186%

1. 'N' in the column 'No. of Iteration' means that the convergent solution could not be obtained under the preset convergent criterion.
2. 'NA' in the column 'Empirical Translaryngeal Pressure Coefficient' means that the empirical data were not available.
3. '*' indicates that the empirical datum may not be accurate. It was excluded when the mean was calculated.
4. '#' denotes the runs for which the computations did not converge.
5. '^' indicates the runs resulting from using a long flow domain for which the Re 700 and 800 cases were not presented in Figure 2A (refer to Footnote 5).

Table 2.

Computational and empirical comparisons for case 2: Constant glottal diameter of 0.04 cm and variable glottal angles. Positive angles refer to a divergent glottal shape, negative angles to a convergent glottal shape

Angle	Reynolds Number	Convergence Criterion Vel.	Convergence Criterion Pre.	No. of Iteration	Computational Translaryngeal Pressure Coefficient	S-G Eq. Prediction Pressure Coefficient	Percentage Difference %	Computational Dimensional Flow: Flow cm^3/s	Computational Dimensional Flow: cmH_2O
40	200	.01	.001	20	1.4124	1.422	0.47	18.6	0.129
	265	.01	.001	18	1.3020	1.305	0.23	24.6	0.209
	355	.01	.001	18	1.2749	1.214	5.02	33.0	0.367
	470	.01	.001	18	1.2251	1.149	6.62	43.7	0.617
	630	.02	.002	37	1.1819	1.097	7.74	58.6	1.070
	830	.02	.002	37	1.1480	1.061	8.20	77.2	1.804
	1130	.03	.002	16	1.1162	1.031	8.26	105.1	3.252
	1500	.03	.003	35	1.0943	1.010	8.35	139.5	5.618
	2000	.05	.005	81	1.0715	0.994	7.80	186.0	9.779
20	200	.01	.001	20	1.3041	1.393	6.38	18.6	0.119
	272	.01	.001	18	1.1864	1.273	6.80	25.3	0.200
	344	.01	.001	18	1.1565	1.203	3.87	32.0	0.312
	488	.01	.001	18	1.0945	1.125	2.71	45.4	0.595
	632	.01	.001	18	1.0642	1.083	1.74	58.8	0.970
	848	.01	.001	17	1.0377	1.047	0.90	78.9	1.703
	1136	.02	.002	17	1.0203	1.020	0.03	105.6	3.004
	1496	.03	.003	35	1.0087	1.000	.87	139.1	5.151
	2000	.06	.006	34	1.0006	0.985	1.58	186.0	9.132
10	200	.01	.001	20	1.4300	1.471	2.79	18.6	0.131
	272	.01	.001	18	1.2736	1.315	3.15	25.3	0.215
	344	.01	.001	18	1.1853	1.224	3.16	32.0	0.320
	488	.01	.001	18	1.1000	1.122	1.96	45.4	0.598
	632	.01	.001	18	1.0255	1.067	3.89	58.8	0.934
	848	.01	.001	17	0.9747	1.020	4.44	78.9	1.599
	1136	.02	.002	17	0.9368	0.984	4.80	105.6	2.758
	1496	.03	.003	16	0.9123	0.959	4.87	139.1	4.658
	2000	.05	.005	34	0.8957	0.939	4.61	186.0	8.174
5	200	.01	.001	20	1.7711	1.742	1.67	18.6	0.162
	272	.01	.001	18	1.5566	1.520	2.41	25.3	0.263
	344	.01	.001	18	1.4301	1.390	2.88	32.0	0.386
	488	.01	.001	18	1.2811	1.246	2.82	45.4	0.696
	632	.01	.001	18	1.1921	1.168	2.06	58.8	1.086
	848	.01	.001	17	1.1080	1.100	0.73	78.9	1.818
	1136	.02	.002	17	1.0390	1.050	1.05	105.6	3.059
	1496	.03	.003	16	0.9846	1.014	2.90	139.1	5.028
	2000	.05	.005	21	0.9369	0.986	4.98	186.0	8.550
0									
	200	.01	.001	20	2.8258	3.663	22.86	18.6	0.258
	272	.01	.001	18	2.4723	2.941	15.94	25.3	0.417
	344	.01	.001	17	2.2632	2.521	10.23	32.0	0.611
	488	.01	.001	17	2.0176	2.052	1.68	45.4	1.096
	632	.01	.001	17	1.8728	1.797	4.22	58.8	1.707
	848	.01	.001	17	1.7354	1.577	10.04	78.9	2.847
	1136	.02	.002	17	1.6238	1.414	14.84	105.6	4.781
	1496	.03	.003	16	1.5328	1.299	18.00	139.1	7.827
	2000	.05	.005	34	1.4527	1.207	20.36	186.0	13.257
-5	200	.01	.001	20	2.0518	2.082	1.45	18.6	0.187
	272	.01	.001	18	1.8390	1.780	3.31	25.3	0.310
	344	.01	.001	18	1.7102	1.605	6.55	32.0	0.462
	488	.01	.001	18	1.5565	1.409	10.47	45.4	0.846
	632	.01	.001	17	1.4650	1.302	12.52	58.8	1.335
	848	.01	.001	17	1.3795	1.211	13.91	78.9	2.263
	1136	.02	.002	17	1.3089	1.142	14.61	105.6	3.854
	1496	.03	.003	16	1.2541	1.094	14.63	139.1	6.404
	2000	.05	.005	21	1.2053	1.056	14.14	186.0	11.000

Table 2. (continued)

Angle	Reynolds Number	Convergence Criterion Vel.	Pre.	No. of Iteration	Computational Translaryngeal Pressure Coefficient	S-G Eq. Prediction Pressure Coefficient	Percentage Difference %	Computational Dimensional Flow Flow cm^3/s	cmH_2O
-10	200	.01	.001	20	1.7900	1.788	0.11	18.6	0.163
	272	.01	.001	18	1.6291	1.565	4.10	25.3	0.275
	344	.01	.001	18	1.5306	1.435	6.66	32.0	0.413
	488	.01	.001	18	1.4126	1.290	9.50	45.4	0.768
	632	.01	.001	17	1.3425	1.211	10.86	58.8	1.223
	848	.01	.001	17	1.2769	1.143	11.71	78.9	2.095
	1136	.02	.002	17	1.2242	1.093	12.00	105.6	3.604
	1496	.03	.003	16	1.1823	1.057	11.85	139.1	6.037
	2000	.05	.005	21	1.1491	1.029	11.67	186.0	10.487
-20	200	.01	.001	20	1.5900	1.611	1.30	18.6	0.145
	272	.01	.001	18	1.4730	1.435	2.65	25.3	0.249
	344	.01	.001	18	1.4012	1.333	5.12	32.0	0.378
	488	.01	.001	18	1.3145	1.219	7.83	45.4	0.714
	632	.01	.001	17	1.2638	1.157	9.23	58.8	1.152
	848	.01	.001	17	1.2162	1.097	10.87	78.9	1.995
	1136	.02	.002	17	1.1761	1.063	10.64	105.6	3.463
	1496	.03	.003	16	1.1472	1.035	10.84	139.1	5.858
	2000	.05	.005	21	1.1229	1.013	10.85	186.0	10.248
-40	200	.01	.001	20	1.4537	1.642	11.47	18.6	0.133
	272	.01	.001	18	1.3645	1.458	6.41	25.3	0.230
	344	.01	.001	18	1.3095	1.351	3.07	32.0	0.353
	488	.01	.001	18	1.2428	1.232	0.88	45.4	0.675
	632	.01	.001	17	1.2035	1.167	3.13	58.8	1.097
	848	.01	.001	17	1.1678	1.111	5.11	78.9	1.916
	1136	.02	.002	17	1.1367	1.070	6.23	105.6	3.347
	1496	.03	.003	16	1.1131	1.040	7.03	139.1	5.684
	2000	.05	.005	21	1.0930	1.017	7.47	186.0	9.975

For more detailed derivations and the complete finite element equations, the reader is referred to the following articles: Taylor & Hughes (1980), Kim (1988a, 1988b), and Kim and Decker (1989).

In the following case studies, two-dimensional geometries were designed as the modified laryngeal airway boundaries to be used within the computational model. The ventricular folds were omitted for simplicity. The glottis was assumed to be symmetric across the centerline so that only half of the glottal airway was considered. Although it is recognized that the flow may not be symmetric in the glottal airway for a diffuser shape, essentially symmetric surface pressure profiles were assumed for the computational simulation.[3] The boundary conditions, as shown in Fig. 1, are listed below:

$$\text{on B1,} \quad v_x = f(y), \quad v_y = 0,$$

$$\text{on B2,} \quad v_x = v_y = 0,$$

$$\text{on B3,} \quad P = 0,$$

$$\text{on B4,} \quad v_y = 0, \quad \frac{\partial v_x}{\partial n} = 0$$

[3]It is understood that incorporating only one-half of the flow domain may exclude the possibility of modeling certain asymmetric flows and wall pressures. This risk was necessary due to computer limitations. Despite this restriction, the computational results satisfactorily match prior empirical data, although it is hypothesized that a complete flow field approach may increase the modeling accuracy.

where n is the unit vector normal to the boundary. The velocity profile of a fully developed Poiseuille flow was applied at the inlet boundary B1.[4]

The Reynolds number for the glottis was defined as:

$$Re = \frac{\rho v D_h}{\mu}$$

where D_h is the hydraulic diameter, defined as 4 times the cross sectional area divided by the perimeter at the minimal glottal diameter cross section, and v is the average velocity at the same location. The non-dimensional pressure coefficient was defined as

$$p^* = \frac{\Delta p}{\frac{1}{2}\rho v^2}$$

where Δp is the dimensional translaryngeal pressure drop.

In this study, a VAXstation 3100 Model 30 with 32 MB and 2.8VUPS was used to execute the computation. The mean CPU time forthe cases listed in Tables 1 and 2 was about 3 hours. The computer time was almost the same as the CPU time.

Case One: Constant Glottal Angle with Variable Minimal Diameters

In this study, a 42° divergent glottis (ref. Fig. 1 showing 21° half-angle) with entrance-exit (axial) length of 0.3 cm, entrance rounding radius of 0.1 cm, and variable minimal diameters of 0.0202, 0.0412, 0.0614 cm was studied. The same dimensions (with vocal fold length of 1.2 cm) and empirical data for simulation comparison were first reported in Scherer (1981) for his Model 3.

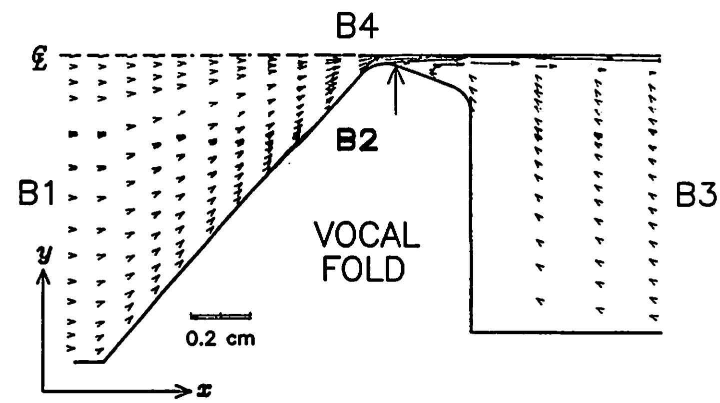


Figure 1. Boundary designations for the computational modeling (see text). B1 corresponds to the inlet, B2 to the vocal fold boundary, B3 to the outlet, and B4 to the midline. Arrow tip direction shows direction of flow, and arrow stem length represents the relative velocity vectors. The condition shown is for steady glottal flow with minimal diameter of 0.0614 cm and Reynolds number of 1000. Dimensions correspond to Model 3 (Scherer & Titze, 1983) in which there was a diverging glottal angle of 42°. The large arrow marks the location of flow separation in the glottis.

[4]A Poiseuille flow distribution was assumed to be reasonable at that location. It was assumed that secondary flows from the subtracheal respiratory system and other non-fully developed flow characteristics would significantly change toward a relatively flat flow profile near glottal entry due to the strong subglottal convergence. Therefore we assumed that the Poiseuille flow distribution would be adequate (as well as convenient and conventional) for the modeling purposes here.

The purpose of this study was to test whether the computational method was capable of simulating glottal pressures that match those of that model, and to examine the flow fields as important concomitants to the pressures.

A relatively short total flow domain of 2.085 cm (ref. Fig. 1) was considered for construction of the computational mesh, where 2 cm was the subglottal lateral diameter.[5] An advantage of a short domain is the savings of computer time. This mesh had 2687 nodes and 630 elements.

There were 52 conditions tested in this first case study, 39 of which corresponded to runs with Model 3 by matching Reynolds numbers. The related information, i.e., the glottal diameters, Reynolds numbers, convergence criteria, number of iterations, computational translaryngeal pressure coefficients, the corresponding empirical data, the percentage differences, and the computational results in dimensional expressions, are listed in Table 1. The maximum Reynolds number for the conditions with convergent solutions was 2510 (two conditions of the 52 did not converge).

Within the computation of each condition, the values of the error norms for the velocities and pressures, observed at the end of each iteration, would approach some constants. The convergence criteria for velocities or pressures were determined by choosing suitable numbers slightly larger than those constants. As the Reynolds number increased, the convergence criteria changed from 0.0005 to 0.04 for velocity, and 0.0005 to 0.004 for pressure. The maximum convergence criterion for pressure of 0.004 was acceptable considering the apparent reasonableness of the results compared to empirical data.

[5]It is generally believed that to obtain more accurate computational results for laminar pipe flow, the downstream flow domain past any cross sectional area change should be as long as is necessary to reestablish the original Poiseuille flow distribution. In line with this convention, we studied the computational model using a relatively long supraglottal tract. This provided a check on the pressure changes within the elongated vocal tract and on the effect on translaryngeal pressure drop.

The glottal airway similar to Model 3 with a relatively long pharynx attached (configured) downstream was used for the construction of the mesh. A total length of 33.4 cm was used, where 32.34 cm was the length of the supraglottal duct and the subglottal lateral diameter was 2 cm. This mesh had 4335 nodes and 1022 elements. A glottal diameter of 0.0614 cm was tested. For low Reynolds number flow, the position of reattachment could be observed clearly at a distance of about 30 cm from the glottal exit. But when Reynolds numbers increased, this observation became impossible because converged solutions of the computational method could not be obtained. This may suggest that as the flow jets out from the glottal exit, it changes from a laminar to a turbulent condition, even when Reynolds number is only approximately 300.

Therefore a shorter flow domain was considered to ensure that the flow condition within the entire flow domain was maintained as laminar flow as assumed by the governing equations. Reasons to support a shorter domain are:

1. In reality, the length of the vocal tract is around 17.5 cm, which is about half the length required for reattachment (30 cm) in a pipe duct. Since the vocal tract is too short to reestablish the Poiseuille flow, the consideration of reattachment is unnecessary for the glottal flow study.

2. According to experimental data with steady flow through static models, the pressure within the pharynx is almost constant. Using a long pharynx to predict negligible pressure differences is therefore not necessary.

3. For a Reynolds number of 800, the translaryngeal pressure coefficient resulting from the longer domain was 1.1468. For the shorter domain used in the first case study, the translaryngeal pressure coefficient was 1.1242. The practical difference between these two values appears insignificant.

Based on the above reasons, the shorter flow domain was used for the entire study. However, if the domain were too short, the computational method would not reflect the downstream flow conditions accurately. The results suggest, however, that the domain length chosen for this study was adequate. The optimal length for the flow domain for a wide variety of realistic laryngeal geometries should be studied further.

Translaryngeal Pressure Drop

The average percentage difference between the empirical data of Model 3 (Scherer, 1981) and the computational results for the three diameter cases combined was 5.19%, and for translaryngeal pressures above 3 cm H_2O was 8.07% (ref. Table 1).

The comparison between the computational and empirical pressure coefficients for the three diameters is shown in Fig. 2. The computational pressure coefficient continually decreases as Reynolds number increases. This is consistent with the laminar flow assumption. For the diameter of 0.0614 cm (Fig. 2A), the computational data fall within the scatter of the empirical data, but for the diameters of 0.0412 cm (Fig. 2B) and 0.0202 cm (Fig. 2C), the computational method predicts slightly lower pressure coefficients compared with the empirical data. The mean percentage difference between the computational results and the empirical data increases as the diameter decreased.

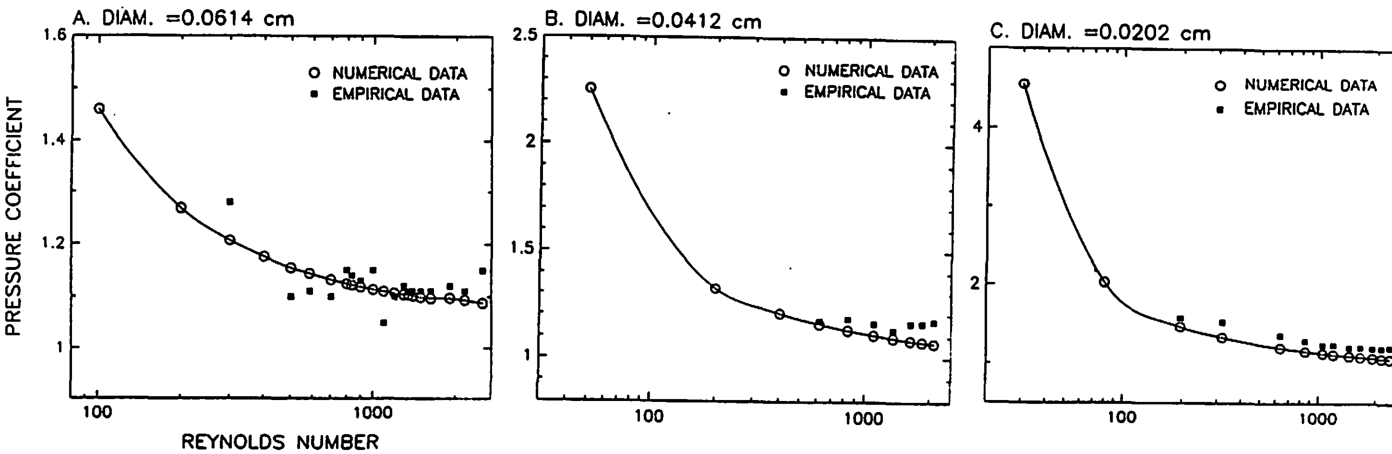


Figure 2. The comparison of computational and empirical (Model 3) translaryngeal pressure coefficients for the diverging glottal angle of 42° and different diameters: (A) 0.0614 cm, (B) 0.0412 cm, (C) 0.0202 cm.

This would be the expected discrepancy between the data for the computational model and the physical model. The physical model was constructed out of polyester resin. Greater viscous flow resistance of the physical model would be consistent with some degree of coarseness of the surfaces, slight nonuniformity of the diameter along the length of the glottis, and the eccentric position of the glottis in the model. The latter refers to the presence of the simulated arytenoid cartilages in the physical model. The cartilaginous glottis was closed so that the open glottal slit was situated anteriorly in the airway, as it is positioned naturally. The effect of eccentricity was ignored in the computational model. Larger discrepancies between the data for the physical model and the computational model as the diameter decreased would seem reasonable due to the closer glottal surfaces and potential increase of relative roughness, nonuniformity, and eccentricity effects.

For this particular diverging glottal configuration, an assumption of 1.1 for the pressure coefficient, suggested by Fant (1982), is consistent with both the computational and empirical data except for the empirical data for the smallest diameter.

Fig. 3 shows a comparison of translaryngeal pressure drop for the computational results, empirical data for Model 3, and two empirical equations: the Ishizaka modified equation (Ishizaka,

1985) and the S-G equation (Scherer & Guo, 1990b). The S-G equation is based on empirical data (using a physical model called Model 5) for 63 different configurations of glottal angle and diameter. Since the S-G equation matches the empirical data of Model 5 with an average percentage difference of 4.45% in the important pressure range of 3 to 50 cm H_2O, it is chosen as an empirical equation for comparison in this study. To avoid confusion, the empirical data for Model 3 are depicted by quadratic equations instead of original data points. The overall percentage differences between the quadratic fitting equations shown in Fig. 3 and the original Model 3 data are 0.16% for the diameter of 0.0202 cm, 0.09% for the diameter of 0.0412 cm, and 0.02% for the diameter of 0.0614 cm.

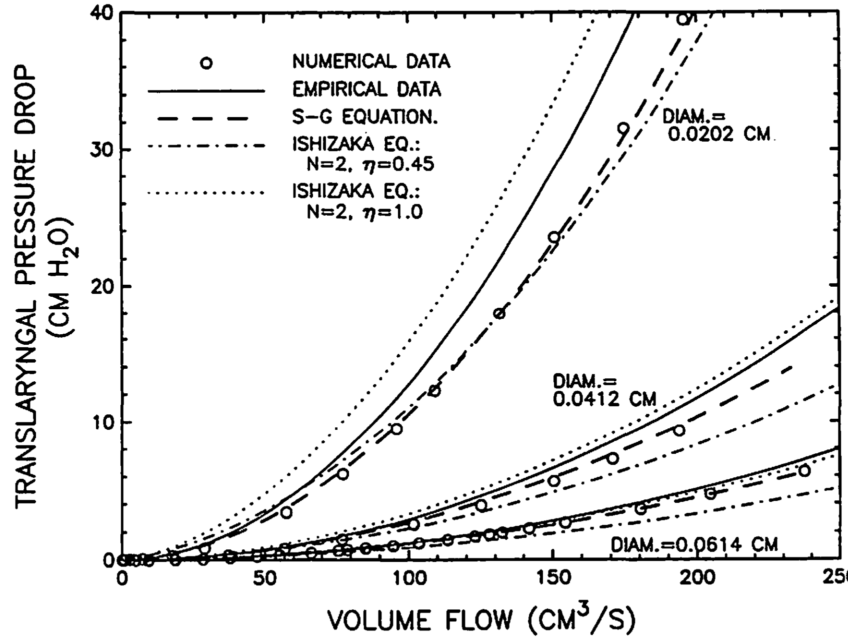


Figure 3. The comparison of translaryngeal pressure drop among the computational results, empirical (Model 3) data, and two empirical equations: the Ishizaka modified equation (Ishizaka, 1985) and the S-G equation (Scherer & Guo, 1990b).

It is shown in Fig. 3 that the computational results match the S-G equation predictions better than the empirical data for Model 3. Since glottal eccentricity was not incorporated in Model 5 and the vocal fold surfaces of Model 5 were constructed more carefully than for Model 3, it is not surprising that the computational method and the S-G equation typically predict lower pressure drop compared with the empirical data. The mean difference between the computational results and Model 3 empirical data for translaryngeal pressure drop between 3 and 15 cm H_2O was approximately 6.8%. However, the corresponding mean difference between the computational results and the predictions according to the S-G equation was 3.0%. This finding constitutes strong support for the validity of the computational method for the relatively large diverging glottal angle studied here.

Ishizaka (1985) supplemented the pressure-flow equations of Ishizaka and Matsudaira (1972) with an η coefficient to account for the losses due to different cross sectional shapes of the glottis. For a glottis divided into 2 vertical (axial) parts, the computational results essentially lie between the lines with η equal to 0.45 and 1 as shown in Fig. 3. As the diameter increases, the computational data better match the Ishizaka equation using a larger η, suggesting a relatively larger pressure loss due to the glottal expansion than for smaller diameters.

Pressure Profiles and Velocity Distributions

The computational velocity vectors for the diameter of 0.0614 cm, the diverging glottal angle of 42°, and a Reynolds number of 1000 are presented in Fig. 1. The large arrow indicates the location of flow separation, which occurs at a distance slightly downstream of the minimal diameter. The point of separation moves upstream as the flow velocity at the minimal diameter increases.

In Fig. 4 the cross sectional velocity profiles within the glottis and the corresponding one-dimensional pressure profile along the vocal fold boundary for the diameter of 0.0412 cm and a Reynolds numbers of 1616 are shown. For this particular condition, the separation (arrow) occurs immediately downstream of (but not coincident with) the minimal diameter. Reverse flow develops downstream of the point of separation. The reverse flow reenters the glottis and eddies are formed. Owing to the lack of empirical data, the validity of the velocity distribution can not be verified. It is common, however, to visualize similar flowpatterns within diffusers (e.g., White, 1979).

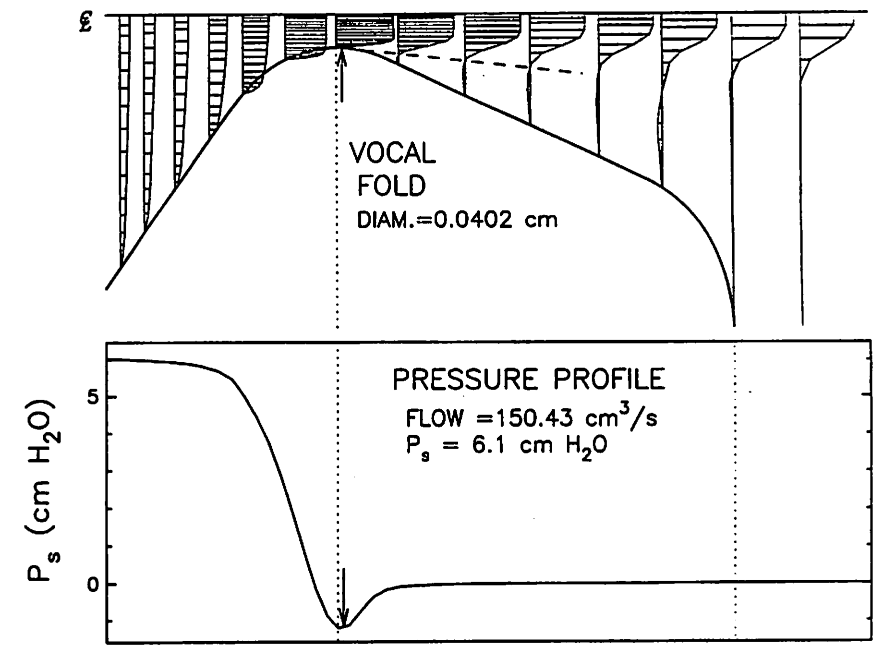


Figure 4. Computationally derived cross sectional velocity profiles within the glottis and the corresponding one-dimensional pressure profile along the upstream-downstream vocal fold boundary. The diameter is 0.0402 cm and the Reynolds number is 1616.

The predicted relationship between flow and surface pressure can be examined in Fig. 4. Due to flow acceleration along the glottal entrance radius (cf. Gauffin & Liljencrants, 1988) and viscosity effects, a significant pressure drop is generated at the entrance to the glottis (minimal diameter). It is immediately followed by pressure recovery caused by the increase in cross sectional area and deceleration of the air flow within the diffuser. Just after the point of separation, the cross sectional area of the flow jet (the boundary of which is estimated by the dashed line in Fig. 4) is seen to expand and the pressure rises quickly. At the point of separation for the condition shown, the pressure rise is 2.5% of the pressure rise from the minimum pressure to the pressure at glottal exit.[6]

Pressure profiles for different diameters are presented in Fig. 5, in which the empirical data for Model 3 are shown for comparison. The Reynolds numbers are around 1630. The pattern of computational pressure profiles are similar for all diameters considered. Pressure profiles also can be seen in the computational works of Iijima et al. (1988, 1990), Liljencrants (1991), and Alipour and Patel (1991).

[6]Figure 4 indicates that the pressure at flow separation for the diverging glottis of 42°was 2.5% of the total recovery from the minimum pressure. In this example, Reynolds number was 1616, glottal diameter was 0.0412 cm, and overall pressure drop was 6.09 cm H_2O. The minimum pressure was -1.22 cm H_2O, the pressure at flow separation was -1.19 cm H_2O, and glottal exit pressure was taken be 0 cm H_2O.

For the relatively large diameter of 0.0614 cm, the computational pressure value at the minimal diameter is higher than the empirical pressure value. It is suggested that the surface of the physical model would not be absolutely smooth, and any added surface resistance may increase the empirical pressure drop at that location. Also, the eccentricity of Model 3 may have created secondary flows and nonuniform pressures in the region of the pressure hole (midway along the glottis).

For smaller diameters, the empirical pressure measurement at the minimal diameter increases relative to the computational value. For the diameter of 0.0202 cm, the empirical data show that the maximum pressure drop is no longer located at the glottal entrance. Compared with the computational results, which consistently creates the maximum drop slightly downstream of the minimal diameter, the variation of the relative value of the pressure drop at the entrance implies the possible effects of a relatively large pressure hole diameter, a slight miscalculation of the pressure hole location (more upstream would result in a higher pressure), or a local irregularity in the diameter of the glottal duct (a larger local diameter would increase the pressure).

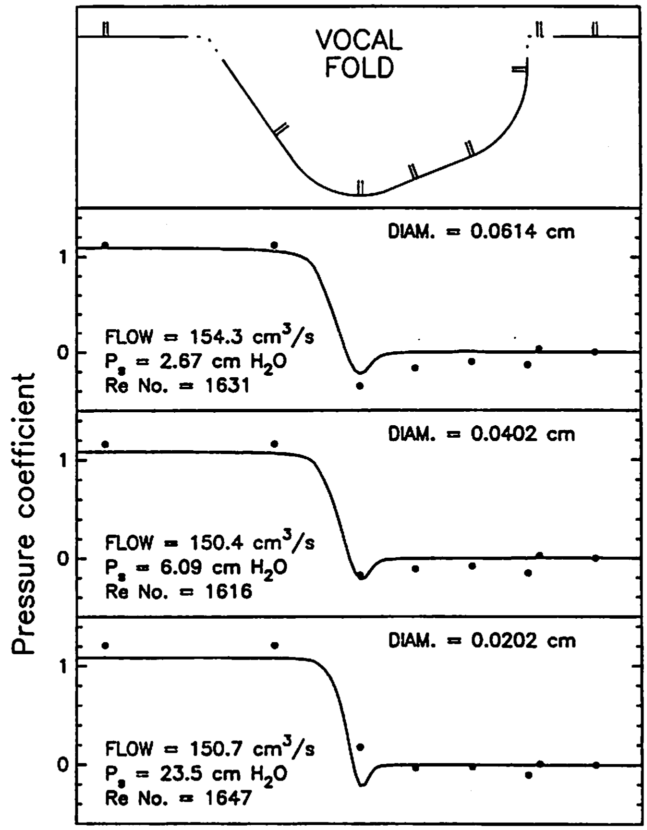


Figure 5. The comparison of pressure profiles on the vocal fold surface with the same glottal angle (diverging 42° but different diameters (0.0614, 0.0402, and 0.0202 cm). The dots indicate results using Model 3 (Scherer, 1981; Scherer & Titze, 1983). The solid lines are the pressure profile predictions using the penalty finite element technique.

Regarding the pressure hole effect, it is well established that if the ratio of the diameter of the pressure hole to the diameter of the duct at the location of the hole is too large, pressure readings may be too high (Rayle, 1958). Some kinetic pressure would add to the static pressure as air impinges into the pressure hole. The diameter of the pressure hole at the glottal entrance of Model 3 was 0.05 cm, while the minimal diameter was 0.0202 cm x 5 = 0.101 cm (Model 3 was a 5-times enlargement of the prototype size). Since the diameter of the pressure hole was half the size of the glottal diameter, a higher pressure reading may have been unavoidable (ref. Scherer, 1981, for further discussion). There apparently are no studies that suggest a correction function that accurately predicts the static pressure when the pressure hole diameter is one-half the duct diameter.

The computational results also suggest that pressure along the glottal surface rises faster just past the entrance than suggested by the empirical data. The eccentricity may have created secondary flows that slightly altered the surface pressures in the physical model. The general finding that pressure does not rise appreciably in the glottis (or, in other words, there is not a large negative pressure at glottal entry) for a large divergent angle is suggested by both the physical and computational results and is consistent with flow separation near the entrance to the glottis. It is further noted that the divergent shape occurs as the glottis is closing during a phonatory cycle, and the pressure dip at glottal entrance may help to close the glottis due to relatively low pressure.

Case Two: Constant Minimal Diameter with Variable Glottal Angle

In this study, the effect of the variation of glottal angle was explored. The minimal glottal diameter was set at 0.04 cm, while the glottal angles of 40, 20, 10, 5 (divergent), 0 (uniform), -5, -10, -20, and -40 (convergent) degrees were used. The computational translaryngeal pressure drop was compared with the S-G equation (Scherer & Guo, 1990b). As mentioned above, the S-G equation was based on empirical results for laryngeal Model 5. Detailed information on the geometries for simulation can be found in Scherer and Guo (1990b, 1991).

Since the tracheal duct air pressure is essentially constant for a few centimeters upstream of the glottal entrance for steady flow, a flow domain with a relatively short upstream distance, as shown in Fig. 6, was used to construct the computational mesh, which had 2367 nodal points and 554 elements. For each angle, 9 different inflow rates were used. The computational results were compared to the empirical predictions derived from the S-G equation for 81 cases listed in Table 2. The average percentage difference was 6.68% (8.86% for translaryngeal pressures greater than 3 cm H_2O; the maximum translaryngeal pressure drop studied was 13.3 cm H_2O.

The computational translaryngeal pressure drop for the different glottal configurations is plotted in Fig. 7. Due to relatively high viscous loss, the uniform glottis corresponds to the largest pressure drop compared to the non-zero angle conditions for any given flow. Relative to the

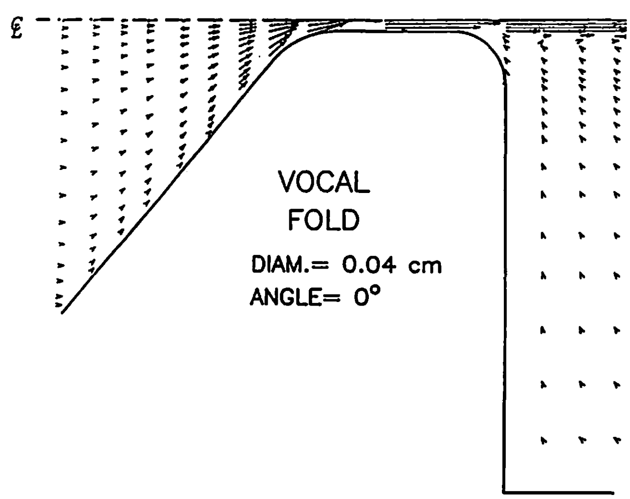


Figure 6. Definition of the flow domain with a relatively short upstream and downstream extent for case 2. The example shows a glottal angle of zero degrees and a Reynolds number of 2000.

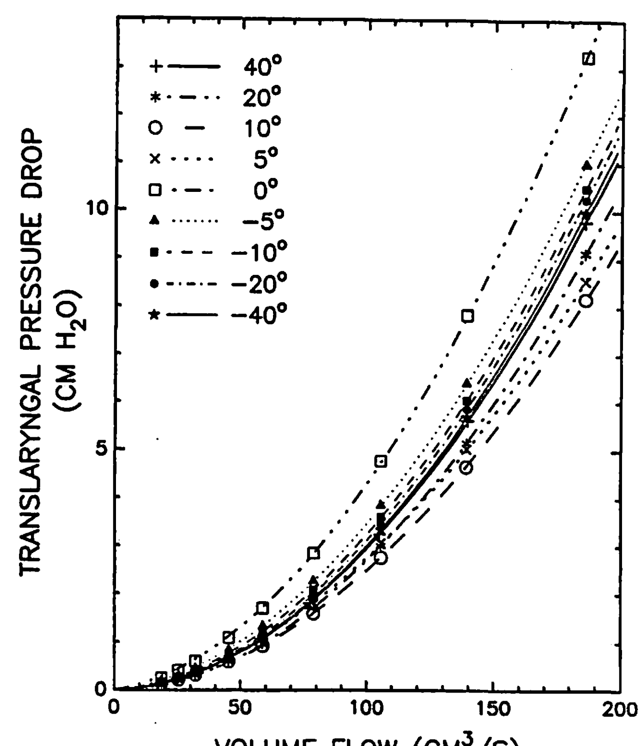


Figure 7. The computational translaryngeal pressure drop for different volume flow rates and glottal angles. The configurations for Model 5 (Scherer & Guo, 1990b, 1991) were used in the computational simulation.

empirical data of Model 5 (Scherer & Guo, 1990b), the computational prediction and empirical data both indicate greatest pressure drop for the uniform glottis for each flow studied. The least pressure drop for any given flow, that is, the least flow resistance, is found to correspond to the 10° divergence for both the computational predictions and Model 5 data for flows greater than about 60 cm^3/s. The 10 degree divergent case is near the optimal angle for pressure recovery within a diffuser (Robertson & Fraser, 1960; Kline, 1959). These results suggest that efficient phonation may involve glottal shaping of about 10° divergence angle near maximum flow and a nearly uniform glottis near glottal closure.

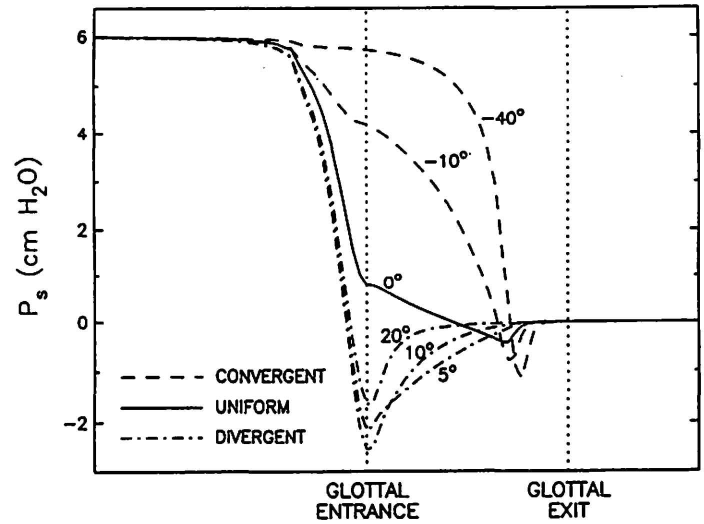


Figure 8. The simulation of pressure profiles for different glottal angles with the same translaryngeal pressure of 6 cm H_2O. The configurations for Model 5 (Scherer & Guo, 1990b, 1991) were used in the computational simulation.

Fig. 8 shows the computational predicted pressure profiles for 6 glottal angles at the same translaryngeal pressure of 6 cm H_2O. During glottal opening within a phonatory cycle, the glottis forms convergent angles and positive pressure is generated along the glottal surfaces except near the rounded exit. The positive pressure would enhance the outward vocal fold movement. During glottal closing, the glottis takes on divergent shapes. The relatively negative pressure generated within the glottis, shown in Fig. 8, would facilitate the movement of the vocal folds inward. This finding is consistent with the asymmetric driving force description of vocal fold oscillation (Titze, 1988). The data illustrated in Fig. 8 suggest that the positive and negative pressures generated through glottal shape change alone can be significant (Scherer & Guo, 1990a; Scherer, 1991).

Discussion

The finite element model presented in this report differs from the existing published methods of Liljencrants (1991), Iijima et al. (1988, 1989, 1992), and the proposed method of Alipour and Patel (1991). Liljencrants (1991) solved Navier-Stokes equations for two-dimensional laminar flow using a vorticity transport formulation, from which vorticity and stream functions resulted. Iijima et al. (1988, 1989, 1992) applied a velocity correction method to solve two-dimensional Navier-Stokes equations in primitive variables. In this method, since the computed velocity does not satisfy the equation of continuity, the computed velocity must be corrected at each iteration (Kawahara and Ohmiya, 1985). The method used by Alipour and Patel (1991) was originally formulated for analysis of turbulent flow. The coordinates and velocity vectors are transformed for the particular problem to be solved, so that the velocities are not expressed in a conventional Cartesian x-y orthogonal coordinate system for the computation. Liljencrants (1991) and Iijima et al. (1990, 1992) have applied their methods to simulate unsteady glottal flow with promising results. In contrast to these other studies, the present study tests the computational method against empirical data for which the same airway boundary conditions were applied to both computational and physical models.

In the penalty finite element method reported here, the primitive variables are directly derived from the computation. A 9-node velocity and 3-node pressure element allows curved boundaries to be defined efficiently and gives enhanced accuracy for pressure (Kim 1988a, 1988b). Using the frontal solver of Irons (1970) for solving the global matrix, the method could be run on a VAXstation instead of a supercomputer.

Although this method may have generated computational results favorably matching the empirical data, its accuracy still needs to be carefully examined. In the second case study, the computational results were compared with the S-G equation, which was derived from data collected using Model 5, a carefully constructed, enlarged plexiglas model of the larynx permitting 63 combinations of glottal angle and diameter. Because of complete glottal symmetry and no eccentricity, it was expected that the computational method should produce the same results as the empirical equation. Some discrepancies exist, however, and may have been created by the computational method.

Fig. 9A is a scatter plot of the computational results and the S-G equation predictions for translaryngeal pressure. The computational results match the S-G equation predictions better for the divergent angles than for the uniform and the convergent angles. The computational method predicts comparatively higher pressure drops for the uniform and convergent angles as Reynolds number increases. This implies that the computational method may assume too much viscous loss for higher Reynolds number flow. In contrast, the computational method predicted too little pressure drop for the diverging glottis Model 3 for small diameter, but this may be explained by the eccentricity and less rigorous fabrication of Model 3. It is also noted that if the flow tends to introduce significant turbulence, the laminar flow assumption made in the computational method used here may not be adequate. For the conditions shown in Figure 9A, the percentage difference between the computational results and the S-G predictions for the divergent conditions is on the average 3.7%. The average percentage difference for the convergent conditions is 8.1%, and for the uniform conditions is 13.1%.

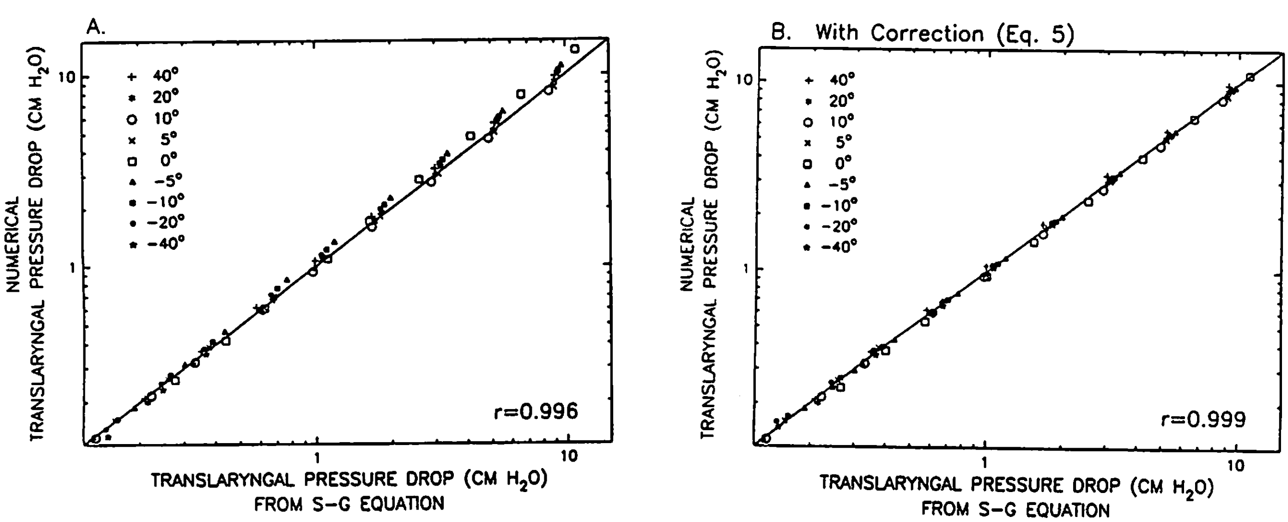


Figure 9. Scatter plots of computational results versus the S-G equation predictions: (A) original comparison, (B) Comparison with correction to computational results using Eq. 6. The pearson correlation r is also given.

For the divergent angles, pressure recovery occurs within the glottis in which the length scale for turbulence is relatively small so that the laminar assumption throughout the entire flow domain may be sufficient to produce acceptable results. For the uniform and convergent angles, flow separation is generated near the exit where sudden expansion exists. As the flow jets out from the exit, it may suddenly become more turbulent, a condition less suitable for the computational method used here. Fig. 9A suggests that the discrepancies between the computational data and the S-G equation predictions appear regular.

For the divergent cases, the average percentage difference for the translaryngeal pressure drop of 3.7% suggests that a correction is not necessary. But for the uniform and convergent cases, a correction derived from the study of the correlation between the computational results and the S-G equation predictions for translaryngeal pressure drop in cm H_2O can be obtained, viz.,

$$P' = (P + 0.03)/C \tag{6}$$

where P' is the corrected translaryngeal pressure drop, P is the translaryngeal pressure drop directly obtained from the computational method, and $C = 1.199 - 0.00766\,X + 0.000119\,X^2$, in which X is the absolute value of glottal angle. Fig. 9B is a scatter plot of the computational results and the S-G equation predictions with the correction (Eq. 6) applied to the uniform and convergent cases. For the translaryngeal pressure drops greater than 3 cm H_2O, the average percentage difference for all angles combined improves from 8.86% to 2.48%.

Conclusion

The penalty finite element computational method, an approach based on laminar flow assumptions, has been applied to realistic laryngeal configurations with good prediction for the translaryngeal pressure drop. Across both studies reported, the overall prediction for translaryngeal pressure drop was 6.2%. The important convergence criteria and sufficient number of iterations for this method were reported. The limited comparison between the computational results and pressure profile empirical data indicates that the computational method appears to usefully predict the pressure along the glottal boundary and has potential for further studies of laryngeal fluid mechanics, but further investigation of three-dimensional pressure gradients due to airway eccentricity is suggested.

Pressure profiles obtained using the penalty finite element method for converging and diverging glottal shapes predict effective positive and negative vocal fold surface pressures, respectively, and emphasize the facilitating effect of the external driving forces during normal vocal fold vibration (Titze, 1988; Scherer & Guo, 1990a, 1993).

Some considerations remain. For a more realistic simulation, a three-dimensional flow domain should be used. The effect of the more realistic eccentricity of the glottis within the laryngeal tract needs exploration. At this stage, minor discrepancies between the computational results and empirical data are apparent for higher flows and pressures, reflecting the possible need for modeling transition and turbulent flows. Despite these areas of extension, the computational model in its present form appears highly effective in predicting significant aerodynamics characteristics within the larynx.

Acknowledgements

We greatly appreciate the thoughtful advice given in the review of an earlier version of this manuscript by Jan Gauffin, Richard McGowan, Fariborz Alipour, Eric Muller, Ralph Ohde, and Hirohisa Iijima.

This research was supported by EPA Agreement CR813113 through the Center for Extrapolation Modeling at Duke University Medical Center, and by grant P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. The research described in this article has been reviewed by the Health Effects Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Agency nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

References

Alipour, F. and Patel, V.C. (1991). "A Two-Dimensional Model of Laryngeal Flow," J. Acoust. Soc. Am., Vol.89, No.4, Pt.2, 1978(A).

Anderson, D.A., Tannehill, J.C., and Pletcher, R.H. (1984). "Computational Fluid Mechanics and Heat Transfer," (Hemisphere, Washington), 132.

Berg, Jw. van den, Zantema, J.T. and Doornenbal, P. Jr. (1957). "On the air resistance and the Bernoulli effect of the human larynx," J. Acoust. Soc. Am., 29, 626-631.

Binh, N. and Gauffin, J. (1983) "Aerodynamics Measurements in an Enlarged Static Laryngeal Model," STL-QPSR 2-3, Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, 36-60.

Crystal, T.H. (1966). A Model of Laryngeal Activity During Phonation. Sc. D. Thesis, Cambridge, MIT.

Davies, P.O.A.L., McGowan, R.S., and Shadle, C.H. (1993). "Practical flow duct acoustics applied to the vocal tract," in Vocal Fold Physiology: Frontiers in Basic Science, edited by I.R. Titze, Singular Publication Group: San Diego, 93-134.

Engelman, M.S., Sani, R.L., Gresho, P.M., and Bercovier, M. (1982). "Consistent vs. Reduced Integration Penalty Methods for Incompressible Media Using Several Old and New Elements," Inter. J. Num. Meth. Fluids, Vol.5, 107-126.

Fant, G. (1982). "Preliminaries to Analysis of the Human Voice Source," STL-QPSR 4/1982, Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, 1-27.

Flanagan, J.L. (1958). "Some properties of the glottal sound source," J. Speech hear. Res., 1, 99-116.

Flanagan, J.L. (1972). Speech Analysis, Systems and Perception. Second Edition, Springer-Verlag: New York.

Fletcher, C.A.J. (1984). "Computational Galerkin Methods," (Springer-Verlag, New York).

Gauffin, J., Binh, N., Ananthapadmanabha, T.V., and Fant, G. (1983). "Glottal Geometry and Volume Velocity Waveform," in Vocal Fold Physiology: Contemporary Research and Clinical Issues, edited by D.M. Bless and J.H. Abbs (College Hill, San Diego), 194-201.

Gauffin, J. and Liljencrants, J. (1988). "The role of convective acceleration in glottal aerodynamics," in Vocal Physiology: Voice Production, Mechanisms, and Function, edited by O. Fujimura, Raven Press, Ltd.: New York, 219-226.

Iijima, H., Miki, N., and Nagai, N. (1988). "Viscous Flow Analyses of the Glottal Model Using a Finite Element Method," J. Acoust. Soc. Am., Vol. 84(S1), S126(A).

Iijima, H., Miki, N., and Nagai, N. (1989). "Fundamental Consideration of Finite Element Method for the Simulation of the Vibration of Vocal Cord," International Conference in Acoustics, Speech and Signal Processing (ICASSP 89), 16.S5.9 (1989-5).

Iijima, H., Miki, N., and Nagai, N. (1990). "Glottal Flow Analysis Based on a Finite Element Simulation of a Two-Dimensional Unsteady Viscous Flow" ICSLP Proceedings, 1990 International Conference on Spoken Language, V.1, The Acoust. Soc. of Japan, Japan, 3.6, 77-80.

Iijima, H., Miki, N., and Nagai, N. (1992). "Glottal Impedance Based on a Finite Element Analysis of Two-Dimensional Unsteady Viscous Flow in A Static Glottis," IEEE Trans. on Signal Processing, V. 40, No. 9, 2125.

Irons, B.M. (1970). "A Frontal Solution Program for Finite Element Analysis," Inter. J. Num. Meth. Eng., Vol.2, 5-32.

Ishizaka, K. (1985). "Air Resistance and Intraglottal Pressure in a Model of Larynx," in Vocal Fold Physiology: Biomechanics Acoustics and Phonatory Control, edited by I.R. Titze & R.C. Scherer, (The Denver Center for the Performing Arts, Denver), 414-424.

Ishizaka, K. and Matsudaira, M. (1972) "Fluid Mechanical Considerations of Vocal Cord Vibration," SCRL (Speech Communications Research Laboratory) Monograph No. 8, (Santa Barbara, California).

Kawahara, M. and Ohmiya, K. (1985) "Finite Element Analysis of Density Flow Using the Velocity Correction Method,' Inter. J. Num. Meth. Fluids, Vol.5, 981-993.

Kim, S.W. (1988a). "A Fine Grid Finite Element Computation of Two¬Dimensional High Reynolds Number Flows," Computers & Fluids, Vol.16, No.4, 429-444.

Kim, S.W. (1988b). "Velocity-Pressure Integrated Versus Penalty Finite Element Method for High-Reynolds Number Flows," NASA CR-179357.

Kim, S.W. and Decker, R.A. (1989). "Velocity-Pressure Integrated Versus Penalty Finite Element Method for High Reynolds Number Flows," Inter. J. Num. Meth. Fluids, Vol.9, 43-57.

Kline, S.J. (1959). "On the Nature of Stall," ASME J. Basic Eng. 81, 305-320.

Liljencrants, J. (1991). "Numerical Simulation of Glottal Flow," in J. Gauffin and B. Hammarberg (eds.), Vocal Fold Physiology: Acoustic, Perceptual and Physiological Aspects of Voice Mechanisms, San Diego: Singular Publishing Group, Inc., 99-104.

Mongeau, L., Coker, C.H., and Kubli, R.A. (1992). "Experimental study of the aerodynamics of a larynx model," J. Acoust. Soc. Am. 92/4, Pt 2, 2391(a).

Rayle, R.E. (1958). "Influence of Orifice Geometry on Static Pressure Measurements," ASME, paper #59-A-234.

Robertson, J.M. and Fraser, H.R. (1960). "Separation Prediction for Conical Diffusers." Trans. ASME, Series D:J. Basic Eng., 82, 1.

Rothenberg, M. (1981). "Acoustic interaction between the glottal source and the vocal tract," in Vocal Fold Physiology, edited by Stevens, K.N. and Hirano, M., Univ. of Tokyo Press: Tokyo, 181-189.

Rothenberg, M. (1983). "An interactive model for the voice source," in Vocal Fold Physiology: Contemporary Research and Clinical Issues, edited by D.M. Bless and J.H. Abbs, College-Hill Press: San Diego, 155-165.

Scherer, R.C. (1981). "Laryngeal Fluid Mechanics: Steady Flow Considerations Using Static Models," Ph.D. Thesis. (The University of Iowa, Iowa City).

Scherer, R.C. (1991). "Physiology of Phonation: a Review of Basic Mechanics," in Phonosurgery: Approaches to Assessment and Treatment, edited by C.N. Ford and D.M. Bless, Raven Press: New York, 77-93.

Scherer, R.C. (1992). "Aerodynamics Assessment in Voice Production," in Assessment of Speech and Voice Production: Research and Clinical Applications, edited by J.A. Cooper, National Institute on Deafness and Other Communication Disorders, NIH., 112¬123.

Scherer, R.C. and Guo, C.G. (1990a). "Effect of Vocal Fold Radii on Pressure Distributions in the Glottis," J. Acoust. Soc. Am. 88(S1), S150(A).

Scherer, R.C. and Guo, C.G. (1990b). "Laryngeal Modeling: Translaryngeal Pressure for a Model with Many Glottal Shapes," ICSLP Proceedings, 1990 International Conference on Spoken Language, V.1, The Acoust. Soc. of Japan, Japan, 3.1, 57-60.

Scherer, R.C. and Guo, C.G. (1991). "Generalized Translaryngeal Pressure Coefficient for a Wide Range of Laryngeal Configurations," in J. Gauffin and B. Hammarberg (eds.), Vocal Fold Physiology: Acoustic, Perceptual and Physiological Aspects of Voice Mechanisms, San Diego: Singular Publishing Group, Inc., 81-90.

Scherer, R.C. and Guo, C.G. (1993). "Intraglottal pressures for pathological glottal configurations using a computational fluid dynamic approach," (in preparation)

Scherer, R.C. and Titze, I.R. (1983). "Pressure-Flow Relationships in a Model of the Laryngeal Airway with a Diverging Glottis," in Vocal Fold Physiology: Contemporary Research and Clinical Issues, edited by D.M. Bless and J.H. Abbs (College Hill, San Diego), 179-193.

Scherer, R.C., Titze, I.R. and Curtis, J.F. (1983a). "Pressure-Flow Relationships in Two Models of the Larynx Having Rectangular Glottal Shapes," J. Acoust. Soc. Am. 73, 668-676.

Scherer, R.C., Titze, I.R., Linville, R., Hueffner, D., and Shaw, K. (1983b). "The effects of vocal fold growths on pressure-flow relationships in the larynx," in Transcripts of the Eleventh Symposium: Care of the Professional Voice, edited by V. Lawrence, The Voice Foundation: New York, 25-36.

Taylor, C. and Hughes, T.G. (1980). "Finite Element Programming of the Navier-Stocks Equations," (Pineridge Press, Swansea).

Titze, I.R. (1986). "Mean Intraglottal Pressure in Vocal Fold Oscillation," J. Phonetics V.14, 359-364.

Titze, I.R. (1988). "The Physics of Small-Amplitude Oscillation of the Vocal Folds," J. Acoust. Soc. Am., 83/4, 1536-1552.

Wegel, R.L. (1930). "Theory of vibration of the larynx," Bell Sys. tech. J., 9, 207-227.

White, F.M. (1979). "Fluid Mechanics," (McGraw-Hill, New York).

NCVS Status and Progress Report - 4
June 1993, 21-27

Modulation of Fundamental Frequency by Laryngeal Muscles during Vibrato

Tzu-Yu Hsiao, M.D.
Department of Otolaryngology, National Taiwan University Hospital
Nancy Pearl Solomon, Ph.D.
Erich S. Luschei, Ph.D.
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

The variations in voice fundamental frequency (F_0) that occur during vibrato production may be produced, at least in part, by modulation of laryngeal muscle activity. We have quantified this relation by using a cross-correlation analysis of the changes in F_0 during vibrato, and the changes either in motor unit firing rate or in gross electromyographic activity from the cricothyroid (CT) and the thyroarytenoid (TA) muscles. Two trained amateur tenors provided the data. Correlations were generally quite strong, thus providing support for previous evidence that fundamental frequency modulation in vibrato involves active modulation of the laryngeal motoneuron pools, especially by the cricothyroid muscle. In addition, phase delays between muscle modulation and changes in fundamental frequency were substantial. This finding may help provide insight regarding the mechanisms responsible for the production of vibrato.

The variations in fundamental frequency (F_0) that occur during the production of vibrato have been reported to be associated with modulation of laryngeal muscle activity. Specifically, examination of acoustic data and gross electromyographic (EMG) data has revealed EMG modulations during normal vibrato primarily for the cricothyroid muscle (1, 2, 3, 4), but also for the thyroarytenoid muscle (2, 3, 5, 6), lateral cricoarytenoid muscle (2, 5, 6), and the posterior cricoarytenoid muscle (3).

Shipp, Doherty, & Haglund (4) provided quantitative analyses of the relation between gross muscle activity of the cricothyroid and F_0 modulation during vibrato. Two key findings were that cricothyroid activity increased approximately 30% when switching from straight tone production to vibrato for low- and mid-range pitches, and that the changes in F_0 followed the changes in cricothyroid activity by approximately 110° (this is termed the "phase delay").

The purpose of the present investigation was to further quantify the relation between the modulation of laryngeal muscle activity with F_0 changes during the production of a natural vibrato. A cross-correlation analysis procedure was used to allow for quantitative evaluation of the correlation between the two signals and the extent of the phase delay. We were particularly interested in the activity of single motor units in the laryngeal muscles. Analysis of single motor units provides more detailed information about the neural control of the muscle, such as firing rates, recruitment and derecruitment thresholds, and burst duration. In this study, we concentrated on the interval between motor unit spikes for vibrato produced at various fundamental frequencies to examine if firing rate was modulated with the vibrato cycle.

Method

Procedures

Two trained amateur singers, both tenors, produced vibrato while sustaining single pitches. Several pitches were sampled throughout their ranges. Bipolar hooked-wire stainless-steel electrodes were inserted percutaneously in the cricothyroid (CT) and thyroarytenoid (TA) muscles. Electrode placement was confirmed by the usual techniques developed previously by Hirano and Ohala (7). The voice was sensed by a microphone placed approximately 30 cm from the lips.

The EMG signal was recorded on an FM data tape recorder and the voice was recorded simultaneously on the tape. These two signals were converted to digital signals and recorded onto a personal computer using data acquisition software (DATAQ Codas) at a sampling rate of 25 kHz per channel. An example of these unprocessed data is provided in Figure 1 (top panel: motor unit activity; bottom panel: voice, average F_0 = 285 Hz). Each record for analysis consisted of 2 s of data.

A curve was created from the voice signal that reflected the changes in F_0 with vibrato production. The fundamental period

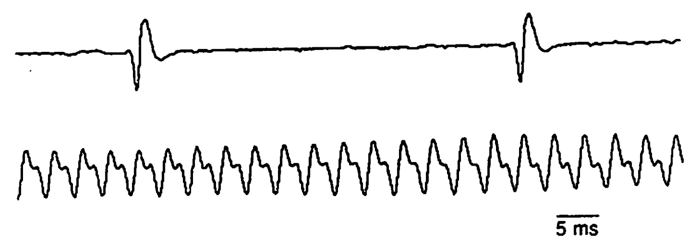


Figure 1. (top) Single motor unit from the CT muscle (top) and voice waveform (bottom) of Singer #1.

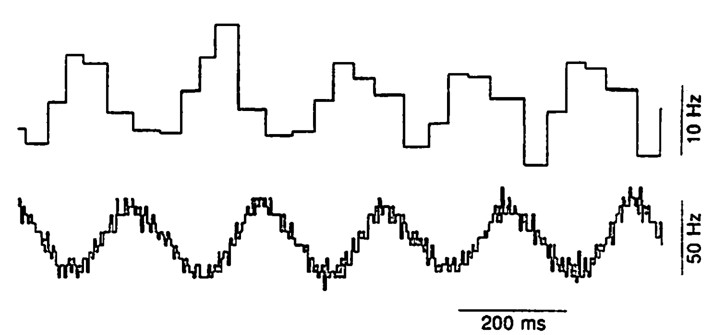


Figure 2. (bottom) Curves representing the motor unit firing rate (top) and the voice F_0 (bottom) derived from the same data sample used in Figure 1. Mean motor unit firing rate = 25.3 Hz; mean voice F_0 = 285 Hz.

of the voice was marked using the peak-detection algorithm of the DATAQ data analysis program for the computer; accurate marking of each cycle was confirmed visually by one of the investigators. Using the computer software, a constant voltage signal was then created, integrated, reset with each peak of the voice F_0 cycle, and held until the next peak. This resulted in a record consisting of short horizontal line segments whose voltage reflected the interval between each vibratory cycle of the vocal folds. The reciprocal of this signal was calculated and plotted, creating a record of variations in F_0 across time with increasing frequency represented by an upward deflection. Because of the reset-and-hold procedure, the curve was offset in time by one cycle period. To correct this, the value of an average F_0 cycle period, considered to be the best estimate of the offset, was subtracted from the time base. The frequency curve for the voice data illustrated in Figure 1 is shown in the bottom half of Figure 2 (average vibrato rate = 4.7 Hz). The small, high-frequency, variations in the frequency curve reflect vocal jitter. The larger, slower fluctuations reflect the vibrato.

Analysis of records with single motor unit activity was similar to the procedure used for the vibrato curve. Motor unit spikes were marked using the computer software's peak-detection algorithm. These marks were used to create a record of unit firing rate, exactly as described previously for the creation of the voice F_0 record. The curve representing motor unit firing rate was corrected by subtracting the value of the average motor unit firing period from the time base. The frequency (firing rate) curve for the motor unit illustrated in Figure 1 is shown in the top half of Figure 2 (average firing rate = 25.3 Hz). The horizontal line segments are longer than those in the voice F_0 curve because the intervals between motor unit spikes were, of course, longer than those between F_0 cycles.

Gross EMG data records were rectified and smoothed with the computer software's moving average algorithm. This created a curve which reflected changes in the amplitude of the EMG over time.

The two frequency curves, for voice and motor unit activity, or the frequency curve for voice and the amplitude curve for gross EMG, were cross-correlated. The results of this procedure on the curves illustrated in Figure 2 are given in Figure 3. The peak of the curve corresponds to a correlation coefficient of 0.86. The phase delay at this peak value is 72 ms, which based on the frequency of the vibrato cycle (4.7 Hz), corresponds to a phase delay of 122°.

Results

For Singer #1, one particular motor unit from the CT was observable for pitches D3 (147 Hz), D4 (294 Hz), and G4 (392 Hz). Twenty-eight data records were made and analyzed for this single motor unit. The average vibrato rate plotted against the average F_0 for each of the 28 records is plotted in Figure 4 (following page). Note that vibrato was faster at the highest of the 3 pitches sung by this subject, a phenomenon reported previously by Horii (8) but not observed by Shipp et al. (9). In Figure 5, the average firing rate of the motor unit is

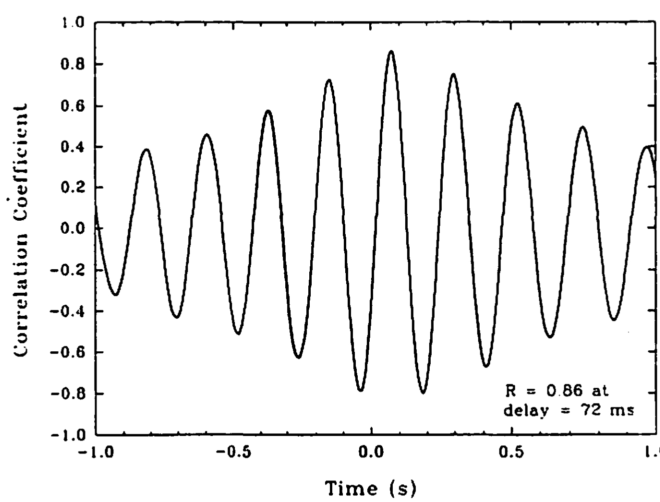
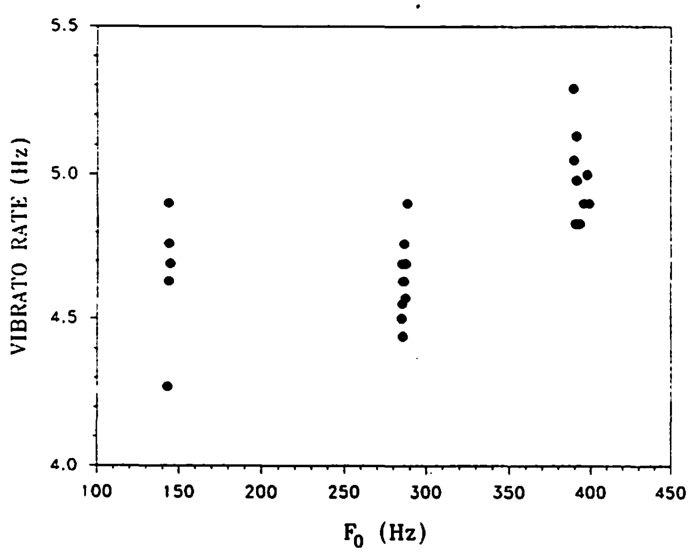


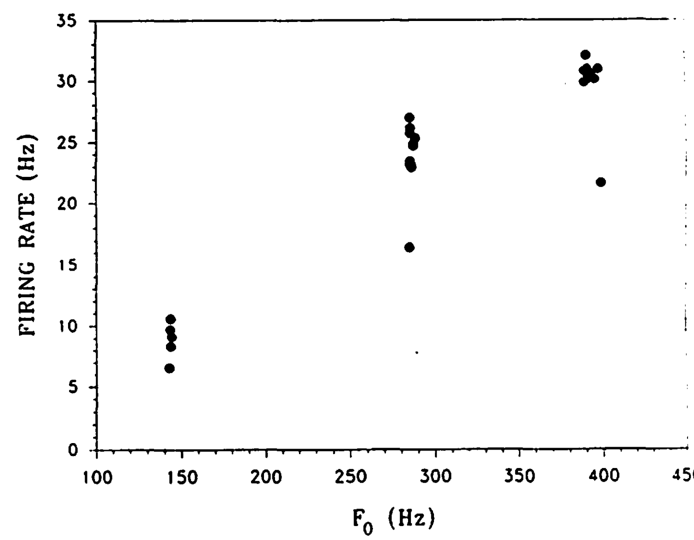
Figure 3. Results of a cross-correlation analysis of the frequency curves illustrated in Figure 2.

plotted against the average F_0 for each record. It is clear from this figure that the motor unit had a slower firing rate at the low pitch than at the higher pitches.

The correlation coefficients from the cross-correlation procedure for each of the records are plotted in Figure 6. The best correlations were found for the middle of the 3 pitches, and were generally good for most of the samples at the higher and lower pitches. The one sample with a poor correlation coefficient ($\underline{R} = 0.11$) actually involved activity of the motor unit which was well coordinated with the vibrato cycle, but which was inconsistent in the number of times it fired per cycle. A portion of this record is provided in Figure 7. Note that the motor unit fired twice per vibrato cycle initially, then switched to a single spike per cycle and then reversed back to double spikes. The cross-correlation procedure does not handle this situation well.

According to the phase delays calculated for the 28 data records, changes in F_0 followed changes in motor unit firing rate by an average of 127.4° ($\underline{SD}$ = 10.7°). In Figure 8, the phase delays are plotted against the vibrato rate. It appears that there may be a weak relationship such that





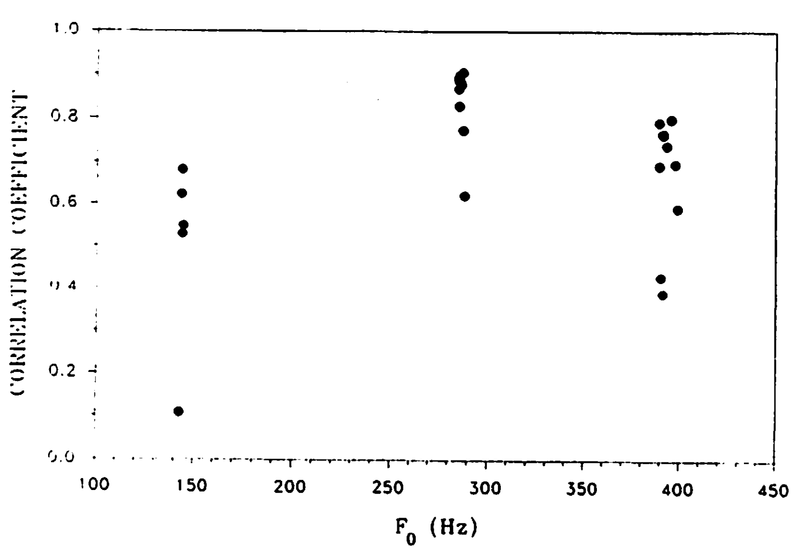


Figure 4. (top left) The rate of the F_0 modulations during vibrato (vibrato rate) plotted according to average F_0 for the 28 data samples from Singer #1. Figure 5. (top right) Motor unit firing rate from the CT muscle of Singer #1 plotted against the average F_0 for the 28 data samples. Figure 6. (at left) Correlation coefficients from the cross-correlation analysis of the two frequency curves (modulations in motor unit firing rate and voice F_0) plotted against average F_0 for the 28 data samples from the CT muscle of Singer #1.

the phase delays tend to be greater at the faster vibrato rates ($r = 0.395$).

For Singer #2, a few single motor units could be detected from the CT data, but only at the low frequency end of his range [G2 (98 Hz) and C3 (131 Hz)]. At these low frequencies, the motor unit discharge was inconsistent probably because the motoneurons were close to their recruitment threshold. In addition, the singer's vibrato was somewhat inconsistent and uneven at these low pitches. We measured 16 2-second records of single motor unit activity from the CT for Singer #2, but only 1 yielded adequate results -- the cross correlation coefficient was 0.64, and the phase delay was 97.3°. Gross EMG data were obtained from both the CT and the TA for this singer. Of the 14 data records for the CT, including productions at 3 pitches (D4, G4, A4), the correlations between gross CT activity and changes in F_0 during vibrato were modest (mean $\underline{R} = 0.5$, $\underline{SD} = 0.13$). The average phase delay was 132° ($\underline{SD}$ = 14°). Activity of the TA was not as well correlated with the F_0 (12 records, mean $\underline{R} = 0.31$, $\underline{SD} = 0.13$). The average phase delay was 140° ($\underline{SD}$ = 15.9°).

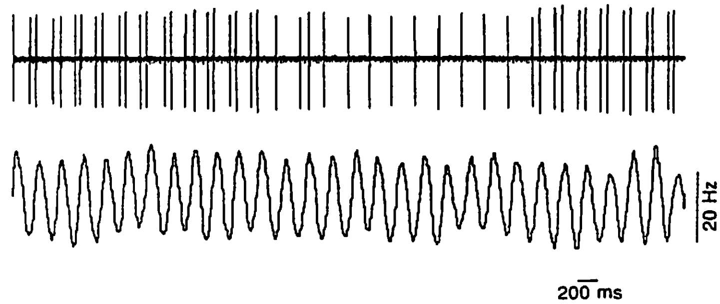


Figure 7. (top) The data sample from the CT of Singer #1 that yielded a correlation coefficient of 0.11. Note the change in the pattern of motor unit activity (top) with the F_0 modulations associated with vibrato (bottom) over time. Mean motor unit firing rate = 6.6 Hz; mean voice F_0 = 143.

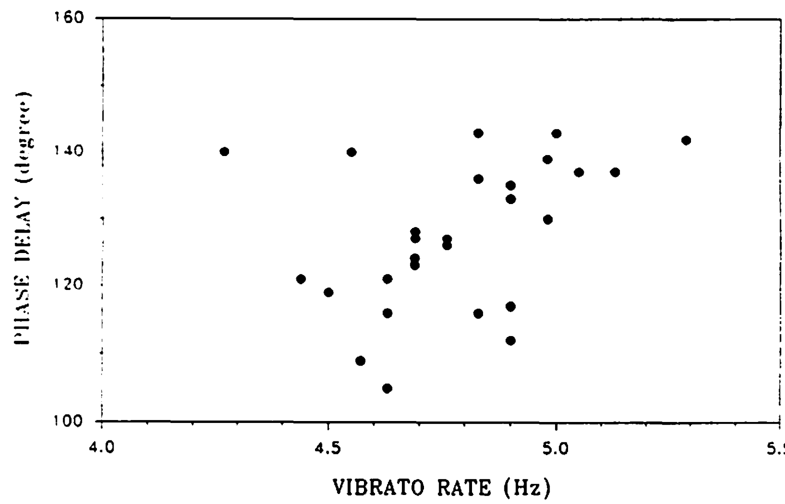


Figure 8. (bottom) The phase delays between the modulation in motor unit firing rate and F_0 modulations during vibrato plotted against vibrato rate for the 28 data samples from the CT muscle of Singer #1. The correlation coefficient for these two variables is 0.395.

Discussion

We studied modulations in the firing rate of a single motor unit from the CT muscle over a wide range of voice F_0, and found that these correlated well with variations of fundamental frequency during vibrato production by a trained tenor. This finding supports previous literature (1, 2, 3, 4). On average, the change in fundamental frequency followed the change in modulations of muscle activity during vibrato by 127°. Shipp et al. (4) also found substantial phase delays between changes in the amplitude of gross EMG in the CT and frequency changes with vibrato.

The fact that this motor unit could be studied over a relatively large range of vocal pitches (one octave and a fourth) was quite remarkable. The CT is thought to be inactive at low pitches and so active at high pitches that recruitment of additional motor units would obscure the view of the motor unit being studied. This may have been a unique opportunity, although we hope that this research can be replicated in the future. Nonetheless, the present findings are consistent with previous literature that involved measures of gross EMG.

Gross EMG data from the CT and TA muscles for Singer #2 correlated with F_0 variations in vibrato, but the correlations were not as strong as those obtained for Singer #1. This may be due to the nature of Singer #2's vibrato -- its frequency excursions (typically 12 Hz or 0.5 semitones) were not as extensive as Singer #1's (typically 30 Hz or 2.0 semitones). It also may be related to the difference in the analyses between the subjects in that the activity of a single motor unit may follow the frequency changes in vibrato more closely than would the amplitude of gross EMG activity. The phase delay results were quite similar. Single motor unit analyses from this subject's data were problematic, as described previously.

The substantial phase delays found for all data records may contribute to the understanding of some of the basic mechanisms that produce vibrato. Studies of muscle physiology provide insight regarding the dynamic characteristics of muscle contraction based on phase information. Mannard and Stein (10) frequency-modulated electrical stimuli to the nerve of the soleus muscle of anesthetized cats. They found that the isometric force response curves for gain and phase were fit well by a critically damped second-order low-pass filter function. This same result was reported recently when the nerve to soleus was electrically stimulated using amplitude, frequency, or a combination of amplitude and frequency modulation (11). Thus, the low-pass model of isometric muscle contraction dynamics appears to be robust.

In the low-pass model of cat soleus described by Baratta et al. (11), modulation at about 2 Hz was associated with a phase lag of 90°. A similar corner, or "cut-off," frequency was obtained for the human soleus muscle (12). If this line of analysis is used in the case of the present study, the 5 Hz variations in motor unit firing rate, or EMG modulation, can be considered an input modulation to the CT muscle and the changes in F_0 can be considered the output of the muscle. The fact that the phase delays were an average of 122° implies the corner frequency of CT is somewhat less than 5 Hz, perhaps closer to 4 Hz.

The fact that most singers have vibrato that is modulated at approximately 5 Hz may result, at least in part, from the basic physiology of the laryngeal muscles. If the input to these muscles were modulated at the same amplitude but at frequencies significantly higher than 5 Hz, the response of the muscles would be too low to produce much F_0 variation. To produce such a rapid vibrato, modulation rates would have to be extreme (i.e., the muscles would turn on and off at high rates), and the nervous system may be incapable of this. Titze et al. (13) have suggested that the production of vibrato involves a peripheral oscillator that shows resonant properties, either because of physical characteristics or feedback control. Perhaps the phase lag associated with 5 Hz modulation is critical to oscillation in this system, whereas the larger lags that would occur with higher vibrato rates would be incompatible with oscillation.

Acknowledgments

This research was supported by Grant No. P60 DC00976-02 from the National Institutes on Deafness and Other Communication Disorders. We gratefully acknowledge Dr. Minoru Hirano for his assistance with data collection, Mr. Douglas Van Daele for his contributions to data analysis, and Drs. Fariborz Alipour-Haghighi and Judith Preston Grayhack for theoretical contributions.

References

1.Mason RM, Zemlin WR. The phenomenon of vocal vibrato. NATS Bull 1966;22;12-17.

2.Hirano M, Vennard W, Ohala J. Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles. Folia Phoniatr 1970;22:1-20.

3.Martin F, Thumfart WF, Jolk A, Klingholz F. The electromyographic activity of the posterior cricoarytenoid muscle during singing. J Voice 1990;4:25-29.

4.Shipp T, Doherty ET, Haglund S. Physiologic factors in vocal vibrato production. J Voice 1990;4:300-304.

5.Niimi S, Horiguchi S, Kobayashi N. The physiological role of the sternothyroid muscle in phonation An electromyographic observation. Ann Bull RILP 1988;22:163-170.

6.Niimi S, Horiguchi S, Kobayashi N, Yamada M. Electromyographic study of vibrato and tremelo in singing. In: Fujimura O (ed). Vocal physiology: Voice production, mechanisms, and functions. New York: Raven Press, 1988:403-414.

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

8.Horii Y. Frequency modulation characteristics of sustained /a/ sung in vocal vibrato. J Speech Hear Res 1989;32:829-836.

9.Shipp, T, Leanderson, R., & Sundberg, J. Some acoustic characteristics of vocal vibrato. J Res in Singing 1980;4:18-25.

10.Mannard A, Stein RB. Determination of the frequency response of isometric soleus muscle in the cat using random nerve stimulation. J Physiol 1973;229:275-296.

11.Baratta R, Zhou B-H, Solomonow M. Frequency response model of skeletal muscle: effect of perturbation level, and control strategy. Med & Biol Eng & Comput 1989;27:337-345.

12.Bawa P, Stein RB. Frequency response of human soleus muscle. J Neurophysiol 1976;39:788-793.

13.Titze IR, Solomon NP, Luschei ES, Hirano M. Interference between normal vibrato and artificial stimulation of laryngeal muscles at near vibrato rates. J Voice 1993;**[this volume]**.

NCVS Status and Progress Report - 4
June 1993, 29-40

Interference Between Normal Vibrato and Artificial Stimulation of Laryngeal Muscles at Near Vibrato Rates

Ingo R. Titze, Ph.D.
Nancy Pearl Solomon, Ph.D.
Erich S. Luschei, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Minoru Hirano, M.D.
Department of Otolaryngology, Kurume University

Abstract

A stabilized tremor hypothesis for vocal vibrato is investigated. The stabilizer is assumed to be a mechanical oscillator that may contain reflex loops. Artificial stimulation of the cricothyroid muscle in one subject showed a well-defined resonance curve of this peripheral oscillator at about 5.0 Hz. Combined artificial stimulation with natural vibrato showed that the vibrato could be entrained by a peripheral stimulus, provided the two frequencies are separated by no more than about ± 0.5 Hz. This suggests that vibrato frequencies are not "hard-wired" centrally, even though a collection of centrally generated tremors may serve as excitation to the peripheral oscillator.

Introduction

The neural mechanisms of vocal vibrato are still being debated, but evidence is accruing in support of the hypothesis that vibrato is a stabilized physiologic tremor in the laryngeal musculature[1]. Through vocal training, singers appear to be able to reduce variability in the rate and extent of a normal 4-6 Hz physiologic tremor. It is not known whether this stabilization of frequency and extent is facilitated by some peripheral mechanism (e.g., a mechanical load or a servo loop acting as a "tuner"), or whether a central oscillator is influenced by training.

A study pertinent to this question was conducted by Rack and Ross[2]. Sinusoidal movement was delivered to joints in the upper extremities in people with Parkinson's disease who exhibited limb tremor. They found that the external stimulus sometimes entrained the tremor and sometimes the two frequencies coexisted independently. Tremor was entrained most consistently when the driving frequency was close to the tremor frequency. The authors argued that reflexes play a major role in the genesis of limb tremor. Abnormal reflexes can give oscillatory instability at the periphery that can then be stabilized by an external oscillator or a mechanical load.

Much earlier than this, von Holst[3] studied the fin movements of fish. He discovered what he called "superposition and the magnet effect". Two coupled oscillators can exhibit totally independent movements, the output being a superposition of the two movements, or the oscillations can "attract" one another like a magnet, moving together in synchrony. This can be related to the 17th century observation by Christian Huygens that two pendulum clocks hanging on a common wall can synchronize each other (Jackson, 1991)[4]. If mechanical coupling through the wall is weak, the clocks tick independently at their own frequencies. At some critical coupling, however, they lock onto each other and move in perfect synchrony. In the language of modern nonlinear dynamics, the attractor for the coupled oscillators has changed from a torus (doughnut-like shape in 3-dimensional phase space) to a limit cycle (a closed loop in 2-dimensional phase space). Such a change in behavior, called a bifurcation, is now well understood in nonlinear dynamics.

Investigations on the larynx have demonstrated that fundamental frequency (F_o) of phonation can be manipulated by external stimulation of laryngeal muscles. This suggests that an artificial vibrato can be induced in the voice. Sapir, Campbell, and Larson[5] showed that stimulation of the cricothyroid (CT) muscle had the greatest effect on F_o, but extrinsic muscle stimulation added to F_o changes. Results were highly consistent across three monkeys that were studied.

Kempster, Larson, and Kistler[6] showed that in five normal human subjects, stimulation of CT and thyroarytenoid (TA) muscles both had the effect of raising F_o. TA had a faster contraction time than the CT, which agrees with *in vitro* results obtained by Alipour et al. and Perlman and Alipour[7,8]. Stimulation of the TA muscle does not always raise F_o, however. Results by Titze, Luschei and Hirano[9] showed that F_o can decrease with TA stimulation when overall CT activity is high and overall TA activity is low. Given this more complex result for TA stimulation, the present study will focus only on CT stimulation.

Relative to the stabilized tremor hypothesis, it is interesting to point out that patients with clinically diagnosed voice tremor have tremorous activity in both TA and CT, as well as in adductor and extrinsic laryngeal muscles[10]. Tremor frequencies during phonation were between 4.7 and 5.2 Hz in that study. Since this narrow frequency range is quite close to typical vibrato frequencies in singers[11], it seems logical to investigate further the relationship between vibrato and tremor.

The purpose of this study was to determine whether the vibrato of a trained singer could be altered by superimposing an artificial vibrato onto the laryngeal muscles. Presumably, if the vibrato frequency were "hard-wired" centrally, the effect of artificial stimulation should simply be additive. In other words, two independent modulations of the muscle force should exist, with their appropriate beat frequencies. On the other hand, if peripheral modulation were to alter the frequency of the neural oscillator, some form of phase locking or entrainment should be expected. We investigated these alternatives.

Methods

A trained amateur singer (tenor) served as the subject. He was seated in a dental chair. Pairs of hooked-wire stainless steel electrodes were inserted into his cricothyroid (CT) and thyroarytenoid (TA) muscles, but only data from the CT are reported here. The accuracy of placement of the electrodes was verified by high-pitched productions. EMG recordings were normal and have been reported previously[9].

The subject was asked to choose a comfortable pitch for which he could produce long sustained vowels *with* and *without* vibrato. This pitch (A_3 = 220 Hz) was given repeatedly as a prompt. The length of the sustained vowels was typically 10-20 seconds at a comfortable loudness.

The hooked-wire electrodes were used for direct unilateral muscle stimulation, using a Grass S-88 stimulator and an associated constant-current stimulus isolation. Monopolar electrical pulses, 0.5 ms in duration, were presented at the following stimulation rates: 2, 3, 4, 5, 5.5, 6, 7, and 8 Hz. The stimulation current was made as large as possible without exceeding the discomfort level for the subject ($\leq$ 5 ma). Once the current was selected, it was held constant throughout the experiment. The stimulation produced a clearly audible frequency modulation in the subject's voice.

Shape and Spectrum of Muscle Twitch

Single-pulse stimulation of laryngeal muscles results in twitches that are superimposed over the natural tetanic contraction of the muscles. The twitches obviously cannot be measured directly in units of force, but it was demonstrated previously[9] that the time course and relative magnitude of a fundamental frequency (F_0) "twitch" is similar to the force twitch. Figure 1 (solid line) shows such an F_0 twitch averaged over 100 repeated stimulations when the stimulation to the CT muscle was at 2 Hz and the subject produced no intentional vibrato. Dashed lines represent the standard deviation over repeated stimulations.

Note that the twitch has a latency of about 20 ms (stimulus was at t = 0), rises to a peak in about 30 ms, and falls to near baseline in another 50 ms. Thus, for a total twitch duration *d* (about 100 ms), the *duty ratio*, *r*, for the artificially induced vibrato is

$$r = dF_s \quad , \tag{1}$$

where F_s is the stimulation frequency. This duty ratio is related to the "sinusoidalness" of the induced vibrato. At F_s = 2 Hz, the duty ratio is 0.2. For such a small duty ratio, one would expect to see a rich harmonic structure in the spectrum of the F_0 contour. As the stimulation frequency F_s increases, the spectrum should become weaker in its harmonic structure. This is demonstrated in Figures 2 and 3 (see following page), where the waveforms and spectra of computer-generated F_0 twitches are shown with F_s as a parameter (2 Hz and 6 Hz, respectively). In Figure 2, the magnitude of the second harmonic is 70% of

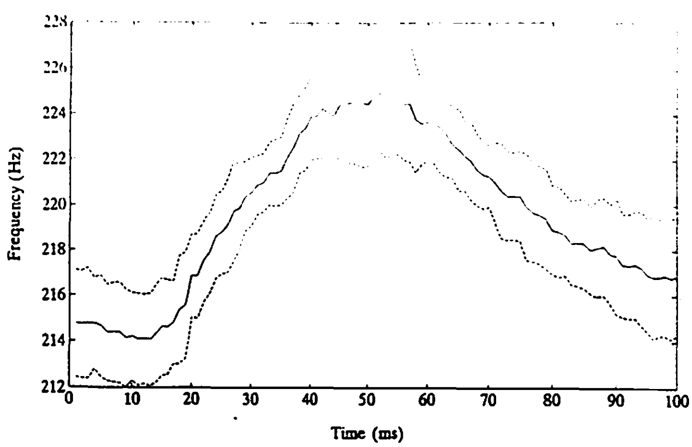
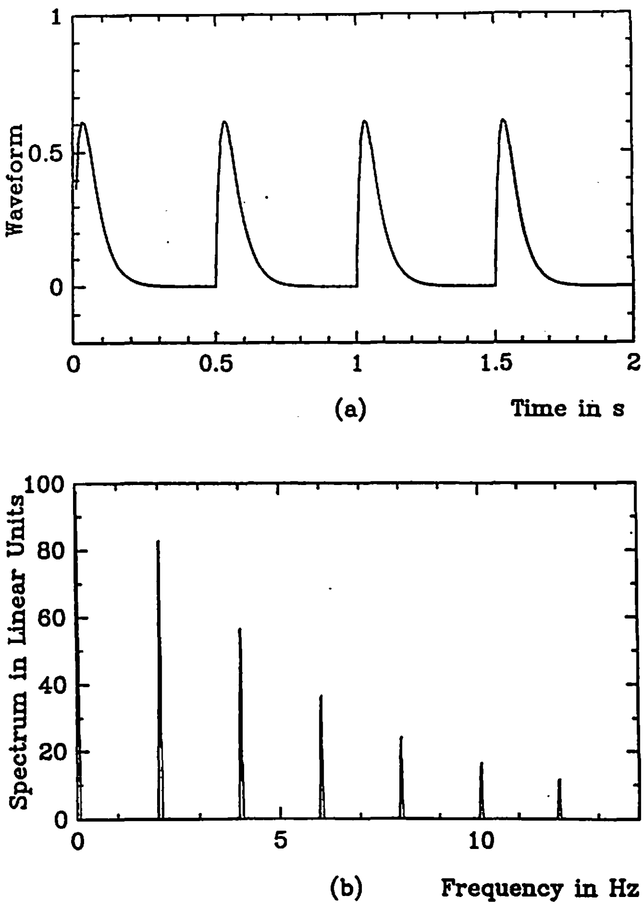


Figure 1. Average F_0 "twitch" pattern obtained by stimulating the cricothyroid (CT) muscle. Dashed lines show standard deviations over many sequential stimulations.

the fundamental and the third harmonic is about 50% of the fundamental. In Figure 3, on the other hand, the second harmonic is already less than 50% of the fundamental.



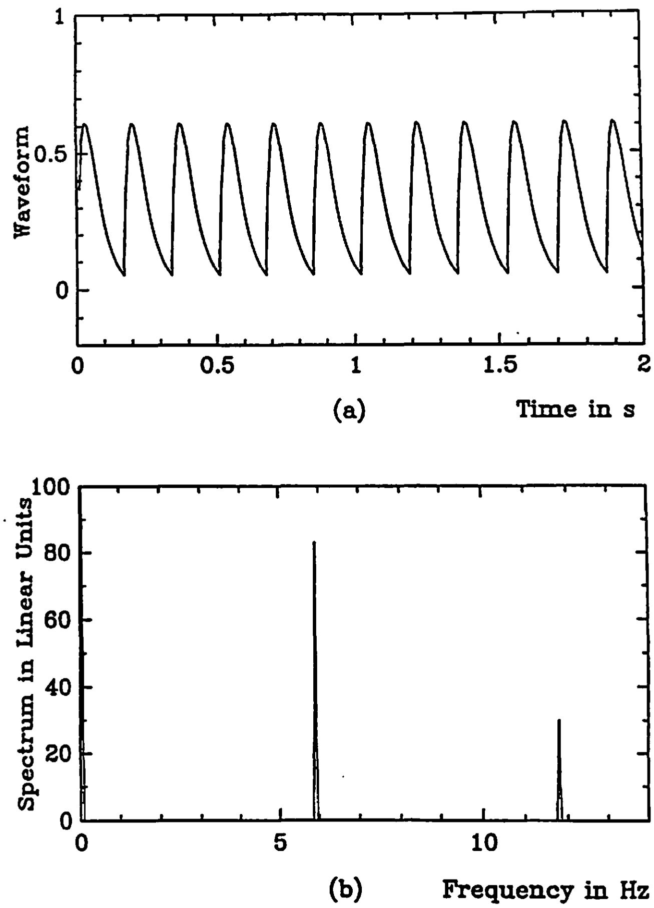


Figure 2.(at left) (a) Simulated twitches and (b) their spectrum, with $F_s = 2$ Hz. Figure 3.(at right) (a) Simulated twitches and (b) their spectrum, with $F_s = 6$ Hz.

Spectral Analysis of F_0 Contours

F_0 contours of the sustained vowels were obtained by triggering a frequency meter on the positive or negative slope of the largest peak of the microphone signal. The output signal of the hardware frequency meter was sampled digitally with a data acquisition software package for an IBM PC (CODAS). The sampling frequency for the F_0 contour was 1000 Hz and the amplitude resolution was 8 bits. Because the F_0 modulation was typically only 1-3% of the mean F_0, amplitude

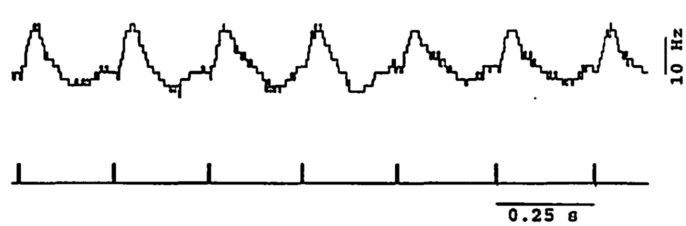


Figure 4. Raw record of an F_0 contour for the stimulation pattern shown below, with $F_s = 4$ Hz.

resolution of the vibrato cycle was somewhat less than ideal (2-3 bits). A typical F_o contour for a 4 Hz stimulation (without natural vibrato) is shown in Figure 4.

Spectra of the F_o contour were computed with an FFT algorithm. These spectra showed the component frequencies of the F_o twitches very clearly. Figure 5(a) shows an example of the normal vibrato spectrum (no stimulation), whereas Figure 5(b) shows an example of the spectrum for 2 Hz stimulation with no intentional vibrato. Note the lack of harmonic structure for the natural vibrato (5.6 Hz) in comparison with the rich harmonic structure for the artificial (pulsed) stimulation. It might be questioned whether the harmonics are primarily a characteristic of the pulse shape (instead of the quantization noise) in the F_o contour, but the spectra were taken over a large enough window of cycles that all quantization irregularities were averaged out.

Over repeated trials, the natural vibrato had a mean frequency of 5.7 Hz with a standard deviation of 0.085 Hz. In a concurrent study[12], it is shown that the natural vibrato is correlated with sinusoidal increases and decreases in the firing rates of motor neurons to the CT muscle, which would explain the simple spectrum in Figure 5.6(a).

In order to describe further details of the experiment, a conceptual model of vocal vibrato needs to be presented. Specific hypotheses will then be tested in relation to this model.

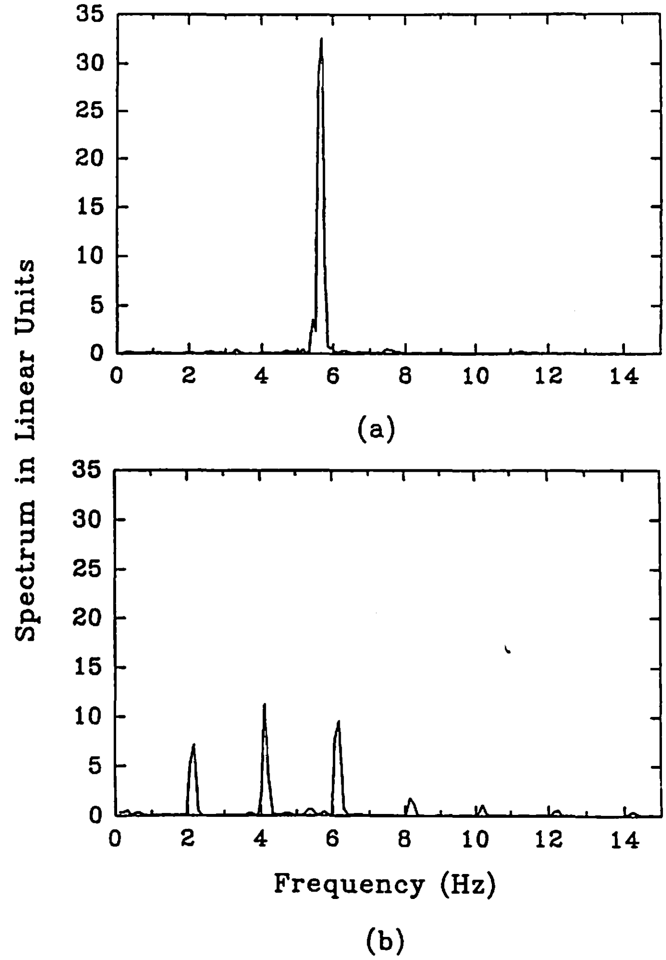


Figure 5. Spectrum of F_o contour. (a) natural vibrato, (b) stimulation at 2 Hz.

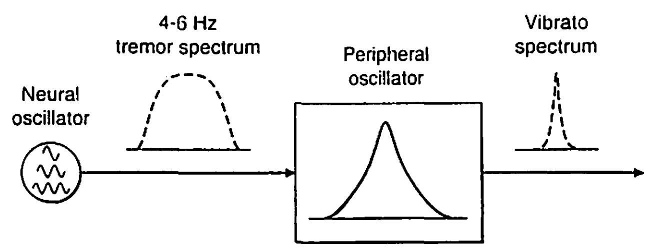


Figure 6. System diagram for the proposed stabilized tremor model of vocal vibrato.

A Model of Vocal Vibrato

Figure 6 shows a system diagram of the proposed stabilized tremor model of vocal vibrato. Main components of the system are a neural oscillator and a peripheral (mechanical) oscillator. The neural oscillator produces a spectrum of frequencies in the 4-6 Hz region, the exact distribution of which is not known. The peripheral oscillator filters the tremor spectrum, producing a more stable vibrato frequency. It is hypothesized that the peripheral oscillator is basically an inertio-visco-elastic load, but not an entirely passive load. There probably are some reflex loops that, with improper gains and phase delays, can make the mechanical system oscillate without a periodic input.

Under normal conditions, however, the peripheral oscillator is assumed to act like a mechanical filter, like the stabilizing bar used by a high-wire gymnast.

Characteristics of the Peripheral Oscillator

The precise biomechanical detail of the peripheral oscillator is not known. One possibility is that it includes the rotation between the cricoid and thyroid cartilages to adjust vocal fold length. Attempts to quantify this rotation have been made[9,13]. Membranous vocal fold length L was derived as a function of normalized cricothyroid muscle activity a_{ct} and normalized thyroarytenoid muscle activity a_{ta}. The result was

$$L = L_o[1 + G(Ra_{ct} - a_{ta})] \quad , \tag{2}$$

where L_o is the rest length (no muscle activity), G is a mechanical gain factor for range of elongation, and R is a torque ratio (maximum counterclockwise torque produced by CT divided by maximum clockwise rotation produced by TA). The possibility for a purely mechanical resonance exists because G and R involve rotational stiffness and inertia between the cricoid and thyroid cartilages. Thus, there may be a natural frequency of oscillation about the CT joint.

Alternatively, or additionally, a reflex loop might establish (or contribute to) this peripheral resonance phenomenon. Note that vocal fold length in equation (2) is sensitive to a difference term $(Ra_{ct} - a_{ta})$ in muscle activity. Such a difference term can provide the necessary feedback for a closed loop system. Assume, for example, that stretch receptors in the vocal folds are sensitive to a length change ΔL brought about by an increase in a_{ct}. In an attempt to regulate L, the reflex system may increase a_{ta}, but with a delayed response. The initial increase in length would then be followed by a delayed decrease in length, causing a fluctuation in the regulation of L. This fluctuation could reach a limit cycle (self-sustained oscillation) if the feedback gain (between the afferent ΔL stimulus and the a_{ta} response) is too large and the delay is too long. Further investigations of these phase delays and gains are presently being conducted.

It should be kept in mind that this is only one of many possible explanations for a peripheral mechanical oscillator. Other candidates might be vertical movement of the laryngeal framework, strap muscle coupling of the larynx to the sternum or hyoid bone, or even an oscillatory movement of the arytenoid cartilages. Much work is needed to clarify these mechanisms and to formulate better hypotheses.

Testing the Peripheral Oscillator

Artificial pulsed stimulation of the CT muscle should produce the resonance effect shown in Figure 7. The peripheral oscillator should selectively filter the line spectrum of a muscle twitch, boosting one or two of its harmonics. This was already demonstrated in Figure 5(b), where the second and third harmonic of the stimulus received a boost. These harmonics were on opposite sides of the resonance frequency in Figure 5(a).

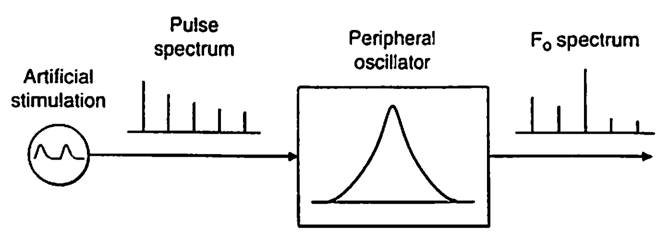


Figure 7. System diagram for artificial stimulation of CT muscle with pulses.

The filtering characteristic of the peripheral oscillator is further explored in Figure 8, where the stimulus frequency is varied from 3 Hz to 5 Hz and the response to several harmonics is shown. Note that the harmonic (or harmonics) closest to 5 Hz always get emphasized. In Figure 8(a), where the fundamental is at 3 Hz, both the fundamental and the second harmonic are boosted. In part (b), the 4 Hz fundamental gets a greater boost than the 8 Hz second harmonic, and in part (c), the 5 Hz fundamental gets a maximum boost because it coincides with the resonance frequency of the oscillator.

The complete response of the peripheral oscillator was constructed in 1 Hz steps. Again, the protocol was to have the subject produce no vibrato and to stimulate the CT with pulses. The amplitude of the stimulus fundamental (F_s) was measured from the magnitude spectra of the F_o contours. The result is shown in Figure 9. Note the tuning near 5.0 Hz, which agrees remarkably well with the mean voice tremor frequency reported by Koda and Ludlow[10].

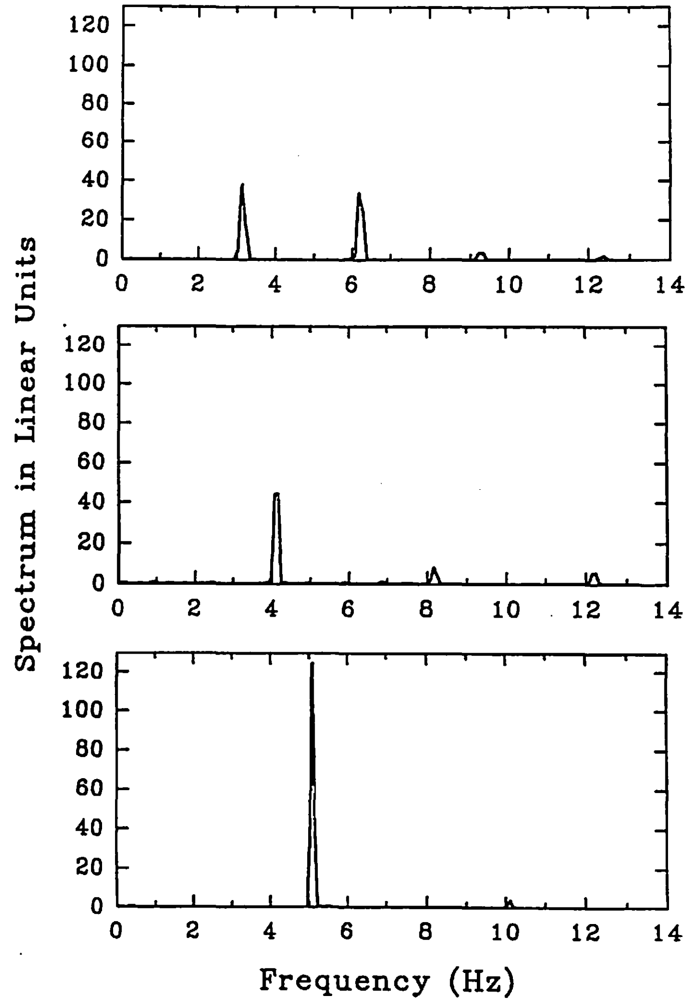


Figure 8. Spectra of F_o contours for stimulation at (a, top) 3 Hz, (b, center) 4 Hz and (c, bottom) 5 Hz.

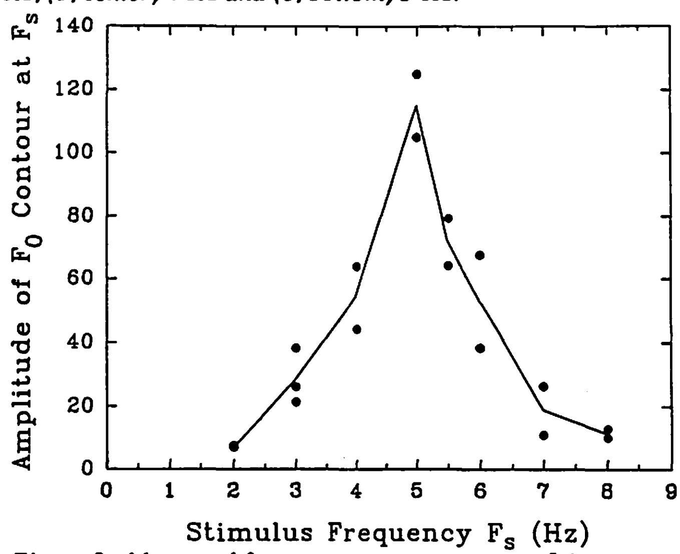


Figure 9. Measured frequency response curve of the peripheral oscillator.

Entrainment Between Natural Vibrato and Artificial Stimulation

A characteristic of coupled oscillators is that one oscillator can entrain another. If one oscillator has a dominant mode of vibration, it can pull a second oscillator into synchrony with it. Applied to our vibrato model, the neural oscillator could entrain the peripheral oscillator, or vice versa. To study this effect, it is desirable to have the frequency of one oscillator under experimental control. Without the use of brain stimulation, however, and without direct manipulation of some of the parameters of the peripheral oscillator (e.g., stiffness, mass, loop gains, etc.), this is a difficult task.

A less ambitious task is to see if the neural-peripheral combination of oscillators can be entrained by artificial stimulation. This adds a third frequency to which either the neural oscillator, the peripheral oscillator, or both can be entrained.

Artificial stimulation should not affect the natural vibrato unless the stimulus frequency F_s is very close to the vibrato frequency. In most cases, the vibrato should remain stable and a simple interference pattern (with beat frequencies) should be observed. When F_s, or any harmonic stimulus frequency nF_s, gets into the 5-6 Hz range, however, an entrainment process should be seen. The vibrato frequency should be "pulled" toward the stimulus frequency.

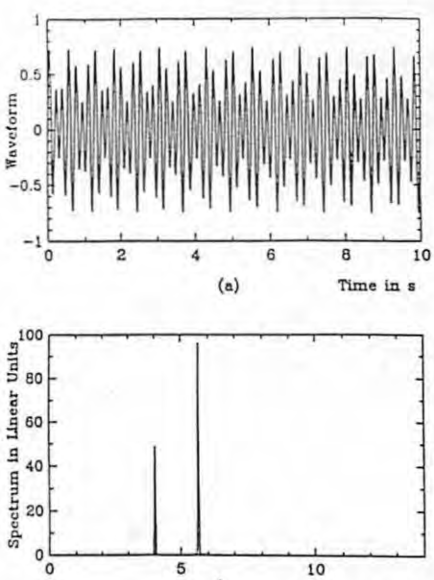
To model a simple theoretical interference pattern, let

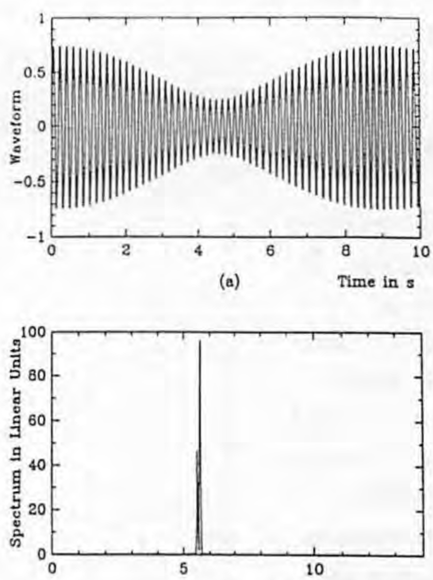
$$F_o = \overline{F_o} + E_s \sin 2\pi F_s t + E_v \sin(2\pi F_v t - \phi)\,, \tag{3}$$

where $\overline{F_o}$ is the mean F_o, E_s and E_v are the modulation extents for stimulated vibrato and natural vibrato, respectively, F_s and F_v are the modulation frequencies, and ϕ is the phase between the two modulations. Figure 10(a) shows a simple interference pattern, computed from equation (3), for two frequencies spaced relatively far apart (F_s = 4 Hz and F_v = 5.61 Hz). The absolute values of E_s and E_v are scaled arbitrarily, but the ratio E_v/E_s is 2.0. The constant $\overline{F_o}$ is not plotted. The phase ϕ was chosen by a random number generator to be somewhere between 0 and 2π.

Note that the amplitude modulation pattern, which never repeats exactly because the frequencies are incommensurate, spans about 3-4 cycles, as dictated by the beat frequency (1.61 Hz). The FFT spectrum is shown in Figure 10(b). Only two frequencies are present, as expected.

Figure 11 shows the same interference pattern when the two frequencies are very close (F_s = 5.5 Hz and F_v = 5.61 Hz). The amplitude modulation pattern now spans about 50 cycles, with a beat frequency of 0.11 Hz. Finally, in Figure 12 the frequencies are 5.61 Hz and 6.0 Hz, respectively. In this case, the beat frequency is 0.39 Hz, showing an amplitude modulation that spans over approximately 14 - 15 cycles. These patterns should all be observed in the F_o contours of the subject if no entrainment was to occur.





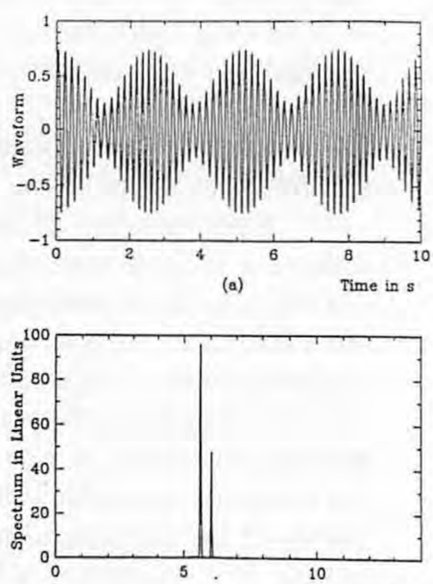


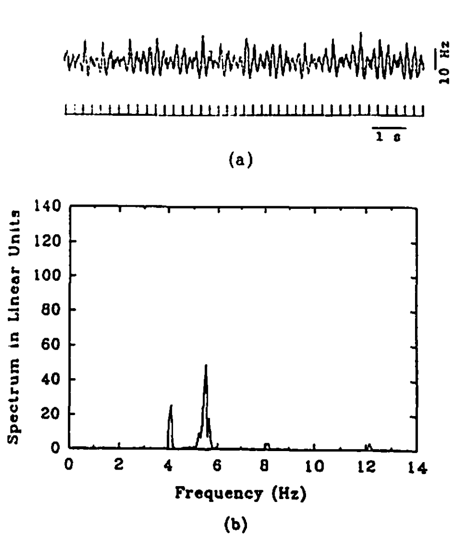
Figure 10 (left). Simple interference pattern for two incommensurate frequencies, 4 Hz and 5.61 Hz. (a) waveform and (b) spectrum. Figure 11 (center). Simple interference pattern for two incommensurate frequencies, 5 Hz and 5.61 Hz. (a) waveform and (b) spectrum. Figure 12 (right). Simple interference pattern for two incommensurate frequencies, 6 Hz and 5.61 Hz. (a) waveform and (b) spectrum.

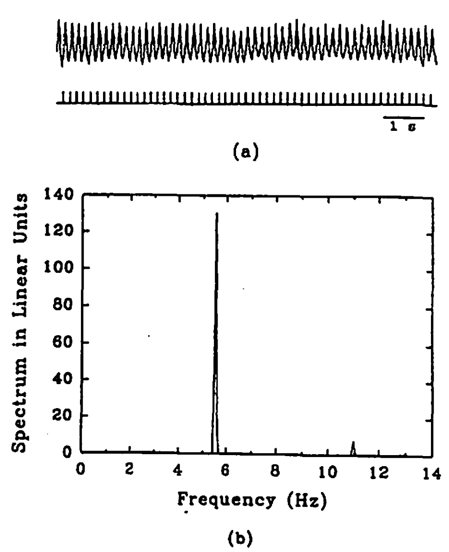
Figures 13, 14, and 15 are subject data that correspond, in stimulation frequency, to the models presented in Figure 10, 11, and 12. Comparing first Figure 13 to Figure 10, we see that there is no entrainment at 4 Hz stimulation. A typical interference pattern is seen in part (a), with the modulation envelope spanning about 4-5 cycles. Spectrally, a 4 Hz stimulation frequency and its harmonics coexist with a 5.7 Hz vibrato frequency, as seen in part (b).

Comparing now Figures 14 to Figure 11, a major difference is seen. No long-term modulation envelope is detected in part (a), and the spectrum in part (b) shows a single frequency at 5.5 Hz. It is clear that the 5.5 Hz stimulus has entrained the 5.7 Hz vibrato. This occurred consistently over four phonation trials. The same pattern was seen for the 6 Hz stimulation (Figure 15). The obvious modulation that was present in Figure 12(a) is clearly not seen in Figure 15(a), and the frequencies have again merged into a single line in Figure 15(b). This occurred in all of four trials. The same kind of entrainment occurred also at 5 Hz stimulation for all trials.

Figure 16 is a summary of the entrainment pattern for all frequencies probed. Vibrato frequency (ordinate), averaged over four trials, is plotted against stimulus frequency (abscissa). Note that entrainment occurs over the narrow interval from 5-6 Hz. Outside of this interval, the average vibrato frequency remains nearly constant.

A final case of interest is shown in Figure 17. Here entrainment to a *harmonic* of F_s was observed, but not consistently. In Figure 17(a), the vibrato is entrained to the third harmonic (6 Hz) of a 2 Hz stimulus. In a subsequent trial, however, entrainment did not occur (Figure 17b). Either the subject was no longer "surprised" by the entrainment effect and prepared for it in the second trial, holding the vibrato frequency more steady, or the third harmonic stimulus was not strong enough to entrain the vibrato in all cases due to inherent noises and threshold effects.





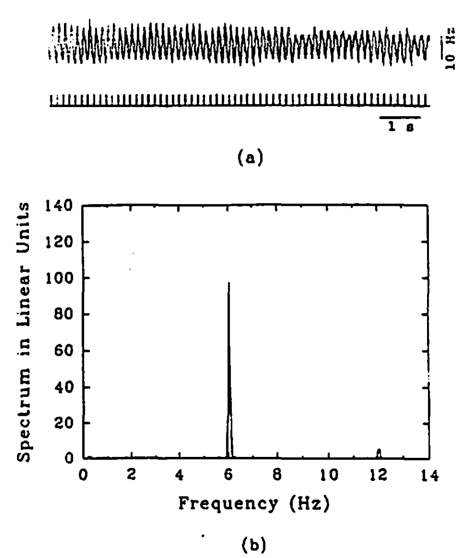


Figure 13 (left). (a) F_o contour and (b) spectrum for 4 Hz stimulation of the CT in the presence of vibrato. Figure 14 (center). (a) F_o contour and (b) spectrum for 5.5 Hz stimulation of the CT in the presence of vibrato. Figure 15 (right). (a) F_o contour and (b) spectrum for 6.0 Hz stimulation of the CT in the presence of vibrato.

Conclusion

Results of this single-subject study indicate that vocal vibrato can be altered by artificial stimulation of the cricothyroid muscle. Specifically, the data suggest that the oscillating mechanism responsible for vocal vibrato can be entrained by an external oscillator that twitches the muscle in a periodic fashion. Entrainment was possible when the stimulus frequency was on the order of $\pm$ 0.5 Hz away from the natural vibrato frequency. For greater separation, entrainment was not observed.

Based on the observations, it is hypothesized that natural vocal vibrato is a stabilized voice tremor. A peripheral oscillator, which acts like a mechanical filter to artificial stimulation, either reduces the variability of centrally-generated tremor frequencies (4-6 Hz) by resonance tuning, or produces its own oscillation (peripherally) by feedback instability. The present experiment could not eliminate one or the other of these possibilities, but it is clear that a peripheral oscillator is involved in either case.

The resonance frequency of the peripheral oscillator is not rigid. It can apparently be modified by voice training or by unusual muscle patterns. Informally, it has been reported that vibrato frequency can change with excitement, fatigue, lack of muscle tones (as in a wobble, perhaps), or disease. Which components of the oscillator are most variable (mechanical stiffnesses, inertial loads, afferent loops, or contractile properties of laryngeal muscles) is not clear. Perhaps studies with tremor patients could shed some light on this question. In tremor patients, the stabilizing effect of the peripheral oscillator may be lost because the peripheral oscillator itself becomes unstable.

Before this study can be conclusive, several additional subjects need to be investigated. This is expected to be difficult, since muscle stimulation with needle insertion is risky for subjects who are vocally trained. Studies of this type will unfortunately always be done with small numbers of subjects.

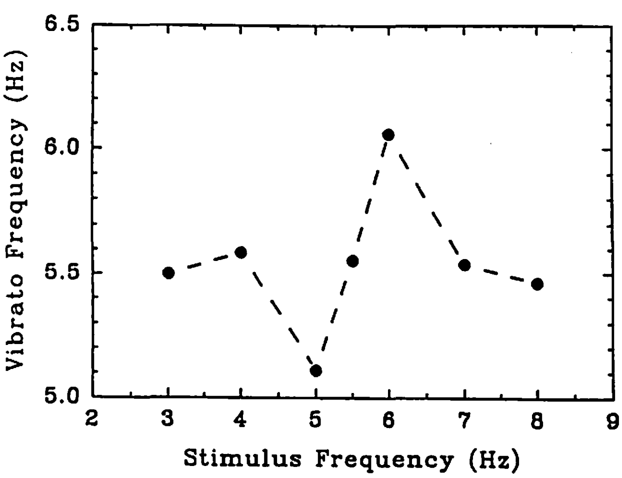


Figure 16. Entrainment curve showing vibrato frequency versus stimulus frequency.

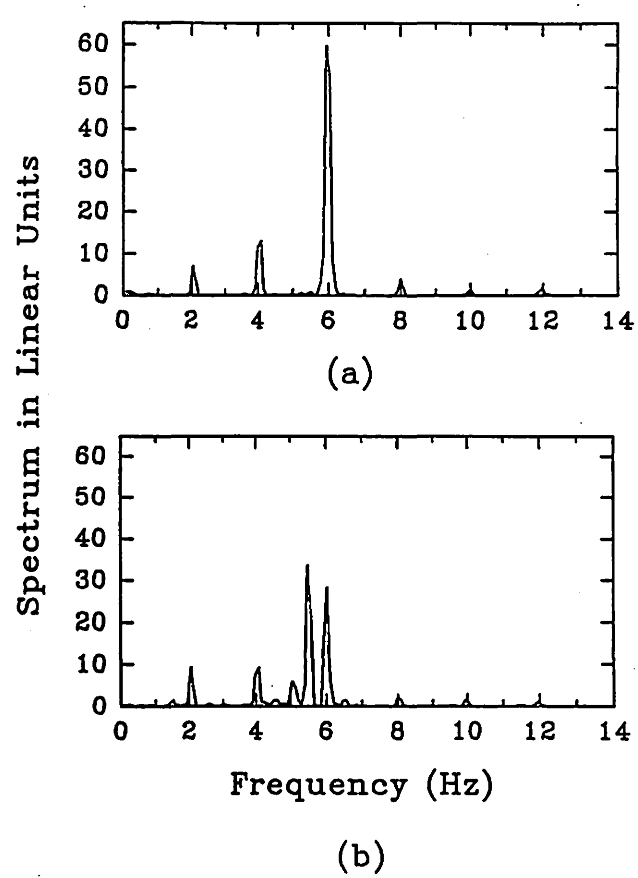


Figure 17. Stimulation of the CT at 2 Hz. (a) third harmonic entrains the vibrato frequency and (b) third harmonic does not entrain the vibrato frequency.

A major shortcoming of the study was the fact that the auditory loop was not manipulated. It would be interesting to mask the input to the auditory system to see if entrainment occurs in the same way. This is left for a future investigation.

Acknowledgments

This work was supported by Grant No. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. We are grateful to Julie Lemke for manuscript preparation and to Mark Peters for graphic support. Special thanks also go to Douglas Van Daele, Kang Liu, Tara Temperly, and Beth Chittick for data analysis.

References

1. Ramig L, Shipp T. Comparative measures of vocal tremor and vocal vibrato. J Voice 1987; 1(2): 162-167.

2. Rack PMH, Ross HF. The role of reflexes in the resting tremor of Parkinson's disease. Brain 1986; 109:115-141.

3. von Holst, E. (1937). The Behavioral Physiology of Animals and Man: Selected Papers, Vol. 1; Coral Gables, University of Miami Press. (Reprinted from Gallistel, CR. The Organization of Action: A New Synthesis, 1980; Hillsdale, NJ: Lawrence Erlbaum Associates).

4. Jackson, E.A. Perspectives of Nonlinear Dynamics. Cambridge University Press, Cambridge, England 1991; p. 203.

5. Sapir S, Campbell C, Larson C. Effect of geniohyoid, cricothyroid and sternothyroid muscle stimulation on voice fundamental frequency of electrically elicited phonation in rhesus macaque. Laryngoscope 1981; 91: 457-468.

6. Kempster G, Larson C, Kistler M. Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency, J. Voice 1988; 2(3): 221-229.

7. Alipour F, Titze I, Durham P. Twitch response in the canine vocalis muscle. JSHR 1987; 30:290-294.

8. Perlman A, Alipour F. Comparative study of the physiological properties of the vocalis and cricothyroid muscles. Acta Otolaryngol (Stockh) 1988; 105:372-378.

9. Titze IR, Luschei ES, Hirano M. The role of the thyroarytenoid muscle in regulation of fundamental frequency. J. Voice 1989; 3(3):213-224.

10. Koda J, Ludlow C. An evaluation of laryngeal muscle activation in patients with voice tremor. Otolaryngology-Head and Neck Surgery (in press).

11. Hakes J, Shipp T, Doherty T. Acoustic characteristics of vocal oscillations: Vibrato, exaggerated vibrato, trill, and trillo. J Voice 1987; 1(4): 326-331.

12. Hsiao T-Y, Solomon NP, Luschei ES, Titze IR. Modulation of fundamental frequency by laryngeal musçles during vibrato. In review.

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

NCVS Status and Progress Report - 4
June 1993, 41-63

Making Typically Obscured Articulatory Activity Available to Speechreaders by Means of Videofluoroscopy

Nancy Tye-Murray, Ph.D.
Karen Iler Kirk, Ph.D.
Lorianne Schum, M.A.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa

Abstract

During the past 25 years, there have been attempts to develop speechreading aids that provide information about nonvisible articulatory gestures. Five experiments were conducted to determine whether increasing the amount of visible articulatory information can affect speech recognition, and whether such aids might be beneficial. The experiments involved videofluoroscopy, which allows placement of the tongue body, lips, teeth, mandible, and often velum to be observed in a naturally-occurring sequence and time course during speech. Subjects were asked to speechread videofluoroscopic and video images. The results suggest that making visible typically obscured supralaryngeal articulator activity by means of videofluoroscopy does not enhance speechreading performance. Normal hearing and cochlear implant users always performed better with the video than videofluoroscopy medium. Subjects performed as well when the tongue was not visible in the videofluoroscopic records as when it was visible. Training did not affect subjects' ability to recognize speech presented with videofluoroscopy. The results pose a challenge to direct-realist theory and motor theory. Both theories posit that perceivers must recover information about articulatory gestures in order to abstract linguistic information from the speech signal.

Recognizing speech through vision alone is extremely difficult. On average, hearing-impaired adults recognize only about 40% of words presented in sentence context (e.g., Tyler and Lowder, 1992), even after many years of hearing loss and speechreading practice. Although speechreading is generally difficult, some individuals demonstrate remarkable skill. For instance, in a group of 65 cochlear implant users, performance on a 100-item sentence test presented in a vision-only condition ranged from 0% to 85% words correct (Tyler and Lowder, 1992).

Many researchers and clinicians believe that the difficulty of the speechreading task stems from the nonvisibility of much of the oral activity that produces speech. Jeffers and Barley (1974) estimate that only 40% of speech gestures are visible on the face. For example, tongue blade and tongue dorsum activity cannot be viewed. The tongue tip is only occasionally visible, and only when the talker faces the speechreader. Implicit in this belief about speechreading difficulty is the assumption that if more articulatory activity was visible, speechreading performance would improve.

Despite the fact that speechreading provides only limited speech information, the visual signal can significantly enhance a degraded auditory signal. Tye-Murray (1992a) found that 16 adult multichannel cochlear implant users recognized an average of 39% of the words in a 100-item sentence test using only audition. When the stimuli were presented in an audition-plus-vision condition, performance improved to 83% words correct.

If speechreading is difficult because we cannot see many articulatory events, then it might be useful to develop speechreading aids that provide information about the events. Individuals with impaired hearing could then use them to improve audiovisual speech recognition. During the past 25 years, there has been sporadic interest aimed toward developing this kind of assistive device (e.g., Upton, 1968; Hochberg, Laroche, Papcun, Thomas, and Zacks, 1991; Goldberg, 1972; Kalikow and Swets, 1972; Kunov and Ebrahimi, 1991; Papcun, Hochberg, Thomas, Laroche, Zacks, and Levy, 1992; Pickett, Gengel, and Quinn, 1974).

Upton (1968) developed the seminal speechreading aid. It was an experimental pair of eyeglasses that provided information about invisible or obscured speech features. Five tiny lights were embedded into the center of an eyeglass lens. By means of a bank of low pass and high pass filters, different lights illuminated in response to voicing (as with vowels), voiced fricatives (in which case, the "voicing light" did not illuminate), voiceless fricatives, voiced stops, and voiceless stops. Upton was optimistic about the eventual usefulness of the display technique although he presented little data. He noted several problems, including the need for extensive training to interpret the light patterns.

Some efforts to develop speechreading aids have attempted to include explicit information about tongue activity in the visual display. These efforts receive impetus from theoretical assertions that during speech perception, abstract articulatory gestures play an important role in the process (Browman and Goldstein, 1987): speech perception may entail identifying the phonologically relevant articulatory gestures that produce the speech signal. Papcun et al. (1992) are presently developing a neural network system to infer glossal behavior from the acoustic signal and then represent it in a visual display. The authors suggest that one application of this work is the development of speechreading aids for those with impaired hearing. Individuals could listen to the degraded signal and simultaneously view information about tongue activity.

To date, little research has been conducted on how speechreading performance is influenced by the introduction of visible evidence of articulatory gestures that are usually not available visually. If it can be demonstrated that increasing the amount of available articulatory information enhances speechreading, then an incentive will exist for developing articulatory-based speechreading aids.

The purpose of the five experiments reported in this investigation was to assess whether increasing the amount of visible articulatory information affects speech recognition performance.

All experiments involved videofluoroscopy and audiovideo recordings. Videofluoroscopy is a radiographic procedure that permits two-dimensional viewing of dynamic articulatory behavior. It allows the placement of articulatory structures such as the tongue, lips, teeth, mandible, and often velum to be observed in a naturally-occurring sequence and time course during speech (Williams, 1989). The images appear as moving x-ray pictures. The tongue and lips appear as nonrigid structures whereas the teeth and mandible appear as rigid structures. When the talker is filmed in profile, videofluoroscopy provides a sagittal view of the supralaryngeal articulators. Photo 1 presents a single frame from a videofluoroscopic film.

Videofluoroscopy and cinefluorography (Moll, 1960; cinefluorographic films cannot visually be distinguished from videofluoroscopic films with the naked eye) have been used extensively by speech science researchers and speech-language pathologists. For example, Kent and Moll (1972) used cinefluorography to demonstrate that the lingual constrictions associated with /g/ are more anterior when the following vowel is /i/ rather than /u/. McWilliams, Morris, and Shelton (1990) describe how videofluoroscopy can be used to obtain clinical descriptions of individuals who have cleft palate.

The popularity of radiographic procedures suggest that they provide a reasonably accurate reflection of articulatory behaviors. However, there are some limitations in how well they convey articulatory information. The videofluoroscopic records present two-dimensional portrayals of three-dimensional structures. They do not provide information about tongue grooving or about asymmetries in the activity of the lateral aspects of the tongue. Although how the tongue displaces over time is telling, the actual site of lingual contact for lingual consonants can only be inferred (Kent & Moll, 1972). McWilliams et al. (1990) note that a "midsagittal view provides information about movement of the velum and posterior pharyngeal wall and about height of the velum and velar relationship to adenoids and posterior pharyngeal wall, but it doesn't provide information about the velum off midline nor about movement of the lateral pharyngeal walls (p. 165)".

In the first two experiments, we provided information about typically nonvisible upper articulator activity to normal hearing subjects in a vision-only condition and to cochlear implant users in an audition-plus-vision condition using videofluoroscopy. In a third experiment, we similarly tested normal hearing subjects who are experienced with viewing videofluoroscopic and cinefluorographic images. Two of these subjects also received training for the task. In the fourth experiment, we asked cochlear implant subjects to speechread videofluoroscopic images when the tongue body was visible and then when the tongue body was obscured by metal fillings in the talker's mouth. Finally, we evaluated whether information about typically nonvisible articulatory activity enhanced cochlear implant users' word recognition performance beyond that obtained in an audition-only condition.

Experiment 1

The purpose of Experiment 1 was to determine how well normal hearing adults recognize speech when watching a videofluoroscopic film. Speech recognition scores achieved with videofluoroscopy were compared to speech recognition scores achieved when subjects watched a video film.

Methods

Subjects. Fourteen adults with normal hearing, 10 female, served as subjects. The subjects reported no history of hearing deficits or uncorrected vision problems. They were recruited through an in-house hospital news bulletin, and they received payment for their participation.

Test stimuli. The test stimuli consisted of one sentence list from the Repeated Sentence Frame Test (Tye-Murray, 1992b). A different ordering of the sentences was used for each of the two test media (i.e., video or videofluoroscopy). The list contains six sets of four sentences that are syntactically identical but differ in two or three keywords. The sentences from each set are randomly presented within the list. With a few exceptions, the keywords within a set have the same number of syllables and are interchangeable. The test is scored as the percentage of keywords correctly identified. There are a total of 49 keywords. The sentence list is presented in Appendix A.

A female talker of general midwestern English produced the test sentences with two different means of filming, the videofluoroscopy medium and the video medium. The talker had no lateral tooth fillings. She produced the sentences with clear speech, pausing 5 seconds between each one. The audio signal was recorded simultaneously with the visual image.

A frame rate of 60 frames per second (Skolnick, 1970) was used for both filming conditions. This rate is rapid enough to preserve the temporal resolution of the signal. In both filming conditions, the talker was filmed from a lateral view so that her profile from head to shoulder was visible.

To obtain the video films, a Panasonic videocamera (model AG-180) was used to record the talker. The camera was mounted on a tripod and the room was well lit. Movement of the talker's lips, jaw, and cheeks was clearly visible.

To obtain the videofluoroscopic films, a lateral view of the talker producing the sentences was recorded using a standard videofluoroscopic system. Skolnick and Cohn (1989) and Williams (1989) describe the videofluoroscopic apparatus. For a lateral view the talker is positioned between a fluoroscope and an image intensifier. The intensifier amplifies the x-ray image so that it is bright enough to be recorded by a camera. A television camera attached to the image intensifier converts the image into the appropriate video format for videorecording. Prior to videofluoroscopic recording the talker's tongue was coated with barium to enhance visibility of the tongue surface. During recording, the talker stood between the fluoroscope and the image intensifier; she was instructed to avoid unnecessary movements while talking.

Practice stimuli. Subjects were shown 10 practice sentences so they would be familiar with a speechreading task and the test sentence structures. The practice stimuli were taken from a different list of the Repeated Sentence Frame Test (Tye-Murray, 1992b), so were similar in structure to the test items, but not identical. A second female talker recorded the practice stimuli. The talker's head and shoulders were recorded with the videorecorder camera from a frontal view. From this perspective, the talker's lips, jaw, teeth, cheeks, and sometimes the tongue tip were visible. Subjects did not practice speechreading videofluoroscopic images.

Procedure. Subjects were tested individually in a darkened room, seated within arm's reach of a Sony Trinitron (Model KW-1926RA) television (19" screen). Prior to testing subjects completed a practice session. The audio channel was turned off so that only visual information was available. Subjects were instructed to repeat each sentence verbatim as best they could. When necessary, the experimenter paused the videotape between the sentences to allow the subject adequate time to respond. No feedback was provided.

Immediately following the practice session, subjects saw each film medium twice in alternating order (e.g., video, videofluoroscopy, video, videofluoroscopy). Two presentations were completed so that we could assess whether or not subjects learned the test materials with repetition. The auditory signal was never presented. One half of the subjects began by seeing the video film and one half began by seeing the videofluoroscopy film. Subjects were instructed to respond by repeating each sentence as best they could. The experimenter circled the keywords that were correctly repeated on a printed score sheet. Subjects received a 5 minute break midway through the experiment. Approximately 50 minutes were required to complete the practice and test sessions.

Subjects received a brief orientation to the videofluoroscopic image (but no training) prior to the first presentation of the videofluoroscopic films. Three sequential single frames were randomly chosen from the test film. For each frame, the experimenter labeled various anatomical structures, including the upper and lower lips, the maxilla, the mandible, the tongue dorsum, blade and tip, and the upper and lower teeth.

Results

The average scores for each film medium, as a function of presentation (first or second), appear in Table 1. Performance was generally poor for both the video and videofluoroscopic test lists, with the average scores never exceeding 7% keywords correct. Scores for the video presentations ranged between 0% and 17% keywords correct. Scores for the videofluoroscopy presentations ranged between 0% and 11% keywords correct. A two-factor (medium, presentation) repeated measures analysis of variance indicated a significant difference between the videofluoroscopic and video media ($F(1, 13)=5.52$, $p=.035$), but no effect of presentation ($F(1, 13)=3.07$, $p=.10$) and no interaction ($F(1, 13)=0.56$, $p=.47$). The means in Table 1 show that subjects performed better with video than videofluoroscopy. They did not learn the test items with repeated presentations.

Table 1.
Means and standard deviations for Experiment 1 (N =14 subjects with normal hearing).

Presentation	Medium: Video (% keywords correct)	Medium: Videofluoroscopy (% keywords correct)
First	5.5 (SD=3.8)	3.1 (SD=3.2)
Second	7.1 (SD=6.2)	5.4 (SD=3.9)

Discussion

Subjects showed no advantage in viewing a videofluoroscopic image as they attempted to recognize sentence materials visually. Several possibilities may account for this finding. First, the subjects in this experiment were not experienced speechreaders. They may not know how to attend to visually presented articulatory information. Experienced speechreaders may better utilize visual information about articulatory gestures.

The second possibility relates to the poor performance of subjects with both video and videofluoroscopy. Subjects may have to perform above a floor level before an advantage for visible tongue activity (and velar activity) becomes apparent.

The third possibility is that subjects in Experiment 1 were unaccustomed to viewing videofluoroscopic images. Experienced viewers and viewers who receive training in speechreading with videofluoroscopy might score better with videofluoroscopy than with video.

Finally, making tongue activity visible might enhance speechreading performance. However, in our experimental paradigm, this enhancement might be masked. Watching videofluoroscopic films is like watching moving x-ray pictures. Having to view the lips and jaw "in shadow" rather than in flesh (as with video) might introduce a performance decrement that masks any advantage that occurs with the addition of typically nonvisible articulatory activity.

Experiments 2-4 were designed to test these possibilities.

Experiment 2

The purpose of Experiment 2 was to determine whether an advantage for videofluoroscopy occurs when performance is above a floor level. Cochlear implant users, who are experienced speechreaders, viewed and heard the sentences so performance would be better than that recorded in Experiment 1.

Subjects. Ten cochlear implant users, five female, served as subjects. Six of the subjects wore the Richards Ineraid cochlear implant (described by Dorman, Dankowski, McCandless, & Smith, 1989) and four wore the Cochlear Corporation Nucleus cochlear implant (described by Seligman, 1987). Subjects were assigned a number, preceded by the letters CI (cochlear implant). Subject data appear in Table 2. Subjects had a hearing impairment for at least 30 years prior to participating in this experiment.

Table 2.
Profile of cochlear implant users who participated in Experiment 2.

Subject	CI*	Age (YRS)	Length of Cochlear Implant Use (MOS)	Age at Hearing Loss Onset (YRS)	Duration of Profound Deafness (YRS)
CI1	I	69	54	20	44
CI2	N	42	12	congenital	40
CI3	I	38	30	4	16
CI4	N	56	54	8	22
CI5	I	70	30	15	10
CI6	I	53	30	4	15
CI7	N	56	18	22	28
CI8	I	73	18	33	25
CI9	I	69	54	28	7
CI10	N	46	30	16	25

CI* = cochlear implant type: I = Ineraid; N = Nucleus

All subjects receive benefit from their cochlear implant. Subjects completed the Iowa Medial Consonant Confusion Test (Tyler, Preece, & Tye-Murray, 1986) in an audition-only, vision-only (filmed with video), and an audition-plus-vision condition within one year of their participation in the current study. The auditory signal was presented at 73 dBA SPL measured at the level of the cochlear implant microphone. This corresponds to a moderately loud conversational level. These scores are presented in Table 3 to illustrate the benefit that these subjects obtain from using information in the three modalities. With one exception, all subjects scored above chance in all conditions. CI7 did not score above chance in the audition-only condition.

Table 3.
Scores on the Iowa Medial Consonant Confusion Test* for the cochlear implant users who participated in Experiment 2.

	(% Consonants Correct)		
Subject	Audition-Only	Vision-Only	Audition-Plus-Vision
CI1	23%	18%	54%
CI2	28%	56%	69%
CI3	32%	37%	63%
CI4	39%	46%	74%
CI5	39%	28%	54%
CI6	55%	65%	69%
CI7	3%	22%	55%
CI8	32%	18%	62%
CI9	49%	28%	76%
CI10	41%	28%	69%

Chance performance = 8% consonants correct

Test and Practice Stimuli. The test and practice stimuli were the same stimuli used in Experiment 1.

Procedures. The procedures were like those used in Experiment 1 with one exception. Subjects heard the auditory portion of the signal while viewing the films. The signal was presented at 73 dBA SPL measured at the cochlear implant earpiece microphone for both the video and videofluoroscopy media. Noise was inherent on the videotape secondary to recording of the films by the room environment and/or the videofluoroscopy camera and/or the dubbing equipment. The signal-to-noise ratio for both the videofluoroscopy and video media was +14 dB. All sound pressure level measures were made with a Bruel and Kjaer Precision Sound Level meter (model 2203) and a Bruel and Kjaer 1/2 inch condenser microphone (type 4132), held at the level of a cochlear implant earpiece microphone of a subject seated one arm's reach from the television monitor. Sound pressure measurements (dBA) were obtained of each stimulus sentences and the inherent noise between each sentence for both the videofluoroscopy and the video conditions. An average speech (signal) level and an average non-speech (noise) level was obtained for each. The average speech level was 73 dBA SPL for each condition and the average noise level was 59 dBA SPL for each condition.

After subjects were oriented to the videofluoroscopic still frames, they were required to label each structure with no prompting or cuing from the clinician.

Table 4.
Means and standard deviations for Experiment 2 (N = 10 cochlear implant users).

Presentation	Medium: Video	Medium: Videofluoroscopy
First	41.7 (SD=26.4)	27.4 (SD=29.1)
Second	48.9 (SD=28.7)	38.9 (SD=27.3)

Results

The results appear in Table 4. Scores for the video medium ranged between 0% and 88% keywords correct. Scores for the videofluoroscopic medium ranged between 0% and 89% keywords correct. A two-factor (medium, presentation) repeated measures analysis of variance showed a significant effect of medium ($F(1,9)=9.7$, $p=.013$) and presentation ($F(1,9)=28.9$), $p=.0004$), and no interaction ($F(1,9)=1.37$, $p=.27$). The video medium yielded better word recognition than the videofluoroscopy medium and word recognition scores improved with list repetition.[1]

It is possible that those cochlear implant users who are able to understand audio, visual, and/or audiovisual speech signals fairly well might perform better with the videofluoroscopy medium than those who are unable to do so. They might have special skill in abstracting information from a degraded or unusual speech signal. To evaluate this possibility, we computed Pearson correlation coefficients between scores from the Iowa Medial Consonant Confusion Test presented in an audition-only, vision-only, or audition-plus-vision condition (Table 3) with scores obtained during the first presentation of the videofluoroscopy Repeated Sentence Frame Test items. None of the three correlations were significant ($p>.05$). However, scores from the first presentation of the videofluoroscopy and the first presentation of the video Repeated Sentence Frame Test items were highly correlated ($r=.868$, $p<.001$). The difference in findings might relate to the difference between recognizing nonsense syllables and sentence materials. However, Tyler, Tye-Murray, and Gantz (1987) have shown that a strong relationship exists between cochlear implant users' ability to recognize nonsense syllables and their ability to recognize words in sentences. In addition or alternatively, the difference might relate to the fact that the same items were presented in the videofluoroscopic and video media conditions in the present experiment, which may have strengthened the relationship between the Repeated Sentence Frame Test scores. Because of the different findings, it is not clear whether an ability to recognize speech in a vision-only, audition-only, or

[1]We were concerned about the apparent skewness of the scores. For this reason, we repeated the analyses on the logs of the scores. The conclusions with these analyses are the same.

audition-plus-vision condition corresponds with an ability to recognize speech presented with videofluoroscopy.

Experiment 3

Watching videofluoroscopic films for the purpose of abstracting speech information is a novel task for most persons. In Experiment 3, the element of novelty was removed by employing four experienced viewers as subjects. Two of the four subjects also received practice in speechreading x-ray images.

Subjects. Subjects were four persons with normal hearing who have had extensive experience with videofluoroscopy or cinefluorography. Cinefluorography presents an x-ray image much like that presented with videofluoroscopy. The primary difference of import to the present experiment is that cinefluorography employs a faster film rate during filming, usually 100 or 150 frames per second. This rate provides good resolution for data analysis purposes, but still provides real-time motion.

Subjects were each assigned a subject number preceded by the letter S, by which they will be identified in this report. All subjects have earned a doctorate degree in speech and hearing sciences. Three subjects, S1, S2, and S3, are clinical speech-language pathologists who use real-time videofluoroscopy three or four times per week to diagnose swallowing disorders. Collectively, these three subjects have viewed more than 1500 videofluoroscopic films, with a range of 100 to 1200 films per subject. The fourth subject, S4, has been involved with cinefluorographic speech research for more than 30 years. Subjects were paid for their participation.

Procedures. The procedures were identical to those for Experiment 1. In addition, S3 and S4 completed additional procedures approximately three months later. These two subjects received training in speechreading x-ray images, and then were retested.

Training for S3 and S4 lasted 4 hours, and was provided over the course of 5-7 days. Two videotapes were constructed for training purposes. Each videotape contained a different randomization of the same 31 phrases or sentences, recorded with cinefluorography. The training items were extracted from cinefluorographic recordings that had been obtained during experimental investigations of normal speech production. The training items were transferred to videotape. A total of four different talkers produced training items, two of whom were female. An example of a training phrase item is, subjects are twelve; an example of a training sentence item is, I saw him eat soup too. None of the test items were practiced during training. All sentences were produced by talkers who had no speech or hearing problems.

S3 and S4 performed two different types of tasks throughout their training program. In the first task, subjects received a training item, and attempted to repeat it. They then read the written text of the item, and received it again. They performed this task both in a vision-only and an audition-plus-vision condition. In the second task, they read the corresponding text before the item was presented, repeated it aloud, received it, and then re-read the text. This task was performed in both a vision and audition-plus-vision condition.

Results

In general, subjects performed poorly with both video and videofluoroscopy. They performed better with video. S1 recognized only three keywords (of a possible 98 keywords) during the two video presentations, and only one keyword during the two videofluoroscopy presentations. S2 recognized eight keywords during the two video presentations, and 5 keywords during the videofluoroscopy presentations. S3 recognized seven keywords during the two video presentations and no keywords in the videofluoroscopy presentations. S4 recognized seven keywords during the two video presentations and three keywords during the videofluoroscopy presentations.

S3 did not change her performance very much following training. S3 recognized five keywords during the two videofluoroscopy presentations following training. Thus, she improved from 0% keywords correct before training to 4% keywords correct following training. S4 showed virtually no change. He recognized four keywords during the videofluoroscopy presentations following training, one more than before training.

Experiment 4

Watching videofluoroscopic films is not equivalent to watching a video image. The lips, jaw, and cheek appear in shadows as opposed to solid flesh. It may be that visible tongue activity can be useful, but that the paradigm of the preceding experiments does not allow us to document this advantage. Having to view lip, jaw, and cheek activity in shadows may introduce a disadvantage to the speechreading task, and may mask any advantage that arises from making tongue activity visible.

In Experiment 4, we introduced a new talker who has metal fillings in almost every lateral tooth. When this talker is filmed with videofluoroscopy, her metal fillings shield her tongue. Only fleeting glimpses of it occur on the film record, as when the talker lowers her jaw extensively and her tongue moves in the space between upper and lower teeth.

As in Experiment 2, we used cochlear implant users and an audiovisual condition to avoid a floor effect. Subjects viewed test materials used in the previous experiments, and also test materials filmed by the second talker. By comparing scores for the two sets of videofluoroscopic test materials, we could surmise the contribution of visible tongue activity to speechreading performance because both sets presented the (potentially) deleterious effects of viewing the articulators in shadow. Any performance difference could be attributed to tongue visibility.

Subjects. Ten cochlear implant users, five female, served as subjects. Eight of the subjects wore the Richards Ineraid cochlear implant and two wore the Cochlear Corporation Nucleus cochlear implant. None of the subjects had participated in Experiment 2. Table 5 presents subject data. All subjects had experienced hearing impairment for at least 25 years.

All subjects receive benefit from their cochlear implant. Subjects completed the Iowa Medial Consonant Confusion Test (Tyler et al., 1986) in an audition-only, vision-only (video), and an audition-plus-vision mode within one year of their participation in the current study. These scores are presented in Table 6 to illustrate the benefit that these subjects obtain from using information in the three modalities. All subjects performed above chance, although this was only marginally true for CI11 in an audition-only condition.

Test and Practice Stimuli. The test and practice stimuli included those used in Experiment 1. In addition, a second female talker with general American dialect recorded a different list from the Repeated Sentence Frame Test (Tye-Murray, 1992b) with both videofluoroscopy and video. The sentences are presented in Appendix B. The list has 46 keywords. This second talker has metal fillings in both upper and lower teeth along the lateral aspects of her oral cavity. The two talkers were assigned the letters V (visible tongue activity) and NV (nonvisible tongue activity). Photo 1 presents a single frame from talker V's film and Photo 2 presents a single frame from talker NV's film. (See center-bound plate).

Table 5.
Profile of cochlear implant users who participated in Experiment 4.

Subject	CI*	Age (YRS)	Length of Cochlear Implant Use (MOS)	Age at Hearing Loss Onset (YRS)	Duration of Profound Deafness (YRS)
CI11	N	63	18	14	26
CI12	I	37	42	3	10
CI13	I	64	42	16	22
CI14	N	65	18	40	13
CI15	I	55	54	21	15
CI16	I	63	54	32	2
CI17	I	59	30	34	4
CI18	I	28	30	congenital	25
CI19	N	72	30	13	1
CI20	I	36	36	7	1

*CI = cochlear implant type: I = Ineraid; N = Nucleus

Table 6.
Scores on the Iowa Medial Consonant Confusion Test*
for the cochlear implant users who participated in Experiment 4.

Subject	Audition-Only	(% Consonants Correct) Vision-Only	Audition-Plus-Vision
CI11	14%	32%	39%
CI12	73%	46%	94%
CI13	32%	28%	54%
CI14	35%	31%	59%
CI15	72%	37%	92%
CI16	76%	35%	87%
CI17	55%	42%	77%
CI18	31%	32%	74%
CI19	54%	30%	92%
CI20	89%	44%	99%

Chance performance = 8% consonants correct

Procedures. The procedures were like those used in Experiment 2, with the following exceptions. Subjects saw the first presentation of the video and videofluoroscopic films of one talker, and then the first presentation of the two films of the second talker. Then they saw the second presentation of the two films for each of the two talkers. One half of the subjects began with the videofluoroscopy medium and one half began with the video medium. Subjects also alternated in which talker they saw first. They received a 5 to 10 minute break in between talker presentations. All conditions presented the audio signal calibrated to 73 dBA SPL. Sound pressure measurements were obtained in the same manner as that described in Experiment 2. The signal-to-noise ratio for both the videofluoroscopy and the video media of talker V was +14 dB (see Experiment 2). The signal-to-noise ratio for the videofluoroscopy condition for talker NV was +10 dB. The signal-to-noise ratio for her video condition was +13 dB.

Results

The results appear in Table 7. For talker V, scores for the two media ranged from 4% to 98% for video and from 0% to 96% keywords correct for videofluoroscopy. Scores ranged from 4% to 98% and from 5% to 86% keywords correct for talker NV for the two media, respectively. A three-factor (talker, medium, presentation) repeated measures analysis of variance indicated significant effects of presentation ($F(1, 8)=24.8$, $p=.001$) and medium ($F(1,8)=22.5$, $p=.002$). The talker effect was not significant ($F(1, 8)=3.2$, $p=.11$), and there were no interactions.

These results indicate that performance was better with the video than videofluoroscopy medium, and that subjects improved their performance with repeated presentation of the test items. In the videofluoroscopy condition, subjects performed as well with the tongue visible as when the tongue was not visible. This finding suggests that there is no advantage to making tongue body activity visible at least via videofluoroscopy.

Table 7.
Means and standard deviations for Experiment 4 (N =10 cochlear implant users).
(V = tongue activity is visible; NV= tongue activity is not visible.)

Presentation	Talker	Medium: Video (% keywords correct)	Medium: Videofluoroscopy (% keywords correct)
First	V	46.2 (SD=25.8)	23.3 (SD=26.9)
	NV	55.4 (SD=24.6)	36.6 (SD=32.7)
Second	V	56.3 (SD=26.9)	30.8 (SD=22.9)
	NV	64.1 (SD=21.5)	39.4 (SD=24.7)

Discussion

One might argue that finding no significant difference between the two videofluoroscopic conditions does not allow us to conclude that tongue body information fails to enhance speechreading performance. The present comparison was confounded by two uncontrolled variables, different talkers and different sentences. However, the absence of a talker effect in the statistical analysis suggests that the two lists of the Repeated Frame Sentence Test and the two talkers are equally difficult to speechread under real-world circumstances. If they are not, then a talker effect for the video condition should have occurred.

Experiment 5

In the first four experiments, we assessed how well subjects perform in a vision-only or an audition-plus-vision condition. In this final experiment, we determined whether videofluoroscopy could enhance cochlear implant users' ability to recognize speech signals in an audition-plus-vision condition. Subjects were tested in an audition-only condition then in an audition-plus-vision condition. Speechreading enhancement was determined by comparing scores from the two conditions. Since speechreading aids will usually serve as a supplement to audition, the results of this experiment speak most directly to the value of incorporating visible tongue behavior into a speechreading aid.

Table 8.
Profile of cochlear implant users who participated in Experiment 5.

Subject	CI*	Age (YRS)	Length of Cochlear Implant Use (MOS)	Age at Hearing Loss Onset (YRS)	Duration of Profound Deafness (YRS)
CI21	I	32	30	24	7
CI22	I	72	67	13	3
CI23	N	45	30	13	29
CI24	N	66	30	43	20
CI25	N	34	42	29	1
CI26	I	72	67	51	8
CI27	N	36	72	20	8
CI28	I	71	30	49	19

*CI = cochlear implant type: I = Ineraid; N = Nucleus

Methods

Subjects. Nine cochlear implant users, five female, served as subjects. None of the subjects had participated in Experiment 2 or 4. Five subjects wore the Nucleus cochlear implant and four wore the Ineraid device. Biographical data are presented in Table 8 and audiological data are presented in Table 9. Subjects had hearing loss for at least 5 years prior to testing. All subjects performed above chance on the Iowa Medial Consonant Test in the three test conditions (audition-only, vision-only, and audition-plus-vision).

Methods. The methods for this experiment were identical to those described in Experiment 2, with the following exception. Subjects were tested in both an audition-plus-vision condition (as in Experiment 2), and in an audition-only condition, with both the video and videofluoroscopy media.

Table 9.
Scores on the Iowa Medial Consonant Confusion Test* for the cochlear implant users who participated in Experiment 5.

Subject	Audition-Only	(% Consonants Correct) Vision-Only	Audition-Plus-Vision
CI21	56%	47%	71%
CI22	15%	37%	53%
CI23	55%	45%	90%
CI24	58%	42%	87%
CI25	83%	44%	94%
CI26	54%	37%	80%
CI27	50%	47%	95%
CI28	60%	35%	94%

Chance performance = 8% consonants correct

Results

On average, subjects scored 58.0% keywords correct (SD = 22.7) with the video medium and 37.8% keywords correct (SD = 24.4) with the videofluoroscopy medium. They scored 10.9% keywords correct (SD = 16.6) when the test items were presented with audition only. The average difference in scores obtained with the video medium (which was presented in an audition-plus-vision condition) and the audition-only signal was 47.1% (SD = 16.7). The average difference between the videofluoroscopy medium and the audition-only signal was 26.9% keywords correct (SD = 14.2). A paired comparison t-test indicated that the enhancement scores obtained with the video medium were significantly greater than those obtained with the videofluoroscopy medium ($t = 4.29$, $\underline{p} < .003$).

Discussion

The videofluoroscopic image provided some enhancement of the auditory signal, but not as much as the video image. The present results do not allow us to determine whether enhancement with videofluoroscopy occurred because (a) subjects attended to the same (possibly degraded) information (e.g., lip and jaw displacement patterns) that they attend to when they view video or (b) because they attended to articulatory information that is typically obscured, such as tongue body activity. The results from Experiment 4 would suggest that the former interpretation is correct. Interestingly, an estimated 50% of all cochlear implant subjects (in Experiments 2, 4, and 5) spontaneously volunteered similar comments to their clinician (the third author) following testing. They complained that the tongue activity displayed in the videofluoroscopic films was "distracting", and gave them difficulty in "speechreading the words". None of the subjects electively commented that the jaw or lips were difficult to see nor difficult to recognize as the structures they are.

General Discussion

The results from the five experiments suggest that making visible supralaryngeal articulator activity that is typically not available visually by means of videofluoroscopy does not enhance speechreading performance. Subjects always performed better with the video medium than videofluoroscopy medium (Experiments 1, 2, 3, and 5). Subjects performed as well when the tongue was not visible in the videofluoroscopy records than when it was visible (Experiment 4). The videofluoroscopic image provided less enhancement of a degraded auditory signal than did a video image (Experiment 5). Training did not affect subjects' ability to recognize speech presented with videofluoroscopy (Experiment 2). These results provide little motivation for the development of articulatory-based speechreading aids.

The conclusion that increasing information about available articulatory behavior does not enhance speechreading performance might be tempered with four considerations. First, it may be that a different representation of the articulators, one in which the articulators appear more as solid mass than x-ray images, might yield different results. Previous investigators have found that some latitude exists in the verisimilitude of the visual display, but that boundaries exist as to how abstract the display may be and still remain helpful. For instance, visual representations of lip movement, such as oscilloscope patterns that modulate according to the ongoing speech signal, can convey linguistic information. This is true even when viewers receive no training (Boston, 1973; Erber, 1979). Summerfield (1979) found that lips painted with luminous make-up improved subjects' ability to recognize speech in noise, although not quite as much as a video display of the entire face. Summerfield also found, however, that when subjects viewed only four 3 mm luminous dots painted on the lips, with the remaining face not visible, they received no benefit.

There is some reason to minimize the importance of this consideration. In the present set of experiments, the tongue body was clearly outlined in the videofluoroscopic display, and had been coated with barium. Ongoing lip activity was clearly visible, and could not be confused with the rigid teeth. The tongue did not appear as an abstraction or as an approximation of true tongue activity. Perhaps different methodology may allow us to state the present conclusions more strongly. In future research, we might superimpose the videofluoroscopic image on th simultaneous video image. In this way, we can retain the exact visual information that is typically available, plus provide information about tongue body activity. If this combined video-videofluoroscopy presentation provides no improvement over the video-only presentation, then our present conclusion will be strengthened.

A second consideration that might temper the present conclusion relates to viewing angle. It is possible that complementing a 0° viewing angle rather than a 90° angle might yield different results. Most individuals, even those who use cochlear implants, rarely speechread a talker who speaks in profile. If somehow additional articulatory information could be provided while the subject viewed the talker head-on, speechreading performance might improve.

This second consideration may also not be important. Although speechreading a talker from a 90° angle is generally thought to be more difficult than speechreading a talker from a 0° angle (Erber, 1974; Neely, 1956; but see IJsseldjk, 1992, who shows no difference for a head-on viewing angle and a 60° viewing angle), some speech can be recognized visually when the talker speaks in profile. For example, Erber (1974) reported percent correct scores of about 64% for words spoken in isolation, at a viewing distance of 6 feet. In the present investigation, many subjects recognized some words in sentences that were filmed with video, although performance in a vision-only condition was generally poor (never exceeding 7% keywords correct, on average).

The third consideration that might temper the present conclusion is that making available forms of nonvisible information other than that provided in the present investigation might prove beneficial. For instance, information about voicing might be presented visually. Speechreaders recognize more speech when they see the talker and hear just the fundamental frequency pattern, than when they only see the talker (Fourcin, Rosen, Moore, Douek, Clark, Dodson, and Bannister, 1979; Rosen, Fourcin, and Moore, 1981). Presenting fundamental frequency information through the tactile modality with a tactile aid can also improve speechreading performance, although not as much as presenting the information acoustically (Boothroyd, 1989).

Perhaps a visual display of fundamental frequency information, as with Upton's (1968) eyeglass display, might enhance speechreading performance. Although Pickett et al. (1974) experimented with a similar apparatus and did not find the results promising enough to pursue its development, recent research suggests potential benefit. Kunov and Ebrahimi (1991) asked eight subjects to speechread consonant-vowel nonsense syllables while wearing a speechreading eyeglass aid that codes voicing, high-frequency energy, and signal envelope cues. The aid enhanced speechreading performance for at least six subjects, after only 1.5 to 2.5 hours of practice.

Finally, the fourth consideration is that more extensive training (more than 4 hours) might result in different findings than those presented in Experiment 3. The subjects who received training, although familiar with moving x-ray images, had never viewed videofluoroscopic films for the purpose of speechreading prior to their participation in the third experiment. It is not unusual for individuals who use tactile aids to receive up to 50 hours of training before they demonstrate enhanced speechreading performance (Reed, Durlach, Delhorne, Rabinowitz, and Grant, 1989). Two sets of experimental findings diminish the importance of this consideration.

First, subjects can often recognize articulatory-based information without training (Erber, 1979; Boston, 1973). For example, Fowler and Dekle (1991) allowed subjects to experience syllables by means of the Tadoma method, wherein they placed a hand on a talker's face and felt lip and jaw movement. While haptically experiencing a syllable, subjects simultaneously heard a different syllable. The results revealed cross-modal influences, even though the subjects did not receive training for the Tadoma method and had not perceived speech haptically prior to the experiment. For instance, an audition-only ga syllable taken from a ba-ga continuum was less ambiguously heard when the subjects felt ga rather than ba.

The second set of findings, tending to negate the fourth consideration, suggests that training has minimal effect on vision-only speech recognition performance (Binnie, 1977; Heider and Heider, 1940; Lesner, Sandridge, and Kricos, 1987; but see Dodd, Plant, and Gregory, 1989; Walden, Erdman, Montgomery, Schwartz, and Prosek, 1981; and Walden, Prosek, Montgomery, Scherr, and Jones, 1977, who report small gains). These findings suggest that subjects may not respond to training aimed toward improving how well they utilize articulatory information presented in a vision-only modality.

In summary, these experiments do not indicate that providing typically nonvisible articulatory information (at least via videofluoroscopy) is beneficial to the speechreader. Of course, the issue is not a closed book, and future researchers employing different experimental procedures might reach different conclusions. Nonetheless, the present evidence weighs against finding benefit.

At least two theories of speech perception, direct-realist theory (Fowler, 1986) and the motor theory (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Liberman and Mattingly, 1985), posit that perceivers recover information about articulatory gestures. In direct-realist theory, articulatory behavior is thought to be uniquely specified by the spectral and temporal patterning of the acoustic and visual signals. Perceivers are capable of recognizing phonetic gestures by attending to the signals. Fowler and Rosenblum (1991) define a phonetic gesture as an organized movement of the articulators that achieves a phonetically relevant goal, such as when the jaw and lips achieve closure for a bilabial stop consonant. When perceivers attend to spoken messages, they are attuned to articulatory gestures and not acoustic segments per se. In fact, the gestures might be conveyed acoustically, visually, or both. As such, "... it matters little through what sense we realize what speech event has occurred" (Fowler, 1986, p. 9; also, Summerfield, 1979). The level of description that best characterizes the objects of perception is not straightforward. In a critique of direct-realist theory, Summerfield (1987) notes that kinematic parameters, such as articulator displacements and velocities, might not be abstracted by perceivers. Rather, the dynamic structure underlying the phonetic gesture might be of relevance, such as the forces that produce the velocity patterns (p. 46).

The motor theory preceded direct-realist theory (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Liberman and Mattingly, 1985). Consistent with direct-realist theory, Liberman and Mattingly (1985) suggest that, "Phonetic perception is perception of gesture" (p. 21). However, the motor theory differs as to what level the perceptual objects are best described. Rather than the kinematic or dynamic structure of the articulatory gestures, perceivers may recognize the talker's speech intent: "Instances of a particular gesture always have certain topological properties not shared by any other gesture ... the gestures do have characteristic invariant properties, as the motor theory requires, though these must be seen, not as peripheral movements, but as the more remote structures that control the movements. These structures correspond to the speaker's intentions. (pg. 23)."

The present results somewhat diminish the strength of theories that claim articulatory gestures are perceived. The argument is as follows. Providing information about articulatory behavior that is typically not available visually did not enhance speechreading performance; in fact, it decreased performance. As such, perceivers appear not to recover nonvisible articulatory gestures during the process of recognizing speech. It is quite possible that the two-dimensional videofluoroscopic display does not adequately capture the kinematics, dynamics, and/or talker's speech intent in a way that can be meaningful to perceivers and that would allow them to recognize the objects of speech perception.[2] For example, Fletcher (1992, pp. 84-89) argues that x-ray data do not capture the fact that vowels are distinguished by linguapalatal contact patterns.[3] Videofluoroscopy also does not provide much evidence of cheek musculature activity. Nonetheless, the present findings pose a challenge to any theory based on the premise that speech perceivers must recover articulatory behavior to abstract meaning (see Diehl and Kluender, 1989; Diehl, Walsh, and Kluender, 1991, for in-depth criticism).

Additional research involving videofluoroscopy might be useful for further evaluating direct realist theory and motor theory. Speech samples could be selected that allow information transmission analysis (Miller & Nicely, 1955) and sequential information analysis (SINFA, Wang & Bilger,

[2]Perhaps one shortcoming of both direct realist theory and motor theory is that they are extremely difficult to test. Articulatory gestures can be conceived in many different ways, and conceptions can be easily amended to accommodate a wide variety of experimental findings. MacNeilage (1991) expresses a similar sentiment: "... how many gestures are there in a given segment, and how do we identify them? ... I have no objection to the use of the notion of gesture as a place-holder, expressing the conviction that at some level speech production must have a finite set of control entities. However, at the present level of development of the concept, I do not see how it can have explanatory status in a view of how *perception* [italics his] occurs." (p. 69).

[3]However, visual displays of linguapalatal contact patterns do not enhance persons' ability to discriminate vowel pairs that are simultaneously presented with the Tadoma method (Reed et al., 1992).

1973). These statistical procedures indicate how well perceivers utilize speech features, such as place, nasality, and friction. Information transmission analysis and SINFA performed on subject responses might indicate whether place of articulation (e.g., alveolar versus velar for consonants; front versus back for vowels) and nasality are conveyed better when tongue body and velum activity are visible rather than not visible. It is not impossible that information transmitted for the place and nasality features will be better for a videofluoroscopic condition than a video condition, even if overall percent of stimuli correct scores are better in the video condition, as we found in the present experiments. Such results might be interpreted as supporting the two theories since information about tongue body and velum would have improved just those items that might be most expected to improve.

In a review of her research with congenitally deaf talkers, Tye-Murray (1992c; also, 1984) suggested that perceivers are aware of and attentive to constraints placed upon the speech signal, constraints that exist by virtue of having been produced by a human vocal tract. The constraints stem both from the anatomy and physiology of the vocal tract, and from how talkers learn to marshal their articulators together for speech production (e.g., they extend and flex the tongue body abundantly, they move the articulators continuously from one open gesture to the next, they manage the breath stream efficiently). The constraints limit how the acoustic signal may be patterned, and perceivers expect these limitations. However, listeners need not recognize articulatory gestures. An apt analogy is what happens when listeners attend to music produced by a trombone. The physical structure of the trombone and the technique that the trombone player employs in maneuvering the shaft and creating an air pressure source constrains any sound that is produced. Listeners perceive individual melodic notes within the context of the global patterning that results from these physical and functional constraints. However, in perceiving a particular note, listeners need not recover how the trombone was configured nor how the player pursed her lips and contracted her rib cage as she produced it.

Articulatory activity that is typically visible may enhance speech recognition because it lawfully corresponds to information in the acoustic signal, and vice versa. For example, the lips visibly achieve closure every time a talker produces a syllable beginning with /p/. It may also provide limited information about articulatory organizational strategies. For example, visible articulatory activity might provide information about the rhythmic alternation of opening and closing gestures but probably not about how the tongue body extends and flexes (see Tye-Murray, 1991). We suggest that making typically obscured articulatory events visible via videofluorography is not helpful because these events are foreign to our experience.[4] As such, recovering the events may hold little significance for the process by which speech perceivers abstract linguistic meaning from a talker's actions.

[4]It does seem possible that information that is typically available through vision can be provided through another modality, provided that it can adequately convey the information. For instance, the Tadoma method can provide information about lip and jaw activity that is usually available through vision, and therefore can enhance speech recognition (Fowler and Dekle, 1992). We would predict, however, that haptic information about tongue displacement during speech production, say if perceivers were to feel the tongue body as a talker spoke he, heh, and hay, would not influence speech perception performance because normally they have limited access to this information visually. This prediction aligns with the results of Experiment 1, where the normal hearing listeners did not improve their speechreading performance when they viewed tongue body activity.

Acknowledgments

This study was supported by research grant DC00242 and DC00976-01 from the National Institutes of Health/NIDCD, grant RR59 from the General Clinical Research Centers Program, Division of Research Resources, NIH, and the Iowa Lions Sight and Hearing Foundation. We thank Kenneth Moll for his suggestions concerning the procedures for Experiment 3, and George Woodworth for statistical consultation, and our four experienced subjects for their participation in Experiment 3.

Appendix A

Sentences included in the first list of the Repeated Sentence Frame Test (Tye-Murray, 1992b). The sentences were presented in random order, but are arranged in sets of four here. Each set corresponds to one sentence frame. Keywords are underlined.

1. The <u>queen</u> is <u>nice</u>.
2. The <u>king</u> is <u>mean</u>.
3. The <u>nurse</u> is <u>fat</u>.
4. The <u>dog</u> is <u>bad</u>.
5. Her <u>mom</u> is in the <u>bedroom</u>.
6. Her <u>friend</u> is in the <u>backyard</u>.
7. Her <u>dad</u> is in the <u>kitchen</u>.
8. Her <u>aunt</u> is in the <u>bathroom</u>.
9. <u>Grandma</u> is eating a <u>hotdog</u>.
10. <u>Mother</u> is eating a <u>cookie</u>.
11. <u>Grandpa</u> is eating a <u>pancake</u>.
12. <u>Father</u> is eating an <u>apple</u>.
13. The <u>birthday cake</u> and <u>plates</u> are on the table.
14. The <u>potatoes</u> and <u>forks</u> are on the table.
15. The <u>bananas</u> and <u>bowls</u> are on the table.
16. The <u>hamburgers</u> and <u>knives</u> are on the table.
17. I <u>want</u> the <u>soup</u>.
18. I <u>cook</u> the <u>eggs</u>.
19. I <u>need</u> the <u>meat</u>.
20. I <u>like</u> the <u>peas</u>.
21. Point to your <u>hair</u> and <u>nose</u>.
22. Point to your <u>arms</u> and <u>face</u>.
23. Point to your <u>hands</u> and <u>ears</u>.
24. Point to your <u>eyes</u> and <u>feet</u>.

Appendix B

Sentences included a second list of the Repeated Sentence Frame Test (Tye-Murray, 1992b). The sentences were presented in random order, but are arranged in frame here. Keywords are underlined.

1. The <u>boat</u> is <u>fast</u>.
2. The <u>train</u> is <u>slow</u>.
3. The <u>plane</u> is <u>big</u>.
4. The <u>bike</u> is <u>small</u>.
5. A <u>balloon</u> is on the <u>chair</u>.
6. A <u>baseball</u> is on the <u>desk</u>.
7. A <u>pencil</u> is on the <u>floor</u>.
8. The <u>cat</u> is looking at the <u>moon</u>.
9. The <u>bear</u> is looking at the <u>tree</u>.
10. The <u>frog</u> is looking at the <u>bug</u>.
11. <u>Ten birds</u> are in the sky.
12. <u>Six clouds</u> are in the sky.
13. <u>Eight kites</u> are in the sky.
14. The <u>silly</u> dog is <u>sleeping</u>.
15. The <u>funny</u> dog is <u>playing</u>.
16. The <u>ugly</u> dog is <u>sitting</u>.
17. The <u>pretty</u> dog is <u>eating</u>.
18. <u>Pick up</u> your <u>hat</u>.
19. <u>Put down</u> your <u>coat</u>.
20. <u>Put on</u> your <u>shoes</u>.
21. <u>Take off</u> your <u>boots</u>.

References

Binnie, C. A. (1977). Attitude changes following speechreading training. <u>Scandinavian Audiology, 6</u>, 13-19

Boothroyd, A. (1989). Developing and evaluating a tactile speechreading aid. In N. S. McGarr (Ed.) <u>Research on the use of sensory aids for hearing-impaired people</u> (pp.101-112). Wash ington, DC: Alexander Graham Bell Association for the Deaf.

Boston, D. W. (1973). Synthetic facial communication. <u>British Journal of Audiology</u>, 7, 95-101.

Browman, C., & Goldstein, L. (1987). Towards an articulatory phonology. <u>Phonology Yearbook, 3</u>, 219-252

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. <u>Ecological Psychology 1</u> (2), 121-144.

Diehl, R. L., Walsh, M. A., & Kluender, K. R. (1991). On the interpretability of speech/nonspeech comparisons: A reply to Fowler. <u>Journal of the Acoustical Society of America, 89</u> (6), 2905-9.

Dodd, B., Plant, G., & Gregory, M. (1989). Teaching lip-reading: The efficacy of lessons on video. <u>British & Journal of Audiology</u>, 23, 229-238.

Dorman, M. F., Dankowski, K., McCandless, G., & Smith, L. (1989). Identification of synthetic vowels by patients using the Symbion multichannel cochlear implant. Ear and Hearing, 10, 40-43

Erber, N. P. (1974). Effects of angle, distance, and illumination of visual reception of speech by profoundly deaf children. Journal of Speech and Hearing Research, 17, 99-112.

Erber, N. P. (1979). Real-time synthesis of optical lip shapes from vowel sounds. Journal of the Acoustical Society of America, 66(5), 1542-44.

Fletcher, S. G. (1992). Articulation: A physiological approach. San Diego, CA: Singular Publishing Group, Inc

Fourcin, A. J., Rosen, S. M., Moore, B. C. J., Douek, E. E., Clark, G. P., Dodson, H., & Bannister, L. H. (1979). External electrical stimulation of the cochlea: Clinical psychophysical, speech perceptual and histological findings. British Journal of Audiology, 13, 85-107

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics, 14, 3-28.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. Journal of Experimental Psychology: Human Perception and Performance, 17(3), 816-828.

Fowler, C. A., & Rosenblum, L. D. (1991). The perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.) Modularity and the motor theory of speech perception (33-59). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldberg, A. J. (1972). A visual feature indicator for the severely hard of hearing. IEEE Trans On Audio & Electroacoustics, Au 20, No. 1, March 1972, 23-28.

Heider, F., & Heider, G. M. (1940). An experimental investigation of lipreading. Psychological Monographs, 52, 124-133.

Hochberg, J. Laroche, F., Papcun, G. Thomas, T., & Zacks, J. (1991). From phonemes to gestures: Non-linear speech recognition at CNLS. CNLS Newsletter, 67, 1-14.

IJsseldijk, F. J. (1992). Speechreading performance under different conditions of video image, repetition and speech rate. Journal of Speech and Hearing Research, 35, 466-471.

Jeffers, J., & Barley, M. (1974). Speechreading (speechreading). Springfield, IL: Charles C. Thomas.

Kalikow, D. N., & Swets, J. A. (1972). Experiments with computer controlled displays in Second Language Learning. IEEE Trans. On Audio & Electroacoustics, Au 20, No. 1, March 1972, 23-28.

Kent, R. D., & Moll, K. L. (1972). Cinefluorographic analysis of selected lingual consonants. Journal of Speech and Hearing Research, 15, 453-473.

Kunov, H., & Ebrahimi, D. (1991). Measurement of performance of a peripheral vision lipreading aid. Scandinavian Audiology, 20, 131-137.

Lesner, S., Sandridge, S., & Kricos, P. (1987). Training influences on visual consonant and sentence recognition. Ear and Hearing, 8, 283-287.

Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1-36.

MacNeilage, P. F. (1991). Comment: The gesture as a unit in speech perception theories. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), Modularity and the motor theory of speech perception (pp. 61-67). Hillsdale, NJ: Lawrence Erlbaum Associates

McWilliams, B. J., Morris, H. L., & Shelton, R. L. (1990). Cleft Palate Speech (2nd Edition). Philadelphia, PA: B. C. Decker, Inc.

Miller, G. A. & Nicely, P. E. (1955). Analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27, 338-352.

Moll, K. L. (1960). Cinefluorographic techniques in speech research. Journal of Speech and Hearing Research, 3, 227-241.

Neely, K. K. (1956). Effect of visual factors on the intelligibility of speech. Journal of the Acoustical Society of America, 28, 1275-1277.

Papcun, G. Hochberg, Thomas, T., Laroche, F., Zacks, J., & Levy, S. (1992). Inferring articulation and recognizing gestures from acoustics with a neural network trained on x-ray microbeam data. Journal of the Acoustical Society of America, 92, 688-700.

Pickett, J. M., Gengel, R. W., & Quinn, R. (1974). Research with the Upton eyeglass speechreader. Procedures of speech communication seminar, Stockholm, 105-110.

Reed, C. M., Durlach, N. I., Delhorne, L. A., Rabinowitz, W. M., & Grant, K. W. (1989). Research on tactual communication of speech: Ideas, issues, and finding. In N. S. McGarr (Ed.) Research on the use of sensory aids for hearing-impaired people (pp. 65-78). Washington, DC: Alexander Graham Bell Association for the Deaf.

Reed, C. M., Rabinowitz, W. M., Durlach, N. I., Delhorne, L. A., Braida, L. D., Pemberton, J. C., Mulcahey, B. D., & Washington, D. L. (1992). Analytic study of the Tadoma Method: Improving performance through the use of supplementary tactile displays. Journal of Speech and Hearing Research, 35, 450-465.

Seligman, P. (1987). Speech processing strategies and their implementation. Annals of Otolaryngology, Rhinology, and Laryngology, Supplement, 96, 71-74.

Skolnick, M. L. (1970). Videofluoroscopic examination of the velopharyngeal portal during phonation in lateral and base projections — A new technique for studying the mechanics of closure. Cleft Palate Journal, 7, 803-816.

Skolnick, M. L., & Cohn, E. R. (1989). Techniques of multiview videofluoroscopy. In M. L. Skolnick (Ed.), Videofluoroscopic studies in patients with cleft palate (pp. 24-48). New York City, NY: Springer-Verlag.

Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.) Hearing by eye: The psychology of lipreading (pp. 3-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tye-Murray, N. (1984). Articulatory behavior of deaf and hearing speakers over change in rate and stress: A cinefluorographic study. Unpublished doctoral dissertation, The University of Iowa, Iowa City, IA.

Tye-Murray, N. (1991). The establishment of open articulatory postures by deaf and hearing talkers. Journal of Speech and Hearing Research, 34, 453-459.

Tye-Murray, N. (1992a). Speechreading training. In N. Tye-Murray (Ed.), Children with cochlear implants: A handbook for parents, teachers, and speech and hearing professionals. (pp. 115-135). Washington, DC: Alexander Graham Bell Association for the Deaf.

Tye-Murray, N. (1992b). Aural rehabilitation and patient management. In R. S. Tyler (Ed.), Audiologic foundations of cochlear implants, (pp. 87-144). San Diego, CA: Singular Publishing Group, Inc.

Tye-Murray, N. (1992c). Articulatory organizational strategies and the roles of audition. Volta Review, 94, 243-260

Tyler, R. S., & Lowder, M. W. (1992). Audiological management and performance of adult cochlear-implant patients. Ear, Nose and Throat Journal, 71(3), 117-128.

Tyler, R. S., Preece, J., & Tye-Murray, N. (1986). The Iowa laser videodisc sentence test without context. Iowa City, IA: University of Iowa Hospitals and Clinics.

Tyler, R.S., Tye-Murray, N., & Gantz, B. J. (1987). The relationship between vowel, consonant and word perception in cochlear-implant patients. In P. Banfai (Ed.), Proceedings of the International Cochlear Implant Symposium, Duren, West Germany (pp. 605-612). West Germany: Fachbuchhandlung.

Upton, H. W. (1968). Wearable eyeglass speechreading aid. American Annals of the Deaf, 113, 222-9.

Walden, B. E., Erdman, S. A., Montgomery, A. A., Schwartz, D. M., & Prosek, R. A. (1981). Some effects of training on speech recognition in hearing-impaired adults. Journal of Speech and Hearing Research, 24, 207-316.

Walden, B. E., Prosek, R. A., Montgomery, A. A., Scherr, C. K., & Jones, C. J. (1977). Effects of training on the visual recognition of consonants. Journal of Speech and Hearing Research, 20, 130-145.

Wang, M. D. & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. Journal of the Acoustical Society of America, 54, 1248-1266.

Williams, W. N. (1989). Radiographic assessment of velopharyngeal function for speech. In K. R. Bzoch (Ed.), Communicative disorders related to cleft lip and palate (pp. 195-210). Boston, MA: College Hill Press.

NCVS Status and Progress Report - 4
June 1993, 65-72

Speaking with the Cochlear Implant Turned On and Turned Off

Nancy Tye-Murray, Ph.D.

Department of Otolaryngology, The University of Iowa

Karen Iler Kirk, Ph.D.

Department of Otolaryngology, Indiana University

George Woodworth, Ph.D.

Department of Statistics, The University of Iowa

Abstrast

Eleven prelingually deafened children spoke 14 words after their cochlear implants had been turned off for several hours. They spoke the words again with their device turned on. Statistical analyses performed on the transcribed utterances revealed that no significant differences occurred between the device-turned-off and the device-turned-on condition. On average, subjects did not change the accuracy with which they produced vowel height, vowel place, initial consonant place, initial consonant voicing, nor final consonant voicing. Subjects who were most likely to recognize vowels in an audiological test were also most likely to demonstrate a decrement in their production of the vowel place feature when their cochlear implant was turned off.

One of the most interesting issues concerning the speech production of young cochlear implant users is whether or not they will learn to use the electrical signal to influence and fine-tune their ongoing speech production. One paradigm for examining this issue is a device-on and device-off paradigm, where children speak after having their device turned off for a length of time and then speak with their device turned on. There is evidence that the acoustic properties of some vowels alter when children speak with their devices turned off. The vowel formants of front-mid vowels may become more centralized, and vowel duration may lengthen [1].

The present investigation had three purposes. First, few investigations have examined whether or not changes between device conditions are linguistically salient. That is, do the shifts that occur when the device is turned off result in a different phonetic percept? Implanted children have limited perceptual skills. In order for them to recognize that their speech production has deteriorated following a device-off condition and after the device has been turned back on, the decrements that occur between a device-on and device-off condition will probably have to be large. This investigation determined whether phonemes or features of articulation sound different when implanted children speak with their devices turned on versus off.

The second purpose was to relate children's overall speech intelligibility to their susceptibility to a device-off condition. It may be that children who have developed relatively good intelligibility have learned to monitor their ongoing speech production auditorally. If this is true, they may demonstrate a performance decrement when they speak with their cochlear implants turned off rather than on, whereas children who have poor intelligibility may not.

The third purpose was to relate children's auditory-only perceptual skills to changes that occur between the device-on and device-off conditions. We reasoned that subjects who demonstrate the best perceptual skills might be most likely to demonstrate a decrement in speaking proficiency when their cochlear implants were turned off. That is, children who have the potential to hear their own speech might be most likely to monitor (and regulate) their own speech auditorally.

Methods

Subjects

Eleven children who use the Nucleus cochlear implant served as subjects. On average, subjects were 8.1 years (SD = 3.4) and 5.8 months (SD = 3.8) at the time of testing. All subjects were prelingually deafened. The average length of cochlear implant use at the time of testing was 34.9 months and ranged from 24 to 48 months. All children use Total Communication.

Device - On/Off Stimuli

The on/off stimulus items were fourteen words contained in the target sentence, "That's the ________." The stimulus words were : bud, tea, D, bat, beet, boat, bed, key, pea, bad, bead, boot, and bee. They were represented by individual pictures and written sentences. Six different randomizations of the stimulus items were created, with subjects randomly assigned to one of the six lists. Subjects recited each item three times in a test condition.

The initial consonants of a subset of the stimuli allowed us to evaluate the features of initial consonant voicing (voiced = D, bee; unvoiced = key, tea, pea) and initial consonant place of articulation (bilabial = pea, bee; alveolar = tea, D; velar = key). The vowels of another subset allowed us to evaluate the features of vowel height (high = bead, boot; mid = bed, boat, bird; low = bud, bat) and vowel place of articulation (front = bat, bed, bead; central = bird, bud; back = boot; boat). Finally, the final consonants of a third subset allowed evaluation of final consonant voicing (voiced = bead, bad, bird, bud, bed; unvoiced = beet, bat, boot, boat).

Intelligibility Speech Stimuli

Children participated in a story-retell activity in order to provide an estimate of speech intelligibility. A clinician told them a story, using four sequential picture cards. The subject then retold the story using Total Communication. The recitation was audiovideo recorded.

A speech-language clinician orthographically transcribed the child's signed message and phonetically transcribed the child's speech. A percent phoneme correct score was computed by referencing the phonetic transcription to the sign transcription [2]. This measure of intelligibility was related to the changes that occurred between the device-on and -off conditions.

Speech Data Collection

Subjects were seated in a quiet room during data collection. They were first shown each picture and instructed to produce each word in the target sentence. If a child could not read, and did not name the picture correctly, the correct target word was modeled for the child. For example, some subjects initially produced the word flower for the target word bud. A few subjects could produce single words, but not the target sentence. For these subjects single word productions were accepted. Once the child could correctly identify all items, data collection began.

All subjects were tested at one of their regularly-scheduled postimplant test intervals, which encompassed an entire day. The implant-off sample was collected in the morning as soon as the subjects arrived at the clinic. Parents were instructed not to let their children use their cochlear implants that morning prior to their arrival. Thus the subjects had been without their cochlear implant for about 12 hours. The implant-on sample was collected after anywhere from one to four hours of cochlear implant use, with the most common interval being four hours. One child mistakenly arrived at the center using her cochlear implant. For this child, the implant-on sample was collected first. She refrained from wearing her cochlear implant for one hour at lunch, and then the implant-off sample was collected.

Samples were collected in a quiet room with a head-mounted Sony microphone, which allowed a constant mouth-to-microphone distance. Prior to data collection the children repeated each item once or twice to determine the appropriate recording level. This level remained unchanged throughout both conditions. Once testing began, each stimulus item was repeated three times. The samples were amplified (Nakamichi X Microphone Mixer [Model MX-100] and Nakamichi Microphone Preamplifier [Model PS-100]) and audiorecorded (Nakamichi Cassette Deck 2).

Sample transcription

A speech-language clinician who is familiar with the speech of talkers with profound hearing impairment (the second author) transcribed the stimulus items. Only the target words were transcribed. For successive subjects the condition transcribed first (device-on or device-off) was alternated. Each target word was first transcribed orthographically. The listener then played the tape a second time and phonetically transcribed the target words.

Transcription reliability

A graduate student in speech-language pathology who is familiar with the speech of Deaf talkers transcribed the samples from two of the subjects (i.e., four samples, two device-on and two device-off conditions). Reliability was computed as the number of agreements between the original transcriber and the graduate student across all six samples. Interjudge reliability was .92.

Intrajudge reliability was computed by having the first transcriber rescore both samples from two subjects. Reliability, computed as above, was .94.

Audiological Testing

Subjects completed the Audiovisual Feature Test and the Children's Vowel Perception Test [3]. The Audiovisual Feature Test consists of 10 monosyllables (e.g., B, C, key), each repeated 6 times in a test list. The clinician speaks the name of an item and the child responds by touching one of the pictures. The responses are used to construct a confusion matrix, and an information transmission analysis [4] is performed to determine how well the child perceives the features of place of articulation, nasality, duration, frication, and voicing. Two subjects did not take the test due to time constraints.

The Children's Vowel Perception Test includes five sets of pictures. Each set has four words (e.g., kite, cat, cut, coat). The clinician speaks an item and the child touches a response from a set of four pictures. The test presents a total of 40 items.

Results

Speaking changes between device-on and device-off

The production data from the device-on and device-off conditions were analyzed in three ways. First, data for each of the three subsets of the speech stimuli (i.e., the initial consonants, the medial vowels, and the final consonants) were grouped across subjects in order to determine whether the group productions differed between the device-on and the device-off conditions. Secondly, the data were grouped by individual phonemes such as initial [p] and medial [i]. This analysis determined whether some phonemes varied between the two test conditions while others did not. Finally, in the third analysis, within-subject comparisons were made to determine whether individual subjects did or did not demonstrate differences in the two test conditions.

The results from the first analysis appear in Tables 1, 2, and 3. Paired comparison *t*-tests indicated that no significant changes occurred between the device-on and the device-off conditions ($p > .05$). On average, subjects were as likely to produce the test phonemes and features correctly when their cochlear implants were turned off as when turned on.

Table 1.
The mean number and percent of initial consonants that were produced correctly and also the accuracy with which the consonant place and voicing features were spoken (5 test syllables X 3 repetitions = 15 possible correct).

Device Condition	Place Feature	Voicing Feature	Total Initial Consonants Correct
ON	9.6 / 64% (SD = 3.2)	8.3 / 55% (SD = 3.0)	4.9 / 33% (SD = 3.6)
OFF	10.9 / 73% (SD = 3.5)	8.6 / 57% (SD = 3.7)	5.0 / 33% (SD - 4.4)

Table 2.
The mean number and percent of vowels that were produced correctly, and also the accuracy with which the vowel place and height features were spoken (7 test syllables X 3 repetitions = 21 possible correct).

Device Condition	Place Feature	Height Feature	Total Vowels Correct
ON	17.5 / 83% (SD = 4.0)	17.0 / 81% (SD = 5.7)	13.3 / 63% (SD = 5.1)
OFF	16.3 / 77% (SD = 2.7)	13.6 / 65% (SD = 3.8)	12.5 / 60% (SD = 11.1)

Table 3.
The mean number and percent of final consonants that were produced correctly, and also the accuracy with which the consonant voicing feature was spoken (9 test syllables X 3 repetitions = 27 possible correct).

Device Condition	Voicing Feature	Total Final Consonants Correct
ON	12.5 / 46% (SD = 11.1)	10.8 / 40% (SD = 11.4)
OFF	13.5 / 50% (SD = 9.6)	10.0 / 37% (SD = 10.7)

This trend was also apparent in the second analysis. No phoneme or feature of articulation was produced correctly more often in the on-condition than the off-condition. Previous research has demonstrated that formant values and phoneme durations for the mid-front vowels, such as [E], may shift when children produce it with their device turned off and then on [1, 5]. The present results suggest that such shifts may not be linguistically significant. Table 4 shows that the place and height features for [E] in /bEd/ were perceived alike in both device conditions.

Table 4.
Mean number and percent correct productions for the place and height features for the vowel [E] and correct productions (1 syllable X 3 repetitions = 3 possible correct).

Device Condition	Place Feature	Height Feature	Total [E] Correct
ON	2.9 / 97% (SD = 0.3)	1.0 / 33% (SD = 1.4)	1.0 / 33% (SD = 1.4)
OFF	2.9 / 97% (SD = 0.3)	1.1 / 37% (SD = 1.4	1.1 / 37% (SD = 1.4)

The third analysis showed that no subject performed better when their device was turned on versus turned off ($p > .05$). Two subjects, CF and MM, actually produced more errors when their cochlear implants were turned on rather than off (CF: $t = -2.38$, $p < .03$; MM: $t = -2.38$, $p < .03$).

Relationship between intelligibility measures and changes in device condition

Pearson correlations were performed between the subjects' intelligibility scores and the difference scores computed between their device-on and device-off conditions for all of the phonemes and for each of the speech features. No significant relationships were revealed between overall intelligibility and the difference scores. Subjects who were more intelligible were no more likely to vary in the number of correct feature or phoneme productions when their cochlear implants were turned on versus when it was turned off.

Relationship between audiologic measures and changes in device condition

Pearson correlations were performed between the difference scores and the results of the information transmission analysis that was performed on the consonant confusion data (from the Audiovisual Feature Test). Pearson correlations were also performed between these difference scores and the percent correct scores from the Children's Vowel Perception Test and the percent correct scores from the Audiovisual Feature Test. Only one relationship reached significance: children who were more likely to recognize vowels correctly on the Children's Vowel Perception Test were also more likely to shift their place of vowel production between the two device conditions ($r = .69$, $p < .03$). This relationship is illustrated in Figure 1.

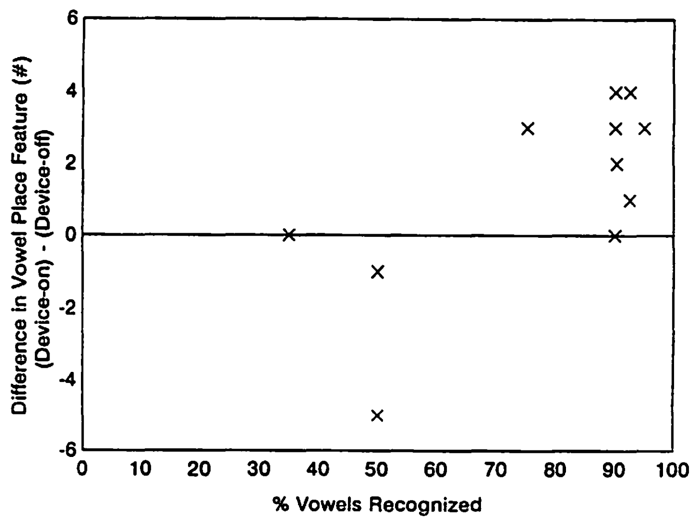


Figure 1. Relationship between the difference scores for the vowel place feature computed between the device-on and -off conditions and the scores from the Children's Vowel Perception Test.

Conclusions

This investigation permits the following conclusions:

1) On average, prelingually deafened children who have two years or more experience with a Nucleus cochlear implant do not change their speaking accuracy when their devices are turned off rather than on, at least in a way that is linguistically salient. Subjects did not change their percent of phonemes correct nor their percent of features correct.

As such, any speaking changes that occur when a child's cochlear implant is turned off are likely to be subtle, and possibly related to changes in how loudly or softly the child speaks.[1] Changes may only be apparent when fine-grained analysis procedures are used, such as acoustic analysis of formants or narrow transcription. (We are currently determining whether movement behaviors change, using strain gauge analysis.) Future investigations might determine how well children with cochlear implants can perceive subtle auditory changes in the speech signal, and explore whether it is possible for them to recognize fine variations in their own auditory feedback.

2) Children who demonstrate relatively good intelligibility do not appear to alter their speech production more between the two device conditions than children who demonstrate poor intelligibility.

3) Children who better recognize vowels are more likely to alter their production of vowel place between device conditions than children who do not. Vowel place, or the location of the tongue dorsum constriction within the vocal tract, is signalled primarily by second formant information [6]. Thus, it may be that children who perform well on the Children's Vowel Perception Test perceive high-frequency information, and can attend to this informa tion in their own auditory feedback for positioning their tongue dorsum.

4) Children's ability to recognize consonants or to utilize consonant feature information appears to be unrelated to how they alter their speech production in a device-off condition. Future investigation may reveal that the production of closed postures of articulation are less likely to be affected by on-line auditory feedback than the production of open postures. Certainly, the acoustic signal that corresponds to closed postures tends to have low amplitude, so it is less likely that this information will be useful for regulating on-going speech production.

Acknowledgments

This study was supported by research grant DC00242 and DC00976-01 from the National Institutes of Health/NIDCD, grant RR59 from the General Clinical Research Centers Program, Division of Research Resources, NIH, and the Iowa Lions Sight and Hearing Foundation.

References

1. E. A. Tobey. Speech production. In R. S. Tyler (Ed.), Cochlear Implants: Audiological Foundations. San Diego, CA: Singular Publishing. (1993), pp. 317-356.
2. N. Tye-Murray, & K. I. Kirk. Vowel and diphthong production by young cochlear implant users and the relationship between the Phonetic Level Evaluation and spontaneous speech. Journal of Speech and Hearing Research. (In press).
3. R. S. Tyler, H. Fryauf-Bertschy, & D. Kelsay. Audiovisual Feature Test for Young Children. Iowa City: The University of Iowa. (1991).

[1]Some children decrease or increase their speaking intensity when their cochlear implant is turned off. This may cause their formant patterns to shift, even though the phonetic complexion of their message does not change.

4. G. A. Miller, & P. E. Nicely. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27, 338-352. (1955).

5. E. A. Tobey, S. Angelette, C. Murchison, J. Nicosia, S. Sprague, S. J. Staller, J. A. Brimacombe, & A. L. Beiter. Speech production performance in children with multichannel cochlear implants. The American Journal of Otology, 12(S), 165-173. (1991).

6. J. M. Pickett. The sounds of speech communication: A primer of acoustic phonetics and speech perception. Austin, TX: Pro-Ed. (1980).

NCVS Status and Progress Report - 4
June 1993, 73-84

Dependence of Phonation Threshold Pressure on Hydration Level

Katherine Verdolini-Marston, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
and
Department of Music, The University of Iowa
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Department of Music, The University of Iowa
and
The Recording and Research Center, The Denver Center for the Performing Arts
Ann Fennell, M.S.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

The purpose of the study was to assess the dependence of phonation threshold pressure (PTP) on hydration level, using a double-blind experimental approach. Twelve adults with normal voices, but without previous voice training, participated as subjects. Each subject received a 4-hour hydration treatment, a 4-hour dehydration treatment, and a 4-hour placebo (control) treatment. Following each treatment, PTPs were measured at low, medium, and high pitches. The results indicated an inverse relation between PTPs and hydration level, as predicted theoretically. However, the effect size was smaller and more variable than in a previous study that did not control for experimental biases and included trained subjects. Pending further, more direct physiological investigations, the results that were obtained are tentatively attributed to changes in vocal fold tissue viscosity associated with changes in hydration level.

Hydration treatments are often included in the clinical management of voice disorders of various etiologies. Patients are advised to increase fluid intake, and to humidify living and working environments. Mucolytic drugs are sometimes also prescribed adjunctively. In the case of lesions such as laryngeal nodules and polyps, the claim is that such treatments may enhance lesion regression, and therefore improvements in a dysphonia (Verdolini, 1988; Verdolini-Marston, 1991). Some empirical evidence consistent with this notion was recently reported on the basis of a double-blind, placebo-controlled study (Verdolini-Marston, Sandage, & Titze, in press). A more general claim is that hydration treatments somehow facilitate phonation, or make it "easier" (see for example, Lawrence, 1981). The present study represents a continuation in a series of studies that investigate this claim from a theoretical perspective. The specific question is whether estimated phonation threshold pressures (PTPs), or the estimated subglottal pressures required to initiate and sustain vocal fold oscillation, inversely depend on hydration level, using a double-blind, placebo-controlled approach. An affirmative answer would be consistent with theoretical predictions, as well as with the clinical claims noted.

Theoretically-based predictions about the effect of hydration level on PTP are based on two successive hypotheses. The first is that hydration directly regulates vocal fold tissue viscosity. The second is that vocal fold tissue viscosity is directly proportional to PTP. Combining the two hypotheses, hydration conditions should decrease PTP, and dehydration conditions should increase it.

Support for the first hypothesis comes from the observation that the viscosity of pure water is about 0.01 poise (P), as compared to the viscosity of vocal fold tissue, which is much higher, ranging from 0-10 P (Titze & Talkin, 1979). It follows that if hydration treatments infuse vocal fold tissue with water or with other low viscosity fluids, superficially or internally, the net viscosity of the tissue should be decreased. Similarly, if dehydration conditions extract low viscosity fluids from the tissue, net tissue viscosity should be increased. These possibilities remain untested assumptions for the present, because direct measures of vocal fold tissue viscosity are currently unavailable for live human subjects.

The second but equally critical hypothesis is based on a theoretical formulation developed to describe small-amplitude vocal fold oscillations in barely abducted vocal folds (Titze, 1988). According to this formulation, the following relation applies to a rectangular glottis:

$$\underline{PTP} = (2k/T)(Bc)(w/2)$$

where $\underline{k}$ is a dimensionless constant equal to about 1.1, $\underline{T}$ is vocal fold thickness, $\underline{B}$ is a damping coefficient that depends directly on vocal fold tissue viscosity, $\underline{c}$ is the velocity of the mucosal wave propagation, and $\underline{w/2}$ is the prephonatory glottal half-width. The focus for the present discussion is the proposed relation between PTP and $\underline{B}$, the damping coefficient. Decreases in vocal fold tissue viscosity (as may occur with hydration treatments) should cause decreases in PTP. Increases in tissue viscosity (that may occur with dehydration treatments) should cause increases in PTP.

Some empirical support is available for the proposed inverse relation between hydration level and PTPs, in both animal and human subjects. In particular, in one previous study that we conducted (Verdolini-Marston, Titze, & Druker, 1990), six adults, most with some amount of prior voice training (4/6) and all with normal voices, each received a 4-hour hydration treatment and a 4-hour dehydration treatment. Following both treatments, PTPs were estimated from oral pressures at low, medium, and high pitches, and were compared to PTPs in a no-treatment control condition. The results indicated a reliable inverse relation between PTPs and hydration level; the lowest pressures were obtained following the hydration treatment, and the highest pressures were obtained following the dehydration treatment, as predicted. Reductions in PTPs with hydration treatments were particularly marked at high pitches. The results were attributed to changes in vocal fold tissue viscosity as already described, because direct measures of tissue viscosity were unavailable.

The results from this study provided preliminary empirical support for the proposed inverse relation between hydration level and PTP. However, at least one important factor restrained a confident interpretation: there was considerable potential for unintentional subject and/or experimenter bias in that study. Most of the subjects (4/6) were explicitly familiar with the experimental hypotheses. Also the experimenter who elicited PTPs was familiar with the research questions, and this experimenter was further informed about subjects' prior treatment condition when PTP measures were made. Thus, unintended subject and/or experimenter biases could have affected the results.

In summary, various lines of evidence and theoretical reasoning point to an inverse relation between hydration level and PTP. The results from one experiment were consistent with this prediction (Verdolini-Marston et al., 1990). However, unintended experimental biases could have affected the results in that experiment. The present study asks whether the same relation is obtained when such biases are controlled.

Methods

Subjects

Subjects were 12 healthy adults who were recruited for pay, mostly from the Department of Speech Pathology and Audiology at our university. Nine females and three males participated. Subjects' ages ranged from 20 to 30 yr (M = 23.8 yr). At the outset of the experiment, all of the subjects expressly indicated unfamiliarity with the hypotheses under investigation. All subjects were non-smokers and denied a history of any voice disorder. None of the subjects had undergone systematic voice training, except for one subject who had received private voice lessons for one semester several years prior to her participation, in the context of a high school choir. The reason for excluding trained voice users was because our impression was that trained voice users often have biases about hydration treatments, that are generally consistent with our experimental hypotheses. All subjects denied active illnesses or significant past illnesses, and specifically denied kidney dysfunctions, asthma, and glaucoma (risk factors for the hydration treatment), hypoglycemia or diabetes (risk factors for the placebo treatment), and pregnancy (a risk factor for any experimental procedure).

Treatments

All treatments were carried out in an environmental chamber that allows for a relative regulation of temperature and humidity level. Across all treatments, the temperature was held constant at approximately 27°C. The hydration treatment was described to subjects as a "Tropical Treatment". It consisted of a 4-hour exposure to a high humidity environment (80 - 98% relative humidity), administration of 2 teaspoons of a mucolytic drug, Robitussin expectorant (guafenisin), at the beginning of treatment and again 1/2-hour prior to treatment termination[1], and encouragements to drink as much water as could be comfortably tolerated. The amount of water that subjects in fact consumed during this treatment ranged from 24 - 128 oz, M= 82 oz.

[1]According to pharmaceutical specifications, the peak effect for this drug occurs 30-60 min following administration.

The dehydration treatment, described to subjects as the "Arizona Treatment", consisted of a 4-hour exposure to a relatively low humidity environment (for most subjects, actual relative humidities ranged from 9 - 32%),[2] administration of 2 teaspoons of a decongestant drug, Dimetapp, at the beginning of treatment and again 1 hour before treatment termination,[3] and encouragements to abstain from any fluid intake. In fact, no subjects drank any fluids during this treatment.

The control treatment was described to subjects as a "Naturalist Treatment". This treatment consisted of a 4-hour exposure to an intermediate environmental humidity (40 - 61%) which subjects were told involved air filtering procedures to improve air purity, the administration of 2 teaspoons of a placebo drug (cherry syrup), described to subjects as an herbal medication and administered at the beginning of the treatment period and again 2 hours before treatment termination, and no instructions regarding fluid intake. For this condition, subjects drank from 0 - 36 oz of fluids, $\underline{M}$ = 15 oz.

For all conditions, medications were provided to subjects from previously unmarked pharmacy bottles that we coded specially for the experiment. Therefore, subjects did not know what drugs they received.

Procedures and PTP measures

Each subject was first provided an overview of the experiment. That is, the treatments were briefly described as a "Tropical Treatment", an "Arizona Treatment", and a "Naturalist Treatment". No information was given about anticipated effects. Following a subsequent collection of demographic information, critical pitches that would be used for PTP trials were individually determined for each subject. A "medium" pitch was determined first. For the determination of this pitch, the subject counted from one to five and the experimenter matched the pitch on the prolonged vowel in the word "three" to a pitch on a portable keyboard. Then, the total pitch range was determined as a basis for what would be the individual subject's "high" and "low" pitches during the experiment. The subject phonated on /a/, starting with an arbitrary intermediate pitch and progressing to successively lower pitches, until both the subject and the experimenter were convinced by repeated trials that the lowest pitch had been produced. Then the subject phonated on /a/ at an arbitrary intermediate pitch level and progressed upwards in pitch, including falsetto, until it was clear by repeated trials that the highest pitch had been produced. The total pitch range in semitones was established on the basis of the minimum and maximum pitches that the subject produced on these trials. The pitch that would be used for that subject for all "high" pitch conditions in the experiment was calculated as the 80th percentile pitch in the total semitone range. The pitch that would be used for all "low" trials was calculated as the 10th percentile pitch in the total semitone range.[4]

[2]The target humidity for the dry condition was 10 - 20% relative humidity. In fact, for 6 subjects, actual humidities ranged from 9 - 19%. However, the environmental chamber failed to consistently maintain pre-set humidity levels for the dry condition. Thus, for 3 subjects, humidities for the dry condition ranged from 29 - 32%, and for 3 final subjects, humidities for this condition ranged from 38 - 47% However, for no subject was there at any time an overlap between the humidities delivered in dehydration, control, and hydration treatment conditions. The point is that the dehydration effect might be underestimated in the present study, because of relatively higher humidities in the dry condition than intended.

[3]Peak effect for this drug occurs 1 - 1 1/2 hr after administration, according to pharmaceutical indications.

[4]In two cases, the "low" pitch or 10th percentile pitch was actually higher than the "medium" or conversational pitch. In these cases, the lower of the 2 pitches was used for all subsequent "low" pitch trials, and the relatively higher pitch was used for all subsequent "medium" pitch trials. Also, in one case, the "low" or 10th percentile pitch was the same as the "medium" or conversational pitch. In this case, the actual pitch used for subsequent "medium" pitch trials was one semitone higher than the "low" (10th percentile) pitch.

After the initial pitch trials were completed, the subject was to instructed how to perform the actual PTP task. The procedures for this task were based on previous work indicating that subglottal pressures during phonation can be estimated from oral pressures during voice stop consonants in consonant-vowel strings (Löfqvist, Carlborg, & Kitzing, 1982; Rothenberg, 1977; Smitheran & Hixon, 1981), assuming a critical rate of syllables production (approximately 92 beats per min) and approximately constant effort levels (Holmberg, Perkell, & Hillman, 1987). Specifically, the subject was trained to produce the consonant-vowel string /pae pae pae pae pae/ at a rate of 92 beats per minute, indicated by a metronome, at the medium pitch, at the high pitch, and at the low pitch. For each pitch condition, the subject was instructed to first produce the repeated syllables at a level just above phonatory threshold, then at a level just below threshold (whispering), and finally at phonatory threshold, or "as quietly as you can". Supra- and subthreshold trials were included to provide the subject with a better sense of "threshold" that might otherwise have been available. However, only data from threshold would be collected. Initial training on the PTP task lasted about 5 - 10 min.

It should be noted that during training and for the remainder of the experiment, the pressure contours that subjects produced for the target threshold tasks were not as smooth as those described by earlier authors for suprathreshold tasks (Holmberg et al., 1987). That is, our subjects did not necessarily produce clear peak steady-states in oral pressures. Arguably, this factor could have resulted in an underestimation of subglottal pressures in the present study. However, in our previous experiment investigating threshold pressures (Verdolini-Marston et al., 1990), we had similarly noted a tendency for pressure contours to be "choppy", despite an emphasis on steady-state performances, and the effects that we were investigating were detected nonetheless. Thus, in the current study we decided not to insist on steady-state contours particularly in light of the fact that the subjects in the current experiment were quite unsophisticated vocally, and it appeared already quite difficult for them to produce threshold utterances at the different pre-determined pitches. Presumably, if subglottal pressures were underestimated in this way, the underestimation would be approximately linear across treatment conditions and thus would not affect the results qualitatively.

Following training, the subject initiated the first treatment session, as already described. (Treatment order was counterbalanced across subjects to control for order effects.) During the treatment, relative humidity and temperature levels were monitored and recorded at approximately 1/2-hour intervals. Within about 10-15 minutes following treatment termination, PTPs were obtained. That is, the subject produced five cycles of suprathreshold, subthreshold, and threshold trials at his/her predetermined "medium" pitch, then at his/her "high" pitch (for both males and females, falsetto phonation was emphasized for high-pitch trials), and finally at his/her "low" pitch. Target pitches were provided with a portable electric piano and also vocally before each trial. Threshold trials (only) were collected at a rate of 2000 Hz and stored for later analysis. Prior to almost each threshold trial, zero-input voltage to the pressure measurement system was monitored on a multimeter and restored to $\pm$ 1 mV or less if required. Also, following each threshold trial, the pressure signal was checked on an oscilloscope display; any signal for which the zero-input offset was perceptibly displaced from the zero-value of the y-axis was recollected. Also trials for which the pitch produced differed by more than one-quarter tone from the target pitch, as perceived by the experimenter, were recollected. Together, no more than about 5% of trials was recollected.

On another day (at least 24 hr later), the subject received a second treatment, and underwent measures as previously. Finally, on a third day, the subject received the final treatment. In all cases, the experimenter who elicited measurements was uninformed about which treatment the subject had just received.

Equipment and measurement extraction

A specially constructed environmentally controllable chamber was used during the administration of all treatments. This chamber allows for a pre-setting of relative humidity and temperature levels. For the measurement of actual humidities and temperatures inside the chamber during treatment, a combination wet-bulb/dry-bulb thermometer was used, and the humidity level was determined from comparative temperature measures.

The basic equipment used to collect PTP measures is shown in Figure 1. The primary device was a Glottal Enterprises (GE) MS100 - A2 pressure/flow measurement system. A 15-gage translabial stainless steel hypodermic needle embedded in a circumferentially vented face mask was connected to the GE pressure transducer. Signals from the main output of the system were relayed to a Hewlett Packard 350D attenuator for attenuation of large (clipped or potentially clipped) signals. Finally, signals were routed from the attenuator to a Hewlett Packard 3466A Digital Multimeter for on-line calibration purposes, and in parallel, to a Gateway2000 386 computer. Hypersignal software (1989) was used to collect and store the pressure data. Signals were recalled using a locally developed software program that indicated the average peak value in the pressure signal for the second and the third peaks of the five-syllable utterance, for the third and the fourth peaks, and for the fourth and the fifth peaks. All values were subsequently transformed to standard pressure units in cm H20, following calibration procedures, and were used in statistical analyses.

A Casio portable keyboard was employed to determine target pitches for PTP trials, and later, to provide these pitches prior to actual experimental trials. A Wittner Taktall Piccolo metronome provided the target rate for syllable productions.

Design

The experiment involved a double-blind, placebo-controlled, counterbalanced design. Subjects were naive to the experimental hypotheses, and the experimenter who elicited post- treatment measures did not know what treatment subjects had just received. The order of treatment conditions was counterbalanced across subjects to control for order effects. Because initial analyses indicated that systematic order effects were not present, order was not included as a factor in further analyses.[5]

As a final methodological note, because the results from the present study will be compared to those from a previous study that we conducted (Verdolini-Marston et al., 1990), it should be pointed out that the primary difference in experimental design between the two

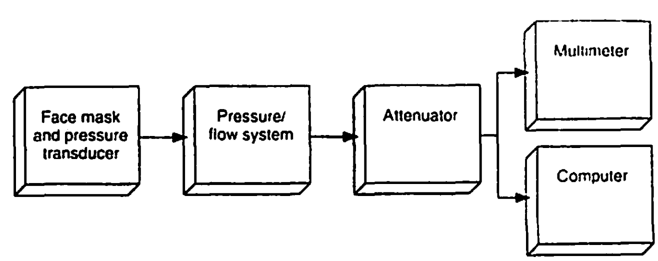


Figure 1. Equipment set-up.

[5]Also, for one subject, in the control condition the wrong high pitch was inadvertently provided. Related trials were excluded from all analyses.

studies was the use of a double-blind design in the current study. Also, trained voice users were excluded in the present study because of their potential for guessing the experimental hypotheses, but were included in the previous study.

Results and Discussion

Table 1 and Figure 2 display the average results for PTP. As predicted, average PTPs were inversely related to hydration level. They were highest following the dehydration condition, they were intermediate following the control condition, and they were lowest following the hydration condition. This trend was consistently observed not only collapsing across pitches, but also within each pitch level. Further, also PTP variability inversely depended on hydration level. With the exception of the dehydration versus the control condition at the medium pitch, the greatest variability in PTPs was observed following the dehydration condition, and the least variability was observed following the hydration condition. This result is indicated by the standard deviations for PTPs, displayed in Table 1.

Table 1.
Average PTPs, in cm H20, at Low, Medium, and High Pitches, following Dehydration, Control, and Hydration Treatments. (Standard Deviations Indicated in Parentheses.)

	TREATMENT CONDITION			
Pitch	Dehydration	Control	Hydration	Average
Low	3.32 (0.83)	3.18 (0.79)	3.09 (0.65)	3.20
Medium	3.23 (0.54)	3.17 (0.63)	3.08 (0.48)	3.16
High	8.43 (4.29)	8.01 (4.12)	7.57 (3.75)	8.00
Average	4.99	4.69	4.58	

Despite the trend noted for average PTPs, a 3 (hydration level) by 3 (pitch level) within-subjects repeated measures Analysis of Variance *(ANOVA)* failed to confirm a reliable main effect of hydration for PTPs, $p < .33$. A primary reason was the large variability in PTP responses for this subject group, particularly at high pitches. Incidental to the purpose of this study, the main effect for pitch was reliable in the same ANOVA, $p < .0001$. As in previous studies (Finkelhor, Titze, & Durham, 1988; Verdolini-Marston et al., 1990; Verdolini-Marston et al., in press), the highest PTPs were produced at the high pitch, and the lowest PTPs were produced at the intermediate pitch.

Although parametric tests failed to confirm a clear influence of hydration level on PTPs, non-parametric tests indicated a reliable effect. That is, according to a simple binomial sign test, the likelihood that the highest PTPs would be obtained for the dehydration condition for all three pitch conditions, and simultaneously that the lowest PTPs would be obtained for the hydration condition for all pitch conditions, was reliably greater than chance ($p < .005$).

Because some subjects in the present study were exposed to higher humidity levels in the dehydration condition than originally intended, separate analyses were conducted for a subset of subjects (N = 6), who represented the subset of subjects who received the lowest humidity levels in the dry condition, counterbalancing for treatment order. The results of these analyses indicated that higher PTPs were not produced by these subjects in the dry condition, as compared to PTPs produced in the dry condition by the subject group as a whole. Therefore, there was no evidence that PTPs for the dry condition were underestimated in this study by the delivery of higher humidity levels in this condition than intended.

Finally, because the present study involved a methodological replication of an earlier study, only controlling for experimenter and subject biases and, as a part of that, using untrained as opposed to trained voice users, it is interesting to directly compare the results from this study to those from the earlier study. Table 2 displays the earlier results.

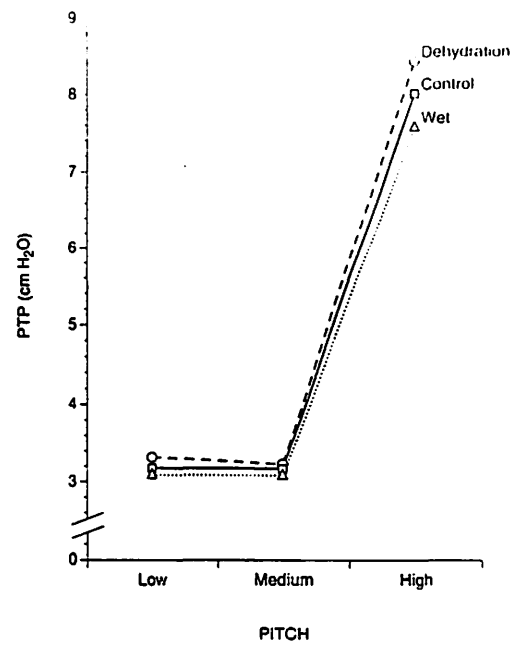


Figure 2. Average PTPs at low, medium, and high pitches, following dehydration, control, and hydration treatments.

Table 2.
Average PTPs, in cm H20, at Low, Medium, and High Pitches, following Dehydration, Control, and Hydration Treatments, from Previous Study (Verdolini-Marston et al., 1990). (Standard Deviations Indicated in Parentheses.)

	TREATMENT CONDITION			
Pitch	Dehydration	Control	Hydration	Average
Low	4.47 (1.32)	3.55 (0.85)	3.01 (0.98)	3.64
Medium	3.81 (0.91)	3.50 (1.17)	3.08 (0.92)	3.44
High	6.93 (1.09)	6.86 (1.29)	4.24 (0.80)	5.96
Average	5.07	4.64	3.44	

Comparing these results to those in Table 1, from the present study, it is clear that the results across the studies were similar, with three exceptions. First, the untrained voice users in the present study produced greater PTPs at the high pitch (but generally somewhat lower PTPs at low and medium pitches), as compared to the more trained voice users in the previous study. Second, and more importantly, variations in PTP performance were remarkably greater for the untrained voice-user subjects in the present study, as compared to variations for the trained voice-user subjects in the previous study, at the high pitches. This result is seen by comparing the standard deviations for PTPs across the two studies (Tables 1 and 2). However, in both studies, PTPs were relatively more stable following the wet condition, and were least stable following the dry condition (again, indicated by standard deviations in Tables 1 and 2). Third, the percent change in PTPs following hydration and dehydration conditions, relative to the control condition, was considerably less in the present study, as compared to the previous study. For the current subjects, the average percent change across pitches was about ±4% (an increase was obtained following the dehydration condition, and a decrease was obtained following the hydration condition). In the previous study, the average percent change across pitches was about ±17% (an increase of about 12% was obtained following the dehydration condition, and a decrease of about 22% was obtained following the hydration condition.) These results are shown in Tables 3 and 4.

Table 3.
Percent Change in PTP, For Dehydration and Hydration Treatments, Relative to Control Condition.

	TREATMENT CONDITION	
Pitch	Dehydration	Hydration
Low	+ 4%	- 4%
Medium	+ 2%	- 3%
High	+ 5%	- 5%
Average	+ 4%	- 4%

Table 4.
Percent Change in PTP, For Dehydration and Hydration Treatments, Relative to Control Condition, from Previous Study (Verdolini-Marston et al., 1990).

	TREATMENT CONDITION	
Pitch	Dehydration	Hydration
Low	+ 26%	- 15%
Medium	+ 9%	- 12%
High	+ 1%	- 38%
Average	+ 12%	- 22%

General Discussion

The results from this study provide some evidence of an inverse relation between hydration level and phonatory effort, as predicted theoretically and as claimed clinically, using a double-blind placebo-controlled approach, and untrained voice users. PTPs, or estimated phonation threshold pressure measures, were greatest following a dehydration condition, and they were smallest following a hydration condition, both within and across three different pitch levels (low, medium, and high). Parametric tests failed to confirm a reliable main effect of hydration level, at least partly because of a large variability in PTP values in our untrained voice-user subjects at high pitches in particular. However, non-parametric tests indicated that the overall hydration effect was a statistically reliable one. At this level, the results are consistent with the theoretical proposal that PTP depends inversely on hydration level (Titze, 1988), and also with the clinical claim that hydration treatments may make phonation "easier" (Lawrence, 1981).

It is interesting to consider why the magnitude of the hydration effect was smaller in the present study, as compared to a previous study that used informed subjects who were mostly trained voice users. One possible reason is that subjects' expectations about hydration effects boosted the effect size in the previous study, because of advertent or inadvertent manipulations of PTPs, consistent with expectations. However, two, more subtle explanations exist, that might be competing or at least complementary ones. First, perhaps the trained voice users in the previous study were able to exert more finely-tuned control over their threshold vocalizations, thus rendering PTPs relatively more sensitive to hydration level. A second interesting possibility is that subjects' expectations in the earlier study produced physiological variations in vocal fold tissue viscosity, in the anticipated direction. There is precedence for such an interpretation, from existing literature in the psychology of medicine literature. For example, in one report (Matysiak & Green 1984), rats received either insulin, glucose, or saline injections on five trials. On a sixth trial, all animals received a saline injection. Blood-glucose levels were evaluated before and after injections. The result was that the rats who had received the insulin or glucose injections on initial trials produced lowered blood-glucose levels, i.e. a conditioned hypoglycemic response, following the final saline injection. Stated differently, expectations about the effect from the final injection produced physiological changes, consistent with the expectations. Other instances of physiological responses to expectations are also reported in the literature, for example, for histamine release (Russell, Dark, Cummins, Ellman, Callaway, & Peeke, 1984) and morphine reaction (Siegel, Hinson, Krank, & McCully, 1982). Based on such findings, with reference to the present series of studies, it seems plausible that when subjects anticipate changes in PTP as a function of hydration level, expectations about the treatments' effects might somehow produce physiological changes that result in greater or lesser tissue irrigation. Of course, this explanation requires considerable further, systematic investigation. However, we think that it has some merit, from both a practical and a theoretical perspective.

In summary, the present study provides some evidence of an inverse relation between phonatory effort and hydration level, using a double-blind placebo-controlled approach and untrained voice users. However, based on the present findings the magnitude of the effect may be smaller and more variable than previously thought, when subject and experimenter biases are controlled, and when untrained as opposed to trained voice users are considered. Pending further, more direct investigations, the results are tentatively attributed to changes in vocal fold tissue viscosity levels with hydration and dehydration treatments. Future studies will focus on more direct physiological assessments of vocal fold tissue viscosity levels with hydration and dehydration treatments, and also on the mechanisms by which both training and expectations about hydration treatments may boost the anticipated effects.

Acknowledgements

This study was supported by Grant No. P60 DC00976 from the National Institute of Deafness and Other Communication Disorders. The authors acknowledge Dr. Carl Gisolfi of the Department of Exercise Physiology for access to the environmental chamber used in the experiment. Also acknowledged are Mr. Marty Milder for programming assistance, Dr. Kevin Spratt for statistical consulting, Dr. Jerry Moon, Dr. Jim Ryan, and Ms. Eileen Finnegan for technical assistance, and Ms. Julie Lemke for secretarial support.

References

Finkelhor, B.K., Titze, I.R., & Durham, P.L. (1988). The effect of viscosity changes in the vocal folds on the range of oscillation. Journal of Voice, 1, 320-325.

Holmberg, E., Perkell, J., & Hillman, R. (1987). Methods for using a noninvasive technique for estimating glottal functions from oral measurements. Cambridge, Massachusetts: Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, 47-58.

Lawrence, V.L. (1981). Handy household hints: To sing or not to sing. NATS Bulletin, 37, 23-25.

Löfqvist, A., Carlborg, B., & Kitzing, P. (1982). Initial validation of an indirect measure of subglottal pressure during vowels. Journal of the Acoustical Society of America, 72, 633-635.

Matysiak, J., & Green, L. (1984). On the directionality of classically-conditioned glycemic responses. Physiology and Behavior, 32, 5-9.

Rothenberg, M.R. (1977). Measurements of air flow in speech. Journal of Speech and Hearing Research, 20, 155-176.

Russell, M., Dark, K.A., Cummins, R.W., Ellman, G., Callaway, E., Peeke, H.V.S. 1984). Learned histamine release. Science, 225, 733-734.

Siegel, S., Hinson, R., Krank, M.D., & McCully, J. (1982). Heroin "overdose" death: Contribution of drug-associated environmental cues. Science, 216, 436-437.

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

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

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

Verdolini, K. (1988). Practice good vocal health and prevent those voice disorders. Choristers Guild LETTERS, XL, 40-44.

Verdolini-Marston, K. (1991). The most common vocal injuries: Nodes and polyps and their causes. Voice and Speech Trainers Association, Inc., 5, 1,7.

Verdolini-Marston, K., Sandage, M., & Titze, I.R. (in press). Effect of hydration treatments on laryngeal nodules and polyps and related measures. Journal of Voice.

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

NCVS Status and Progress Report - 4
June 1993, 85-92

Anterior Commissure Microwebs Associated with Vocal Nodules: Detection, Prevalence, and Significance

Charles N. Ford, M.D.
Diane M. Bless, Ph.D.
Guillermo Campos, M.D.
Mark Leddy, M.S.
Clinical Science Center, The University of Wisconsin-Madison

Abstract

Vocal fold nodules are a common cause of dysphonia generally attributed to vocal abuse. Anterior commissure microwebs have been reported as an incidental finding in surgical patients with nodules. In a series of 105 nodule patients evaluated at the University of Wisconsin Clinical Science Center voice laboratory (1987-1992) microwebs were identified in 11 patients. Ten of these patients with microwebs and nodules failed to respond to voice therapy and the microweb was identified at direct microlaryngoscopy. Microwebs were identified in 10 of the 20 patients who underwent surgery for vocal nodules. The presence of these tiny shelves of tissue might be coincidental or might be related to coexistent vocal fold pathology. Changes induced by anterior commissure microwebs, including shortening of the vibrating segment and possible restriction in vertical excursion, theoretically alter the mechanics of vocal fold vibration to favor nodule formation. Symptoms occurring early in life, in a patient with nodules whose hoarseness is refractory to voice therapy, suggests the presence of an occult microweb. Detection requires a high index of suspicion, observation during maximal vocal fold abduction, and clearing of secretions from the anterior commissure. Definitive identification is facilitated by gentle separation of the anterior vocal folds during direct microlaryngoscopy.

Introduction

Vocal nodules are one of the most common causes of dysphonia. A spectrum of histologic changes has been described associated with the condition and there are some similarities to vocal polyps [1,2]. In both conditions the changes are limited to the epithelium and superficial lamina propria (Reinke's space). It is generally accepted that vocal nodules are the result of faulty or excessive vocal use. Excessively high pitch and loudness increase the mechanical trauma to the vibrating vocal fold tissues and both have been implicated in the formation of nodules [3-6]. Voice therapy directed at modifying those factors thought to cause vocal fold trauma generally resolves the problem.

It is unclear why some individuals fail to respond to voice therapy. Although some patients comply with careful vocal hygiene programs, both the dysphonia and nodules persist. These patients often require surgical intervention. Among those patients treated with microsurgery, Bouchayer and Cornut [7] described finding microwebs at the anterior commissure in 23% of patients with nodules. The prevalence of microwebs in this group of patients suggests a relationship between refractory nodules and anterior commissure microwebs. The rate of occurrence of microwebs in otherwise normal larynges is unknown and the association of microwebs in other laryngeal conditions has not been addressed. It is probable that these tiny tissue bands are the cause or effect of trauma associated with the formation of vocal nodules.

If microwebs play a role in the development or maintenance of vocal nodules, one might expect to see frequent references to this condition in the literature. The paucity of references to microwebs may be explained by several factors: (1) microwebs are uncommon; (2) they are difficult to detect with routine clinical examination of the larynx; and (3) they are easily overlooked during microsurgery and may be dismissed as inconsequential by the surgeon.

Materials and Methods

All microlaryngeal surgical procedures performed by the senior author at the University of Wisconsin Clinical Science Center (UW-CSC) during a 5-year period (1987-1992) were reviewed. Videotape recordings revealed anterior commissure web formation of varying degrees, associated surgical trauma, polypoid disease, papillomata, sulcus vocalis, and laryngeal injury. An attempt was made to identify anterior commissure microwebs in all patients with vocal nodules who underwent microlaryngoscopy. During the same period, the records of all patients evaluated in the voice laboratory were reviewed and videotapes of all stroboscopically-studied patients with a diagnosis of nodules were studied.

All adult patients underwent similar evaluations that included: (1) voice recording for acoustic analysis and perceptual judgment; (2) a complete otolaryngologic examination including either indirect laryngoscopy or flexiblefiberoptic visualization with the 3.5-mm Olympus scope; and (3) a videostroboscopic examination using a Bruel and Kjaer 4914 laryngeal videostroboscope coupled to a Wolf 90 -angled rigid endoscope to assess vocal fold motion and glottic closure. Aerodynamic assessment was accomplished with the aid of a Nagashima PS 77 phonatory function analyzer.

For patients who had videostroboscopic recordings, the adduction quotient was used as an index of the length of time the vocal folds were approximated. For the purposes of this study, the adduction quotient was calculated from the videostroboscopic recordings by dividing the total number of video frames in which the vocal folds were closed (fully adducted) by the number of frames they were open (opening and closing phases). All frames were selected from the phonation produced at normal pitch and loudness levels. Four consecutive "apparent cycles" from the videostroboscopic recording were used for analysis in each patient. It is assumed that the adduction is the time when there is the greatest potential for mechanical trauma to the vocal folds. In general, louder and more hyperfunctional phonation patterns are associated with larger adduction quotients.

For those patients whose vocal nodules were managed with surgery, direct microlaryngoscopy was performed with a Dedo rigid laryngoscope and a Wild Leitz surgical microscope provided magnification. The microsurgical instruments used included the heart-shaped forceps and angulated scissors designed by Bouchayer for Micro-France[R] (Instumentarium, Quebec). Typically the microweb and the associated nodules were found in a plane just inferior to the lead edge of the vocal folds (Photo 3; see center-bound plate). Although it is difficult to be certain that such microwebs cause nodules or that they can be permanently eliminated by simple surgical lysis, it seemed reasonable to perform a sharp precise lysis during microsurgical excision of the nodules (Figure 1). In the youngest patient —a 3-year-old boy with small, fusiform nodules (Photo 4; see center-bound plate)— we elected to sharply lyse the microweb, leaving the nodules to resolve spontaneously with the aid of voice therapy. In other instances where one discrete nodule was found and the apposing vocal fold showed only minimal epithelial thickening, lysis of the microweb and unilateral nodule excision seemed sufficient.

Results

Most of the 105 patients reviewed responded well to voice therapy and did not require further management; this group included one patient who appeared to have a small microweb. There were 20 patients (19%) who failed voice therapy and underwent microlaryngoscopy. Of those, 10 (50%) were found to have anterior commissure microwebs. Identification of these microwebs was usually made after aspirating a small pool of mucus at the anterior commissure. Gentle separation of the anterior vocal folds was required to optimally demonstrate the membrane-thin web.

The typical patient with nodules and microweb was a 25-year-old female with hoarseness since early childhood who had failed to respond to repeated trials of voice therapy. Patients were six females and four males; ages ranged from 3 to 45 years (mean = 24.9). The clinical course and range of management can be illustrated by briefly presenting two cases.

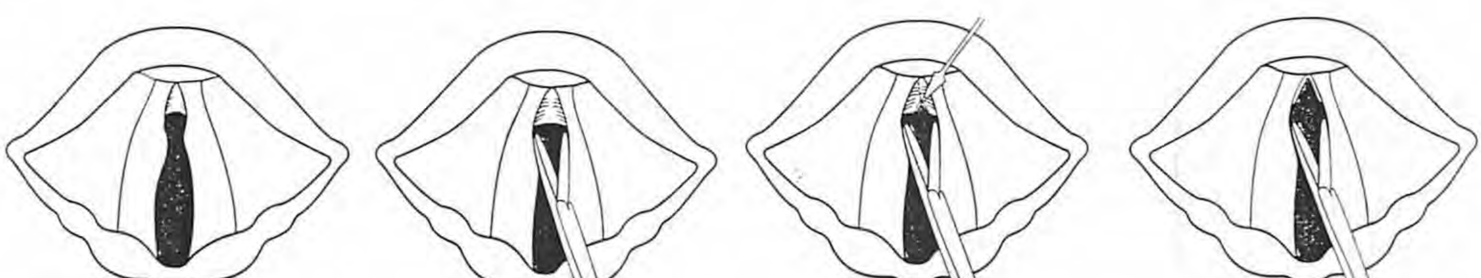


Figure 1A-D. A simplified sketch shows (A) microweb associated with nodules; (B) separation of the vocal folds rendering the web taut—a maneuver that may be necessary to reveal a small web; (C) eccentric lysis of the microweb; and (D) the appearance after the web is cut and the edges retract.

Case Presentations

Case 1

This 32-year-old female music teacher had had problems with hoarseness for as long as she could remember. Vocal-abusive behavior included screaming as a child, and later performing as the lead singer with a rock group. Intermittent voice therapy resulted in slight transient improvement. Associated medical conditions included allergic rhinitis and a nasal septal deviation.

Indirect laryngoscopy revealed bilateral vocal nodules with two discrete nodular formations on the left, apposing one broader-based nodule on the right that extended from the anterior to middle third of the vocal fold. Laryngeal videostroboscopy showed the nodular lesions, uneven medial margins, slight edema of the vocal folds, incomplete posterior glottic closure, and reduced mucosal wave and amplitude of vibration. The adduction quotient was 0.15. Web formation was not appreciated preoperatively.

At surgery an anterior commissure microweb was present 2 mm inferior to the plane of the lead edge of the vocal folds. The nodules were based on the inferomedial vocal fold margins and appeared hyperkeratotic. Each nodule was resected using minimal traction and sharp scissor-dissection technique. The microweb was resected by separating the vocal folds anteriorly and sharply dividing the web to the right of midline. Minimal bleeding was controlled with topical 1:1,000 epinephrine. Histopathologic evaluation showed vocal nodules with acanthosis and parakeratosis. Postoperatively the patient improved with voice therapy and at 6 weeks there was no evidence of recurrent nodules or web formation.

Case 2

The parents of this 3-year-old boy noted that since birth his voice had been abnormal. His cry had been harsh and he continued to exhibit a raspy breathy phonation. We were unable to visualize the larynx during outpatient clinical examination, but because the history was suggestive of a congenital lesion, we undertook direct microlaryngoscopy. At surgery we found non-discrete fusiform nodular swellings at the junction of the anterior and middle third of the vocal folds. A well defined, thin, anterior commissure microweb was also identified (Photo 4; see center-bound plate). Because of the patient's young age and the immature appearance of the nodules, we elected to leave the nodules in place but to resect the microweb. Six months postoperatively his voice was clearly improved but further voice therapy was suggested because of continued vocal-abusive behavior.

Discussion

In this study, patients with vocal nodules were studied over a 5-year period to determine the prevalence of associated anterior commissure microwebs. Comprehensive examinations documented vocal nodules in 105 patients. Many of these patients were referred to UW-CSC after having failed behavioral therapy. Twenty of these patients (19%) eventually underwent microlaryngoscopy and were carefully examined for associated microwebs. The microwebs appeared to be thin shelves of mucosa immediately inferior to the plane of the vocal folds at the anterior commissure. Mucus pooling was usually present on the superior aspect of the microweb and filling the anterior commissure.

Detection of Microwebs

Our results are consistent with the generally accepted belief that voice therapy is the primary modality of treatment for vocal nodules. Of a clinical population of 105 nodule patients, only 20 were taken to surgery. All of these surgical patients had failed voice therapy and some were suspected of having congenital lesions. Anterior commissure microwebs were present in ten of the surgical group. Failure to respond to voice therapy appears to be a strong indicator that associated occult pathology exists and, if the vocal folds appear otherwise normal, careful examination of the anterior commissure may reveal microweb formation.

On the basis of this experience we believe that the prevalence of microwebs is much greater than appreciated. We offer four suggestions that should be helpful in making the diagnosis of anterior commissure microwebs:

(1) maintain a high index of suspicion, especially in patients who develop hoarseness at an unusually early age and in any nodule patient who fails to respond to an adequate trial of voice therapy;
(2) use magnification;
(3) clear the secretions from the anterior commissure; and
(4) attempt to view the inferior aspect of the anterior commissure either by retraction at direct microlaryngoscopy or by having the patient fully abduct with an inhalation maneuver during laryngeal videoendoscopy.

Prevalence of Microwebs

Bouchayer and Cornut found microwebs in 23% of 93 microlaryngoscopic procedures done for nodules[7] and in four of 12 singers with nodules[8]. It is likely, however, that the true incidence among all nodule patients is much lower, for most nodules do not require surgical removal. In the ten cases reported here, certain characteristics suggest that they are not typical nodule patients. The children had long-standing voice abnormalities and were suspected of having congenital lesions. The adults had long-standing hoarseness that failed to resolve with intensive voice therapy; some patients had problems such as allergy or reflux, in addition to vocal abuse, that may have contributed to their vocal fold abnormalities. These associated findings are perhaps common to patients who ultimately undergo surgery but are not typical of all nodule patients. For this reason it is not possible to generalize about the prevalence of microwebs in the non-operative population.

It is tempting to try to explain the prevalence of microwebs in nodule patients by ascribing an etiologic role of webs in nodule formation. The cases presented in this report certainly suggest a strong tendency for webs to coexist with nodules in patients who, for one reason or another, have become surgical candidates. Clearly not all patients with vocal nodules have microwebs, yet microwebs are sometimes associated with other congenital vocal fold lesions. We observed a microweb in one patient with an apparent congenital sulcus vocalis deformity. It is of interest that Bouchayer[9] reviewed 100 successive cases of microwebs and, whereas 78 were patients with nodular lesions, 22 were patients with other lesions such as polyps, cysts, and scars. A review of the similarities and differences between various lesions asthey affect the mechanical properties of the vibratory segment of the vocal folds might help us to understand the development of nodules and the possible contribution of microwebs in the pathogenesis of nodules.

Significance of Microwebs

The association of microwebs with nodules or other forms of vocal fold pathology might be purely coincidental. Indeed, such a small, wispy membrane of tissue seems hardly sufficient to alter vocal fold vibration or induce focal trauma. It is possible that factors responsible for inducing nodules that are refractory to voice therapy might also lead to the formation of anterior commissure microwebs. It is even conceivable that microwebs could be caused by persistent vocal nodules. The finding of anterior commissure microwebs in 50% of operative nodule patients in this series suggests that this is more than a coincidental occurrence.

Arnold[1] expressed the concept that nodules and polyps of the vocal fold develop not only because of chronic irritation but also as the result of several associated factors. He described three categories of factors involved in the pathogenesis of these disorders: (1) <u>predisposing factors</u> that are essentially constitutional, (2) <u>precipitating factors</u> such as allergic reactions, emotional crises, and hormonal imbalances, and (3) <u>aggravating factors</u> like alcohol and tobacco. One of the most important predisposing factors is the overall anatomical structure of the larynx. Those anatomical factors often associated with an excellent voice —a roomy resonating chamber with widely open laryngeal vestibule, a flat epiglottis allowing easy visualization, and overall symmetry— were less commonly found in patients with diseased larynges. Arnold contended that vocal nodules tend to occur in patients who have "unfavorably built organs". It is possible that the peculiarly high incidence of vocal nodules in prepubescent males and adult females reflects the smaller anatomical configuration of the larynx in these groups; this is more plausible than the anachronistic explanation (offered only two decades ago), that women are prone to develop nodules because of "the frustration of motherhood, making it difficult for women to restrain the volume of their vocal rebuke"[2].

Although vocal abuse clearly plays a role in the formation of nodules, prevalence figures relative to gender and age suggest that additional factors must be considered. Short membraneous vocal folds might be a predisposing factor for nodule formation in a speaker who habitually uses a loud voice or in an untrained singer. A shorter vibratory segment causes greater impact at the point of contact where nodules occur during routine phonation; mechanical trauma would be increased with extremes of loudness and pitch. Studies of the prevalence of vocal nodules[10,11] suggest the influence of age and gender. Kahane[12] demonstrated a "high level of morphological congruence" in the prepubertal larynx, even though prepubertal males developed nodules more frequently than females (3:1 male:female ratio). This has been explained on the basis of behavioral differences. Boys speak louder 13 and engage in more aggressive and forceful vocal behavior[14]. Following puberty laryngeal growth is greater in males, accounting for a 63% increase in vocal fold length compared to only a 34% increase in females[12]. The occurrence of nodules in adult females is greater than males, with a reversal of the 3:1 ratio found in children.

The presence of anterior commissure webs alters the normal endolaryngeal configuration by shortening the anterior-to-posterior length of the vocal fold. This loss is essentially in the membranous, vibrating portion of the vocal fold. The mechanical result of a microweb is probably greater than might be anticipated by the loss of length; the most important factor may be the specific location of the lesion and how it alters the functional histoarchitecture in the anterior vocal fold. Study of longitudinal sections of vocal folds indicates that the intermediate layer of the lamina propria forms a modified mass of elastic tissue as the anterior commissure tendon is approached. This modified connective tissue arrangement is called the <u>anterior macula flava</u>. Hirano believes it functions like a cushion and serves "to protect the ends from mechanical damage which may result during vibration of the vocal folds"[15]. Such function would likely be impaired by the presence of a microweb bridging the vocal folds across the anterior commissure. This would effectively shorten the membraneous vocal fold so that greater subglottal pressures and amplitude of vocal fold vibration would be necessary for a given loudness level. Such changes might be expected to result in a prolonged closed phase with increased vocal fold tissue contact and focal tissue trauma that favor development of nodules. Microwebs might not only shorten the effective length of the membraneous vocal fold, but they might restrict vertical movement in the anterior segment and override the protective function of the anterior macula flava. The role of an anterior commissure microweb in the spectrum of events favoring the formation and persistence of vocal nodules can be described as an equilibrium in which vocal abuse and the presence of a microweb impede recovery, often leading to surgical intervention (Figure 2).

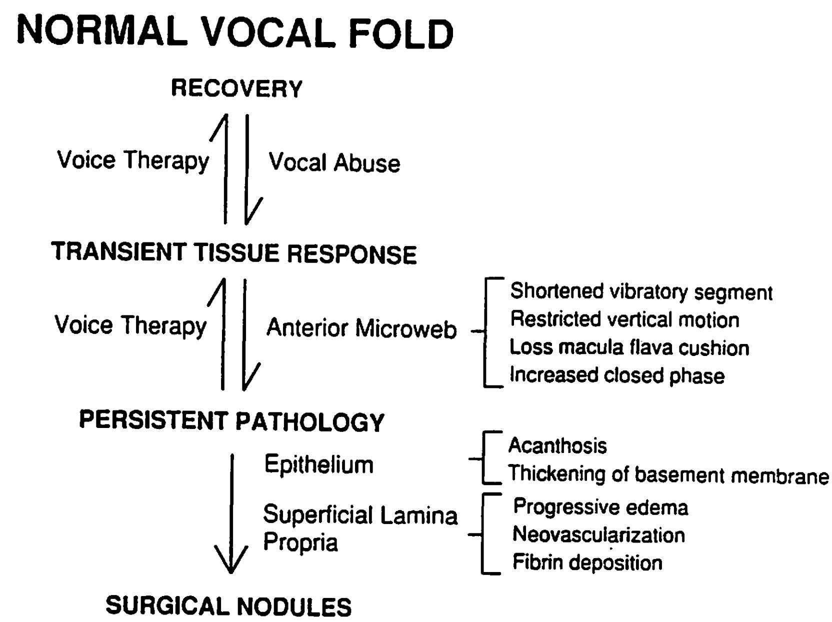


Figure 2. This is a diagrammatic representation of the equilibrium between nodule formation and recovery. The role of anterior commissure microweb in addition to vocal abuse appears to favor nodule formation and impede recovery.

Conclusions

Most vocal nodule patients respond to voice therapy and never undergo microlaryngoscopy. Patients with nodules that require surgery to relieve symptoms are a unique group often exhibiting associated pathology. One frequent associated finding is a thin mucosal web on the inferior surface of the vocal folds at the anterior commissure. In this study, ten anterior commissure microwebs were identified in 20 nodule patients undergoing microlaryngoscopy. Such microwebs are difficult to identify and require deliberate measures for detection. Increased awareness of the possible existence of anterior commissure microwebs should help establish their prevalence in patients with nodules and other related vocal fold pathology. They may be coincidental findings, secondary to the same factors that cause nodules, or causally related to the nodules. It is theoretically possible that anterior commissure microwebs shorten the vibratory segment, impede the cushioning effect of the macula flava, and restrict the vertical excursion of the vocal folds. Although webs might alter the biomechanical and vibratory properties of the vocal fold, further research in the laboratory is necessary to elucidate any relationship between anterior commissure microwebs, vocal fold vibration, and the probability of nodule formation.

Acknowledgement

This work was supported by the National Center for Voice and Speech through Grant P60 00976 from the National Institute on Deafness and Other Communication Disorders.

Bibliography

1. Arnold, G. E.: Vocal Nodules and Polyps: Laryngeal Tissue Reaction to Habitual Hyperkinetic Dysphonia. JSHD, 27:205-217, 1962.
2. Strong, M. S., and Vaughan, C. W.: Vocal Nodules and Polyps —The Role of Surgical Treat ment. Laryngoscope, 81:911-923, 1971.
3. Brodnitz, F. S.: Vocal Rehabilitation in Benign Lesions of the Vocal Cords. JSHD, 23:112-117, 1958.
4. Rubin, H. J., and Lehrhoff, I.: Pathogenesis and Treatment of Vocal Nodules. JSHD, 27:150-161, 1962.
5. Morrison, M. D., Rammage, L. A., Belisle, G., and Nichol, H.: Muscular Tension Dysphonia. J. Otolaryngol., 12:302-307, 1983.
6. Hirano, M., and Kurita, S.: Histologic Structure of the Vocal Fold and Its Normal and Pathological Variations. In: ·Vocal Fold Histopathology·, Kichner, J.A. (Ed.). College Hill Press, San Diego, 1986.
7. Bouchayer, M. and Cornut, G.: Microsurgery for Benign Lesions of the Vocal Folds. ENT Journal, 67:446-466, 1988.
8. Cornut, G., and Bouchayer, M.: Phonosurgery for Singers. J. Voice, 3:269-276, 1989.
9. Bouchayer, M.: Congenital Lesions of the Vocal Fold. Proceedings of First Scandinavian Seminar on Phonosurgery, Stockholm, 1990.
10. Senturia, B.H. and Wilson, F.B.: Otorhinolaryngic Findings in Children with Voice Deviations. Annal. Otol. Rhinol. Laryngol., 77:1027-1042, 1968.
11. Silverman, E. and Zimmer, C.H.: Incidence of Chronic Hoarseness Among School-Age Chil dren. J. Speech Hearing Dis., 40:211-214, 1974.
12. Kahane, J.C.: Growth of the Human Prepubertal and Pubertal Larynx. J. Speech Hearing Res., 25:446-455, 1982.
13. Susser, R.D.: Vocal Intensity Levels of Normal and Voice-Disordered Children. Masters Thesis, Univ. Wisconsin-Madison, 1980.
14. De Pietro, S.: Rough and Tumble Play: A Function of Gender. Develop. Psychol., 17,1:50-58, 1981.
15. Hirano, M.: Clinical Examination of the Voice. Springer-Verlag, New York, 1981.

NCVS Status and Progress Report - 4
June 1993, 93-111

Comparison of Aerodynamic and Electroglottographic Parameters in Evaluating Clinically Relevant Voicing Patterns

K. Linnea Peterson, M.D.
Department of Otolaryngology - Head and Neck Surgery, The University of Iowa
Katherine Verdolini-Marston, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Julie M. Barkmeier, M.A.
Department of Otolaryngology - Head and Neck Surgery, The University of Iowa
and
Department of Speech Pathology and Audiology, The University of Iowa
Henry T. Hoffman, M.D., M.S.
Department of Otolaryngology - Head and Neck Surgery, The University of Iowa

Abstract

The purpose of the present study was to identify one or more aerodynamic or electroglottographic (EGG) measures that distinguishes between clinically relevant voicing patterns for nodule pathogenesis and regression: a presumably pathogenic pattern (pressed voice), a neutral pattern (normal voice), and two presumably therapeutic patterns (resonant voice and breathy voice). Trained subjects with normal voices produced several tokens of each voice type on sustained vowels /a/, /i/, and /u/. For each token, maximum flow declination rate (MFDR), AC flow, and minimum flow were obtained from inverse-filtered airflow signals, and closed quotient and closing time were obtained from electroglottographic (EGG) signals. The results indicate that for /a/ and /i/ (but not for /u/), the closed quotient provided a sensitive tool for distinguishing the voice types in physiologically interpretable directions. Further, post-hoc analyses confirmed a direct relation between the closed quotient and videoscopic ratings of laryngeal adduction, which previous work links to nodule pathogenesis and regression.

Vocal fold nodules are benign space-occupying lesions which develop at the midpoint of the membranous portion of the vocal folds. Nodules appear to be caused by intraglottal trauma (1, 2, 3, 4), although a direct link has not been unequivocally established. The causal pattern that has been most convincingly argued involves two successive hypotheses. First, chronic adductory hyperfunction during phonation produces high vocal fold contact stress (force of collision per unit area of tissue contact) during phonation and second, high vocal fold contact stress causes local trauma and, over time, chronic lesions (5). Jiang (6) provided evidence simultaneously consistent with both hypotheses in an experiment using excised canine hemilarynx preparations. In his study, increased arytenoid approximation during induced phonation resulted in increased contact stresses along the excised vocal folds, and in general the greatest contact stresses occurred at the midpoint of the membranous folds.

Not surprisingly, behavioral therapy for nodule reduction generally focuses on voicing patterns assumed to minimize focal trauma, by limiting intraglottal contact stress. Examples include "breathy" voice and "resonant" voice (7, 8, 9, 10, 11). "Breathy" voice is associated with the auditory perception of breathiness, or audible air escapage during phonation. "Resonant" voice is associated with the proprioceptive perception of oral vibratory sensations on or near the alveolar ridge and other facial plates, and in healthy voices, the auditory impression of a "ringing" voice. Preliminary analyses from an ongoing study indicate that in fact the training of both breathy and resonant voice types may provide some benefit in the behavioral management of nodules (11). The results from such studies may be clinically useful. However, the direct investigation of the physiological mechanisms that regulate nodule formation and regression is currently limited by the lack of quantitative measurement tools for *in vivo* situations. The purpose of the present study was to address this limitation by identifying one or more quantitative measures that may effectively distinguish between four critical voice production types in clinically and physiologically interpretable directions: pressed voice (a voice type presumed to be pathogenic), normal voice (a neutral voice type), and resonant and breathy voice (presumably therapeutic voice types). The results could be applied to future physiological studies investigating these voicing patterns and their contribution to the pathogenic process. The results may also provide some information about physiological factors in nodule pathogenesis and regression.

The measures that were selected for investigation included a series of aerodynamic and electroglottographic (EGG) measures, previously described by other authors (5, 12-19), that may reflect vocal fold contact stress, or force per unit area of vocal fold contact. The aerodynamic measures were derived from inverse-filtered airflow signals, and included maximum flow declination rate (MFDR), AC flow, and minimum flow. The EGG measures were closed quotient and closing time.

MFDR indicates the maximum rate of vocal fold deceleration during closing. According to some authors, this measure may reflect the force of vocal fold contact during phonation (5). AC flow indicates the amount of modulated air during phonation. This measure varies inversely with MFDR (20), and thus might corroborate the findings for MFDR or even prove to be a more sensitive measure. Minimum airflow, which reflects the amount of unmodulated airflow through the glottis during phonation, may inversely indicate vocal fold contact area, or the denominator in the contact stress parameter (10). Regarding the EGG measures, closed quotient (proportion of the glottal cycle during which the folds are together) may indicate the relative degree of medial compression or arytenoid "pressing", and therefore variations in contact stress. In fact, according to at least one report, increases in the closed quotient correspond to increases in laryngeal adduction or "pressing" (21), and presumably, following Jiang's evidence, also increases in intraglottal contact stress (6). EGG closing time might reflect vocal fold collisional forces, assuming constant frequencies and amplitudes.

An effective measure was defined *a priori* as a measure that statistically distinguishes between the voice types in clinically and physiologically interpretable directions. That is, a conceptual plot was envisioned for each dependent variable (aerodynamic and EGG) over three critical spaces: a pathogenic space (pressed voice), a neutral space (normal voice), and a therapeutic space (resonant and breathy voice). An effective measure would satisfy the following criteria: (a) when plotted against these voicing patterns, over the boundary from the pathogenic to the neutral space, the curve would reflect a statistically significant discontinuity that is physiologically interpretable with respect to pathogenesis, and (b) from the neutral to the therapeutic space, the curve would reflect either no change, continued discontinuity in the same direction as from the pathogenic to the neutral space, or a reversed discontinuity the value of which does not equal or exceed the value indicated in the pathogenic space, regardless of which therapeutic voice type is considered.

As a final introductory comment, it should be noted that the voice types used in this investigation were *assumed* to represent pathogenic, neutral, and therapeutic production modes, based on clinical impressions and also previous data (6,11). However, no attempt was made to actually cause lesions, nor to reverse them, in the present study. Such manipulations will be left to other studies.

Methods

Subjects

Seven adults with extensive voice training (six singers and one actor) participated as volunteers in the study. Trained voice users were selected because of their presumed ability to volitionally vary voice production patterns. Subjects included four males and three females, ranging in age from 21 to 40 years (average age = 28.2 yr). Only nonsmokers and subjects without current illness were included. The average length of previous voice training was 9.4 years, with a range of 5-15 years. All subjects were engaged in regular voice performance requiring loud voice output in large halls, and on questioning, all confirmed familiarity with "resonant voice" (one of the voicing patterns evaluated in this study). All subjects had a negative history of voice disorders, despite their demanding performance schedules, except for one subject (Subject 7) who had incipient contact ulcers six months prior to his participation in the study. This subject was asymptomatic at the time of his participation and exhibited no evidence of dysphonia as assessed by the examiners. Thus his results were retained and are included with those for other subjects. Subjects were unaware of the specific purpose of the study.

Equipment

The equipment used to collect aerodynamic and EGG measures is shown in Figure 1 (following page). A Rothenberg circumferentially vented face mask connected to the input of a Glottal Enterprises amplifier (MS100-A2) was used to collect airflow signals. The main output of the amplifier was routed through an attenuator (Hewlett Packard 350D), and then in parallel to: (1) a voltmeter (Hewlett Packard 3466A Digital Multimeter) for on-line calibration of airflow signals; (2) a model PC-108M SONY Digital Audio Tape (DAT) recorder for permanent data storage; and (3) to a Gateway2000 386 computer for direct data collection using Hypersignal software. EGG signals were collected with a SynchroVoice Electroglottograph, connected to the DAT recorder for long-term storage of signals. Acoustic signals were also collected, using a Sony Electret Condenser Microphone (ECM-44B) powered by a Symetrix SX202 Dual Mic pre-amplifier. As for airflow and EGG signals, the acoustic signal was conducted to the DAT recorder for storage and possible later analysis. The three outputs from the DAT (airflow, EGG, and acoustic signals) were connected to a DATA 6000 oscilloscope for monitoring of active signals.

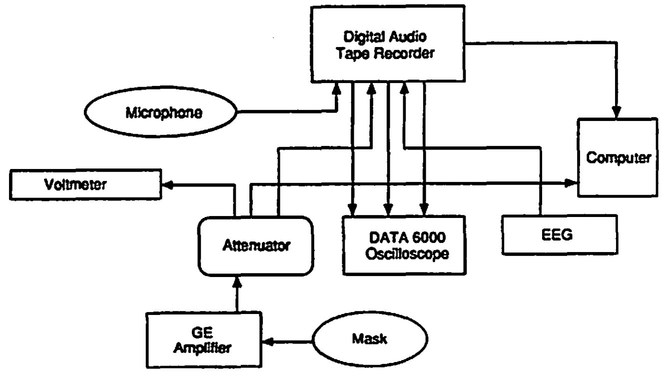


Figure 1. Equipment setup.

Aerodynamic and EGG waveforms were analyzed using CSpeech (Version 3.1) software after digitizing the signals into Hypersignal. A Casio keyboard was used both to extract an average conversational pitch prior to experimental trials and later, to provide a constant target pitch for these trials.

Videostroboscopic images were obtained using a 90 degree R. Wolf 4450.47 rigid telescopic endoscope in conjunction with a Karl Storz 9000 Mini Solid State CCD Video Camera and a Bruel and Kjaer Rhino-Larynx Stroboscope light source, Type 4914.

Equipment and software management for data collection, and explanation of aerodynamic, EGG, and videoscopic measures

During data collection, aerodynamic signals were simultaneously digitized into the computer and stored on a DAT tape. EGG and acoustic signals were stored on the DAT tape and digitized at a later time. Both airflow and EGG input channels to the DAT were set to a sampling frequency of 10 KHz, and the acoustic input signal was set at a sampling frequency of 20 KHz. The voltages of all three input channels to the DAT were monitored for signal saturation throughout the experiment. The voltmeter, set on a +/- 20 volt scale and receiving input from the airflow system, was also monitored constantly and rezeroed as necessary to correct for drift in the airflow signal. Zero-input voltages were noted on the voltmeter for each phonation token during data collection.

For airflow analyses, the data were digitized in Hypersignal at 10 KHz, then each airflow signal was imported to CSpeech. All subsequent manipulations of the airflow signal were done using this software program. The initial portion of the signal, which represented zero flow input, was first measured to obtain a baseline offset reading. For AC and minimum flow, three 20-ms sections from the midportion of each signal were chosen randomly for analysis, and each of the segments was analyzed in the same way. The selected segment was inverse-filtered by first differentiating it, followed by linear prediction correlation analysis, integration, and a final linear prediction correlation analysis. The middle peak from the resulting sample display was selected, and the peak value for this wave was measured to indicate peak flow in liters per second (l/s) after subtracting the baseline offset value. The trough of the same cycle was measured to reflect minimum flow, also in l/s, again correcting for any nonzero offset (Figure 2). AC flow was later calculated by subtracting the minimum flow value from the peak flow value. After peak and minimum flow values were obtained, the waveform was then differentiated again and the minimum point on the resulting middle wave was measured to reflect maximum flow declination rate (MFDR), in liters per second^2 (l/s^2).

Following data collection, the EGG signals were also recorded into Hypersignal at 10 KHz and then imported into CSpeech. Again, for each token imported, three 35-ms segments were selected randomly from the midportion of each signal for independent analyses, and each was analyzed in the following way. The waveforms were oriented so that the open portion of the glottal cycle was represented at the top of the wave. A middle waveform in the display was selected for measurement of closed quotient, open quotient, and closing time, as indicated in Figure 3. The measurements of the closed and closed quotients were taken at 65% of the amplitude of the waveform as described by Rothenberg and Mahshie (17). EGG data were discarded for two subjects (Subjects 5 and 6) due to uninterpretable signals. EGG was not performed on one subject (Subject 2).

Videostroboscopic equipment was managed in the standard clinical fashion for collection of videostroboscopic images. Videoscopic images were analyzed after the results for aerodynamic and EGG measures were evaluated, to confirm or disconfirm the hypotheses about the physiology of the voice types that emerged based on these findings. Specifically, videoscopic images of subjects' larynges during the production of pressed, normal, resonant, and breathy tokens on the vowel /i/ were presented in random order, without soundtrack, to two speech pathologists with extensive experience in videoscopic evaluation, and to two attending laryngologists at our university. The viewer-judges rated the degree of perceived laryngeal adduction for each voice type on an ordinal scale from -5 (extreme hypoadduction) to +5 (extreme hyperadduction). All judges were uninformed about the intended voice types for each segment at the time of viewing.

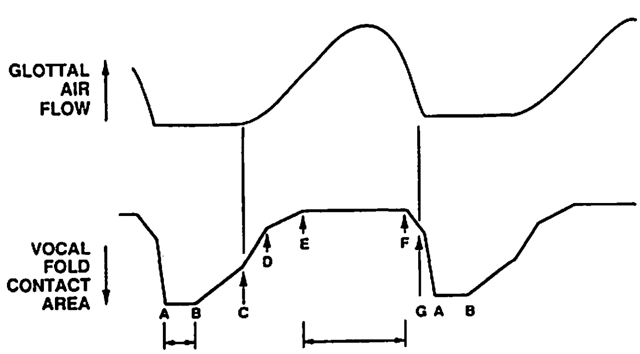


Figure 2. Airflow signal measurements.

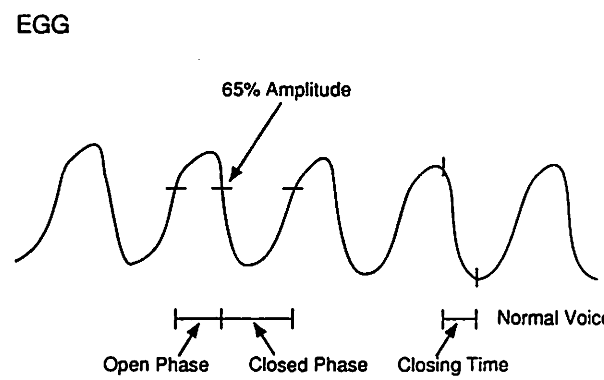


Figure 3. Electroglottographic measurements.

Procedures

Subjects were tested individually. Experimenters maintained consistent roles throughout the data collection. For each subject, the approximate average fundamental frequency was first obtained during a rote speech task (counting). That is, the subject counted out loud from one to five, and his/her pitch on the sustained vowel in the word "three" was matched to a keyboard pitch and subsequently confirmed (in frequency) using CSpeech. The pitch identified in this way was used for all subsequent vowel productions for that subject. Next, the subject was given a demonstration of the four voice types that would be required for critical trials: pressed, resonant, normal, and breathy voice (the model for breathy voice was relatively quiet). For this study, and based on the reports of previous authors (6, 11), pressed voice was assumed to be a pathogenic voice type, and resonant and breathy voice were considered therapeutic voice types, and will be discussed as such for the remainder of the paper. After receiving models of the four voice types from the examiner, the subject practiced the voice types on the vowel /o/, on the designated pitch provided vocally and with a portable keyboard. This procedure was repeated until both the subject and the experimenter were satisfied that the target voice types could be consistently produced. For most subjects, only one practice trial was required for each voice type. During practice trials, the subject was also trained to place the airflow mask firmly over the nose and mouth one second after data collection was initiated (the first second of data collection provided a zero-input signal to be used for calibration during subsequent analyses), to hold the mask firmly in place for four seconds, and then to sustain the target production at the designated pitch for four seconds. The EGG collar and the audio microphone were in place throughout. The subject was then provided with a list of tokens for the experimental trials. The list indicated a random ordering of three different vowels (/i/, /a/, and /u/) and within each vowel, a random ordering of four voice types (pressed, normal, resonant, and breathy), including three tokens of each voice type. Thus, each subject produced a total of 36 voice tokens (3 vowels x 12 tokens per vowel), as trained. The orders of vowel and voice type were varied across subjects. All subjects were encouraged to repeat any token that they judged to be a poor exemplar of the target voice type. (Only three tokens were repeated in the entire experiment). As a validity check, two blind and independent experimenters indicated on separate scoring sheets the voice type perceived for each trial. Later analyses indicated a 79% agreement across trials. That is, on approximately 79% of trials, both of the judges perceived the voice type that the subject in fact intended to produce. All tokens produced were included in the analyses.

The vowels produced in this way provided simultaneous airflow and EGG and acoustic signals for analysis in this study. Following these productions, videoscopic images were obtained of the four voice types. That is, subjects produced at least one token of each of the four voice types during the vowel /i/, under videostroboscopic examination. Videoscopic tokens were also judged for voice type by two independent, blind raters. Analyses of these tokens resulted in 93% agreement across trials (i.e. on 93% of trials, both judges perceived the voice type that the subject intended to produce). Tokens in question were noted immediately and the subject was asked to repeat the trial. No token required more than one repetition. Only valid tokens were included in subsequent analyses.

Analyses and design

For aerodynamic measures, the primary analyses of interest involved the vowel /a/ because of this vowel's suitability for inverse-filtering, related to a high-frequency first formant. Aerodynamic analyses for the vowel /i/ and /u/ were also carried out, because airflow signals for these vowels were collected simultaneously with EGG signals. Although the aerodynamic analyses for /i/ and /u/ were of limited conceptual interest because of their low-frequency first formant encroaching on the fundamental frequency especially in female voices, the results of analyses for these vowels are nonetheless reported. For EGG analyses, the analyses for all vowels were of conceptual interest. Videoscopic analyses involved only the vowel /i/, as this was the only vowel that was assessed endoscopically.

Because of the focus on different vowels depending on the particular measure, and also because initial statistical analyses indicated interactions between vowel and voice type, for both aerodynamic and EGG measures the analyses reported here involve the results for the three vowels assessed independently, with an emphasis on those particularly relevant for the measure in question. For all measures, statistical analyses involved paired comparisons between all four voice types, with a total of six comparisons (pressed vs normal, pressed vs resonant, pressed vs breathy, normal vs resonant, normal vs breathy, and resonant vs breathy). For aerodynamic and EGG measures, the analyses involve a randomized block design with one within-subject fixed factor, voice type (x 4). For videoscopic measures, analyses also included a random factor of judge (x 4).

To reiterate, the criteria for identifying an effective measure used the plot of the measure over voice type: (a) from the pathogenic to the neutral space, the curve would reflect a statistically significant discontinuity that was physiologically interpretable with respect to pathogenesis, and (b) over the boundary from the neutral space to the therapeutic space, the curve would reflect either no change, continued discontinuity in the same direction as from the pathogenic to the neutral space, or a reversed discontinuity the value of which did not equal or exceed the value indicated in the pathogenic space, regardless of which therapeutic voice type was considered. It turns out that to satisfy these criteria, the paired comparisons that had to indicate significant differences (in physiologically interpretable directions) were comparisons between pressed voice and all other voice types (pressed vs normal, pressed vs resonant, and pressed vs breathy). The remaining paired comparisons might or might not differ.

Results

As already noted, airflow and videoscopic data were obtained for all 7 subjects, but interpretable EGG signals were obtained only for 4 subjects. The results for both data sets (N = 4 and N = 7) are indicated as appropriate in the tables and figures, and are discussed in the text. Tables 1 - 4 and 6, and Figures 4 - 11 display the average values obtained across subjects. Individual results for the closed quotient of the EGG signal are indicated in Table 5, because analyses indicated that these values were of particular interest. The significance level was set at $p < .05$ for all statistical tests.

MFDR

Average values for MFDR as a function of voice type are shown in Table 1 and in Figure 4. The general trend was that the highest values were obtained for pressed and for resonant voice, and lower values were obtained for normal and breathy voice. If MFDR reflects the potential for intraglottal trauma, the relatively high value obtained for one presumably therapeutic voice type (resonant voice) is counter-intuitive. Statistical analysis confirmed that MFDR failed to differentiate between the voice types according to the pre-established criteria. Focusing on the vowel /a/, the primary reason was a failure to reliably distinguish between the pathogenic voice type (pressed voice) and one therapeutic voice type (resonant voice) for sample sizes of both N = 4 and N = 7, and further, between the pathogenic voice type (pressed voice) and the second therapeutic voice type (breathy voice) for the sample of N = 4. Also for the vowels /i/ and /u/, which were actually not of conceptual interest for this or other airflow measures, for both N = 4 and N = 7, several of the paired comparisons between pressed voice and other voice types did not indicate significant differences, or revealed differences in the wrong direction: the three significant distinctions that were obtained between pressed voice and resonant voice (of a possible four distinctions) indicated *greater* MFDRs (and presumably collisional forces) for a therapeutic voice type (resonant voice), as compared to the pathogenic voice type (pressed voice).

Table 1.

Average maximum flow declination rate (MFDR) in liters/sec.² for four voice types. Significant paired comparisons also indicated with asterisk ($p < .05$).

VOICE TYPE

VOWEL	Pressed	Normal	Resonant	Breathy
/a/ N=4	-303.28	-232.34	-303.39	-262.62
N=7	-301.70	-227.89	-293.11	-198.54
/i/ N=4	-168.45	-157.48	-205.40	-174.30
N=7	-207.29	-172.30	-238.31	-162.89
/u/ N=4	-202.23	-172.16	-237.62	-175.86
N=7	-214.29	-186.05	-241.83	-159.21

SIGNIFICANT PAIRED COMPARISONS

VOWEL	Pressed/ Normal	Pressed/ Resonant	Pressed/ Breathy	Normal/ Resonant	Normal/ Breathy	Resonant/ Breathy
/a/ N=4	*			*		
N=7	*		*	*		*
/i/ N=4		*		*		
N=7	*	*	*	*		*
/u/ N=4				*		*
N=7	*	*	*	*	*	*

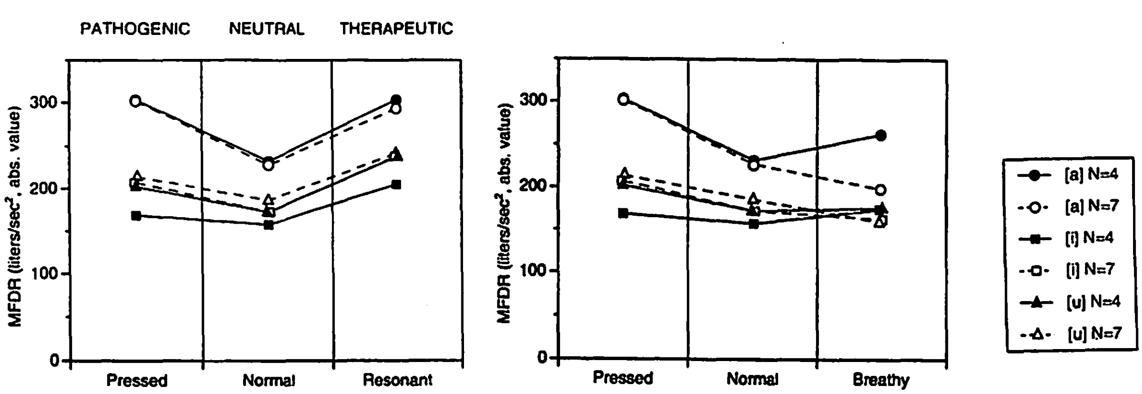


Figure 4. Average maximum flow declination rate (MFDR) in liters/sec2 (absolute value), as a function of voice type, for vowels [a], [i], and [u] (N=4 and N=7).

Table 2.
Average AC flow in liters/sec, for four voice types
Significant paired comparisons also indicated with asterisk (p<.05).

VOICE TYPE

VOWEL	Pressed	Normal	Resonant	Breathy
/a/ N=4	0.196	0.199	0.223	0.233
N=7	0.192	0.198	0.216	0.224
/i/ N=4	0.173	0.176	0.232	0.220
N=7	0.198	0.196	0.246	0.231
/u/ N=4	0.206	0.213	0.257	0.193
N=7	0.209	0.214	0.247	0.210

SIGNIFICANT PAIRED COMPARISONS

VOWEL	Pressed/ Normal	Pressed/ Resonant	Pressed/ Breathy	Normal/ Resonant	Normal/ Breathy	Resonant/ Breathy
/a/ N=4			*		*	
N=7		*	*		*	
/i/ N=4		*	*	*	*	
N=7		*	*	*	*	
/u/ N=4		*		*		*
N=7		*		*		*

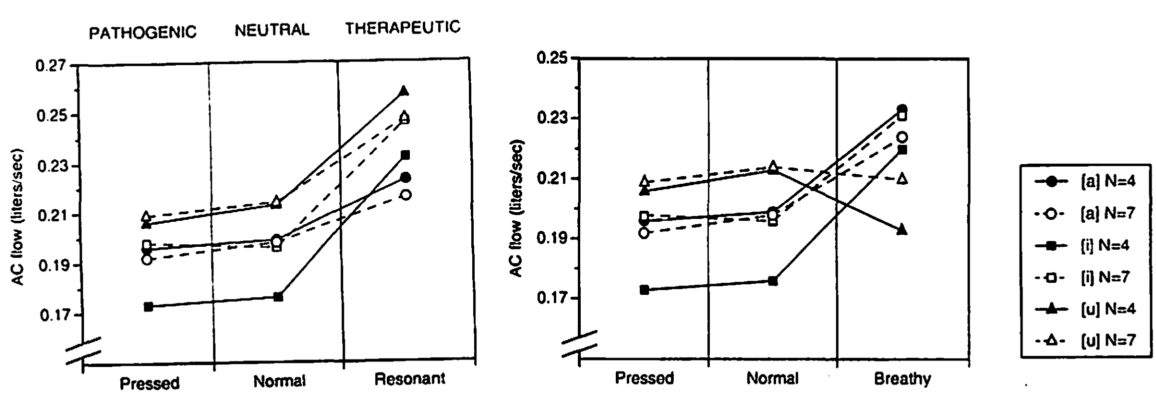


Figure 5. Average AC flow, in liters/sec, as a function of voice type, for vowels [a], [i], and [u] (N=4 and N=7).

AC flow

Table 2 and Figure 5 show the average values for AC flow. The general trend was that relatively high AC flow values were obtained for resonant and for breathy voice, and relatively lower values were obtained for normal and pressed voice. If AC flow reflects the potential for tissue injury, also this result is counterintuitive. Statistical tests further confirmed that this measure did not effectively distinguish the voice types according to pre-established criteria. Across all vowels, for both N = 4 and N = 7, none of the paired comparisons between pressed and normal voice were reliable. Further, the significant comparisons that were obtained between pressed and resonant voice, and between pressed and breathy voice across vowels were in the wrong direction, compared to what was anticipated: as already noted by inspection of the data, AC flows were *higher* for the therapeutic, as compared to the pathogenic voice types.

Because of the covariation of AC flow with MFDR reported by previous authors (20), the relation between these measures was evaluated for our data. Pearson *r*s were calculated between AC flow and MFDR, using individual data and collapsing across vowels and voice types. For both sample sizes, the correlations were negative and reliably different from chance: $r = -.75$, $p = .0001$ for N = 4, and $r = -.71$, $p = .0001$, for N = 7. Thus, our data generally reflect the same inverse relation between MFDR and AC flow described by the previous authors (20).

Minimum flow

For minimum flow, the results were no more encouraging than for the other two aerodynamic measures. The average results for this measure are shown in Table 3 and Figure 6. Minimum flow was generally low for pressed, normal, and resonant voice types, and was considerably higher for breathy voice. For the vowel /a/, the only comparison between pressed voice and the other voice types that indicated a statistically significant difference was the comparison between pressed voice and breathy voice (N = 4 and N = 7). Similar results were seen for the other vowels, with the addition of a statistically significant difference between pressed voice and normal voice for /u/ (N = 4). Thus, minimum flow was discarded as a useful investigatory tool, in particular because it failed to consistently distinguish between the pathogenic voice type and neutral voice, and between the pathogenic voice type and one therapeutic voice type (resonant voice).

Note that with the exception of breathy voice, average minimum flows were negative. We feel quite confident that this finding does not reflect calibration errors, not only because we monitored for signal drift during data collection but also because for each token produced, a zero-input flow value was obtained and subtracted from all subsequent values for minimum flow. Our interpretation is that negative minimum flows in this study reflect incompletely filtered vocal tract resonance phenomena, or perhaps laryngeal depression during vocal fold closure. In any event, negative minimum flows do not reflect a backflow of air across the glottis during phonation.

Closed quotient

The average values for the (EGG) closed quotient are shown in Table 4 and in Figure 7. Across all vowels, the highest closed quotients were obtained for pressed voice, intermediate values were obtained for normal and for resonant voice, and the lowest values were obtained for breathy voice. For /a/ and /i/, both of which were of interest, statistical analyses indicated that this measure acceptably distinguished the voice types according to the pre-established criteria. For both of these vowels, all of the paired comparisons between pressed voice and the other voice types indicated significant differences. Further, the differences were in a direction that was physiologically interpretable with respect to the pathogenic continuum. Assuming that as indicated in previous studies, increases in the closed quotient reflect increasing laryngeal adduction (21), and further that increasing adduction leads to increased local contact stress (6), pressed voice might be associated with the greatest contact stresses and normal, resonant, and breathy voice with relatively lower contact stresses. The relation between the closed quotient and laryngeal adduction in the present study was investigated with follow-up videoscopic analyses, discussed shortly.

Table 3.
Average minimum flow in liters/sec for four voice types.
Significant paired comparisons also indicated with asterisk ($p < .05$).

VOICE TYPE

VOWEL	Pressed	Normal	Resonant	Breathy
/a/ N=4	-0.03	-0.02	-0.03	0.09
N=7	-0.03	-0.02	-0.04	0.07
/i/ N=4	-0.04	-0.04	-0.06	0.06
N=7	-0.05	-0.04	-0.07	0.06
/u/ N=4	-0.06	-0.04	-0.07	0.04
N=7	-0.06	-0.05	-0.06	0.05

SIGNIFICANT PAIRED COMPARISONS

VOWEL	Pressed/ Normal	Pressed/ Resonant	Pressed/ Breathy	Normal/ Resonant	Normal/ Breathy	Resonant/ Breathy
/a/ N=4			*		*	*
N=7			*		*	*
/i/ N=4			*		*	*
N=7			*	*	*	*
/u/ N=4	*		*	*	*	*
N=7			*	*	*	*

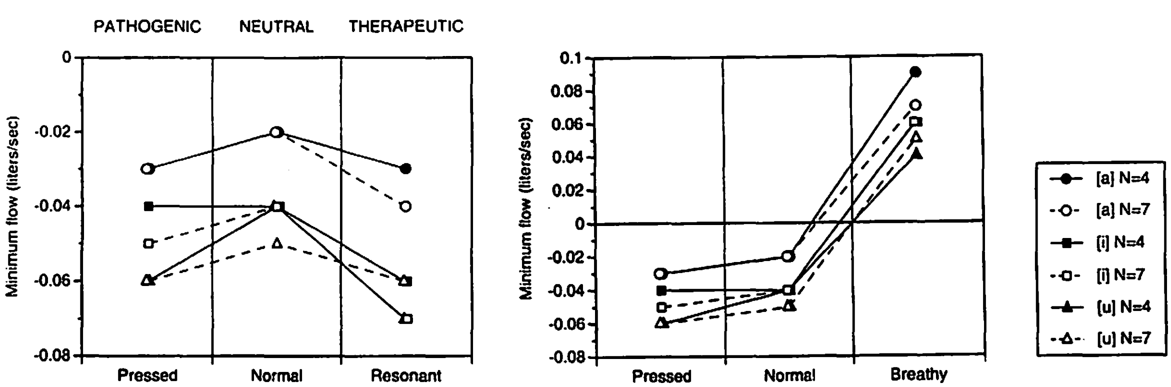


Figure 6. Average minimum flow, in liters/sec, as a function of voice type, for vowels [a], [i], and [u] (N=4 and N=7).

Table 4.
Average closed quotient in proportion of total cycle for four voice types.
Significant paired comparisons also indicated with asterisk ($p < 0.5$).

VOICE TYPE

VOWEL	Pressed	Normal	Resonant	Breathy
/a/ N=4	0.61	0.54	0.54	0.48
/i/ N=4	0.64	0.57	0.55	0.47
/u/ N=4	0.59	0.56	0.58	0.44

SIGNIFICANT PAIRED COMPARISONS

VOWEL	Pressed/ Normal	Pressed/ Resonant	Pressed/ Breathy	Normal/ Resonant	Normal/ Breathy	Resonant/ Breathy
/a/ N=4	*	*	*		*	*
/i/ N=4	*	*	*		*	*
/u/ N=4			*		*	*

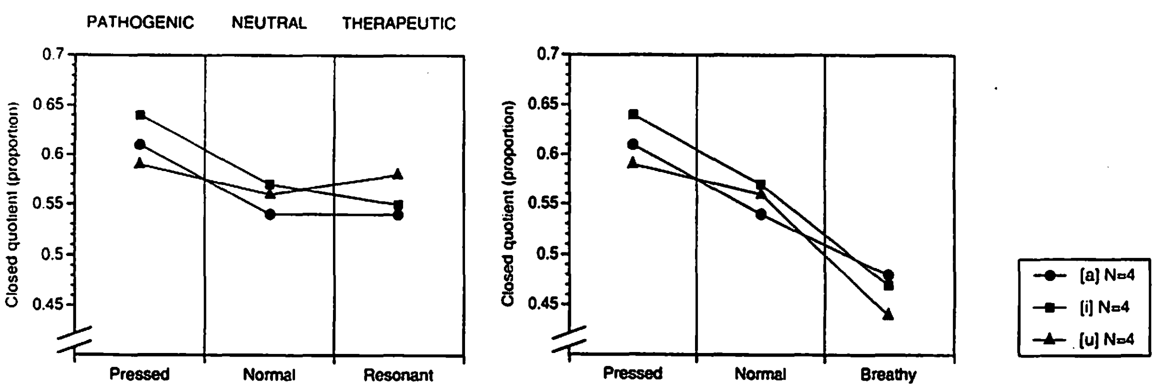


Figure 7. Average closed quotient, in proportion in total cycle, as a function of voice type, for vowels [a], [i], and [u] (N=4 and N=7).

For the vowel /u/, the closed quotient was unacceptable as a useful measure to distinguish the voice types. For this vowel, the closed quotient failed to consistently distinguish between pressed voice and all other voice types. Specifically, the comparisons between pressed and normal voice, and between pressed and resonant voice, did not indicate significant differences.

Because the average values for the closed quotient effectively differentiated the voice types for /a/ and /i/, individual data were inspected for these vowels and are displayed in Table 5. For the vowel /a/, for three of the four subjects for whom EGG data were available, the closed quotient was greatest for the pathogenic voice type (pressed voice). For the vowel /i/, for two of four subjects the quotient was the greatest for this voice type. For both /a/ and /i/, Subject 3 was an exception to the trend noted (in addition to Subject 7 for /i/).

Table 5.
Individual data for closed quotient in proportion of total cycle, for four voice types. (Vowels /a/ and /i/ only, N = 4).

VOICE TYPE

VOWEL/ SUBJECT	**Pressed**	**Normal**	**Resonant**	**Breathy**
/a/ S.1	0.61	0.56	0.53	0.49
S.3	0.54	0.61	0.58	(1)
S.4	0.61	0.46	0.49	0.50
S.7	0.61	0.54	0.57	0.46
/i/ S.1	0.61	0.54	0.53	0.51
S.3	0.67	0.67	0.62	0.46
S.4	0.70	0.49	0.46	0.45
S.7	0.58	0.58	0.59	0.45

(1) Missing data

Closing time

The notion that the EGG closing time might reflect force of vocal fold closure was based on the assumption of a constant frequency and amplitude of vocal fold vibration across voice types. Although frequencies were kept constant within subjects by the experimental protocol, AC flow data indicated that the assumption of a constant amplitude was not met. Therefore, this measure was discarded as a useful investigatory tool.[1]

Videoscopic evaluation and comparison of visual ratings with closed and open quotients.

Videoscopic evaluations of the larynx were conducted as a follow-up to the results for the EGG closed quotient, to determine if as reported by previous authors (21), the closed quotient might reflect laryngeal adduction level. Before comparing the closed quotient to the videoscopic ratings, the consistency of laryngeal ratings across judges was assessed, and differences in laryngeal ratings were also evaluated as a function of voice type. Two independent, two-way analyses of variance (ANOVAs) were conducted (for N = 4 and N = 7), with voice type (4) as a within-subjects factor and judge (4) as a random factor. In both ANOVAs, the effect of judge was unreliable. Thus, judges appeared consistent in their ratings of laryngeal adduction. The results for the laryngeal ratings themselves are indicated in Table 6 and in Figure 8. Adduction ratings clearly varied as a function of voice type, in the anticipated direction. Both when four and seven subjects were considered, the highest adduction ratings were obtained for pressed voice, consistent with the impression

[1]Note that an attempt was made to normalize closing time data by amplitude. For this manipulation, AC flow, which reflects vocal fold amplitude, was divided by closing time, for each subject, for each token produced. The resulting variable approximately represented average vocal deceleration during closing. This variable is a poor one because AC flow was used as the numerator,

of hyperadduction, intermediate ratings close to neutral were obtained for normal and for resonant voice, and the lowest ratings were obtained for breathy voice, consistent with hypoadduction. Paired comparisons confirmed that for both N = 4 and N = 7, the ratings for pressed voice differed significantly from those for all other voice types. Further, for N = 4 (but oddly, not for N = 7), normal and breathy ratings also differed, as did resonant and breathy ratings. Thus, focusing in particular on the distinction between pressed voice and other voice types, laryngeal adduction ratings provided critical distinctions between the voicing patterns.

Table 6.
Average ratings of laryngeal adduction for four voice types, based on videoscopic views (-5 = extreme hypoadduction, 0 = neutral, +5 = extreme hyperadduction).
Significant paired comparisons also indicated with asterisk ($p < .05$)

VOICE TYPE

VOWEL	Pressed	Normal	Resonant	Breathy
/i/ N=4	2.38	0.19	0.47	-1.66
N=7	2.55	-0.18	0.02	-1.21

SIGNIFICANT PAIRED COMPARISONS

VOWEL	Pressed/ Normal	Pressed/ Resonant	Pressed/ Breathy	Normal/ Resonant	Normal/ Breathy	Resonant/ Breathy
/i/ N=4	*	*	*		*	*
N=7	*	*	*			

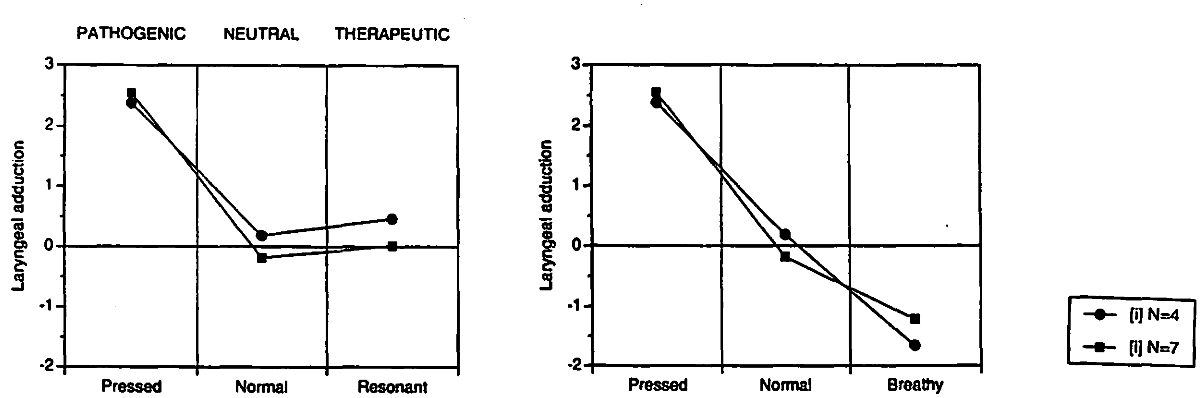


Figure 8. Average laryngeal adduction ratings (-5 = extreme hypoadduction, +5 = extreme hyperadduction), as a function of voice type, for the vowel [i] (N=4 and N=7).

To specifically assess the relation between the closed quotient and laryngeal adduction ratings, correlations were calculated between these measures for the four subjects for whom both EGG and videoscopic measures were available. First, correlations were calculated using average data for both the closed quotient and for laryngeal ratings. These results are shown in Figure 9. Comparing the average closed quotient for /i/ to the average adduction rating (also for /i/ which was the only vowel assessed endoscopically), increases in the closed quotient corresponded to increases in adduction ratings, Pearson $r = .98$, $p < .01$. A similar pattern was noted for the comparison between the closed quotient for /a/ and the adduction ratings for /i/, Pearson $r = .996$, $p < .01$. (Relations between the EGG closed quotient for /u/ and laryngeal adduction ratings were not assessed, because the results for /u/ closed quotients were unreliable.) Stated differently for the vowels assessed, there was a remarkable linear relation between the average closed quotient and average adduction rating, both decreasing across the continuum from the pathogenic to the therapeutic voice types. Correlations using individual data also indicated significant relations. Considering individual closed quotients for /i/ and comparing them to individual adduction ratings for /i/, Pearson $r = .59$ ($p < .02$). Comparing individual closed quotients for /a/ and comparing them to individual adduction ratings for /i/, Pearson $r = .70$ ($p < .01$).

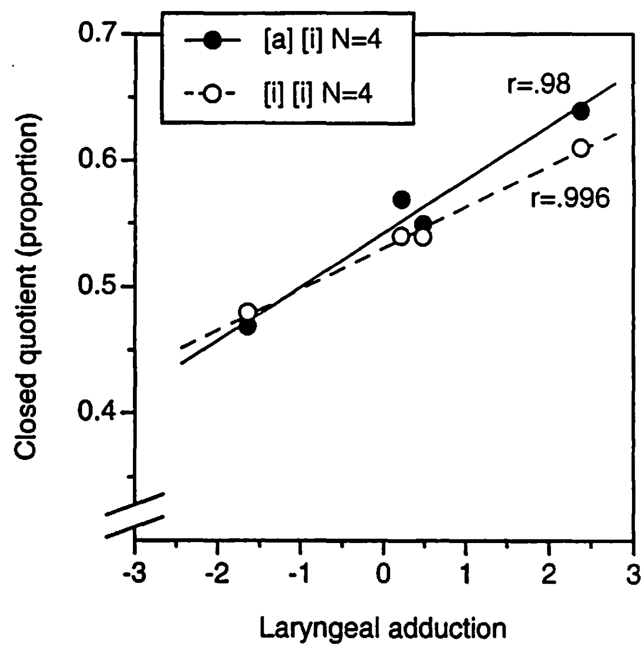


Figure 9. Average closed quotient in proportion of total cycle, for [i] and [u] (N=4) as a function of average laryngeal adduction ratings for [i] (N=4, -5 = extreme hypoadduction, +5 = extreme hyperadduction).

Discussion

Based on the results from the present study, for the vowels /a/ and /i/, one physiological measure was identified that may effectively distinguish voice types that appear relevant for nodule formation and regression: the EGG closed quotient. A series of aerodynamic measures, including MFDR, AC flow, and minimum flow, as well as another EGG measure, closing time, failed to acceptably differentiate between canonical tokens of pathogenic, neutral, and therapeutic voice types for any of the vowels, as did the EGG closed quotient for the vowel /u/.

In greater detail, average EGG closed quotients for /a/ and /i/ were reliably greatest for a voicing pattern that may be pathogenic, pressed voice. Closed quotients were intermediate for a neutral voicing pattern (normal voice) and for another voicing pattern that may be therapeutic (resonant voice, 11), and were the smallest for another possibly therapeutic voicing pattern (breathy voice, 11). High closed quotients were seen for pressed voice not only for the subject group as a whole, but also for most individual subjects.

Evidence from other authors suggests a link between the EGG closed quotient and laryngeal adduction (21). Confirmation of this relation was sought in the present study by comparing closed quotients for /i/, which was assessed videoscopically, to laryngeal adduction ratings for the same vowel, and also by comparing closed quotients for /a/ (which was not assessed videoscopically) to laryngeal adduction ratings based on the vowel /i/. The results of these comparisons agree with those reported by the previous authors, indicating linear relations between the closed quotient and laryngeal adduction ratings, both when group and individual data were considered. The link to nodule pathogenesis and regression was provided by Jiang (6), whose data suggest an increased likelihood of trauma with increasing adduction.

Thus, the main points that emerge from this study are (a) clinically important voicing patterns may be effectively distinguished by the closed quotient from the EGG signal, for sustained vowels /a/ and /i/ (but not /u/), and (b) a critical dimension in the pathogenic-therapeutic voice use continuum may be laryngeal adduction level.

At a general level, our results are consistent with a framework proposed by Hillman and colleagues (5). According to this framework, laryngeal adduction is a critical dimension for pathogenesis. However, our results do not agree with one of these authors' specific conclusions. In the study by Hillman and colleagues (5), subjects with nodules and polyps generally produced higher MFDRs than normals, whereas voice disordered subjects without lesions produced normal MFDRs. The conclusion was that MFDR may predict the potential for pathogenesis, by reflecting force of vocal fold collision. In our study, MFDRs for the pathogenic voicing pattern were not reliably different from those for one putative therapeutic pattern (resonant voice) for the vowel of particular interest (/a/), and for the other vowels (/i/ and /u/), MFDRs were reliably *greater* for the same therapeutic voicing pattern as compared to the pathogenic pattern. The difference in the results across studies is likely attributable to a difference in voice tasks, and also to subject differences. In the study by Hillman and colleagues, voice disordered subjects produced their own spontaneous (and possibly pathogenic) voicing pattern at soft, medium, and loud levels. Intentional variations in voice across a pathogenic-therapeutic continuum were not included. In our study, vocally healthy subjects produced not only spontaneous ("normal") patterns, but also patterns assumed to be pathogenic and therapeutic. In our study, the inclusion of "resonant voice" in particular was critical, because it produced results indicating that MFDR may not distinguish well between pathogenic and non-pathogenic voicing patterns.

Although our results for MFDR indicate this difference with respect to one previous study, we did replicate the report of a reliable covariance between MFDR and AC flow, reported by other authors. Specifically, as reported by Sundberg, Scherer, and Titze (20), our data suggest a significant inverse relation between MFDR and AC flow. However, as already noted, in our study neither of these measures provided much insight about mechanisms regulating nodule pathogenesis and regression. The data for AC flow are particularly interesting in this regard. Although most of the paired comparisons between the presumably pathogenic and therapeutic voice types indicated significant differences, the data were skewed in a direction that is difficult to interpret within the pathogenic-therapeutic framework. AC flows were *smallest* for the pathogenic voice, and were *greatest* for the therapeutic voice types. If the significance of AC flow is a direct reflection of vocal fold collisional force, this result is problematic to interpret. Why should AC flows be greatest for a therapeutic voice type, and least for a pathogenic voice type? We do not have any answer to this question and it is an intriguing one. However, our findings agree with a general approach to voice training and therapy suggested by Sundberg (22), which emphasizes "flow mode", or an overall high utilization of glottal flows during phonation. Our findings also indicate that *high* amplitude vocal fold vibrations may be associated with therapeutic voicing patterns, in some cases, as opposed to low amplitude vibrations, as often assumed.

Returning to the main finding from our study, and that is the sensitivity of the EGG closed quotient for distinguishing clinically relevant voicing patterns, it is important to point out the limitations of our results and the remaining questions. First, before accepting the EGG closed quotient as a useful physiological investigatory tool, the present results should be confirmed in follow-up studies using both normal and pathological subjects. EGG signals are sometimes poorly obtained. Thus, the resulting measures may be available only in a restricted proportion of cases. In fact, in our study, valid signals were obtained only for four of six subjects for whom EGG was attempted. Factors that may contribute to poor signals include thyroid cartilage angle, relative fat and soft tissue in the neck, electrode slippage, and the presence of laryngeal pathology, resulting in poor long-term signal conduction across the larynx. Particularly the possibility of poor EGG signals for pathological subjects indicates the need for follow-up studies with this population, for whom evaluations seem particularly relevant.

Second, from the present study it is not clear why the EGG closed quotient distinguished between the voicing patterns for /a/ and /i/, but not for /u/. The finding of vowel differences for these measures should be confirmed or disconfirmed with follow-up studies, and if confirmed, the finding should be explained. However, if the present results are replicated, it should be kept in mind that the closed quotient may be meaningful for physiological investigations only for a restricted set of vowels.

Third, although the present results are consistent with the hypothesis that laryngeal adduction level is an important parameter for nodule development and reversal, as suggested by previous authors (for example, Hillman et al., 5), based on the present study the reasoning is based on inference only. In our study, subjects produced voice types *presumed* to be pathogenic or therapeutic, but there was no independent verification of this assumption. Longitudinal causal studies are needed for confirmation or disconfirmation of the relation between pressed voice, normal voice, resonant voice, and breathy voice, and pathogenesis. The main contribution of the present study is that perhaps the identified EGG measure (closed quotient) could be used as an investigatory tool in future studies of this type.

In conclusion, the results from this study indicate that the closed quotient from EGG signals for the vowels /a/ and /i/ may be useful for future physiological investigations of clinically relevant voicing patterns. The results also indicate that this quotient varies with intralaryngeal adduction level and by extension, contact stress. However, follow-up studies are needed to confirm the present results over a larger subject sample size including normal and pathological subjects, and to investigate the possible relation between the closed quotient and intraglottal contact stress. Assuming that the present findings are replicated, also longitudinal studies are needed to determine if in fact nodule development and regression can be predicted by the closed quotient.

Acknowledgements

This work was supported in part by grant No. P60 DC 00976 from the National Institute on Deafness and Other Communication Disorders. The first and second authors contributed equally to the project. Ingo Titze is thanked for his comments on an earlier version of this paper, Jon Lemke and Kice Brown for assistance with statistical management, Mark Peters and Kay Klein for assistance with graphics, and Julie Lemke and Phyllis Palmer for technical support.

References

1. Epstein SS, Winston P, Friedmann I, Ormerod FC. The vocal cord polyp. J Laryngol Otol 1957; 71:673-688.

2. Gillman T. Treatise on collagen. In: GOULD Biology of Collagen. New York: Academic Press, Vol 2, 1968:334.

3. Sonninen A, Damste PH, Jol J, Fokkens J. On vocal strain. Folia Phoniat 1972; 24:321-336.

4. Kleinasasser O. Pathogenesis of vocal cord polyps. Ann Otol Rhinol Laryngol 1982; 91:378.

5. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughn C. Objective assessment of vocal fold hyperfunction: An experimental framework and initial results. J Speech Hear Res 1989; 32:373-392.

6. Jiang J. A methodological study of hemilaryngeal phonation and the measurement of vocal fold intraglottal pressure and impact stress. Unpublished doctoral dissertation; The University of Iowa, 1991.

7. Boone DR. The Voice and Voice Therapy. Englewood Cliffs, NJ: Prentice-Hall, 1977.

8. Colton RH, Casper JK. Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment. Baltimore: Williams & Wilkins, 1990.

9. Cooper M. Modern Techniques of Vocal Rehabilitation. Springfield, IL: Charles C. Thomas, 1973.

10. Verdolini-Marston K. Voice Therapy. Paper presented at Iowa Speech-Language-Hearing Association Convention, Des Moines, Iowa, October 1991.

11. Verdolini-Marston K, Burke MK, Lessac A, Glaze L, Caldwell E. Efficacy of Resonance and Confidential Voice Therapy for Treatment of Laryngeal Nodules and Polyps. Paper accepted for presentation at the 21st Annual Symposium: Care of the Professional Voice, Philadelphia, PA, June 1992.

12. Fog Pedersen M. Electroglottography compared with synchronized stroboscopy in normal persons. Folia Phoniat 1977; 29:191-199.

13. Childers DG, Krishnamurthy AK. A critical review of electroglottography. CRC Critical Reviews in Biomedical Engineering 1985; 12:131-161.

14. Childers DG, Hicks DM, Moore GP, Alsaka YA. A model for vocal fold vibratory motion, contact area, and the electroglottogram. J Acoust Soc Am 1986; 80:1309-1320.

15. Scherer RC, Druker DG, Titze IR. Electroglottography and direct measurement of vocal fold contact area. In: Vocal Physiology: Voice Production, Mechanisms and Functions. New York: Raven Press, LTD, 1988:279-291.

16. Scherer RC, Gould WJ, Titze IR, Meyers AD, Sataloff RT. Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. J Voice 1988; 2:230-244.

17. Rothenberg M, Mahshie JJ. Monitoring vocal fold abduction through vocal fold contact area. J Speech Hear Res 1988; 31:338-351.

18. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. J Acoust Soc Am 1988; 84:511-529.

19. Colton RH, Conture EG. Problems and pitfalls of electroglottography. J Voice 1990; 4:10-24.

20. Sundberg J, Scherer R, Titze IR. Phonatory control in male singing. A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. J Voice (in press).

21. Scherer R, Vail V. Measures of laryngeal adduction. J Acoust Soc Amer (suppl 1) 1988; 84:S81(A).

22. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. J Speech Hear Res 1989; 32:556-565.

NCVS Status and Progress Report - 4
June 1993, 113-119

Measuring the Impact of Prevention on Vocal Abuse: University Cheerleaders

Florence B. Blager, Ph.D.
Ronald C. Scherer, Ph.D.
Lawrence R. Brown, D.M.A.
The Recording and Research Center, The Denver Center for the Performing Arts
Jon N. Lemke, Ph.D.
Department of Preventive Medicine, The University of Iowa

Introduction

Cheerleading -- the "all-American" sport image -- has been described as extremely damaging to the voice (1,2,3,4,5). However, even among those who have described this voice use as damaging, attempts to establish the factors in cheerleading that impact destructively on the vocal mechanism have been inconclusive. Factors that have been discussed which differentiate cheerleaders with and without laryngeal pathology or dysphonia include marked pitch elevation (3), hard vocal attack (1), depressed pitch range (6), and constitutional vulnerability (7) within the pathology group. Methods of evaluating voice change have included voice intensity measures (7,10), respiratory volumes (9), and questionnaires (2,4). Training in appropriate breath support and voice production techniques has been suggested to prevent damage to the voice and larynx (1,5,8,10).

This study was an attempt to measure the acoustic characteristics of cheerleaders' voices throughout the cheerleading season, give the cheerleaders information about appropriate voice techniques and vocal hygiene, and evaluate the impact of the information on the quality of voice produced by the cheerleaders.

Methods

Sixteen cheerleaders, eight male and eight female, comprised the cheerleading squad at the University of Colorado at Boulder during the 1990-1991 sports season, September, 1990, through April, 1991.

History Questionnaires were filled out at the beginning of the study (8/15/90 and 8/27/90) and Follow-up Questionnaires at the end (5/5/91). Voices were recorded at four different times: Time 1, before the cheerleading season started (8/15/90); Time 2, after cheerleading camp and after the first football game (8/27/90); Time 3, after information on voice was presented and after cheerleading at football games (9/10/90); and Time 4, at the end of the cheerleading season following the last basketball game (5/5/91).

Information on voice was presented between Times 2 and 3 in two sessions: Session one (8/27/90) covered techniques to warm-up the voice and to relieve vocal tension. Session two (9/1/90) covered structure and function of the voice, prevention of abusive vocal habits, use of breath support, development of vocal resonance, and medical care of voice problems. Laryngoscopic examination of the larynx was performed at session two (9/1/90) using recorded stroboscopy.

Acoustic recordings at Times 1, 2 and 3 were made at The Recording and Research Center at The Denver Center for the Performing Arts. Subjects were recorded in an IAC booth using an AKG C451EB(CK22) microphone and a Digital Sound Corporation DSC240 amplifier. The acoustical signal was video encoded through a Sony Digital Audio Processor PCM-601ESD, and recorded on videotape. The cheerleaders were recorded at Time 4 at the University of Colorado Communications Disorders andSpeech Sciende department in Boulder in an IAC booth. Voices were recorded using a TEAC MDX-3R reel-to-reel tape recorder and a Sennheiser MD-44IU microphone. The subjects stood approximately one foot from the microphone.

For this study each person was instructed to produce 15 sustained /ɑ/ tokens in a normal, steady voice using a comfortable pitch and loudness.

Acoustic analyses of the sustained /ɑ/ vowels included the "short term" (cycle-to-cycle) perturbation measures jitter and shimmer, and the "long term" perturbation measures called the coefficient of variation for frequency (CVF) and the coefficient of variation for amplitude (CVA) (11). Also, the harmonics-to-noise ratio (HNR) (12), the harmonic spectral slope (HSS), the slope in dB per octave of the average calculated microphone spectrum, and the average fundamental frequency (Fo) were also calculated.

A repeated measures analysis of variance was performed on the acoustic analyses for the group of the five males who recorded at Times 2, 3 and 4. Only two females returned for the final recording, which gave insufficient data to control for gender differences.

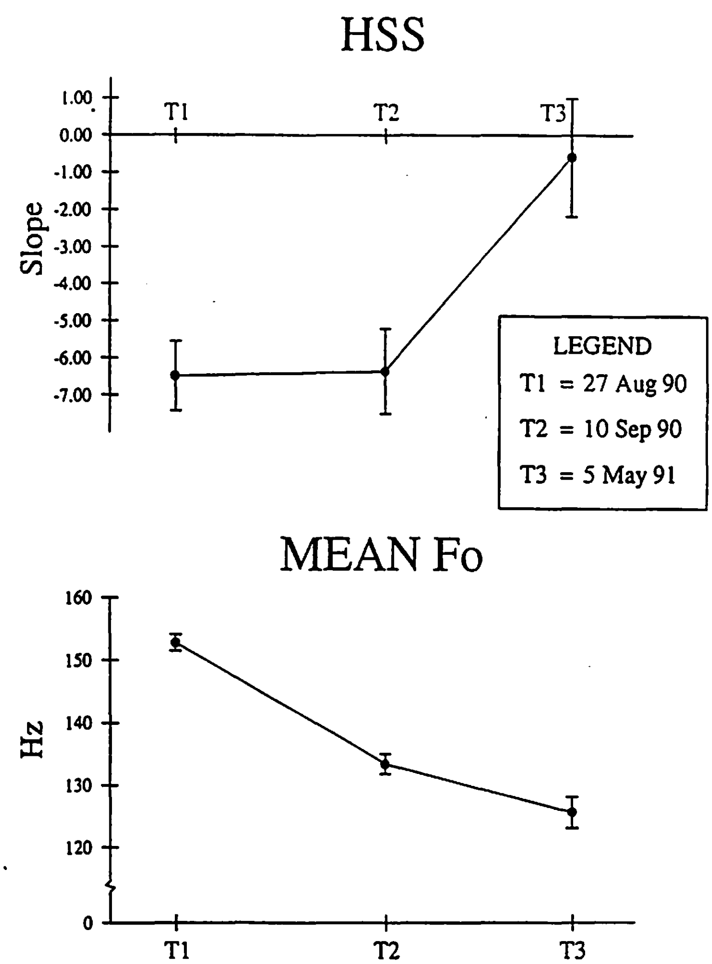


Figure 1. Harmonic spectral slope (HSS) and fundamental frequency (Fo) across three recording dates for 5 male cheerleaders. These measures changed statistically significantly (p<.05).

Results

Only two measures showed significant change over time: a decrease in slope of the HSS and a drop in the fundamental frequency for the five males (see Figure 1, previous page). There were no significant changes in the values of the voice perturbation measures.

The laryngoscopic examinations at session 2 (9/1/90) after the first cheering at the first football game did not show significant tissue abnormality, although the perceptual quality of some of the voices reflected hoarseness.

Information from the History and Follow-Up Questionnaires is shown in Charts 1, 2 and 3. Information on the History Questionnaire (Chart 1) indicates that the students had cheered with laryngeal problems and impaired voice quality in the past. The Follow-Up Questionnaire (Chart 2) reflects changes in voice use during cheerleading and fewer vocal problems experienced after cheerleading. The Follow-Up (Chart 3) suggests that cheerleading practice was less abusive than cheerleading during games.

Chart 1.

History Questionnaire Responses -- Before Cheerleading Season

Number	16 (8 male, 8 female)
Mean Age	20.0
Mean Number of Residences before college	4.56
Training in Body Work	yes - 12, no - 4
Types of Training	Males Athletics - 3 Martial Arts - 2 Females Dance - 5 Gymnastics - 6
Most Significant Training for Cheerleading	Males Athletics - 4 Females Gymnastics - 5
Currently Weightlifting	yes - 15
Learned Voice Warm-Ups in the Past	yes - 5
Present Health (August 15, 1990)	Excellent - 10 Good - 5 No answer - 1
Alcohol Use	yes - 13
Mean Age Started	Males - 17.8 Females - 16.5
Amount/week	Males - 10.5 "drinks" Females - 3.8 "drinks"
Had cheered at games in the past	
With laryngeal problems in the past	yes - 11
With impaired voice quality	yes - 9

Chart 2.
Follow-up Questionnaire Responses After Cheerleading Season

Number 9	(6 male, 3 female)
Mean Cheering/week	6 hours
What do you remember from the voice workshop presentations?*	Vocal Hygiene: Drink Water - 6 Use water for phlegm - 6 Don't clear throat - 4 Other hygiene points - 2 Voice Quality: Warm up voice - 7 Other voice exercises - 4
What information have you used?*	Drink water - 4 Various Voice exercises - 4
How much water do you drink?	Average of 6.27 glasses/day
Has voice changed since cheerleading season began?	Males no - 2 yes - 4 Females no - 3
If so, how would you describe the change?*	Improved - 3 Use stretches - 1

What changes have you made?*

Male	
Don't scream; yelling is now loud enough	1
Direct voice	1
Lower, more constant tone	1
Tone vibrates in nose and face	1
Use diaphragm	2
Female	
Deeper, more forceful sound	1
Sound with breath support	1

What has been the effect on your voice?*

Male	
Voice not as gone after game	2
Voice fatigued, but still have it	1
Don't get hoarse as much	1
Much improvement	1
Don't get sore throats like before	1
Female	
Stronger, louder	1
Sounds better	1
Lasts longer	1
Feels better	1
Voice rarely feels tired or strained	1

*Open-ended question

Chart 3.
Follow-up Questionnaire Responses. Sensations in Throat and Voice Quality.

Which of the following do you experience?
(1= Never, 2= Sometimes, 3= Frequently, 4= Always)

	Males (6)		*Females (3)*	
	End of Practice Day	End of Cheerleading Day	End of Practice Day	End of Cheerleading Day
a. Throat feels tired	1.16	1.66	1.33	1.66
b. Throat is painful	1.16	1.42	1.33	1.66
c. Phlegm in throat	1.33	1.83	2.33	2.66
d. Feel "something" in throat	1.2	1.2	2.0	2.0
e. Clear throat a lot	1.5	1.66	2.0	2.0
f. Voice is hoarse	1.5	2.42	1.33	1.33
g. Can't speak loudly	1.0	1.50	1.33	1.33
h. Lose voice	1.0	1.33	1.0	1.0
	1.22	1.73	1.58	1.71

Discussion

Cheerleaders are at great risk for voice damage because of the required use of loud voice during sports games and squad practices. It could be expected that the voice quality may deteriorate during the sports season. The voice perturbation analyses of the group were not statistically significantly different over time and did not support general voice quality deterioration. In addition, the laryngoscopic examination did not reveal laryngeal damage, contrary to the perceptually hoarse voice quality of some of the cheerleaders.

The significant findings were the decrease of both the fundamental frequency and the spectral slope. Fo decreased by about three semitones between August 27, 1990 and May 5, 1991, which may reflect greater comfort with the cheerleading task. The decrease in spectral slope may reflect use of slightly louder voices or more efficient vocal production.

The voice care sessions may have increased the attention of the cheerleaders to their voices and may have contributed to the prevention of voice deterioration as suggested by no statistical change in values of the perturbation measures. The results suggest that it may be inappropriate to assume that a few sessions during which the cheerleaders strongly attended to information on voice care and function would have the effect of <u>improving</u> the voice over the duration of their sports season. However, no significant change in voice perturbation values may be thought of as a positive finding.

Those cheerleaders who had previous voice "training," defined here as prior voice lessons or choir singing, appeared to have more stable voices (lower perturbation values) than those without training as shown by the "long term" perturbation measures HNR, CVA, and CVF in Figure 2 (following page). Except for one subject with no training, those with training had perturbation values lower than those without training. Future designs may wish to control for prior singing and speaking voice training and experience.

Information from the Follow-Up Questionnaires suggests that the subjects recalled and used information on both voice techniques and vocal hygiene. This is in contrast to the first author's follow-up study in a health care setting with high risk voice users who recalled and used information on vocal care and hygiene, but did not recall information on or use of exercises to increase voice resonance, breath support, or voice placement (13).

Future designs would be strengthened if a simultaneous control group were studied rather than a historical control group. The control group would be cheerleaders who receive no information on voice. Also, using identical questionnaires for pre- and post tests would define the parameters of voice change more specifically.

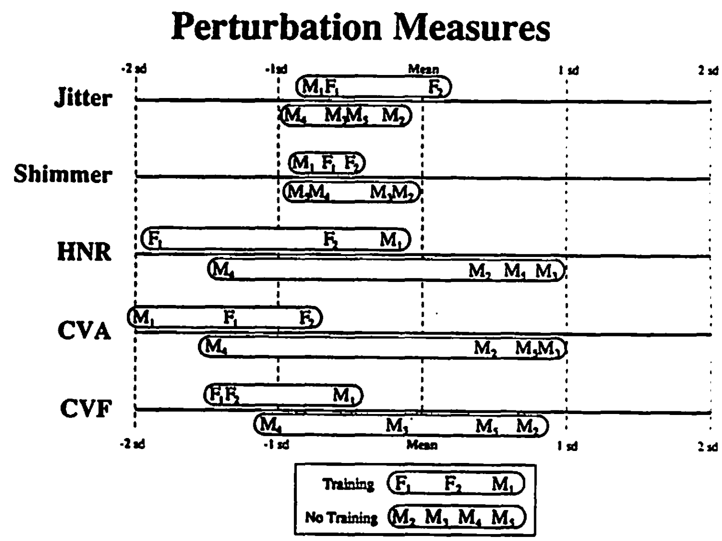


Figure 2. Voice analysis profiles for two groups of cheerleaders, those with a history of some voice training (three subjects), and those without (four subjects), according to their history questionnaires. The perturbation measures are given for the last recording time. Means and standard deviations refer to the normative data given in Reference 11. Male: M, Female: F.

Difficulty in consistent attendance at all recording sessions resulted in acquiring limited data. The related factors for this -- travel distance, scheduling, more encouragement, etc. -- should be controlled more rigorously in future studies.

It has been hypothesized that knowledge of voice use and care would help prevent vocal abuse in cheerleaders. The findings of the present study appear to support this. Such intervention may help prevent what may be a source of serious and at times long term damage to young voices.

Acknowledgments

The authors would like to thank Marilyn A. Hetzel, Ph. D. and Raymond P. Wood II, M.D., for their work in presenting information on voice, and Yoshiyuki Horii, Ph.D., for his help in recording the subjects during Time 4. Support from NIDCD grant P60 DC00976 is gratefully acknowledged.

References

1. Case JL: Clinical Management of Voice Disorders. Rockville, Maryland, Aspen Systems Corporation, 1984.

2. Gillespie S and Cooper B: Prevalence of speech problems in junior and senior high schools: Journal of Speech and Hearing Research 16:739-743, 1973.

3. Wilson KD: Voice Problems of Children. Third Edition. Baltimore, Williams & Wilkins, 1987.

4. Campbell S, Reich A, Klockars A, et al: Factors associated with dysphonia in high school cheerleaders. Journal of Speech and Hearing Disorders, 53: 175-185, 1988.

5. Andrews M and Shank KH: Some observations concerning the cheering behavior of school-girl cheerleaders: Language, Speech and Hearing Services in Schools 14: 150-156, 1983.

6. Sander E and Ripich D: Vocal fatigue. Annals of Otology, Rhinology and Laryngology 92: 141-145, 1983.

7. McHenry MA and Reich AR: Effective airway resistance and vocal sound pressure level in cheerleaders with a history of dysphonic episodes. Folia Phoniatrica 37:223-231, 1985.

8. Aaron VL and Madison CL: A vocal hygiene program for high-school cheerleaders. Language, Speech, and Hearing Services in Schools 22: 287-290, 1991.

9. Reich AR and McHenry MA: Respiratory volumes in cheerleaders with a history of dysphonic episodes. Folia Phoniatrica 39:71-77, 1987.

10. McHenry MA: Vocal SPL control in cheerleaders with a history of frequency, acute, cheering-related, dysphonic episodes. Doctoral Dissertation, University of Washington, 1983.

11. Scherer R., Gould WJ, Titze I, Meyers A and Sataloff R: Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. J. of Voice 2:230-244, 1988.

12. Yumoto E, Gould WJ and Baer T: Harmonics-to-noise ratio as an index of the degree of hoarseness. J. Acoust. Soc. Am. 71:1544-1550, 1982.

13. Blager FB: Follow-up study of high-risk voice users. American Speech-Language-Hearing Association, National Convention, Seattle, Washington, September, 1991.FIGURE CAPTIONS

NCVS Status and Progress Report - 4
June 1993, 121-134

Stroboscopic, Acoustic, Aerodynamic, and Perceptual Analysis of Voice Production in Normal Speaking Adults

Diane M. Bless, Ph.D.
Department of Communicative Disorders, The Waisman Center, and Department of Surgery, Division of Otolaryngology
University of Wisconsin-Madison
Leslie E. Glaze, Ph.D.
Midwest Voice and Speech Institute, St. Paul, Minnesota
Dawn Biever Lowery, Ph.D.
Ear, Nose, Throat and Head and Neck Surgeons, Inc., Columbus, Ohio
Guillermo Campos, M.D.
Facultad de Medicina, Hospital Universitario de San Ignacio, Bogota, Columbia
Robert C. Peppard, Ph.D.
Speech Language Pathology and Audiology, Northeastern University

Introduction

The purpose of this study was to examine the effect of chronological age on a range of vocal function measures in adult males and females with no communication deficits. In the last decade, technological advances in measurement of voice production have generated cautious enthusiasm from speech pathologists, speech scientists, and laryngologists who wish to use objective parameters to enhance assessment and to monitor treatment effects in patients with voice disorders. In an effort to determine the validity of these measures, numerous examiners have reported on both physiologic and perceptual measures of voice production. Physiologic measures include aerodynamic and acoustic parameters; perceptual measures consist of auditory and visual perceptual ratings of voice quality and the vocal fold vibratory pattern (McGlone and Hollien, 1963; Fitch and Holbrook, 1970; Ryan, 1972; Hollien and Shipp, 1972; Wilcox and Horii, 1980; Hammarberg, Fritzell, Gauffin, Sundberg, and Wedin, 1980; Honjo and Isshiki, 1980; Hirano, 1981; Kent and Burkard, 1981; Ramig and

Ringel, 1983, Morrison, Rammage, Belisle, Pullan, and Nichol, 1983; Linville and Fisher, 1985; Askenfelt and Hammarberg, 1986; Gelfer, 1988).

These studies have frequently been constrained by one or more factors, including limitation to a few examination parameters, or to a single sex or age level. Although some vocal function profiles have compared performance in young adults and geriatrics, they have largely ignored speakers in their middle years. Moreover, previous voice studies have not always considered demographic characteristics, such as health status, relevant laryngoscopic examination, and prior professional voice use. This omission may seriously limit the applicability of normative data. For example, the singer's literature has provided adequate evidence that differences exist between vocal function measures in trained and untrained speakers, which must be accounted for in any normal data base (Schutte and Miller, 1983; Peppard, Bless, and Milenkovic, 1988).

Additionally, the reported findings have not all been in agreement regarding age-related changes for such features as habitual frequency and intensity levels. These discrepancies may be explained in part by differences in such factors as the physical condition of the population sampled (Ramig and Ringel, 1983; Young, Bless, McNeil, and Braun, 1983), equipment differences (Titze and Scherer, 1985) and test protocol (Turner, 1988). This disagreement in the literature may indicate that existing data do not adequately reflect the underlying complexity of phonation, nor the variability in vocal function measures across normal subjects. Regardless of the source of the discrepancies, these limits make it difficult to apply the current "normative" values to the clinical population or to a better understanding of voice production.

Thus, basic questions remain about the vocal performance of normal speakers on a series of phonatory function tasks, including: what constitutes normal voice? and, how does performance vary with advancing age? We sought to address those issues by conducting a cross-sectional study of normal voice in adults. We collected stroboscopic, acoustic, aerodynamic, perceptual, laryngoscopic, and demographic data on males and females between the ages immediately post-puberty through senescence. These preliminary data begin to establish an age-referenced data base of normal phonatory function measures. Results from a standard vocal function battery are reported for measures of airflow rate, frequency, intensity, perturbation, audio perceptual judgments, and vibratory movement patterns for 146 "normal" speaking adults on sustained vowel production and connected speech tasks.

Methods

Subjects

Sixty-seven male and 79 female normal speakers ranging in age from 16 to 92 years were recruited from local bulletins. A "normal" subject was defined as a speaker with a negative history of speech, voice or hearing problems, no formal speaking or singing training, and no chronic health problems. Subjects who reported prior voice problems or injury to the throat or neck were excluded. Smokers were not excluded from the study (unless they exhibited voice problems) to achieve a group that was most representative of the general population.

Recording and Measurement

Five types of data were collected in this study: demographic, videostroboscopic laryngeal imaging, audio recordings, aerodynamic measurements, and indirect laryngeal examination. Data collection procedures were described in detail elsewhere (Peppard, Bless, and Milenkovic, 1988).

Case history information was obtained using a standard battery of questions about demographic characteristics including age, sex, height, weight, social and medical history, voice usage, and specific details about smoking, alcohol and caffeine consumption. A laryngoscopic examination was conducted by a licensed otolaryngologist specializing in vocal pathology, who reported the general appearance of the vocal folds and laryngeal structures, and noted any abnormality, including inflammation, supraglottic hyperfunction, erythema, or excessive mucous.

Stroboscopic recordings were made using a Bruel and Kjaer Stroboscopic Light Generator coupled to a Wolf 90 degree angle rigid endoscope and camera and a Hitachi 1/2" videorecorder. Subjects produced the /i/ vowel at normal, high and low pitch, and at normal, loud and soft levels.

Acoustic recordings were conducted with subjects seated in a sound proof booth, using a Nagra LVII reel-to-reel tape recorder (Model # 1716903) and a cardioid microphone (Model #666), with a constant mouth to microphone distance of 12 inches. The voice sample included sustained vowels /o/ and /i/ at comfortable pitch and loudness, maximum and minimum pitch and loudness levels, conversational speech, description of the standard "Cookie Theft" picture (Goodglass and Kaplan, 1972) and repetitions of the sentence "The blue spot is on the key again."

Aerodynamic vocal function measures were recorded using a facemask attached to the Nagashima PS-77 Phonatory Function Analyzer, which measures simultaneous fundamental frequency, intensity, and airflow. Habitual, minimal, and maximal frequency and minimal intensity were measured on sustained /o/ vowels. Maximal intensity was measured on a shouted phrase "Hey you!" Vocal fold diadochokinesis tasks were conducted using rapid repetitions of the vowel /i/ and the consonant-vowel syllable /hi/.

Data Analysis

Following the recording, three raters with extensive experience in videostroboscopic analysis viewed subject tapes and rated glottic closure pattern, shape of epiglottis, presence of mucous, supraglottic activity and vibratory features including amplitude, mucosal wave, phase symmetry and closure, and regularity on a numerical scale (Bartell and Bless, 1987).

For acoustic analysis, tape recordings were digitized from a Sony reel-to-reel playback (TC-654-4) using CSpeech (Milenkovic, 1987) a locally designed signal processing technique for an IBM PC-AT. Acoustic analysis of sustained /o/ and /i/ vowels were conducted using one second length midportions of the sustained vowels. Means, ranges, and standard deviations were derived for frequency, jitter, shimmer, and signal-to-noise ratio. Also, the difference between the highest and lowest frequency was calculated and converted to a semitone range.

Audio recordings of the Cookie Theft picture description and sentence "The blue spot is on the key again" were used for perceptual judgments. Tapes were played to two voice pathologists who had extensive experience in assessment and treatment of voice disorders. All samples were rated on a seven-point scale for sixteen perceptual bipolar attributes based on a perceptual battery adapted from Gelfer (1988) to measure the normal speaking voice.

Subject ages and individual means for acoustic and aerodynamic measures and perceptual and stroboscopic ratings were entered into a statistical program (Minitab 7.0). Simple regression analyses with analysis of variance were conducted to assess the direction and significance of relationships between each of the individual data parameters and age. Male and female data were analyzed separately, as there was no expectation that both sexes would perform similarly on these vocal function tasks. The p value was determined to be significant at a level of .05 or less. Additionally, r^2 values were identified to estimate the amount of variance in each dependent measure accounted for by the independent variable, age (Kleinbaum, Kupper, and Muller, 1988).

Results

Results for the male and female groups are presented for each data measure in the following Tables I through IV. Table V (see page 129) presents a listing of all measures which yielded significant regression with age ($p < .05$), followed by the r^2 value, signifying the amount of variance in the data accounted for by the dependent variable, age. In all analysis categories, female data demonstrated a greater number of significant age-related effects than males.

Table I presents the mean, standard deviation, and absolute range of vibratory measures taken from the videostroboscopic analysis, measured on a six point scale ranging from 0 (normal) to 5 (most abnormal). Graphs A and B present a summary of laryngoscopic findings for glottic closure and shape of epiglottis. Glottal closure patterns revealed a higher prevalence of a posterior glottal gap in this sample of females (78%) compared to males (27%). Other closure patterns, such as anterior and spindle-shaped gaps, appeared only in older subjects in females, but across all ages in males. For example, several spindle-shaped gaps were identified in young males. Sixty-four percent of females had a flat epiglottis, with the remaining 36% rated as crescent-shaped. In males, these ratings were divided among three shapes: flat, 35%, crescent, 52%, and omega-shaped, 13%.

Table I.
Stroboscopic Vibratory Features
(6 point scale)*

	MALES (N=64)				FEMALES (N=73)			
	mean	sd	min	max	mean	sd	min	max
Supraglottic activity	.47	.66	0	3	.39	.66	0	3
Vocal fold edge	.47	.21	0	1	.19	.39	0	1
Amplitude	.36	.55	0	2	.28	.59	0	2
Mucosal wave	.31	.47	0	1	.22	.48	0	2
Nonvibrating portion	0	0	0	0	.01	.11	0	1
Phase closure	.68	1	-1	3	-.05	.99	-3	3
Phase symmetry	1.9	1.7	0	5	1.9	2.0	-1	5
Periodicity	1.7	1.6	0	5	1.5	1.8	0	5

* Ratings made from 0 (normal appearance and movement) to 5 (most abnormal)

Although 97% of the females and 89% of the males were rated as having normal vocal fold appearance on stroboscopic examinations, laryngoscopic examinations revealed that many subjects had deviations from what is typically called "normal." Subjects exhibited slight edema, hyperemia, subclinical involvement of one vocal fold, and incomplete closure of the glottis during voice production. Excessive or remarkable mucous was noted in 60% of females, and 78% of males in both strobo- and laryngoscopic examinations.

Stroboscopic ratings of vocal fold movement patterns demonstrated the greatest concordance between males and females, and the highest r^2 values of all data measures. This finding suggests that variability in stroboscopic visual-perceptual measures was more strongly attributed to changes in chronological age than were other physiologic or perceptual voice measures. Both sexes displayed significant age-related changes in supraglottic activity, vibratory amplitude, mucosal wave, and phase symmetry. Females also demonstrated significant change with age in vibratory periodicity.

Tables IIa, IIb and III present the mean, standard deviation, and absolute and interquartile ranges of the acoustic and aerodynamic parameters for males and females. Wide variability in task performance is evidenced by the large standard deviations and absolute ranges. Therefore, interquartile ranges (values which fell at the 25th and 75th percentiles of the full distribution) were included to allow three useful interpretations. First, the symmetry and the spread of values can be examined for each measure. Second, potential outliers may be identified outside the 25 to 75% range. Third, by excluding extreme values, we can appreciate the majority "cluster" of values established by this normal subject sample for each of these phonatory function measures (Rosner, 1982).

Table IIa.
Male Acoustic Measures
(N=58)

	mean	sd	min	max	I*	III*
Vowel /ɑ/						
Fo (Hz)	106	18	74	171	94	117
Jitter (msec)	.07	.07	.01	.5	.03	.06
Shimmer (%)	4.9	5.5	1.4	41	2.7	5.5
SNR (dB)	17.2	3.8	7.2	27	14.8	19.7
min track	.45	1.6	0	11	0	0
max track	7.8	26	0	166	0	3
min err	.05	.39	0	3	0	0
max err	.8	2.9	0	19	0	0
Vowel /i/						
Fo (Hz)	112	20	74	185	98	121
Jitter (msec)	.05	.05	.01	.33	.03	.05
Shimmer (%)	3.26	2.6	.8	15.5	1.4	3.8
SNR (dB)	19.5	3.7	10.7	27.6	17.2	22.1
Min track	.3	1.2	0	7	0	0
Max track	5.15	23	0	170	0	2
Min err	0	0	0	0	0	0
Max err	2	9	0	67	0	1
High fo (Hz)	482	244	33	1190	278	640
Low fo (Hz)	77	33	25	225	55	92
Semitone range	31	12.6	7	66	22	40
Intensity (dB)†	70.5	4	62.5	85.5	68	73
High Io (dB)	98	3.6	86	110	96	101
Low Io (dB)†	62	4.3	60	80	60	65

*Values which fell at the first and third interquartile ranges.
†Intensity measures (dB SPL) made on Nagashima PS-77 Phonatory Function Analyzer are restricted to a lower limit of 60 dB. Therefore, no statistical results are influenced by machine limits.

Table IIb.
Female Acoustic Measures
(N=61)

	mean	sd	min	max	I*	III*
Vowel /ɑ/						
Fo (Hz)	193	27	117	262	178	211
Jitter (msec)	.06	.14	.01	.98	.01	.03
Shimmer (%)	4.7	6.3	1.1	37	2.2	4.1
SNR (dB)	18.3	4.05	6.6	25	16	21
min track	.5	1.8	0	12	0	0.
max track	14	52	0	306	0	1
min err	.03	.25	0	2	0	0
max err	1.43	4.4	0	19	0	0
Vowel /i/						
Fo (Hz)	202	28	132	284	183	223
Jitter (msec)	.03	.06	.008	.4	.01	.02
Shimmer (%)	2.5	2.3	.8	14	1.3	2.8
SNR (dB)	22	3.6	10.6	31	20.9	24.5
Min track	.39	1.5	0	10	0	0
Max track	8.81	31	0	163	0	1
Min err	.01	.13	0	1	0	0
Max err	4	14	0	73	0	0
High fo (Hz)	634	228	185	1190	439	758
Low fo (Hz)	137	45	34	226	112	163
Semitone range	26	10	7.5	51	17	34
Intensity (dB)†	68	4	60	82	66.5	72
High Io (dB)	94	6	80	106	90	98
Low Io (dB)†	61	1.7	60	70	60	62

*Values which fell at the first and third interquartile ranges.

†Intensity measures (dB SPL) made on Nagashima PS-77 Phonatory Function Analyzer are restricted to a lower limit of 60 dB. Therefore, Io statistical results are influenced by machine limits.

Acoustic measures for females revealed significant differences in frequency range, mean and range of intensity, and mean jitter, shimmer, track and err. Male effects were limited to frequency and semitone ranges and mean intensity, with no significant changes with age in perturbation measures. In aerodynamic measures, vocal fold diadochokinesis and airflow rate changed significantly with age in females, while male data revealed no significant effects.

Table III.
Aerodynamic Measures

	mean	sd	min	max	I*	III*
MALES (N=67)						
Flow rate /ɑ/†	119	62	30	320	80	150
Flow rate /i/†	140	89	20	540	90	171
DDK /i/~	5	0.9	3	8	4.6	5.6
DDK /hi/~	5	1.3	2	9	4	6
FEMALES (N=78)						
Flow rate /ɑ/†	115	48	10	220	80	155
Flow rate /i/†	113	51	8	250	80	140
DDK /i/~	5	1.2	2.5	9	4.3	5.7
DDK /hi/~	4.9	1.5	1.8	9	4	6

*Values which fell at the first and third interquartile ranges.
† Flow rates are measured in cc/second.
~Diadochokinesis rates measured on rapid repetitions of syllables /i/ and /hi/.

Table IV presents the mean, standard deviation, and absolute ranges for bipolar perceptual ratings from a seven-point scale of voice quality in males and females. Perceptual judgments of normal voice consistently rated older voices less favorably than younger samples. Both sexes were rated more negatively with age on five features: young/old, smooth/rough, clear/hoarse, steady/shaky, and like/don't like. Overall, female data revealed ten features that regressed significantly with age; male data yielded six significant effects.

Demographic data showed the subjects to be overall in good health, non-smokers, non- or light drinkers, having one to two colds per year, not working in vocally demanding environments, and consuming approximately 400 mg of caffeine daily. These and other demographic categories were examined to identify subjects who might be at risk for developing vocal pathology, including allergies, arthritis or joint problems, psychological dysfunction, and neurological problems. Surprisingly, there was no correspondence between subjects who were targeted as potentially at risk (e.g., reported alcohol, tobacco, or drug use, heavy vocal use, or psychological problems) and statistical outliers for any of the vocal parameters. In all cases, subjects who were identified at risk had vocal function measures that fell within one standard deviation of the mean, and demographic information for statistical outliers revealed no vocally compromising behaviors.

Table IV.
Auditory-Perceptual Measures of Voice Quality
(7 point scale)*

	MALES (N=62)				FEMALES (N=79)			
	mean	sd	min	max	mean	sd	min	max
low/high pitch	1.9	.6	1	3	3.7	.9	2	6
soft/loud	4.3	.8	2	6	4.1	.9	2	6
strong/weak	4.9	.9	3	7	4.2	.9	2	6
smooth/rough	3.7	1.3	1	6	4	.9	2	6
pleasant/ unpleasant	3.5	1	2	6	4.2	.9	2	6
resonant/shrill	2.8	.8	1	5	3.7	.8	2	6
clear/hoarse	3.4	1	1	5	3.8	1	2	6
unforced/ strained	3.7	1	1	6	4	.9	1	5
soothing/harsh	3.7	.9	2	6	3.9	.9	1	6
melodious/raspy	3.5	1	2	6	4.1	.9	2	6
full voice/ breathy	3.1	1.1	1	5	4.1	.8	2	6
nasal/denasal	4	.2	3	4	3.8	.5	2	4
animated/ monotonous	3.8	.9	2	6	3.8	1	2	6
steady/shakey	3.1	.8	2	5	3.5	1.1	2	6
young/old	3	1.3	2	6	4	1.6	2	7
like/don't like	4.2	1.4	2	7	3.6	1.1	2	7

*Ratings made on bipolar scales from 1 (e.g., lowest pitch) to 7 (highest pitch).

Discussion

The data reported here raise questions about what constitutes a "normal" voice. All subjects included in this study had perceptually normal voices, by their own admission and judgement of the examining clinicians. Nevertheless, findings that might be considered "abnormal" were present among all data sectors. Laryngoscopic examinations revealed a large percentage of subjects to have slight edema or local inflammation, in the presence of normal acoustic, aerodynamic, perceptual, and stroboscopic results. Stroboscopic observations indicated lack of complete glottal closure, reduced vibratory amplitude and mucosal wave, and aperiodicity. In acoustic and aerodynamic measures of sustained vowels, variability was large, with perceptually normal speakers measuring as much as two standard deviations from the mean. Yet consistently, audio perceptual measures of subjects' sentence productions were rated as normal.

Table V.
Significant Regression with Age
($p<.05$)

	MALES	(r^2)	FEMALES	(r^2)
Stroboscopy				
	supraglottic	(11)	supraglottic	(30)
	amplitude	(7)	amplitude	(35)
	wave	(11)	wave	(52)
	phase symmetry	(20)	phase symmetry	(20)
			regularity	(55)
Acoustics				
	high Fo	(7)	high Fo	(7)
	/ɑ/ Io	(8)	/ɑ/ Io	(7)
	/i/ Io	(14)	/i/ Io	(16)
	semitone range	(12)		
			high Io	(15)
			low Io	(33)
			/ɑ/ jitter	(11)
			/ɑ/ shimmer	(8)
			/ɑ/ track	(16)
			/ɑ/ err	(7)
			/i/ track	(12)
			/i/ err	(7)
Aerodynamics			/ɑ/ flow	(7)
			DDK /i/	(16)
Perceptual				
	smooth/rough	(7)	smooth/rough	(12)
	clear/hoarse	(8)	clear/hoarse	(12)
	steady/shaky	(11)	steady/shaky	(32)
	young/old	(83)	young/old	(83)
	like/don't like	(12)	like/don't like	(5)
	melodious/raspy	(12)		
			strong/weak	(5)
			soothing/harsh	(9)
			nasal/denasal	(6)
			full voice/ breathy	(6)
			pleasant/unpleasant	(15)

Perceptual judgments of the vocal folds under stroboscopic and indirect laryngoscopic examinations differed in their assessment of "normal" appearance. Several possibilities may explain this discrepancy. First, stroboscopic recordings are made under different lighting conditions, and preclude assessment of color changes, such as erythema, in the laryngeal structures. Second, the overriding majority of "abnormal" laryngoscopic findings in this subject sample were described as minimal; some difference was noted, but the finding was insufficient to warrant a clinical diagnosis. Third, the intent of stroboscopic assessment is to judge movement patterns in the laryngeal structures and vocal fold vibration. It is possible that the slight deviations noted by the otolaryngologist during indirect laryngoscopic examinations are compensated during vibration, and not visible. Laryngoscopic and stroboscopic examinations provide separate information contributing to the total assessment of laryngeal function. Similarly, otolaryngologists and voice pathologists represent complementary, but not identical disciplines, which might account for some differing evaluation criteria. All of these factors underscore the need for a team approach to achieve competent evaluation of the larynx.

A great variation in degree of vocal fold closure was evident across subjects, with different types of incomplete glottal closure observed among females and males. These findings seem to have important implications for laryngeal modeling research, and may explain why closed portions of EGG waveforms are sometimes difficult to define, especially in females (Childers, Smith, and Moore, 1984). Formerly, glottal gap has been attributed to vocal weakness or pathology (Morrison, M, Rammage, L, Belisle, G, Pullan, B, and Nichol, H, 1983). Results from this normal speaker sample support a contrary finding that some degree of incomplete glottic closure may be considered typical of adult female closure patterns (Biever and Bless, 1989), and a variant in males (Bless, Biever, and Shaikh, 1986). Apparently, glottal gaps do exist in young adults and geriatric speakers who have no voice complaints and otherwise normal vocal function measures. Therefore, we suggest that without concomitant aberrant measures of vocal function, glottal gaps be termed "incomplete approximations," and that the term "glottal insufficiency" be reserved for those gaps associated with voice pathology, including excessive airflow, perceptual breathiness, or other abnormal vocal function measures.

To further explore glottal closure patterns, a preliminary subgroup of data were selected for more detailed analysis. For each sample, a frame depicting maximum closure in the vibratory cycle was selected from the videostroboscopic recording for digitization and measurement using a VAX workstation coupled with a Data Translation-IRIS board. A locally designed software program (Ulmer, University of Wisconsin Vocal Function Laboratory) allowed the investigators to measure visual landmarks on the glottal image, and determine a ratio between the length of the posterior glottal gap relative to the entire glottal length for each subject. These findings will be presented fully in a separate report.

None of the subjects with aberrant glottal gap or extreme acoustic or aerodynamic measures would be targeted as likely to have voice problems based on case history or laryngoscopic information. Thus, we begin to acknowledge the range of anatomical and voice performance levels within the normal speaking population that are not necessarily founded in vocal abuse or environmental factors. Indeed, other unmeasured factors, such as learning, motor control, physical and mental health, anatomy, and behavioral patterns may contribute to the variability of "normal" vocal function (Aronson, 1976).

Statistically significant age-related changes were revealed for some measures, but all parameters except stroboscopy were marked by large random variability not accounted for by age. Other investigators have shown previously that vocal function measures may exhibit large variability in the aging population (Ramig, Scherer, and Titze, 1985) and that some vocal measures are more sensitive to age than others (Wilcox and Horii, 1980; Ramig and Ringel, 1983; Biever and Bless, 1989; Brown, Morris, and Michel, 1989). The explanations for differences in age-sensitivity are unknown.

What is certain is that theories of aging seeking to explain the physiologic process of senescence allow for multiple influences, including anatomic, cardiovascular, pulmonary, hormonal, immunologic, musculoskeletal, neurologic, and psychogenic changes (Chodzko-Zajko and Ringel, 1987). While specific effects of these aging changes on voice has not been determined, evidence in voice literature indicates that each of these factors may play a role in vocal function performance (Aronson, 1976; Hirano, 1983; Kahane, 1983; Linville and Fisher, 1985; McGeer and McGeer, 1986; Hixon, 1987; Abitbol, et al, 1989; Griffiths and Bough, 1989; Higgins and Saxman, 1989; Orlikoff, 1989;). Thus, future studies must recognize this complex relationship between chronologic age and physiologic aging patterns to help discern the source of variability in vocal function measures.

Although no statistical comparisons were conducted to assess sex-based differences, data reported here are consistent with the anatomical and physiological claims that there are male-female differences in vocal performance (Hirano, 1981). Interestingly, anatomical and histological evidence generally presents greater age-based changes in males than females (Kahane, 1983). Conversely, our vocal function data display more significant effects in female subjects. The reason for this discrepancy is unclear, and will be explored further in an expanded report of these data.

Limitations of the Present Study

Our subject sample had certain communal characteristics which may not provide wholly generalizable results across all recording locales. All speakers were native Midwesterners using a General American English dialect (Tiffany and Carroll, 1972). Over 95% percent of all speakers were Caucasian. All subjects were of sufficient health, mobility, and economic means to come to the recording site independently. None suffered from chronic health problems, and denied a presence or history of voice problems. This general subject status cannot be wholly representative of all "normal" speakers (particularly in geriatric populations), due to inherent sampling bias in the exclusion criteria. Thus, we caution interpretation of these data be restricted to healthy Caucasian speakers.

In recent exploration of age-related changes in physical performance, much has been reported about the utility of referencing a "physiologic age" as well as a chronological age. This study did not attempt to make any external estimate of physiologic age (e.g., blood measures or pulmonary function), but rather, relied entirely on chronological age. In the absence of any large age-referenced data base for normal vocal function, it was unknown whether chronological age may reflect phonatory changes adequately across age levels. Studies have shown significant differences in vocal function between subjects in good and poor physical condition. However, contrasts of different "physiological" ages required sampling of subjects in deteriorating health for comparison with robust subjects (Ramig and Ringel, 1983), and therefore may demonstrate another type of sampling bias. In this study, we limited our definition of "normal speakers" to persons who did not have chronic medical problems, even in geriatric years.

These data suggest that a chronological age reference in combination with comprehensive and well-elicited case history information is clinically useful independent of a physiologic measure. Nonetheless, large variability in measures not explained by age is evident in these data. Further investigation may determine whether reports on the variable rate and extent of physiologic aging can explain the differences within older subjects not accounted for by chronological age alone. A possible concordance between resilience in motor performance in healthy, active older adults (Fries and Crapo, 1981; Spirduso, 1982) and improved vocal performance in these geriatric speakers must still be determined.

Conclusions

The purpose of this article is to present results of vocal function measures collected on normal speakers from young adult through geriatric years. Sources and measures of age-related changes in phonatory function are complex and cannot be explored fully in this report. Likewise, differences in male-female vocal performance have not been addressed here. Both topics are worthy of further discussion and will be presented separately in other papers.

As future studies continue the process of determining age effects in vocal function measures for normal speakers, it will be important to reconsider the possible influence of the select aging variables, including chronologic vs. physiologic age, gender, and demographic factors. The need remains for longitudinal studies and careful replication using similar elicitation methods, voice sample, and recording and analysis equipment. This study sought to begin the process of establishing a comprehensive normative sample of adult speakers. The findings presented here seem to underscore two overriding observations: the marvelous ability of the human vocal mechanism to compensate across a wide range of age and speaker history, and the need to view normal laryngeal structure and vocal function on a variable continuum.

Acknowledgements

This research was supported in part by Grant No. P60 DC00976-03 from the National Institutes on Deafness and Other Communication Disorders and NINCDS Grant RO1-NS24859-01. The authors wish to thank the following individuals for their contributions to the completion of this project: Linda Rammage, Danna Koschkee, Mark Leddy, Ebonie Howard, Colleen Quinn, Christy Bemis, and Yvonne Annis.

References

Abitbol J, de Brux J, Millot G, Masson MF, Mimoun OL, Pau H, Abitbol B. Does a hormonal vocal cord cycle exist in women? Study of vocal premenstrual syndrome in voice performers by videostroboscopy-glottography and cytology on 38 women. J VOICE 1989; 3:157-162.

Aronson AE. CLINICAL VOICE DISORDERS. New York: Thieme, Inc., 1985.

Askenfelt A, Hammarberg B. Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. J SPEECH HEAR RES 1986; 29:50-64.

Bartell T, Bless DM, and Ford CN. Stroboscopic examination of voice: a comparison of two rating procedures. Proceedings of International Conference on Voice. Kurume, Japan, 1986; 2:35.

Biever DM and Bless DM. Posterior glottal gap in normal young adult and geriatric women. In review, J VOICE.

Biever DM and Bless DM. Vibratory characteristics of the vocal folds in young adult and geriatric women. J VOICE 1989; 3:120-131.

Bless DM, Biever DM, Shaikh A. Comparison of vibratory characteristics of young adult males and females. In: Proceedings of International Conference on Voice. Kurume, Japan, 1986; 2:46-54.

Brown WS, Morris RJ, and Michel JF. Vocal jitter in young adult and aged female voices. J VOICE 1989; 3:113-119.

Childers DG, Smith AM, and Moore GP. Relationships between electroglottography, speech, and vocal cord contact. FOLIA PHON 1984; 36:105-118.

Chodzko-Zajko WJ and Ringel RR. Physiological aspects of aging. J VOICE 1987; 1:18-26.

Fitch JL and Holbrook A. Modal vocal fundamental frequency of young adults. ARCH OTOLARYNGOL 1970; 92:379-82.

Fries JF and Crapo LM. VITALITY AND AGING. San Francisco: Freeman Press, 1981.

Gelfer MP. Perceptual attributes of voice: development and use of rating scales. J VOICE 1988; 2:320-326.

Glaze LE, Bless DM, Milenkovic P, and Susser RD. Acoustic characteristics of children's voice. J VOICE 1988; 2:312-319.

Goodglass H and Kaplan E. THE ASSESSMENT OF APHASIA AND RELATED DISORDERS. Philadelphia: Lea and Febinger, 1972.

Griffiths C and Bough ID. Neurologic diseases and their effect on voice. J VOICE 1989; 3:148-156.

Hammarberg B, Fritzell B, Gauffin J, Sundberg J, and Wedin L. Perceptual and acoustic correlates of abnormal voice qualities. ACTA OTOLARYNGOL 1980; 90:141-145.

Higgins MB and Saxman JH. Variations in vocal fundamental frequency across the menstrual cycle. J VOICE 1989; 3:233-243.

Hirano M. CLINICAL EVALUATION OF VOICE. New York: Springer-Verlag, 1981.

Hirano M, Kurita S, and Nakashima T. Growth, development, and aging of the vocal folds. In: VOCAL PHYSIOLOGY, Bless DM and Abbs JA, eds. San Diego: College-Hill Press, 1983.

Hixon T. RESPIRATORY FUNCTION IN SPEECH AND SONG. Boston: College-Hill Press, 1987.

Hollien H and Shipp T. Speaking fundamental frequency and chronologic age in males. J SPEECH HEAR RES 1972; 15:155-159.

Honjo I and Isshiki N. Laryngoscopic and voice characteristics of aged persons. ARCH OTOLARYNGOL 1980; 106:149-150.

Kahane JC. A survey of age-related changes in the connective tissues of adult human larynx. In: Bless DM and Abbs JH, eds. VOCAL PHYSIOLOGY. San Diego: College Hill Press, 1983.

Kent RD and Burkard R. Changes in the acoustic correlates of speech production. In: Beasley DS Davis GA, eds. AGING: COMMUNICATION PROCESSES AND DISORDERS. New York: Grune & Stratton, 1981.

Kleinbaum DG, Kupper LL, Muller KE. APPLIED REGRESSION AND OTHER MULTIVARIATE METHODS. Boston: PWS-Kent Publishing Company, 1988.

Linville S and Fisher H. Acoustic characteristics of perceived versus actual vocal age in controlling phonation by adult females. J ACOUS SOC AMER 1985; 78:40-48.

McGeer PL and McGeer EG. Aging and neurotransmitter systems. In Finch CE, Potter DE Kenny AD, eds. PARKINSON'S DISEASE, vol 2, AGING AND ENDOCRINE RELATIONSHIPS. New York: Plenum, 1986: 41-58.

McGlone R and Hollien H. Vocal pitch characteristics of aged women. J SPEECH HEAR RES 1963; 6:164-170.

Milenkovic P. Least mean square measures of voice perturbation. J SPEECH HEAR RES 1987; 30(4):529-38.

Morrison M, Rammage L, Belisle G, Pullan B, and Nichol H. Muscular tension dysphonia. J OTOLARYNGOL 1983; 12:302-306.

Orlikoff RF. Vocal jitter at different fundamental frequencies: a cardiovascular-neuromuscular explanation. J VOICE 1989; 3:104-112.

Peppard RC, Bless DM, and Milenkovic P. Comparison of young adult singers and nonsingers with vocal nodules. J VOICE 1988; 2:250-260.

Ramig L and Ringel R. Effects of physiological aging on selected acoustic characteristics of voice. J SPEECH HEAR RES 1983; 26:22-30.

Ramig LA, Scherer RC, Titze IR. Acoustic correlates of aging. Research Report 1, Denver Center for Performing Arts, 1985.

Rosner B. FUNDAMENTALS OF BIOSTATISTICS. Boston: Duxbury Press, 1986.

Ryan W. Acoustic aspects of the aging voice. J GERONTOL 1972; 27:265-268.

Schutte H and Miller R. Differences in spectral analysis of a trained and untrained singer. NATS BULLETIN, Nov/Dec 1983; 22-26.

Spirduso WW. Physical fitness in relation to motor aging. In: Mortimer JA, Pirozzolo FJ, Maletta GJ, eds. THE AGING MOTOR SYSTEM. New York: Praeger Publishers, 1982.

Tiffany WR and Carrell J. PHONETICS: THEORY AND APPLICATION. New York: McGraw Hill, Inc., 1977.

Titze IR, Horii Y, Scherer RC. Some technical considerations in voice perturbation measurements. Research Report 1, Denver Center for the Performing Arts, 1985.

Turner RG. Protocol performance. EAR AND HEARING 1988; 9(4):178-189.

Wilcox K and Horii Y. Age and changes in vocal jitter. J GERONTOL 1980; 35:194-98.

Young M, Bless DM, McNeil M, and Braun S. Relation of physical activity level to respiratory-laryngeal changes in voice. FOLIA PHON 1983; 35:185.

NCVS Status and Progress Report - 4
June 1993, 135-149

Parkinson's Disease: Longitudinal Changes in Acoustic Parameters of Phonation

Judith B. King, Ph.D.
Northern Arizona University
and
The Recording and Research Center, The Denver Center for the Performing Arts

Lorraine Olson Ramig, Ph.D.
Department of Communication Disorders & Speech Science, University of Colorado-Bolder,
The Recording and Research Center The Denver Center for the Performing Arts
and
Lee Silverman Center for Parkinson's Disease and Movement Disorders, Mayo Clinic Scottsdale

Jon H. Lemke, Ph.D.
Department of Preventative Medicine and Environmental Health, The University of Iowa

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

Abstract

This study was designed to longitudinally investigate the extent of phonatory variability in the voices of untreated (speech) patients with idiopathic Parkinson's disease and to document the effect of learning in this group of subjects. Microphone recordings were made from fourteen patients over time to measure the following acoustic variables: maximum duration sustained phonation (in seconds: secs), maximum fundamental frequency range (in semitones: st), mean fundamental frequency range (in Hertz: Hz) and fundamental frequency variability (in semitone standard deviation: stsd) in reading. During the course of the study none of the subjects received speech or voice

therapy. Growth curve models were constructed to simultaneously estimate a constant rate of decline of the variables over time (excluding the influence of baseline values) and to estimate the effect of learning due to baseline assessments.

Statistically significant constant decline in weighted mean values were found to exist over time on the variables maximum and mean duration of sustained vowel phonation, maximum and mean fundamental frequency range and mean fundamental frequency variability in reading. The results of this study reveal not only describe longitudinally the phonatory variability among this group of subjects, but also quantify the declining nature of that variability. In addition, a strong learning experience as a result of the baseline assessment was documented. A significant learning effect existed for all variables except fundamental frequency variability in reading. Furthermore, the extent to which learning continues after baseline assessment may actually underestimate the magnitude of normal decline in untreated patients with Parkinson's disease.

These consolidated, longitudinal descriptive data for these fourteen untreated subjects provide, 1) additional insight into the continuum of pathophysiologic changes that occur in the phonatory mechanisms of these patients, and 2) a point of reference for comparative studies from which to gauge the magnitude of treatment differences (and potential carryover) over the same time intervals for treated patients with Parkinson's disease.

Parkinson's disease, a progressive, degenerative neurological disorder resulting from nigrostriatal dopamine deficiency (Hornykiewicz, 1966; Hornykiewicz and Kish, 1986) affects nearly one million people nation wide (American Parkinson's Disease Association, 1990). Eventually, seventy-five percent of the individuals who develop the disease will manifest a speech disability which will limit their full participation in society (Hoberman, 1958; Canter, 1961; Barbeau, Dushay, and Spiegel, 1965; Martin, Loewenson, Resch and Baker, 1973; Logeman, Fisher, Bashes and Blonsky, 1978; Oxtoby, 1981; Streifler and Hofman, 1984, Ramig, Bonitati, Lemke and Horii, in submission).

According to reports of a number of individuals with Parkinson's disease, it is the progressive deterioration of speech and the patient's inability to effectively communicate with family, friends and employers that has the greatest negative impact on quality of life, self-esteem and hope for the future (Barbeau et al., 1965; Oxtoby, 1981; Ansel, 1985; MacCarthy and Brown, 1989; Ramig et al., in submission).

Classic descriptions of the voices of patients with Parkinson's disease include breathy, weak, monotonous voices, reduced in pitch variability and loudness, with short rushes of speech and imprecise consonants (Darley, Aronson and Brown, 1969 a,b; Critchley, 1981; Aronson, 1985). Muscular rigidity and hypokinesia within the laryngeal, respiratory and articulatory mechanisms have been associated with these classic perceptual Parkinson voice descriptors (Mueller, 1971). For example, bowed vocal folds resulting from rigidity in laryngeal musculature have been associated perceptually with the breathy voice quality, reduced loudness and abnormally short phonatory duration (Hansen, Gerratt and Ward, 1984). Respiratory muscular rigidity in combination with reduced vocal fold adduction have been associated with reduced maximum sustained phonation (Canter, 1965; Critchley, 1981). The perceptual manifestation of monotony of pitch and reduced range of fundamental frequency have been associated with muscular rigidity within the cricothyroid muscles which affect pitch change (Aronson, 1985). Reduced intelligibility associated with imprecise consonant articulation has been associated with a reduced range of accelerated (Netsell,, Daniel and Celesia, 1975; Hirose, Kiritani, Ushijima and Sawashima, 1981) or rigid (Hunker, Abbs and Barlow 1982; Weismer, 1984) movement of the articulators.

Given the substantial frequency of occurrence of laryngeal disorders among patients with Parkinson's disease (Logeman et al., 1978) and the well-documented phonatory variation among these patients (Canter, 1963; 1965), progressive phonatory capabilities should be quantified for both research and clinical purposes. Quantification of phonatory changes would further define the range and progression of problems facing individuals diagnosed with Parkinson's disease and provide baseline data from which to consider treatment-related changes.

Historically, traditional clinical efforts at speech therapy for individuals with Parkinson's disease have not been effective (Canter, 1963; Sarno, 1968; Allan, 1970; Greene, 1980; Aronson, 1985; Rubow and Swift, 1985; Weiner and Singer, 1989). As a result, currently many patients are neither referred for nor do they receive speech treatment, and their voices continue to deteriorate as the degenerative course of the disease progresses (Morley, 1955; Logeman et al., 1978).

However, a recent method of intensive, short-term (one month) voice therapy designed to target the underlying pathophysiologic mechanisms of voice has documented improved functional communication abilities in a group of forty patients with idiopathic Parkinson's disease (Ramig, Bonitati, Lemke and Horii, in submission). This behavioral treatment (focusing on the pathophysiologic laryngeal and respiratory mechanisms underlying the disordered voices in patients with Parkinson's disease) documents not only improved functional communication abilities but also suggests generalization of increased communication effort beyond phonation to improved articulatory precision as well (Ramig, Fazoli, Scherer and Bonitati, 1990). However, no information exists about the extent of variability for the same measures among untreated (no speech therapy) Parkinson subjects over a similar longitudinal span of time.

If the underlying bases for therapy-related improvement is to be thoroughly understood, the extent of phonatory variability among untreated subjects must be considered within the same clinical time trial and for the same measures as treated subjects. Descriptive data regarding acoustic measurements of phonation among untreated subjects with Parkinson's disease may provide: (1) additional insight into the continuum of pathophysiologic changes that occur in the phonatory mechanisms of these patients, and (2) provided further support for well-documented treatment-related changes observed in patients with Parkinson's disease following intensive voice therapy.

Because motor variability, which is well document in Parkinson's disease, is a clinical issue in assessing treatment effect and can be related, among other things, to learning, it was considered important to gauge the range of motor speech variability in these patients. The purpose of this study was to describe the extent of variability in acoustic parameters of phonation over time in a group of untreated (speech) subjects with idiopathic Parkinson's disease and to document the effect of learning in this untreated patient group.

Method

In order to clinically describe the extent of variability in acoustic parameters of phonation and to formulate a hypothesis regarding the pathophysiologic progression of mechanisms underlying the disordered voice of untreated (speech) subjects with Parkinson's disease, the protocol of this study was designed to consider three elements: (1) phonatory disorders reported in Parkinson's disease, (2) hypothesized underlying laryngeal and/or respiratory pathophysiology and (3) objective acoustic and physiologic variables. The framework and rationale for these three elements are summarized in Table 1 (Ramig et al., in submission).

Table 1.
Framework and Rationale for Perceptual Characteristics of Speech; Hypothesized Laryngeal and/or Respiratory Pathophysiology; Acoustic/Physiologic Variables Measured

Perceptual characteristics of speech	Hypothesized laryngeal and/or respiratory pathophysiology	Acoustic, physiologic 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 (Critchely, 1981)	Maximum breath, duration of sustained vowel phonation (sec)
Reduced pitch variability; monopitch (Logemann, et al., 1978; Aronson, 1985).	Rigidity cricothyroid muscle (Aronson, 1985).	Maximum range of fundamental frequency (st). Variability of fundamental frequency in connected speech (stsd).

The breathy, weak voices of individuals with Parkinson's disease have been associated with bowed vocal folds (Hansen et al., 1984). The acoustic impact of these combined respiratory and laryngeal pathophysiologic conditions was assessed using objective acoustic measures of maximum duration of sustained vowel phonation (seconds).

Rigidity in the cricothyroid muscles, responsible for controlling pitch change, has been associated with the monotonous voices of individuals with Parkinson's disease (Aronson, 1985). The impact of this pathophysiologic laryngeal condition was assessed using ratings of variability of fundamental frequency (STSD) in reading (Horii, 1987). (STSD reflects frequency variation in reading associated with intonation.) In addition, the maximum (st) and mean (Hz) ranges of fundamental frequency were calculated from maximum high and low vowel "ah" phonations. (ST maximum range reflects maximum limits of the system.)

Table 2.
Age, Stage of Parkinson's Disease, Primary Speech Symptoms, and Anti-Parkinson Medications for the Treated Subjects

Age (years)[1]	
Minimum	49
1st quartile	62
Median	67
3rd quartile	71
Maximum	82
Stage[2]	
I	3
II	8
III	2
IV	0
Primary Speech Symptoms[3]	
harsh voice	4
hoarse voice	3
hypernasal	1
imprecise articulation	6
increased rate	1
monotone	3
palialia	1
reduced loudness	10
vocal tremor	3
loudness decay	1
Ant- Parkinson Medications[4,5]	
carbidopa/levodapa	14
trihexyphenidyl HCL	1
amantadine hydrochloride	1
bromocriptine	1
pergolide mesyloate	2
seligiline hydrochloride	3

[1] Age information was unavailable for one subject
[2] Stage of disease information was unavailable for one subject
[3] Primary speech symptoms were unavailable for one subject
[4] Two subjects (14%) were taking additional medications for co-morbid conditions
[5] Medication information was unavailable for one subject

Ramig et al. (in submission) advocated measurement and training of maximum duration sustained vowel phonation and maximum fundamental range because these tasks: (1) stimulate subjects to increase their phonatory effort level, (2) stimulate subjects to sustain an increased effort level over time, (3) can be taught to subjects at most cognitive levels, (4) can be clinically analyzed efficiently to provide on-line feedback, (5) can be practiced independently by subjects, and (6) can be quantitatively analyzed even in cases of dysphonic voices which are common in Parkinson's disease. In addition, because these tasks have been associated with documented phonatory and perceptual changes in treated (speech) subjects (Ramig et al., in submission), they are the reasonable variables to use to quantify similar measures in untreated (speech) subjects with Parkinson's disease.

Subjects

Fourteen untreated (speech) patients (11 males and 3 females) diagnosed as having idiopathic Parkinson's disease participated as subjects in this study. Their ages, sex, stage of clinical disability (Hoehn and Yahr, 1967), self-reported primary speech symptoms and Anti-Parkinson medications are summarized in Table 2. The participants selected for this study responded to a Parkinson's disease support group newsletter advertisement which solicited patients with idiopathic Parkinson's disease to act as volunteer subjects. Given the participants' desire to act as untreated (speech) subjects, they also agreed not to participate in speech therapy services throughout the course of the study. However, at the conclusion of the study, speech therapy was made available to them. To the extent that was possible, voice recordings were collected on each subject at the same time of day post-medication during recordings. All control subjects were residents of the Denver-Boulder, Colorado area. Three of the subjects had been semiprofessional singers prior to the diagnosis of Parkinson's disease.

Data collection

Data collection instruments and methods used in this study were selected to reflect physiologic changes that support speech. The availability of the instruments and methods provides the basis for reaching objective, quantitatively-based conclusions regarding the extent of phonatory variability for individuals with Parkinson's disease. In addition, the procedures used in this study have been used routinely in studying normal (Ramig and Ringel 1983; Scherer et al., 1986; 1987) and disordered voices (Ramig et al. 1985; Ramig et al., 1990).

The Data collection protocol included voice and speech tasks from which the previously described acoustic measures were made. The rationale for these measures for subjects with Parkinson's disease is summarized in Table 1.

Voice and speech data were collected in the acoustics laboratory of The Recording and Research Center of the Denver Center for Performing Arts (RRC-DCPA). Voice recordings were obtained with each subject seated in a sound-treated booth (IAC-401-A) and fitted with a headset from which a microphone (AKG 190 E) powered by an AKG preamplifier was suspended 15 cm in front of the lips. After preamplification (AT 1M-1000), the microphone signals were digitized and video encoded through a Sony PCM 601 esD Digital Pulse Code Modulator and recorded onto a 1/2" videocassette tape (using a Panasonic AG VHS videocassette recorder).

In order to obtain acoustic voice data for analyses of the variable maximum duration sustained vowel phonation, subjects were asked to sustain the vowel /a/ for as long and steady as possible on a single deep breath. However, because of potential variability associated with maximum performance testing (Kent et al., 1987; Gelfer, 1989) as well as the variability associated with phonatory performance among individuals with Parkinson's disease, subjects were trained on this task until the experimenter determined that performance limits had been achieved. A timer with a second hand was placed in front of the subjects to increase their motivation and self-monitoring of each phonation. Encouragement was provided by the experimenter following each of the six to nine maximally sustained phonations collected from each subject. The task was terminated if the experimenter determined that, because of subject fatigue, maximum performance had been reached following four of five successive repetitions.

To obtain acoustic voice data for analyses of maximum fundamental frequency range, subjects were instructed to produce an /a/ at their highest sustainable fundamental frequency level (including falsetto) and at their lowest sustainable fundamental frequency level (without using vocal fry) (Hollien, Dew and Philips, 1971, Ramig and Ringel, 1983). the experimenter demonstrated both the step-wise and gliding (glissando) methods of production (Reich, Frederickson, Mason and Schlauch, 1990; Prater and Swift, 1984) from which the subjects could choose and again they were trained until both the experimenter and subject were satisfied that maximum performance limits were achieved. This task was repeated successively five to eight times within each recording session.

To obtain acoustic voice data for analyses of fundamental frequency and its variability, subjects were instructed to read the standard, phonetically balanced 'Rainbow Passage' (Fairbanks, 1960) using a comfortable rate, pitch and loudness. Although oral reading cannot be considered to be identical to spontaneous speech, the reading of the 'Rainbow Passage' is a feasible procedure to obtain uniform samples of connected speech from which objective measure could be derived (Canter, 1963).

Speech data analysis

The data analysis procedures selected for this project were chosen in order to maximize the relationship among objective measures, physiologic bases and functional communication. Furthermore, the analysis procedures have well established validity and reliability (Horii, 1987). Acoustic data were analyzed at the RRC and in the Communication Disorders and Speech Science Department at the University of Colorado-Boulder. Twenty percent of the data were re-analyzed to assess measurement reliability.

From the previously digitized and video encoded videocassette recordings, a Magnavox VHS videocassette recorder was used to play back the 1/2" videotapes which were input into a Sony PCM 501 ES digital audio processor. The analog signals were then input into a Panasonic SV 3700 professional digital audio tape deck and recorded onto Maxell DM 60 digital audio tapes (DAT) in order to interface the signals with a digital oscilloscope for further analysis.

In order to obtain measures of maximum duration sustained vowel phonation, each phonation was input into a digital oscilloscope (Data Precision Model 611, Data 6000). Cursors were placed to mark the monitor-displayed zero crossing preceding the first negative-going peak at the onset and zero crossing following the final positive-going peak at the offset of each vowel (Ramig et al., in submission).

Each attempt at maximum high or low phonation was low-pass filtered above the predicted fundamental frequency and input to the digital oscilloscope which provided measures of maximum fundamental frequency range in Hertz (HZ). The maximum high and low phonations were then converted to express the maximum range in semitones (st) (Ramig et al., in submission).

Measures of fundamental frequency and fundamental frequency variability during reading were derived from the digitized signals at 10,000 samples per second using a 12-bit analog-to-digital converter into a PDP 11/34 computer. Analyses of mean fundamental frequency (Hz) and fundamental frequency variability in semitone standard deviation (STSD) were derived by using a computer program Unidentified Fundamental Frequency Analyzer (UFO) (Horii, 1987). This program uses a peak-picking technique to analyze the digital data.

Statistical analysis

Growth curve models were constructed to simultaneously estimate a constant rate of decline of the variables over time (excluding the influence of baseline values) and to estimate the effect of learning due to baseline assessment (the difference between the baseline levels and the intercept of the line). The parameterization of this model for each subject follows:

$$E(Y/X) = \beta_0 + \beta_1 {*}Z + \beta_2 {*}Z{*}Days$$

where Z = 0, baseline
1, not baseline

The estimates are best understood by substituting the possible values of Z.

Substituting 0 for Z, provided the expected value of Parkinson's disease patients at baseline:

$$E(Y/X) = \beta_0$$

Substituting 1 for Z, provided the expected value of Parkinson's disease patients across the follow-up assessments:

$$E(Y/X) = (\beta_0 + \beta_1) + \beta_2 {*}DAYS$$

The estimate of $(\beta_0 + \beta_1)$ is the intercept of the line determined from the follow-up assessments. The estimate of β_2 is the slope of that line.

This parameterization provided two direct tests. First, in order to determine if there is a learning effect due to assessment, the following hypotheses were tested:

$$H_0: \beta_1 = 0 \text{ vs. } H_A: \beta_1 > 0.$$

If the null hypothesis could not be rejected, the β_1 was set equal to zero and the following model fit:

$$E(Y/X) = \beta_0 + \beta_2 {*}Days.$$

Second, whether or not β_1 was in the odel or set equal to zero, a decline in the outcome over time was tested with the following hypothesis:

$$H_0: \beta_2 = 0 \text{ vs. } H_A: \beta_2 > 0.$$

The models were independently fit for each subject. The final estimate $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ where weighted means of the subject specific estimates.

$$E(Y/X) = \beta_0 + \beta_2 {*}Days.$$

These weights are functions dependent upon the extent of follow-up from each individual.

Results

Acoustic variables

Duration of sustained vowel phonation and fundamental frequency range were considered in two ways, given the recommendation that measures of maximum performance be studied with caution (Kent et al., 1987.) The single maximum duration or range and the mean of the three best durations or ranges were derived for each variable in each recording condition.

Mean fundamental frequency range (st). Parameter estimates for growth curve analyses for weighted mean fundamental frequency range data for 11 of the 14 subjects are summarized in Table 3 (Three of the 14 subjects were lost for this measure due to too few follow-up assessments.) Figure 1 denotes the value of -0.00190 st (P=0.0001) for the estimated slope. This is equivalent to a decline of 0.695 st/year. The mean baseline value is 25,131 st and the mean learning effect is 1.891 st (P<0.0001). The mean learning effect is the difference between the intercept and the mean baseline value. The learning effect cannot be defined as the difference between the baseline measurement and the first follow-up since that difference includes both the learning effect and the effect of time which accounts for any additional decline with the progression of the disease. If the rate of decline remains constant, the projected length of time for this group of subjects to return to baseline values of mean fundamental frequency range was estimated to be 33.1 months.

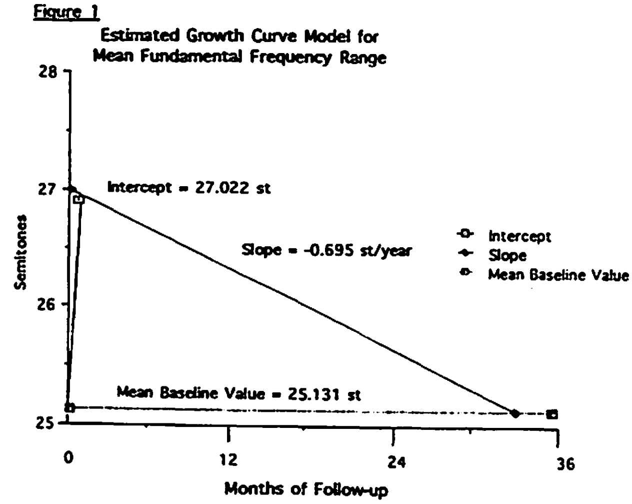
Table 3.
Parameter Estimates for Growth Curve Analyses for Mean Fundamental Frequency Range

Variable	N	Parameter	Estimate	Standard Error	P-Values
Mean					
Fundamental	11		25.131	0.247	<0.000
Frequency			1.891	0.283	<0.0001[1]
Range			-0.00190	0.000317	0.0001[2]
(st)					

[1] Mean baseline value significantly less than intercept.
[2] Significant decrease.

Maximum fundamental frequency range (st). Parameter estimates for growth curve analyses for weighted maximum fundamental frequency range data for 12 of the 14 subjects are 26.94 st, 2.633 st (P<0.001) and -0.00235 st (Intercept, 30-day slope and 1-year slope estimates from growth curve analyses for maximum fundamental frequency range, summarized in Table 4B, revealed a significant decrease of -0.856 st/year. This value, derived from the estimated slope decreased of -0.00235 st/day multiplied by 365 days, is plotted in Figure 2. The projected length of time for this group

Figure 1



of subjects to return to baseline values of maximum fundamental frequency range was estimated to be 37.4 months. the effect of learning is captured in Figure 2.

Table 4.
Parameter Estimates from Growth Curve Analyses for Maximum Fundamental Frequency Range

Variable	N	Parameter	Estimate	Standard Error	P-Value
Maximum	12		26.946	0.307	<0.0001
Fundamental					
Frequency			2.633	0.352	<0.0001[1]
Range			-0.00235	0.000377	<0.0001[2]
(st)					

[1] Mean baseline value significantly less than intercept.
[2] Significant decrease.

Figure 2

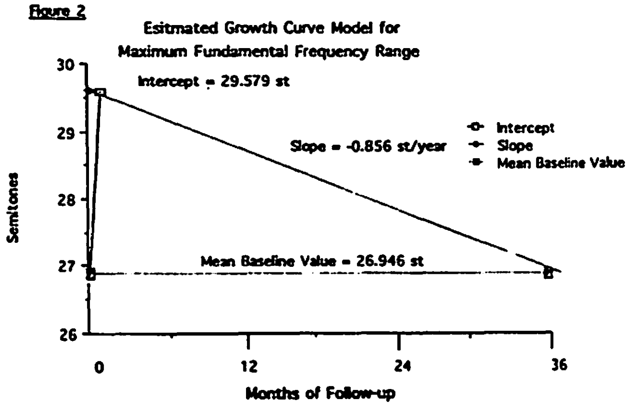


Figure 3

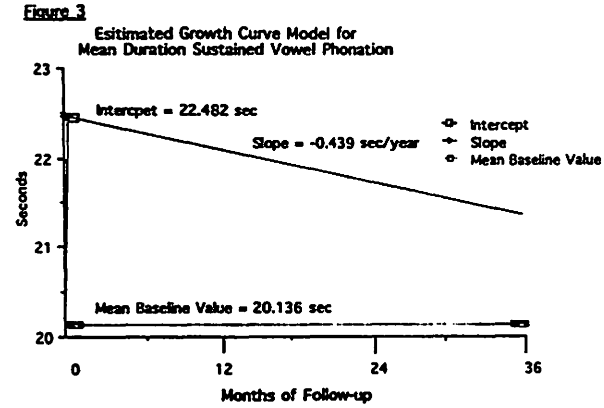


Table 5.
Parameter Estimates from Growth Curve Analyses for Mean Duration Sustained Vowel Phonation

Variable	N	Parameter	Estimate	Standard Error	P-Value
Mean	14	β_0	20.136	0.218	<0.0001
Duration					
Sustained		β_1	2.364	0.251	<0.0001[1]
Phonation					
(sec.)		β_2	-0.00120	0.000278	<0.0004[2]

[1] Mean baseline value significantly less than intercept.
[2] Significant decrease.

Mean duration sustained vowel phonation (seconds). Parameter estimates for growth curve analyses for weighted mean duration sustained phonation data for the 14 subjects are summarized in Table 5. Intercept, 30-day slope and 1-year slope estimates from growth curve analyses for mean duration sustained phonation revealed a significant decrease of -0.439 seconds/year. This value, derived from the estimated slope decrease of -0.00120 seconds/day multiplied by 365 days, is plotted in Figure 3. The projected length of time for this group of subjects to return to baseline values of mean duration sustained phonation was estimated to be 65.0 months.

Maximum duration sustained vowel phonation (seconds). Parameter estimates for growth curve analyses for weighed maximum duration sustained phonation data for the 14 subjects are summarized in Table 6. Intercept, 30-day slope sand 1-year slope estimates from growth curve analyses for maximum duration sustained phonation revealed a significant decrease of -0.530 seconds/year. This value, derived form the estimated slope decreased of -0.00145 seconds/day is multiplied by 365 days. The projected length of time for this group of subjects to return to baseline values of maximum duration sustained phonation was estimated to be 50.8 months. The effect of learning is captured in Figure 4.

Table 6.
Parameter Estimates For Growth Curve Analyses for Maximum Duration Vowel Sustained Phonation

Variable	N	Parameter	Estimate	Standard Error	P-Value
Maximum Duration	14	β_0	21.664	0.244	<0.0001
Sustained Phonation		β_1	2.214	0.282	<0.0001[1]
(sec.)		β_2	-0.00145	0.000312	<0.0002[2]

[1] Mean baseline value significantly less than intercept.
[2] Significant decrease.

Fundamental frequency variability (semitone standard deviation: stsd). Parameter estimates for growth curve analyses for weighted fundamental frequency variability data for 10 of the 14 subjects are summarized in Table 7. Intercept, 30-day slope and 1-year slope estimates form growth curve analyses for fundamental frequency variability revealed a decrease of -0.030 st/year which was not significant. This value, derived from the estimated slope decrease of -0.0000816 st/day is multiplied by 365 days.

Figure 4

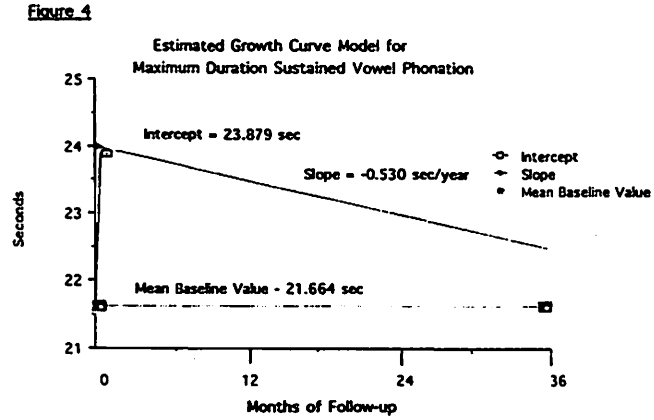


Table 7.
Parameter Estimates for Growth Curve Analyses for Fundamental Frequency Variability

Variable	N	Parameter	Estimate	Standard Error	P-Value
Fundamental Frequency	10	β_0	1.972	0.0115	<0.000[1]
Variability		β_2	-0.0000816	0.0000294	<0.0108[2]

1 This model includes β_1 since there is no significant learning effect: ($\hat{\beta}_1$ is equal to -0.000517: P-value = 0.426.
2 Significant decrease.

Discussion

Given the substantial frequency of occurrence of laryngeal disorders among patients with idiopathic Parkinson's disease (Logemen et al., 1978), and the well-documented phonatory variation among these patients (Canter, 1963; 1965), it is important to quantify these longitudinal changes in acoustic parameters of phonation. While the results of this study provide information regarding the relationship between the neuropathology associated with the progressive, deteriorating nature of Parkinson's disease and its effect on the phonatory capabilities observed among untreated (speech) subjects, there is also evidence to support that to longitudinally study phonatory capabilities of patients with Parkinson's disease it to intervene in the natural progressive deterioration of their speech. Because researchers have questioned for years the influence of research design for future studies.

Describing the range of variability and overall deterioration of the potentially unstable measures of maximum performance (Kent et al., 1987) and notoriously variable unstable characteristics among untreated (speech) subjects' voices over time provides a point of comparison from which to gauge the magnitude of treatment differences for the same variables over similar time frames in a treated group of Parkinson's disease patients (Ramig et al., in submission).

Specifically, Ramig et al. (in submission) reported that the magnitude of treatment differences in maximum performance speech tasks in their group of speech-treated patients with Parkinson's disease far exceeds the longitudinal changes in phonation reported for these fourteen subjects for the same variables. The most logical underlying physiological changes that would accompany the improved phonatory changes in Ramig's treated group of subjects would involve increased vocal fold adduction and respiratory support (sustained vowel phonation) and increased range of motion of the cricothyroid and thyroarytenoid muscles (increase in maximum fundamental frequency range) (Ramig et al., in submission).

Because phonatory capability and variability, which is well documented in Parkinson's disease, is a clinical issue in assessing treatment effect and can be related to learning, longitudinal phonatory changes in these 14 patients was documented. While there was no direct evidence to show that the results of the present study did no reflect motivational deficits, when compared to either normals (Canter, 1963; 1965; Logemann et al., 1978; Metter and Hanson, 1986) or the speech-treated group described by Ramig et al. (in submission), they clearly indicate that reduced physiological support for speech was characteristic of this group of 14 Parkinsonian subjects. For years researchers have questioned the influence of learning from baseline assessments within the context of clinical trials. The results of the present study document a very strong learning experience as a result of the baseline assessment. The constant rates of change are significant for all five measures of phonatory capability. To estimate these rates of change, growth curve models were constructed to exclude the baseline values in order to allow simultaneous estimation of the slope with estimates of the learning effect and a variance based upon baseline and follow-up data. Because the majority of these 14 subjects were in Stage 1 or 11 of Parkinson's disease, estimates of deterioration of phonatory capabilities to more severe stages cannot be extrapolated.

For mean and maximum fundamental frequency range and mean and maximum duration of sustained vowel phonation, the subjects demonstrated learning effects with a magnitude that was so substantial that very few had values that returned to their baseline level over the course of the study. If the rate of decline in phonatory capability since the first follow-up remains constant, on would expect to wait almost three years before subjects decline to their baseline fundamental frequency ranges and at least four years before they decline to the baseline durations of sustained vowel phonation. Additional learning effects beyond the subjects' baseline experiences is only speculation. If there are additional but decreasing learning effects, then the rate of decline in this study is overestimated and it would take even longer for subjects to return to base line values. In short, if learning continues after the first follow-up assessment, the normal decline in untested patients with Parkinson's disease has been underestimated.

Given the learning effects exhibited by these subjects in Stage 1 and 11, the sooner Parkinson's patient are exposed to assessment of their phonatory capabilities, the better. However, clinical experience has taught us to cautiously consider that subjects who volunteer as "controls" (receive no drug or behavioral treatment) may be less severely involved and/or motivated differently than another group of subjects who seek experimental treatment. Nonetheless, when the influence of baseline values were excluded in growth curve analyses, the magnitude of the normal decline was statistically significant in four of the five variables measured for this group of subjects. This observation supports similar evidence of reduced physiologic speech support among other patients with Parkinson's disease (Canter, 1963; 1965a and b; Logemann et al, 1978).

There are several important ramifications for all clinical studies of treatments of speech therapy for Parkinson's patients. First, randomization of patients to treatments is crucial. One would be well advised to control for stage and sex by matching. Second, due to the learning effect, two pre-treatment assessments are required. One would not want the learning effect first independent of treatment. However, variability of treatment effects may make it difficult, if not impossible, to separate the learning effect from the treatment effects if there is only one pre-treatment assessment. It is also more cost effective to have the extra measurement for each subject than to increase the sample size of subjects, each requiring multiple follow-up assessments. The uses of two pre-treatment assessments also provides one with a good run-in, that is, it serves as a trial to determine which subjects of eligible subjects are more likely to return for later follow-up assessments. A good run-in will reduce the proportion of subjects lost in follow up.

Several questions remain to be answered as a result of the present sample of subjects from this study. Are all subjects capable of demonstrating similar learning? If not, what factors influence the magnitude of the learning effect? How long does it take for subjects to demonstrate a maximum learning effect? How extensive is the learning effect from follow-up assessments one and two? What are the reasons for the dramatic learning effects? If follow-up is extended, will the rate of decline remain constant or become curvilinear?

It will also be important to similarly document the progressive changes in more sensitive phonatory measures and to identify which measures of voice are sensitive to learning and which ones are not. The consolidated descriptive dingdings for these fourteen untreated (speech) patients over time will also provide a point of reference for comparative studies from which to gauge the magnitude of treatment differences (and potential carryover) over the same time intervals for treated patients with Parkinson's disease. Continued longitudinal measures will enhance the understanding of not only the progression of Parkinson's disease——and but its effect on the phonatory capabilities observed among untreated (speech) subjects And finally, because individuals with Parkinson's disease report that it is the progressive deterioration of speech and their inability to effectively communicate that has the greatest negative impact on quality of life, self-esteem and hope for the future, the impact of the overall deterioration of functional communication cannot, then, be neglected in continuing research.

Acknowledgments

This work was supported by Grant DC00976 from the National Institute on Deafness and Other Communication Disorders. The authors would like to acknowledge the following people for thei individual and varied contributions to this papers: From the Roecording and Research Center of The Denver Center for the Performing Arts, Mary Serkowski, Wendy Savoy, Stefanie Countryman, Larry Brown, and Vern Vail; from Northern Arizona University, Heidi Stein.

References

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

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

Canter, G. J. (1965). Speech characteristics of patients with Parkinson's disease: II. Physiological support for speech. J. Speech Hear. Dis. 30(1), 44-49.

Critchley, E. M. R. (1981). Speech disorders of Parkinsonism: a review. J. Neurol. Neourosurg. Psych. 44, 751-758.

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

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

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

Green, H. C. L. (1980). The Voice and Its Disorders. London: Pritman medical.

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

Hirose, H., Kiritani, S., Usnijima, T. and Sawashima, M. (1981). Patterns of dysarthric movements in patients with Parkinosnism. Folia Phonaiatr. 33, 204-215.

Hoberman, S. G. (1985). Speech techniques in aphasia and Parkinonsim. J. Mich. St. Med. Soc. 57, 1720-1723.

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

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

Horii, Y. (1987) Personal Communication.

Hornykiewicz, O. and Kish, S. J. (1966) Metabolism of brain dopamine in human parkinonism: neurochemical and clinical aspects. In E. Costar, L. Cote, and N. Yahr (Eds.) New York: Raven Press.

Hornykiewicz, O. and Kish, S. J. (1986). Biochemical pathophysiology of Parkinson's disease. In M.D. Yahr and K.J. Bergman (Eds.) Adv. Neurology 45, 19-34.

Hunker, C. J., Abbs, J. H., and Barlow, S. M. (1982). The relationship between Parkinson rigidity and hypokinesia in the orofacial system: Aquantitative analysis. Neurol. 32(7), 749-754.

Kent, R. D., Kent, J.F. and Rosenbek, J. C. (1987). Maximum performance tests of speech produc tion. J. Speech Hear. Dis. 52, 367-387.

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

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

Netsell, R., Daniel, B., and Celesia, G. G. (1975). acceleration and weakness in Parkinsonian dysarthria. J. Speech Hear. Dis. 40, 1970-1978.

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

Reich, A. R., Frederickson, R. R., Mason, J. A., and Schlauch, R. S. (1990). Methodological vari ables affecting phonational frequency range in adults. J. Speech Hear. Dis. 55, 124-131.

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

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

Streifler, M., and Hofman, s. (1984). Disorders of verbal expression in Parkinsonism. J. Acoust. Soc. Am. 75, 570-580.

Weiner, W. J., and Singer, C. (1989). Parkinson's disease and nonpharmacologic treatment pro grams. J. Am. Geriat. Soc. 37, 359-363.

Weismer, G. (1984). Articulatory characteristics of parkionsonian dysarthria: segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In M. McNeil, J. Rosenbeck, A. Aronson(Eds.) The Dysarthrias: Physiology, Acoustics, Perception, Management. San Diego: College Hill Press. pp. 101-130.

NCVS Status and Progress Report - 4
June 1993, 151-167

Differential Phonatory Characteristics of Women with Amyotrophic Lateral Sclerosis

Edyth A. Strand, Ph.D.
Department of Speech and Hearing Science, The University of Washington
Eugene H. Buder, Ph.D.
Department of Speech and Hearing Science, The University of Washington
Kathryn M. Yorkston, Ph.D.
Department of Rehabilitation Medicine, The University of Washington
and
Department of Speech and Hearing Science, The University of Washington
Lorraine O. Ramig, Ph.D.
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

Abstract

Patients with specific neural subsystem involvement are often reported to present with particular perceptual characteristics of voice. This has been true even for diseases such as amyotrophic lateral sclerosis (ALS), a progressive disease in which mixed neurologic signs are present. This paper reports acoustic data on the phonatory performance of four women diagnosed with ALS who all had initial bulbar signs and progressive phonatory deterioration. The data presented in this paper demonstrate that the phonatory characteristics of women with ALS are not uniform, but in fact may vary greatly from patient to patient.

Respiratory and laryngeal function, and therefore phonatory performance, are often impaired as a result of neurologic disease (Darley, Aronson and Brown, 1975; Aronson, 1985; Ramig, Scherer, Titze & Ringel, 1988). Perceptual descriptions of the various voice problems associated with different neurologic disorders are numerous. Patients with specific neural subsystem involvement (e.g. Parkinson's disease) are often reported to present with particular perceptual characteristics of voice. However, Ramig, et al., (1988) point out that there are insufficient data to describe the laryngeal pathophysiology underlying the phonatory performance in neurologic disorders. This is especially true for diseases such as amyotrophic lateral sclerosis (ALS), a progressive disease in which mixed neurologic signs are present. The central purpose of this paper is to demonstrate, through acoustic analyses, that the phonatory characteristics of women with ALS are not uniform, but in fact may vary greatly from patient to patient. This acoustic analysis of phonatory performance allows us to formulate hypotheses regarding different respiratory/laryngeal pathophysiology as well as to investigate different compensatory strategies of patients (Ramig and Scherer, 1992).

ALS involves the degeneration of motor neurons in the cortex, brainstem and the spinal cord. The patient may exhibit paralysis and amyotrophy, bulbar disturbances, or pyramidal tract syndrome involving exaggerated reflexes (Bonduelle, 1975). Onset of the disease is characterized by widely varying symptomatology, depending on the site of motor neuron involvement, and as the disease progresses, symptoms may overlap. Upper motor neuron signs (UMN) include muscle weakness, increased muscle tone, hyperreflexia, extensor plantar reflexes and pseudobulbar palsy. Lower motor neuron (LMN) symptoms include muscle weakness, muscle atrophy and diminished or absent deep tendon reflexes. The usual clinical course is progressive with death occurring by three years in 50% of cases (Mulder, 1980; Tandan and Bradley, 1985). The degeneration of upper and lower motor neurons results in spastic and flaccid symptoms in the cranial and spinal muscles (Carpenter, McDonald & Howard, 1978). This degeneration does not proceed uniformly. As a result, flaccidity or spasticity may predominate at any given time. During the course of the disease, an assortment of neurologic signs, including dysarthria and phonatory deficits, develop as manifestations of the different neuronal structures involved.

Bulbar involvement, which affects speech, swallowing and voice production, is the primary initial symptom in about 25% to 30% of patients (Bonduelle, 1975; Dworkin and Hartman, 1979; Yorkston, Strand, Miller, Hillel, and Smith, 1993). The speech of individuals with ALS is typically classified as a mixed dysarthria due to the presence of both upper and lower motor neuron involvement. Upper motor neuron degeneration will cause symptoms of spasticity in the facial and laryngeal muscles. Flaccidity results from degeneration of lower motor neuron pools in the brainstem which innervate muscles used in articulation (e.g. facial nucleus, hypoglossal nucleus) and those used in phonation (nucleus ambiguus). While symptoms associated with both flaccidity and spasticity are often present at some point during disease progression, the relative influence of each to phonation will vary across patients and within individual patients (Carrow, Rivera, Mauldin, & Shamblin, 1974; Darley, Aronson & Brown, 1969 a,b).

Abnormalities in respiratory/laryngeal function are commonly noted at some point during disease progression, and changes in voice quality are among the most frequently noted parameters of the dysarthria exhibited by many ALS patients (Darley, Aronson & Brown, 1975; Ramig, Scherer, Klasner, Titze & Horii, 1990). This is especially true for those patients who exhibit early bulbar signs and for whom changes in phonatory performance may continue to be a prominent symptom throughout the course of the disease. The nature of the voice deficit has been described by listeners according to quality and stability of phonation. Perceptual descriptions of voice quality associated with ALS dysarthria include: wet hoarse quality; strained-strangled quality, breathiness, harshness and hypernasality. Perceptual descriptions of vocal stability include tremor and long-term fluctuation in both frequency and intensity. These perceptual observations of changes in voice quality and stability are often reported in the clinical research literature. In contrast, the body of literature providing acoustic quantification of the articulatory and phonatory performance of ALS speakers is still quite small. This is surprising since such analyses may yield insight regarding the nature of the movement disorder in dysarthria (Weismer, 1984; Kent, Kent, Weismer, Sufit, Rosenbek, Martin & Brooks, 1990; Weismer, Martin, Kent and Kent, 1992) and provide sensitive quantification of the result of laryngeal pathophysiology (Ramig et al., 1988).

Acoustic studies to date have examined a variety of the features of the phonatory performance of ALS speakers, including aspects of fundamental frequency, intensity, cycle to cycle variability and tremor. Measures of fundamental frequency provide information regarding the nature of the periodicity of vocal fold vibration. Frequency and intensity contours for sustained phonation and for phrase production allow examination of the ability of the speaker to maintain a steady state phonation and volitionally control changes in frequency and intensity for stress. Measures of spectral energy provide information regarding laryngeal efficiency across individual speakers.

Measures of cycle to cycle variability are frequently reported in the phonation literature to describe cycle to cycle changes in frequency and intensity. They have been used to describe characteristics of dysarthric phonation, and as indices of deterioration of phonatory function. However, these perturbation measures are often difficult to interpret, due to the influence of data collection methods and equipment (Titze and Winholtz, in review); strategies of the speaker for sustaining phonation, and the use of varying analysis algorithms across laboratories (Titze, Horii, and Scherer, 1987). However, with carefully designed and described procedures, such dependent variables become replicable and should allow examination of differences in vocal stability across individual speakers.

A brief review of the acoustic literature examining the speech and voice of speakers with ALS illustrates the utility of acoustic measures in the study of ALS dysarthria. Kent, Kent, Rosenbek, Weismer, Martin, Sufit and Brooks, (1992) presented acoustic quantification of phonatory performance in 10 female ALS speakers. Their acoustic measures of phonatory function for sustained vowel production indicated abnormalities in fundamental frequency, and vocal stability. All ten ALS speakers were reported to exhibit fundamental frequency values that were either plus or minus two standard deviations from a control group mean. ALS subjects varied among themselves in measures of vocal stability. Five of the ten ALS subjects exhibited shimmer and jitter values that were two standard deviations above the control group mean, even though intelligibility scores were high. Five others exhibited stability measures that did not differ significantly from the control group mean. The authors suggest that acoustic analyses may facilitate early detection of articulatory and laryngeal signs of disease, but caution that group data do not always apply to all individuals who may vary greatly among themselves.

Group means reported for males by Kent, Kim, Weismer, Kent, Rosenbek, Sufit, Brooks and Workinger (in review) indicated higher fundamental frequency for male ALS speakers than a control group. In their analysis of vocal stability, they reported no differences in jitter or shimmer for male ALS speakers versus a control group, but significantly higher group means for women ALS speakers on both variables. Their data provide additional evidence that individual differences (in this case gender differences) need to be taken into account when reporting group data.

Two longitudinal studies have indicated individual variability in the changes in phonatory performance during disease progression. Ramig, Scherer, Klasner, Titze, and Horii (1990) measured the phonation of a male ALS speaker five times over a six month period. They reported significantly decreased phonatory stability in session five compared to sessions one, two and three, as seen in measures of increased shimmer, jitter and coefficient of variation for both amplitude and frequency. The authors attribute these changes in vocal stability to changes in respiratory as well as laryngeal control mechanisms that occurred due to disease progression. These physiologic changes contributed to both long-term and cycle to cycle variability.

Kent, Sufit, Rosenbek, Kent, Weismer, Martin, and Brooks (1991) reported a decrease in mean fundamental frequency over a two year period for a 53 year old woman with ALS. The authors note that any conclusions regarding short term vocal stability, however, are tenuous because measures of shimmer, jitter, and signal to noise ratio were variable across the 7 test sessions during the two year period. Unlike the Ramig et al. study, the measures of phonatory stability did not differ greatly from values obtained for a control group. This could be due to the fact that the measures were made over different sample lengths and with the use of different software.

Acoustic analysis has also been used to quantify the rapid voice tremor or "flutter" sometimes associated with the phonation of speakers with ALS. Aronson, Ramig, Winholtz, & Silber (1992) reported that ALS speakers, with perceptible tremor or flutter, produced frequency and amplitude modulations that varied in amplitude as compared to control subjects who exhibit low levels of modulations. While some patients exhibited more amplitude modulation, other exhibited more frequency modulation. Still others exhibited frequency and amplitude modulations that were equal and in phase. The authors suggest that the neural origin of these modulations are peripheral rather than central and may be due to loss of motor units resulting in intermittent absence of motor unit firing in the intrinsic laryngeal muscles.

Although the body of acoustic data quantifying the perceptual observations of voice quality in ALS is still quite small, acoustic work to date has illustrated that such analysis has much to offer. Together, these data have indicated that acoustic measures facilitate early detection, provide sensitive indices of disease progression and lead to specific hypothesis about underlying mechanisms of pathophysiology. These data also illustrate the variability in voice characteristics among ALS speakers and suggest that examination of individual phonatory performance may take us further in formulating hypotheses about the physiologic laryngeal changes that occur in ALS.

Observations of ALS subjects participating in our clinical research project have corroborated the intersubject variability in ALS voice production that has been recently noted in the literature. The variability of phonatory quality and stability exhibited by these ALS subjects motivated further study of the laryngeal involvement in ALS. It was of interest to acoustically describe phonatory characteristics among patients in order to determine what specific acoustic parameters contributed to the differences heard in vocal quality. The acoustic descriptions also allow one to begin to formulate specific hypotheses concerning physiologic changes and compensatory behaviors that may be contributing to differences in phonatory performance.

Method

Subjects

Four women (BS; CS; MB; and JL), diagnosed with ALS and ranging in age from 41-70 participated in this study. All four were seen for initial evaluation at the Neuromuscular Speech and

Swallowing Disorders clinic at the University of Washington Medical Center within three months of the diagnosis. All four had early bulbar symptoms and were dysarthric but intelligible at initial visit. Their patterns of articulatory and phonatory impairment, however, varied. Descriptive data for each subject at the time of recording (which was not at time of first visit) are noted in Table 1. All subjects were native English speakers with no previous speech, language or voice problems and no history of neurologic deficit prior to the diagnosis of ALS. No sensory deficits were reported by any of the four women. There was no history of respiratory dysfunction other than that associated with the ALS. While there were differences in sentence intelligibility, vital capacity and velopharyngeal function, the most striking perceptual characteristic, differentiating the four women, as judged by two speech/language pathologists, was phonatory performance (Table 2).

Table 1.
Descriptive Data at Time of Recording: Sentence Intelligibility (Yorkston, Buekelman and Traynor, 1984), Vital Acapacity, Rating of Velopharyngeal (VP) Function, and Vocal Quality

	Subject			
	CS	JL	BS	MB
Age	53	70	41	63
Months Post Diagnosis	11	7	2	9
Sentence Intelligibility (percentage)	27	41	100	38
Vital Capacity (liters)	2.4	1.1	3.2	1.5
VP Function	Severely Reduced	Moderately Reduced	Severely Reduced	Moderately Reduced
Voice Quality	Breathy	Strained-Strangled	Inconsistent Harshness and Flutter	Consistent flutter: inconsistent harshness
Tongue Strength and Movement	Severely Reduced	Moderately Reduced	Good	Moderately Reduced

Table 2.
Perceptual Description of Voice at Time of Recording

Subject	Sustained Phonation
CS	Consistent breathy quality; little variation in frequency; consistently low intensity; hypernasal.
JL	Very short phonation time; voice breaks; hyperadductive, strained-strangled quality with great variation in intensity.
BS	Tremulous quality, characterized by short term fluctuations in frequency and intensity; hypernasal.
MB	Regular alterations in frequency and intensity; inconsistent harshness.

Procedure

Three of the four subjects were recorded in a IAC sound treated booth for digital audio recording, using a Panasonic (SV3700) digital audio recorder. The fourth subject was recorded in a quiet examination room at UW Medical Center using a Casio (model DA-7) digital audio recorder. A Head mounted microphone (Shure, model 274-016) allowed for a consistent 12 cm. mouth to microphone distance for all subjects and recording levels were kept constant.

A sentence production task (Yorkston, Beukelman & Traynor, 1984) was administered for intelligibility measures. Each subject was asked to produce two samples of the vowel /a/ at a comfortable pitch and loudness level, for as long as they could sustain phonation. This vowel was chosen because the production of the open back sound places limited demands on tongue movement and posture, which varied among the four subjects. The sample was kept to two productions due to the fact that two patients fatigued very easily and one subject tended to go into laryngospasm after repeated sustained phonation. Each subject rested for as long as they wished between vowel productions. They then read 6 CVC words in isolation and in the carrier phrase "say _____ again". The carrier phrase was produced as a model by the examiner to ensure the subjects would attempt to put the stress on the second (target) word.

Data Analyses

Acoustic analyses of phonatory function was completed with CSpeech, version 4.0, a speech analysis software program (Milenkovic, 1992). Each sustained phonation was digitized at 22 Khz. Acoustic measures completed were: mean fundamental frequency and standard deviations during sustained phonation and phrase production; pitch and intensity contours during sustained phonation and phrase production; and measures of phonatory stability (shimmer, jitter and signal to noise ratio) in sustained phonation and in the word "top" produced in a carrier phrase. Fundamental frequency was measured over the entire phonation. Phonatory stability (shimmer, jitter, signal to noise ratio) was measured from three 500 msec samples from the sustained phonation. A 500 msec sample was measured 500 msec from initial glottal pulse in order to avoid onset phenomena. A second 500 msec sample was analyzed at midpoint of phonation. The third 500 msec sample was analyzed just prior to the last 500 msec of phonation in order to avoid offset characteristics. The sample size of 500 msec was chosen because the frequent aperiodicity of these voices caused errors to occur in the automatic analysis program for larger samples.

Results

Waveforms and Frequency and Intensity Contours

Figure 1 illustrates the frequency and intensity contours for sustained vowel production. In each quadrant of the figure, a subject's waveform is depicted on the top, fundamental frequency (f0) contour in the middle, and root mean square (RMS) intensity contour on the bottom. Time and value scales for each subject are held constant and the four sets are vertically and horizontally aligned on these scales to allow cross-subject comparisons. For each sample, the subjects were given the same instruction: "say the vowel /a/; make it strong and steady and hold it for as long as you can."

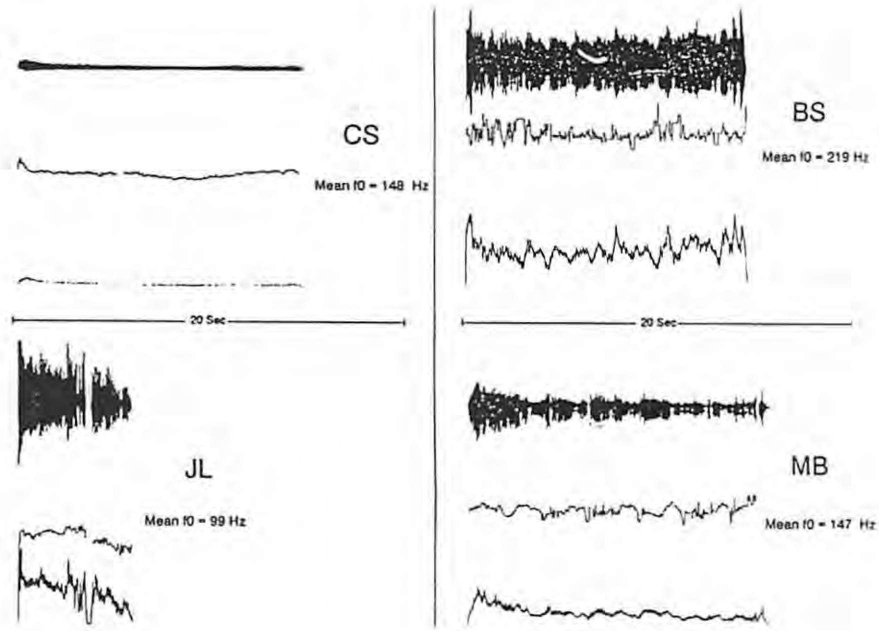


Figure 1. Waveform, frequency contour and intensity contour for one sample of sustained phonation for each of the four ALS speakers.

The longest duration of phonation was produced by MB (15.8 sec). CS and BS produced duration of 14.7 and 14.5 respectively. JL produced only 5.9 seconds of phonation. Frequency and intensity contours are presented to illustrate variability exhibited among the four women in these parameters. Absolute intensity values are not reported as calibration equipment was not available for all subjects. Because the recording levels and microphone to mouth distance were kept constant, visual inspection of relative intensity levels among the four women is still informative.

The waveforms for the sustained phonations of two subjects (BS and MB) illustrate short and long term variability in intensity and frequency. Perceptually, BS exhibits a tremulous quality. This is characterized in the waveform by irregular fluctuations in frequency and intensity. She exhibits a rapid tremor in both frequency and intensity contours of 6 to 6.5 Hz. She also exhibits a longer term fluctuation of 1.1 to 1.25 Hz, specifically in intensity. MB's most salient perceptual quality is a regular fluctuation in both pitch and loudness. Her waveform and contours illustrate this 2.2 to 2.6 Hz fluctuation in both frequency and intensity. MB also exhibits a very rapid tremor component of 8 to 10 Hz seen primarily in intensity.

Very similar acoustic patterns are illustrated in the waveforms and fundamental frequency and intensity contours for the phrase "say top again" (Figure 2). Phrase durations were: 2630 msec for BS; 5259 msec for MB; 4375 msec for CS; and 4747 msec for JL.

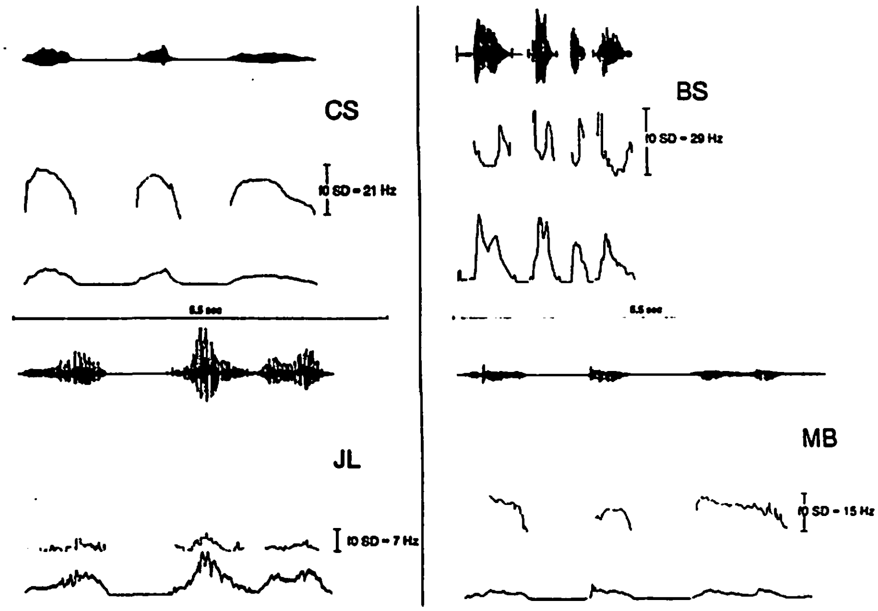
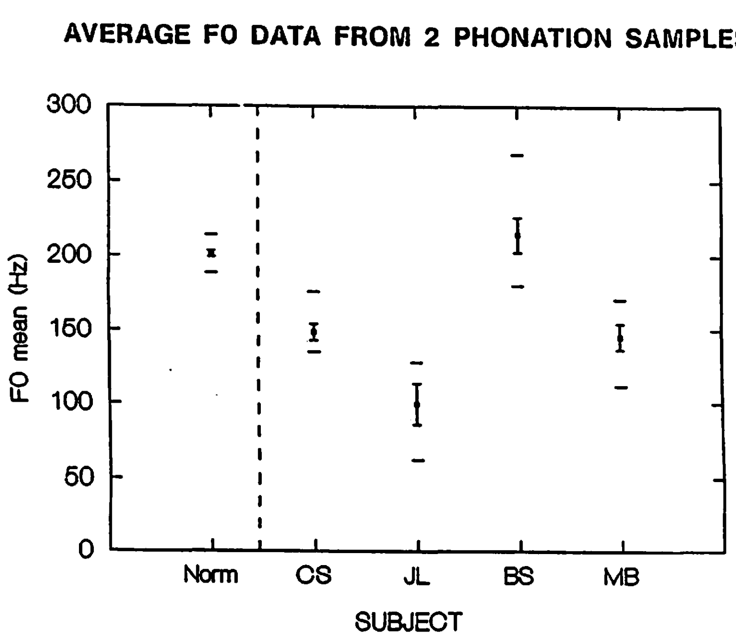


Figure 2. Waveform, frequency contour and intensity contour for one production of the phrase "say top again" by each of the four ALS speakers.

Fundamental Frequency and Intensity Variation Data

Figure 3 illustrates the average mean fundamental frequency values, standard deviations and ranges for an average of two samples of sustained phonation for each ALS speaker and for one control subject (who was 62) who is representative of published norms for f0. This control example is included only as a reference for f0 values typically seen in women with normal phonation. The



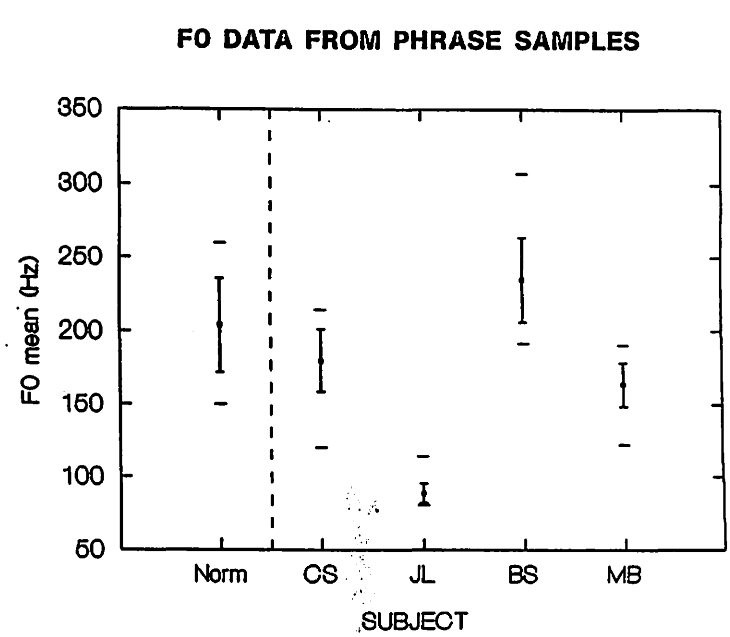


Figure 3 (left). Average fundamental frequency for two samples of sustained phonation for one control subject (NORM) and for the four ALS speakers. Figure 4 (right). Average fundamental frequency for one production of the phrase "say top again" for one control subject (NORM) and the four ALS speakers.

four ALS speakers vary greatly in f0 values and deviate by at least 2 standard deviations from normal values. Note the small standard deviation and range for the control speaker, and the greater, but differing profiles for the ALS speakers.

Differences in the ability of the four speakers to vary f0 for stressed words in phrase production is shown in Figure 4 . The control subject exhibits a range of 110 Hz. Three of the ALS speakers exhibit reduced range: CS = 94 Hz; JL = 33 Hz; and MB = 31 Hz. BS exhibits a range of 115 Hz (although this may be due to involuntary register shifts). Table 3 lists the fundamental frequency values (minimum, maximum, mean and sd) for sustained phonation and for one production of "say top again" for each speaker.

Table 3.
Mean Fundamental Frequency Values Averaged for Two Samples of Sustained Phonation and One Sample of Phrase Production

Subject	Sustained Phonation				"Say Top Again"			
	$X F_o$	Min. F_o	Max. F_o	SD F_o	$X F_o$	Min. F_o	Max. F_o	SD F_o
Control	201.5	188.5	214.0	2.2	204.0	150.0	260.0	32.0
CS	148.5	134.5	175.5	5.7	180.0	120.0	214.0	21.2
JL	99.5	62.0	127.5	13.9	89.0	81.0	114.0	6.6
BS	214.5	180.0	268.5	11.6	235.0	192.0	307.0	29.0
MB	145.5	112.0	171.0	9.0	164.0	122.0	191.0	15.1

Cycle to Cycle Variability

Results of analysis of percent jitter, percent shimmer and signal to noise ratio were derived from three 500 msec samples within one production of /a/. This was done to illustrate the variability within and among subjects. Careful rules were established to allow readers to judge the validity of these measures. The perturbation results are summarized within 100 msec tokens; 5 values are thereby obtained from each 500 msec sample. The autocorrelation based perturbation algorithm implemented in CSpeech (Milenkovic, 1987) notes an error count for each token analyzed. An "error" refers to any point in the token at which the algorithm encounters unclear pitch periodicity; in such cases an approximate value is taken and the error flag is set to alert the user. Each token reported here had three or less errors. Of the 60 tokens measured, 54 had no errors; 4 had one error; 1 had two errors; and 1 had three errors. The six token values with error reports were of the same magnitude as those with no errors. In order to maintain this level of accuracy, the 500 msec sample was occasionally shifted 25 to 200 msec in order to avoid passages in which extreme aperiodicity yielded high error counts. One subject's (JL) extreme aperiodicity required that we use a 400 msec sample at the end of her phonation. While the results may therefore underrepresent passages of extreme cycle-to-cycle instability, we are confident that the measures are a valid application of the Milenkovic algorithm.

Table 4.
Percent Jitter, Percent Shimmer and Signal-To-Noise Ratio (dB) Values for Three 500 msec Samples of a Sustained Phonation for the Four ALS Speakers

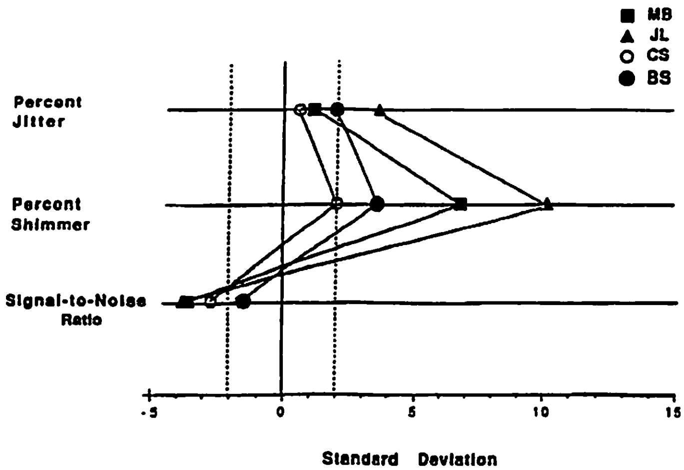
Subject	Beginning	Middle	End	Mean
	Percent Jitter (SD)			
Control	.18 (.05)	.38 (.08)	.71 (1.07)	.42
MB	.62 (.08)	1.93 (1.41)	.75 (.19)	1.10
JL	1.72 (.33)	1.42 (.66)	2.45 (.64)	1.86
CS	.31 (.03)	.76 (.24)	.91 (.20)	.66
BS	.85 (.14)	1.11 (.98)	1.68 (1.26)	1.20
	Percent Shimmer (SD)			
Control	1.80 (.12)	3.00 (.18)	3.02 (1.90)	2.60
MB	5.37 (.68)	13.12 (7.87)	5.01 (2.12)	7.80
JL	6.66 (1.25)	13.34 (6.64)	11.04 (4.03)	10.30
CS	2.57 (.39)	4.86 (.96)	4.93 (.80)	4.12
BS	2.93 (.56)	4.63 (3.59)	8.32 (6.19)	5.29
	Signal-To-Noise Ratio (SD)			
Control	23.20 (1.40)	16.40 (.10)	24.50 (6.50)	21.30
MB	14.20 (.50)	9.40 (2.80)	12.10 (.70)	11.90
JL	11.50 (.80)	13.60 (2.80)	8.80 (1.20)	11.30
CS	17.80 (.70)	12.90 (.30)	12.40 (.40)	14.30
BS	19.80 (2.20)	18.10 (5.50)	14.50 (3.00)	17.40

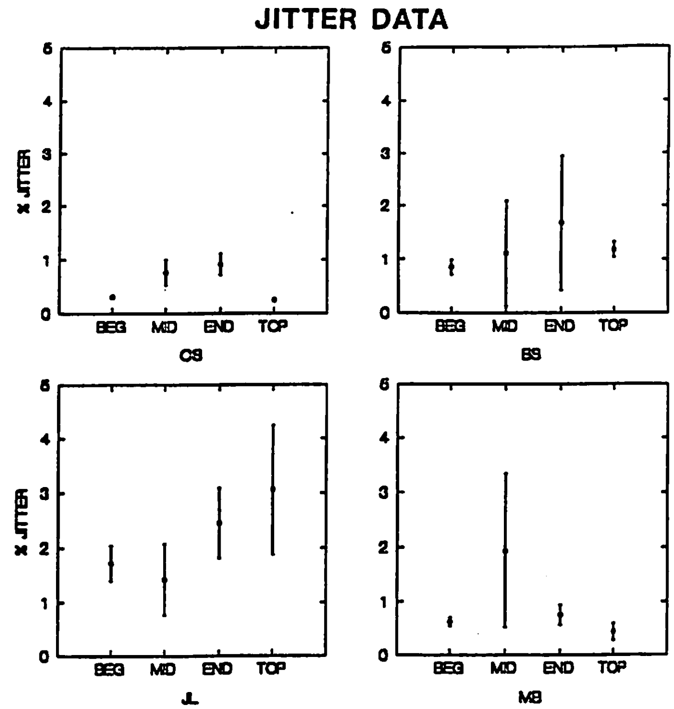
Results of the perturbation measures for sustained phonation are summarized in Table 4. All four ALS speakers exhibited greater shimmer and jitter values than the control subject. (Our control subject values are very similar to control group values reported by Ramig et al., (1990), and Kent, et al., (1992). It is to be expected that the ALS perturbation values would be different than those exhibited by non neurologically impaired voices. The control values are reported here as a point of reference, to illustrate different patterns of vocal stability among the four ALS voices. Figure 5 illustrates the variable deviation of mean perturbation data of the four ALS speakers from the control perturbation means. The dotted lines indicate two standard deviations above and below the control mean which is represented by the solid vertical line. Only one subject (JL) exhibited jitter values that were more than 2 standard deviations above the control mean. All four ALS speakers exhibit shimmer values that were 2 standard deviations or more above the control mean. Two subjects (JL and MB) exhibited SNR values that were greater than 3 standard deviations below the control subject mean. CS was greater than 2 standard deviations.

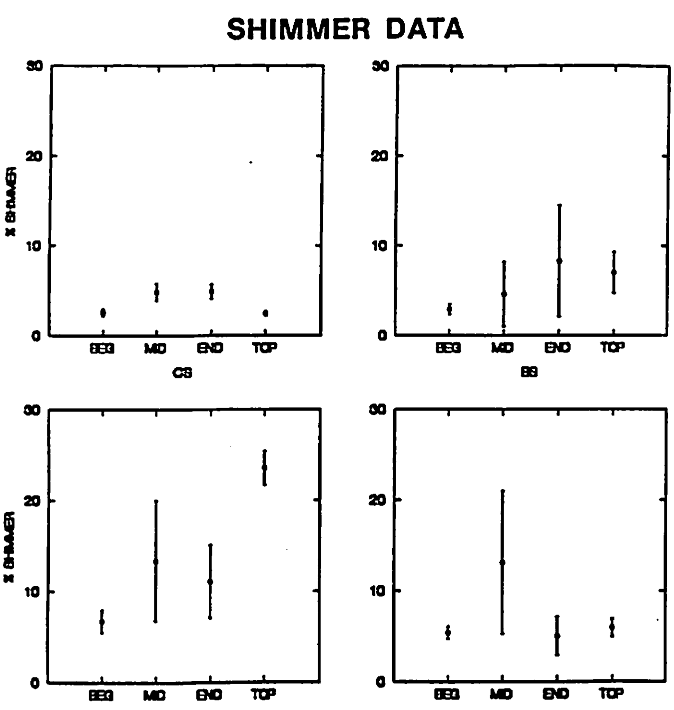
Perturbation values for intraword phonation (Table 5) illustrate vocal stability in short samples (200 msec) of phonation during production of the vowel in the word "top produced in the carrier phrase "say top again". (A 100 msec sample was used for the control speaker due to her shorter vowel duration. As a result the analysis program used only 1 token, so there is no sd reported.) Figures 6 and 7 illustrate percent jitter and percent shimmer for the three samples analyzed during sustained phonation and within the word "top" for each of the four ALS speakers. It is of interest to examine how perturbation measures varied among the four speakers at different points in sustained phonation, as well as how vocal stability within a short sample of phonation in connected

Table 5.
Perturbation Means and Standard Deviations for 200 msec of Phonation in "Top" During Production of Phrase "Say Top Again."

Subject	% Jitter (SD)	% Shimmer (SD)	SNR (SD)
Control	.23	1.29	27.60
MB	.43 (.16)	5.93 (.98)	12.90 (1.00)
JL	3.05 (1.31)	22.65 (2.46)	7.30 (.80)
CS	.25 (.01)	2.31 (.25)	20.40 (1.10)
BS	1.18 (.14)	7.06 (2.30)	16.60 (.70)







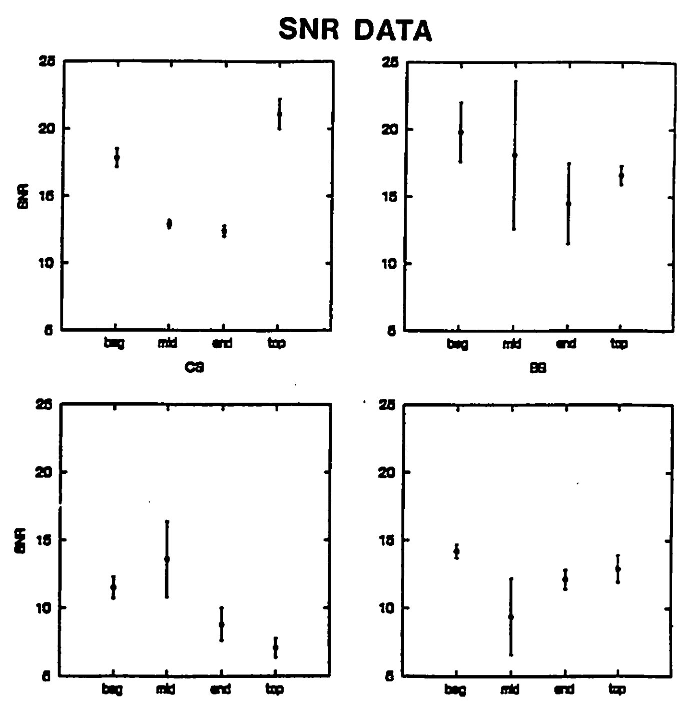


Figure 5 (upper left). Number of standard deviations above or below the control speaker mean for percent jitter, percent shimmer and signal to noise ratio. The dotted lines indicate two standard deviations above and below the control mean which is represented by the vertical solid line. Figure 6 (upper right). Percent jitter for three 500 msec samples taken at the beginning, middle and end of one sustained phonation and during 200 msec of the vowel / ɔ / in the word "top" produced in the phrase "say top again" for each of the four ALS speakers. Figure 7 (lower left). Percent shimmer for three 500 msec samples taken at the beginning, middle and end of one sustained phonation and during 200 msec of the vowel / ɔ / in the word "top" produced in the phrase "say top again" for each of the four ALS speakers. Figure 8 (lower right). Signal to noise ratio for three 500 msec samples taken at the beginning, middle and end of one sustained phonation and during 200 msec of the vowel / ɔ / in the word "top" produced in the phrase "say top again" for each of the four ALS speakers.

speech would differentiate the speakers. The range of perturbation values across the three samples taken during one sustained phonation varied among subjects. Jitter results for samples taken at the beginning, middle and end of phonation indicate that CS exhibits the smallest range (.60) compared to the other three ALS speakers (MB = 1.31; JL = 1.03 and BS = .83). A similar pattern of results was noted for shimmer. CS again exhibits the smallest range (2.36); MB (8.11); JL (6.68); and BS (5.39). For both jitter and shimmer, three of the four ALS speakers produced perturbation values for the word top that was within the range (or lower) exhibited during sustained phonation. JL, however, exhibits greater shimmer and jitter values during production of the vowel within a word. Figure 8 illustrates the SNR values for each subject at the three points during phonation and in the word "top". The SNR ranges were: CS=5.4; MB=4.8; JL=4.8; and BS=5.3.

Discussion

Acoustic analysis has been used to describe aspects of phonatory function in four women who exhibited changes in voice quality due to ALS. These findings, as predicted, demonstrate differences in characteristics of vocal function among the four ALS speakers. These acoustic analyses are discussed in relation to how they might reflect neuromuscular events (consequences of pathophysiology) in the larynx, as well as how they may reflect compensatory strategies of the patient.

The fundamental frequency data from these four speakers are consistent with Kent et al., (1992) who report the f0 values for 10 women with ALS to all vary (+ or -) at least two standard deviations from the normal group means. In this study, three of the four women exhibited lower than normal fundamental frequency, while one subject exhibited f0 values that were more than two standard deviations above the Kent normal group mean and our control subject. So, while changes in F0 seem ubiquitous in ALS, the direction of these changes is not always consistent. Fundamental frequency relates to a variety of laryngeal parameters, including decreased muscle mass, reduced crico-thyroid function, and degree of spasticity or flaccidity of laryngeal muscles. The fundamental frequency data, along with examination of waveforms and intensity and frequency contours, suggests some explanation for the varying perception of vocal quality heard among these four speakers.

JL perceptually produces a strained-strangled vocal quality and a very short phonation time. Her waveform indicates hyperadductive laryngeal vibration (as evidenced by elevated intensity and instability). Her extremely low F0, the relatively high intensity, and the inability to cease phonation during the voiceless interval in the word "top" suggests hyperadduction of laryngeal muscles. This may be due to increased spasticity of the vocal folds, to decreased strength of laryngeal abductors, or perhaps compensatory adduction of ventricular folds. The frequency contours for phrase production would also lead us to predict loss of crico-thyroid function. While the intensity contour shows some variation in intensity, there is little variability in frequency. Because of the nature of the instructions and modeling, and the consistent effort of the subject (who by self-report was not depressed), we believe this is suggestive of an inability to volitionally raise frequency for stressed words.

CS shows a marked contrast to JL. She perceptually exhibits consistently breathy phonation with low volume. Her waveform illustrates weak adduction and low intensity, suggesting flaccidity in laryngeal muscles. Her contours illustrate very little frequency and intensity variation during the production of about 16 seconds of continuous phonation. Although her vital capacity is adequate for speech, her intensity is consistently weak and is the least variable among the four women. The phrase contours for CS suggest that she may be unable to volitionally raise F0 or intensity. These data plus her breathy quality suggests general weakness in laryngeal muscles.

While JL presents with a "hyperadductive" and CS with a "hypoadductive" set of acoustic parameters, MB and BS suggest a third category of laryngeal dysfunction. Both MB and BS exhibit fluctuation in intensity and frequency, and periods of harshness or hyperadductive vibration accompanied by periods of relative breathiness.

Perceptual descriptions of the voice of BS include tremor, high pitch, hypernasality and inconsistent harshness. Her sustained phonation waveform illustrates the inconsistent fluctuations in frequency and amplitude. The contours illustrate generally high frequency and intensity, but with much fluctuation in both. Unlike those seen for JL and CS, the frequency contours for phrase production illustrate BS's ability to vary f0. The spikes noted in the frequency contour seem to be register shifts. They are consistently noted in her speech and not artifact and may be due to postural compensation strategies to reduce nasal airflow.

Perceptually, MB exhibits inconsistent harshness and a striking, somewhat regular variation in pitch and loudness. Her sustained phonation waveform illustrates this unusual long term fluctuation in both frequency and intensity which occurred throughout sustained phonation. It was difficult from the acoustic signal alone to determine to what degree respiratory function (versus laryngeal function) was contributing to the long term modulation. Examination of laryngeal function during sustained phonation with videostroboscopy showed regular increases in the degree of laryngeal adduction (characterized especially by severe ventricular fold approximation with anterior/posterior compression) followed by regular periods of less complete closure.

Measures of shimmer, jitter and signal to noise ratio are designed to examine source characteristics, and may also be useful in examining individual differences in laryngeal vibration. JL exhibits the largest cycle to cycle variability in frequency and intensity. Ramig et al., (1988) notes that periodic variation may depend largely on the integrity of the muscles governing vocal fold elasticity. JL's perturbation results are consistent with the observation of hyperadductivity and laryngeal spasticity. The large amplitude variation noted in JL's RMS contour may be indicative of deficits in laryngeal adductory-abductory muscle function. This would be consistent with the earlier observation of her inability to cease voicing during voiceless stops. CS, whose waveform indicates weak, steady and breathy phonation also exhibits the lowest jitter and shimmer values of the four subjects (although still higher than the control subject values). These results, in addition to the other data, suggest that while CS produces weak adduction, the periodicity of vibration is not as compromised as the other three women.

Variability among subjects was noted not just in mean perturbation data, but also at different points during sustained phonation. Examination of these differences allows more specific description of intrasubject and intersubject variability and allows inferences about respiratory/laryngeal patterns at different points in sustained phonation. All four subjects exhibited lower standard deviations at the beginning of sustained phonation for both shimmer and jitter, indicating that fatigue of laryngeal muscles or changes in respiratory support may affect stability as phonation is sustained. While CS, JL, and BS have a tendency to increase variability as they sustain phonation, MB exhibits the largest standard deviations in the middle of sustained phonation for both jitter and shimmer. This may be because the middle sample coincided with one of her fluctuating periods of increased amplitude and frequency.

Shimmer and jitter values measured from the 200 msec sample in the word "top" indicated greater stability than overall sustained phonation for all subjects except for JL. Interestingly, she exhibits the least stability in word production. While this single observation is insufficient to warrant interpretation, it is clear that variability exists across tasks as well as across subjects, lending further support to our argument that phonatory impairment in ALS is complex and idiosyncratic.

While direct relationships between acoustic patterns and perceptual phenomena should not be expected (Weismer, 1984), acoustic analysis of dysarthric speech can provide specific information relating to possible physiological contributions to perceptual characteristics. Such analysis may therefore aid in determining patterns of motor neuron degeneration, facilitate early detection of disease and provide a tool for monitoring disease progression. This in turn, can lead the clinician to better decisions regarding earlier and better patient education and counseling and lead to more efficacious treatment (e.g. compensation strategies to improve intelligibility).

Together the perceptual descriptions and acoustic data presented in this paper indicate great variability in phonatory performance and vocal quality among these four women. These differences may be due to several factors: the relative degree of spasticity vs. flaccidity in the laryngeal musculature; differential involvement of particular laryngeal and/or respiratory muscles; and differing respiratory/laryngeal strategies for compensatory performance. While acoustic data alone cannot determine the relative degree of spasticity vs. flaccidity or confirm specific patterns of individual laryngeal muscle involvement, these data, coupled with perceptual description, lead us to specific hypotheses which can then be tested with more invasive procedures such as videostroboscopy. With continued work in acoustic analysis, directed toward multiparametic assessment, these data may be used as a less invasive yet valid way to make inferences about laryngeal pathophysiology and its impact on the quality of phonation.

Kent et al., (1992) suggest that conclusions reached from group data do not necessarily apply to all individuals. Our data are consistent with this notion. Nonetheless, 70% of the dysarthria data reported in the literature during the last ten years is based on group data (Strand and Yorkston, in press). Further, 61% of those researchers reporting group data used the medical etiology (rather than characteristics of the dysarthria) to define the group. It is common for both clinicians and researchers to identify the dysarthria (and the phonatory performance) of ALS speakers as being a mixed flaccid/spastic "type". The description of perceptual characteristics and the acoustic data presented in this paper suggest that there is no one classification or diagnostic category that adequately represents such a grouping. Further, reporting group data in the absence of individual description of phonatory performance may be misleading. Within subject longitudinal research, examining the physiologic and acoustic changes accompanying perceptual changes during disease progression offers more promise for continuing research

Acknowledgement

This research was supported in part by NIDCD grant 1KO8 DC00043-01A1 and by NIDCD P60 DC00976.

References

Aronson, A. (1985). Clinical voice disorders: An interdisciplinary approach. New York: Thieme.

Aronson, A., Ramig, L., Winholtz, W., and Silber, S. (1992). Rapid voice tremor, or "flutter" in amyotrophic lateral sclerosis. Annals of Oto. Rhinol. Laryngol., 101, 511-518.

Bonduelle, M. (1975). Amyotrophic lateral sclerosis. In P. Vincken, G. Gruyn, and J. DeJong (Eds.), Handbook of clinical audiology, Vol. 22, pp. 281-338. New York: American Elsevier.

Carpenter, R.J., McDonald, T.J., and Howard, F.M. (1978). The otolaryngolic presentation of amyotrophic lateral sclerosis. Otolaryngology, 86, 479-484.

Carrow, E., Rivera, V., Mauldin, M., and Shamblin, L. (1974). Deviant speech characteristics in motor neuron disease. Archives of Otolaryngology, 100, 212-219.

Caruso, A. and Burton, E. (1987). Temporal acoustic measures of dysarthria associated with amyotrophic lateral sclerosis. Journal of Speech and Hearing Research, 30, 80-87.

Darley, F., Aronson, A. and Brown, J. (1969a). Differential diagnostic patterns of dysarthria. Journal of Speech and Hearing Research, 12, 246-270.

Darley, F., Aronson, A. and Brown, J. (1969b). Clusters of deviant speech dimensions in the dysarthrias. Journal of Speech and Hearing Research, 12. 462-497.

Darley, F.L., Aronson, A.E., and Brown, J.R. (1975). Motor speech disorders. Philadelphia, PA: W.B. Saunders.

Dworkin, J.P. and Hartman, D. (1979). Progressive speech deterioration and dysphagia in patients with amyotrophic lateral sclerosis: A case report. Arch. Phys. Med. Rehabil., 60, 423-425.

Kent, R., Kent, J., Weismer, G., Sufit, R., Rosenbek, J., Martin, R., and Brooks, B. (1990). Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. Journal of Speech and Hearing Disorders, 55, 721-728.

Kent, R., Sufit, R., Rosenbek, J., Kent, J., Weismer, G., Martin, R., and Brooks, B. (1991). Speech deterioration in amyotrophic lateral sclerosis: A case study. Journal of Speech and Hearing Research, 34, 1269-1275.

Kent, J., Kent, R., Rosenbek, J., Weismer, G., Martin, R., Sufit, R., and Brooks, B. (1992). Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. Journal of Speech and Hearing Research, 35, 723-733.

Kent, R.D., Kim, H-H., Weismer, G., Kent, J.F., Rosenbek, J., Sufit, R., Brooks, B., and Workinger, M. (in progress). Laryngeal dysfunction in neurological disease: Amyotrophic lateral sclerosis, Parkinson's disease, and stroke.

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

Milenkovich, P. (1992). CSpeech Ver. 40. University of Wisconsin-Madison, Department of Electrical Engineering.

Mulder, D. (Ed.) (1980). The diagnosis and treatment of ALS. Boston: Houghton Mifflin.

Ramig, L.O. and Scherer, R.C. (1992). Speech therapy for neurologic disorders of the larynx. In Blitzer, A., Brin, M., Sasaki, C.T., Fahn, S., and Harris, K. (Eds.), Neurologic Disorders of the Larynx. New York: Thieme Medical Publishers.

Ramig, L., Scherer, R., Titze, I., and Ringel, S. (1988). Acoustic analysis of voices of patients with neurological disease: A rationale and preliminary data. Annals of Otology, Rhinology and Laryngology, 97, 164-172.

Ramig, L., Scherer, R., Klasner, E., Titze, I., and Horii, Y. (1990). Acoustic analysis of voice in amyotrophic lateral sclerosis: A longitudinal case study. Journal of Speech and Hearing Disorders, 55, 2-14.

Strand, E. and Yorkston, K. (in press). Description and classification of dysarthric individuals: A ten year review. In Till, J. (Ed.), Motor Speech Disorders, Advances in Assessment and Treatment. Baltimore: Paul H. Brookes, Inc.

Tanden, R. and Bradley, W. (1985). Amyotrophic lateral sclerosis: Part 1. Clinical features, pathology and ethical issues in management. Annals of Neurology, 18, 271-280.

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

Titze, I.R., Horii, Y., and Scherer, R.C. (1987). Some technical considerations in voice perturbation measurements. Journal of Speech and Hearing Research, 30, 252-260.

Weismer, G. (1984). Acoustic descriptions of dysarthric speech: Perceptual correlates and physiological inferences. In J.C. Rosenbek (Ed.), Current view of dysarthria: Nature, assessment and treatment; seminars in speech and language, pp. 293-314. New York: Thieme-Stratton.

Weismer, G., Martin, R., Kent, R., and Kent, J. (1992). Formant trajectory characteristics of males with amyotrophic lateral sclerosis. Journal of the Acoustical Society of America, 91, 1085-1098.

Yorkston, K., Beukelman, D., and Traynor, C. (1984). Computerized Assessment of Intelligibility of Dysarthric Speech. Tigerd, OR: C.C. Publications.

Yorkston, K., Strand, E., Miller, R., Hillel, A., and Smith, K. (1993). Speech deterioration in amyotrophic lateral sclerosis: Implications for the timing of intervention. Journal of Medical Speech Pathology, Vol. 1; 1, 35-46.

NCVS Status and Progress Report - 4
June 1993, 169-176

Simultaneous Functional Laryngeal Stridor and Functional Aphonia in an Adolescent

Marshall E. Smith, M.D.
Department of Otolaryngology, The Children's Hospital, Denver, Colorado
Department of Otolaryngology/Head and Neck Surgery, The University of Colorado Health Sciences Center
The Recording and Research Center, The Denver Center for the Performing Arts

Kathe Perez Darby, M.A.
Department of Otolaryngology/Head and Neck Surgery, The University of Colorado Health Sciences Center

Kevin Kirchner, M.D.
Division of Pulmonary Medicine, The Children's Hospital, Denver, Colorado

Florence B. Blager, Ph.D.
Department of Otolaryngology/Head and Neck Surgery, The University of Colorado Health Sciences Center
National Jewish Center for Immunology and Respiratory Medicine, Denver, Colorado
Recording and Research Center, The Denver Center for the Performing Arts

Abstract
A case is reported of a sixteen year old female who developed simultaneously both functional laryngeal stridor and functional aphonia. After appropriate diagnosis was made, the stridor resolved promptly with behavioral intervention. The aphonia, however, persisted for five months. Speech therapy was also eventually successful in restoring normal voice. Therapy was aided by temporary voice restoration via unilateral recurrent laryngeal nerve block with lidocaine. Sources of contributing associated psychological and physiological stress in both functional disorders were identified. The case illustrates application of a contemporary clinical framework for evaluation and treatment of voice disorders that is pertinent to the pediatric population.

Functional and psychogenic phonatory disturbances are well recognized.[1,2] In addition, disruptions in the laryngeal function of respiration without manifested organic causes have been described and reported on in detail in the last two decades.[3-20] This report addresses these clinical entities which both affect laryngeal function and describes their simultaneous presentation in an adolescent patient.

First described as the misnomer of "Munchausen's stridor"[3], a plethora of terms have been applied to describe functional laryngeal stridor (FLS). Commonly associated with underlying or past psychological stress but not a symptom of malingering, it is felt in many cases to be a conversion disorder.[9,11,15] The prototypical presentation is of a young to middle aged adult female who develops acute respiratory distress and dyspnea. The symptoms are felt to be asthma attacks. The patient undergoes an extensive regimen of pharmacotherapy without relief. Acute attacks prompt emergency room visits; extreme cases result in a tracheotomy. Oximetry studies and blood gases are normal, despite the severity of symptoms. Laryngoscopic findings reveal adduction of the vocal folds on inspiration during spells of distress, with normal laryngeal respiratory function between spells. This clinical disorder has also been reported in children.[4,5,9,14,17,18,21-28] Of nineteen reported pediatric cases, fifteen were females from ages ten through nineteen.

Functional aphonia (FA) is a well described voice disorder, characterized by a whispered voice during conversation and phonation without the finding of laryngeal pathology. It is uncommon, but certainly not rare. In a series of one hundred consecutive patients with voice complaints, four were diagnosed with FA, including one child.[29] In a series of 416 patients with functional voice disorders, 7% had FA.[30] Reports of FA corroborate that 1) it is much more common in females than males, 2) it presents in ages from childhood to the elderly, and 3) subjects may exhibit a sense of detachment or unconcern about their voice problem.[1,2,30,31] FA is less commonly reported in children. A recent review of the English literature encountered 12 reported cases of patients under 16 years of age.[32] A recent survey[33] added 14 more pediatric cases, twelve of which were females.

The association of functional disorders of the larynx affecting both respiration and phonation might be expected but has not been generally reported. Martin et al[15] observed that some patients with FLS who have co-existing asthma developed voice problems, presumably related to impaired breath support. This association has not been documented or studied in detail.

We report a case of an adolescent female who developed simultaneously both FLS and FA. The case is unique in several other respects, the persistence of the aphonia for five months and the finding of co-existing reflux laryngitis. Another notable aspect included the use of recurrent laryngeal nerve lidocaine block to temporarily relieve the aphonia and supplement behavioral speech therapy.

Case Report

A sixteen year old female presented to the The Children's Hospital emergency room with cough, hoarseness, and increasing shortness of breath for three days. On examination she was found to have a hoarse, barely audible voice, poor air exchange with faint wheezing, and the absence of stridor. The patient had been diagnosed with mild bouts of asthma in the past. A diagnosis of severe asthma and tracheitis/bronchitis was made and she was treated with aggressive nebulization treatments, steroids, and antibiotics. After clinical improvement she was discharged the following day. Her voice also improved after this episode. Five weeks later the symptoms recurred and she re-presented to the emergency room, again with symptoms of increasing shortness of breath and aphonia/dysphonia. A presumptive diagnosis of spasmodic croup was made. A pulmonary medicine consultation was obtained and on further evaluation, the patient's "wheezing" was localized to the neck. Fiberoptic laryngoscopy revealed inspiratory adduction of the vocal folds and obstruction at the level of the glottis, as well as excessive expiratory laryngeal adduction. She was diagnosed with FLS. After two days of observation and conservative treatment the patient was discharged in no respiratory distress, but still with a whispering voice.

The patient was referred to speech pathology at National Jewish Hospital where a large population of patients with FLS are seen. A social history revealed that the patient's parents divorced when she was two years old. A stepfather lived in the home most of the patient's childhood. The patient and her mother reported that during the patient's childhood the stepfather frequently criticized her and "told her she was stupid". Two past episodes of physical abuse involving the stepfather and patient were reported, but no history of sexual abuse was elicited. The stepfather left the home one year ago, after these abusive episodes. The patient reported that life at home improved for her after the stepfather left. Psychological evaluation and counseling was offered but declined by the patient's mother.

She underwent a series of behavioral exercises directed toward improvement of breathing, prevention of episodes of acute respiratory distress, and treatment of hyperfunctional voice disorder. No further bouts of acute shortness of breath developed but the whispering aphonia persisted.

The patient's aphonic (whispering) voice continued for four months. She was referred to the otolaryngology service at The Children's Hospital for another opinion regarding her voice. The new school year was soon to begin and the patient and her mother did not want her to begin school without a voice. In addition, the patient also related a present and prior history of gastrointestinal upset, took antacid tablets on a daily basis, and had undergone an upper GI radiographic study two years previously, reported normal.

Examination revealed an adolescent female speaking in a whispering, aphonic voice. No voicing could be elicited on various maneuvers. Fiberoptic laryngoscopy revealed that during phonation the true and false folds of the larynx were tightly adducted anteriorly with a small open posterior glottic chink. Laryngeal activity was normal during respiration, cough, and valsalva. Mild erythema of the right vocal process of the arytenoid and posterior glottic mucosa was observed, as well as pooling of secretions in the posterior larynx and pyriform sinuses.

For evaluation of reflux laryngitis, the patient underwent 24 hour two-channel ambulatory pH probe monitoring. This study was markedly positive for gastroesophageal reflux. An anti-reflux behavioral regimen and H_2 blocker therapy were initiated.

The patient was studied in the voice laboratory. As a diagnostic probe and to initiate therapy, a unilateral recurrent laryngeal nerve block with lidocaine was performed. This prevented the larynx from assuming a "whispering posture" and allowed the patient to phonate. With the speech pathologist in attendance, the videocamera and monitor connected to the fiberscope were used as a biofeedback tool to instruct the patient on phonation without hyperadduction.

The patient underwent initial therapy while the lidocaine effect was still present to apply voice relaxation techniques. This gave the patient confidence that the therapy techniques could work and reassured her mother that she did "have a voice" in spite of its absence for five months. The aphonia recurred after the lidocaine wore off. After two more sessions of voice therapy the aphonia abated and normal voicing returned. The patient's normal voice has remained without relapse on six month followup.

Discussion

We prefer to use the term "functional" in the description of laryngeal stridor and aphonia for several reasons. It appropriately targets the abnormal behavior and its remediation (speech therapy is the mainstay of effective treatment for both functional laryngeal stridor and functional aphonia) without neglecting associated psychological relationships. Labels such as "vocal cord dysfunction" and "laryngeal dyskinesia" inappropriately imply an organic or neurologic abnormality of the larynx, and they do not describe the problem in laryngeal function (eg. respiration, phonation, swallowing). Likewise, though both the disorders of functional laryngeal stridor and functional aphonia have been called "psychogenic", this label may be accurate but less desirable. A "psychogenic" connection between the current symptoms and either particular recent events or chronic problems is not made by the patient.[15] Psychological findings and personality structures of these patients are quite heterogeneous.[15,34-36] Direct confrontation with the patient and family about psychologic etiology of symptoms for which primary and/or secondary gain often exists is met with resistance.[15] Therefore, a behavioral approach which targets correction of the abnormal laryngeal behavior (ie. restore appropriate function) is favored,[15,30,33] not neglecting in the context of therapy a sensitive consideration of the psychiatric symptoms.[15,35]

Morrison[37] has proposed a model for evaluation of voice problems that fits well into this context in the approach to both functional laryngeal stridor and functional aphonia. Dysphonias are described as a symptom complex, not as a single diagnostic entity. The evaluation of dysphonia involves the identification of all the factors contributing to the voice problem. These may include behavioral, psychological, medical, neurologic, and other causes. Weightings may be given to the relative contribution of these factors, but the key aspect is the recognition that multiple interrelated causes frequently contribute to the resulting voice disorder.[30] A patient's psychological and behavioral response may contribute as much or more to the voice problem than the instigating medical "disease".

This model applied well to our case. The patient had underlying psychological stresses related to a history of physical and emotional abuse. She exhibited "alexithymic" traits, commonly seen patients with either functional laryngeal stridor or functional aphonia,[35,38] relating to inadequacies in the verbal expression of feelings such as anger, fear, or sadness. Two medical stresses to the larynx were identified: an acute stress related to the initial upper respiratory infection, and the chronic stress of occult gastroesophageal reflux (GER) laryngitis. GER has been associated with many laryngeal disorders, including chronic laryngitis.[39] The patient's GER may have been coincidental; however, we speculate that it may have contributed to the problem in making the larynx vulnerable to subsequent stresses.

The persistence of this patient's aphonia symptoms is uncommon though not unusual. In two of fourteen pediatric aphonics reported by Harris and Richards[33] symptoms persisted beyond five months. In contrast to the usual prompt response of functional aphonia patients in one therapy session, Harris and Richards noted a gradual response in children that took several treatment sessions, 50% requiring more than two months of therapy.[33] Prolonged behavioral treatment is often required in adult patients with functional dysphonia (not aphonia).[30,31] It was in this context of prolonged persistence of our patient's symptoms that a recurrent laryngeal nerve (RLN) lidocaine block was considered. RLN lidocaine block has been used as a therapeutic trial in patients with spasmodic dysphonia to assess potential for RLN section.[40] In this case the procedure 1) reassured and gave confidence to the patient and her mother that, after five months of aphonia, she did have a voice, and 2) provided the opportunity for the patient to re-learn the "feel of phonation" after a prolonged period of hyperfunctional laryngeal posturing. Its successful use in this case does not mean that this intervention is advocated or necessary in every case of functional aphonia; indeed, most cases are successfully treated with speech therapy alone. It is, however, an additional tool that may be available for selected or difficult cases.

Conclusions

1. Functional laryngeal stridor and functional aphonia presented simultaneously in an adolescent female, the aphonia lasted for five months.

2. Recognition of the various psychological, behavioral, and medical factors that may contribute to functional laryngeal disorders provides a conceptual framework for addressing treatment of this symptom complex. Evaluation and treatment require a collaborative effort between physician and speech pathologist.

3. Adolescents with functional aphonia may have prolonged symptoms and require more prolonged treatment.

4. In prolonged functional aphonia recurrent laryngeal nerve lidocaine block may be helpful to reassure the patient that a voice is present and to assist successful voice therapy.

Acknowledgement

This work was supported, in part, by grant P60 DC00976 from the National Institute on Deafness and other Communicative Disorders (NIDCD).

References

1.Aronson AE, Peterson HW, Litin EM: Voice symptomatology in functional dysphonia and aphonia. J Speech Hear Disord 29:367-380, 1964

2.Brodnitz FS: Functional aphonia. Ann Otol Rhinol Laryngol 78:1244-1258, 1969

3.Patterson R, Schatz M, Horton M: Munchausen's stridor: non-organic laryngeal obstruction. Clin Allergy 4:307-310, 1974

4.Rogers JH, Stell PM: Paradoxical movement of the vocal cords as a cause of stridor. J Laryngol Otol 92:157-8, 1978

5.Cormier YF, Camus P, Desmeules MJ: Non-organic acute upper airway obstruction: description and a diagnostic approach. Am Rev Respir Dis 121:147-150, 1981

6.Appleblatt NH, Baker SR: Functional upper airway obstruction: a new syndrome. Arch Otolaryngol 107:305-6, 1981

7.Kellman RM, Leopold DA: Paradoxical vocal cord motion: an important cause of stridor. Laryngoscope 92:58-60, 1982

8.Downing ET, Braman SS, Fox MJ, et al: Factitious asthma: physiological approach to diagnosis. JAMA 248:2878-2881, 1982

9.Christopher KL, Wood RP, Eckert RC, et al: Vocal-cord dysfunction presenting as asthma. NEJM 308:1566-70, 1983

10.Rodenstein DO, Francis C, Stanescu DC: Emotional laryngeal wheezing: a new syndrome. Am Rev Respir Dis 127:354-356, 1983

11.Starkman MN, Appelblatt NH: Functional upper airway obstruction: a possible somatization disorder. Psychosomatics 25:327-333, 1984

12.Myears DW, Martin RJ, Eckert RC, et al: Functional versus organic vocal cord paralysis: rapid diagnosis and decannulation. Laryngoscope 95:1235-1237, 1985

13.Ramirez-R J, Leon I, Rivera LM: Episodic laryngeal dyskinesia: clinical and psychiatric characterization. Chest 90:716-721, 1986

14.Kivity S, Bibi H, Schwarz Y, et al: Variable vocal cord dysfunction presenting as wheezing and exercise-induced asthma. J Asthma 23:241-244, 1986

15.Martin RJ, Blager FB, Gay ML, et al: Paradoxic vocal cord motion in presumed asthmatics. Sem Respir Med 8:332-337, 1987

16.Brown TM, Merritt WD, Evans DL: Psychogenic vocal cord dysfunction masquerading as asthma. J Nerv Ment Dis 176:308-310, 1988

17.Skinner DW, Bradley PJ: Psychogenic stridor. J Laryngol Otol 103:383-385, 1989

18.Snyder HS, Weiss E: Hysterical stridor: a benign cuase of upper airway obstruction. Ann Emer Med 18:991-994, 1989

19.Sim TC, McClean SP, Lee JL, et al: Functional laryngeal obstruction: a somatization disorder. Am J Med 88:293-295, 1990

20.Tousignant G, Kleiman SJ: Functional stridor diagnosed by the anaesthetist. Can J Anaesth 39:286-289, 1992

21.Rogers JH: Functional inspiratory stridor in children. J Laryngol Otol 94:669-670, 1980

22.Smith MS: Acute psychogenic stridor in an adolescent athlete treated with hypnosis. Pediatrics 72:247-248, 1983

23.Kattan M, Ben-Zvi Z: Stridor caused by vocal cord malfunction associated with emotional factors. Clin Pediatrics 24:158-160, 1985

24.Barnes SD, Grob CS, Lachman BS, et al: Psychogenic upper airway obstruction presenting as refractory wheezing. J Pediatrics 109:1067-1070, 1986

25.Kissoon N, Kronick JB, Frewen TC: Psychogenic upper airway obstruction. Pediatrics 81:714-717, 1988

26.Logvinoff MM, Lau KY, Weinstein DB, et al: Episodic stridor in a child secondary to vocal cord dysfunction. Ped Pulmonol 9:46-48, 1990

27.Ophir D, Katz Y, Tavori I, et al: Functional upper airway obstruction in adolescents. Arch Otolaryngol Head Neck Surg 116:1208-1209, 1990

28.Rusakow LS, Blager FB, Barkin RC, et al: Acute respiratory distress due to vocal cord dysfunction in cystic fibrosis. J Asthma 28:443-336, 1991

29.Koufman JA, Isaacson G: The spectrum of vocal dysfunction. Otolaryngol Clin NA 24:985-988, 1991

30.Koufman JA, Blalock PD: Functional voice disorders. Otolaryngol Clin NA 24:1059-1073, 1991

31.Bridger MWM, Epstein R: Functional voice disorders: a review of 109 patients. J Laryngol Otol 97:1145-1148, 1983

32.Froese AP, Sims P: Functional dysphonia in adolescence: two case reports. Can J Psychiatry 32:389-392, 1987

33.Harris C, Richards C: Functional aphonia in young people. J Laryngol Otol 106:610-612, 1992

34.Kinzl J, Biebl W, Rauchegger H: Functional aphonia: a conversion symptom as defensive mechanism against anxiety. Psychther Psychosom 49:31-36, 1988

35.Kinzl J, Biebl W, Rauchegger H: Functional aphonia: psychosomatic aspects of diagnosis and therapy. Folia Phoniatr 40:131-137, 1988

36.Aronson AE, Peterson HW, Litin EM: Psychiatric symptomatology in functional dysphonia and aphonia. J Speech Hear Disord 31:115-127, 1966

37.Morrison MD: Psychiatric and neurologic interaction in dysphonic patients. Presented at the Fifth Annual Pacific Voice Conference, San Francisco, October, 1992

38.Freedman MR, Rosenberg SJ, Schmaling KB: Childhood sexual abuse in patients with paradoxical vocal cord dysfunction. J Nerv Ment Dis 179:295-298, 1991

39.Koufman JA: The otolaryngologic manifestations of gastroesophageal reflux disease. In Richter JE (ed): Ambulatory Esophageal pH Monitoring: Practical Approach and Clinical Applications, Igaku-Shoin, New York, Tokyo, 1991, pp 129-149

40.Dedo HH: Recurrent laryngeal nerve section for spastic dysphonia. Ann Otol Rhinol Laryngol 85:451-459, 1976

NCVS Status and Progress Report - 4
June 1993, 177-193

Analysis of Vocal Disorders with Methods from Nonlinear Dynamics

Hanspeter Herzel, Ph.D.
Fachbereich Physik, Humboldt - Universität Berlin
David Berry, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Ingo Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Marwa Saleh, M.D.
Faculty of Medicine, Ain Shams University, Egypt

Abstract

Several authors have recently demonstrated the intimate relationship between nonlinear dynamics and observations in vocal fold vibration (Mende, Herzel & Wermke, 1990; Herzel & Wendler, 1991; Titze, Baken & Herzel, 1992). The aim of this paper is to analyze vocal disorders from a nonlinear dynamics point of view. Basic concepts and analysis techniques from nonlinear dynamics are reviewed and related to voice. The voices of several patients with vocal disorders are analyzed using traditional voice analysis techniques and methods from nonlinear dynamics. The two methods are shown to complement each other in many ways. Likely physiological mechanisms of the observed nonlinear phenomena are presented, and it is shown how much of the terminology in the literature describing rough voice can be unified within the framework of nonlinear dynamics.

Introduction

Recently several authors have underscored the importance of using a nonlinear dynamics approach to better understand vocal fold vibration. It has been argued that several observations in newborn cries (Mende, Herzel, & Wermke, 1990), children's noncry vocalizations (Robb & Saxman, 1988), vocal fry phonation (Titze, 1989; Herzel, 1992), and pathological voices (Herzel &

Wendler, 1991; Titze, Baken, & Herzel, 1992) are intimately related to the concept of bifurcations and deterministic chaos.

We begin by reviewing several of the observations in the literature that bear a close resemblance to nonlinear dynamics. First, "diplophonia" and "dicrotic patterns of vocal fold vibrations" have been observed for several centuries (see Cooper, 1989). Quite similar findings have been reported for several decades in newborn cry studies based on sound-spectrographic investigations (Sirviö & Michelsson, 1976). Period-doublings and biphonation also appear in noncry vocalizations of young children (Robb & Saxmann, 1988). Occasionally, irregularities are also observed in normal healthy voices (Moore & von Leden, 1958; Dolansky & Tjernlund, 1968; Klatt & Klatt, 1990). In particular, some vocal fry phonation is characterized by glottal pulses of alternating amplitudes or by irregular trains of pulses (Hollien & Michel, 1968; Scherer, 1989).

Several authors report the appearance of subharmonics in cases of unilateral or localized vocal fold lesions. Periodicities of 2-12 pitch periods are found by Koike (1969) for patients with laryngeal neoplasms. For a female speaker with papilloma, period-doubling was termed "zig-zag sections in the fundamental frequency curve" (Askenfelt & Hammarberg, 1986). In a recent paper, subharmonics have been identified for patients with cysts and nodules (Remacle & Trigoux, 1991).

Clear indications of bifurcations and chaos can be found also in studies on unilateral laryngeal nerve paralysis (Ishizaka & Isshiki, 1976; Hammarberg, Fritzell, Gauffin & Sundberg, 1986; Smith, Berke, Gerratt & Kreiman, 1992). In these cases, the desynchronization of the right and left vocal fold is presumably the origin of bifurcations.

Most neurobiological disorders of the larynx lead also to phonatory instabilities (Ramig & Scherer, 1991). Diplophonia has been reported for patients with Shy-Drager Syndrome (Ludlow, Coulter & Gentages, 1983) and spasmodic dysphonia (Titze et al. 1992). In a study by Wieser (1981) octave jumps, two pitches and noise segments are already considered as characteristics of spastic dysphonia. Sudden drops in fundamental frequency of approximately one octave are found in patients with Huntington's Disease and also for individuals at risk for this disease (Ramig, Scherer, Titze & Ringel, 1988).

The appearance of period-doublings and other nonlinear phenomena in such a wide variety of voice signals suggests these observations may need to be analyzed from a rather general point of view, i.e. as manifestations of nonlinear dynamics. The aim of this paper is to provide further evidence of bifurcations and attractors in voice signals, to test the applicability of several quantitative measures and to discuss some questions of terminology.

A framework for analysis is provided by first sketching the basic concept of nonlinear dynamics. Attractor types are classified and the most important bifurcations are discussed. Following this general discussion, several records from pathological voices are analyzed from the point of view of nonlinear dynamics. It is shown that the described phenomena occur in various diseases, e.g. nodules, cancer, paralysis and hyperfunctional dysphonia. It is also shown that much information about bifurcations can be gained from a careful inspection of the time-series, spectrograms and particularly the fundamental frequency and amplitude contours. Moreover, phase portraits and Poincaré sections turn out to be useful in some cases.

In the final discussion, possible physiological mechanisms of nonlinear phenomena in voice are listed. Also, an attempt will be made to clarify and standardize the various terms which have been introduced in the literature.

Basic Concept of Nonlinear Dynamics

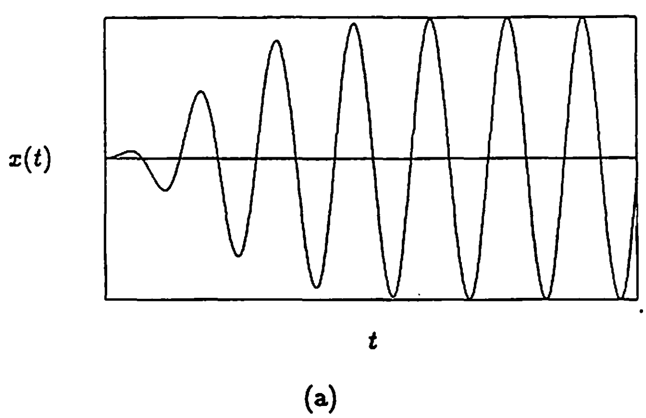
In this section, we give only a brief summary of the main ideas prerequisite to an understanding of the results in the next sections. A more detailed introduction to the theory of nonlinear dynamics as applied to phonation can be found in a chapter of a recent book (Titze, 1992). Also, the texts by Bergé, Pomeau, & Vidal (1986), Holden (1986) and Glass & Mackey (1988) are recommended.

Phase Space

Conventional time series analysis is carried out either in the time domain or the frequency domain. Nonlinear dynamics is based on an embedding of the time series in a "phase space". The most convenient way to embed the series is to introduce delay-coordinates:

$$\underline{x}(t) = \{x(t), x(t+\tau), \ldots, x(t+(m-1)\tau)\} \quad . \tag{1}$$

With these coordinates, a scalar time series generates an orbit in an m-dimensional phase space. Obviously, the choice of the delay time τ and of the embedding dimension m introduces some arbitrariness. However, several features of the dynamics (e.g., attractor dimensions or Lyapunov exponents) are essentially independent of the specific embedding parameters. Figure 1 illustrates the embedding of a time series into a three-dimensional phase space ($m = 3$).



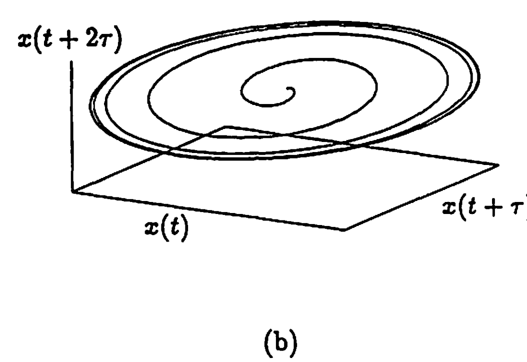


Figure 1. An example of the embedding of a time series (a) into phase space (b). In this case, the resultant attractor in (b) is a limit cycle.

Attractors

If the external parameters of a dissipative dynamical system are held constant, the orbit in phase space approaches an asymptotic regime after some initial transient. The geometrical object in phase space corresponding to the asymptotic behavior is termed an attractor. In Figure 1, for example, the attractor is a closed curve, or a "limit cycle". The spiraling out in this figure corresponds to the initial transient. Attractors can be classified as follows:

a) stable stationary states
b) limit cycles (periodic oscillations as in Figure 1)

c) tori (superposition of two or more oscillations with incommensurable frequencies)

d) chaotic attractors (nonperiodic behavior)

It has been recognized in the last two decades that complicated nonperiodic oscillations ("chaos") can occur in nonlinear systems without any random perturbations. Thus, the discovery of "deterministic chaos" provides a new approach for understanding irregular observations. The aim of this paper is to study vocal disorders from this point of view.

Bifurcations

Up to now we have assumed fixed external conditions. If parameters of the dynamical system are varied, qualitative changes of the attractors might appear. Such transitions are termed "bifurcations".

An example is the onset of self-sustained oscillations due to increasing subglottal pressure (see e.g. Titze, 1988). Such a transition from a steady state (resting fold) to periodic oscillations (phonation) is a manifestation of a "Hopf bifurcation".

For simplicity, we assume here that normal phonation corresponds to a limit cycle and we do not discuss at this point the omnipresent deviations from periodicity of a normal healthy voice (jitter and shimmer on the order of a percent). Then, the question arises whether or not observed sudden transitions to other regimes can be interpreted in terms of bifurcations of a limit cycle.

The most important of such bifurcations are period-doublings and secondary Hopf bifurcations. In the first case, a new limit cycle with roughly double the original period becomes stable, i.e. in the time-domain we may observe an alternation of large and small amplitudes (or pitch periods). In the spectral domain one would observe the appearance of subharmonics of the original pitch.

A secondary Hopf bifurcation corresponds to a modulation of the signal with another independent frequency, i.e. a torus appears in phase space. It will be shown in this paper that these bifurcations can be found in pathological voices.

Another frequently observed bifurcation is the sudden jump from the original limit cycle to another limit cycle with another period and amplitude. Different attractors may coexist in nonlinear systems and, therefore, even extremely tiny changes of parameters (e.g., muscle tension) may lead to abrupt jumps to other regimes.

The theory of nonlinear dynamics predicts that period-doublings, bifurcations, tori and coexisting limit cycles are often accompanied by "deterministic chaos" at adjacent parameter values. Thus, the detection of bifurcations in voice signals supports the idea that some of the irregularities in vocal disorders are manifestations of chaotic dynamics.

Voice Samples

Our study is based on the analysis of 95 dysphonic patients with various pathologies. The patients can be grouped as follows (Saleh, 1991):

A) Dysphonia with minimal associated pathological lesions

1)	Nodules	14 patients
2)	Polyps	8 patients
3)	Cysts	5 patients
4)	Reinke's edema	8 patients

B) Organic dysphonia

1)	Paralysis	16 patients
2)	Neoplastic dysphonia	8 patients

C) Functional dysphonia

1)	Phonasthenia	11 patients
2)	Hypofunctional dysphonia	4 patients
3)	Hyperfunctional dysphonia	11 patients
4)	Hyperfunctional childhood dysphonia	6 patients
5)	Mutational voice disorders	4 patients

For each patient, three utterances of the sustained vowel "a" have been digitally recorded with a sampling rate of 20 kHz and a resolution of 16 bits (Saleh, 1991).

The aim of our study is not a detailed and comprehensive analysis of all these cases but a first estimation of the relevance of the discussed nonlinear phenomena. Moreover, we want to test what statistical measures are appropriate for an analysis of bifurcations and chaos.

In about one-fourth of the patients, we have found indications of bifurcations or attractors. These phenomena have been found in patients from all three groups discussed above. In a previous study (Herzel & Wendler, 1991) bifurcations and chaos were detected in about one-half of the patients. In that case, all patients had very hoarse voices. In the present analysis, the voice quality ranged from normal to severely hoarse; therefore, the frequency of occurrence of nonlinear phenomena was somewhat less in this data set.

Methods

In this section, we briefly describe the techniques used to detect and quantify the nonlinear phenomena. The applicability of the methods is exemplified using characteristic recordings of our samples of pathological voices.

First of all, perceptual evaluation provides clues as to whether or not nonlinear phenomena are present in the voice signals. For example, episodes of subharmonic vocalization are perceived as intermittent roughness. If subharmonic components or modulation frequencies are sufficiently low (say about 50 Hz) the acoustic signal is often perceived as vocal fry phonation. Although vocal fry is considered linguistically normal (Hollien & Michel, 1968), fry is often symptomatic in voice pathology (Scherer, 1989). Indeed, episodes of low-frequency modulations have been found in several of our recordings.

According to the theory of psychoacoustics (Zwicker & Fastl, 1990), modulations around 70 Hz lead to the perception of roughness. Consequently, our observations of subharmonic vocalizations, low-frequency modulations and chaos are intimately related to rough voice quality since these phenomena typically lead to an increased amount of low-frequency components.

Spectrographic Displays

Besides listening, the careful inspection of acoustic waveforms, EGG signals and spectrograms is helpful for the detection of bifurcations and chaos. Subharmonics, biphonation (i.e. two independent frequencies) and chaotic segments have been effectively identified using narrow-band spectrograms (Sirviö & Michelsson, 1976; Kelman, 1981; Robb & Saxmann, 1988; Mende et al. 1990).

Figure 2 shows abrupt transitions to subharmonic regimes spectrographically, which can be interpreted as bifurcations. The patient (with a hyperfunctional childhood disorder) has a characteristic fundamental frequency around 300 Hz; however, a strong subharmonic corresponding to a period-tripling appears between 300 and 800 ms. While the relative energy of the signal around 100 Hz is so low that it cannot be seen on the spectrogram, the spacing of the harmonics reveals that the subharmonic is, indeed, present. Between 1200 and 1800 ms, the subharmonic appears again. Moreover, low-frequency modulations around 20 Hz are noticeable.

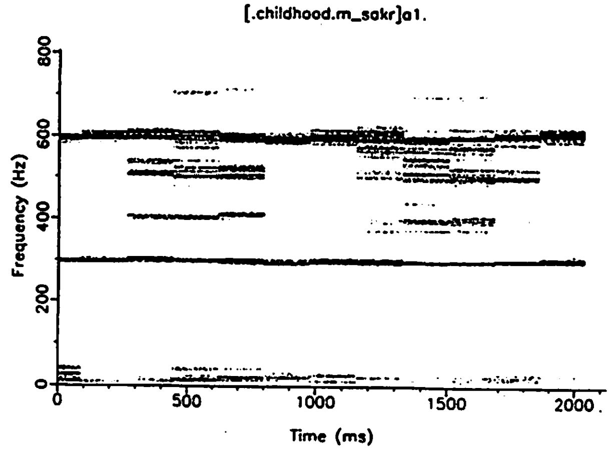


Figure 2. Narrow-band computer spectrogram for a patient with a hyperfunctional childhood disorder. Abrupt transitions to subharmonics are visible. Note, especially, the subharmonic corresponding to period-tripling from 300-800 ms and from 1200-1800 ms. In several instances, a low-frequency modulation around 20 Hz also appears.

Waveform Displays

An acoustic waveform displaying sudden transitions, also interpreted as bifurcations, is shown in Figure 3. The figure displays 1000 ms of an utterance from a patient with unilateral recurrent nerve paralysis. Parts (a)-(e) are serial segments in time. Note that (b) and (d) show amplitude modulations, whereas (a), (c) and (d) look relatively normal.

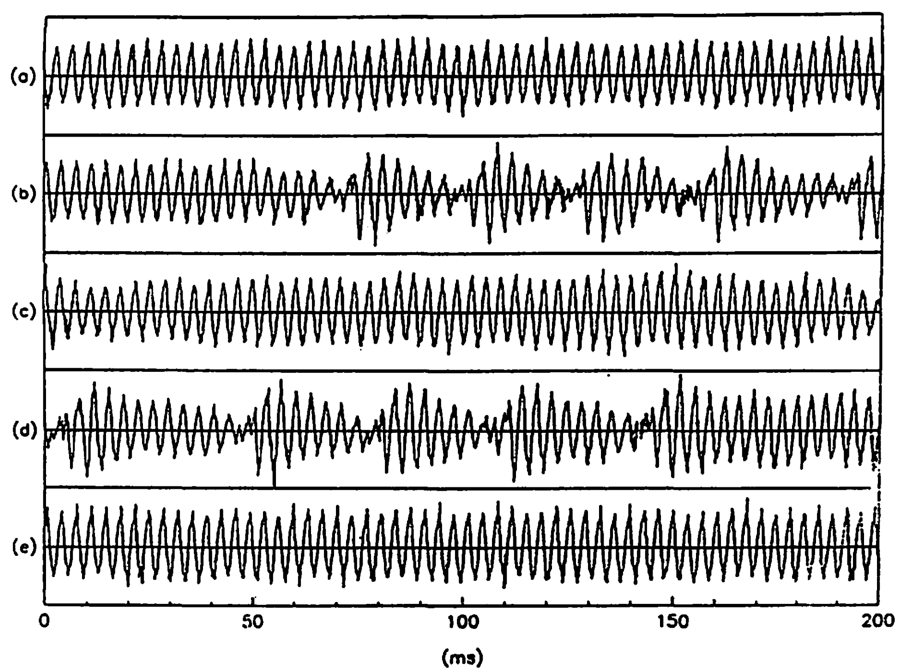


Figure 3. A 1000 ms utterance (with 200 ms per line) from a patient with laryngeal paralysis. Bifurcations to low-frequency modulations are shown in lines (b) and (d).

F_o and Amplitude Contours

Traditional perturbation analysis, based on the detection of fundamental frequency and amplitude, can be exploited for the characterization of nonlinear phenomena. We applied a software package called GLIMPES for the extraction of F_o and amplitude. In the first step, the acoustic (or EGG) signal was low-pass filtered with a cut-off frequency slightly above F_0 (typically at 200 Hz for a male voice). The zero-crossings of the filtered signal were used to locate the approximate period boundaries. Based on these markers, the precise estimates for F_0 were found from the unfiltered speech signal using a least mean squares or waveform matching technique. Details concerning the algorithms can be found elsewhere (Titze & Liang, in review). Since the cutoff frequency of the low-pass filter and the allowed range of the fundamental frequencies can be varied interactively, the algorithm provides the necessary flexibility and user-control for analyzing even severely hoarse voices. After extraction of the periods, the differences between the maximum and minimum value within each cycle are computed and referred to as amplitudes.

Figure 4 shows representative examples of the contours (left panels) and their corresponding autocorrelation functions (right panels) derived from patients with nodules and paralysis. In Figure 4a, a period-doubling is clearly visible, whereas Figure 4b shows a low-frequency modulation around 30 Hz, which was audible as vocal fry.

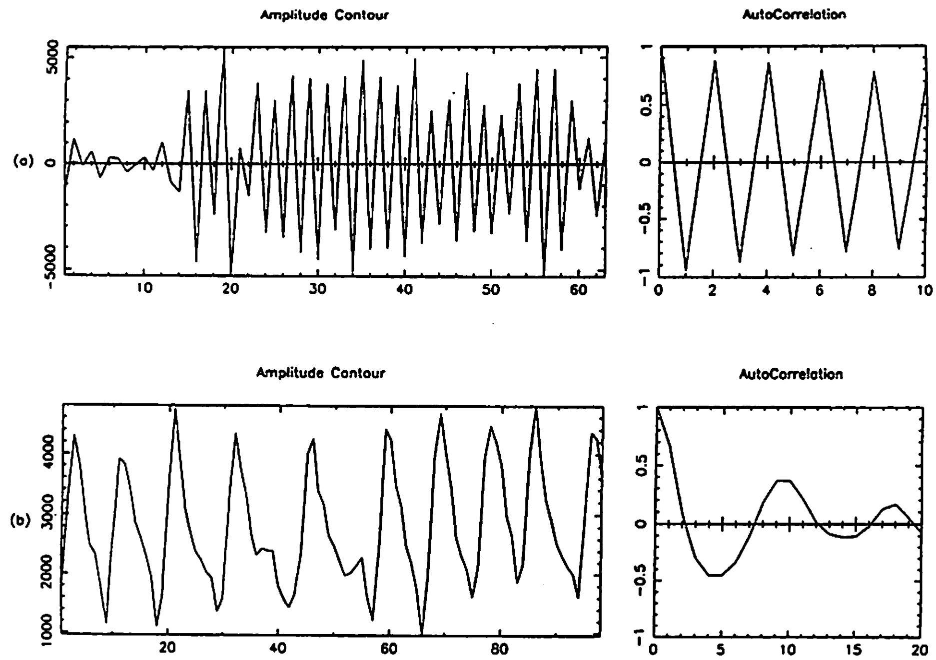


Figure 4. Amplitude contours (and corresponding autocorrelation functions) clearly illustrating (a) a period-doubling for a patient with nodules and (b) a low-frequency modulation around 30 Hz for a patient with paralytic dysphonia. The linear trend remover described in the text was applied to the contour in (a). Figure 4b is derived from another utterance of the speaker in Figure 3.

From the F_0 and amplitude contours, various perturbation measures can be derived (Pinto & Titze, 1990). However, traditional jitter and shimmer analysis may have limited value for our purposes. In particular, some average perturbation measures cannot distinguish between short-correlated turbulent noise and modulations of the signal over several cycles. In order to detect such modulations, which are characteristic for subharmonic vocalization and biphonation, the autocorrelation function of the F_0 (or amplitude) contour has proven useful. Autocorrelation functions of the contours have been used by others (Koike, 1969; Imaizumi, 1986) to describe rough voices.

If the cut-off frequency of the low-pass filter for F_0 detection is above the normal fundamental frequency, period markers are set according to this characteristic frequency even if bifurcations do occur. Consequently, period-doublings appear as alternating periods (or amplitudes), and modulations are seen as lower F_0 (amplitude) fluctuations.

Since voice signals exhibit common nonstationarities such as long-term trends and drifts, analysis of the F_0 (amplitude) contour is mostly based on second order perturbation functions:

$$F_i = \frac{f_{i+1} + f_{i-1}}{2} - f_i \tag{2}$$

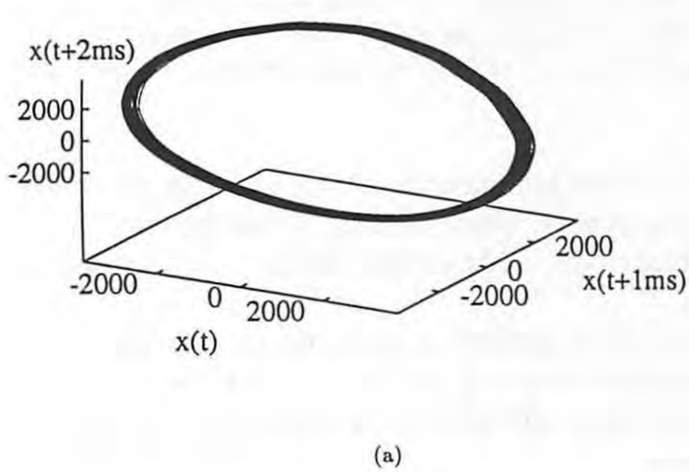
$$A_i = \frac{a_{i+1} + a_{i-1}}{2} - a_i \tag{3}$$

which measure the deviations from a linear trend (Pinto & Titze, 1990). Here f_i (or a_i) denotes the i-th extracted frequency (or amplitude). Usually, the second order perturbation functions are more stationary than the original contours and contain, nevertheless, information on period-doublings, modulations, etc.

Phase Portraits

If relatively stationary segments of the signal have been found, a more detailed analysis of the corresponding attractors is possible. A first step is the representation of the signal in a reconstructive phase space. For this purpose, delay-coordinates are useful (Equation 1).

For visualizing the attractor, embedding dimensions of $m = 2$ or $m = 3$ are appropriate. The delay-time is chosen as a fraction of the pitch period, which gives a proper representation of the attractor. Figure 5 shows some representative phase portraits. Figure 5a shows a limit cycle resulting from a sample of "normal" phonation. Figure 5b shows a more complicated attractor which takes the shape of a cylindrical surface. This relatively complicated attractor was found in a patient with laryngeal nerve paralysis (see also Figure 7 referring to the same patient).



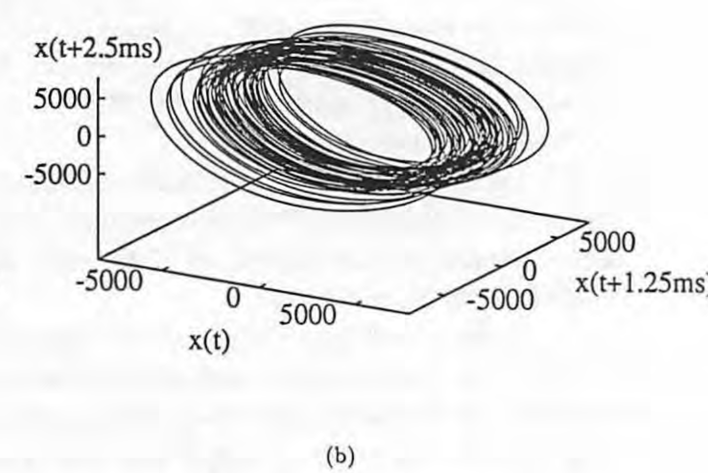


Figure 5. Phase portraits generated from (a) normal phonation and (b) a patient with laryngeal paralysis. The signals have been low-pass filtered slightly above the fundamental frequency.

Attractors can be characterized by discrete maps as well. For example, the intersections of orbits in phase space with a predefined plane (Poincaré plane) can be analyzed. If the attractor is a two-dimensional torus, such a Poincaré section yields a closed curve of intersection points.

Next Amplitude or Next Period Maps

Topologically equivalent to such Poincaré sections are next amplitude or next period maps. Such maps are easily derived from the F_o (or amplitudes) contours. It has been exemplified that such maps give valuable information on the attractor (Lorenz, 1963; Titze et al. 1992). Next amplitude (or period) maps can reveal additional correlations in the contours that are not necessarily covered by the autocorrelation function. An example of a next amplitude map, derived from the amplitude contour of Figure 4b (paralytic dysphonia), is shown in Figure 6.

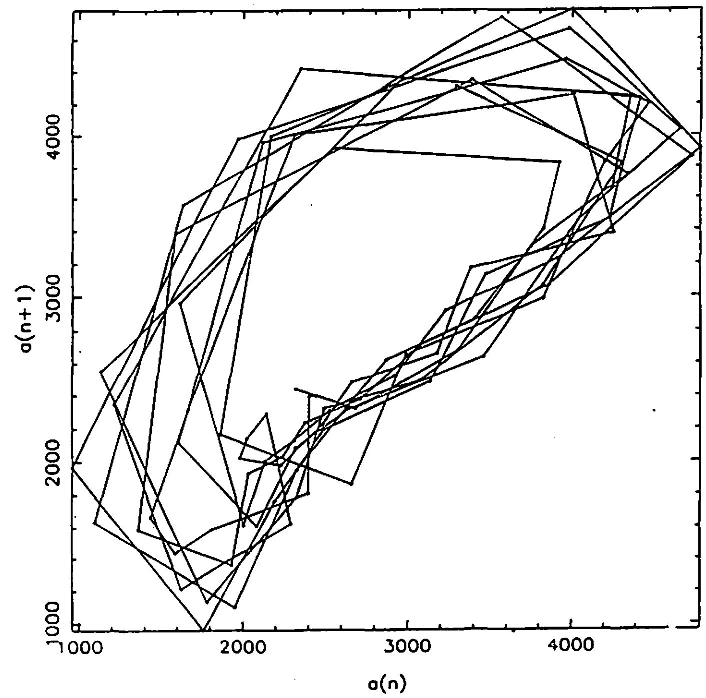


Figure 6. A next amplitude map which approximates a closed, ellipsoidal curve, suggesting that the corresponding attractor in phase space is a torus (i.e., there are two independent frequencies in the system).

Despite nonstationarities, the trajectory shown in Figure 6 approximates a closed, ellipsoidal curve. This suggests that the corresponding attractor in phase space is a two-dimensional torus, as might be expected assuming independence of the low-frequency modulation and the fundamental frequency.

Beside these methods to visualize the attractors, methods have been derived to describe the attractors quantitatively with fractal dimensions and Lyapunov exponents (Bergé et al. 1983; Eckmann & Ruelle, 1985; Mayer-Kress, 1986; Holden, 1986). It has been shown (Mende et al. 1990; Herzel & Wendler, 1991; Titze et al. 1992) that these techniques apply to sufficiently stationary phonatory signals. However, these relatively sophisticated methods are beyond the scope of this paper.

Analysis of Characteristic Voice Samples

In this section, sustained vowels of four patients are analyzed in detail in order to (1) provide further evidence of bifurcations and chaos in phonatory signals and (2) illustrate the application of the techniques described above. First, vocalizations from a male patient with paralytic dysphonia, previously introduced in Figure 5b, are further analyzed. The voice sounded breathy and slightly rough. Figure 7a shows an unfiltered acoustic waveform with pronounced amplitude fluctuations (shimmer around 5%). Subharmonics or bifurcations are not immediately visible. However, the amplitude contour, and especially the corresponding autocorrelation function in Figure 7b, indicate that the amplitude modulations are not random.

This observation can be substantiated using next amplitude (period) maps (see Figure 7c). Due to noise and nonstationarities, the structure is not as obvious as in Figure 6. However, with enumeration of the points, a rotational motion becomes visible. The rotation angle is about 120°, which indicates that the dynamics is close to period-three motion. Indeed, period-three is also detected by the autocorrelation (Figure 7, right panel).

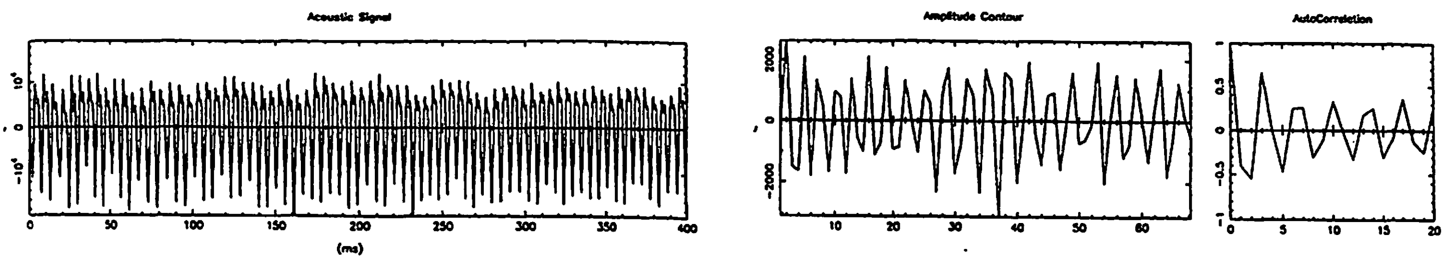
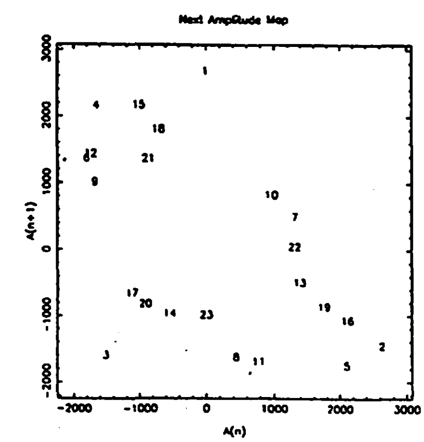


Figure 7. Displays of an (a - upper left) acoustic waveform, (b - upper right) amplitude contour and autocorrelation function and (c - right) an enumerated next-amplitude plot from a patient with paralytic dysphonia .



Summarizing the above observations, we can conclude that the hoarseness of the patient is not only due to turbulent noise, but also due to complicated vibratory pattern of the folds, i.e., period-tripling. The subharmonics are also clearly visible in the EGG records. This case demonstrates that often several sources of hoarseness superimpose on each other.

Indications of period-doubling or tripling and of low-frequency modulations have been found also in the vocalizations of four other patients of paralytic dysphonia (see e.g. Figures 3, 4b and 6).

From a total number of eight cancer patients, we have found three with indications of bifurcations. However, these voices were relatively unstable and, therefore, a detailed quantitative analysis was difficult due to nonstationarity. A characteristic example of a vowel with various transitions is presented in Figure 8. There are low-frequency modulations as well as abrupt transitions to other regimes which might be related to bifurcations.

Even though Aronson, Peterson and Litin (1964) have already mentioned diplophonia, low-pitched segments and "intermittent squeaking noises" for functional voice disorders, these vocal disorders have been studied less extensively. Out of 36 speakers, we have found six with clear signs of bifurcations. As a first example, we discuss a voice with hyperfunctional childhood dysphonia. A spectrogram of the first utterance of this speaker was already shown in Figure 2. All three recorded utterances exhibit audible low-frequency episodes. Figure 9 displays a time series with various transitions. Again, amplitude and pitch contours are helpful tools for the analysis. Figure 10a is an amplitude contour showing a fairly strong period-tripling. Figure 10b shows the fundamental frequency contour near the end of an utterance, at which point a low-frequency modulation appears which was perceived as vocal fry.

Finally, we present the voice of a male speaker with hyperfunctional dysphonia (Figure 11). The signal displays two interesting transitions. First, towards the end of line (c), a period-doubling is visible. Then, on line (d), an amplitude modulation of about 40 Hz is seen. Consequently, vocal fry episodes can be perceived. These bifurcations are also reflected in the corresponding amplitude contour. Similar transitions occur in a second utterance of the same speaker. The example in Figure 11 reflects a general observation that often several bifurcations occur within a single vocalization if the borderline of normal phonation is reached.

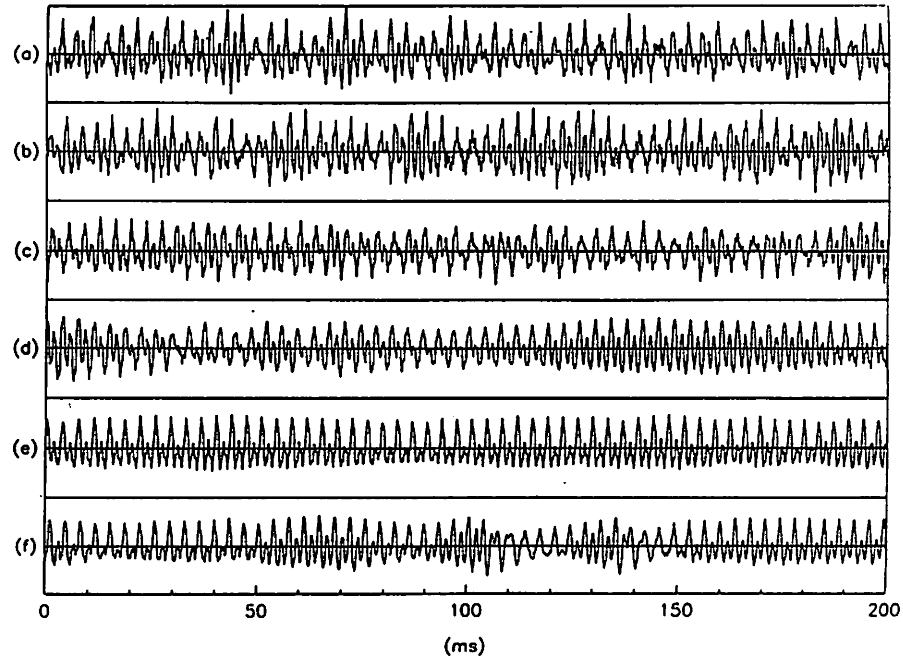


Figure 8. A characteristic utterance from a cancer patient showing beating-like segments and sudden jumps to other phonatory regimes.

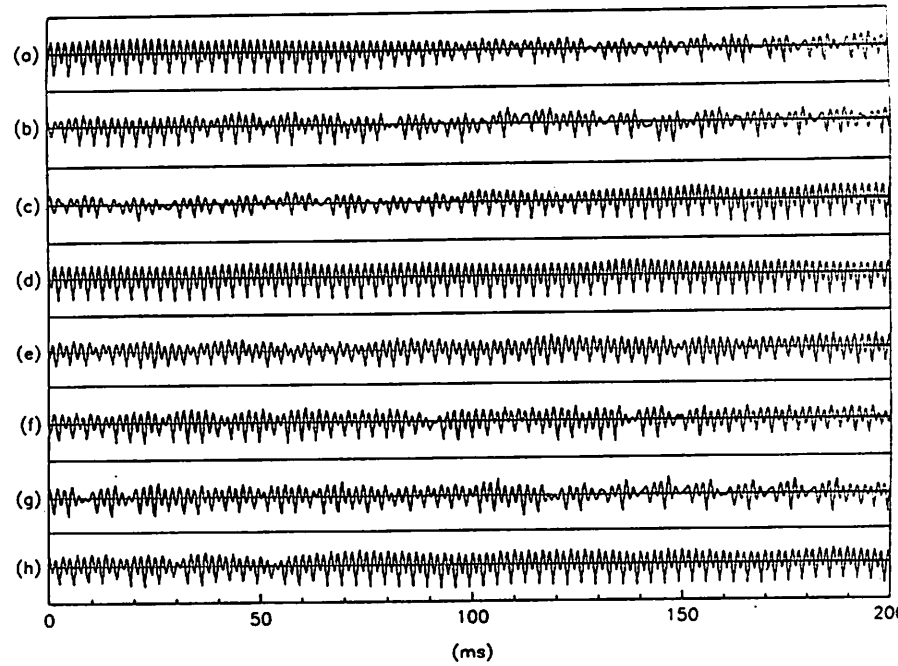


Figure 9. An utterance from a patient with hyperfunctional childhood dysphonia showing various transitions.

The phonatory signals analyzed in this section have illustrated that a variety of bifurcations are found in pathological voices. Possible mechanisms leading to bifurcations and chaos will now be discussed.

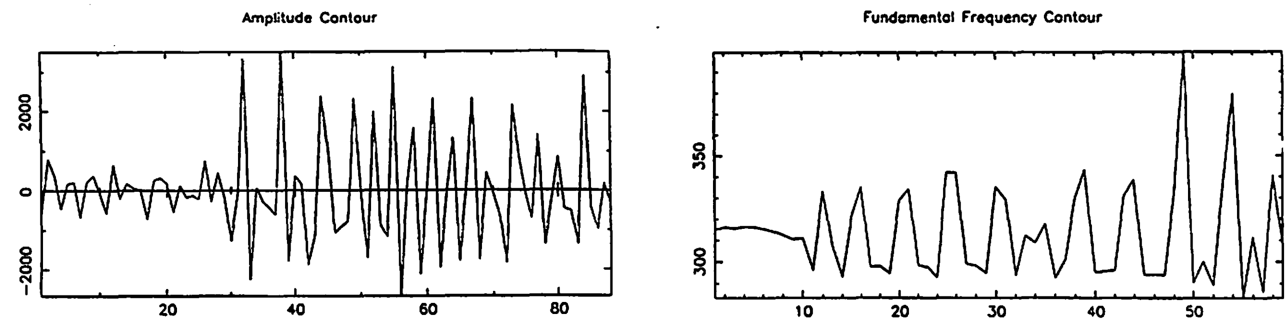


Figure 10. (a-left) A segment of an amplitude contour from a patient with hyperfunctional childhood dysphonia. A transition to a subharmonic regime close to period-tripling appears to be present, (b-right) a segment of a fundamental frequency contour for the same patient. An onset of a modulation is present which is perceived as vocal fry.

Discussion

Chaos and bifurcations have been reported in many acoustic systems besides the voice. Examples are bubbles in water driven by a sound field (Lauterborn & Parlitz, 1988) and woodwind musical instruments (Maganza & Caussé, 1986; Keefe & Laden, 1991). But the vocal folds are particularly prone to chaotic behavior because of nonlinearities in airflow and tissue dynamics (Titze, Baken & Herzel, 1992).

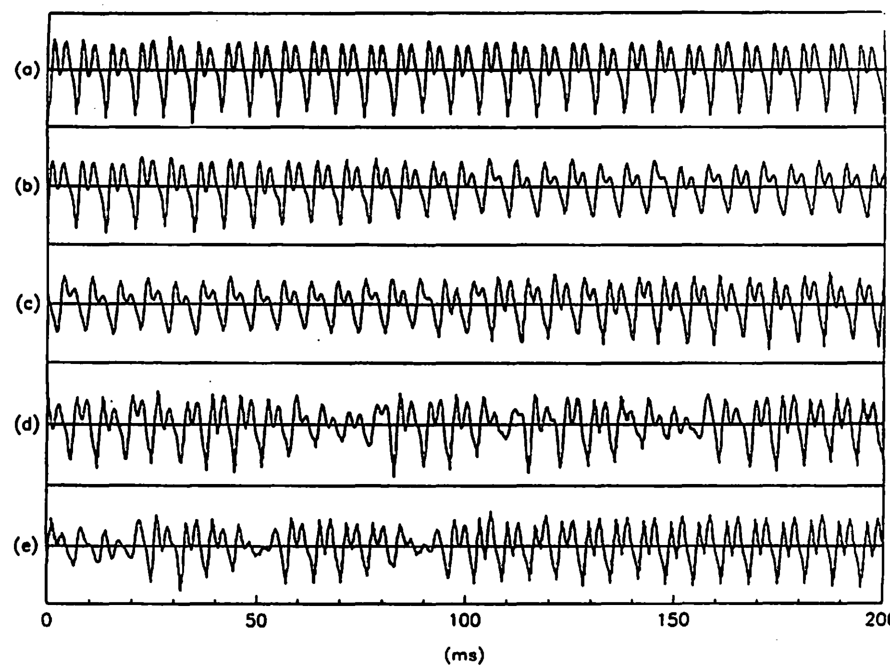


Figure 11. An acoustic waveform from a male speaker with hyperfunctional dysphonia showing two interesting transitions: a period-doubling toward the end of line (c), and an amplitude modulation of about 40 Hz on line (d).

Mechanisms

Qualitatively, the origin of bifurcations and low-dimensional attractors can be understood as follows: normal phonation corresponds to an essentially synchronized motion of all vibratory modes. A change of parameters such as muscle tension or localized vocal fold lesions may lead to a desynchronization of certain modes resulting in bifurcations and chaos. The desynchronization of the following modes might be of particular relevance:

*desynchronization motion of the left and right vocal fold (e.g. for localized vocal fold lesions and unilateral paralysis)

*desynchronization of horizontal and vertical modes (such "symmetric chaos" might be relevant for an understanding of vocal fry phonation)

*interaction of true and false vocal folds (e.g., for pressed voice)

*interaction of vocal fold vibrations with sub- and supraglottal acoustic resonances.

Obviously, quite a lot of experimental and theoretical work is necessary to assign the various acoustic observations to these physiological mechanisms. However, nonlinear dynamics provides a framework to understand the underlying mechanisms. If low-dimensional attractors can be identified as the origin of a rough voice, then the phenomena are due to the nonlinear interaction of a few degrees of freedom. It is already known (Glass & Mackey, 1988; Titze et al. 1992) that period-doubling, frequency jumps, beating like modulations and chaos can be found in a system of two coupled oscillators. Thus, it makes sense to look for the specific modes that are desynchronized.

Moreover, vocal fry or simulated creaky voice might be a good starting point to understand the physiological mechanisms of complicated vocal fold vibrations. Preliminary simulations with a finite element model of the vocal folds (Titze & Alipour-Haghighi, 1992) and a two-mass model (Herzel, Steinecke, Mende & Wermke, 1991) suggest that vocal fry can be generated due to the desynchronization of the horizontal movement of the folds dictated by the body and the vertical motion which is very slow due to a lax cover. The computer simulations show strong correlations with the high-speed film observations of Moore and von Leden (1958). A detailed analysis of the computer simulations in terms of bifurcations and chaos is in the planning stages.

Terminology

In the literature, a rich variety of terms are used to describe nonlinear phenomena. Among them are dicrotic dysphonia, diplophonia, creaky voice, subharmonic vocalization, double harmonic break, biphonation, octave jump and noise concentrations. "Double harmonic break" refers to the sudden appearance of subharmonics, and "biphonation" describes two independent pitches. "Noise concentrations" probably correspond to chaotic episodes (Mende et al. 1990). Sometimes, authors assign different meaning to these terms. For example, diplophonia is often used to describe alternating amplitudes of glottal pulses, while in other instances it is used to denote the appearance of two fundamental frequencies. Thus, an important contribution of nonlinear systems theory might be a well-defined terminology to classify the observations.

According to nonlinear systems theory, stationary deterministic signals can be assigned to three attractor types: limit cycle (periodic), torus (two or more independent frequencies) or chaotic attractors (nonperiodic). Sudden transitions due to parameter changes can be interpreted as bifurcations. For example, an octave jump is related to a period-doubling bifurcation and the appearance of a modulation might be a manifestation of a secondary Hopf bifurcation.

Summary

Our study was devoted to specific features of voice signals, such as sudden frequency jumps, appearance of a second frequency and noisy episodes. It was demonstrated that these phenomena can be identified in various kinds of voice disorders, but can also appear also in normal healthy voices and in childrens' vocalizations. Nonlinear dynamics provides an approach to understand these features in a unified manner.

Chaos theory guided our choice of analysis procedures. In addition to conventional time and frequency domain techniques, the embedding of a time series in a reconstructed phase space often yielded new information and insight. Attractors were also visualized with phase portraits and next period (amplitude) maps. For example, modulations (i.e. tori in our terminology) were detected as rotational motion in next-amplitude plots. A more sophisticated attractor analysis would involve the estimation of dimensions and Lyapunov exponents (Titze et al. 1992). However, these techniques would require relatively long, stationary time series, which are often not available.

Much information was obtained from conventional F_o and amplitude contours and their corresponding autocorrelation functions. In these plots, period-doublings appeared as zig-zag sections and low-frequency modulations were easy to visualize. Sometimes it was desirable to remove long-term trends by taking the second derivative of the contours.

We claim that the nonlinear dynamics approach offers new ways of understanding certain features of voice signals. Various observations can be interpreted from a physical perspective. Sources of irregularity can often be traced back to the desynchronization of principle vibratory modes.

Acknowledgement

This work was supported by Grant No. P60 00976 from the National Institutes on Deafness and Other Communicative Disorders. We gratefully acknowledge Julie Lemke for manuscript preparation.

References

Aronson, A., Peterson, H., Jr., & Litin, E. (1964). Voice symptomatology in functional dysphonia and aphonia. Journal of Speech and Hearing Disorders, 4, 367-381.

Askenfelt, A.G., & Hammarberg, B. (1986). Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. Journal of Speech and Hearing Research, 29, 50-64.

Baken, R. (1990). Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice. Journal of Voice, 4, 185-197.

Bergé, P., Pomeau, Y., & Vidal, C. (1986). Order Within Chaos. New York: Wiley.

Cooper, D.S. (1989). Voice: A historical perspective. Journal of Voice, 3, 187-203.

Dolansky, L., & Tjernlund, P. (1968). On certain irregularities of voiced speech waveforms. IEEE Trans. AU-16, 51-56.

Eckmann, J.-P., & Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. Rev. Modern Physics, 57, 617-656.

Glass, L., & Mackey, M. (1988). From Clocks to Chaos. Princeton: University Press.

Hammarberg, B., Fritzell, B., Gauffin, J., & Sundberg, J. (1986). Acoustic and perceptual analysis of vocal dysfunction. Journal of Phonetics, 14, 533-547.

Herzel, H. (1992). Bifurcations and chaos in voice signals. In T. Kapitaniak (Ed.) Chaos: The New Idea. Cambridge: University Press.

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

Herzel, H., Steinecke, I., Mende, W., & Wermke, K. (1991). Chaos and bifurcations during voiced

Herzel, H., Steinecke, I., Mende, W., & Wermke, K. (1991). Chaos and bifurcations during voiced speech. In E. Mosekilde & L. Mosekilde (Eds.), Complexity, Chaos and Biological Evolution. New York: Plenum Press.

Holden, A.V. (Ed.) (1986). Chaos. Manchester: University Press.

Hollien, H., & Michel, J. (1968). Vocal fry as a phonational register. Journal of Speech and Hearing Research, 11, 600-604.

Imaizumi, S. (1986). Acoustic measures of roughness in pathological voice. Journal of Phonetics, 14, 457-462.

Ishizaka, K., Isshiki, N. (1976). Computer simulation of pathological vocal-cord vibration. Journal of the Acoustical Society of America, 60, 1193-1198.

Keefe, D.H., & Laden, B. (1991). Correlation dimension of woodwind multiphonic tones. Journal of the Acoustical Society of America, 90, 1754-1765.

Kelman, A.W. (1981). Vibratory pattern of the vocal folds. Folia Phoniatrica, 33, 73-99.

Klatt, D.H. & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality vibrations among female and male talkers. Journal of the Acoustical Society of America, 87, 820-841.

Koike, Y. (1969). Amplitudes modulations in patients with laryngeal diseases. Journal of the Acoustical Society of America, 45, 839-844.

Lauterborn, W., & Parlitz, U. (1988). Methods of chaos physics and their application to acoustics. Journal of the Acoustical Society of America, 84, 1975-1993.

Lorenz, E.N. (1963). Deterministic nonperiodic flow. Journal Atmospherical Sciences, 20, 130-141.

Ludlow, C., Coulter, D. & Gentages, F. (1983). The differential sensitivity of frequency perturbation to laryngeal neoplasms and neurophathologies. In D. Bless & J. Abbs (Eds.), Vocal Fold Physiology - Contemporary Research and Clinical Issues. San Diego: College Hill Press.

Maganza, C., & Caussé, R. (1986). Bifurcations, period doublings and chaos on clarinet-like systems. Europhysics Letters, 1, 295-302.

Mayer-Kress, G. (1986). Dimensions and Entropies in Chaotic Systems. Berlin: Springer.

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

Moore, G. & van Leden, H. (1958). Dynamic variation of the vibratory pattern in the normal larynx. Folia Phoniatrica, 10, 205-238.

Pinto, N.B. & Titze, I.R. (1990). Unification of perturbation measures in speech signals. Journal of the Acoustical Society of America, 87, 1278-1289.

Ramig, L.O. & Scherer, R.C. (1991). Speech therapy for neurological disorders of the larynx. NCVS Report 1, 167-190.

Ramig, L.A., Scherer, R.C., Titze, I.R., & Ringel, S.P. (1988). Acoustic analysis of voices of patients with neurologic disease: A rationale and preliminary data. Annals of Otology, Rhinology and Laryngology, 97(2), 164-172.

Remacle, M., & Trigoux, I. (1991). Characteristics of nodules through the high-resolution frequency analyzer. Folia Phoniatrica, 43, 53-59.

Robb. L., & Saxman, J. (1988). Acoustic observations in young children's noncry vocalizations. Journal of the Acoustical Society of America, 83, 1876-1882.

Saleh, M.M.H. (1991). Acoustic Analysis of Voice in Certain Types of Dysphonia. Unpublished Ph.D. thesis. Ain Shams University.

Scherer, R. (1989). Physiology of creaky voice and vocal fry. Journal of the Acoustical Society of America, 86(S1), S25(A).

Sirviö, P., & Michelsson, K. (1976). Sound-spectrographic cry analysis of normal and abnormal newborn infants. Folia Phoniatrica, 28, 161-173.

Smith, M.E., Berke, G.S., Gerratt, B.R. & Kreiman, J. (1992). Laryngeal paralyses: theoretical considerations and effects on laryngeal vibration. Journal of Speech and Hearing Research, 35, 545-554.

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

Titze, I.R. (1989). Acoustic of creaky voice. Journal of the Acoustical Society of America, 86 (S1), S26(A).

Titze, I.R., & Alipour-Haghighi, F. (1992). Three source models for voice synthesis. Submitted to Journal of the Acoustical Society of America.

Titze, I., Baken, R., & Herzel. H. (1992). Evidence of chaos in vocal fold vibration. In I. Titze (Ed.), Vocal Fold Physiology: New Frontiers in Basic Science. San Diego: Singular Publishing Group.

Titze, I.R. (Ed.) (1992). Vocal Fold Physiology: New Frontiers in Basic Science. San Diego: Singular Publishing Group.

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

Wieser, M. (1981). Periodendaueranalyse bei spastischen Dysphonien. Folia Phoniatrica, 33, 314-324.

Zwicker, E., & Fastl, H. (1990). Psychoacoustics. Berlin: Springer.

NCVS Status and Progress Report - 4
June 1993, 195-200

Fundamental Frequency and Amplitude Perturbation in Reconstructed Canine Vocal Folds

Jack J. Jiang, M.D., 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
David B. Wexler, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa
Steve D. Gray, M.D.
Department of Otolaryngology-Head and Neck Surgery, The University of Iowa

Abstract
A submucosal fat autograft was implanted within the cover of injured vocal folds of five dogs. The implant occurred six weeks after unilateral mucosal excision had been performed. Three months postoperatively the larynges of these animals were excised and their phonation compared to those of normal dog larynges and to other larynges with mucosal excision (but without fat-grafting). Radiated acoustic pressure from the artificially driven larynges was recorded and digitized at 20 kHz with 16 bits resolution. Amplitude and fundamental frequency perturbations were extracted from a segment of phonation to assess the stability of the acoustic signals from the three groups. It was found that fat-augmentation after mucosal excision reduced amplitude and frequency perturbation measures. There was no significant difference between fat-augmented and normal vocal folds. The acoustic measures were also positively correlated with phonation threshold and phonation efficiency measures reported earlier. Result suggests that submucosal fat autograft implantation within an injured vocal fold cover cannot only restore the "ease" of phonation, but also the stability of phonation, which is a component of vocal quality.

Introduction

This paper is a follow-up to an earlier study by the same authors (Wexler, et al. 1989) on a phonosurgical procedure involving a submucosal autograft of fat in an injured canine vocal fold. The purpose of the earlier study was to investigate if a functional lamina propria could be maintained by grafting. After surgery and an extended recovery period, the larynges were harvested and phonated on a laboratory bench. It was found that the fat-augmented vocal folds had lower phonation threshold pressure, greater vocal intensity, and greater glottal efficiency than the injured vocal folds without fat-grafting. These results suggested that this phonosurgical procedure improves laryngeal function, at least as far as measures of "ease of phonation" are concerned.

Normal vocal function is not represented by "ease" of phonation alone, but also by stability of phonation. It was deemed appropriate, therefore, to follow up with some perturbation measures. A series of publications have appeared that relate to specific or general laryngeal diseases, and the perception of vocal disorders, to a variety of perturbation measures (Simon, 1927; Horii, 1975, 1979; Yumoto, Gould and Baer, 1982; Hanson, Gerratt, and Ward, 1983; Askenfelt and Hammarberg, 1986). It seems that jitter, shimmer, and the signal to noise ratio increase in most pathological cases.

Phonatory stability is sensitive to structural characteristics of the vocal folds. Scarring of vocal fold tissue can result in irregular vibration, which can manifest itself as hoarseness or loss of sonority. A series of studies by Hirano and Kakita (1985) and more recent theoretic work by Titze, Jiang and Druker (1988) have supported a body-cover theory of vocal fold vibration. In this theory, the vocal fold is represented as a layered structure having a cover, which consists of the surface epithelium and the underlying superficial lamina propria, and a body, consisting of the deeper lamina propria and the vocalis muscle. A mathematical analysis of the body-cover model shows that the compliant nature of the lamina propria is critical to normal vocal fold vibration (Titze, et al. 1988). Further physiologic significance of the cover in phonation is discussed by Cooper (1988). Animal studies have demonstrated that loss of the superficial layer and chronic fibrotic changes in the lamina propria occur after vocal fold stripping (Leonard, et al. 1985). Therefore, phonosurgical procedures should be conducted with maximum prevention of scarring and maximum restoration of a functional lamina propria.

The purpose of this study was to use acoustic perturbation analysis to investigate the effects of a vocal fold augmentation procedure on the stability of phonation in scarred canine vocal folds. It was our hypothesis that if phonation stability could be restored in these damaged vocal folds, along with intensity and glottal efficiency described earlier, then the quality of the sound would be improved along with "ease of phonation".

Method

Thirteen female canines weighing 15-20 kg were used in the earlier study. Four animals served as the unoperated group, hereafter referred to as the normal group. The other nine animals underwent a right true vocal fold mucosal excision endoscopically. The excisions were directed along the leading vocal fold edge, at a depth to include the upper-to-middle lamina propria. After the mucosal excision, four of the nine animals had no further operative intervention. They are referred to as the mucosal excision group.

After six to eight weeks of healing from the mucosal excision, the remaining five of the nine mucosal excision animals underwent a right vocal fold fat-autograft augmentation. This group will be referred to as the fat-graft group. Through a midline laryngofissure approach, a submucosal fat autograft was implanted within the injured vocal fold cover.

Coronal-section photomicrographs were taken from the vocal folds. In the normal group, the lamina propria was well developed (Photo 5, left side; see center-bound plate). In the mucosal excision group, reepithelialization occurred, but a deficiency of the lamina propria thickness remained apparent two months after excision. In the fat-graft group, the implant remained localized near the leading edge of the vocal fold after two months (Photo 5, right side).

All larynges of the three groups of animals were harvested three months postoperatively. The performance of the larynges was studied with an excised larynx setup (Baer, 1975; Durham, Titze, and Scherer, 1987). A humidified air column with regulated flow provided a subglottic airstream while the subglottal pressure was monitored. Radiated acoustic pressure was recorded on the audio channel of a Panasonic VCR and digitized at 20 kHz with 16 bits resolution (DSC 240 A/D converter) on a VAX station 3200. Four tokens of phonation were used for each larynx. Each token was one second in duration. A waveform matching algorithm (Milenkovic, 1987; Titze and Liang, in press) was utilized to evaluate the frequency and amplitude stability (jitter and shimmer) of every token for all three groups.

Results

Figure 1 shows shimmer and amplitude coefficient of variation ACV (standard deviation divided by the mean) of the microphone signal for each group. Shimmer was computed as the mean rectified adjacent-cycle difference between peak-to-peak amplitudes, whereas ACV was computed as the root mean squared difference from the mean (Pinto and Titze, 1990). Comparatively, shimmer relates more to the short term fluctuation in amplitude, whereas the coefficient of variation relates more to long term fluctuations. Error bars indicate the standard error of the mean for each measure (0.05 for shimmer and 0.06 for CV). It is seen that the fat-augmented vocal fold group has significantly lower amplitude perturbation measures than the mucosal excision group and was essentially equivalent to the normal group.

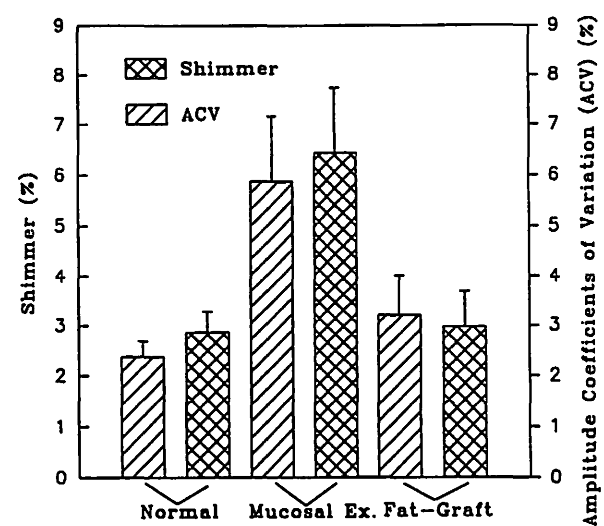


Figure 1. Coefficient of peak-to-peak amplitude variation (ACV) and shimmer obtained from the acoustic radiated pressure for the three experimental groups.

Figure 2 shows jitter and the coefficient of variation of fundamental frequency FCV. Again, the fat-augmented vocal folds had significantly lower values ($p < 0.05$ for jitter and $p < .06$ for FCV) than the mucosal excision group. Values for both normal and fat-graft groups were less than 1%, which is typical for normal human steady phonation (Horii, 1979; Titze, Horii, and Scherer, 1987).

A third measure, the noise-to-harmonic ratio (which is the inverse of the harmonic-to-noise ratio described by Yumoto et al. 1982) is shown in Figure 3. This shows most dramatically the difference in regularity of vibration between the mucosal excision group and the two other groups. An approximate -20 dB noise-to-harmonic ratio (or +20 dB harmonic-to-noise ratio) is typical for human phonation. The mucosal excision group had about 25-30 dB more noise energy in the acoustic signal than either the normal group or the fat graft group.

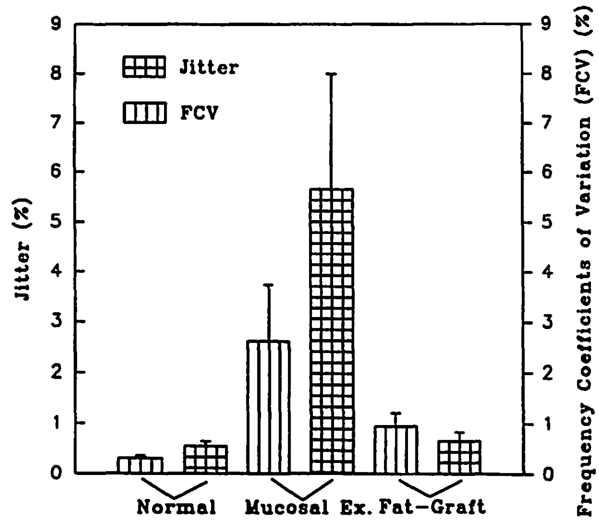


Figure 2. Coefficient of variation of fundamental frequency (FCV) and jitter obtained from the acoustic radiated pressure for the three experimental groups.

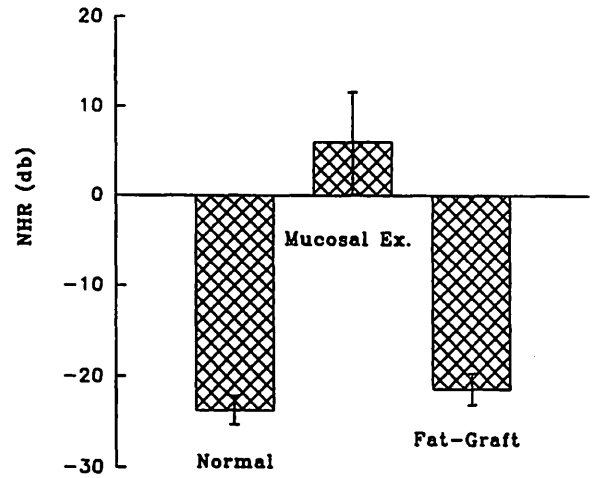


Figure 3. Noise-to-harmonic ratio obtained from the acoustic radiated pressure for the three experimental groups.

Discussion and Conclusion

As an addendum to results reported in an earlier study, the present results suggest that submucosal fat autograft implantation in an injured vocal fold cover can restore the stability (periodicity) of the acoustic output. This is in addition to the improvement in "ease" of phonation described earlier in terms of vocal efficiency and phonation threshold pressure. Because acoustic stability is a component of vocal quality, the result indicate that the fat autograft may improve overall phonation quality of a damaged vocal fold, although no formal quality judgements were made.

Generally, the purpose of reconstructive procedures is to make the vocal folds more symmetric. Restoration of symmetry should include both geometric and viscoelastic properties of the vocal folds. It appears that a fat-graft can offer some improvement in both tissue properties and vocal fold shape.

Acknowledgement

This work was supported by Grant No. DC00976 from the National Institute on Deafness and Other Communication Disorders. The authors are grateful for this support, and also for the assistance in manuscript preparation by Julie Lemke.

References

Askenfelt, A. and Hammarberg, B. (1986). Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. Journal of Speech and Hearing Research, 29, 50-64.

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

Cooper, D.S. (1988). The laryngeal mucosa in voice production. Ear, Nose, and Throat Journal, 332-352.

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

Hanson, D.G., Gerratt, B.R., and Ward, P.H. (1983). Glottographic measurement of vocal dysfunction, a preliminary report. Annals of Otology, Rhinology and Laryngology, 92, 413-420.

Horano, M., and Kakita, Y. (1985) Cover-body theory of vocal vibration. In: Daniloff RG, (Ed.) Speech Science. San Diego: College-Hill Press, 1-46.

Horii, Y. (1975). Some statistical characteristics of vocal fundamental frequency. Journal of Speech and Hearing Research, 18, 192-201.

Horii, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. Journal of Speech and Hearing Research, 22, 5-19.

Leonard, R.J., Kiener, D., Charpied, G., and Kelly, A. (1985). Effects of repeated stripping on vocal fold mucosa in cats. Annals of Otology, Rhinology, and Laryngology, 94, 258-262.

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

Pinto, N., & Titze, I.R. (1990). Unification of perturbation measures in speech analysis. Journal of the Acoustical Society of America, 87(3), 1278-1289.

Simon, C. (1927). The variability of consecutive wavelengths in vocal and instrumental sounds. Psychological Monographs, 26, 41-83.

Titze, I., Horii, Y., and Scherer, R. (1987). Some technical considerations in voice perturbation measurements. Journal of Speech and Hearing Research, 30, 252-260.

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

Titze, I., and Liang, H. (in press). Comparison of F_0 extraction methods for high precision voice perturbation measurements. Journal of Speech and Hearing Research.

Wexler, D., Gray, S., Jiang, J., and Titze, I. (1989). Phonosurgical studies: Fat-graft reconstruction of injured canine vocal cords. Annals of Otology, Rhinology and Laryngology, 98, 9, 668-673.

Yumoto, E., Gould, W., and Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. The Journal of the Acoustical Society of America, 71, 1544-1550.

NCVS Status and Progress Report - 4
June 1993, 201-207

Fundamental Frequency Stability in Functional Dysphonia

M. Nasser Kotby, M.D.
Faculty of Medicine, Ain Shams University, Egypt
Ingo R. Titze, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Marwa M. Saleh, M.D.
Faculty of Medicine, Ain Shams University, Egypt
David A. Berry, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

Functional dysphonia is a term applied to voice disorders for which there is an absence of apparent structural change in the larynx. The aim of this work was to investigate how functional dysphonia may differ acoustically from other types of dysphonia. Fundamental frequency profiles for steady vowels were generated using a software program called GLIMPES (Glottal Imaging by Processing External Signals). The fundamental frequency variations were found to be unimodal in normal individuals. In contrast, the variations for dysphonic patients were sometimes bi-modal or multi-modal due to the presence of subharmonics or low-frequency modulations. The appearance of these patterns was generally related to the severity of the dysphonia rather than to its etiology.

Introduction

The term "functional dysphonia" is usually applied when a voice complaint is present with no apparent structural change in the larynx(1). Functional dysphonia has long been recognized(2,3) and variably classified by several authors(4,5,6). Although it represents a large division of voice disorders, this category of dysphonia has not received due attention in the field of acoustic analysis. Most studies have focused on organic lesions (neoplastic and paralytic dysphonia) and minimal associated pathological lesions [MAPLs] (polyps and nodules). Perhaps, the only acoustic study that focused exclusively on functional dysphonia was that of Klingholz and Martin(7). It compared jitter and shimmer values in hypo- and hyperfunctional dysphonia.

While acoustic analysis has not been very successful in the etiological categorization of voice disorders(8), it has been found useful for assessing the degree of the vocal pathology(9) and in monitoring the effect of intervention(10).

The aim of this study was to investigate the fundamental frequency stability in steady vowels for patients with functional dysphonia and compare them to those of a normal group.

Methods

The study was conducted on 155 subjects: 40 normal and 115 dysphonic. Results of only a few interesting cases are reported here, however. The normal subjects had no history of laryngeal problems. The etiological categorization of dysphonic patients is shown in Table 1. This was the classification suggested by Kotby(6). The normal subjects consisted of 20 males and 20 females, ranging from 18 to 42 years of age. The dysphonic subjects included 49 females and 66 males, with ages ranging from 8 to 72 years.

Table 1.
The Number of Patients Within Each Category and Subtype of Dysphonia

A. Functional Dysphonia	48 patients
1. Phonasthenia	12 patients
2. Hyperfunctional Dysphonia	10 patients
3. Hypofunctional Dysphonia	8 patients
4. Mutational Voice Disorder	8 patients
5. Hyperfunctional Childhood Dysphonia	10 patients
B. MAPLs Dysphonia	40 patients
1. Nodules	14 patients
2. Polyps	8 patients
3. Cysts	8 patients
4. Reinke's edema	10 patients
C. Organic Dysphonia	27 patients
1. Paralytic Dysphonia (neurological disorders)	17 patients
2. Neoplastic Dysphonia	10 patients

The following procedures were used in the assessment of every patient:

Elementary Diagnostic Procedures:

These included a patient interview, an auditory-perceptual assessment of the patient's voice (APA), and a preliminary visual examination of the larynx through indirect laryngoscopy. The character of the voice as analysed through the APA was described according to a modified GRBAS scale[11,6]. It covers the following items: overall grade, strained, leaky, breathy, rough (irregular). The division of dysphonic patients into those with mild, moderate and severe degrees of dysphonia is shown in Table 2.

Table 2.
Degree of Severity of Dysphonia in Each Pathological Subtype
(auditory perceptual assessment)

	Mild	Moderate	Severe
A. Functional Dysphonia	18	24	6
1. Phonasthenia	12	0	0
2. Hyperfunctional	0	6	4
3. Hypofunctional	0	8	0
4. Mutational	3	5	0
5. Hyperfunctional Childhood	3	5	2
B. MAPLs Dysphonia	13	13	14
1. Nodules	6	4	4
2. Polyps	3	3	2
3. Cysts	2	4	2
4. Reinke's Edema	2	2	6
C. Organic Dysphonia	5	5	17
1. Paralytic (neurologic)	5	5	7
2. Neoplastic	0	0	10

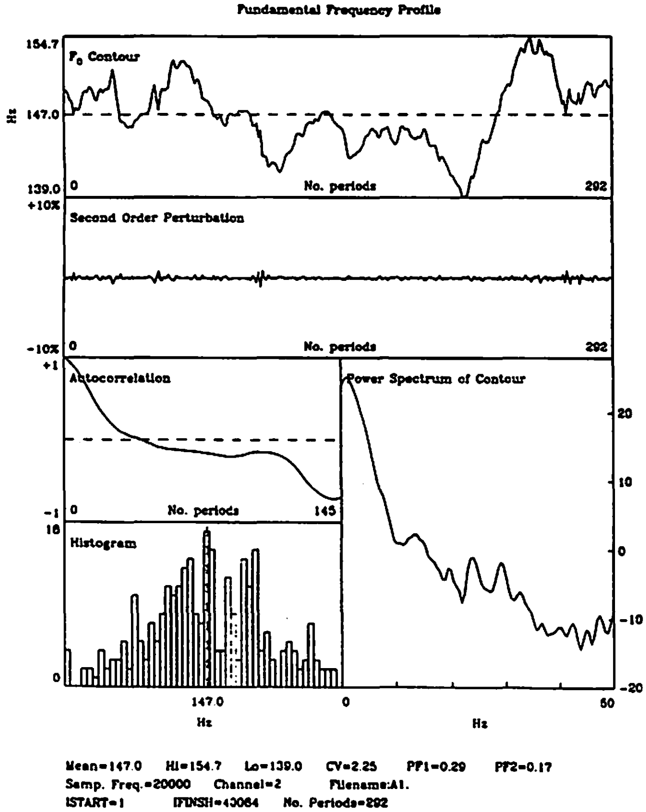
Clinical Diagnostic Aids:

The clinical diagnostic aids included: (1) Augmentation and documentation of the visual examination of the vocal folds by a Zeiss Opmi 9-RC microscope coupled to a Sony Umatic video-cassette recorder VO-7630 and a Hitachi camera 5050. (2) High-fidelity voice recording in a sound treated booth.

Additional Instrumental Diagnostic Procedures:

Acoustic and electroglottograpic signals were obtained from each subject. The recordings were made using a cardiod dynamic microphone (National WM-306N); a wide-flex amplifier (National WA-25); an electroglottograph (EG 830, F-J Electronics A/S Holte Denmark) together with a Pulse Code Modulator (Sony PCM-501 ES) that digitizes and video-encodes the signals for recording onto a VHS video-cassette recorder.

Each subject was asked to phonate a sustained /a/ vowel at a comfortable pitch and loudness and to hold it steady for as long as possible. This task was repeated three times. Detailed analysis of the recordings was done using "GLIMPES" (Glottal Imaging by Processing External Signals), on a VAX 11/750 computer. Prior to analysis, the voice signal was low-pass filtered (24 dB/octave) with a Wavetek filter (model 432) at a frequency of 10 kHz. The filter also offered a 20 dB gain to the signal. The signal was then digitized at a 20 kHz sampling frequency, using a 16-bit analog-to-digital converter [Digital Sound Corporation (DSC) 240 Data Conversion System].



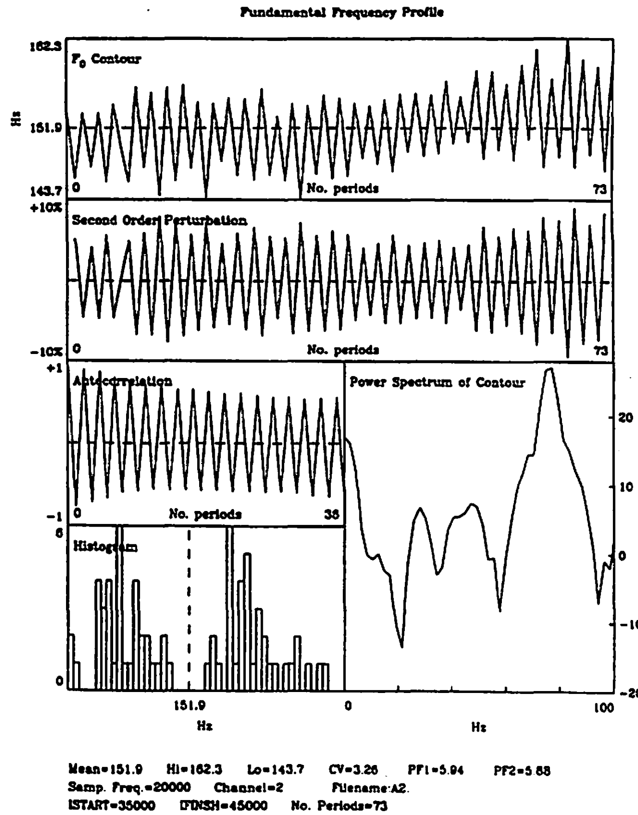
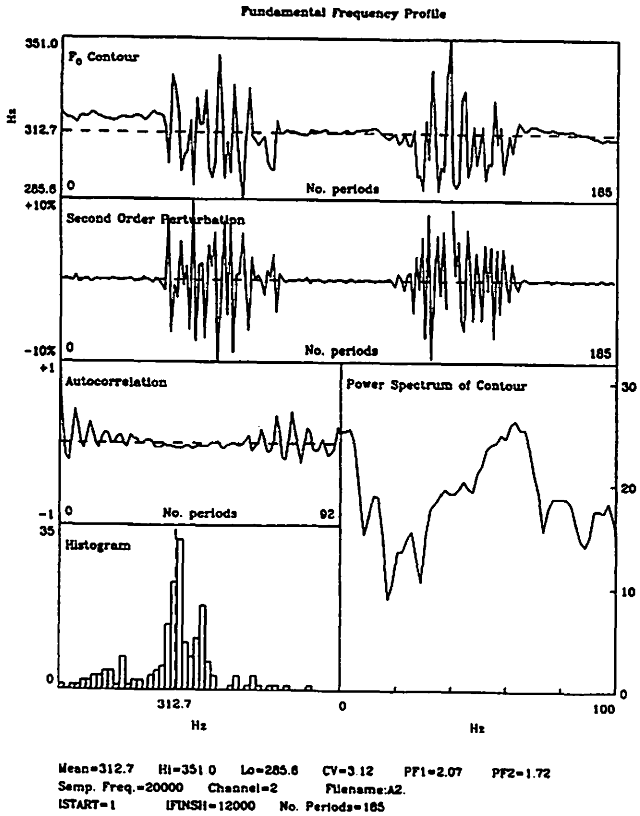


Figure 1 (left). A typical fundamental frequency profile from a sample of normal phonation. The histogram portrays a uni-modal, Gaussian-like distribution of the contour values. This is typical of utterances of persons with no dysphonia or with only mild dysphonia. Figure 2 (right). A fundamental frequency profile from a patient with hyperfunctional dysphonia of moderate severity. A strong, relatively stationary period-doubling is present in this utterance. Such period doublings are common in patients with moderate dysphonia.

To analyze an utterance, the GLIMPES program first extracted the fundamental frequency (or amplitude) contour. Based on this contour, a voice profile was generated which displays: 1) the contour (either frequency or amplitude); 2) a second-order derivative of the contour (a version of the original contour in which low-frequency trends have been suppressed); 3) a histogram, displaying the distribution of instantaneous frequencies or amplitudes; 4) a power spectrum of the contour displaying the low-frequency components of the contour; 5) an autocorrelation function of the contour, or equivalently a time-domain representation of contour power spectrum; and 6) various average perturbations measures, such as the coefficient of variation, first-order mean rectified value, second-order mean rectified value, higher harmonic percentage, and noise-to-harmonic percentage[12]. Figure 1 shows a typical frequency profile for normal phonation.

Results

In general, the fundamental frequency distributions for the normal group were uni-modal and Gaussian-like, as illustrated by the histogram in the lower left corner of Figure 1. This distribution suggests random perturbations, probably a result of neurologic jitter[13] or mucus on the vocal folds. Gaussian distributions are suggestive of high-dimensional movement of tissue, mucus, or unsteady airflow in or around the glottis. A moderate amount of this is expected. The F_0 contour (top of the



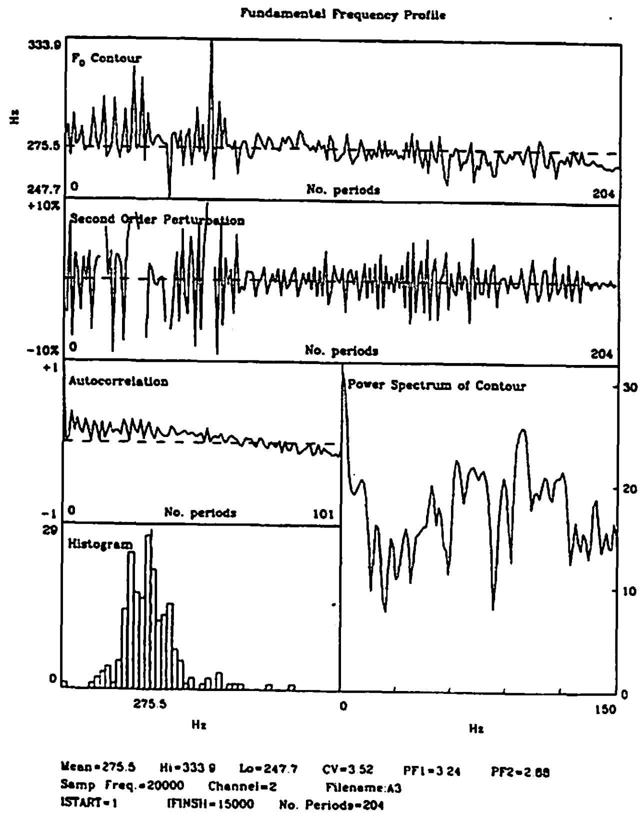


Figure 3 (left). A fundamental frequency profile from a patient with hyperfunctional childhood dysphonia of moderate severity. An intermittent, low-frequency modulation around 65 Hz appears in the contour. This is a less-stationary example of a low-frequency structure in the contour of a patient with a moderate dysphonia. Figure 4 (right). A fundamental frequency profile from a patient with a severe neoplastic dysphonia. A strong 100 Hz frequency component is clearly seen in the contour power spectrum, but is nearly masked by the high "noise" floor in the spectrum--perhaps caused bynonstationarities in the contour. In general, well-defined, low-frequency structures are much harder to identify in patients with severe dysphonia.

figure) shows no specific pattern. This is confirmed by the autocorrelation function, which usually highlights any periodic structure in the F_0 contour. The second-order perturbation function[14], which is designed to tease out short-term variations, is unremarkable. Finally, the power spectrum of the contour (lower right) shows no prominant frequencies.

In contrast, the F_0 variations for dysphonic patients were sometimes bi-modal or multi-modal due to the presence of subharmonics or low-frequency modulations. Figure 2 shows the frequency profile for a male patient with a hyperfunctional disorder of moderate severity. The contour shows an alternating sequence of high and low F_0 (about 145 Hz to 155 Hz), a common manifestation of period-doubling. This can be confirmed by observing the strong frequency component at 76 Hz (half of an average F_0 of 152 Hz) in the power spectrum. In the histogram, the period-doubling appears as a bi-modal distribution. As a third example (Figure 3), an intermittent, low-frequency modulation of about 65 Hz appears in the fundamental frequency contour. This contour is from a female patient with a hyperfunctional childhood dysphonia, again of moderate severity. The manifestation of low-frequency modulations is not unique to functional disorders; such structures were found in all the dysphonic categories. Rather, the appearance of complex patterns was related to the severity of the disorder: they were most often observed in dysphonia of moderate severity. The contours of non-dysphonic and mildly dysphonic patients generally did not contain substantial low-frequency structures, although subtle drifts, trends, and other nonstationarities were found in these groups as well.

Severe dysphonia usually contained strong nonstationarities. Consequently, if well-defined structures did exist within these dysphonia, they were often masked by these nonstationarities or noise. An example of this is shown in Figure 4, which is a fundamental frequency profile of a male patient with a severe neoplastic dysphonia. In this figure, the 100 Hz frequency component (close to period-tripling) is nearly masked by a substantial "noise" floor in the contour power spectrum.

Summary

Only a small sampling of the F_0 distribution patterns have been analyzed to date. It appears from this sampling that the fundamental frequency variations in functional dysphonia differed from normals, but were similar to the variations found in other pathological groups. Well-defined subharmonic and low-frequency structures were common in all dysphonic types of moderate severity. In this study, however, functional dysphonia had the greatest percentage of patients which fell into the range of moderate severity. Thus, subharmonic and low-frequency modulations appear to be at least as characteristic of functional dysphonia as of other types of dysphonia.

Acknowledgement

This study was supported, in part, by NIDCD grant No P60 DC00976. The authors thank Julie Lemke for manuscript preparation.

References

1. Murphy AT. Functional voice disorders. Foundations of Speech Pathology Series, 1964.

2. Froeschels, E. Laws in the appearance and the development of voice hyperfunctions. Journal of Speech Disorders, 1940; V:1-4.

3. Greene, JS. Functional speech and voice disorders. J. Nerv. Ment. Dis., 1942; 95:299.

4. Perello, J. La theorie muco-ondulatoire de la phonation. Ann. D'otolaryngol., 1962; 79:722-725.

5. Damste PH, Lerman JW. An introduction to voice pathology. Functional and Organic. Charles Thomas Publisher, 1975.

6. Kotby, MN. Voice disorders: recent diagnostic advances. Egypt. J. Otolaryngol., 1986; 3(1):69-98.

7. Klingholz F, Martin F. Quantitative spectral evaluation of shimmer and jitter. JSHR, 1985; 28:169-174.

8. Ludlow CL, Celia JB, Connor NP, Coulter DC, Lee YJ. The validity of using phonatory jitter and shimmer to detect laryngeal pathology. Laryngeal Function in Phonation and Respiration, Thomas Baer, Clarence Sasaki, and Katherine Harris (eds.), College-Hill Press, 1985.

9. Scherer RC, Gould WJ, Titze IR, Arlen DM, Sataloff RT. Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. J. of Voice, 1988; 2(3):230-244.

10. Hirano M. Objective evaluation of human voice. Clinical aspects. Folia Phoniatrica, 1989; 41:89-144.

11. Hirano M. Clinical examination of voice. Springer-Verlag, Wein, 1981.

12. Ramig LO, Scherer RC, Titze IR, Horii Y. Acoustic analysis of voice in amyotrophic lateral sclerosis: a longitudinal case study. JSHD, 1990; 55:2-4.

13. Titze IR. A model for neurologic sources of aperiodicity in vocal fold vibration. JSHR 1991; 34: 460-472.

14. Pinto N, Titze IR. unification of perturbation measures in speech analysis. JASA 1990; 87(3); 1278-1289.

NCVS Status and Progress Report - 4
June 1993, 209-227

A Preliminary Study on Two Methods of Treatment for Laryngeal Nodules

Katherine Verdolini-Marston, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
and
Department of Music, The University of Iowa
Mary Katherine Burke, M.F.A.
Old Globe Theatre, San Diego
Arthur Lessac
Lessac Institute, Santa Monica
Leslie Glaze, Ph.D.
Midwest Voice and Speech Institute, Saint Paul
Elizabeth Caldwell
Department of Speech Pathology and Audiology, The University of Iowa

Abstract

This experiment represents a preliminary stage of inquiry in a series of planned studies. Adult females with nodules participated as paid subjects. Some subjects received "confidential voice therapy", some received "resonant voice therapy", and some received no therapy (control condition) over an approximately two-week period. Pre- and post-measures provided some evidence of a benefit from therapy. Baseline measures were then repeated two weeks after therapy was terminated, for therapy subjects. The final results indicated that compliance scores (which reflected the degree to which subjects appeared to use the therapy technique outside the clinic) were relevant for predicting the likelihood of benefitting from therapy. However, the specific type of therapy that was administered was not. The primary importance of these findings is that they point to a need to assess compliance factors in future efficacy studies. The findings also emphasize the critical role of compliance in clinical voice practice, in general.

Voice therapy is generally viewed as an important aspect of treatment for laryngeal nodules, either alone or in conjunction with microsurgery. Some empirical support for this view has been reported over the past several years (see for example, Kotby, El-Sady, Basiouny, Abou-Rass, & Hegazi, 1991; Lancer, Syder, Jones, & Boutillier, 1988; Murry & Woodson, 1992). Large-scale, controlled studies are still needed. The purpose of the present study was to obtain preliminary information about the effectiveness of two therapy modalities, "confidential voice therapy" and "resonant voice therapy", for the treatment of nodules. The information would be useful in planning future, larger studies. In this preliminary experiment, the focus was the role of therapy type (confidential voice versus resonant voice therapy) and of compliance (the extra-clinical utilization of the voicing technique taught in therapy) in predicting the likelihood of a benefit therapy. If compliance factors turned out to be relevant, post-hoc information would be used to assess what factors might affect compliance.

General Description of Therapy Types

Confidential voice therapy is often utilized by speech/language pathologists for the treatment of nodules. According to Colton and Casper (1990), the focus of this type of therapy is the elicitation of a minimal intensity, low effort, and somewhat breathy phonation mode, as if speaking confidentially at close range. Videoscopic examinations show that at least breathy phonation is associated with a persistent glottal gap during phonation, at variable locations along the anteroposterior dimension of the folds (see Casper, Colton, Woo, & Brewer, 1989, 1990). Confidential voice therapy is considered particularly useful during initial stages of treatment, for maximum early reduction of lesions and of dysphonia.

Resonant voice therapy is perhaps more regularly encountered in performance domains such as theatre and classical music. As described by various authors, including Cooper (1973) and Lessac (1967), resonant voice involves vibratory sensations on the alveolar ridge and other facial plates during phonation. These sensations apparently arise from a relative acoustic "tuning" of the supraglottic cavities to the glottal source spectrum (see for example, Raphael & Scherer, 1987). In part related to the tuning phenomenon, voice output is not necessarily quiet with this approach. Our videoscopic observations and those from other clinics (for example, Rammage, personal communication) indicate that resonant voice is generally produced with relatively complete anteroposterior vocal fold closure during phonation. The likely reason is that more or less complete (and rapid) vocal fold closure is a prerequisite for the production of a sufficient harmonic series, for availability to supraglottic enhancement (see for example, Gauffin & Sundberg, 1989; Titze & Sundberg, 1992).

Framework for Basic Experimental Question

As noted, the focus in this initial study regarded the role of therapy type and of compliance in predicting the likelihood of a benefit from therapy. This question is related to the proposal that, to be effective, these or any voice therapy methods should be sound from at least three perspectives: a physiological perspective, a learning perspective, and a compliance perspective (Verdolini-Marston, Burke, Lessac, Glaze, & Caldwell, 1992). That is, the method must emphasize a physiology that has the potential for reversing pathogenic voicing patterns. The method must also be "learnable". And finally, the method must be one that subjects are likely to utilizeoutside the clinical situation. How can the two methods under investigation be considered, from these three perspectives?

Regarding the physiological perspective, the results from a recent study by Peterson, Verdolini-Marston, Barkmeier, and Hoffman (submitted) provided information about both "breathy voice" (similar to confidential voice, in the present study) and "resonant voice". In that study, both of these voicing modalities were produced with a generally abducted laryngeal posturing, as compared to "pressed voice" (a presumed pathogenic voicing mode). Breathy voice resulted in a frank abductory posture, and resonant voice resulted in a slightly abducted or barely adducted posture. "Pressed voice" resulted in a clearly hyperadducted laryngeal posturing. Thus, lesion reduction might be favored with both confidential and resonant voice therapy approaches, as compared to pressed voice (which for many patients may represent a pathogenic factor), because of a direct relation between vocal fold adduction level and intraglottal impact force (Jiang & Titze, 1992), and further, between intraglottal impact force and pathogenesis (see also Jiang & Titze, 1992). Due to the relatively greater abduction associated with confidential as compared to resonant voice, the confidential voice therapy method might be physiologically superior in its potential for reversal of lesions.

From the learning perspective, there is no obvious a priori reason that either the confidential voice or the resonant voice therapy approach should be more effective than the other. A learning model that honors skill acquisition principles, such as the relevance of attention to perceptual information (Nissen & Bullemer, 1987; Verdolini-Marston & Balota, submitted), consistent repetitions (Schneider & Fisk, 1982), and knowledge of results (see Schmidt, 1988, for review), could be used with either approach, as could a learning model that violates the same principles. The effectiveness of either type of therapy would be anticipated to vary with the learning model, to some degree.

From a compliance perspective, there are also no clear-cut predictions. Health-care studies indicate that compliance with therapeutic regimens may depend on a series of factors, including "self-efficacy", or the patient's perception of his/her ability to carry out the prescribed behavior (for a discussion of "self-efficacy", see Bandura, 1977; see also Poll & De Nour, 1980). In the context of voice therapy, self-efficacy would involve the patient's perception of his/her technical ability to produce the type of voicing trained. From this viewpoint, confidential voice therapy might result in greater compliance than resonant voice therapy, because of the informal impression that most patients are able to produce confidential voice to the conversational level with little training, as compared to resonant voice, which tends to require more training. However, another consideration might offset this possible advantage, for confidential voice therapy. By definition, confidential voice involves quiet phonation. Thus, patients might fail to utilize this type of voicing in situations requiring more normal loudness levels. In contrast, patients might more readily utilize the resonant voice technique even in situations requiring louder phonation, because this production mode allows for relatively loud output.

In summary, from a physiological perspective, confidential voice therapy might be expected to result in greater reductions in laryngeal nodules and in dysphonia than resonant voice therapy, although both may have the potential for producing benefits. From the learning and compliance perspectives, less confident predictions can be made. The experiment reported here focuses on the role of physiological factors (therapy type) and of compliance factors (extra-clinical utilization of the therapy technique) in determining relatively longer-term benefits with therapy. Specifically, first the outcomes of confidential and resonant voice therapy are assessed as compared to the outcome for a control group, following a two-week intensive therapy period. To address the question of fundamental interest, the likelihood of a relatively longer-term benefit with therapy is then assessed as a function of therapy type and compliance measures. If the results indicate a role of compliance, a secondary, post-hoc question regards what factors might in turn affect compliance.

Methods

Subjects. The original subject group consisted of 18 adult females who were recruited from college sororities to participate in the study for pay. Sororities were considered optimal recruiting environments because of the perceived high density of voice disorders among this population at our university. Although all 18 subjects did participate in the protocol, only 13 yielded a complete set of utilizable data.[1] These are the subjects described in this report. The average age of this subject set was 20 years, with a range of 18-22 years. Subsequent to their recruitment and before initiation of the protocol, a diagnosis of laryngeal nodules was confirmed for each subject by an otolaryngologist. The average time since onset of symptoms was 4.5 years, with a range of .33 years (4 months) to 14 years. Therefore, nodules were considered chronic for all subjects. None of the subjects had received prior voice therapy, and at the outset of the experiment all denied previous voice training of any type. By self-report, none of the subjects had ever smoked regularly, although five smoked sporadically, socially (Subjects MS, JO, KG, LP, and JR). For these subjects, the number of pack years, or packs per day times number of years, ranged from .04 to 1.14 years. With the exception of one subject, all denied active illnesses, ongoing medical conditions, or regular medications. The exception was Subject MR, who took .125 mg of Synthroid daily for control of hypothyroidism. Hearing screening indicated pure tone thresholds of 20 Db or lower in the better ear, for 11 of the 13 subjects. For the remaining two subjects, thresholds were not obtained, but there was no indication of hearing loss for these subjects.

Therapists. As an attempt to control for therapist effects, two therapists each provided both confidential voice therapy and resonant voice therapy to different subjects. Both therapists were nationally certified in speech/language pathology, with membership in the American Speech-Language-Hearing Association. One was a doctoral-level speech/language pathologist (LG) with about five years of experience in voice therapy. She had prior experience with the confidential voice therapy approach, and expressed confidence in its effectiveness for the treatment of nodules. Before her participation, she had no experience with the resonant voice method. However, she considered that it might also be effective. The other therapist (AL) had practiced as a voice trainer and therapist for about 60 years, and had developed the resonant voice approach used in the protocol (the Lessac approach). Not surprisingly, he believed that resonant voice therapy was effective in the treatment of nodules. Prior to his participation in the study, he had no experience with the confidential voice therapy approach, and he was frankly skeptical about its therapeutic potential.

Cross-training of therapists and general description of therapies provided. For two days immediately preceding the initiation of therapy, the two therapists engaged in cross-training for confidential voice and resonant voice therapy, for a total of about 7 hours. On the first day of cross-training, the therapists worked with each other and with another experimenter for about 3-4 hours. The next day, five patients (who were not involved in the data set reported here) were brought in for the therapists to work with, with reciprocal feedback, for another 3-4 hours.

The confidential voice therapy method was modelled after descriptions by Colton & Casper (1990). For this type of therapy, the focus was (a) the production of a minimal effort, minimal

[1] One subject developed the flu and could not come in for the final measurement session. Three subjects' videoscopic views of the larynx were altered during dubbing procedures, or were unavailable, and one subject's audio tapes were unavailable. No subject was excluded on the basis of the results obtained. The data for excluded subjects were carefully examined, and it was determined that, as best we can estimate, the pattern of results reported in this study was not qualitatively affected by their elimination.

loudness, and slightly breathy phonation mode, as if speaking confidentially at short range, without perceptible perilaryngeal "squeezing", and (b) general body relaxation, obtained for example by shaking out the jaw and shoulders and by breathing easily. The resonant voice therapy method was modelled after descriptions by Lessac (1967). For this type of therapy, the focus was (a) the production of concentrated vibratory sensations on the anterior palate during phonation, using an "inverted megaphone" facial posture (slightly expanded pharynx and a slight forward stretch in facial muscles, with labial protrusion), and (b) upper body relaxation, using manual manipulations to reverse any obvious head, neck, or shoulder tensions, and to obtain good head and neck alignment.

In addition to these distinguishing characteristics, common aspects across the two therapy types were also emphasized. Therapists were instructed to display an overall attitude of confidence in both methods, and to characterize both methods to subjects as well-established and successful clinical approaches that are quite easy to implement. Also for both methods, therapy was to start with work on sounds in isolation (typically vowels for the confidential voice approach, and both consonants and vowels for the resonant voice approach). Then therapy was to proceed to the conversational level and real-life applications outside the therapy room, as appropriate. Finally, for both methods, exercises were to be as interesting as possible, numerous repetitions of target behaviors were to be included, and frequent positive feedback about subjects' performance was to be provided.

All therapy sessions were videotaped for possible later reference, and an experimenter observed most sessions, on-line, as well. Informally, these observations indicated that the therapies were largely delivered as trained. In particular, although there may have been different biases about the effectiveness of the methods across the two therapists, observations of therapy sessions did not reveal any obvious evidence of such biases.

Procedures and design. When the therapists had completed cross-training, subjects initiated their participation in the protocol. Subjects first received general (logistical) information about the protocol. Then they were instructed to observe general voice hygiene practices. That is, subjects were instructed to limit alcohol and caffeine intake, smoking, and "heavy voice use" (for example, yelling). They were provided daily logs to indicate their adherence to these directives. After they received these instructions, subjects underwent a series of voice and laryngeal examination procedures described shortly. All test procedures were conducted by an experimenter who was uninformed about which therapy treatment subjects were about to receive, for subjects who received therapy. (For logistical reasons related to the therapists' availability, it was necessary to run control subjects in the protocol some time after therapy subjects had completed their participation. Thus, the experimenter was aware of control subjects' status as controls.)

Subjects who would receive therapy were then assigned to a therapy group. In the original subject set, six subjects were assigned to the confidential voice therapy group (three subjects were assigned to LG and three were assigned to AL), and six subjects were assigned to the resonant voice therapy group (again, three subjects were assigned to LG and three were assigned to AL). The assignments were made on the basis of severity ratings made prior to the initiation of the protocol, from videoscopic views of the larynx with audio, such that each therapy x therapist cell in the experiment would be comprised of subjects with the same approximate degree of hoarseness and nodule severity. (Control subjects were also subsequently selected so that the severity distribution would be about equivalent for this group, as compared to the two therapy groups.) However, as already noted, the data for some of the subjects were subsequently excluded from analysis because of missing data: five subjects were retained for the confidential therapy group, of which two received therapy from LG and three from AL, and three subjects were retained for the resonant therapy group, of which two received therapy from LG and one from AL. (Five subjects were retained for the control group.) The result was a data set that regrettably reflected different initial severity distributions across the groups.

For therapy subjects, therapy was then initiated the same day, or the following day. (Control subjects were simply asked to follow general voice hygiene measures, as instructed, and to return for follow-up measurements 12 days later.) Therapy subjects received a total of nine therapy sessions across an approximately 12-day period. For each subject, eight of the sessions were one-hour individual sessions, and one was a one-hour group session together with all other subjects in the same therapy condition (confidential or resonant voice).

The rationale for administering therapy over an approximately two-week period was related to the preliminary nature of the present study. Usually, voice therapy is administered over a longer time-span, with one or two sessions per week. For the present, initial inquiry, we elected to support a briefer, more intensive protocol. We anticipated that reliable changes in voice and in the larynx could be obtained within the two-week time-period, based on the findings from another efficacy study that we conducted recently, on hydration treatments (Verdolini-Marston, Sandage, & Titze, in press).

Following the final therapy session, i.e. about 12 days following therapy initiation, therapy subjects underwent the same voice and laryngeal testing procedures as previously. They also filled out questionnaires that assessed their perceptions about therapy. Control subjects received post-tests of the voice and of the larynx (but did not fill out questionnaires) at the same time interval. Therapy subjects then returned two weeks later for a final measurement session, so that their relatively longer-term status could be assessed. Control subjects did not return for a later follow-up.

A final note is that voices and larynges tended to vary with day of the week in our subject population. Status tended to be relatively worse following weekends, and relatively better following the work week. To control for this factor, and to maximize the likelihood of detecting changes in this initial experiment, we administered all pre-treatment measures earlier in the week (mostly on a Sunday or Monday), and all post-treatment measures later in the week (mostly on a Friday). As noted, gains would be maximized by this approach. However, they would be maximized equivalently across groups. Thus, any group differences would be attributable to the independent variables of interest.

Measures and equipment. The primary pre- and post-measures of the voice and of the larynx consisted of (a) phonatory effort ratings, (b) auditory-perceptual ratings of voice, and (c) visual-perceptual ratings of the larynx. Acoustic measures of voice (jitter, shimmer, and signal-to-noise ratio) and aerodynamic measures (phonation threshold pressure) were also made, as supportive measures. However, these measures are not reported here because their inclusion would increase the number of subjects that would have to be excluded because of missing or uninterpretable data.[2] Other critical measures in the experiment, reported here, were (d) estimates of adherence to instructions about general voice hygiene practices (for all subjects), (e) measures of subjects' perceptions about therapy methods (for therapy subjects), and (f) estimates of ongoing compliance following

[2] Interpretable acoustic and aerodynamic data were not obtained for one control subject, for the baseline measurement session, and interpretable acoustic data were not obtained for another control subject, for the two-week measurement session. As best we can estimate, the overall pattern of results was not qualitatively affected by the omission of acoustic and aerodynamic data. Specific information is available from the first author on request.

therapy termination, i.e. of the relative continued utilization of the therapy technique after therapy was discontinued (for therapy subjects).

Measures of phonatory effort were based on a procedure originally described by Colton and Brown (1972) and Wright and Colton (1972), that we modified for this and other experimental and clinical procedures. For these measures, subjects were required to rate the perceived effort of phonation, in general, on a magnitude estimation scale on which "100" indicated "a comfortable amount of effort during phonation", 200 indicated "twice as much effort as comfortable" (i.e. "very effortful"), and "50" indicated "half as much effort as comfortable" (i.e. "very easy").

Auditory-perceptual ratings of the voices were based on tape recordings made during the reading of a standard passage, "A Man and his Boat." Recordings were made in a quiet room, using an AKG C 460 B condenser microphone powered by a Symetrix SX202 Dual Mic pre-amplifier. Signals were routed to a Realistic STA-785 Digital Synthesized Am/Fm Stereo Received/Amplifier and a Panasonic Digital Audio Tape (DAT) Deck SV-3500. During recordings, the microphone-to-mouth distance was approximately 7 in, and recording levels were monitored for an approximately constant value on the VU meter of the DAT recorder. After the experiment was completed, three recordings from each subject were played successively to four listener-judges (speech pathologists) independently, in free-field, over Realistic Minimus 7 speakers. For subjects who received therapy, the recordings included one segment made just prior to the initiation of therapy, one made following therapy, about 12 days later, and one made about two weeks following the termination of therapy. For control subjects, who did not receive therapy, the recordings included one segment from before participation, one segment from 12 days later, and then a repeat of one of the other two segments. (A repeat was included for control subjects so that three segments would be presented, consistent with the number presented for therapy subjects. The presentation of only two segments might have provided a "clue" about control subjects' status as somehow different.) Across subjects, the order of presentation of the three segments was counterbalanced to control for order effects. All of the listener-judges reported normal hearing, they were uninvolved in any other phase of the experiment, and they were unfamiliar with the subjects and subjects' specific conditions (pre- versus post, and confidential versus resonant voice versus control condition). The judges independently rated each taped segment on an ordinal scale, on which 1 = "healthy voice", 2 = "mildly impaired voice", 3 = "moderate", 4 = "moderate-severely impaired voice", and 5 = "severely impaired voice."

Visual-perceptual measures of the larynx were based on views of the larynx obtained with an R. Wolf 4450.47 90 degree rigid videoscope, connected to a Karl Storz 9000 Mini Solid State CCD Video Camera and a Brüel and Kjaer Rhino-Larynx Stroboscope light source, Type 4914. Views were monitored on a NEC autocolor PM-1971A color television monitor at the time of collection. After all data were collected, videoscopic segments were played back using a SONY CVM Trinitron video monitor. Four independent judges (two speech pathologists with experience in videoscopic evaluation of the larynx and two otolaryngology residents at our institution) were presented with three, approximately 5-sec video segments in succession. For subjects who received therapy, the recordings included one segment made just prior to the initiation of therapy, one made following the termination of therapy, about 12 days later, and one made about two weeks following termination of therapy. For control subjects, who did not receive therapy, the recordings included one segment from before participation, one segment from 12 days later, and also a repeat of one of the other two segments. (Again, a repeat was included to avoid a distinction in the number of segments presented, as compared to the number presented for therapy subjects.) The order of segments was counterbalanced across subjects. All 5-second segments were selected as "the best 5-second view of the larynx" from longer, taped segments, by a research assistant who was not involved in any other aspect of the experiment, and without regard to treatment condition. Also none of the judges who rated the segments were involved with any other phase of the experiment, they were unfamiliar with the subjects, and they were uninformed about subjects' treatment conditions upon viewing. Each judge independently rated each video segment on a 5-point ordinal scale, on which 1 = "healthy vocal folds", 2 = "mild nodules/polyps or related", 3 = "moderate", 4 = "moderately severe", and 5 = "severe nodules/polyps or related". The term "related" was intended to refer to characteristics that may accompany nodules, for example erythema.

For all subjects, daily logs indicated subjects' adherence to the instructions about general voice hygiene practices during the initial two-week period of the protocol. For therapy subjects, measures of subjects' perceptions about therapy were obtained using a questionnaire that was administered upon therapy termination. The questionnaire included a set of 12 questions, that we thought might be in some way interesting in the subsequent interpretation of the results. At the time they were formulated, the questions were motivated more by practical than by rigorous theoretical considerations. The questions are indicated in Appendix A. Subjects responded to all questions with ordinal ratings from "1" (maximum negative response) to "5" (maximum positive response).

Finally, for therapy subjects, estimates of ongoing compliance (continued utilization of therapy techniques following therapy discontinuation) were made on the basis of responses to an open-ended question: "To what extent have you continued to use [the therapy technique] [since therapy was terminated]?" Subjects' responses were literally transcribed by one of the experimenters. For scoring purposes, the transcriptions were later dichotomized by a clinician who was uninvolved in any other aspect of the protocol. A score of "0" indicated a report of early discontinuation of the therapy method following therapy termination, and a score of "1" indicated a report of at least some continued use of the method. Dichotomized compliance scores were used because they were considered to provide the appropriate level of differentiation in this preliminary study.

Results and Discussion

<u>Adherence to general voice hygiene instructions</u>. Table 1 indicates that on most parameters, subjects in all groups appeared to follow instructions about general voice hygiene practices to about the same degree, at least on the basis of daily logs that they filled out. On average, subjects in all groups drank about one alcoholic beverage per day (although for all subjects, drinking tended to cluster on weekends, without drinking during the week). Subjects drank about one or one-and-a-half cups of a caffeinated beverage per day, and they did not smoke or smoked an average of less than a third of a cigarette per day. The primary discrepancy between the groups appeared to come from amount of "heavy voice use". Subjects in the confidential voice therapy group appeared to restrict heavy voice use the most (to about 0.1 hr/day, on average), subjects in the resonant voice group were intermediate (about 0.3 hr/day), and subjects in the control group appeared to engage in heavy voice use the most (about 0.70 hr/day). The relatively high value for the control group was not necessarily attributable to the lack of participation in therapy, for all subjects: excluding Subject 85, who reported an average of 2.62 hr of heavy voice use per day, the average "heavy voice use" for the control group was about 0.2 hr/day (less than for the resonant voice therapy group).

Table 1.
Subjects' Logs on Compliance with General Voice Hygiene Practices, Over First Two-Week Period of Protocol. Alcohol (average drinks/day), Caffeine (average cups/day), Smoke (average cigarettes/day), and Heavy Voice Use (average hours/day). Number of Days That Subjects Filled Out Logs Indicated in Parentheses.

Group/Subject	Alcohol	Caffeine	Smoke	Heavy Voice Use
Confidential				
MS (13)	1.31	0.15	0.92	----
BB (12)	1.67	2.08	----	0.25
JO (11)	1.09	1.18	0.64	----
MR (12)	----	1.75	----	0.08
SN (09)	1.11	2.56	----	0.33
Average	1.04	1.54	0.31	0.13
Resonance				
CS (07)	1.00	0.29	----	0.57
SC (07)	----	0.07	----	0.29
SD (12)	1.25	2.08	----	0.14
Average	0.75	0.81	0.00	0.33
Control				
WK (13)	2.27	0.77	----	0.28
AH (13)	----	1.00	----	2.62
KG (12)	1.25	4.33	1.33	0.23
LP (13)	0.54	0.92	----	0.08
JR (12)	0.58	0.42	----	0.25
Average	0.93	1.49	0.27	0.69

Overall effect of therapy. The first question was whether therapy produced any benefit, as compared to the control condition. There was some evidence that it did. The raw data are displayed in Tables 2-4. Because of different severity distributions across the groups in this initial study, the data were primarily evaluated non-parametrically. That is, for each subject, a "+" score was used to indicate a numeric improvement in the phonatory effort measure, in the auditory-perceptual rating of voice, and in the visual-perceptual rating of the larynx, from baseline to the two-week follow-up. Tables 2-4 also show these "+" values, as well as the proportion of therapy and control subjects who improved on each measure. The same proportions are indicated graphically in Figure 1.

Combining all therapy subjects (confidential and resonant voice therapy subjects), effort measures improved for 6 of 8 subjects across the initial two-week period (P = .75), auditory-perceptual ratings improved for 5 of 8 subjects (P = .63), and visual-perceptual ratings improved for 6 of subjects (P = .75). In contrast, for the control group, effort measures improved for only 1 of 5 subjects (P = .20), auditory-perceptual ratings similarly improved for only 1 of 5 subjects (P = .20), and visual-perceptual ratings improved for 3 of 5 subjects (P = .60). Thus, a greater proportion of therapy subjects improved on all three measures, as compared to the proportion who improved in the control condition. At this level, there was some evidence of an overall benefit from therapy. However, using a binomial test, this result must be considered a trend only, because of a statistical floor effect: the lowest possible p-value was obtained, with three "successes" (superior results for the therapy groups) out of three "events" (measures), p = .13.

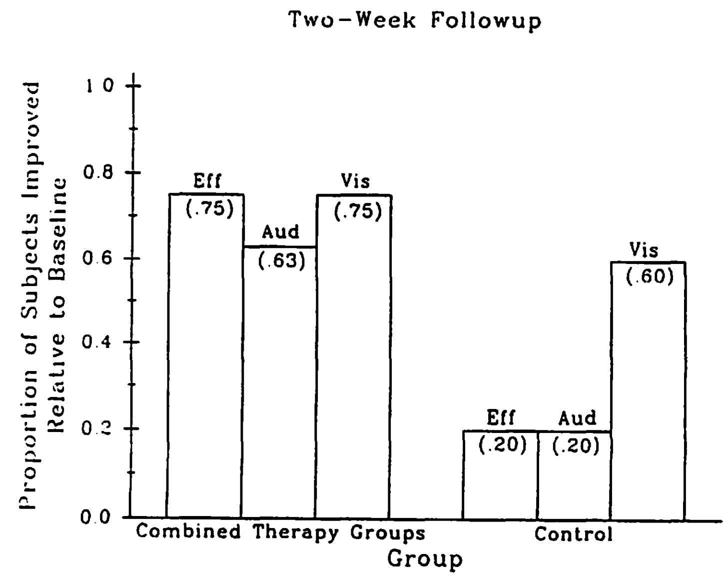


Figure 1. Proportion of therapy and control subjects improved on effort measures, auditory-perceptual ratings of voice, and visual-perceptual ratings of the larynx, from baseline to two-week follow-up.

Returning to the raw data for a moment, it is noteworthy that control subjects appeared to deteriorate markedly in phonatory effort over the initial two-week period. This finding points to the possibility that sometimes, symptoms of laryngeal nodules may be degenerative, without treatment.

Effect of therapy type versus compliance. The primary question of interest regarded the role of therapy type and of compliance (extra-clinical utilization of the therapy technique following therapy discontinuation) in predicting the likelihood of relatively longer-term improvements with therapy. The relevant raw data are again displayed in Tables 2-4, and also in Table 5. From baseline to two-week follow-up, 4 of 5 subjects in the confidential voice therapy group improved on effort measures (P = .80), 2 of 5 improved on auditory-perceptual ratings (P = .40), and 4 of 5 improved for visual-perceptual ratings (P = .80). From baseline to four-week follow-up, for the same group, 3 of 5 subjects improved on effort measures (P = .60), 3 of 5 improved on auditory-perceptual ratings (P = .60), and 2 of 5 improved on visual-perceptual ratings (P = .40). For the resonant voice therapy group, from baseline to two-week follow-up, 2 of 3 subjects improved on effort measures (P = .67), 3 of 3 subjects improved on auditory-perceptual ratings (P = 1.00), and 2 of 3 improved on visual-perceptual ratings (P = .67). From baseline to four-week measures, 2 of 3 subjects improved on effort measures (P = .67), 2 of 3 improved on auditory-perceptual ratings (P = .67), and 2 of 3 also improved on visual-perceptual ratings (P = .67). Considering these results together, at the simplest level, there appeared to be a slight overall advantage for the resonant therapy group at both two- and four-week measures. Table 5 shows compliance scores. Three of 5 subjects in the confidential voice therapy group reported some continued utilization of the therapy technique following therapy discontinuation. Three of 3 subjects in the resonant voice therapy group reported some continued utilization.

Table 2.
Effort Measures. Plus Sign Indicates Improvement Relative to Baseline, and P Refers to Proportion Improved.

Group/Subject	Baseline	Two-Week Measures	Four-Week Measures
Confidential			
MS	100	100	100
BB	100	50 +	75 +
JO	140	80 +	45 +
MR	110	100 +	100 +
SN	125	120 +	163
Average	115	90 (P=.80)	97 (P=.60)
Resonance			
CS	125	90 +	75 +
SC	200	125 +	115 +
SD	100	100	100
Average	142	105 (P=.67)	97 (P=.67)
Combined Therapy Groups Averages	125	96 (P=.75)	97 (P=.63)
Control			
WK	100	200	
AH	100	200	
KG	125	150	
LP	150	100 +	
JR	100	100	
Average	115	150 (P=.20)	

The critical statistical analyses involved three separate Analyses of Variance, one for effort measures, one for auditory-perceptual ratings, and one for visual-perceptual ratings. For each measure, simple dichotomous improvement scores were used for each subject (1 = improvement, 0 = no improvement), and dichotomous compliance scores were also used (1 = reported continued utilization of the therapy technique, 0 = reported early discontinuation of the technique). Despite the trend noted for a slightly better outcome following resonant therapy, the results of these analyses failed to confirm a main effect of therapy type, for any of the measures ($p = .81$ for effort measures,

Table 3.
Auditory-Perceptual Ratings. Plus Indicated Improvement Relative to Baseline, and P Refers to Proportion Improved.

Group/Subject	Baseline	Two-Week Measures	Four-Week Measures
Confidential			
MS	2.28	2.93	2.65
BB	3.03	2.65 +	1.88 +
JO	3.00	3.75	2.38 +
MR	2.08	1.63 +	1.88 +
SN	2.48	2.95	3.00
Average	2.57	2.78 (P=.40)	2.36 (P=.60)
Resonance			
CS	2.20	2.13 +	1.45 +
SC	4.50	3.05 +	2.63 +
SD	2.48	2.20 +	2.93
Average	3.06	2.46 (P=1.00)	2.34 (P=.67)
Combined Therapy Groups Averages	2.76	2.66 (P=.63)	2.35 (P=.63)
Control			
WK	3.20	2.50 +	
AH	2.05	2.88	
KG	3.33	3.33	
LP	1.20	1.60	
JR	1.27	1.50	
Average	2.21	2.36 (P=.20)	

$p = .81$ for auditory-perceptual measures, and $p = .51$ for visual-perceptual measures). However, the main effect of compliance scores was reliable, for effort measures ($p = .03$) and for auditory-perceptual measures ($p = .03$). A weak trend towards a compliance effect was also shown for visual-perceptual measures, but this trend was statistically unreliable ($p = .22$). The interaction between therapy type and compliance was also unreliable, for all of the measures. Finally, although there was a trend for resonant therapy subjects to have more consistently good compliance scores, this trend was not confirmed with a correlational measure ($r = .45$, $p = .27$).

Table 4.
Visual-Perceptual Measures. Plus Indicates Improvement Relative to Baseline, and P Refers to Proportion Improved.

Group/Subject	Baseline	Two-Week Measures	Four-Week Measures
Confidential			
MS	1.63	1.43 +	1.88
BB	2.50	2.18 +	2.43 +
JO	2.23	2.13 +	1.88 +
MR	1.75	1.25 +	1.88
SN	1.50	1.50	2.08
Average	1.92	1.70 (P=.80)	2.03 (P=.40)
Resonance			
CS	2.25	3.58	2.78
SC	3.63	2.45 +	2.25 +
SD	3.00	2.50 +	2.63 +
Average	2.96	2.84 (P=.67)	2.55 (P=.67)
Combined Therapy Groups Averages	2.31	2.13 (P=.75)	2.23 (P=.50)
Control			
WK	2.38	2.33 +	
AH	2.00	2.13	
KG	4.13	4.13	
LP	2.93	2.88 +	
JR	2.33	1.63 +	
Average	2.75	2.62 (P=.60)	

Based on these results, compliance scores appeared to partially predict the likelihood of longer-term improvements in phonatory effort and in auditory-perceptual ratings with therapy, but therapy type did not. A trend in the same direction was obtained for visual-perceptual measures, but this result was not confirmed statistically.

Factors that may have affected compliance. Table 6 shows individual subjects' responses to questions about therapy, when therapy was terminated. As noted in this table, subjects' responses tended to be skewed in a positive direction, as is often the case for questionnaire data of this type. Table 6 also displays the correlations between subjects' responses about therapy and compliance scores, already indicated in Table 5. At a numeric level, the strongest correlations were obtained for the questions about how interesting therapy was, and about how technically able subjects thought they were in producing the voice type trained (r = .58 and .52, respectively). However, even these correlations failed to confirm reliable relations (p = .13 and .19). Skewed questionnaire data undoubtedly contributed to relatively small correlational values.

General Discussion

The findings from the present study are preliminary, and are mainly primarily useful for planning future, larger N studies. The primary contribution is that the need is emphasized for considering compliance factors in future studies. Compliance measures (reported extra-clinical utilization of the

Table 5.
Subjects' Scores for Continued Utilization of Technique Following Therapy Termination. (0 = Early Reported Discontinuation; 1 = Reported Continued Use.)

Group/Subject	Score
Confidential	
MS	0
BB	1
JO	1
MR	1
SN	0
Resonance	
CS	1
SC	1
SD	1

Table 6.
Subjects' Responses to Questionnaire Regarding Perceptions about Therapy (See Appendix A; 1= Maximum Negative Response, 5 = Maximum Positive Response).
Correlations with Compliance Scores Also Indicated (r), as well as Significance Levels (p).

	QUESTIONS											
Group/ Subject	Overall Improvement	Like Therapy	Voice Easier	Voice Clearer	Accept Voice Type	Can be Yourself	Therapy Fun	Therapy Interesting	Able to Use Technique	Willing to Use Technique	Predicted Future Use	Recomm. Therapy to a Friend
Confident.												
MS	5	4	5	4	4	4	4	4	3	4	4	4
BB	4	4	4	3	4	3	5	5	3	3	3	5
JO	5	5	4	4	4	5	4	5	5	5	4	5
MR	3	5	4	4	3	2	4	4	3	3	4	4
SN	3	4	3	4	3	3	3	4	3	4	3	5
Average	4.0	4.4	4.0	3.8	3.6	3.4	4.0	4.4	3.4	3.8	3.6	4.6
Resonance												
CS	4	4	4	4	4	3	3	4	4	3	3	4
SC	4	4	4.5	4	3	4	3	5	4	4	3	5
SD	4	5	5	4	4	4	5	5	4	5	5	5
Average	4.0	4.3	4.5	4.0	3.7	3.7	3.7	4.7	4.0	4.0	3.7	4.7
$r_{compliance}$	.00	.45	.22	-.22	.15	.00	.28	.58	.52	-.09	.10	.15
p	1.00	.27	.60	.60	.72	1.00	.51	.13	.19	.83	.81	.72

therapy technique following therapy termination) were relevant for predicting the likelihood of a relatively long-term benefit from voice therapy, for phonatory effort measures and for auditory-perceptual measures of the voice. (For visual-perceptual measures of the larynx, a similar trend was obtained, but the trend was not confirmed statistically.) In contrast, there was little evidence that either confidential voice therapy or resonant voice therapy produced a greater likelihood of a benefit. Stated differently, both types of therapy examined had about the same likelihood of producing improvements, provided that subjects actually applied the therapy method outside of the clinic.

The finding of "no difference" in the effectiveness of the two therapy types is not a confident one, at this point. The obvious reasons are that therapy was administered to too small a subject set, for too short a time-period, for treatment differences to be seen. These reasons are certainly valid, and should be considered in future studies. However, reliable differences in treatments' outcomes were obtained in another study that we recently conducted, on hydration treatments (Verdolini et al, 1992), using a similar subject set size (N = 6), and providing each of two treatments for an even shorter time-period (one week). What is the basis for the discrepancy in the sensitivity to treatment differences, across the studies? We consider that there are two primary reasons. First, in the earlier study, subjects' severity distributions across treatment were controlled by the use of a within-subjects design. In contrast, in the present study, which used a between-subjects design, there were large differences in initial severity distributions across the experimental groups. This factor restricted statistical sensitivity. Second, in the earlier study, most of the subjects were actual clinical patients, who participated without pay. In the present study, subjects were recruited from outside environments for pay. Thus, there may have been motivational differences that affected the consistency of the results across subjects, leading to a further compromise of statistical power in the present study. (See Nisbett and Wilson, 1977, for a discussion of extrinsic and intrinsic motivators and their effect on attitudes and behavior.) The point is that future studies should ensure similar initial severity distributions across experimental groups, and should utilize clinically motivated subjects.

Returning to the findings for compliance, a secondary, post-hoc question regarded what factors might affect compliance, as estimated by subjects' reports. The only factors that appeared even weakly relevant were: (a) subjects' perceptions about how interesting therapy was, and (b) subjects' perceptions about their technical ability to actually produce the type of voicing that was trained. The trend for this second factor is consistent with the suggestion that "self-efficacy" may affect compliance with health-care regimens in general (see for example, Poll & De Nour, 1980). However, the relations of both factors to compliance scores were relatively small ones, that were not confirmed statistically in the present study. Future studies should improve on statistical sensitivity by using questionnaires about therapy that are less susceptible to the production of skewed data.

A further finding of interest is that although compliance scores predicted improvements in phonatory effort and in the sound of the voice, these scores did not strongly predict improvements in laryngeal status. Inherent in this observation, voice and laryngeal status did not necessarily covary. Examination of individual data confirms this conclusion. In some cases, an improved laryngeal status did not correspond to an improvement in voice measures. In other cases, the converse was true. The point is that laryngeal status did not entirely govern voice status (see also Verdolini-Marston et al., in press, for a similar finding).

In summary, the important finding that emerged from this preliminary study was that compliance factors are important in predicting the likelihood of a benefit from voice therapy for treatment of nodules. This finding should be considered in future efficacy studies, and perhaps in clinical practice as well. Future studies should also obviously include a larger subject set, perhaps recruited from a clinical population, and with similar initial severity distributions across experimental groups. Finally, therapy should be administered over a longer time-period, to allow for a better generalization to the typical clinical situation.

Acknowledgements

This work was partially supported by University of Iowa Junior Faculty Seed Grant G98 to authors KVM and MKB, and by Grant No. P60 DC 00976 from the National Institute on Deafness and Other Communication Disorders. The authors acknowledge Janina Casper for her valuable comments at various points before and during the experiment. This acknowledgement in no way implies her support of the study. The authors also acknowledge Sharon Lindsey, Kay Klein, Mark Peters, and Robin Michel for technical support, and Kevin Spratt for statistical consultations.

References

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

Casper, J.K., Colton, R.H., Woo, P., & Brewer, D. (1989, 1990). Investigation of selected voice therapy techniques. An amalgamation of two papers presented at the 18th and 19th Symposia: Care of the Professional Voice, Philadelphia, PA.

Colton, R.H., & Brown, W.S. (1972). Some relationships between vocal effort and intra-oral air pressure. Paper presented at the 84th meeting of the Acoustical Society of America, Miami, FL, November, 1972.

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

Cooper, M. (1973). Modern techniques of vocal rehabilitation. Springfield, IL: Charles C. Thomas.

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

Jiang, J.J., & Titze, I.R. (1992). Measurement of vocal fold intraglottal pressure and impact stress (1992). National Center for Voice and Speech Status and Progress Report, 3, 21-38.

Kotby, M.N., El-Sady, S.R., Basiouny, S.E., Abou-Rass, Y.A., & Hegazi, M.A. (1991). Efficacy of the Accent Method of Voice Therapy. Journal of Voice, 5, 316-320.

Lancer, M., Syder, D., Jones, A.S., & Le Boutillier, A. (1988). The outcome of different management patterns for vocal cord nodules. The Journal of Laryngology and Otology, 102, 423-427.

Lessac, A. (1967). The use and training of the human voice: A practical approach to speech and voice dynamics. New York: Drama Book Publishers.

Murry, T., & Woodson, G.E. (1992). A comparison of three methods for the management of vocal fold nodules. Journal of Voice, 6, 271-276.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Nissen, M.J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. Cognitive Psychology, 19, 1-32.

Peterson, K.L., Verdolini-Marston, K., Barkmeier, J., & Hoffman, H. (1993). Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. Paper submitted to Annals of Otology, Rhinology, and Otolaryngology.

Poll, I.B., De Nour, A.K. (1980). Locus of control and adjustment to chronic haemodialysis. Psychological Medicine, 10, 153-157.

Raphael, B.N., & Scherer, R.C. (1987). Voice modifications of stage actors: Acoustic analyses. Journal of Voice, 1, 83-87.

Schmidt, R.A., (1988). Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics Publishers, Inc.

Schneider, W., & Fisk, A.D. (1982). Concurrent automatic and controlled visual search: Can processing occur without resource cost? Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 261-278.

Titze, I.R., & Sundberg, J. (1992). Vocal intensity in speakers and singers. The Journal of the Acoustical Society of America, 92, 2936-2946.

Verdolini-Marston, K., & Balota, D.A. (submitted). The role of elaborative and perceptual integrative processes in perceptual-motor performance. Paper submitted to Journal of Experimental Psychology: Learning, Memory, and Cognition.

Verdolini-Marston, K., Burke, M.K., Lessac, A., Glaze, L., & Caldwell, E. (1992, June). Efficacy of resonance and confidential voice therapy for treatment of laryngeal nodules and polyps. Paper presented at the Twenty-first Annual Symposium: Care of the Professional Voice, Philadelphia, PA.

Verdolini-Marston, K., Sandage, M., & Titze, I.R. (in press). Effect of hydration treatments on laryngeal nodules and polyps and related voice measures. Journal of Voice.

Wright, H.N., & Colton, R.H. (1972). Some parameters of autophonic level. Paper presented at the American Speech and Hearing Association Convention, November, 1972.

Appendix A

Questionnaire to therapy subjects, regarding their perceptions of therapy

1) How much do you think that your recent voice therapy <u>improved</u> your voice, overall?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

2) How much did you <u>like</u> the therapy approach that was used?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

3) To what extent did therapy make your voice <u>easier</u>?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

4) To what extent did therapy make your voice <u>clearer</u>?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

5) To what extent do you <u>accept</u> the type of voice use that was trained in therapy?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

6) To what extent do you feel you can <u>be yourself</u>, using your voice the way you learned to use it in voice therapy?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

7) How much <u>fun</u> was voice therapy for you?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

8) How interesting was therapy for you?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

9) To what extent are you technically able to use the technique trained in therapy, in "real life" situations?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

10) To what extent are you willing to use the technique trained in therapy in "real life" situations?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

11) To what extent do you think you will use the technique trained in therapy in "real life" situations in the future?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

12) To what degree would you recommend this type of therapy to a friend who is having voice problems?

1	2	3	4	5
Not at all	Very little	Somewhat	Quite a bit	A lot

NCVS Status and Progress Report - 4
June 1993, 229-235

Sulcus Vocalis in Laryngeal Cancer: A Histopathologic Study

Meijin Nakayama, M.D.
Charles N. Ford, M.D.
James H. Brandenburg, M.D.
Diane M. Bless, Ph.D.
Division of Otolaryngology-Head and Neck Surgery, Department of Surgery, The University of Wisconsin-Madison

Abstract

The incidental finding of sulcus vocalis in surgical specimens of patients with laryngeal cancer prompted this review. Sulcus deformities were histologically identified in 28 of 58 (48%) whole-mount coronal serial-sectioned laryngeal specimens procured from laryngeal cancer patients. The lesions were analyzed, described, and graded. A control group of 20 larynges, obtained from autopsies of patients without known laryngeal pathology, were similarly processed and whole-mount histologic sections studied. Four of these specimens (20%) also demonstrated sulcus deformities. In the control group, the shape and location of the sulci were similar, but the lesions were smaller than in the cancer group. The sulcus lesions revealed chronic inflammation of the subepithelial tissues with vascular ingrowth and fibrosis of the superficial lamina propria (Reinke's space); in the cancer group the sulcus was usually on the opposite vocal fold, where irritation from the tumor might be anticipated. Although the etiology of the sulci remains controversial, these findings suggest that irritation and inflammation might play a role in the pathogenesis of sulcus vocalis.

Introduction

Sulcus vocalis or vocal fold sulcus is a fine longitudinal furrow along the medial edge of the vocal fold. The sulcus might extend over the entire length of the fold or might be limited to a segment of variable length and depth, ranging from a faint line to a deep partition of the fold[1]. Typical clinical findings of sulcus have been described as medial furrows on the membranous vocal fold edges, bowed vocal folds, glottic incompetence during phonation, and excessive ventricular fold adduction[2]. At times these factors result in mild or moderate dysphonia, which might require surgical correction.

The etiology of sulcus is controversial; hypotheses suggest that it is either a congenital malformation or an acquired disorder. The congenital theory is supported by the observation that some sulcus patients have been dysphonic for as long as they could remember and lesions have been identified in some children. Bouchayer et al.[3] support a congenital theory because of the frequent association of sulcus vocalis with epidermoid cysts and mucosal bridges. They hypothesize that it is due to faulty development of the fourth and sixth branchial arches. The acquired theory has been supported by the higher incidence of sulcus in the geriatric population and by its frequent concurrence with laryngeal inflammatory disease such as acute or chronic laryngitis, vocal fold polyps, etc[4,5]. There have been reports of a few cases of sulcus vocalis associated with laryngeal cancer, but the histopathologies were not addressed in detail[2,4]. Because of the relatively small number of the sulcus patients and the difficulties of conducting thorough histopathologic investigations, morphological evidence for these hypotheses of pathogenesis are limited.

Since 1969, the senior author (JHB) has conducted research in which whole-mount laryngectomy specimens have been serially sectioned in the coronal plane; to date, two studies regarding to the features of cancer invasion have been completed and reported[6,7]. During histologic examination of the cancer specimens, attention was drawn to the uninvolved vocal fold because of the high incidence of incidental mucosal deformities. By carefully examining these deformities in 58 laryngeal cancer specimens, and by similarly processing and analyzing 20 control laryngeal specimens, we evaluated the morphological characteristics of sulcus deformities.

Materials and Methods

Between 1969 and 1992, we processed approximately 150 laryngectomy specimens as whole-organ serial sections according to the techniques described by Tucker[8]. Larynges were first decalcified by using 1.5% nitric acid in 10% formalin solution. This solution was daily changed until the calcium precipitation test was negative; generally this process took about 4 weeks. After decalcification, specimens underwent dehydration, infiltration, and, finally, embedding in paraffin. Each specimen was then serially sectioned in the coronal plane from the posterior edge of the thyroid lamina to the thyroid notch. Histologic sections were prepared by serially cutting 8-micron sections after advancing the block 50 times at 20-micron intervals. One laryngeal specimen, therefore, consisted of 25 to 35 ordered sections, the gap between any one section and its neighboring section was approximately 1 mm. After we excluded the partial laryngectomy cases, the cases with no uninvolved vocal folds, and the cases with incomplete medical information, 58 of the 150 cases were available for this study. There were 49 males and nine females, aged 38 to 90 years (mean=59 years) (Photo 6; see center-bound plate). Of the 58 cases, 24 had bilateral and 34 had unilateral uninvolved folds, consequently, 82 folds were available for the analysis.

To validate and reinforce the associated findings in the cancer specimens, 20 control laryngeal specimens (40 vocal folds) were obtained from consecutive autopsies and processed in an identical fashion; there were 14 males and six females, aged 24 to 81 years (mean=53 years) (Photo 6). No patients had had known laryngeal pathology or terminal long-term tracheal intubation.

For histologic evaluation of the mucosal deformities, we defined "sulcus" as an isolated invagination, seen within the squamous epithelium of the true fold, that extends deeper than the full thickness of the squamous epithelium. By limiting the definition to irregularities in the squamous epithelium, we eliminated mucosal deformities frequently seen at the upper and lower transitional zone of the squamous-respiratory epithelium. Considering only invaginations deeper than the full thickness of the epithelium prevented preparation-artifact epithelial irregularities from being misidentified as sulcus[4].

To evaluate the morphological features of sulcus, we classified the lesions into three groups based on the depth and shape of the invagination (Photo 7): (I) Superficial type: Sulcus confined to the superficial layer of lamina propria (Reinke's space) (Photo 8); (II-a) Deep type: Sulcus penetrates the superficial layer and approximates the intermediate and deep layers (vocal ligament) (Photo 9); (II-b) Pouch type: Deep type in which type sulcus exhibits a pouch-shaped configuration (Photo 10).

We evaluated the subepithelial tissues adjacent to the sulcus with particular attention to the extent of fibrous and vascular proliferation. In the superficial layer of normal lamina propria, the fibrous tissue has a loose and homogenized structure, and vascular ingrowth is either very scarce or limited to a few fine capillaries[5]. Based on these descriptions, we defined "increased" fibrosis as irregularly thick or dense fibrous proliferations of fibrous tissue in the superficial lamina propria (Plate 11); "increased" vascular ingrowth was defined as the appearance of multiple enlarged capillaries microscopically identified at low power (Photo 12). Chronic inflammatory changes associated with the sulcus lesions were considered secondary to mechanical irritation when observed adjacent to supraglottic cancers or in the vocal fold opposite the cancer (Photo 13). The location of the sulcus upon the vocal fold was also evaluated. The length of the sulcus was estimated on the basis of the number of serial sections showing sulcus and the calculated distance between each section (1 mm).

Results

Differences in vocal fold morphology were noted between cancer and control specimens. Of the 58 cancer specimens, 28 (48%) had sulcus deformities, whereas only four of the 20 control specimens (20%) showed sulci (Chi square, P=0.027) (Table 1). Bilateral sulci were found in seven of the 28 sulcus-positive cancer cases and one of the four control cases; unilateral sulci were discovered in 21 cancer cases and in three controls. Our results showed that, among the 82 uninvolved folds from the cancer specimens, 35 (43%) manifested sulcus deformities, whereas only five of the 40 folds (13%) from the control specimens revealed sulci.

Among the 35 sulci from cancer specimens, 15 (43%) were classified as superficial, 14 (40%) as deep, and six (17%) were pouch type. All five sulci (100%) from control specimens were classified as superficial type (Table 2) (Photo 14).

Examination of the adjacent subepithelial structure (Reinke's space) revealed that among the 35 sulcus folds from cancer specimens, 22 (63%) showed increased vascular ingrowth and 31 (89%) had increased fibrosis, while of those 47 folds without sulcus, 20 (43%) revealed increased vascular ingrowth and 28 (59%) showed increased fibrosis (Chi

Table 1.
Sulcus Deformities in Cancer and Control Specimens

		Sulcus Deformities			
		(+)		(−)	
	Cases				
CANCER	58	28	(48%)	30	(52%)
CONTROL	20	4	(20%)	16	(80%)
				P=0.027*	

Table 2.
Morphological Classification of Sulcus Deformities

		Sulcus Type		
		I	II-a	II-b
	Sulcus Folds			
CANCER	35	15 (43%)	14 (40%)	6 (17%)
CONTROL	5	5 (100%)	0 (0%)	0 (0%)

square, P=0.004). A more obvious correlation could be seen among the 40 control folds. Of the five sulcus folds, four (80%) had increased vascular ingrowth and all five (100%) showed increased fibrosis, whereas of those 35 folds without sulcus, only three (9%) revealed increased vascular ingrowth (Chi square, P=0.0008) and 20 (57%) showed increased fibrosis (Table 3).

Evaluation of the sulcus deformities for possible cancer-irritation causation revealed that among the 42 *irritation-positive* cases from the cancer group, 24 (57%) showed sulcus deformities, while of those 16 negative cases, only four (25%) showed sulcus lesions (Chi square, P=0.007) (Table 4).

The most common location for sulcus vocalis was the middle one-third of the membranous vocal fold. Among the 35 sulci from cancer specimens, 25 (71%) were located in this portion. The estimated length of sulcus deformities in the cancer specimens ranged from 3 to 13 mm (mean=7 mm), while the sulcus from control specimens ranged from 4 to 6 mm (mean= 5 mm). The mean length of the sulcus did not vary with the type.

Discussion

Although the first complete description of "sulcus vocalis" was made in 1901[9], the clinical literature concerning sulcus remains limited[10]. Histopathologic study of sulcus vocalis is uncommon because of the difficulties of obtaining pathologic specimens. Sulcus is a benign condition, so larynges are never removed for sulcus vocalis. Furthermore, excisions are always conservative because there is already decreased tissue mass in sulcus

Table 3.
Subepithelial Structure and Sulcus Deformities

	Sulcus	Vocal Folds	Subepithelial Structure: Reinke's space — Increased Vascularity (+)	(–)	Increased Fibrosis (+)	(–)
CANCER	(+)	35	22 (63%)	13 (37%)	31 (89%)	4 (11%)
	(–)	47	20 (43%)	27 (57%)	28 (59%)	19 (41%)
				P=0.069		P=0.004*
CONTROL	(+)	5	4 (80%)	1 (20%)	5 (100%)	0 (0%)
	(–)	35	3 (9%)	32 (91%)	20 (57%)	15 (43%)
				P=0.0008*		P=0.064

Table 4.
Cancer Irritation and Sulcus Deformities

	Irritation	Cases	Sulcus Deformities (+)		(–)	
CANCER	(+)	42	24	(57%)	18	(43%)
	(–)	16	4	(25%)	12	(75%)
CONTROL		20	4	(20%)	16	(80%)
					P=0.007*	

vocalis[5] and it is important to conserve normal adjacent structures to achieve optimal voice results. The whole-mount coronal sectioning of larynges, therefore, is an ideal way of studying sulcus; it facilitates the in situ observation of the sulcus and enables the investigation of surrounding submucosal tissues. In previous studies, Ishii et al.[11] identified five sulcus cases in 200 autopsy larynges and Shin[4] also found five sulcus cases in 1200 autopsy larynges. He processed them into whole-mount sections together with one cancerous larynx. The number of specimens available for this type of study, however, are limited. Incidental findings of numerous sulcus deformities in our laryngeal cancer specimens accordingly prompted us to carry out this study.

By our classification, 57% of the sulci from the cancer group were either deep or pouched type (II-a, II-b), whereas all sulci from the control group were superficial type (I). Even if we regarded these superficial sulci as the result of processing artifacts and excluded them, we still found 20 deep-type sulci (57%) among the cancer group. This observation suggests that the sulcus deformities seen in the cancer specimens were closely related to the pathologic condition typically described as sulcus vocalis.

One of the advantages of conducting a whole-mount section study is the ability to investigate the submucosal structures in sulcus. Such investigation can promote a better understanding of the pathogenesis of the sulcus. In our study, we assessed factors related to the inflammatory tissue reaction. Two factors vascular and fibrous proliferation were found exclusively related to the reactions adjacent to the sulcus. These two reactions are commonly seen in the chronic inflammation. In the cancer group, increased fibrous tissue was the dominant finding, while in the control group, increased capillary growth was the primary finding. Our findings were similar to those in existing histopathologic studies, which claimed that extensive fibrosis and edema are typical of the tissue surrounding sulci[4,11-13]. Chronic inflammation of the submucosal tissues was usually found on the opposite vocal fold, where irritation from the cancer mass might be anticipated. In our cancer specimens, irritation-positive cases appeared to closely relate to the sulcus. In 28 sulcus-positive cancer specimens, 16 (57%) revealed possible irritation from the contralateral cancer mass and eight (29%) suggested irritation from an adjacent supraglottic cancer mass.

Conclusion

Careful histologic examination of whole-mount laryngeal specimens has demonstrated a substantial incidence of sulcus vocalis deformities in both cancerous larynges and non-diseased controls. Sulci were more common in the cancer group, suggesting increased susceptibility to the formation of sulcus in this group. The depth of sulcus lesions was greater in the cancer group and the site of lesion was usually opposite or adjacent to the tumor. Just deep to the epithelial invagination, the superficial lamina propria demonstrated increased capillary ingrowth and fibrosis. Chronic inflammation associated with the presence of tumor in adjacent or opposed vocal fold tissues might be a factor in the pathogenesis of sulcus vocalis. These observations suggest that sulcus vocalis is more common than anticipated and, at least in some cases, it appears to be an acquired lesion.

Acknowledgement

This research was supported in part by Grant No. P60 DC00976-03 from the National Institute on Deafness and Other Communication Disorders.

The authors acknowledge, with grateful thanks, Nick Quartuccio, otolaryngology technician, for preparing the whole-organ serial sections, Marta Voytovich, MD, for offering the autopsy larynges, Kennedy Gilchrist, MD, for histopathologic advice, Tom Cook, MS, for statistical assistance, and Celeste Kirk for editorial advice.

Bibliography

1. Arnold G.E.: Dysplastic dysphonia: Minor anomalies of the vocal cords causing persistent hoarseness. LARYNGOSCOPE, 68:142-158, 1958.

2. Itoh T., Kawasaki H., Morikawa I., et al.: Vocal fold furrows: A 10-year review of 240 patients. Auris Nasus Larynx (Tokyo), 10(supplement):S17-S26, 1983.

3. Bouchayer M., Cornut G., Witzig E., et al.: Epidermoid cysts, sulci, and mucosal bridges of the true vocal cord: A report of 157 cases. LARYNGOSCOPE, 95:1087-1094, 1985.

4. Shin Y.: Clinical and pathological investigations of sulcus vocalis. Otologia (Fukuoka), 22:819-835, 1976.

5. Hirano M.: Phonosurgery: Basic and clinical investigations. Otologia (Fukuoka), 21:239-442, 1975.

6. Brandenburg J.H., Condon K.G., Frank T.W.: Coronal sections of larynges from radiation-therapy failures: A clinical-pathologic study. Otolaryngol Head Neck Surg, 95:213-218, 1986.

7. Nakayma M., Brandenburg J.H.: Clinical underestimation of laryngeal cancer: predictive indicators. Arch Otolaryngol Head Neck Surg, (accepted for publication).

8. Tucker G.F.: A histological method for the study of the spread of carcinoma within the larynx. Ann Otol Rhinol Laryngol, 70:910-921, 1961.

9. Salvi G.: De alcune anomalie della laringe umana in individe dilinguenti. Archivo de Psichiatria, 22:369, 1901.

10. Lee S.T.S., Niimi S.: Vocal fold sulcus. J Laryng Otol, 104:876-878, 1990.

11. Ishii H., Baba T., Kawabata I.: Clinical observations on the sulcus vocalis. J Otolaryngol Jpn, 70:911-919, 1967.

12. Loire R., Bouchayer M., Cornut G., et al.: Pathology of benign vocal fold lesions. Ear, Nose and Throat J, 67:357-362, 1988.

13. Colton R.H., Casper J.K., Hirano M.: Understanding voice problems: A physiological perspective for diagnosis and treatment. Williams & Wilkins, Baltimore, pp.59-66, 101-102, 1990.

NCVS Status and Progress Report - 4
June 1993, 237-248

Effects of Voice Disorders on Patient Lifestyle: Preliminary Results

Elaine Smith, M.P.H, Ph.D.
Department of Preventive Medicine, The University of Iowa
Sara Nichols, M.S.
Department of Preventive Medicine, The University of Iowa
Jon Lemke, Ph.D.
Department of Preventive Medicine, The University of Iowa
Katherine Verdolini, Ph.D.
Department of Speech Pathology and Audiology, The University of Iowa
Steven D. Gray, M.D.
Division of Otolaryngology, The University of Utah
Julie Barkmeier, M.S.
Department of Speech Pathology and Audiology, The University of Iowa
Heather Dove, M.S.
Division of Otolaryngology, The University of Utah
Henry Hoffman, M.D.
Department of Otolaryngology, The University of Iowa

Abstract

There is little known about the effects that voice disorders have on disrupting the every day routine and long-term lifestyle of individuals afflicted with these conditions. We have developed a questionnaire designed to elicit information not only about symptoms and clinical diagnoses of voice disorders, but also about the effects these conditions may have on the ability of voice impaired individuals to function at work, with their families, in social interactions, and on their psychological well-being. This paper will describe the preliminary results of these findings.

Introduction

On the basis of research of other chronic diseases, it is well recognized that psychosocial effects are likely to be felt by individuals who recently develop a chronic disease as they attempt to adjust to a changed body image or to a permanent disability (1,2). Studies of patients who survive with cancer or permanent physical impairment indicate that there are major adjustments associated with lifestyle, including changes in income and status associated with loss of work or interaction with one's colleagues. These coworkers serve not only as a reference status group, but also as a group who provide friendship and help with daily coping of life's stresses (3,4). Thus the benefits of this group may be lost when they are needed most. Research suggests that the social and family life may suffer as well, not only due to the loss in income and role identification as parent or provider, but also due to changes in roles that the patient must make in these other areas of his/her life. The effect of the disease subsequently changes the roles and expectations of other family members, leading to further stress with disruption of lifestyle.

Similar studies of the lifestyle and personal problems among patients with voice disorders have not been conducted. To evaluate these effects, we first needed to develop a questionnaire specifically designed for this disorder. Included in the self-administered questionnaire was information about symptoms, clinical diagnosis, and effects the disorder has on the patient's work patterns, social/community activities, family/personal interactions, and psychological well-being.

Materials and Methods

Since September, 1991, patients aged 18 or older have been enrolled from the Voice Clinics in the Department of Otolaryngology, U. Iowa Hospitals and Clinics, College of Medicine, or Department of Otolaryngology - Head and Neck Surgery, College of Medicine, U. of Utah. The patient population includes people who live in the area as well as individuals who are referred from throughout the states of Iowa or Utah to these specialty voice clinics. The various voice disorders were diagnosed in the two voice clinic populations based on laryngeal and speech pathology examinations. The voice disorder groups consist primarily of polyps, nodules, dysphonia, and laryngeal trauma. Diagnoses included nodules, polyps, laryngitis, vocal fold paralysis, spasmodic dysphonia, laryngeal trauma, and a small number of other vocal disorders. Those who were seeking a diagnosis for their voice symptoms were asked to participate by completing a questionnaire regarding the effects their voice symptoms or disorder had on their life. The speech pathologists (J.B., H.D.) reviewed the study with potential participants using an approved Human Subjects Consent form and information summary. Less than 5% refused or were unable to complete the questionnaire. Included in this analysis are the initial 113 participants who completed an elongated questionnaire that was being tested for reliability. Questions about lifestyle effects were tested for reliability by using multiple questions with slightly different wording, then comparing similarity in responses based on the Kappa and McNemar's statistics (S. Nichols, unpubl. data).

The lifestyle questionnaire included information about: 1) functional impact of voice problems in various domains (work, social interactions, personal life, and psychological well-being), 2) current occupation and work-related changes due to vocal symptoms, 3) characteristics and date of initial symptoms, severity of current symptoms, and clinical diagnoses, 4) potential risk factors (eg, alcohol, tobacco, chemical/physical exposures), 5) familial history of voice disorders, and 6) age, gender, occupation, income, education, and race. A medical history review form was devised to document the specific voice disorder, and to examine prior medical conditions or other risks (eg, therapeutic drugs, surgeries) that may be associated with the development of the disorder.

There were five areas of questioning asked about the effect that voice problems had on their work/career (Table 4). About a year after enrollment began, an additional question was asked to determine whether patients (n=30) had to change their job or thought they may need to change it as a result of their voice impairment. Additional questions which may be applicable to social interactions and communication were asked also (eg., difficulties being understood on the phone, conversations with background noise; Table 4). There were nine areas of questioning in this group. Many questions we asked about social interaction problems are applicable to problems that may be encountered with family members/friends as well. We included an additional set of two questions that would apply only to these more personal interaction groups. There were five areas of interest in the psychological category, asking about feelings of self esteem, depression, embarrassment, or self-esteem effects related to work/career.

The age groups in Table 1 were categorized on the basis of different phases in life related to schooling and work/career status. The youngest age (< 21) was still primarily in school and only a few had started working. Most of those working outside of the home were between the ages of 21-65. Those between ages 21-39 were considered to be in the early stages of their career, particularly since many had received advanced degrees. Individuals aged 40-65 were considered to be more established in their careers or job, while those >65 yrs. included primarily retired individuals and housewives. Using the Department of Transportation codes, occupational groups were defined by a three digit code and, because of the small sample size, were divided into five occupational categories: professional/clerical, craftsman/laborer, student, retired, or housewife.

Table 1.
Sociodemographic Characteristics

	%	(N)
Gender		
Female	72.7	(78)
Male	27.3	(30)
Age[1]		
< 21	23.0	(26)
≥ 21-39	37.2	(42)
40-65	22.1	(25)
> 65	17.7	(20)
Education[1]		
≤ 12	38.9	(44)
13-16	43.4	(49)
> 16	17.7	(20)
Income		
< $20,000	33.3	(37)
$20,000 < 40,000	21.4	(24)
≥ $40,000	28.6	(32)
Refused	4.5	(5)
Unknown	12.5	(14)

[1] In years

Because of the small sample size, associations between specific voice disorders and sociodemographic or risk factors would not provide statistically valid assessments, and thus are not described in this study. These will be examined at a later date when additional patients have been enrolled. Thus our results reflect the characteristics of voice disorders as a group by sociodemographic and other risk factors (Tables 1 and 2). Responses to each lifestyle question were categorized in a Likert type scaling; and two groups were devised for the multivariate analysis on responses of Not at all/Very Little versus Moderately/Very Much/Extremely. We used Mantel-Haenszel to examine the bivariate analyses and conditional logistic regression to examine the multivariate relationship between individual questions regarding lifestyle effects and symptoms, severity of symptoms, age, gender, education, income, and potential etiologic risks associated with occupation, alcohol and tobacco. Odds ratios and 95% confidence intervals (CI) were determined from these procedures (5).

Table 2.
Risk Factors

	%	(N)
Occupation		
Prof/Clerical	27.4	(31)
Crafts/Laborer	14.2	(16)
Retired	15.0	(17)
Housewife	14.2	(16)
Student	29.2	(33)
Smoking		
Current	11.6	(13)
Ex	26.8	(30)
Never	61.6	(70)
Alcohol		
Current	63.3	(62)
Ex	10.2	(10)
Never	26.5	(26)

Results

Sociodemographic Characteristics

Table 1 shows the frequency of gender, age, education, and income of the participating patients. Almost three-fourths of the voice patients were female. Most clinic patients were caucasian (95.6%), with the remainder black or hispanic (data not shown). There was an even distribution of age groups seen in the Voice Clinics. Almost all of the retired group were over age 65. Half of the housewives were between 21-65 and the other half were >65. As expected, students were evenly divided in the younger two age groups. Most of the population had at least a high school degree (85.8%, n=97); 61.1% had additional college or graduate education with 64-74% ages 21-65, and 25% ages >65. Income was a mix of low, middle, and higher levels with a portion of the population, usually those less than age 20 unaware of the family income.

Risk Factors

Patients were asked about current occupation and use of tobacco and alcoholic beverages as potential risks of voice problems (Table 2). Professional/clerical and student groups contained the highest representation. Teachers accounted for about 8% (n=9) of those classified in the occupation groups (coded as professional/clerical). No other specific occupational group was identified with a high frequency. Few individuals in the retired group were noted to be younger than the expected retirement age (<65). While the Housewife work status was exclusively female, the Retired group was almost exclusively male.

Ever use of tobacco products other than cigarettes was very low and are not reported here. As indicated in Table 2, only a small proportion of the population was a current cigarette smoker. Ever use of tobacco increased with age (40-65, 64%) dropping to 25% in the oldest age group (p=.02). This contrasts to current use in which those ages 21-39 were heaviest users and few over age 40 currently smoked (p=.01). Men were more likely to be current and ex-smokers than women (p=.03). Likewise, those employed as craftsmen (p=.007) or laborers (p=.02) were more likely to be smokers compared with other work status groups. Use of cigarettes was not statistically associated with severity of any vocal symptoms except for an aching voice (p=.05).

Table 2 shows the frequency of alcoholic use. There were no significant gender-related usage patterns. Use of alcoholic beverages also was evaluated by type: beer, wine, or liquor (data not shown). Beer was more likely to be drunk than other beverages, followed by liquor, then wine. Among beer drinkers, 45.5% were current users, 16.4% former users, and 38.2% Never users. There was an inverse relation between age and ever use of beer, with 85% of the youngest age (<20 yrs) more likely to have ever drunk this form of alcoholic beverage (p=.0001) or to have been current users (p=.001) compared with other age groups. Among users of liquor, 40.0% were current, 19.0% were former, and 41.0% were never users. Again the youngest group was most likely to report ever using this substance (p=.01) and to be currently using liquor (p=.01). For wine drinkers, 34.0% were current drinkers, 23.6% former, and 42.5% never drinkers. There were no age-related wine usage patterns. Among occupational groups, students were most likely to be current drinkers of beer (p=.03) and liquor (p=.08).

Characteristics of Voice Problems

When asked if they thought they currently had a chronic voice problem, 23.4% responded "NO". Nonetheless, 37.6% responded that they had had a chronic voice problem in the past. For most patients, once their voice problem started, it persisted (89.8%). Only about a quarter of the patients had symptoms for less than 12 months; an equal percentage had problems for 5 years or more. About one year after the initial questionnaire began to be administered to patients, we included a question about family history of voice problems. Two patients responded in the affirmative (6.7%).

Vocal Symptoms and Severity

The most commonly reported symptom at the time of diagnosis was hoarseness (73.4%), followed at some distance by a tired, weak, or effortful voice (Table 3). Another voice symptom noted was the inability to sing high, but not low, notes (54.1 vs 22.9% respectively). Yet among those >age 65, the inability to sing low notes was more frequent: 55% compared with 13.0-19.2% among the younger age groups (p=.002). Using logistic regression to control for confounders, effects of age continued to be seen with voice spasms (p=.01) with those >40 yrs old were more likely to state that they had symptoms of vocal spasms. Alcohol use was associated with reporting a tired voice (p=.04). No demographic or other risk factors were significantly associated with reporting voice symptoms.

Table 3.
Voice Symptoms and Severity

	%	(N)
Symptom		
Hoarseness	73.4	(83)
H-Notes Difficult	54.1	(61)
Tired Voice	46.8	(53)
Weak Voice	45.9	(52)
Effortful	44.0	(50)
Breathy	39.4	(45)
L-Notes Difficult	22.9	(26)
Severity[1]		
Effortful	66.9	(75)
Rough	62.1	(69)
Tiring	56.3	(63)
Scratchy	52.8	(57)
Uncomfortable	57.6	(64)
Ache	48.2	(54)

[1] % Based on "moderate-extreme" positive responses

Patients were asked which symptoms interfered more severely with their overall daily life (Table 3). The symptom of effortfulness was most likely to be reported as having severe effects on the patient's daily life (66.9%). The next most commonly cited symptoms were a voice that was rough (62.1%), uncomfortable (57.6%), tired/fatiguing (56.3%), scratchy (52.8%). We then examined whether severity was associated with specific demographic or potential risk factors for voice disorders, using logistic regression. Women were more likely than men to report severity with roughness (p=.05), tiredness (p=.03) and scratchiness (p=.02). Alcohol use was inversely associated with severity of vocal effort (p=.01) and discomfort (p=.01), whereas cigarette use increased severity of scratchiness (p=.08), and achiness (p=.05). There were no differences in responses to severity of vocal effects by age, income, or education.

Since the presence of a symptom at diagnosis may not be associated with feeling that it severely affects daily life, we evaluated this association. Those with a hoarse voice at diagnosis were more likely to report severity associated with symptoms of scratchiness (p=.02), roughness (p=.04), or effortfulness (p=.01); those with voice spasms at diagnosis were more likely to note severity with roughness (p=.005), and effortfulness (p=.002). Those who indicated having a tired voice were more likely to respond with higher severity levels for effortfulness (p=.08), voice discomfort (p=.03), and feeling tired from using their voice (p=.02). A symptom of effortfulness or vocal weakness at diagnosis was associated with feeling a severe effect upon using the voice (p=.01; p=.06). Those who responded that they were unable to sing low notes were more likely to report severity from scratchiness (p=03), and roughness (p=.009).

Lifestyle Effects

The focus of this study was the effect of voice disorders on daily or long-term lifestyle activities 1) at work or in pursuing a career, 2) communicating in social interactions or maintaining community interactions, 3) interacting with family or friends, and 4) altering self-esteem at work or psychological well-being in other parts of their life (Table 4). Tables 5 and 6 show the results of conditional logistic regression based on the same model, separating demographic, chemical, and physical risks (Table 5) from symptom risk factors (Table 6). Since education and income were not significant risks in the bivariate analyses, they were excluded from the logistic regression models. Because of the small sample sizes in the occupational categories, we also were unable to examine them in association with specific voice problems.

Age was significantly associated with most of the lifestyle effects. Smoking was not significant in any of the multivariate models; nor were symptoms of hoarseness, tired voice, and high-notes, and these variables were excluded from Table 6. Communication problems associated with voice disorders in social settings, with family and friends, and with psychological well-being most heavily affected the elderly.

Table 4.
Frequency of Adverse Effects of Voice Disorders on Lifestyle

Effect[1]	%	(N)
Work		
Past job/career	51.5	(51)
Current: limit job decisions	33.3	(37)
Current: limit job performance	45.6	(41)
Current: change job career	31.4	(27)
Future job/career options	74.1	(69)
Social		
Interactions: negative	70.5	(79)
Interact differently	57.3	(63)
Limit participation	60.8	(68)
Avoid social situations	54.5	(60)
Trouble being understood	49.1	(55)
Phone conversation problems	50.9	(57)
Repeat statements	54.9	(61)
Can't express self	50.0	(55)
Conversations: background noise	60.0	(66)
Family/Friends		
Interact differently	50.4	(57)
Family/Friends annoyed	32.4	(36)
Psychological		
Overall	53.6	(60)
Professional/job	55.5	(51)
Depressed	62.5	(70)
Emotional	52.7	(59)
Embarrassed	53.1	(59)

[1] % based on "moderate-extreme" positive responses

Table 5.
Demographic, Physical, and Chemical Risks[1] of Voice on Lifestyle

Effect[2]	Age OR (95% CI)	Sex OR (95% CI)	Alcohol OR (95% CI)
Work			
Past job/career	--	--	--
Current: limit job decisions	--	--	--
Current: limit job performance	--	--	--
Current: change job career	--	--	--
Future job/career options	--	--	--
Social			
Interactions: negative	1.04 (1.01,1.07)	--	--
Interact differently	1.04 (1.01,1.08)	--	0.27 (0.09-0.77)
Limit participation	1.04 (1.01,1.04)	--	--
Avoid social situations	--	--	--
Trouble being understood	1.03 (1.01,1.06)	--	--
Phone conversation problems	1.04 (1.01,1.06)	--	--
Repeat statements	1.03 (1.01,1.05)	--	--
Can't express self	1.04 (1.01,1.07)	--	--
Conversations: background noise	1.03 (1.00,1.06)	--	0.26 (0.09,0.75)
Family/Friends			
Interact differently	1.02 (1.01,1.04)	2.42 (0.99,5.95)	--
Family/Friends annoyed	--	--	--
Psychological			
Overall	1.03 (1.01,1.06)	--	--
Professional/job	--	--	--
Depressed	--	--	0.34 (0.12,0.97)
Emotional	1.05 (1.02,1.08)	--	--
Embarrassed	1.03 (1.01,1.05)	--	--

[1] controlling for age, gender, smoking, alcohol, voice symptoms at diagnosis
[2] % based on "moderate-extreme" positive responses

Work/Future Career. When asked if voice problems had an effect on their past job or career options, 51.5% of the participants concurred and 13.3% had either thought about changing or actually changed their job as a result of severe voice impairment. When asked if their condition currently affected their job performance, the percentage was similar: 45.6%. Almost three-fourths of the participants felt that their condition would limit their future job or career options. Finally, about 56% had a negative self-image in relation to their job or professionally.

In the bivariate analyses, current and future work problems significantly differed by age group. The youngest and oldest age groups were less likely than those between ages 21-65 to state that their voice problems caused job-related problems. Future career decisions were adversely affected as well, but the association increased from the youngest age and dropped off only in the oldest age (p=.001). This same age effect was noted for the two current work-related issues: limiting the way they performed at work (p=.02), and limiting their ability to change their profession/work

Table 6.
Symptom-Related Risks[1] of Voice on Lifestyle Effort

Effect[1]	Weakness OR (95% CI)	Effort OR (95% CI)
Work		
Past job/career	--	--
Current: limit job decisions	2.77 (1.03,7.46)	--
Current: limit job performance	--	--
Current: change job career	--	--
Future job/career options	4.34 (1.34,14.04)	--
Social		
Interactions: negative	--	2.53 (0.95,6.69)
Interact differently	--	--
Limit participation	--	--
Avoid social situations	--	--
Trouble being understood	--	--
Phone conversation problems	--	--
Repeat statements	--	--
Can't express self	--	--
Conversations: background noise	--	--
Family/Friends		
Interact differently	--	--
Family/Friends annoyed	--	--
Psychological		
Overall	2.44 (1.01,5.92)	--
Professional/job	2.54 (0.98,6.60)	--
Depressed	--	--
Emotional	--	--
Embarrassed	--	--

[1] controlling for age, gender, smoking, alcohol, voice symptoms at diagnosis
[2] % based on "moderate-extreme" positive responses

(p=.07). After controlling for the effects of other confounders and risk factors, however, age was no longer a significant risk for work-related problems (Table 5).

Social Interactions. Although this category refers to vocal problems associated with social participation, because of the nature of the questions, this section also includes issues that may be pertinent to the patient's work/career efforts or family interactions. Overall, 70.5% felt that social interactions were negatively affected by their voice problem (Table 4). For example, 54.5% avoided social situations and over 60% otherwise limited these interactions. When they did participate in social events, people with voice impairment interacted differently. Background noise created the most severe problem and they often were asked to repeat statements because they were not clear to others. In phone conversations over half of the patients also found they were difficult to understand. Regardless of the medium used, half the patients felt they could not express themselves.

Most of the social interaction issues were significantly associated with increasing age (Table 5), with those >65 the most severely affected (OR=1.04; CI:1.01, 1.07). The oldest age group most often reported that voice problems caused them to act differently than they would like and lead to limiting their social participation. The elderly also were more likely than other age groups to report: 1) feeling that they could express themselves because of their voice problems, 2) having requests to repeat themselves because their voice was not clear, 3) responding that people have trouble understanding them when they talk, 4) feeling plagued by difficulties in being clearly understood on the telephone, and 5) having people express difficulty understanding them when there was a noisy environment, as expressed by over 70% of the elderly compared with only 23% in the youngest age. Despite these problems, they were no more likely to avoid social situations compared with other age groups. Use of alcohol seemed to lessen these adverse reactions in social interactions, notably for interacting differently.

Interactions with Family/Friends. We asked additional questions that applied only to these more personal interaction groups (Table 4). More than half of the patients interacted with family/friends differently than they wanted to because of their condition and almost a third felt their family/friends were annoyed by their voice problem. Overall however, interactions with family/friends received fewer negative responses than interactions with other groups or in other settings.

The multivariate results suggest that again, the oldest age group was most likely to report the worst effect, feeling a need to interact differently due to their voice problem (OR=1.02; CI:1.01-1.04). Men were somewhat more likely to indicate interacting differently with family/friends because of their voice problems (OR=2.42; CI:0.99-5.95).

Self Esteem. Five areas examined in this category asked about personal feelings of self esteem or those related to work/career (Table 4). Although over half stated that they suffered severe adverse psychological effects (53.6%), this rate belies the effect. A larger proportion stated they had depression, emotional pain, embarrassment, or adverse work-related self-esteem. Increasing age was associated with an overall negative self-esteem, and having adverse emotional feelings about their voice (Table 5). As seen previously for other responses to work-related effects, negative self-esteem on the job increased with age and dropped off in those >age 65. Alcohol seemed to ameliorate feelings of depression.

Symptom-Related Effects on Lifestyle

Voice symptoms were identified that adversely affected the work life of these patients. Table 6 shows that having a weak voice affected the way current career/job decisions were made. Even more detrimental, patients indicated that their voice disorder would limit their future career options because of severe effort required in speaking. Effort required in talking also negatively affected social interactions but no other specific symptom affected this part of their life. No specific symptom was associated with impairing interactions with family or friends. A weak voice was associated with diminishing one's overall psychological well-being and self-esteem at work.

Discussion

This is the first study to assess the effects of voice problems on lifestyle in a large group of patients. Although many health care providers may not consider voice disorders a serious, or life threatening condition, the results of this study indicate that notable adverse effects on lifestyle and quality of life are experienced by this diseased population, as exemplified by the data shown in tables 4-6. We would expect that the working age population would be most affected in their careers or on the job. In fact, 8% of the patient group listed their occupation as teacher at the primary level. Yet in Iowa they comprise only 1% of the total population, suggesting that their work activities place them at high risk of developing voice disorders. It also is evident that various aspects of lifestyle we examined are age-dependent, as evidenced by the fact that the elderly were the most likely age group to report the inability to communicate with family, friends, or others in social interactions, and that this would be viewed as severely compromising their lifestyle. Problems associated with social isolation have been identified as a major issue in studies of the elderly. The inability to speak adequately or to be understood clearly must be frustrating to this age group, and, in addition to their increased likelihood of having other medical problems, may contribute to their social isolation. Since many elderly have limited mobility, the phone is a major source of communication with the outside world, for social and medical needs. Further study is needed to verify these findings and to examine mechanisms for disease prevention or earlier diagnosis, with emphasis on these potential high-risk groups.

It is unclear why the ratio of women to men was 3:1 diagnosed with voice problems in these clinics. It was unrelated to current work status, and thus greater time available to seek medical care, since employment rates were similar. Previous studies have indicated that in certain diagnostic categories, women are more at risk than men (eg., nodules, polyps, Reinke's edema). Since we are continuing to evaluate patients for impaired lifestyle effects, we will reexamine the prevalence at a later date for specific diseases.

Since starting this study, we began addressing three additional issues in the questionnaire. The first issue is associated with work-related effects. Previously, we only asked patients to specify their current job. Since it is also important to determine whether voice disorders may have required currently employed patients to change their occupation, job, or work-related activities, and whether those who defined themselves as Retired or as Housewife may have quit work due to their voice, we have included information about the association between voice problems and changes associated with prior jobs and work activities. Information about changing job status or activities is important in evaluating economic as well as emotional effects that voice disorders may have on the individual. The second issue that we have recently begun addressing is in regards to a potential genetic cause of voice disorders. Patients now are asked to identify first or second degree family members with a voice disorder. The third issue involves inclusion of a comparison, "normative," group to that of the voice disorder population. We also are including a control group to evaluate the prevalence of voice symptoms, frequency, and duration, in comparison to the voice clinic population. This comparison group includes individuals who have not sought a diagnosis for symptoms of a voice disorder and are solicited from the same otolaryngology clinics as the diseased group to control for the effects of referral patterns and other potential biases associated with health seeking behavior. They are companions of patients who are being treated in the Otolaryngology clinics, and are asked similar questions as those assessed in the voice disorder group regarding voice symptoms, and lifestyle effects associated with their "normal" voice. This will establish the prevalence of not only of voice symptoms in a normative population, but also lack of lifestyle disruption among a normative population presumed to be without disease.

Additional research is needed to validate this study based on two specialized voice clinics of primarily Caucasian populations. Further study is needed to evaluate quality of life issues with a larger sample size of the diseased population to determine whether and how diagnosed categories of voice disorders are more likely to be disruptive, and whether certain social, demographic (eg., age), or current occupation influence the adverse side effects associated with voice disorders. The questionnaire could be used as a mechanism to assess treatment outcome, from the patient's viewpoint. In addition, the information gained from this questionnaire will be useful in developing public education and clarifying misinformation that people have regarding the seriousness of specific voice symptoms and their subsequent effects on quality of life. Future studies should evaluate the prevalence of specific voice disorders not only in those who seek, but also in those who do not seek, medical care, due to inadequate education about the association between symptoms and disease, financial limitations, or inadequate medical triage.

Acknowledgement

This research was supported by Grant Nos. K08 DC00036 and P60 00976 from the National Institutes of Health.

References

1. Kaplan RM, Criqui MH. Behavioral epidemiology and disease prevention. New York, Plenum Press. 1983.
2. Holahan CK, Holahan CJ, Belk SS. Adjustment in aging: The roles of life stress, hassles, and self-efficacy. Health Psychol 1984; 3:81-89.
3. Krause N. Social support, stress, and well-being among older adults. J. Gerontol 1986; 41:512-519.
4. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985; 98:310-357.
5. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: Wiley, 1981.

NCVS Status and Progress Report - 4
June 1993, 249-265

Acoustics of the Tenor High Voice

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
Sharyn Mapes, M.M.
Westminster Choir College, Princeton
Brad Story, B.S.
Department of Speech Pathology and Audiology, The University of Iowa

Abstract
The spectra of six tenors were analyzed at high pitches, F_4 to B_4. Because of the wide separation between harmonics, formant frequencies could not be extracted in the traditional way. Rather, an analysis by synthesis technique was used to match the spectra of a model to the measured spectra, using parameter optimization. Results suggest that tenors maintain their first formant frequencies well above the fundamental for all vowels except [u]. The purpose of this seems to be to distribute the acoustic energy between harmonics 2, 3, and 4 rather than to boost the fundamental. Tuning the first formant to the fundamental is a technique used effectively by sopranos but seems to be deliberately avoided by tenors.

Introduction
Tenors often have difficulty producing high pitches with consistent timbre and loudness. When vocal technique is excellent, however, audiences find the tenor voice unusually thrilling. Composers of romantic opera have often assigned heroic roles to the tenor, partially because this voice category offers a rare combination of beauty, excitement, and risk. Excitement stems from the fact that tenor production appears energetic, often produced with high vocal fold tension and high lung pressure. But this leads also to the factor of risk. Tenors, more than any other voice category, are notorious for "cracking" on high notes (changing register involuntarily). Audiences know this and therefore regard tenor production like a high-wire balancing act in the circus, or a triple jump in the skating rink. There is always a finite chance for a major disaster.

The purpose of this paper is to explore some of the acoustic properties of the high tenor voice. In particular, frequency spectra are modeled and interpreted on the basis of source and filter characteristics. Questions of particular interest are: (1) can the traditional linear source-filter theory explain the power spectra observed for tenors? (2) if so, are the source and filter characteristics remarkably different from those found in speech? and (3) do tenors reinforce particular harmonics with vowel modification?

With regard to vowel modification, most classically-trained singers find themselves in the dilemma of having to choose between three conflicting goals: maximizing vocal intensity, maintaining uniform quality (timbre) and maintaining word intelligibility. In the upper part of the fundamental frequency (F_o) range, the harmonics are widely spaced and formants are not uniformally energized for all vowels. Some vocalists claim that they modify the vowels deliberately to get more intensity or to preserve quality. On the other hand, if word intelligibility has to be maintained, some variation in intensity or timbre may be unavoidable over a wide fundamental frequency range.

The choice between uniform intensity and phoneme integrity often depends on the style of the song. For a ballad, a country western song, or an art song, it is generally more important to preserve the verbal message. Hence, vowels and consonants are maintained more speech-like. In an operatic aria, on the other hand, the verbal message is sometimes compromised in lieu of uniform intensity or timbre. The present study is intended to shed some light on the trade-offs made by classically trained tenors.

Recordings

Parts of the study were conducted at Westminster Choir College in Princeton, New Jersey. Six tenors participated as subjects. None were full-time professionals at the time the study was conducted, but all were in training for a career in vocal performance. The singers were asked to perform repeated sequences of connected vowels [a-e-i-o-u] at several pitches, starting with C_4 (261.63 Hz) and ascending the diatonic scale to B_4 (493.88 Hz). The duration of each vowel was approximately 1 second (as metered by a human conductor). Each pitch was prompted on a piano prior to the 5-vowel sequence. The setting was somewhat natural; the singers stood in front of a small audience (about 25 people) in a medium-size classroom. No equipment was attached to the body, and the singers stood 50 cm from a microphone mounted on a stand.

In addition to the high-pitched vowel sequences, where vowel modification was expected to occur naturally as a result of training, the tenors were asked to perform the same vowel sequence on one low pitch, C_3 (130.81 Hz). This was done in four modes to establish some estimate of their speech formant frequencies and possible migrations away from these frequencies. Mode 1 simulated normal speech vowels, mode 2 simulated a lowered larynx posture, mode 3 a deliberate nasalization, and mode 4 the articulatory posture they thought they assumed when they sang high pitches.

Recordings were made by direct A/D conversion (Macintosh II CX with Mac Audios single channel 12 bit audio conversion system and MacSpeech Lab). The sampling frequency was 10 kHz. Prior to each recorded 5-vowel sequence, a practice sequence was performed by the subject to allow for adjustment of maximum signal level without distortion.

Broad-band spectrographic analysis (MacSpeech Lab II) was used to determine the formant frequencies at the low pitch (C_3). For the higher pitches (C_4 to B_4), a narrow-band analysis was used to capture the energy level of each harmonic. The strongest harmonics were typically 40-50 dB above the noise floor. All harmonics of interest (0 to 4000 Hz) were at least 10 dB above the noise floor. Harmonic energy was measured at three points in the vibrato cycle: highest, middle, and lowest frequency. These three values were then averaged to get a representative stable spectrum for each vowel. All total, 120 spectra were analyzed (6 subjects x 5 vowels x 4 pitches).

Spectrographic analysis did not allow for accurate determination of formant frequencies at high pitches because the response of the vocal tract was undersampled by too few harmonics. A modeling and optimization technique (described below) was used instead to estimate the formant frequencies at the high pitches of interest.

Analysis

The classical source-filter theory of vowel production assumes that the frequency spectrum of the mouth output pressure is the product of the glottal (source) spectrum and the vocal tract transfer function. Figure 1(a) shows a model glottal spectrum (described below) whose harmonic amplitudes decrease at a rate of approximately 12 dB/octave. Figure 1(b) shows a vocal tract and radiation transfer function for a neutral vowel (also described below), and Figure 1(c) shows the product of the glottal spectrum and the transfer function; namely the output spectrum. The vocal tract is modeled as an all pole filter, which in this case is truncated to include only 6 poles (formants). However, a correction factor is included to compensate for the effect of missing higher poles (Fant, 1959).

The low fundamental frequency used by the tenors (130.81 Hz) was used for the glottal source simulation in Figure 1. This low F_o produces enough harmonics so that the "shape" of the vocal tract transfer function is reasonably well represented by the output spectrum (i.e. the transfer function is adequately "sampled" by the harmonics). If F_o is much higher, however, the number of harmonics in the glottal spectrum is reduced and consequently the shape of the transfer function is not as apparent in the output spectrum. Figure 2 shows the same spectra for a fundamental frequency of 493.88 Hz (the highest F_o that all the tenors sang). It is clear that the resultant output spectrum does not represent the shape of the transfer function as faithfully as that produced with the lower pitch.

The output spectrum in Figure 2(c) is representative of the spectra obtained experimentally in the present study. Each spectrum con-

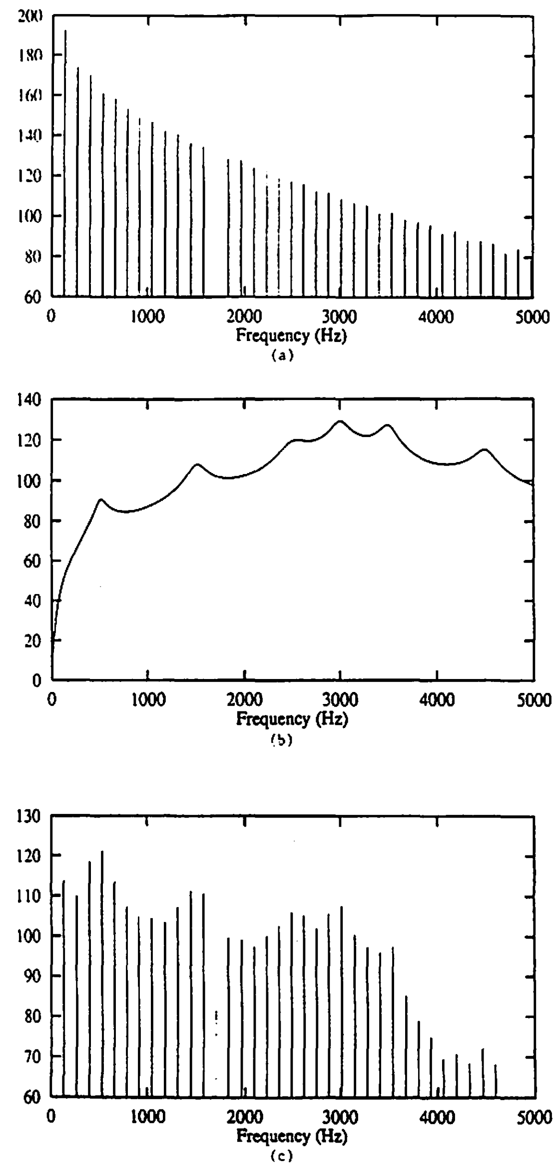


Figure 1. (a) Magnitude spectrum of a glottal flow pulse modeled in the text; F_o = 131 Hz, Q_o = 0.6, Q_t = 3.0, (b) vocal tract transfer (filter) function, including radiation at the lips, for a neutral vowel, (c) the combination (product) of the source and filter spectra.

sisted of 7 to 12 measurable harmonics. An analysis-by-synthesis strategy was employed to interpret the 120 spectra obtained in the experiment. The formant frequencies and glottal parameters of the model were optimized by minimizing the mean square error (Err) between the harmonic amplitudes h_i of the tenors and $\overline{h}_i$ of the model,

$$Err = \sum_{i=1}^{N} (h_i - \overline{h}_i)^2 = minimum \quad . \tag{1}$$

A software package called MINPAC was used for this optimization, which adjusted the values of the formants, an amplitude scale factor, and two glottal source parameters. The criterion for satisfactory optimization was no significant change in the match between the two spectra (visually) after about 50 to 100 iterations.

When the optimization was finished, the final six formant values represented the most likely filter function used by the subject during the vowel production. Optimized values of two glottal parameters (open quotient and skewing quotient) also provided some information about the glottal source. These results were used to answer the three questions set forth in the introduction. The details of the model will now be explained.

Glottal Waveform and Spectrum

The glottal pulse model is illustrated in Figures 3 and 4. Part (a) of each figure shows one cycle of the waveform and part (b) shows the spectral envelope, each for four parameter values. In Figure 3 the flow pulse is parametrized by an open quotient (duty ratio) Q_o that ranges from 0.4 to 1.0. In Figure 4 the flow pulse is parametrized by a skewing quotient Q_s that ranges from 1.0 to 4.0.

The airflow model is based on the premise that glottal flow u(t) is the product of mean air particle velocity v(t) and the minimum cross-sectional area a(t) in the glottis. A relatively simple model of the glottal flow can be constructed by assuming, as Rothenberg (1981) did, that the glottal flow impedance is represented by a nonlinear kinetic resistance R_g and a lumped inertance I of the vocal tract air column.

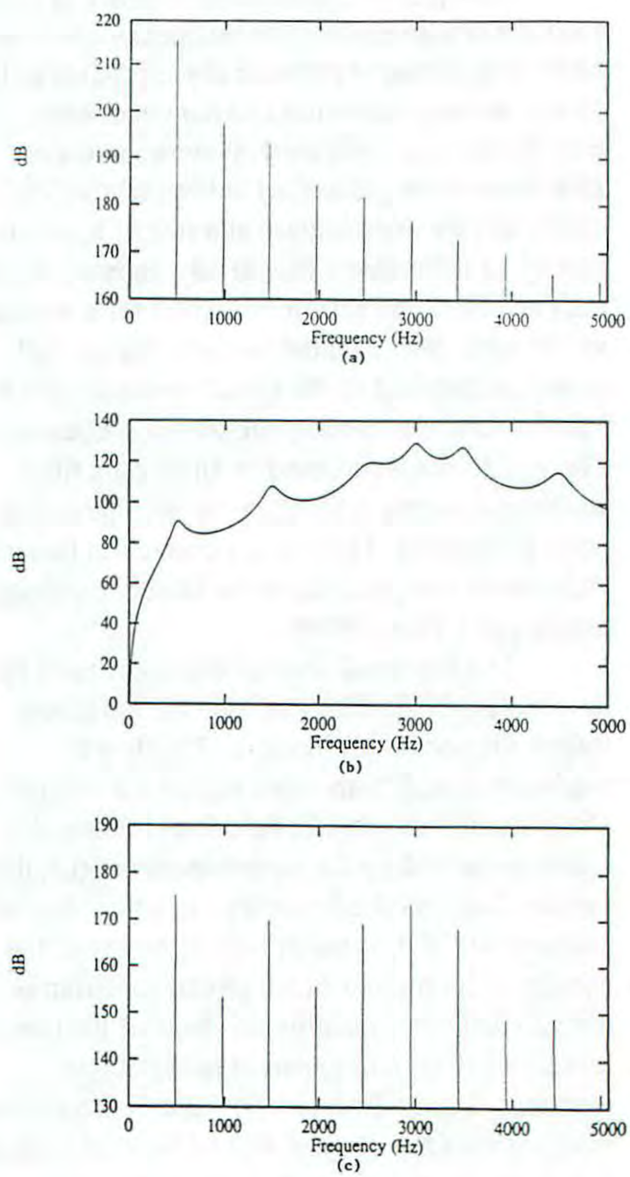
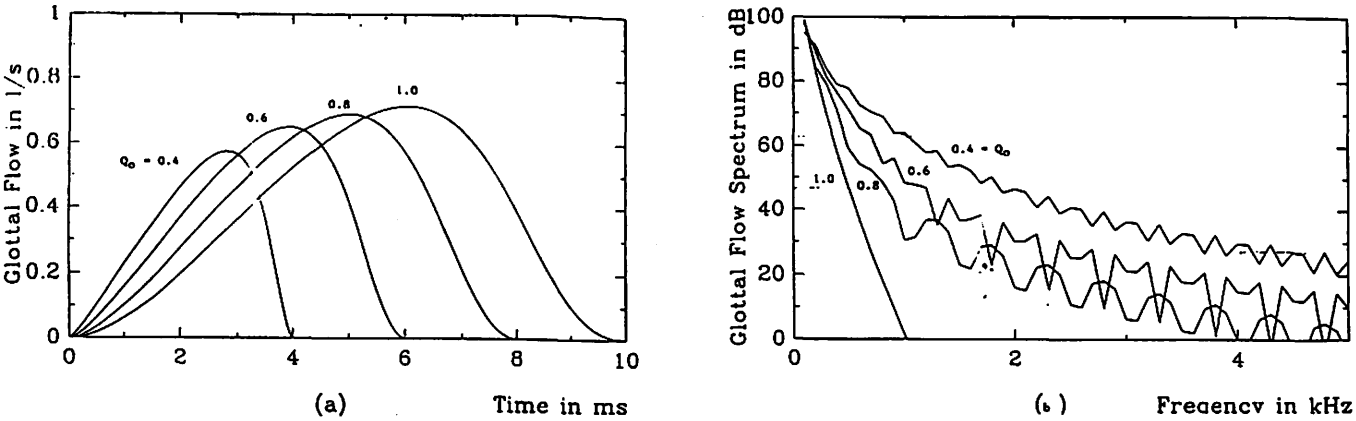


Figure 2. (a) Magnitude spectrum of a glottal flow pulse modeled in the text; F_o = 493.88 Hz, Q_o = 0.6, Q_s = 3.0, (b) vocal tract transfer (filter) function, including radiation at the lips, for a neutral vowel, (c) the combination (product) of the source and filter spectra.



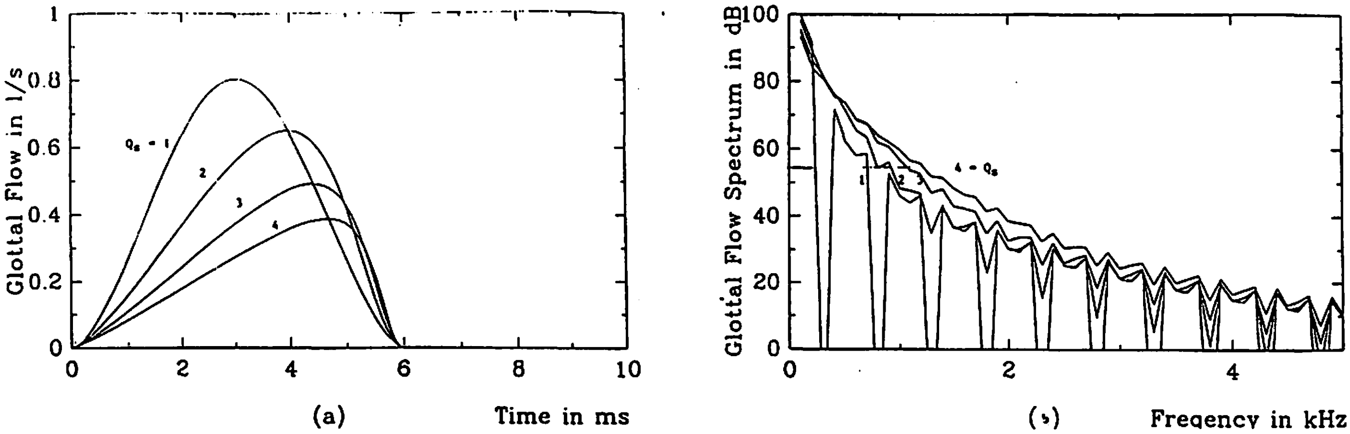


Figure 3. (a - upper left) One cycle of the glottal flow pulse with open quotient Q_o as the parameter, (b - upper right) the corresponding spectral envelopes. Figure 4. (a - lower left) One cycle of the glottal flow pulse with skewing quotient Q_s as the parameter, (b - lower right) the corresponding spectral envelopes.

Then

$$P_L = R_g u + I\dot{u} \tag{2}$$

$$= \frac{1}{2} k_t \rho |u| u / a^2 + I\dot{u} \quad , \tag{3}$$

where P_L is the lung pressure (no subglottal system is included), k_t is a transglottal pressure coefficient (Scherer et al., 1983; Gauffin et al, 1983; Scherer & Guo, 1990), ρ is the density of air, u is the glottal flow, a is the minimum glottal area (in the direction of flow), and $\dot{u}$ is the time-derivative of the flow.

Equation (3), a nonlinear differential equation in u, is the physical basis of the flow pulse. Solutions for u are obtainable numerically. Replacing u(t) by the sampled variable u_n and by $(u_n - u_{n-1})/\Delta t$, an algebraic solution for u_n is

$$u_n = av_o \left[\pm(1 + \delta^2 + \frac{2\delta}{av_o} u_{n-1})^{\frac{1}{2}} - \delta \right] \quad , \tag{4}$$

where

$$v_o = (\frac{2P_L}{k_t\rho})^{\frac{1}{2}} \tag{5}$$

and

$$\delta = \frac{I}{2P_L\Delta t} av_o = \gamma av_o \quad . \tag{6}$$

The quantity v_o is the no-load (I = 0) particle velocity in the glottis, and δ is an inertial load factor that delays the buildup of flow when the glottis opens. For $\delta = 0$, the solution for u is simply av_o, as can be seen from equation (3). The flow is proportional to the glottal area *a* in this no-load case. For $\delta = 0$, typical values of I and P_L can be estimated from the literature, but it is simpler to adjust the quantity $\gamma = I/(2P_L\Delta t)$ for optimizing the waveform skewing. This will be discussed later.

The glottal area is sometimes approximated by a raised, truncated sinusoid (Lin, 1987; Titze, 1988) or by a triangular wave (Rothenberg, 1981; Fant, 1982). Truncation (half-rectification) results from vocal fold collision, and raising of the function above the zero line results from abduction of the vocal processes. Typically, the glottal area varies from a half-rectified sinusoid (or triangular wave) in normal voice to a nearly full sinusoid in breathy voice. This is captured by a simple formula

$$a(\theta) = Max[0, \sin^{\beta}\theta] \quad , \tag{7}$$

where

$$\theta = \frac{\pi t}{Q_o T_o} \qquad 0 \leq \theta \leq \pi \tag{8}$$

Parameters are the period T_o, the open quotient Q_o and an exponent β. All of the parameters can be non-integer numbers. Equation (7) produces a constant (square pulse) for $\beta = 0$ and a sinewave for $\beta = 1$, both of which have a pulse duration given by Q_oT_o. For $\beta = 2$ and $Q_o = 1$, the formula produces the function $\sin^2(\theta t/2) = 1 - \cos\theta$, which is a raised cosine. The exponent β is thus a measure of the "softness" of onset and offset of the pulse. It governs the value of the derivative of $a(\theta)$ at opening and closing.

A *skewing quotient* is defined as

$$Q_s = \frac{T_p}{T_n} \quad , \tag{9}$$

where T_p is the time of increasing flow (positive flow derivative) and T_n is the time of decreasing flow (negative flow derivative). This parameter is used for more direct control of the waveshape. A rule is used to relate this ratio to the parameter γ of the model,

$$\gamma = 80{,}000\,(Q_s - 1.0) \quad s/m^3 \quad . \tag{10}$$

A second rule

$$\beta = Q_o + 1.0 \tag{11}$$

quantifies the degree to which waveform rounding at opening and closing co-varies with the open quotient. These rules are empirical and are based on experience in working with the model.

A value u_o is assigned to the maximum value of the non-interactive flow av_o in equation (4). Physiologically, this is the appropriate amplitude control parameter because it is directly related to lung pressure P_L (through v_o in equation 4). The only disadvantage is that the exact peak-to-peak value of flow is not known *a priori* for $\delta > 0$. This value can, of course, be determined computationally during simulation for later scaling purposes.

With equations (10) and (11), the optimization program can be given control over Q_o and Q_s, which were illustrated parametrically in Figures 3 and 4. Note that Q_o affects primarily the spectral slope, as seen in Figure 3(b). For $Q_o = 0.4$, the spectral slope is on the order of 12 - 15 dB/octave, whereas for $Q_o = 1.0$ the roll-off is 30 - 50 dB/octave, suggesting that the waveshape approximates a sinusoid.

The skewing quotient Q_s has less effect on overall spectral slope, but governs the depth of the spectral valleys, as seen in Figure 4(b). A perfectly symmetric pulse ($Q_s = 1.0$) has various harmonics missing, whereas a highly skewed pulse ($Q_s = 4.0$) has significant energy in all the harmonics. For the case shown ($F_o = 100$ Hz and $Q_o = 0.6$), the third, eighth, thirteenth, eighteenth, twenty-third. . . harmonics are depressed, with the depth of the valleys being a direct function of Q_s. It is also interesting to note that the spacing between the valleys is determined by Q_o. When $Q_o = 0.5$ (not shown), the valleys are closest in spacing (every other harmonic is depressed). Spacing increases on both sides of $Q_o = 0.5$. This is nothing more than the well-known (sinx/x spectral pattern obtained in Fourier analysis of periodic pulses of varying widths.

For tenors producing high notes, loss of a single harmonic can be critical, especially if this harmonic is near a formant. The tendency, therefore, is to maintain pulse skewing under all conditions, as will be shown later.

Vocal Tract Transfer Function

According to Fant (1960) the vocal tract transfer function for non-nasalized vowels can be modeled as

$$\frac{P_r(s)}{U_g(s)} = sRr\left[\left(s + \frac{R_r}{I_r}\right)\prod_{i=1}^{\infty}\left(1 - \frac{s}{s_i}\right)\left(1 - \frac{s}{s_i^*}\right)\right]^{-1} = H(s) \quad , \tag{12}$$

where s is complex frequency and

$$s_i = -\omega_i/2Q_i + j\omega_i \tag{13}$$

is the i-th complex pole (formant) of the vocal tract, with ω_i being the radian formant frequency and Q_i being the resonant quality factor of the formant. This is the classical source-filter transfer function. If the source spectrum $U_g(s)$ is known, then the power spectrum can be calculated.

Higher Pole Correction

Implementation of the source-filter transfer function in an analysis procedure requires that the number of poles be truncated from infinity to some reasonable number. For this study, we have chosen to use six poles. In theory, however, an infinite number of poles do exist and any truncated transfer function must include a correction for the effects of the higher poles. The complete transfer function would assume the form

$$|H(s)| = K_{r6} \prod_{i=1}^{6} H_i(s) \tag{14}$$

where K_{r6} is the higher pole correction factor (HPC).

An equation for the HPC is derived in Fant (1959). For a six pole approximation to the transfer function, it is given as

$$20\log_{10}(K_{r6}) = 0.361(x_1)^2 + 0.00041(x_1)^4 \tag{15}$$

where $x_1 = 4*f*L_e/c$. Frequency is represented by f and L_e is the effective length of the vocal tract, which is the sum of the actual vocal tract length and an end correction $(\Delta L = 0.8\sqrt{A/\pi})$ resulting from the radiation inertance of the tract and the cross section area (A) at the open end.

Table 1.

Vowel	True F_7	Hypothetical F_7
[a]	6100	9100
[o]	5900	6500
[u]	6300	3500
[i]	6100	6500
[e]	6000	6760

However, the use of a truncated transfer function with the HPC has some limitations. According to Fant (1959), one cannot expect the HPC to be valid for formant configurations in which F7, F8, F9,. . . are located at significantly different frequencies than they would be if the tract were a uniform tube with first resonance F1. Table 1 shows approximate first formants (F1) for 5 vowels, the "true" F6, and the hypothetical F6 that would occur if the tract were a uniform tube with a first formant equal to F1. It is clear that the "true" F6 for [u] and [a] is much different than the hypothetical F6. Thus, we cannot expect the analysis to be very successful at matching the model and experiment for all vowels.

Table 2.
Initial Conditions

Vowel	F_1	F_2	F_3	F_4	F_5	F_6
[a]	600	1200	2400	2800	3500	5000
[e]	400	1500	2400	2800	3800	4800
[i]	500	1800	2400	2800	3800	4800
[o]	500	900	2400	3300	3800	4800
[u]	240	800	2200	2600	3500	5000
Pitch	**Frequency (Hz)**					
F_4	349.23					
G_4	392					
A_4	440					
B_4	493.88					
$Q_o = 0.6$						
$Q_s = 3.0$						

Optimization Procedure

The parameters optimized by MINPAC were:

- six formant frequencies (F1 - F6)
- an amplitude scale factor
- the glottal open quotient (Q_o)
- the glottal skewing quotient (Q_s).

Seed values were selected from the experimental data for F_0 and vowel (Table 2). A set of Q_i-values (formant bandwidths) were selected for each formant, and initial conditions for Q_o and Q_s were 0.6 and 3.0 respectively, as shown in Table 2. It is important that the initial formant values are given in an arrangement that is reasonable for the particular vowel under analysis. For example, formant values resembling those of an [u] would not be expected to work for an [a] vowel.

The parameters were constrained to stay within a reasonable range to avoid nonsense solutions, such as negative formant values or glottal parameters that have no physical interpretation. The constraints imposed on each of the parameters were as follows:

$$0.3 < Q_o < 0.8$$
$$1.0 < Q_s < 6.5$$
$$0 < F_1 < 1000 \text{ Hz}$$
$$500 < F_2 < 2200 \text{ Hz}$$
$$1700 < F_3 < 3500 \text{ Hz}$$
$$2300 < F_4 < 4000 \text{ Hz}$$
$$2800 < F_5 < 5000 \text{ Hz}$$
$$3000 < F_6 < 9000 \text{ Hz.}$$

An additional constraint was imposed on the formant values,

$$F_1 < F_2 < F_3 < F_4 < F_5 < F_6 \quad .$$

The amplitude scale factor was not constrained.

The Q_i-values (formant frequencies divided by formant bandwidths) were not optimized by the software optimizer. Rather, they were under user control. Typical values were (5, 10, 10, 10, 20, 20), but if the resulting match between model and experiment was not as good as desired, the user had the choice of altering the formant bandwidths and running the model again. Initially, we thought that we could let the optimizer find the formant bandwidths, but this was not successful. Optimizing 6 formant frequencies, 6 bandwidths, 2 glottal parameters, and a scale factor creates an extremely complicated parameter surface that contains many local minima in which the model can get "stuck". Also, changing bandwidths can be a way of quickly changing formant amplitudes (a large amount of gain for a small change in parameter value), and the optimizer would tend to use formant bandwidths to adjust spectral slope.

Testing the Model

Before the 120 tenor spectra were analyzed, the model was tested by using an articulatory synthesizer (Liljencrants, 1985) as the ideal "subject". The source in this synthesizer was identical to the one described herein, but the vocal tract was simulated by wave reflection techniques. Mouth output spectra were synthesized for the same vowels and pitches used in the experimental study. The magnitude of the vocal tract transfer function was then computed. Figure 5(a) shows an example of a mouth spectrum in which the fundamental frequency was 440 Hz and the vowel was [o]. Figure 5(b) shows the corresponding transfer function.

Since the formant frequencies are known from the computer transfer function, the mouth spectrum can be used as an input to the optimizer to compare the extracted values with the known values. Figure 6 shows best and worst results of this optimization. In Figure 6(a), the peaks of the harmonics in Figure 5(a) were redrawn as vertical lines for the vowel [o]. The asterisks representthe model match to the lines and the open circles represent the extracted source spectrum. We can see that the fit between the known and extracted harmonic amplitudes is excellent.Other vowels and pitches also produced good spectral matches and formant values, with the exception of the [u] vowel (Figure 6b). The fit between model and experiment for this case was poor. The difficulty in adequately modeling [u] is based entirely on the limitations of the higher pole correction to which Fant (1959) alluded and which was described earlier in this paper. To maintain a complete set of cardinal vowels, the optimized formant values for [u] were included in subsequent data sets, but they must be interpreted with some reservations. In general, the source-filter optimization model appears to provide an adequate model for the high tenor voice.

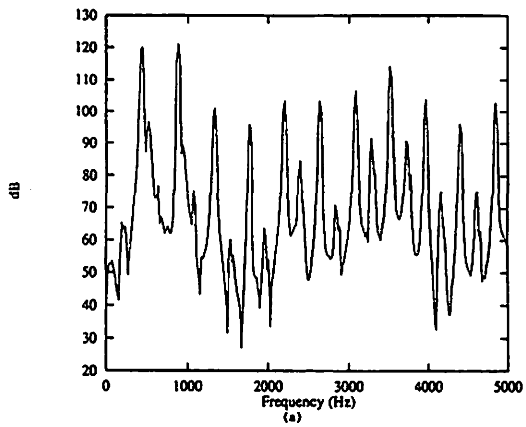


Figure 5. (a - left) Mouth spectrum for the vowel [o] obtained from an articulatory synthesizer for which the source function was known, (b - right) the corresponding vocal tract transfer function.

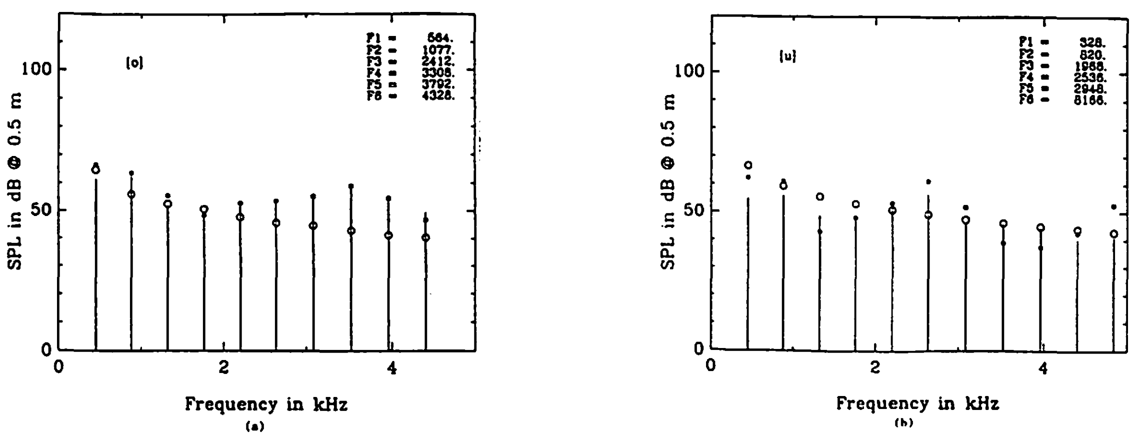


Figure 6. Mouth spectra obtained from an articulatory synthesizer (lines) matched by the optimizer (asterisks) for a pitch A_4 = 440 Hz. The open circles represent the source spectrum (a - left) the vowel [o], which shows the best results, (b - right) the vowel [u], which shows the worst results.

Results and Discussion

The 120 experimental vowel spectra were analyzed with the procedure just described. Figure 7 shows results for one subject (JD) singing the same two vowels [o] and [u] as the articulatory synthesizer in Figure 6. Note that the spectrum is matched extremely well for [o] in Figure 7(a). Extracted formant frequencies are listed in the upper right hand corner, as well as the open quotient value Q_o and the skewing quotient value Q_s. For the vowel [u] in Figure 7(b), the results

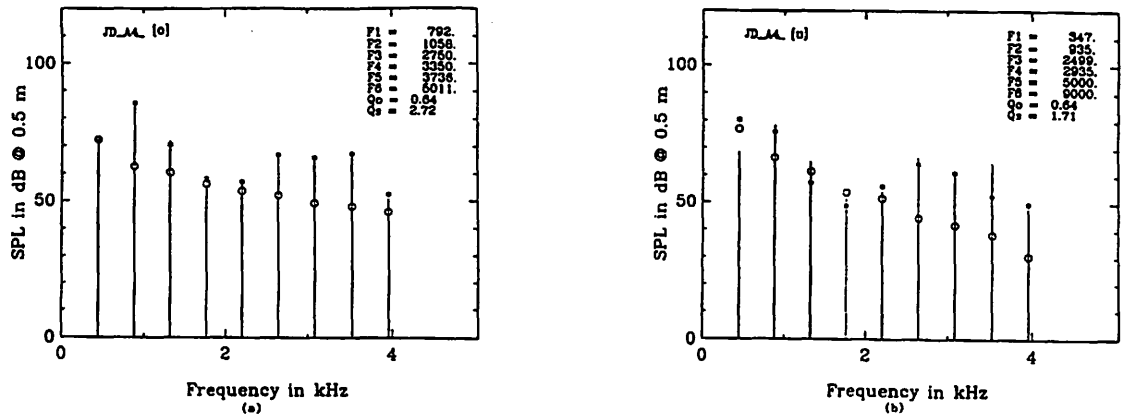


Figure 7. Mouth spectra obtained for subject JD (lines) matched by the optimizer (asterisks) for a pitch A_4 = 440 Hz. The open circles represent the source spectrum (a - left) the vowel [o], which shows the best results, (b - right) the vowel [u], which shows the worst results.

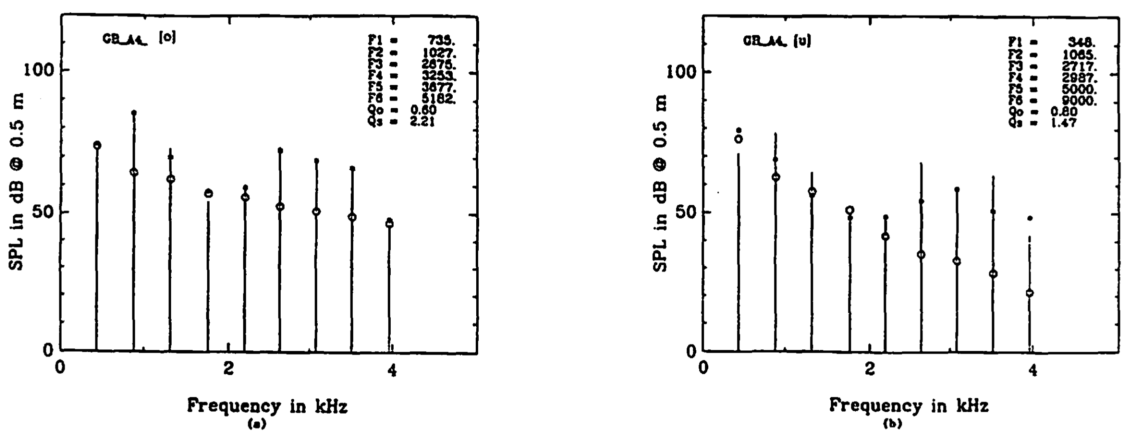


Figure 8. Mouth spectra obtained for subject GB (lines) matched by the optimizer (asterisks) for a pitch A_4 = 440 Hz. The open circles represent the source spectrum (a - left) the vowel [o], which shows the best results, (b - right) the vowel [u], which shows the worst results.

appear more questionable, but perhaps useful. For [i], [e] and [a] vowels, results were generally as good as those for [o].

Figure 8 shows the same picture one more time (A_4 = 440 Hz on the vowels [o] and [u]), but for a different tenor (GB). This subject did not have as much training as subject JD and wasjudged by one of the authors to be struggling with the high tenor technique. The amount of training and general quality of the sound did not seem to influence the optimization procedure.

To demonstrate the full range of vowel spectra obtained for all the tenors, Figure 9 shows the traditional Peterson and Barney (1952) representation of F_2 and F_1. Figure 9(a) shows speech-like vowels (C_3 = 130.81 Hz) for the six tenors and Figure 9(b) shows singing vowels for the high pitches (F_4 to B_4) across all subjects. Note that the vowel spaces are generally enlarged from

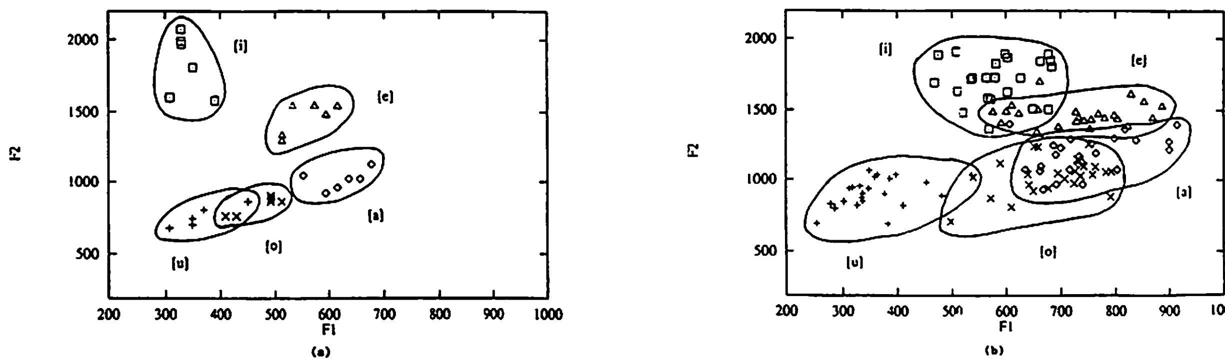


Figure 9. Vowels for six tenors in the traditional F_1-F_2 plane. (a - left) speech-like vowels at C_3 = 130.81 Hz, (b - right) singing vowels at F_4, G_4, A_4 and B_4 combined into one chart.

speaking to singing, and F_1 is significantly raised for all vowels. The raising of F_1 is probably related to increased mouth opening for singing, which was described by Sundberg (1977) for sopranos. It may also be related to some larynx raising for singing versus speech, although this technique has been more debated for the male high voice.

Figure 10 shows a superposition of speaking and singing vowel spaces in the F_1-F_2 plane. The shifting to the right (solid boundary) in singing with respect to speaking (dashed boundary) is clear evidence of F_1 raising. Also evident is a general broadening of the horse-shoe-like vowel space, suggesting more vowel centralization (shifting toward the neutral vowel [3] for singing.

Figures 11 and 12 show the spectral changes for [o] associated with rising pitch. For each of the three subjects shown in parts (a)-(c), F_0 is the parameter that identifies a curve. The four curves represent the high notes F_4, G_4, A_4 and B_4. It is seen that the tenors raise both F_1 and F_2 fairly consistently with rising F_0 for this vowel, but there is no apparent consistency for the other formants. The attempt for

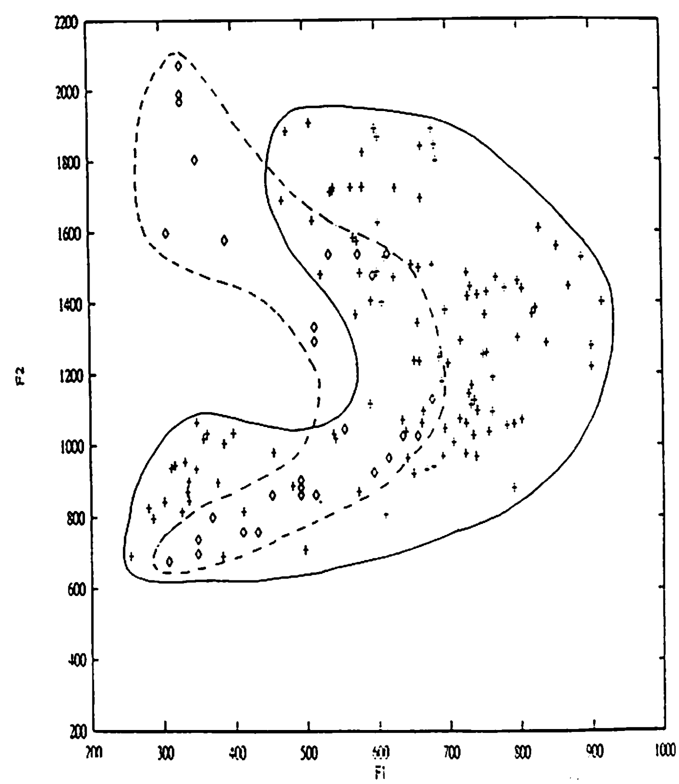


Figure 10. Superposition of speech-like (dashed outline) and singing (solid outline) vowel spaces in the F_1-F_2 plane. Data are for six subjects across five vowels. For the singing vowels, there were four pitches (F_4, G_4, A_4 and B_4)whereas the speech-like vowels there was only one pitch (C_3 = 130.81 Hz).

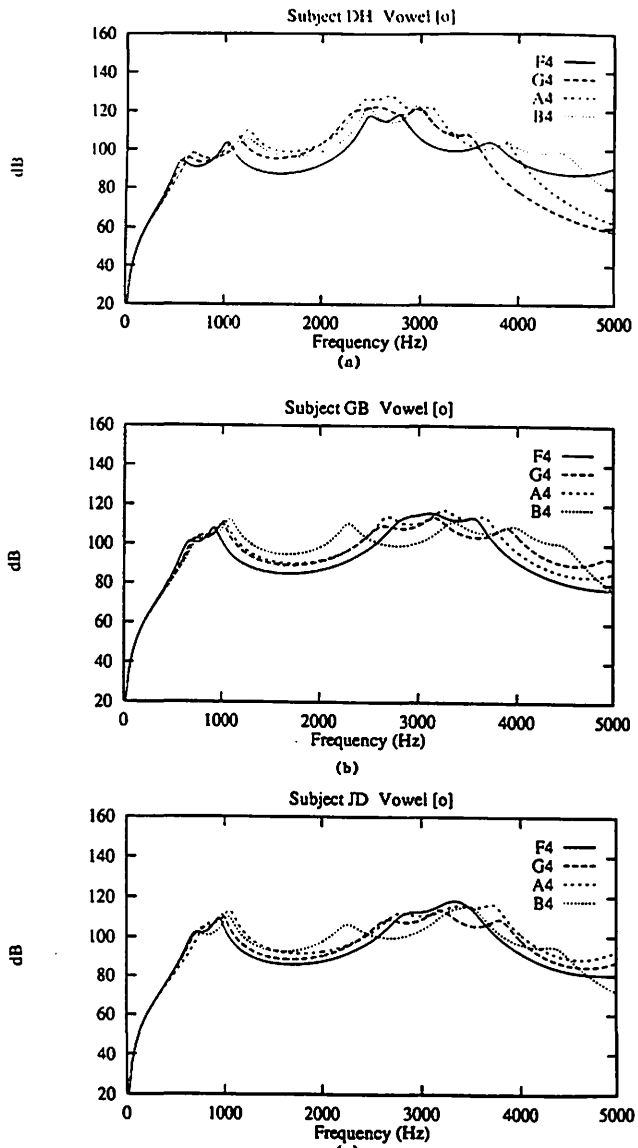


Figure 11. Spectral changes of the vowel [o] for rising F_o in three tenors (DH, GB and JD). The F_o values are F_4 = 349.23 Hz (solid lines), G_4 = 392 Hz (longer dashes), A_4 = 440 Hz (shorter dashes) and B_4 = 493.88 Hz (dots).

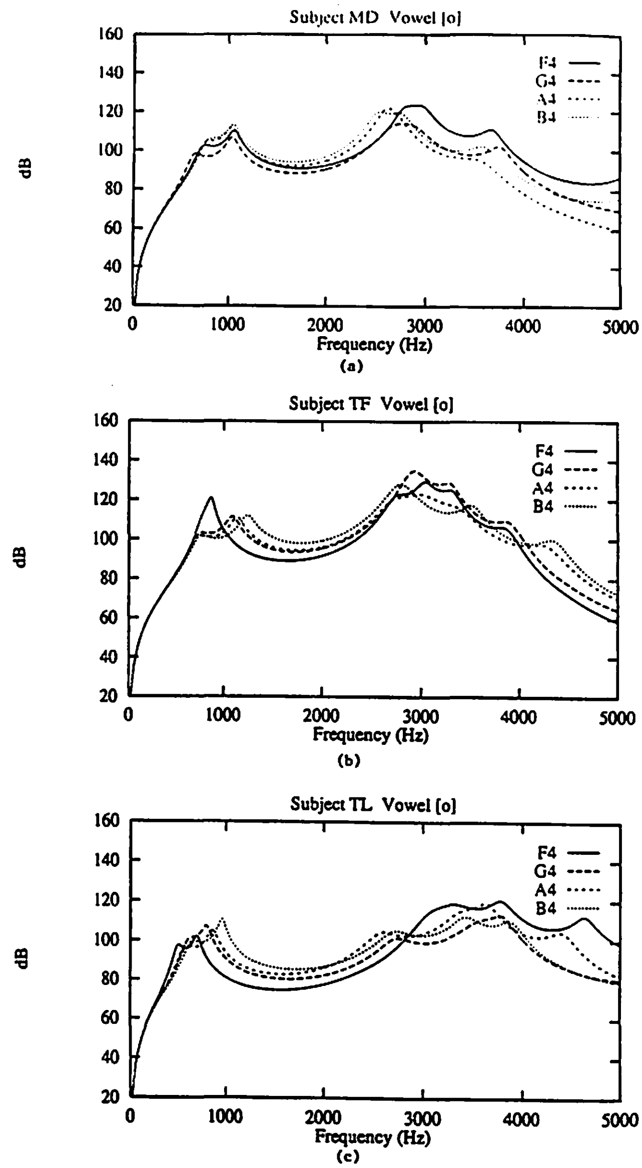
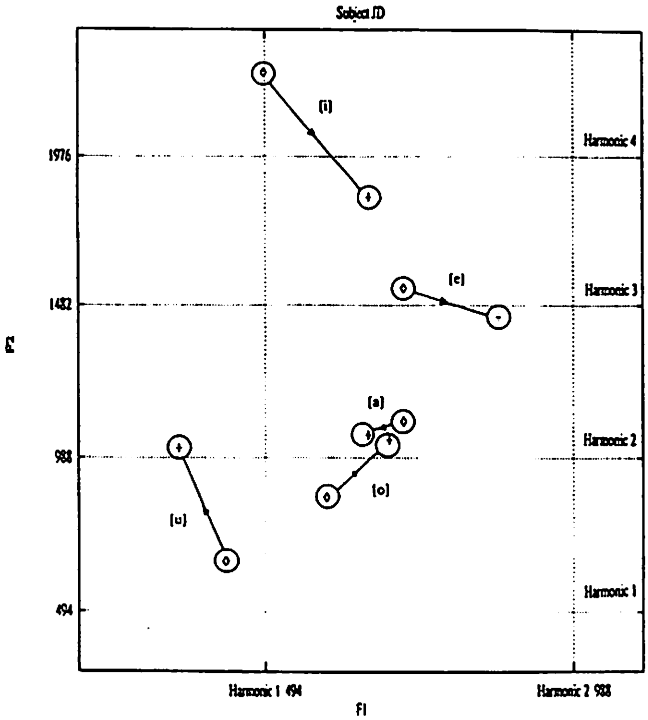


Figure 12. Spectral changes of the vowel [o] for rising F_o in thee tenors (MD, TF and TL). The F_o values are F_4 = 349.23 Hz (solid lines), G_4 = 392 Hz (longer dashes), A_4 = 440 Hz (shorter dashes) and B_4 = 493.88 Hz (dots).

higher formants seems to be to bunch the formants together to maintain a cluster near 3000 Hz, the singer's formant region (Sundberg, 1977).

As a final data set, we address the subject of formant tuning. We return to the F_1-F_2 plane, but this time display only the results of two subjects (the least skilled in Figure 13 and the most skilled in Figure 14). Both cases are for the highest note (B_4 = 493.88 Hz). Harmonic grid lines are



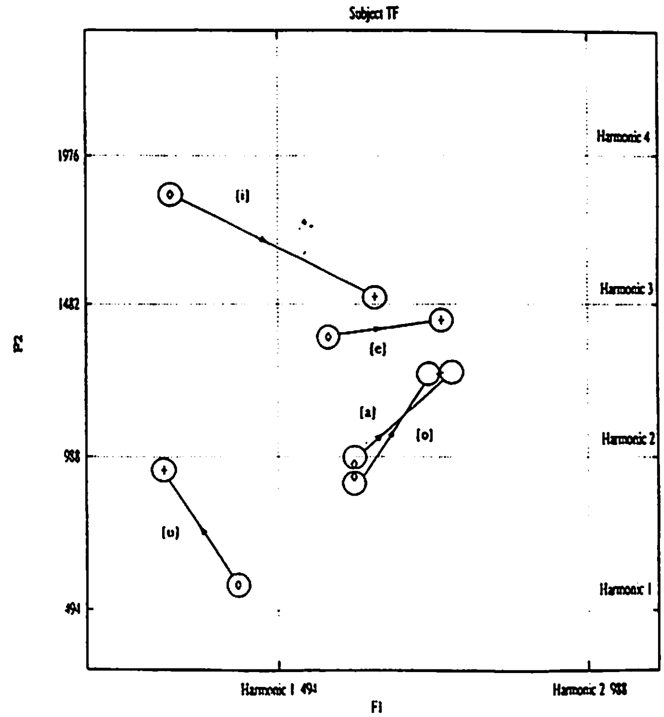


Figure 13. (left). Shifting of formant frequencies F_1 and F_2 from in relation to target source harmonics (grid lines). The shift is from speech-like F_o (C_3 = 130.81 Hz) to singing at the highest F_o (B_4 = 493.88 Hz). The subject is JD.
Figure 14 (right). Shifting of formant frequencies F_1 and F_2 from in relation to target source harmonics (grid lines). The shift is from speech-like F_o (C_3 = 130.81 Hz) to singing at the highest F_o (B_4 = 493.88 Hz). The subject is TF.

shown to determine if the subjects deliberately adjusted F_1 or F_2 to lie on or near one of the harmonics. Arrows point from speaking vowels to singing vowels.

For subject JD, the least skilled of the subjects, Figure 13 shows no obvious attempt to tune F_1 or F_2 to a single harmonic, especially not to the fundamental. If there is a trend at all, it is toward strengthening the second, third and fourth harmonics collectively. Interestingly, this is also the trend for the most skilled tenor, TF, in Figure 14. There is a slightly more noticeable trend toward strengthening the third harmonic in the vowels [a] and [o]. It appears that the intent is to get strong, broad spectral regions rather than strong isolated harmonics.

Conclusions

To return to the original questions, it appears that a traditional source-filter analysis can explain the power spectra of the high tenor voice. The source characteristics are not remarkably different from those found in speech. Open quotients were estimated between 0.3 and 0.8, with the majority of values ranging between 0.4 and 0.6. This is in agreement with direct measurements made on a different set of tenors (Titze & Sundberg, 1992; Sundberg, Titze, and Scherer, 1993). Skewing quotients ranged between 1.0 and 6.0, with the majority of values ranging between 2.0 and 4.0. This is also not remarkably different from speech or from previous measurements on singers.

All of the tenors studied here tended to raise their first formant as F_0 was increased. Only for the vowel [u] was F_0 ever above F_1. None of the tenors reinforced the fundamental with formant tuning. This is particularly remarkable since F_0 and F_1 have a natural range of overlap in the high tenor range (350 - 500 Hz). The objective seems to be to maintain a high concentration of energy in the second, third, and fourth harmonics, rather than the fundamental, to retain the characteristic male sound. Any attempt to boost the fundamental, or any other single harmonic, would make the sound characteristically female.

The high-wire act that tenors seem to perform, then, is not to do anything acoustically different at high pitches, but rather to produce sound in the same consistent way throughout the entire pitch range. What that requires physiologically is beyond the scope of this discussion. Suffice it to say that the voice has a natural tendency to "flip over", to change to falsetto register for many tenors. Reversing this tendency is what pedagogy for the high male voice seems to be all about.

Acknowledgment

This research was supported by Grant No. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders.

References

Fant, G. (1959). Acoustic analysis and synthesis of speech with applications to Swedish. Reprint from Ericsson Technics, 1.

Fant, G. (1960). Acoustic Theory of Speech Production. The Hague: Mouton & Company

Fant, G. (1982). The voice source-acoustic modeling. Quarterly Progress and Status Report, STL-QPSR 4/1982. Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden, pp. 28-48.

Gauffin, J., Binh, N., Ananthapadmanabha, T.V., & Fant, G. (1983). Glottal geometry and volume velocity waveform. In D. Bless & J. Abbs (Eds.), Vocal Fold Physiology, Contemporary Research and Clinical Issues, pp. 194-201. San Diego: College Hill Press.

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

Lin, Q. (1987). Nonlinear interaction on voice production. Quarterly Progress and Status Report,STL-QPSR 1/1987. Speech Transmission Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden, pp. 1-12..

Rothenberg, M. (1981). Some relations between glottal airflow and vocal fold contact area. Proceedings of the Conference on the Assessment of Vocal Pathology, ASHA Reports, Rockville, MD, No. 11., 88-96.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24(2), 175-184.

Scherer, R., & Guo, C. (1990, November). Laryngeal modeling: Translaryngeal pressure for a model with many glottal shapes. Proceedings from the International Conference on Spoken Language Processing, Kobe, Japan.

Scherer, R., Titze, I.R., & Curtis, J.F. (1983). Pressure-flow relationships in two models of the larynx having rectangular glottal shapes. Journal of the Acoustical Society of America, 73(4), 669-676.

Sundberg, J., Titze, I., & Scherer, R. (1993). Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. Journal of Voice, 7(1), 15-29.

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

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

Photo 1. A single frame from a videofluoroscopy film. (Taken from Talker V, Experiment 4.)

Photo 2. A single frame from a videofluoroscopy film of Talker NV (Experiment 4). The exposure for the head is darker than that used when filming Talker V to eliminate information about tongue root movement.

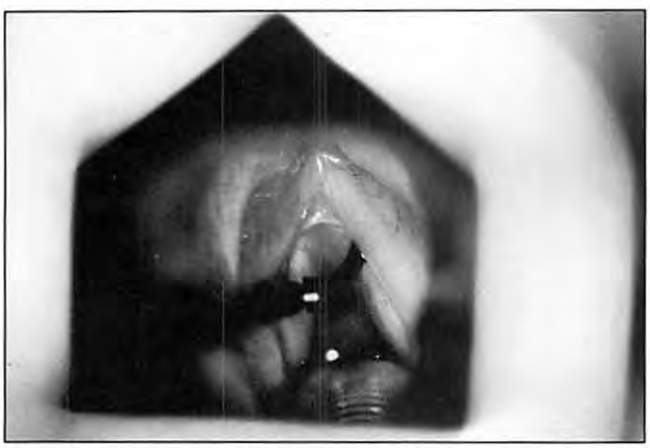


Photo 3. An anterior commissure microweb is demonstrated by gentle separation of the vocal folds at the anterior commissure with a curved microforceps. This is a typical appearance of a microweb after secretions have been cleared. The web is very thin and located on the infraglottic surface of the vocal folds.

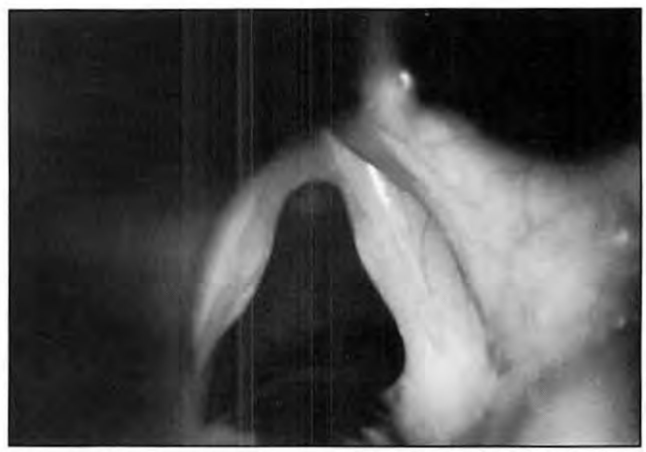


Photo 4. The endoscopic appearance of the larynx in Case 2 reveals an anterior commissure microweb associated with bilateral fusiform nodular lesions.

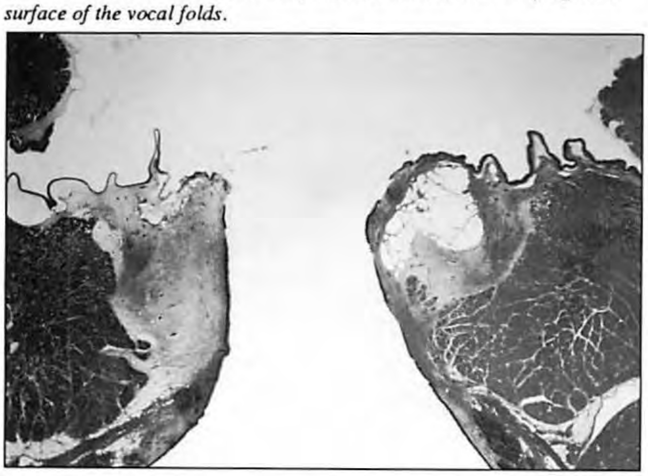


Photo 5. Coronal section through normal vocal fold (left) and fat-grafted vocal fold (right).

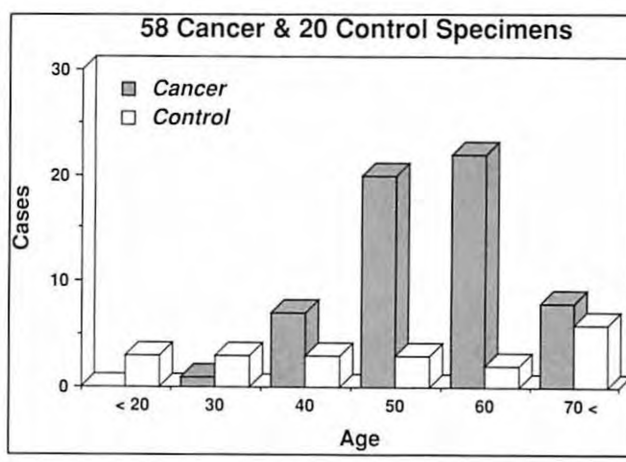


Photo 6. Age distributions of 58 cancer and 20 control specimens.

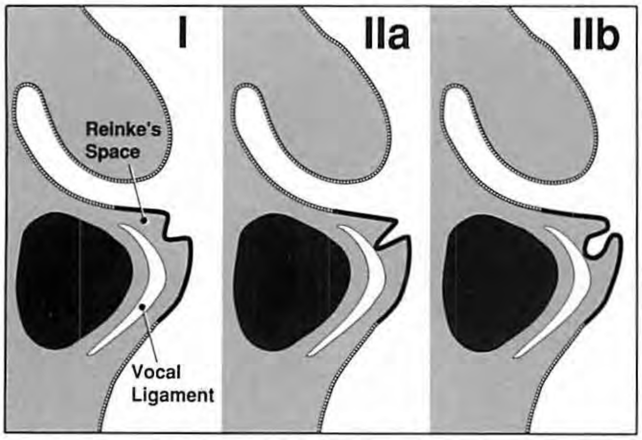


Photo 7. Schematic illustrations of the morphological classification of sulcus deformities.

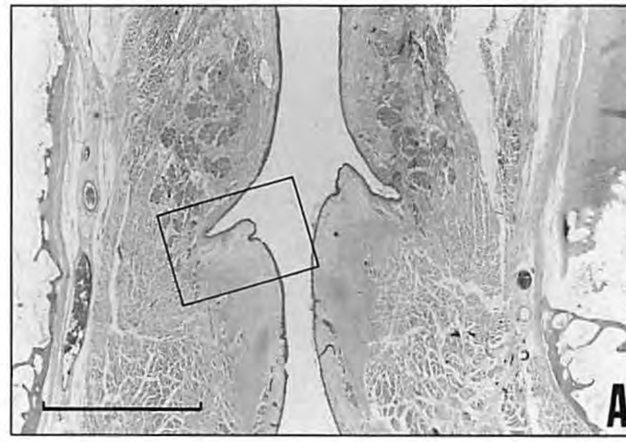


Photo 8A. 70-year-old male, T2 supraglottic cancer. A: Coronal specimen showed bilateral superficial type (I) sulcus (scale=5 ram).

Photo 8B: Same patient: High power view of the left type I sulcus (H&E, original magnification x20).

Photo 9A. 43-year-old male, T4 supraglottic cancer. A: Specimen showed deep type II-a sulcus on the left vocal fold, type I sulcus on the right (scale=5 ram).

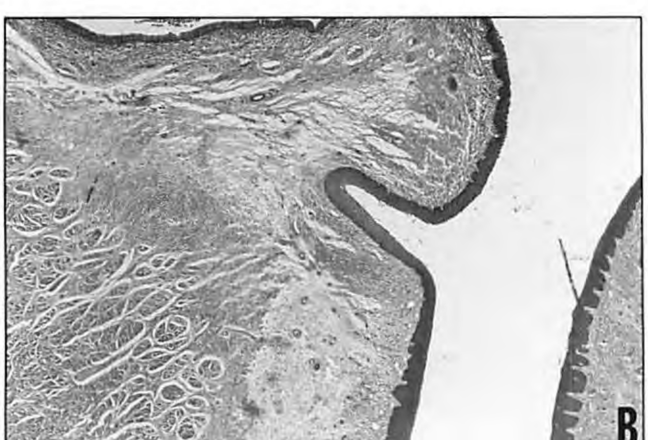


Photo 9B: Same patient. High-power view of the left type II-a sulcus (H&E, original magnification x20).

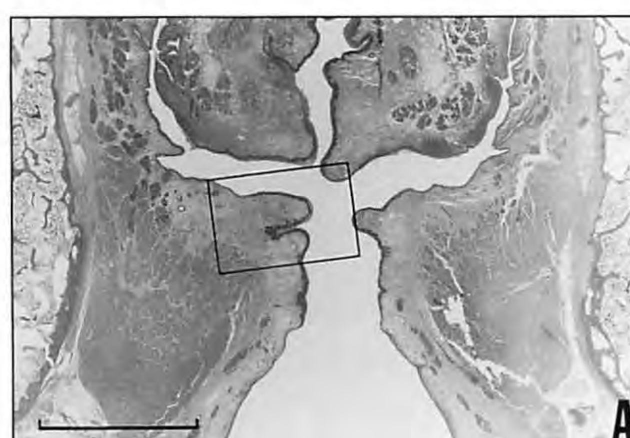


Photo 10A. 61-year-old male, T2 supraglottic cancer. A: Specimen showed pouch type II-b sulcus on the left vocal fold (scale=5 mm).

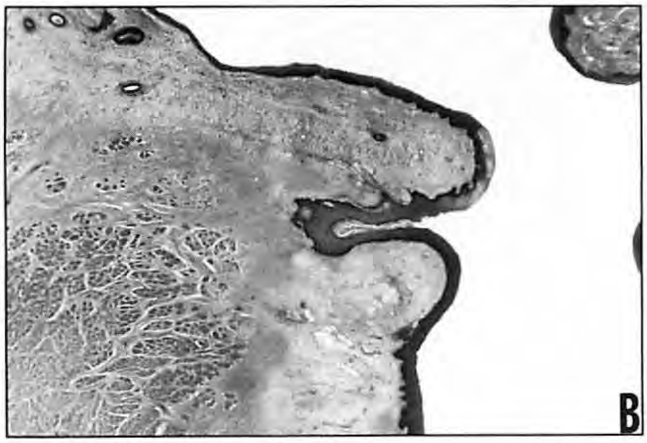


Photo 10B: Same patient: Highpower view of the left type II-b sulcus (H&E, original magnification x20).

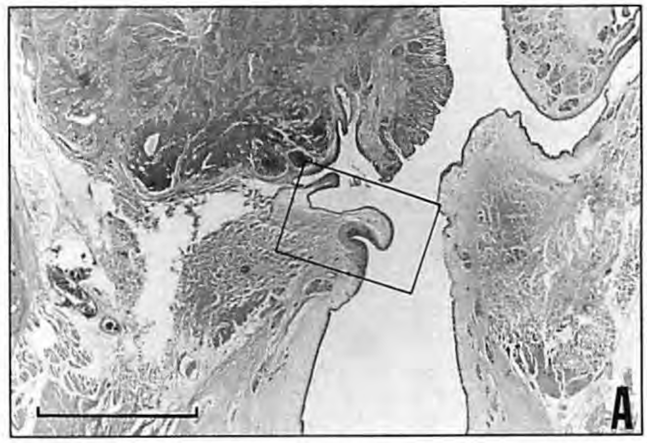


Photo 11A. 65-year-old male, T3 supraglottic cancer. A: Large supraglottic cancer mass neighboring the left vocal fold, which exhibited type II-a sulcus (scale=5 mm).

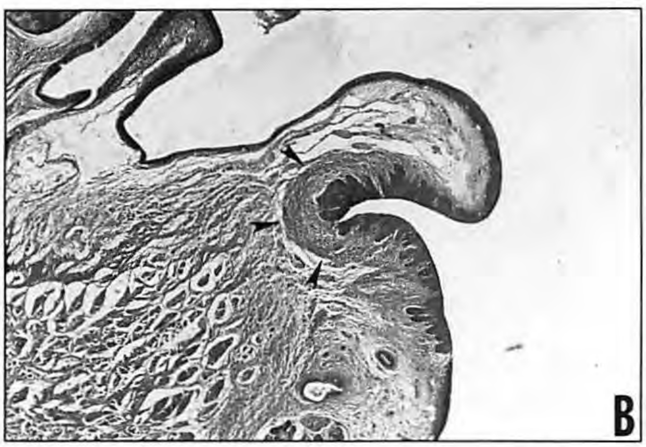


Photo 11B: Same patient: High-power view of the left fold showed dense fibrous proliferation (arrows) adjacent to the sulcus (H&E, original magnification x20).

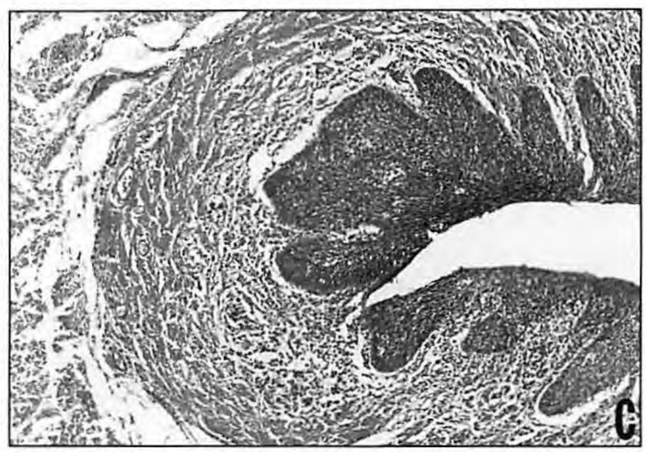


Photo 11C: Same patient: Higher power view of the fibrous proliferation (x100).

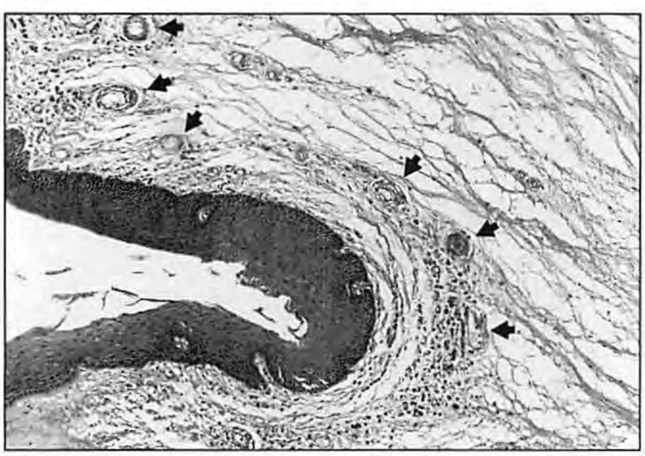


Photo 12: Same case as in Photo 8. High-power view of the right sulcus showed multiple enlarged capillaries (arrows) adjacent to the sulcus, indicating "increased" vascular ingrowth (H&E, original magnification x100).

Photo 13A: 62-year-old male, T3 glottic cancer. Type II-a sulcus on the right vocal fold (arrow) with large pedunculated cancer mass opposite it (scale=5 mm).

Photo 13B: 59-year-old male, T3 supraglottic cancer. Type II-b sulcus on the right vocal fold (arrow) with large cancer mass adjacent to it (scale=5 mm).

Photo 14: 67-year-old male, control case. Type I superficial sulcus on the right vocal fold (arrow) (scale=5 mm).

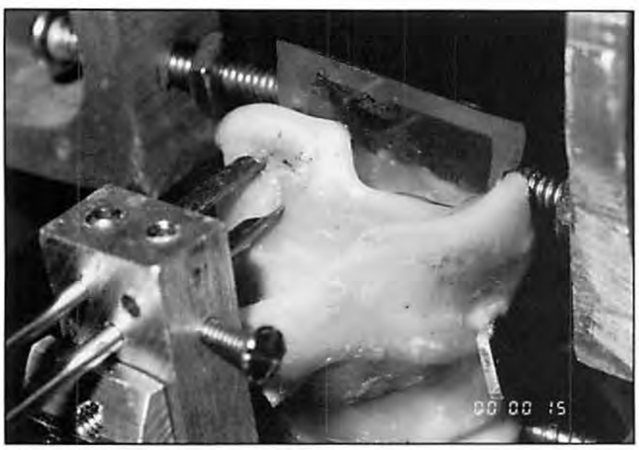


Photo 15: Hemilarynx preparation.

Photo 16: Vocal fold contact against glass plate.

Part II

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

NCVS Status and Progress Report - 4
June 1993, 271-280

Toward Standards in Acoustic Analysis of Voice

Ingo R. Titze, Ph.D.

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

Abstract

A brief summary of the purpose of standardization is given. This is followed by a few suggestions about methods and procedures in acoustic voice analysis that could benefit from consensus, if not fully developed standards. In particular, better agreement on test utterances, techniques for recording voices, and methods for extracting fundamental frequency is called for. Some new databases and calibration materials are in need of development, with better portability and shareability in mind.

Introduction

This article serves two purposes. The first is to review the rationale for standards in any field, recalling the benefits and possible liabilities that are part of standardization. The second purpose is to propose a few procedures that could (in time) be standardized in a specific area of acoustic analysis of voice, namely assessment of laryngeal control in phonation.

There is no claim to originality of any of the material presented here. Many articles and books have been written on every aspect of standardization. But every field of science somewhere has to rediscover the steps that lead to better communication, better product control, and greater economy in having several people at different locations working on similar things. This is what standardization is really all about: fostering a team approach, even though the players don't always practice or play together on the same field.

General Discussion of Standards

In What Way are Standards Helpful?

The rationale for standardization is outlined clearly by Sullivan[1]. First, standards *educate.* In the process of writing a standard, unusual attention is given to detail and to proper definition of terms. This allows newcomers to the field to be calibrated to the level of the science and to some of its history. Often ideals are set forth for future achievement.

Second, standards *simplify.* In industry, the number of stock sizes, shapes, and material components are limited to make the selection process easier. This then also simplifies accounting, advertising, storing, and other handling procedures. By reducing the variety of the processes, more attention can be given to improving the quality of a few.

This leads to the third point. Standards *conserve.* They save time, money, and effort. Better tools can be designed when multiple copies of the same item are produced. Henry Ford is credited with mass-production of the automobile. Although consumers are often intrigued by handcrafted, custom-made products, they also appreciate the value of assembly line merchandise.

Finally, standards help to *certify.* Institutions that give professional degrees and accreditation usually have standardized some criteria for giving a stamp of approval. Also, when patents or trademarks are sought by individuals or companies, a description is needed that provides some disclosure of the idea, design, or procedure. This allows the consumer the option to put his trust into an accreditation or protection agency, or to test the product himself.

In What Way are Standards not Helpful?

There is also a downside to standardization; in particular, premature standardization or overstandardization can be a hindrance to progress. Just as an economy can be stifled by too much regulatory action, so can a scientific field be slowed down by too much zeal for order. Premature standardization may squelch personal initiative and entrepreneurship. There are always individuals who prefer to go against the grain. By putting the steering wheel on the right side of the car, or by turning an electroglottogram (EGG) upside down, clever dissenters discover something that the rest of the world did not see. Standards should therefore not be a threat to those who do not wish to conform; they should at most be a slight inconvenience.

The inconvenience may actually become an unnecessary burden, however. Those who have the flexibility to deal with large varieties of processes and products may find it limiting (or even frustrating) not to be able to have it their way. An example is the carpenter who can only get certain sizes of lumber to build an odd-shaped structure. He either has to cut pieces that are too long or make undesirable seams by combining two shorter pieces.

In addition, standards may be confusing or erroneous. Ambiguities may have been created by the writers of a standard that may make it difficult to interpret even the intent, much less the detailed procedure. Everyone then uses their own interpretation. The result is an ineffective standard and a cynical attitude about standardization. In a worst case scenario, an entire field may be lead astray by a hidden error. For this reason, standards require frequent reviewing and updating.

The Process of Standardization

The process of standardization usually begins with an individual, or a small group, perceiving a need to unify and simplify. In some cases, it is a need to protect the public from certain hazards. In either case, the decision is made early whether the standard should be *voluntary* or *mandatory*. Mandatory standards require a regulatory agency for enforcement, which makes the process much more complicated and expensive. Fortunately, the greater percentage of standards is voluntary. Basically, the field then regulates itself and requires no outside arbitration.

The keyword in voluntary standardization is *consensus*. Majority rule does not get the job done. All minority opinions must be heard exhaustively until there is no reasonable argument against the standard. Otherwise, disgruntled individuals or groups will crusade against the standard and voluntary adherence becomes improbable. Ultimately, of course, a large majority will sway the dissenters simply on the basis of convenience and economics.

An appropriate professional society is sought out to establish a working group. This society typically has a liaison with the American National Standards Organizations (ANSI) or the International Standards Organizations (ISO) who provides guidance in policy and procedures. The working group reads all the pertinent literature and begins to formulate some proposals. Representation is then sought from industry, education, government, consumers, and any other groups that may have an interest in the standard. Lengthy discussions ensue, the aim being consensus at every step of the way.

Areas Where Some Consensus May be Achievable

Acoustic analysis of voice is performed for many reasons. Some of the more obvious ones are telecommunications (speech transmission), basic studies in speech science and linguistics, and assessment of disorders in human communication. The technical needs for these applications are quite varied and cannot easily be described in a single paper. The focus here will be the use of acoustic voice analysis to determine the *phonatory capabilities of the larynx*. Phonatory capability includes, as a minimum, control of pitch, loudness, phonation mode, and register. We consider these four variables to be fundamental in control of the larynx.

If phonatory capability is to be assessed acoustically, each of the control variables must have an acoustic counterpart, a quantity that is measurable from a microphone signal. Table 1 shows a pairing of the control variables with primary acoustic variables, the first two being rather obvious. As far as phonation mode is concerned, it has been difficult to find a single parameter that quantifies changes in spectral envelope when the voice changes from breathy to pressed. The strength of the fundamental component, relative to the harmonics, is one useful indicator. Register is even more difficult to define in terms of one or two spectral or temporal parameters. A temporal gap is a primary feature of pulse (fry) register, whereas an abrupt spectral slope transition is a primary feature of a modal-falsetto register change.

It is clear that none of the control variables have a single acoustic dimension. Loudness, for example, is governed not only by intensity, but also by spectral balance, and pitch is influenced by both intensity and spectrum. The non-orthogonality between control variables and acoustic variables is an inconvenience, but not a major theoretical stumbling block.

Following the concepts developed in motor control of limb and body movement, each of the control variables can presumably be assessed in terms of (1) strength, as measured by the range of a control variable, (2) accuracy in executing a planned movement patterns, (3) stability in maintaining a posture, and (4) speed of executing a task. For these assessments, it is useful to develop a battery of test utterances from which the relevant acoustic measures can be extracted.

Table 1.
Phonatory Control Variables and Acoustic Correlates

Control Variables	Acoustic Variables
loudness (soft to loud)	intensity
pitch (low to high)	fundamental frequency
mode (breathy to pressed)	fundamental to harmonics ratio
register (falsetto/modal/pulse)	temporal gap or spectral slope

Design of a Vocal Treadmill (Test Utterances)

The traditional clinical goals of constructing test utterances are to determine (1) how voice impacts on speech intelligibility and communication effectiveness, and (2) what insight can be gained about laryngeal health or general body condition. An additional goal, a pedagogical one, would be to determine (3) how the effectiveness of vocal training can be quantified.

Historically, clinicians have used a battery of test utterances that progress from primitive vocalizations to isolated syllables or words to complete sentences or paragraphs. Almost everyone is in agreement that the tasks must reveal control of pitch, loudness, and some aspect of vocal quality. In addition, the interaction between respiratory, phonatory, and articulatory components of speech are important to most clinicians. The following list includes nearly all of the test utterances that have been described in the literature:

1. Sustained expiration with no major vocal tract constriction to assess airflow management (tidal volume, vital capacity, inspiratory and expiratory reserves)
2. Prolonged [s] and [z] to compare airflow management with one or two vocal tract con strictions
3. Sustained vowels, typically some combination of [a], [e], [i], [o] or [u], at
   a) comfortable pitch and loudness
   b) prescribed pitch and loudness (matched tones)
   c) gliding pitch and loudness
4. Prolonged vowels to assess phonatory endurance at
   a) comfortable (habitual) pitch and loudness
   b) prescribed pitch and loudness (high, medium, low)
5. Singing of scales on vowels (including falsetto register) to determine phonation range and stability
6. Repeated syllables, such as [pæ] or [pa]
7. Counting or using some emotionally neutral utterance like "ah-hum" to assess normal speaking pitch and intensity. A vowel in one of the isolated words can be prolonged to assess F_o more accurately. Variations include fast, slow, soft, loud, high, low counting
8. Chant talk, an alternation and mixture of speaking and singing to obtain optimal pitch

9. Speaking of prescribed sentences to simulate situations such as:
    a) soft conversation, near whisper
    b) normal conversation
    c) loud conversation, or shouting
    d) exaggerated articulatory movement (clear speech)
    e) accelerated rate (rapid speech)
    f) simulated psychological stress or emotion
    g) during or after physical exertion (running, lifting, pushing, etc.), or relaxation (meditation, etc.)
    h) assuming different roles (lecturer, boss, subordinate, etc.)
10. Oral reading of a passage (with variations as in 9)
11. Conversational speech (with variations as in 9)

Assessment of vocal control across clinics, studios, and laboratories could benefit from some consensus reached in the use of diagnostic utterances, particularly if they also were to serve as a vocal treadmill. By "treadmill" we mean that some of the utterances would be taxing enough to expose both the strengths and the weaknesses of the larynx in control of pitch, loudness, mode of phonation, and registration. The use of vowels and conversational speech alone does not seem to do that.

Table 2 is a proposed set of utterances of this type. The top half of the table lists a set of nonspeech utterances and the bottom half lists some speech utterances. The primary aim in constructing this table was completeness, not brevity at this point. The battery is probably too hefty to be used for clinical diagnostics, but trimming is a second task. As it stands, the table includes most of the utterances used historically, but expands the list significantly in the direction of dynamic testing, that is, using phonatory *glides* to allow assessment of coordinated muscle activity in the larynx and respiratory system.

All utterances are customized to the individual's Voice Range Profile (VRP). This must be obtained first to establish the bounds for further testing. Low, medium, and high pitch can then be defined as some percentage of the F_o range, say 10%, 50%, and 90%. The same can be done to define soft, medium, and loud intensity. With these definitions, sustained vowels are elicited at strategic locations within the VRP to determine phonatory stability. This is followed by [s] and [z] consonants and, finally, by a series of pitch, loudness, adduction, and register glides.

In the second half of the table, speech material is used with increasing phonetic, emotional, and artistic complexity. After traditional counting, an all-voiced sentence is first used to test F_o control independent of adductory control. Three loudness conditions can be used: (1) as speaking into someone's ear, (2) as speaking across the dinner table, and (3) as shouting across a busy street. This is followed by a sentence with frequent voicing onset and offset, again at different loudness. The Rainbow Passage and the "Cookie Theft" description are then administered as *de facto* standards. At this point, the treadmill advances to some parent-child speech. It is expected that exaggerated F_o, intensity, and register patterns will emerge in this test as subjects mimic typical parentese. Further testing of extreme F_o and intensity patterns (with highly expressive vocalizations) comes with a dramatic recitation, such as one of Shakespeare's soliloquies, or recounting of a highly emotional experience. Finally, a portion of a familiar song ("Happy Birthday") is sung in both modal and falsetto register to examine "heavy" and "light" production in a singing mode.

Table 2.
Proposed Test Utterances

NONSPEECH	
Voice Range Profile	defines test frequencies and intensities (low = 10% of F_o range, medium = 50% of F_o range, high = 90% of F_o range; soft = 10% of intensity range, medium = 50% of intensity range, loud = 90% of intensity range)
Sustained [a], [i, [u] Vowels	1. low, soft, 2s 2. low, loud, 2s 3. high, soft, 2s 4. high, loud, 2s 5. medium high, medium loud, 2s 6. comfortable pitch and loudness, 2s 7. comfortable pitch and loudness, maximum duration
Sustained [s] Consonant	comfortable pitch and loudness, maximum duration
Sustained [z] Consonant	comfortable pitch and loudness, maximum duration
Pitch Glides	1. low-high-low, one octave, 0.1 Hz 2. low-high-low, one octave, 2.0 Hz 3. low-high-low, one octave, maximum rate
Loudness Glides	1. soft-loud-soft, 0.1 Hz 2. soft-loud-soft, 2.0 Hz 3. soft-loud-soft, maximum rate
Adductory Glides [a] and [ha]	1. onset-pressed-offset, 0.1 Hz 2. onset-pressed-offset, 2.0 Hz 3. onset-pressed-offset, maximum rate
Register Glides	1. modal-pulse-modal, 0.1 Hz 2. modal-falsetto-modal, 0.1 Hz 3. modal-falsetto-modal, maximum rate, as in yodeling
SPEECH	
Counting from 1 to 100, comfortable pitch and loudness	
All voiced sentence, "Where are you going?", soft, medium, loud	
Sentence with frequent voice onset and offset "The blue spot is on the key again", soft, medium, loud	
Oral reading of "Rainbow Passage"	
Descriptive speech, "Cookie Theft" picture	
Parent-child speech, "Goldilocks and The Three Little Bears"	
Dramatic speech involving deep emotions (fear, anger, sadness, happiness, disgust)	
Singing part of "Happy Birthday to you", modal and falsetto register	

A major unanswered question is whether or not a person's ability to speak or sing can in any way be assessed with nonspeech tasks. One would hope that a wide range of pitch and loudness in the Voice Range Profile, for example, would predict highly expressive intonation, stress, and loudness patterns in speech, but there is no guarantee of that. For assessment of voice disorders, large inaccuracies in pitch and intensity glides should be a predictor of abnormal prosodic contours in speech, but again, this remains an open research question.

Table 3.
Nonspeech Tasks for Assessment of Laryngeal Motor Control

	range	stability	speed	accuracy
pitch	Voice Range Profile	Sustained Vowels	Rapid Pitch Glides	Slow Pitch Glides
loudness	Voice Range Profile	Sustained Vowels	Rapid Intensity Glides	Slow Intensity Glides
adduction	Slow Adductory Glides	Sustained Vowels	Rapid Adductory Glides	Slow Adductory Glides
register	Slow Register Glides	Sustained Vowels	Rapid Register Glides	Slow Register Glides

Table 3 shows how range, stability, speed, and accuracy of the primary control variables of the larynx might be assessed with selected portions of the nonspeech utterances. Range of pitch and loudness are assessed with the Voice Range Profile (VRP). Range of adduction and register are assessed with the slow glides. Stability of posturing is assessed with the sustained vowels, and speed and accuracy are assessed with the rapid and slow glides.

A major problem with administration of any task is to determine the cognitive ability of the subject or patient. Do they understand what they are supposed to do? Ample dialogue between the clinician and the patient can, of course, minimize this problem. But what if administration of the tasks is to be automated? Is it sufficient to give examples by video tape, audio tape, or by some other type of pre-programmed message? Some consensus along these lines would be helpful.

Another problem may be a perceptual deficit of the subject. Are auditory or visual stimuli that may be used to elicit the vocal tasks adequately perceived by the listeners? If not, it may be necessary to conduct some auditory or visual training to acquaint subjects with octave glides, crescendos, diminuendos, register changes, or other vocal exotics.

Once the task is clear to the subject, there is still the question of practice to get the level of performance that is most useful for diagnostic purposes. The objective here is not necessarily to get top level performance, but there should be sufficient success to deem the task completed. If a treadmill is to be useful, a person must be able to walk on it at some speed and for some length of time. But the speed of the mill or the length of the walk may not have to be the ultimate (or the only) measures of physical fitness. Other tests can be made, such as a heartrate, blood pressure, and general laboratory analysis. Likewise, subtle vocal irregularities (tremor, involuntary register shifts, intermittent phonation, subharmonics) may give the complete story of vocal fitness. The vocal treadmill acts as a catalyst to expose the system to induce "vocal stumbles".

Measurement of Fundamental Frequency (F_o)

Extraction of F_o is necessary for most of the acoustic measures obtained from the vocal utterances. There is still much uncertainty among investigators with regard to extraction of F_o and its variability. The field could benefit from some consensus with regard to the following:

1. Clarification of the meaning of F_0 when signals show evidence of bifurcations, chaos, or highly stochastic behavior
2. Determination of an upper limit of the extent of perturbation for which jitter and shimmer measures have validity
3. Selection of appropriate microphone type and placement (with respect to the mouth) for the highest fidelity F_0 extractions
4. The effect of noise, room acoustics, and source-receiver stationarity on jitter and shimmer measures

Some work is underway in these areas. For example, methods of nonlinear dynamics are being applied to voice signals to classify them in terms of well-known attractors[2,3]. It appears that the signals fall into three distinct categories; periodic with small random perturbations, periodic with subharmonic structure and modulation, and nonperiodic (chaotic). Traditional jitter and shimmer analysis can be applied only to the first category, but it is not yet clear how to determine the boundary between the categories.

Some work is also being done to select the most appropriate microphones[4]. Preliminary results suggest that microphone sensitivity, shielding from groundloops, and distance from the mouth can all affect measured jitter and shimmer of normal voices. This is significant in light of the earlier study by Doherty and Shipp[5] on the effect of tape recorders on jitter and shimmer measurements. A complete set of recommendations is now needed for the entire recording system, including the room environment.

Database of High Fidelity Recordings

Once consensus has been reached on test utterances and recording techniques, it would seem appropriate to build a new database of voice recordings. This database would include representative numbers of normal and abnormal populations "stepping onto the vocal treadmill" and following the protocol agreed upon. This database should then be made available to a large number of researchers and clinicians for comparative studies.

Synthetic Calibration Materials

In addition to the database of human voice recordings, the field would benefit from a randomly accessible library of synthetic voice samples. This library could be distributed on DAT tape, CD ROM or over computer networks. Table 4 shows an outline of how such a library might be structured. Agreement would be reached on some constants, such as stimulus duration, magnitude, and perhaps onset and offset duration. Parameters to be varied would include waveform type, fundamental frequency, sampling frequency, vowel, modulation (type, frequency, and extent), and additive noise. A triangular (or ramp) modulation would provide the synthetic equivalent of glides produced by human subjects. This would allow ample comparisons between natural and synthetic voice stimuli.

File Structures

In order to make databases less site and machine specific, it would be advisable to agree on a highly flexible file structure. One suggestion[6] has been to use only single channel binary data strings (16 bits, with perhaps a later expansion in mind). No delimiters would be used for data records, and the file header would contain only character strings. The header would have an agreed-upon fixed length that is known to everyone.

Table 4.
Synthetic Voice Samples

Constants:	
Duration:	5s
Onset:	0.1s
Offset:	1.0s
Magnitude:	±20,000 peak to peak (computer units)
Parameters:	
Signal Type:	sinusoids; glottal flow, mouth flow, mouth pressure, and contact area analogs
Fundamental Frequency:	approximately 100 Hz, 200 Hz, 400 Hz, but non-integer
Sampling Frequency:	20 kHz, odd-ball (44.1)
Vowels:	[a], [i], [u]
Modulation Type:	sinusoidal, Gaussian, ramp; AM and FM for each
Modulation Frequency:	1 Hz, 5 Hz, 50 Hz, $F_o/2$, $F_o/3$
Modulation Extent:	0.05%, 0.5%, 5%, 10%, 50%
Additive Noise:	40 dB, 80 dB SNR

The text strings in the header would contain typical file specifications, such as file length, sampling frequency, patient or subject information, date of recording or generation, or any other useful information. Users of the shared databases would need to program their computers to read the unformatted text strings rather than field-specific number-character combinations (the typical header organization).

Nomenclature

Some agreement on nomenclature would be helpful, but this is not as important as other foregoing considerations. It was argued before(7) that an attempt should be made to use accepted engineering and statistic terminology, such as root-mean-squared, mean rectified, mean squared, variance, coefficient of variation, modulation index, etc. in making actual calculations on the voice signal. Terms like jitter, shimmer, tremor, vibrato, flutter, etc. are best retained as generic terms that give a qualitative feeling for the variability. Assigning numbers to them may be taking a step backwards.

Terminology of nonlinear dynamics has been helpful in sorting out some of the ill-defined terms for vocal quality. For example, the term diplophonia is ascribed to a voice for which two independent pitches are perceived. The attractor in phase space for such a signal is a torus, assuming that two incommensurate frequencies indeed exist. On the other hand, creaky voice often refers to phonations for which subharmonic frequencies occur. The attractor for this signal is a bifurcated limit cycle. If two or more pitches are perceived for these phonations, they should be in octave ratios. Creaky voice should therefore not be confused with diplophonia. Creaky voice becomes vocal fry (or pulse or register) when a subharmonic frequency dominates (perceptually) and is below about 70 Hz.

The term Voice Range Profile has recently been adopted by the International Association of Logopedics and Phoniatrics as a display of intensity range versus fundamental frequency. The display has also been referred to as *Stimmfeld* (German for voice area) or phonetogram in the literature. It would be helpful to settle on the name of this display.

Conclusion

Although it is difficult to identify procedures and measurement techniques in voice analysis that are clear candidates for formal standardization at this point, much progress can be made by reaching some preliminary consensus. In particular, the field is bogged down by lack of specificity in test utterances that have universal appeal and proven diagnostic value. Once such test utterances are defined, work can proceed more rapidly in establishing the technical criteria for extraction of fundamental frequency, intensity, and spectral measurements. Shared calibration materials and databases would then be extremely helpful.

Acknowledgement

This work was supported by grant No RO1 DC00387 from the National Institutes on Deafness and Other Communication Disorders. The manuscript was prepared by Julie Lemke.

References

1. Sullivan CD. Standards and Standardization: Basic Principles and Applications. New York: Marcel Dekker, Inc., 1983.

2. Herzel H, Berry D, Titze I, Saleh M. Analysis of vocal disorders with methods from nonlineary dynamics. JSHR in review.

3. Titze IR, Baken R, Herzel H. Evidence of chaos in vocal fold vibration. In: Titze IR, ed. Vocal Fold Physiology: Frontiers in Basic Science. San Diego: Singular Publishing, 1993.

4. Titze IR, Winholtz W. The effect of microphone type and placement on voice perturbation measures. JSHR in press.

5. Doherty E, Shipp T. Tape recorder effects on jitter and shimmer extraction. JSHR 1988;31:485-490.

NCVS Status and Progress Report - 4
June 1993, 281-290

Tongue Function Testing in Parkinson Disease: Indications of Fatigue

Nancy Pearl Solomon, Ph.D.
Donald A. Robin, Ph.D.
Daryl M. Lorell, M.S.
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

The role of weakness and fatigue in motor speech disorders is relatively unexplored. We have developed three techniques for the evaluation of the tongue for strength, endurance, and effort. The current report is part of a larger project aimed at examination of these functions in relation to speech and its disorders. The clinical relevance of these measures is being evaluated as well. The first two techniques, measures of strength and endurance, have been the focus of much of our work. The third measure, pertaining to the perception of effort, is a new addition to our test battery.

All three techniques use the Iowa Oral Performance Instrument (IOPI, Breakthrough, Inc.) to obtain measures of pressure exerted by the tongue and the hand. The IOPI is a pressure-sensing instrument with a digital display (in kPa) and an LED display. Detailed descriptions of the IOPI can be found in Robin, Somodi, and Luschei (1991) and Robin, Goel, Somodi, and Luschei (1992). The tongue is tested because of its obvious relation to speech. The preferred hand is tested to ascertain if the findings are generalized to other parts of the body.

For measures of maximum strength, subjects squeeze a small air-filled bulb as hard as they can. The tongue bulb is placed against the hard palate immediately posterior to the alveolar ridge. The subject pushes against the bulb with the anterior dorsum of the tongue. The hand bulb is placed in the palm of the preferred hand. The subject wraps his or her fingers around the bulb and squeezes.

Endurance is assessed by measuring the subjects' ability to maintain a pressure exertion on the bulb at 50% of their maximum strength. The IOPI is set up, by means of the digital display, so that the middle light on the LED display corresponds to 50% of the maximum pressure. The subject is instructed to squeeze the bulb so that the middle light is illuminated, and to keep that light on as long as possible. When the subject can no longer maintain 50% of maximum pressure, the trial is terminated. This measure should be a good indication of fatigue because, by definition, fatigue is a "failure to maintain the required or expected force" (Edwards, 1981).

Speech samples are recorded routinely as part of the tongue function evaluation. These involve syllable, word, phrase, and sentence repetitions for intelligibility analysis. In addition, a description of the "Cookie Theft" picture (Boston Diagnostic Aphasia Examination, Goodglass & Kaplan, 1981) is used for speech rate (calculated as the number of syllables produced per second excluding pauses >250 ms) and perceptual characteristics. Possible relationships between tongue function and speech impairment can be explored using this information.

To date, several studies from our laboratory have used a standardized protocol for the assessment of tongue and hand strength and endurance (Robin et al., 1991) in normal and disordered speakers. Lorell, Solomon, Robin, Somodi, Luschei, and Rodnitzky (1992) reported weaker than normal tongues but normal tongue endurance for a group of subjects with mild to moderate Parkinson disease. Lorell and Robin (unpublished observations) have documented reductions in tongue strength related to motor system degeneration in a patient with amyotrophic lateral sclerosis. Lorell, Robin, Somodi, Solomon, and Luschei (1992) observed reductions in tongue strength as a function of normal aging. Robin et al. (1991) found that children with developmental apraxia of speech had normal tongue strength but significantly reduced endurance compared to normally speaking children. In contrast, Robin et al. (1992) documented increased endurance times in subjects with exceptional skill with the tongue, including debaters and trumpet players.

We have recently developed the third measure, an index of perceived effort, which we believe may have clinical utility. The evaluation of sense of effort was motivated by the obvious increase in effort that occurs during the endurance task. The fact that increased effort invariably accompanies tasks involving sustained muscular contractions with a constant force output has lead to revised definitions of fatigue that include both peripheral and central components. Enoka and Stuart (1985) define fatigue as "a progressive increase in the effort required to exert a desired force and the eventual progressive inability to maintain this force in sustained or repeated contractions" (p. 2281). The perception of effort is thought to be due to an awareness of descending motor drive by means of central feedback (e.g., Gandevia, 1982; McCloskey, 1981). If effort is high and prolonged, fatigue is experienced irrespective of the state of muscle contraction. This is referred to as "central fatigue."

Reports of effort and fatigue in people with neurologic disease, in combination with an inability to sustain a task, may implicate a mismatch between descending motor commands and force production at the periphery. It has been suggested that the performance of complex motor patterns rely on the programming of effort as an integral part of volitional output (McCloskey, 1981). Neilson and O'Dwyer (1984) hypothesized that abnormal internal feedback may play a part in the abnormal movements of people with athetoid cerebral palsy.

The task we have developed for assessing perceived effort involves squeezing the IOPI bulb to various levels of effort (from 10% to 90% of maximum in 10% increments). A visual display, in the shape of a thermometer, was used to cue subjects by marking the target effort level before each trial. Effort levels were tested in random order. A study using this technique was conducted with 20 young, healthy subjects (Somodi, Robin, & Luschei, submitted). Findings indicated a highly consistent sense of effort within and across subjects that related well to the actual pressures produced. The mathematical function describing the relation between effort and pressure was a 3rd-order polynomial. Specifically, the function was continuously rising with a relatively flat mid-section. This suggests that the low and high extremes of effort are more sensitive for pressure changes than is the range from approximately 30% to 80% of maximum effort. Of interest was that knowledge of 100% effort was not needed to calibrate submaximal effort levels. Similar results were obtained for both the tongue and hand.

People with Parkinson disease often report feelings of fatigue and increased effort while performing daily activities of life (Mayeux, Stern, Williams, Cote, Frantz, & Dyrenfurth, 1986; McDowell, 1971). Weakness, fatigue, and perceptions of effort in the use of the tongue may relate to some of the speech characteristics of hypokinetic dysarthria, especially imprecision of articulation (Canter, 1965; Darley, Aronson, & Brown, 1975; Logemann, Fisher, Boshes, & Blonsky, 1978; Solomon & Hixon, in press). In our study of 23 people with mild to moderate Parkinson disease (Lorell, Solomon, et al., 1992), we were surprised to find no difference between subjects with Parkinson disease and control subjects for tongue and hand endurance. To further explore the basis for the perceptions of fatigue, we conducted follow-up studies with 3 of the subjects who reported fatigue. Subjects were again tested for strength and endurance and a speech sample was collected. In addition, they performed the task for perceived effort.

In this preliminary report, we present the cases of 3 subjects with mild parkinsonism and complaints of fatigue, and data from 3 neurologically normal control subjects matched for sex, age, and weight. A list of subject characteristics and results for strength and endurance is provided in Table 1. Unfortunately, we did not test the control subjects for perceptions of effort; instead, comparison data from the study of young healthy adults were used (Somodi et al., submitted).

Table 1.
Subject Characteristics and Test Results

Subject	Age (yrs)	Weight (kg)	Strength tongue (kPa)	Strength hand (kPa)	Endurance tongue (s)	Endurance hand (s)	Speech Rate (syllables/s)	Overall Speech
Mrs. S.	65	79.1						
Eval. 1			70	112	45	83	5.62	mild-to-moderate
Eval. 2			77	129	54	31*	5.14	mild-to-moderate
Control	64	77.7	53	169	43	60	4.79	normal
Mrs. H.	71	55.4						
Eval. 1			29	149	19; 55*	52	4.98	mild
Eval. 2			27	137	9*	55	5.03	mild
Control	72	55.0	50	126	32	30	5.27	normal
Mr. N.	43	92.2						
Eval. 1			53	273	6	14	4.23	mild
Eval. 2			39	239	18	39	3.36	mild
Control	45	92.2	70	156	25	24	5.44	normal

* See text for explanations.

Mrs. S.

Pertinent History and Neurologic Examination. Mrs. S., a 65 year old ambidextrous woman, was diagnosed with idiopathic Parkinson disease on September 13, 1991. At that time, her neurologist assessed the disease severity as mild (Stage 2 on the Hoehn and Yahr scale, 1967). Levodopa-carbidopa (25/100, t.i.d.) and Vitamin E were prescribed. Mrs. S. reported no change in her speech or facial expression since the onset of parkinsonian symptoms in the left arm 2 years earlier.

Initial Evaluation. Our evaluation on the same day (9/13/91), before medications were taken, revealed moderately masked facial expression and mild-to-moderate hypokinetic dysarthria. Her voice was markedly monopitch, speech rate during a picture-description task was fast (5.62 syllables/s with pauses excluded), phrases produced on one breath were long, and articulation was mildly imprecise. These characteristics are consistent with mild-to-moderate hypokinetic dysarthria. The matched-control subject's speech was normal, and her interpause speech rate was 4.79 syllables/s.

Evaluations of tongue and hand strength and endurance were conducted. Using the IOPI protocol, maximum tongue pressure was 70 kPa and maximum right-hand pressure was 112 kPa. Endurance at 50% of the maximum pressure was 45 s for the tongue and 83 s for the right hand. These data are comparable to those for the matched-control subject whose tongue strength was 53 kPa, hand strength was 169 kPa, tongue endurance was 43 s, and hand endurance was 60 s. In addition, Mrs. S.'s strength measures are similar to, and endurance measures are better than, those obtained from neurologically normal women of the same age (Lorell, Robin, et al, 1992). We conclude from this evaluation that, although Mrs. S. exhibits mild Parkinson disease and mild-to-moderate hypokinetic dysarthria, she has normal strength and endurance of the tongue and hand.

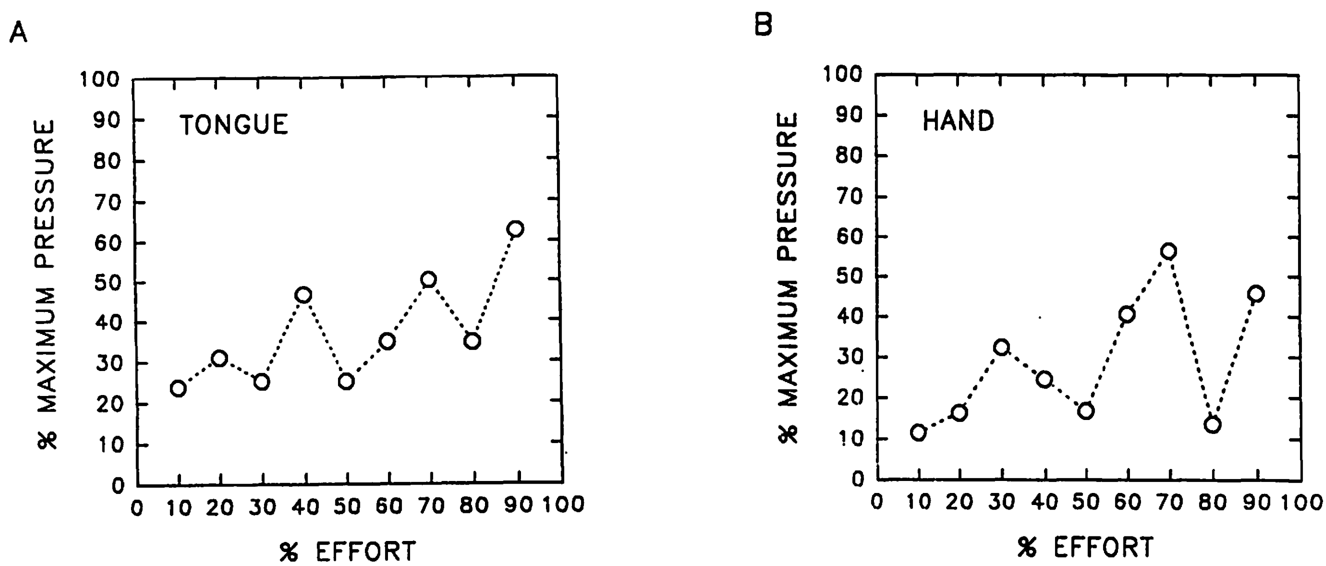


Figure 1. Effort results for Mrs. S. Pressures produced as a percentage of maximum pressure for various levels of effort for the tongue (A - left) and hand (B - right). Maximum pressure: tongue = 70 kPa, hand = 112 kPa.

Follow-up Evaluation. We evaluated Mrs. S. again on 1/31/92. Testing began 4 hours after her morning dose of levodopa-carbidopa was taken. Similar speech characteristics as those observed during the initial session were noted. Maximum tongue strength was 77 kPa, and hand strength was 129 kPa. Endurance was 54 s for the tongue. Mrs. S. maintained 50% of maximum pressure with the hand for 31 s, but admitted to terminating the task early due to generalized fatigue. Endurance was tested at the end of this 1 hour session. Except for hand endurance, these results are similar to those obtained during the initial evaluation.

Before the endurance tests were conducted, perceptions of various levels of effort were assessed using the protocol described by Somodi et al. (submitted) and summarized previously. The results from one trial with the tongue are plotted in Figure 1a, and for the hand, in Figure 1b. For the tongue, the pressures exerted generally increased with increasing effort levels, but the pattern was more scattered than is typically seen for normal subjects (Somodi et al., submitted). The data are even more variable for the hand, indicating little relation between pressure and effort except at the lowest effort levels. Interestingly, for both the tongue and hand, the pressures tended to be in the lower portion of the total range available. Non-neurologically impaired young adults used a wider range of the available pressures for this task (20%-90%), and exerted relatively high pressures when asked to use high levels of effort (Somodi et al., submitted). Thus, although strength and endurance measures were normal, the perception of effort appeared to be abnormal.

Mrs. H.

Pertinent History and Neurologic Examination. Mrs. H., a 71 year old right-handed woman, was diagnosed with idiopathic Parkinson disease in May, 1988. On November 19, 1991, her neurologist assessed the severity of disease as mild-to-moderate (Stage 2.5 on the Hoehn and Yahr scale). Mrs. H., an avid tennis player for 20 years, has had to reduce her activity level recently. Mrs. H. reported that her speech is a little slower and quieter than it used to be.

Initial Evaluation. We evaluated Mrs. H. on November 19, 1991. She had taken a tablet of levodopa-carbidopa (either 25/100 or control-release 50/200; this information was unavailable because Mrs. H. was a subject in a double-blind medication study) 2 hr 40 min before testing. Her speech intelligibility was excellent, although mild articulatory imprecision was noted. Inconsistent vocal fry was perceived in her voice. Interpause speech rate was 4.98 syllables/s. The control subject's speech was normal and speech rate was 5.27 syllables/s.

Maximum tongue strength was 29 kPa, and maximum hand strength was 149 kPa. Endurance at 50% of maximum pressure was 19 s, and 55 s on a second trial, for the tongue, and 52 s for the hand. The tongue endurance task was repeated because Mrs. H. had difficulty controlling the pressure with her tongue. The matched-control subject's tongue strength was 50 kPa, hand strength was 126 kPa, tongue endurance was 32 s, and hand endurance was 30 s. The control-subject's values, with the exception of hand endurance which is reduced, are similar to data from other women of this age (Lorell, Robin, et al., 1992). Mrs. H.'s strength and endurance were normal for the hand. However, tongue strength was markedly reduced compared to normal and tongue endurance times were variable.

Follow-up Evaluation. We re-evaluated Mrs. H. for tongue and hand function on March 4, 1992. Mrs. H. reported worsening of motor signs, now involving the left leg in addition to the previously affected right arm and leg. She did not notice a change in her speech since her previous evaluation.

Testing began 15 minutes after levodopa-carbidopa was taken. Her speech intelligibility, articulation, and speech rate were essentially unchanged compared to the initial session. However, the vocal fry was absent.

Maximum strength was 27 kPa for the tongue and 137 kPa for the hand. Mrs. H. was able to maintain 50% of maximum pressure with the tongue for only 9 s, apparently because of difficulty controlling the pressure. Hand endurance was 55 s. Except for tongue endurance, these findings are consistent with the initial evaluation. Evaluation for the perception of various levels of effort revealed relatively typical findings for the hand but markedly unusual findings for the tongue. The results are illustrated in Figure 2. Pressures exerted by the hand increased with increasing effort (Figure 2B). The results for the tongue appear to be almost random (Figure 2A). It should be noted that, for both the tongue and hand, most of the pressures exerted are relatively high. Based on the results of this evaluation, it appears that Mrs. H. has significant difficulty controlling and monitoring tongue movements for non-speech tasks. However, her speech was very good and may not be affected by these abnormalities.

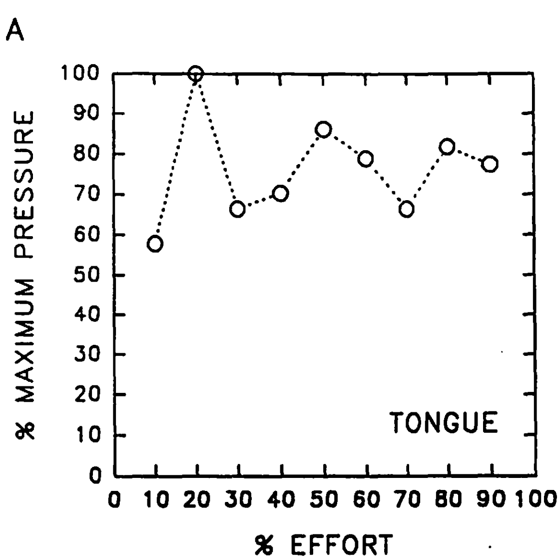


Figure 2. Effort results for Mrs. H. for the tongue (A - left) and hand (B - right). Maximum pressure: tongue = 31 kPa, hand = 149 kPa.

Mr. N.

Pertinent History and Neurologic Examination. Mr. N., a 43 year old left-handed man, reported noticing difficulty with movement on the right-side of the body in April, 1988. In February, 1991, he was diagnosed with parkinsonism that was considered to be secondary for etiology because of a history of drug-abuse (meperidine hydrochloride). One year after diagnosis (1/31/92), his neurologist described the disease severity as mild-to-moderate (Stage 2.5), and Mr. N. was on disability leave from his job as a warehouse worker. Mr. N. reported that his speech was mildly affected; vocal intensity decreased progressively during running speech and some "slurring" occurred.

Initial Evaluation. We evaluated Mr. N. for tongue function and speech on January 31, 1992. He took levodopa-carbidopa (25/250) 4 hr 30 min before testing. His speech was completely intelligible, interpause speech rate was slow (4.23 syllables/s), and voice quality was breathy. The matched-control subject had normal speech and an interpause speech rate of 5.44 syllables/s.

Tongue strength was assessed at 53 kPa, and left-hand strength was 273 kPa. Endurance at 50% maximum pressure was 6 s for the tongue and 14 s for the hand. The matched-control subject's tongue strength was 70 kPa, preferred (left) hand strength was 156 kPa, tongue endurance was 25 s, and hand endurance was 24 s. The control subject's data are typical for men his age with the exception of markedly reduced hand endurance (Lorell, Robin, et al., 1992). These data reveal that Mr. N.'s tongue strength was somewhat lower than normal, hand strength was much greater than normal, and endurance times for both the tongue and hand were substantially reduced.

Follow-up Evaluation. One month later (2/25/92), 3 hr after taking his antiparkinsonism medication, Mr. N. was re-evaluated. His speech was notably impaired. Intelligibility was still excellent, but monotony of pitch and loudness were pervasive, and interpause speech rate was markedly slower (3.36 syllables/s). Voice quality was similarly breathy.

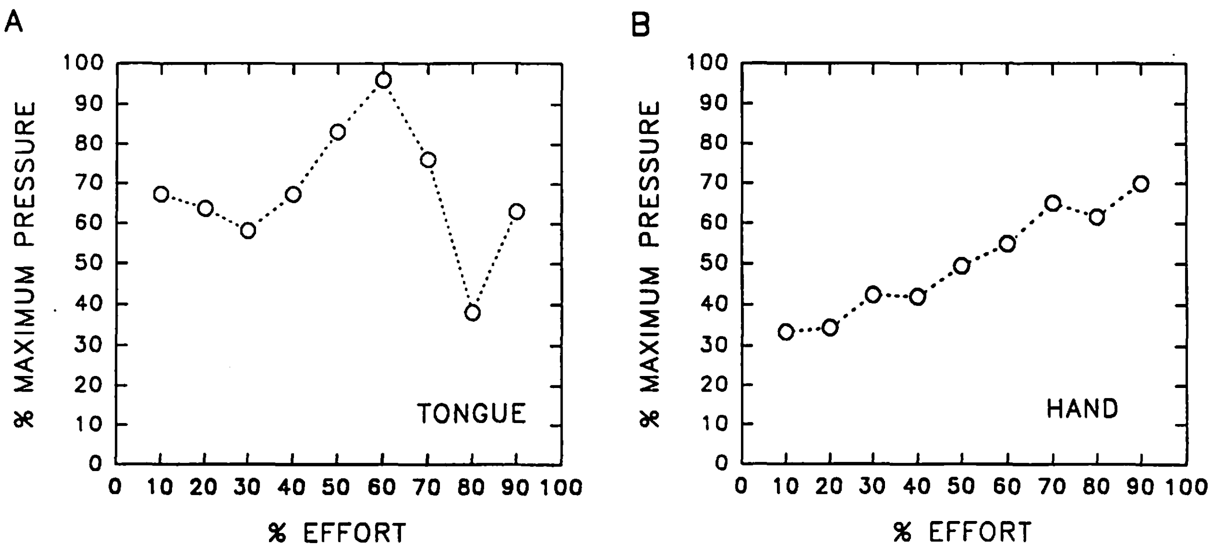


Figure 3. Effort results for Mr. N. for the tongue (A -left) and hand (B - right). Maximum pressure: tongue = 39 kPa, hand = 239 kPa.

Tongue strength was determined to be 39 kPa and hand strength was 239 kPa. Both of these values are lower than those from the initial evaluation. Endurance at 50% maximum pressure was 18 s for the tongue and 39 s for the hand. Although these endurance times are somewhat better than those obtained previously (possibly because the maximum pressures were lower), they remain abnormally brief.

The pressure-effort relationship for the hand (Figure 3 b) was relatively linear with pressure increasing with effort. The function is similar to that obtained from healthy young subjects, except that the pressures exerted at the effort extremes were restricted to the middle of the pressure range

(33% to 70%) rather than showing the usual sensitivity to pressure differences. The findings for the tongue are quite unusual (Figure 3a). There does not appear to be a systematic relationship between effort and pressure. Most attempts at various effort levels resulted in pressures of approximately 60% to 80% of the tongue's maximum pressure. The exceptions to this general finding are bizarre -- Mr. N. exerted 96% of his maximum pressure when asked to give 60% effort, and 38% of maximum pressure for 80% effort.

The results from Mr. N.'s follow-up evaluation confirm that tongue strength and endurance were abnormally reduced, and suggest that he has little awareness of actual pressures exerted by his tongue. It is tempting to speculate that Mr. N.'s abnormally slow speech rate may be related to these abnormalities. This notion is supported by the decrease in both speech rate and tongue strength from the first to the second session.

Summary

The relations between strength, endurance, and perceptions of effort are unclear. Our ultimate goal is to assess these features of motor control in the tongue and to determine their relation to speech characteristics. Previous evaluations of tongue strength and endurance in a variety of normal and disordered populations have provided interesting clinical and basic information. A technique for the study of effort perceptions was added recently to the standard evaluation procedures. This measure may reveal clues for the explanation of abnormally low endurance levels and symptoms of fatigue during speech.

In this report, we presented case studies of 3 people with mild parkinsonism and symptoms of generalized fatigue. The results of this preliminary report provide intriguing patterns of findings. Two profiles are particularly interesting. First, one subject, Mrs. S., had normal strength and endurance in the tongue and hand, but her perception of effort was abnormal. Besides the marked variability in her responses to the effort task, the pressures produced were relatively low even at high levels of effort. This finding is consistent with Mrs. S's general perception of fatigue. Second, the 2 subjects with low strength and low or abnormally variable endurance of the tongue (Mrs. H. and Mr. N.) exhibited abnormally variable responses to the effort task and produced pressures near their maximum strength across the effort range. Typically, individuals with weakened muscles have an increased sense of effort (Gandevia, 1982; McCloskey, 1981), which may explain this pattern of results. Because of low tongue strength, they may perceive a need to produce near maximal pressures to accomplish tasks of any effort level.

The tongue function evaluation involves non-speech tasks, and their relation to speech proficiency is unknown. Based on these 3 cases and a previous investigation, we can speculate that tongue endurance may correlate with speech rate. In the present report, Mrs. S. had normal tongue endurance and a fast speech rate, and Mr. N. had low tongue endurance and a slow speech rate. Similarly, neurologically normal speakers with fast speech rates had better than normal tongue endurance (Robin et al., 1992). Following this logic, we would have expected Mr. N.'s tongue endurance to be better when his speech rate was faster, but this did not occur. Obviously, these preliminary observations are insufficient for drawing conclusions about speech rate and tongue endurance. Another puzzling finding was that Mrs. H.'s speech rate and intelligibility were normal, yet all three tongue function tests were abnormal. Perhaps the abnormalities were too mild to affect speech, or Mrs. H. was able to compensate successfully for these impairments. The present results reinforce the need to explore further the relation between speech and non-speech tongue function measures.

We are beginning to study a larger group of people with Parkinson disease, especially in more advanced stages of the disease, using a similar protocol for tongue function testing. We will examine tongue function in terms of the relations between strength, endurance, and effort, and correlations between these features and specific speech characteristics. In the end, we hope to better understand fatigue and how to cope with it in the management of people with neuromotor speech disorders.

Acknowledgments

This research was supported by Grant No. R03 DC01182-01 and Grant No. P60 DC00976-02 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

Breakthrough, Inc. 131 Technology Innovation Center, Oakdale, Iowa 52319.

Canter, G. 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.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). Motor speech disorders. Philadelphia: W. B. Saunders Co.

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.

Gandevia, S. C. (1982). The perception of motor commands or effort during muscular paralysis. Brain, 105, 151-159.

Goodglass, H., & Kaplan, E. (1981). The assessment of aphasia and related disorders (2nd ed.). Philadelphia: Lea & Febiger.

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

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., Robin, D. A., Somodi, L. B., Solomon, N. P., & Luschei, E. S. (1992, November). Tongue strength and endurance in the elderly [abstract]. Asha, 34, 175.

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.

Mayeux, R., Stern, Y., Williams, J. B. W., Cote, J., Frantz, A., & Dyrenfurth, I. (1986). Clinical and biochemical features of depression in Parkinson's disease. American Journal of Psychiatry, 143, 756-759.

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.

McDowell, F. H. (1971). The diagnosis of parkinsonism or Parkinson syndrome. In F. H. McDowell & C. H. Markham, (Eds.), Recent advances in Parkinson's disease (pp. 163-174). Philadelphia: F. A. Davis.

Neilson, P. D., & O'Dwyer, N. J. (1984). Reproducibility and variability of speech muscle activity in athetoid dysarthria of cerebral palsy. Journal of Speech and Hearing Research, 27, 502-517.

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.

Solomon, N. P., & Hixon, T. J. (in press). Speech breathing in Parkinson disease. Journal of Speech and Hearing Research.

Somodi, L. B., Robin, D. A., & Luschei, E. S. (submitted). "Sense of effort" within the oral motor system.

NCVS Status and Progress Report - 4
June 1993, 291-301

Mechanical Stress in Phonation

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

Abstract

Mechanical stress is always encountered in phonation. This includes tensile stress, shear stress, impact stress during collision, maximum active contractile stress in laryngeal muscles, inertial stress, and aerodynamic stress (pressure). Order of magnitude calculations reveal that tensile stress can reach the greatest value (near 1.0 MPa), contractile stress is next in size (near 100 kPa), and aerodynamic stress is relatively small (1-10 kPa). Inertial stress and impact stress are greater than aerodynamic stress, but less than contractile stress. Excessive collision and acceleration may be responsible for the greatest tissue damage, even though they do not account for the greatest stresses. This is because they act perpendicular to the direction of tissue load-bearing fibers and are applied directly to mucosal tissue.

Introduction

This paper was motivated by the life-long investigations of voice disorders by Moore (1971). Moore convinced this author, and many others in the field, that voice disorders and voice physiology are one and the same topic. One cannot understand normal vocal fold vibration with a disregard for abnormal vibration patterns. This is becoming increasingly more evident in light of modern views of nonlinear dynamics. Chaotic behavior is part of a normal self-oscillating vocal fold system, and "normal" vibration patterns can be seen when the vocal folds are visually impaired.

Nowhere is this gray boundary between normal function and dysfunction more evident than in the assessment of mechanical stress in phonation. It is generally assumed that excessive mechanical stress can lead to organic disorders. But how much is excessive? Although repeated collision of the vocal folds is likely to be the primary cause of vocal nodules, we cannot produce much sound without colliding the vocal folds. So what is a tolerable amount? We also believe that persistent "pressing" together of the arytenoid cartilages is a cause of contact ulcers, and rubbing vocal fold tissue with foreign objects (such as an endotracheal tube) is thought to cause granulomas. But, again, we have no criteria for how much mechanical stress is abusive.

Benign lesions are apparently a reaction to, and ultimately a fortification against, mechanical insult to vocal fold tissues. Little is known, however, about the kind of stress that routinely occurs during vocal fold vibration, and how that stress is distributed within the tissues. If the stress fields were known for various phonation types, perhaps some strategies for healthy, non-abusive voice production would become clearer. Specifically, it would be desirable to weigh phonation type and increases in vocal loudness and against increases in tissue stress to obtain a cost/benefit ratio for certain vocal productions.

The present study is an extension of the summary article on vocal strain written by Sonninen, Damsté, Jol & Fokkens (1972), who summarized the types of mechanical loads and deformations found in the vocal folds. Twenty years later, we can shed only a little additional light on these critical questions: (1) What types of stresses occur in vocal fold vibration, (2) how large are they under normal and maximal effort phonation and (3) is there a likelihood that tissue damage will occur from a specific type of stress?

Types of Mechanical Stress and Relative Magnitudes

We begin by identifying the various types of mechanical stress encountered in vocal fold vibration, with an attempt to estimate their magnitudes. It should be understood that these estimates are order-of-magnitude. In many cases, experiments are yet to be done to verify the numbers.

Tensile Stress

By far the greatest stress applied to vocal fold tissues is a tensile stress (Figure 1). This is applied primarily to the longitudinal (anterior-posterior) fibers of the vocal ligament by the cricothyroid muscle. An estimate of the maximum tensile stress is obtained by assuming that collagenous and elastic fibers vibrate in a string-like fashion. Then

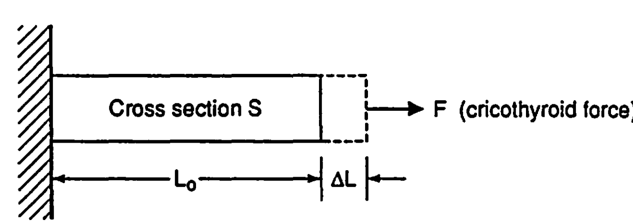
$$\sigma = 4L^2F_o^2\rho \quad , \tag{1}$$

where σ is the tensile stress, L is the length of the membranous vocal fold, F_o is the fundamental frequency, and ρ is the tissue density (1040 kg/m^3). For a female singing a high C (F_o = 1046 Hz) and a membranous length of 0.01 m, the stress is

$$\sigma = 4(10^{-4})(1046)^2(1040) \sim 500\ kPa \tag{2}$$

For higher notes in the coloratura repertoire, frequencies up to 1500 Hz are possible, for which the tensile stress would exceed 1.0 MPa. If the ligament cross-section is about 1 mm x 1 mm, then the tension is on the order of 1 N, or about 1/2 lb.

For typical speech fundamental frequencies, especially for males, the tensile stress is much lower. At 100 Hz, for example, σ is two orders of magnitude lower than at 1000 Hz, giving a stress of about 5 kPa.



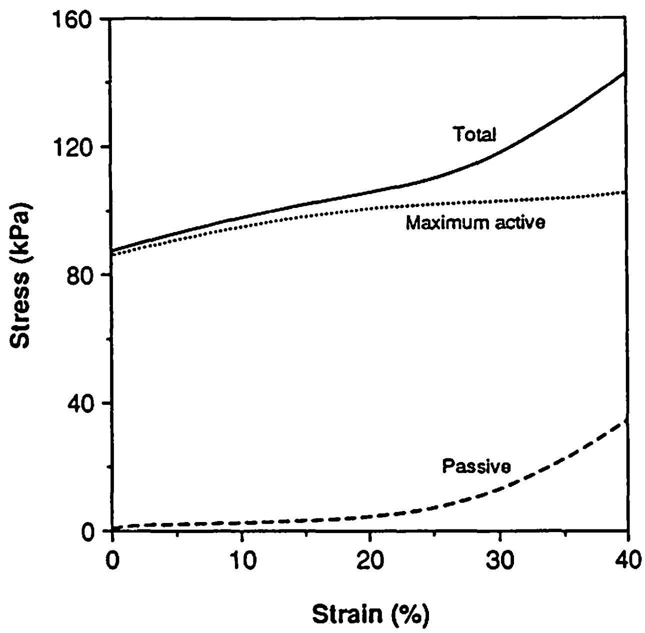
Tensile stress = σ = F/S

Tensile strain = ε = $\Delta L/L_0$

Figure 1. Illustration of tensile stress and strain.

Maximum Active Contractile Stress in Muscles

The maximum active contractile stress in the thyroarytenoid (TA) and cricothyroid (CT) muscles of canines has been measured by Alipour-Haghighi, Titze, & Perlman (1989; 1991). Results are shown in Figure 2 for the TA and Figure 3 for the CT. Note that these stresses vary with vocal fold length (strain), but have a range of values of about 30-115 kPa. The TA muscle reaches a maximum active stress of 100 kPa at 20% elongation, whereas the *pars recta* of the CT reaches the same value at about 30% elongation.



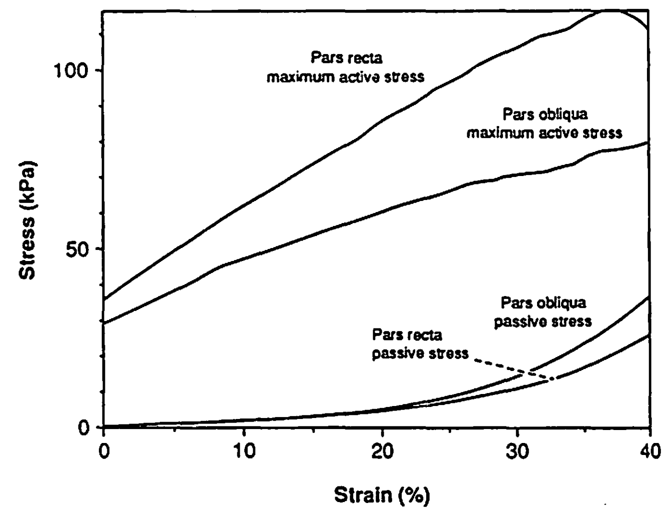
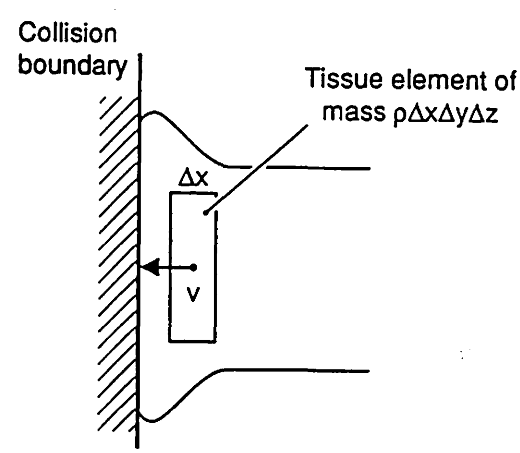


Figure 2 (left). Maximum active contractile stress (dotted line) of the TA muscle of canine larynges. The total stress (solid line) is assumed to be the summation of the two stresses (after Alipour-Haghighi et al. 1989). Figure 3 (right). Maximum active contractile stress of two components of the CT muscle of canine larynges. The passive stress is also shown (after Alipour-Haghighi et al. 1991).

One might ask how the CT can produce a stress of 1.0 MPa in the vocal ligament when its maximum contractile stress is only about 100 kPa. The answer lies in a transformation of cross-sectional area. The force applied to the ligament equals the force produced by the CT muscle, but since the ligament has a smaller cross-sectional area, the stress is magnified by a factor of 10 or so.

Collision Stress Between Vocal Folds

The collision stress between the vocal folds can be estimated from basic physical principles (Figure 4). Assuming the mass of a tissue element at the medial surface to be



$$\text{Impact stress} = (\rho \Delta x) \frac{\Delta v}{\Delta t}$$

Figure 4. Illustration of impact stress.

$$m = \rho \Delta x \Delta y \Delta z \quad , \tag{3}$$

where ρ is the tissue density given above and ΔxΔyΔz is a small volume then, from Newton's second law, the average collision force over an impact interval Δt is

$$F = m \Delta v / \Delta t \quad , \tag{4}$$

where Δv is the change in velocity during impact. Jiang & Titze (in review) estimated the impact interval to be on the order of

$$\Delta t = \frac{T_o}{10} \tag{5}$$

where T_o is the fundament period.

The velocity change in equation (4) can be estimated by assuming sinusoidal motion with amplitude A and radian frequency $\omega = 2\pi F_o$. The maximum velocity, which occurs near impact, is then

$$v = \omega A = 2\pi F_o A \quad . \tag{6}$$

This velocity is reduced to zero during the collision interval, such that

$$\Delta v = v - 0 = 2\pi F_o A \tag{7}$$

Substituting (3), (5), and (7) into (4) yields

$$F = 20\pi A F_o^2 \rho \Delta x \Delta y \Delta z \quad . \tag{8}$$

If ΔyΔz is taken to be the impact surface and Δx the depth of the vibrating tissue, then the collision stress is

$$\sigma = \frac{F}{\Delta y \Delta z} = 20\pi A F_o^2 \rho \Delta x \quad . \tag{9}$$

For an amplitude of vibration of 10^{-3}m, a depth of vibration of 10^{-3}m, and an F_o of 200 Hz, this stress is 2.6 kPa. Peak impact stresses on the order of 0.5-5.0 kPa were measured by Jiang & Titze (in review) and are shown in Figure 5 as a function of subglottal pressure. Similar increases in contact stress were obtained by Reed, Doherty and Shipp (1992) with a piezoelectric transducer placed between the vocal folds of a human subject.

Values of F_o in Figure 5 are below 200 Hz. For frequencies approaching 1000 Hz, one would expect much higher impact stresses if the amplitude and depth of vibration were to remain constant. Typically, however, both amplitude and depth of vibration decrease with F_o, making the exact stress uncertain. One expects that the impact stress could increase an order of magnitude, to about 50 kPa, for high F_o and high subglottal pressure.

Inertial Stress

The inertial stress is similar to the impact stress, except that acceleration and deceleration occur without collision. Assuming again that the tissue moves sinusoidally with amplitude A, i.e.,

$$x = A \sin \omega t \quad , \tag{10}$$

then the maximum acceleration is $\omega^2 A$. The maximum inertial stress is mass ($\rho \Delta x \Delta y \Delta z$) times acceleration ($4\pi^2 F_o^2 A$) divided by the surface area ($\Delta y \Delta z$), which simplifies to

$$\sigma = 4\pi^2 A F_o^2 \rho \Delta x \quad . \tag{11}$$

This stress is about half of the impact stress, on the order of 1-2 kPa for normal phonation.

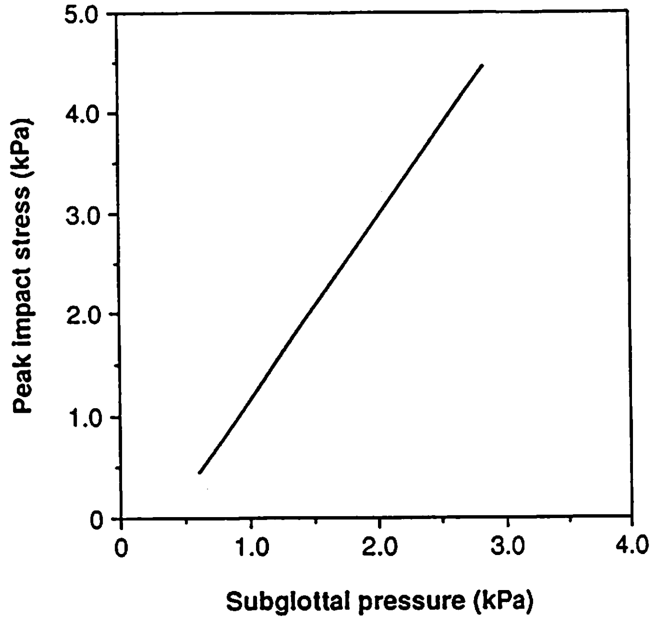


Figure 5. Peak collision stress versus subglottal pressure in a canine hemilarynx (after Jiang & Titze, in review).

Aerodynamic Stress

The mean aerodynamic driving stress is the mean intraglottal pressure, which has been estimated as

$$P_g = P_i + \left(1 - \frac{a_2}{a_1}\right)(P_s - P_i) \tag{12}$$

in the open phase (Titze, 1988). Here P_i is the input pressure to the vocal tract (supraglottal pressure), a_1 is the glottal entry area, a_2 is the glottal exit area, and P_s is the subglottal pressure. The greatest pressure occurs for a highly convergent glottis, where $a_2/a_1 \approx 0$, in which case $P_g \approx P_s$.

In high effort phonation, P_s can be as high as 5-6 kPa, and occasionally reaches 10 kPa (Schutte, 1980). This can be taken as the upper limit on aerodynamic stress.

Arytenoid Contact Stress

Maximum contact stress between the arytenoid cartilages has been estimated by Rethi (1897) and Kakeshita (1927) to be on the order of 50-100 kPa in animals who had the adductor muscles fully contracted. This range would tend to agree with the maximum active stresses measured in the CT and TA muscles of canines. Adduction stress for normal phonation is much lower than that, however, even for so-called "pressed voice". Values reported by Scherer, Cooper, Alipour & Titze (1985) were on the order of 1-5 kPa.

Shear Stress at Anterior and Posterior Macula Flava

The shear stress in the ligament increases with the amplitude of vibration. If the ligament is pinned at the endpoints (anterior and posterior macula flava), then the shear stress τ can be written as

$$\tau = \mu \sin\theta \quad , \tag{13}$$

where μ is the shear modulus and θ is the shear angle at the endpoints (Figure 6). The shear angle can be expressed in terms of the vibrational amplitude and the membranous length of the vocal folds. Thus, if

$$x = A \sin \pi y / L \tag{14}$$

is the displacement of a lowest string mode in the anterior-posterior (y) direction, A is the amplitude (at the center) and L is the membranous length, then

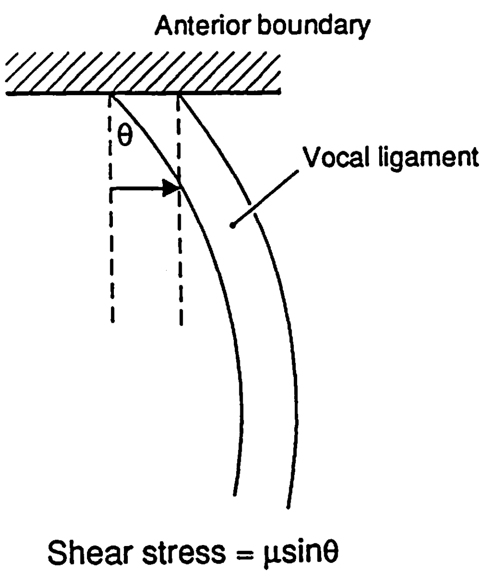
$$\frac{dx}{dy}\Big|_{y = O,L} = \frac{\pi A}{L} = \tan\theta \approx \sin\theta \quad . \tag{15}$$

For an amplitude of 0.2 mm and a length of 10 mm, sin θ = 0.6.

The shear modulus μ of the collagen fibers in the macula flavae is not known to this author, which makes it impossible to estimate the absolute value of the shear stress. All one can say is that it increases with amplitude of vibration and decreases with vocal fold length. The A/L ratio is a key variable for dynamic shear stress.

Summary of Relative Magnitudes of Stresses

A summary chart of the relative magnitudes of the various mechanical stresses is given in Figure 7. Estimates are made for conditions of normal phonation (clear bars) and maximum stress (cross-hatched bars). Note the large maximum tensile stresses relative to all other stresses. Note also the relatively small aerodynamic stresses. Active contractile, arytenoid contact, impact, and inertial stresses are intermediate in size.



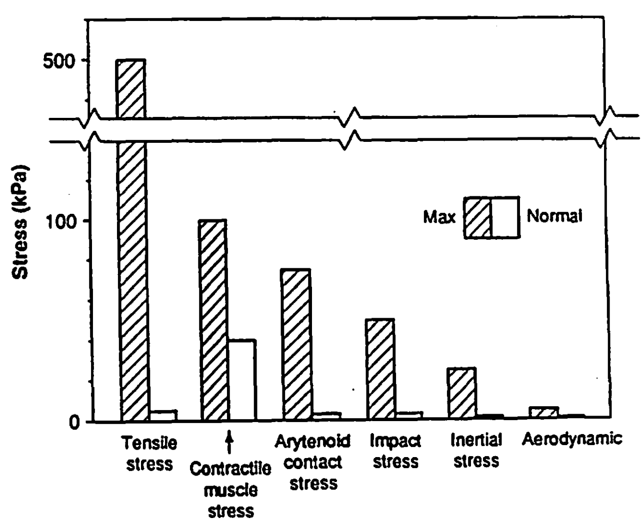


Figure 6 (left). Illustration of shear stress in vocal ligament. Figure 7 (right). Summary of the relative magnitudes of various mechanical stresses in vocal fold tissues.

Damage Criteria

Damage to vocal fold tissue may occur from any of the mechanical stresses described. Based on the relative size of the tensile stress in relation to other stresses, one would reason that the vocal ligament would be at the highest risk for damage, particularly at high pitches. Excessive tension would appear to rupture the tissue fibers. If the vocal ligament has comparable strength to other ligaments in the body, however, it is well protected against this type of rupture. The following discussion will clarify this point.

Rupture Due to Tensile Stress

As an analogy with other body tissues, consider the anterior cruciate ligament of the knee, which has been studied extensively by Noyes & Grood (1976). In size, this ligament is larger than the vocal ligament. The authors report an average length of 27 mm and an average cross-sectional area of 44 mm^2. This is about twice the length of the vocal ligament and about ten times the cross-section. Nevertheless, since stress and strain are essentially independent of sample size, the comparison is useful.

Figure 8 shows the mechanical response of the anterior cruciate ligament under tensile stress. Strain was applied at a rate of 100% per second. The attempt was to simulate conditions in sports, where extension, flexion, and torsion occur quite rapidly. Tissue age was a parameter in the study. One group of subjects was in the 16-26 year age bracket, while another was in the 48-86 year bracket. Note that the older group displayed a lower maximum stress (13.3 MPa as compared to 37.8 MPa for the younger group). Furthermore, this maximum stress occurred at a lower strain (30% instead of 44%). Thus, not only could the younger ligament support more tensile stress, but it could support this stress at a greater elongation.

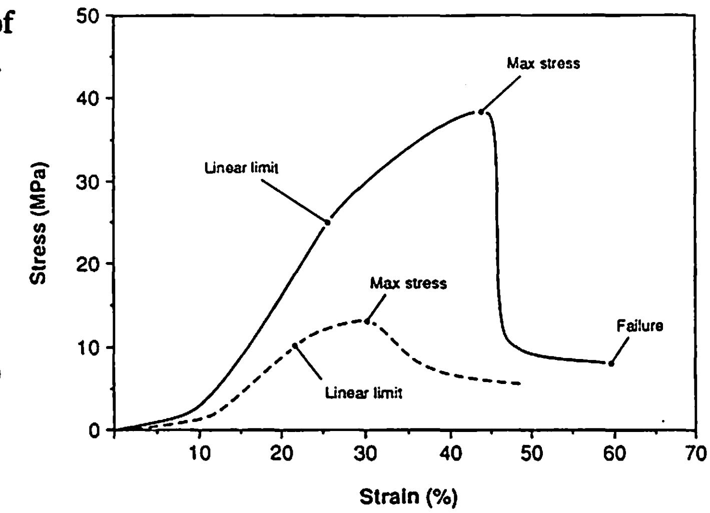
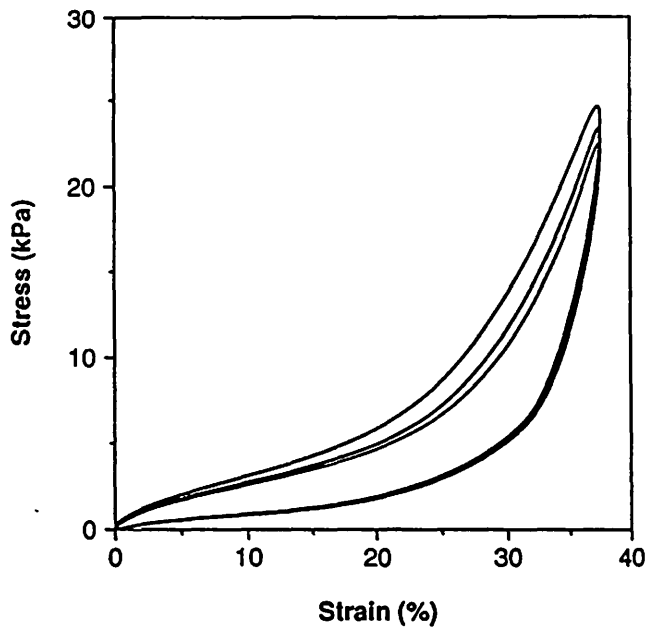


Figure 8. Stress-strain curve for anterior cruciate ligament in humans (after Noyes & Grood, 1976). Solid line is for younger specimen (16-26 years) and dotted line for older specimen (48-86 years).

Failure of the ligament occurred differently in the two age groups. For the younger group, collagen fibers were ruptured throughout the length of the ligament. This caused the stress to decrease rapidly as shown. For the older group, failure was primarily due to bone avulsion. This took place more gradually and is probably the cause of earlier failure with increased strain.

It is important to contemplate the overall stress magnitude that this knee ligament can support. In the "linear" range, stresses can be 10-20 MPa. This is 40-80 times the stress we estimated for soprano singing a high C. If the vocal fold ligament is made up of the same type and density of collagen fibers as this knee ligament, one would conclude that there is little chance for failure of the vocal ligament by excessive CT contraction.

The vocal ligament probably serves to protect other tissues in the vocal fold from rupture. If we can assume that 30% is a strain limit on the vocal folds, which tends to agree with the estimates of van den Berg (1958), then the mucosa and the TA muscle will not be elongated to their yield points. Figures 9 and 10 show stress-strain curves for the canine TA muscle and vocal fold cover, respectively. These curves were obtained by cyclic stretch-release methods at a rate of 1 Hz (similar to the rate of the knee ligament elongation). Note that 30-40% elongation was possible without tissue damage. Thus, the presence of a vocal ligament would tend to keep strains in the "safe" range for muscle and mucosal tissue.



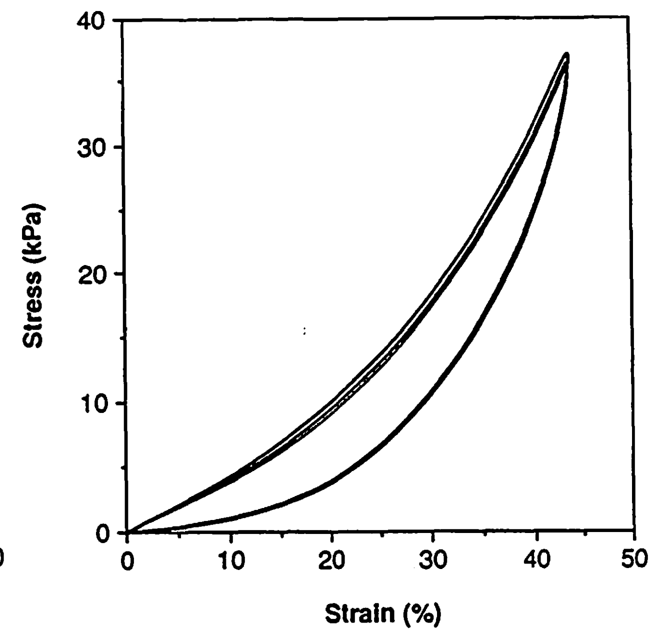


Figure 9 (left). Stress-strain curve for the thyroarytenoid muscle of canines during cyclic stretch and release.
Figure 10 (right). Stress-strain curve for the vocal fold cover of canines during cyclic stretch and release.

Excessive Vibration

In his Handbook of Human Vibration, Griffin (1990) lists five disorders associated with hand-transmitted vibration:

1) vascular disorders
2) bone and joint disorders
3) peripheral neurologic disorders
4) muscle disorders
5) other (e.g., central nervous system)

These disorders are usually associated with tool use. They may not apply at all to vocal fold vibration, but it is worth while to make a few comparisons. Parameters for consideration are vibration magnitude, frequency, and duration.

The primary symptom of a vascular disorder is "white finger", which is basically the sign of reduced circulation. Circulation is impeded by rapid acceleration and deceleration of the tissue. Estimates of maximum acceleration of vocal fold tissues within the vibratory cycle were implicit in equation (1),

$$a = \omega^2 A = 4\pi^2 F_o^2 A \quad . \tag{16}$$

For a fundamental frequency of 200 Hz and a vibrational amplitude of 0.001 m, there is a peak acceleration of 1600 m/s^2, with a root mean squared (r.m.s) value of 1100 m/s^2. If such a vibration were to occur continuously for 30 minutes, it would fall well into the *unacceptable* region of exposure according to ISO standards (Griffin, 1990, p. 647). As shown in Figure 11, accelerations of even one tenth this value would be unacceptable (see data point in relation to the line of safety). Phonation is never continuous, of course, for 30 minutes. This makes the dosage criterion difficult to assess, but it cannot be ruled out that excessive vibration could lead to some vascular disorders in the larynx.

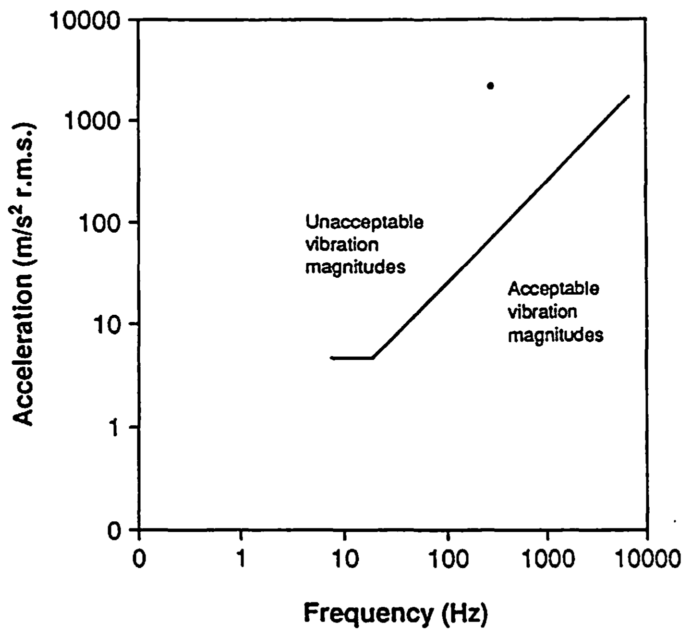


Figure 11. ISO standard for hand-transmitted vibration (after Griffin, 1990). Data point suggests that vocal fold vibration could exceed the maximum recommended dose.

Impact Stress

Assessment of safe limits of impact stress is also difficult. We do not have well-controlled experiments that show the tissue damage as a direct result of collision. The studies by Gray, Titze & Lusk (1987) and Gray & Titze (1988) are beginnings. Continuous phonation was maintained (by an artificial air supply) for several hours at high intensity in anesthetized canines. Electron-microscopic examination showed destruction of surface microvillae, of squamous epithelial cells, and of the basement membrane zone. Future investigations should quantify this type of destruction in terms of measured impact stress. It would also be extremely valuable to know the time course and degree of the healing process.

Conclusion

The largest mechanical stresses in vocal fold vibration are the tensile stresses required for pitch increase. Estimates are that they may reach values on the order of 1 MPa. This would be excessive for epithelial or muscular tissue, but seems to be normal for ligamental tissue. We assume, therefore, that a well-developed and healthy vocal ligament provides a "safety-valve" for other, perhaps more injury prone, tissues in the vocal folds. The ligament limits elongation and assumes most of the tensile stress at high pitches.

The ligament cannot protect as well against vocal fold collision, however, where the stress is transverse to the fibers. Here the softer tissues of the lamina propria are exposed and do absorb most of the impact stress. Some evidence of destruction has been reported, and the fact that vocal fold nodules occur bilaterally at the point of maximum impact stress (Jiang & Titze, in review), attest to this.

The possibility that safe limits of tissue acceleration are exceeded in prolonged phonation needs further exploration. Exposure durations are shorter than what is typically reported for hand-transmitted vibration with power tools, but acceleration magnitudes are comparably larger. This leaves the integrated dosage uncertain.

Finally, shear stresses and aerodynamic stresses were discussed briefly, but no damage estimates could be made in this preliminary study. Aerodynamic stresses are very small in comparison to tensile stresses and maximum contractile stresses of the laryngeal muscles, which leads one to speculate that aerodynamic stresses of themselves do not pose a damage risk.

Acknowledgement

This study was supported by grant No. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders. The author thanks Julie Lemke for manuscript preparation and Mark Peters for graphic support.

References

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

Alipour-Haghighi, F., Titze, I.R., & Perlman, A. (1991). Tetanic response of the cricothyroid muscle. Annals of Otology, Rhinology & Laryngology, 100(8), 626-631.

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

Gray, S., & Titze, I.R. (1988). Histologic investigation of hyper-phonated canine vocal cords. Annals of Otol. Rhinol. and Laryngol., 97, 381-388.

Griffin, M.J. (1990). Handbook of Human Vibration. New York: Academic Press.

Jiang, J., & Titze, I.R. (in review). Measurement of vocal fold intraglottal pressure and impact stress.

Kakeshita, T. (1927). Über eine neue methode zur messung der beim stimubrandverschluss winkende krafte, I. Mitt. Pflügers Arek. für die Gesamunte Physiologie 215, 19

Moore, P. (1971). Organic Voice Disorders. Englewood Cliffs, NY: Prentice-Hall.

Noyes, F.R. & Grood, E.S. (1976). The strength of the anterior cruciate ligament in humans and rhesus monkeys. J. Bone Joint Surgery, 58-A, No. 8, 1074-1082.

Reed, C.G., Doherty, E.T., & Shipp, T. (1992). Direct measurement of vocal fold medial forces. ASHA, 34:131(A).

Rethi, L. (1897). Die stimmbandspannung, experimentell geprüft. Sitzgsber. Akadem. Wissenschapt Wien, Math. Naturwissenshaft Kl. III 106, 244.

Scherer, R.C., Cooper, D.S., Alipour-Haghighi, F., & Titze, I.R. (1985). contact pressure between the vocal processes of an excised bovine larynx. In J. Titze & R. Scherer (Eds.), Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control (pp. 292-303). Denver Center for the Performing Arts.

Schutte, H. (1980). The Efficiency of Voice Production. Groningen: Kemper.

Sonninen, A., Damsté, P.H., Jol, J., & Fokkens, J. (1972). On vocal strain. Folia phoniat., 24, 321-336.

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

van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. J. Speech Hearing Res. 1, 227-244.

NCVS Status and Progress Report - 4
June 1993, 303-310

Current Topics in Voice Production Mechanisms

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

Abstract

A brief overview of current research in voice production mechanisms is given. The self-oscillatory behavior of the vocal folds is clarified, along with the identification of preferred tissue modes. Fundamental frequency control is reviewed as a coordinated activity between cricothyroid and thyroarytenoid muscles and lung pressure. Lung pressure is shown to be the primary regulator of vocal intensity. For eventual application to the study of voice disorders, a method of measurement of contact stress on a hemilarynx preparation is described.

Introduction

Voice production continues to be an active field of research. It has traditionally attracted a broad mix of basic and clinical scientists. Voice is fascinating to investigators because it is very close and personal on the one hand, yet elusive and mysterious on the other. How can two pieces of flesh collide with each other to produce such a large inventory of sounds? Basic scientists (engineers, physicists, physiologists) are intrigued by the complex nonlinear and interactive mechanisms that are found in vocal fold vibration. Clinicians wonder why the vocal instrument works so well for some people, but fails so miserably for others.

In this paper, some currently active topics of research in voice production mechanisms will be highlighted. This is only a sampling, however, of many interesting research areas. Three specific topics have been selected: the mechanism of vocal fold oscillation, control of pitch and intensity, and vocal fold contact stress. It is hoped that the last of these areas will lead to future discoveries in clinical voice.

Vocal Fold Oscillation

The mechanism of vocal fold oscillation has been described in many textbooks, but only recently has the energy transfer between fluid flow and vocal fold tissue vibration been elucidated. In particular, it has been found that the vocal folds vibrate in specific modes, like violin strings or drum heads. Not all of the modes, however, are easily excited by the airstream in the glottis. This sets up a situation of *preferred modes* of vibration for self-oscillation of the vocal folds. These preferred modes become obvious when the mechanism of energy transfer is understood.

Mechanism of Energy Transfer

Figure 1 shows schematic representations of the vocal folds in coronal cross-section. Part (a) shows the configuration for outward (lateral) movement of the vocal folds and part (b) shows the configuration for inward (medial) movement. On the left side of each sketch is a 3-mass representation of the tissue, with m being the mass of the body and m_1 and m_2 being masses of the cover. On the right side, the body-cover structure is highlighted, with the body being stippled and the cover being unstippled. Airflow is indicated with dotted lines. The most important discovery about vocal fold self-oscillation is that in a *convergent* glottis, illustrated in part (a), the pressure P against the medial surface of the vocal folds is considerably larger than for a *divergent* glottis (part b). This discovery comes from aerodynamic theories of nozzles and diffusers and has led to quantitative treatments of vocal fold oscillation[1,2,3].

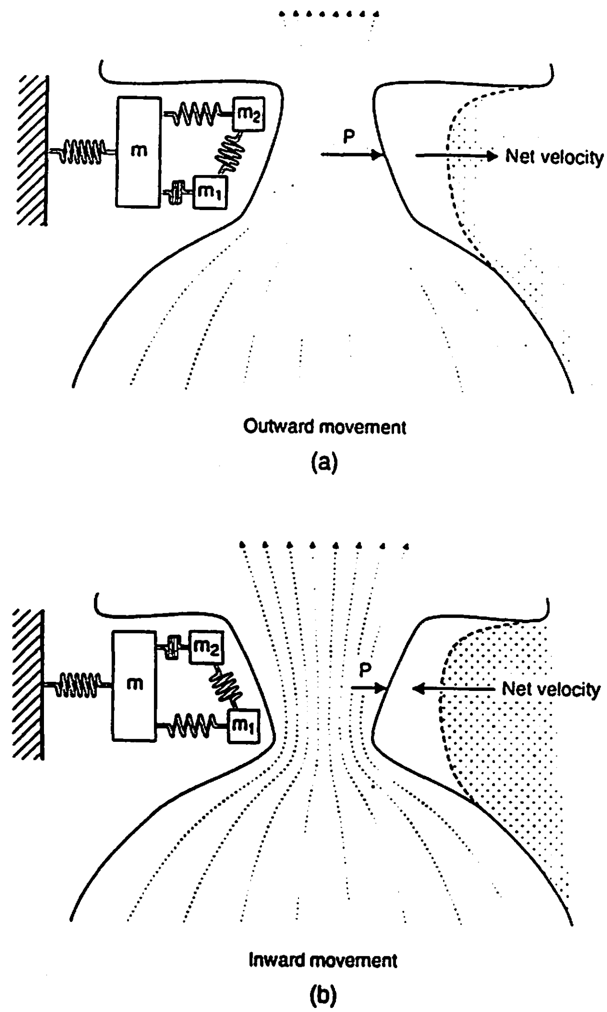


Figure 1. Airflow and pressure for two glottal shapes.

During outward movement, then, when the net velocity of the tissue is to the right (see arrows), a large pressure reinforces this motion by being in phase with the velocity. On medial movement, however, the aerodynamic pressure is reduced and the tissue can return to the midline with minimal interference from the flow. The key in the entire series of events is that the net driving pressure is basically in phase with the tissue velocity. Energy is thereby transferred from the airflow to the vocal folds. This energy overcomes the natural damping of energy by friction in the tissue and thereby sustains the oscillation.

Excitable Tissue Modes

To obtain an alternately convergent and divergent glottal shape for self-oscillation, the tissue must have a natural mode that allows this alternation of geometry. Titze and Strong[4] have studied the basic tissue modes in the vocal folds and have given them labels. Figure 2 shows a summary of what the mode structure is. On the left side of the figure, the medial edges of the vocal folds are represented as strings from the superior view. Dashed lines indicate the prephonatory glottal configuration and solid lines indicate the maximum excursions of the edge. Proceeding from top to bottom, we see first a mode in which there is a half wavelength between anterior to posterior boundary points. In the second figure from the top, we see two half wavelengths and in the third figure we see three half wavelengths. In standard mode nomenclature, an index 1, 2, and 3 can be assigned to represent the number of half wavelengths between the two endpoints of the vocal folds.

In addition, there can be a variation of tissue movement in the vertical direction. This can be represented by another index. In the top three illustrations on the right side of Figure 2, the tissue is moving uniformally from top to bottom, meaning there is no variation in the z direction. For such modes, the index 0 is assigned. Thus, an "m-0" mode has *m* half wavelengths horizontally and no wavelengths vertically. In the fourth illustration shown at the bottom of the figure, however, we see that the top portion of the vocal folds moves in opposite direction to the bottom portion. This then represents a case of nonuniform tissue movement in the vertical direction. An approximate half wavelength is represented by this mode pattern and hence, the index 1 is assigned to this z variation. An "m-1" mode has *m* half wavelengths horizontally and 1 half wavelength vertically. As an example, the bottom illustration in Figure 2 is a 1-1 mode. This is the most prevalent mode in vocal fold vibration. It is the mode that is easily excited by an airflow passing over the surface of the tissue, like wind passing over a water surface.

In most normal situations, more than one mode is excited simultaneously. It has been observed that typically a combination of the 1-0 mode and the 1-1 mode represents normal movement of vocal fold tissues. Higher modes, such as the 2-0, 3-0, 2-1 or 3-1, are sometimes observed in pathological voices, especially when asymmetries in the left and right vocal fold occur or when there are lesions on the vocal folds. Generally speaking, normal vocal fold vibration is made up of a few simple modes, whereas pathologic vibration tends to have a greater number of higher modes. These modes create complex airflow patterns that are often perceived as a rough voice.

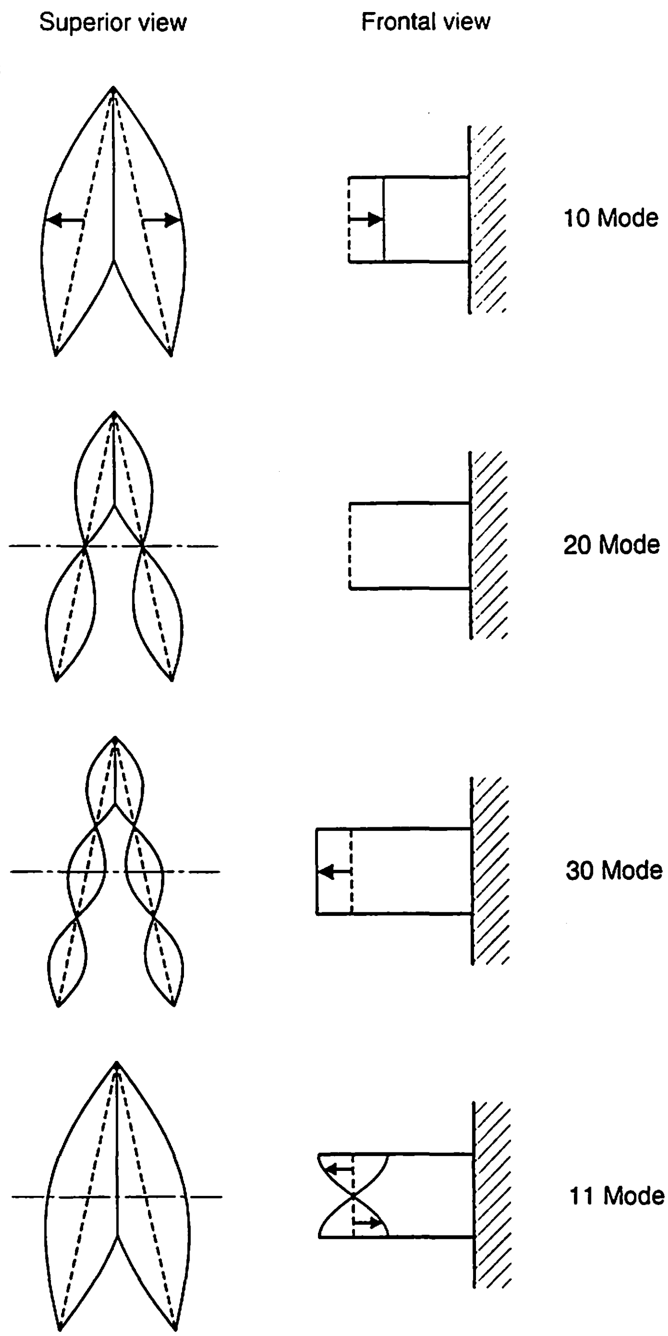


Figure 2. Normal modes of the vocal folds.

Phonation Threshold Pressure

It has been found that self-oscillation of the vocal folds can only occur after a threshold lung pressure has been reached[5,6,7]. This phonation threshold pressure is determined by the mechanical and geometrical properties of the vocal folds. Titze[3] has shown that a convergent prephonatory glottis has a greater threshold than a rectangular or slightly divergent prephonatory glottis. In the same article, it was also shown that tissue viscosity effects the phonation threshold pressure. In other words, tissues that are highly viscous tend to require more lung pressure to be set into oscillation. From a clinical point of view, it is important to maintain the cover of the vocal folds in a state of low viscosity by ample irrigation and lubrication with mucous and internal fluids.

Figure 3 shows measurements of phonation threshold pressure as a function of fundamental frequency. The different curves represent different measurements obtained in the literature. Curve G was obtained by Gramming[8] whereas curve V was from the data of Verdolini-Marston et al[7]. Solid lines are for males and dashed lines for females. The primary observation is that phonation threshold pressure, which is on the order of 0.3 kPa in the low part of the F_o range, rises to more than 0.5 kPa in the upper part of the F_o range.

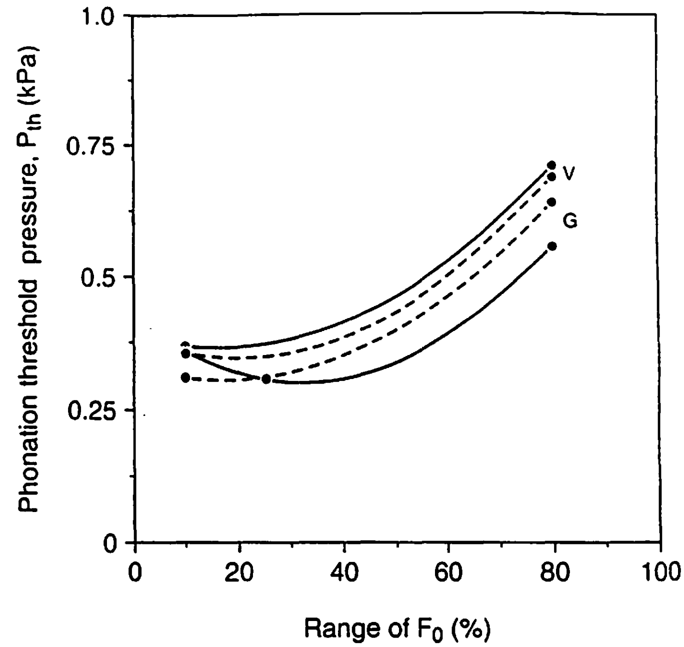


Figure 3. Phonation threshold pressure.

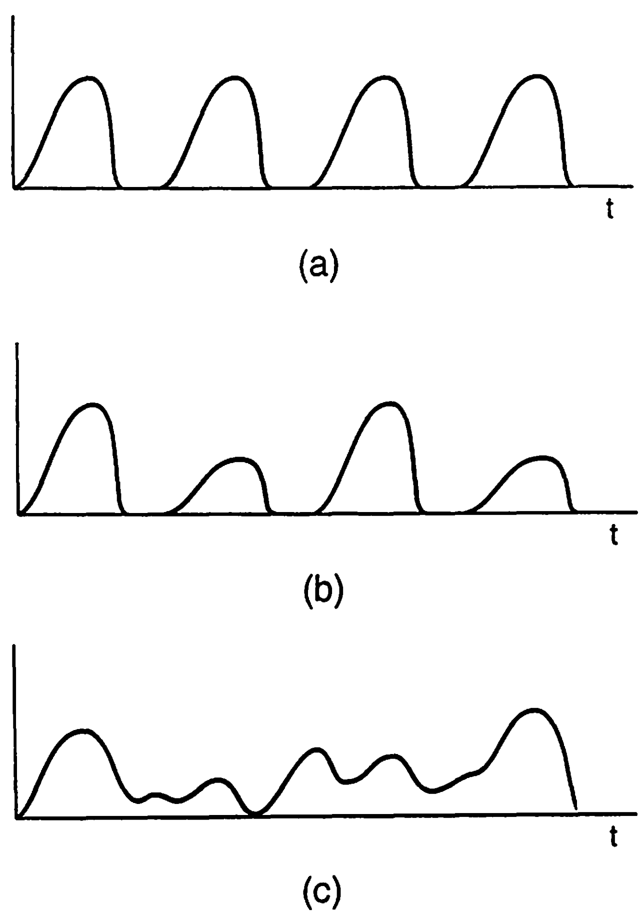


Figure 4. Glottal flow waveforms for a) periodic, b) subharmonic and c) chaotic vibration.

Bifurcation and Chaos

Recent studies by Herzel, Berry, Titze & Saleh[9] and Titze, Baken & Herzel[10] have shown that vocal fold vibration can become chaotic under certain conditions. Primary factors in producing chaos are nonlinearities in vocal fold tissues and asymmetries between the left and right vocal fold. When the modes of the left and right vocal folds are sufficiently different not to be *entrained* by the airflow, a desynchronization takes place between the two coupled oscillators. This desynchronization often begins with the appearance of a subharmonic or a modulation in the flow, but ultimately results in total aperiodicity. Figure 4 shows how a normal flow pattern in part (a) *bifurcates* to a subharmonic structure in part (b) and then to chaotic vibration in part (c). Some bifurcations are typical in nonlinear dynamical systems. As asymmetry and nonlinearity in a system of coupled oscillators increase, the order of events is typically from periodic motion to subharmonic vibrations, followed by bifurcations to more complex structure, and finally to chaotic movement.

Control of Pitch and Intensity

Much has been learned in the last two decades about the control of fundamental frequency (F_o) by the use of electromyography[11,12,13]. Generally speaking, it has been established that the cricothyroid muscle (CT) is a major muscle for control of F_o. The involvement of the thyroarytenoid muscle (TA) has been less clear. Recent studies with the use of biomechanical models of the laryngeal framework, however, have shed considerable light on this topic[14,15].

Muscle Activation Plot

Figure 5 shows a Muscle Activation Plot (MAP) for the intrinsic laryngeal muscles. On the vertical axis is normalized cricothyroid muscle activity a_{ct}, a quantity ranging between 0 and 1. On the horizontal axis, we see thyroarytenoid muscle activity a_{ta}, also normalized to a maximum of 1. The figure shows bands of constant fundamental frequency as indicated by the shaded regions. These bands represent combinations of activity a_{ct}, a_{ta} and lung pressure P_L that can produce the same fundamental frequency. Note that at the left boundary of each band, the lung pressure is higher than at the right boundary. Thus, if a vocalist wishes to perform a crescendo at a constant pitch, the muscle activities in the cricothyroid or thyroarytenoid have to be reduced as the lung pressure increases from the right boundary to the left boundary. This requires very delicate coordination of three groups of muscles.

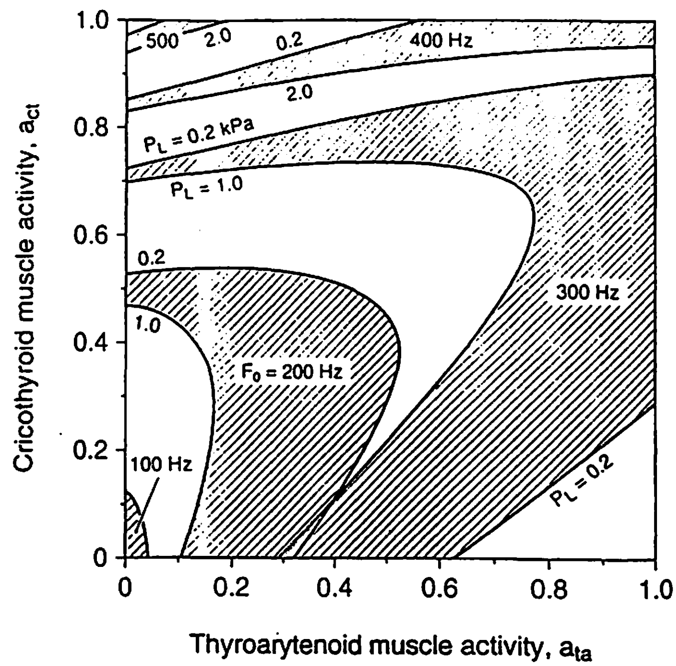


Figure 5. Muscle activation plot for F_o control.

Note also that F_o can be increased more effectively with a_{ta} than a_{ct} in the lower part of the Muscle Activation Plot. Maximum change of F_o occurs from left to right, perpendicular to the bands. In the upper portion of the MAP, however, change of F_o is more effective with a_{ct} than a_{ta}. Thus, when a singer or any vocalist attempts to produce wide ranges of fundamental frequencies, considerable difference in strategy must be employed for F_o change in different parts of the range. This is, of course, what much of vocal training is all about.

Voice Range Profile

Control of vocal intensity is often assessed with a Voice Range Profile[16,17]. This is shown in Figure 6(a), where sound pressure level in dB is plotted vertically and fundamental frequency F_o is plotted horizontally. The open circles represent data from human subjects[8], whereas the smaller dots represent data generated by computer. This diagram basically shows how *range* of intensity varies with different fundamental frequencies. In part (b) of the figure, the corresponding lung pressure is shown for the same range of F_o. Note that at the low F_o, a small range of lung pressure produces a large range of sound pressure level. As higher fundamental frequencies are attempted, a greater range of lung pressure is needed to get a similar range of sound pressure level. The bottom boundary of the stipples represents the phonation threshold pressure previously discussed, whereas the upper boundary represents the maximum lung pressure used in phonation.

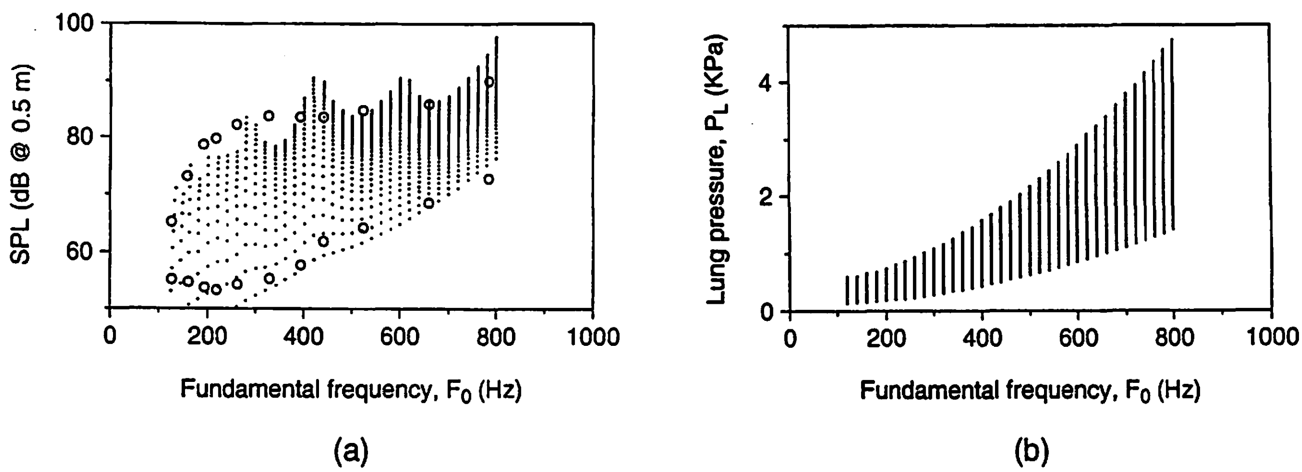


Figure 6. a) Voice range profile and b) corresponding lung pressure range.

Vocal Fold Contact Stress

Research is in progress to quantify the contact stress in vocal fold vibration. This research may lead to a better understanding of the etiology of vocal fold lesions produced by excessive or abusive phonation.

Phonation of a Hemilarynx

To quantify the contact stress between the colliding vocal folds, Jiang and Titze[18] developed a hemilarynx preparation. This is shown in Photo 15; see center-bound plate. An excised larynx is cut in half and mounted such that the vocal fold can collide with a plane of glass rather than with the opposing vocal fold. The plane of glass houses a small pressure transducer that can sense not only the aerodynamic pressure between the glass and the tissue, but also the contact pressure. An added advantage of this hemilarynx procedure is that the motion of the vocal folds can be viewed from the sagittal aspect through the glass. Videostroboscopy has been used to get a clearer picture of the contact pattern between the vocal fold and the opposing glass surface. Photo 16 shows one frame of such a videostroboscopic observation. The upper light band seen in the photograph is the medial surface of the tissue colliding with the glass. With this experimental procedure, it is possible to quantify the relation between movement of the vocal fold, contact stress at particular points on the medial surface, and the contact area as determined by an electroglottograph (EGG).

To show an example of the type of pressure pattern that has been measured, Figure 7 is a composite drawing of the pressure waveform and some sketches of vocal fold movement. Note the strong peak that is observed in the pressure waveform at the point where vocal fold contact begins (at time = 0.0125s). This peak relaxes rapidly after full contact has been achieved and the tissue has deformed to accommodate the collision After the collision, the vocal folds begin to open and a broader peak in pressure is observed (near time = 0.016s). This is now the aerodynamic pressure, which reduces gradually and may become slightly negative prior to vocal fold closure (time = 0.019s). The existence of a negative pressure in the glottis has been called the Bernoulli effect.

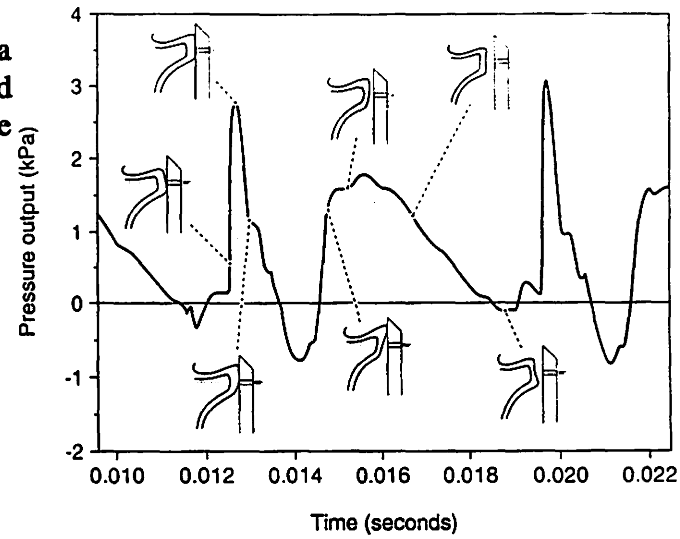


Figure 7. Intraglottal pressure measurements.

Note that the measured pressure patterns in Figure 7 corroborate the claims made earlier with regard to vocal fold self-oscillation. A divergent glottis (at time = 0.019s) has a much lower pressure than a convergent glottis (at t = 0.015s). This was illustrated in Figure 1 with arrows against the surface. The utility of this type of experimentation is that the magnitudes of collision stresses and aerodynamic pressures can be studied simultaneously. This will hopefully allow researchers to get a better quantification of what abusive phonation might be.

Conclusion

A sampling of a few studies on the mechanism of phonation indicate that much has been learned in recent years about the manner in which vocal folds vibrate. Some of the facilitating factors and some of the impeding factors for vocal fold oscillation have been mentioned. It remains to be seen which modes of vibration offer the least amount of damage to vocal fold tissues while producing the greatest amount of vocal power. Studies are presently underway to determine the stress distribution within tissue layers for various modes of vibration.

Acknowledgement

This study was supported by Grant No P60 000976 from the National Institutes on Deafness and Other Communication Disorders. The author appreciates the assistance of Julie Lemke in manuscript preparation and Mark Peters in graphic support.

References

1. Ishizaka K, Matsudaira M. Fluid Mechanical Considerations of Vocal Cord Vibration 1972; Monograph 8, Speech Communication Research Labs, Santa Barbara, CA.

2. Stevens K. Physics of laryngeal behavior and larynx modes. Folia Phonetica 1977; 34(4): 264-279.

3. Titze I. The physics of small-amplitude oscillation of the vocal folds. JASA 1988; 83(4): 1536-1552.

4. Titze IR, Strong WJ. Normal modes in vocal cord tissues. JASA 1975; 57(3): 736-744.

5. Baer T. Investigation of Phonation Using Excised Larynxes 1975; Unpublished Ph.D. dissertation. MIT, Cambridge, MA.

6. Finkelhor B, Titze I, Durham P. The effect of viscosity changes in the vocal fold on the range of oscillation. J. Voice 1988; 1(4): 320-325.

7. Verdolini-Marston K, Titze I, Druker D. Changes in phonation threshold pressure with induced conditions of hydration. J Voice 1990; 42(2): 142-151.

8. Gramming P. The Phonetogram. An Experimental and Clinical Study 1988; Department of Otolaryngology, University of Lund, Malmõ, Sweden.

9. Herzel H, Berry D, Titze I, Saleh M. Analysis of vocal disorders with methods from nonlinear dynamics (in review).

10. Titze I, Baken R, Herzel H.. Evidence of Chaos in Vocal Fold Vibration. In: I Titze (ed), Vocal Fold Physiology, Frontiers in Basic Science. San Diego, CA: Singular Publishing Group, Inc. 1992; 143-188.

11. Hirano M, Vennard W, Ohala J. Regulation of register, pitch, and intensity of voice. Folia Phonetica 1970; 22: 1-20.

12. Atkinson J. Correlation analysis of the physiological factors controlling fundamental voice frequency. JASA 1978; 63(1): 211-222.

13. Shipp T, McGlone R. Laryngeal dynamics associated with voice frequency change. JSHR 1971; 14: 761-768.

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

15. Titze I, Luschei E, Hirano M. The role of the thyroarytenoid muscle in regulation of fundamental frequency. J Voice 1989; 3(3): 213-224.

16. Coleman R, Mabis J, Hinson JK. Fundamental frequency-sound pressure level profiles with adult male and female voices. JSHR 1977; 20: 191-204.

17. Schutte H. Seidner W. Standardizing voice area measurement/phonetogram. Folia Phoniatrica 1983; 35: 286-288.

18. Jiang J, Titze I. (in press). Measurement of vocal fold intraglottal pressure and impact stress. J. Voice.

NCVS Status and Progress Report - 4
June 1993, 311-324

Voice Research as an Integral Part of The Denver Center for The Performing Arts

Ingo R. Titze, Ph.D.

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

Introduction

This is a brief report describing the Recording and Research Center, a division of the Denver Center for the Performing Arts. The Recording and Research Center (RRC) is unique in the world in that it is physically and administratively an integral part of a major performing arts complex in the United States. Typically, research laboratories are affiliated with universities and industry, but this innovative approach allows science to be in direct contact with the arts, thereby providing a practical service.

An overview is given here about the purpose, history, and structure of the RRC after one decade of existence. There are comments about specific research projects, the personnel involved, and sources of funding. The final section is a selected list of publications by RRC personnel in this first decade.

Purpose

The primary purpose of the RRC is to conduct basic and applied research that will aid performing artists, particularly actors and singers, who use voice as a tool of trade. The demands on the human voice are enormous in some professions, causing many fine performers to terminate a career early or to curtail their professional activities. Research is needed to get a better grasp on vocal fatigue, the origin of voice disorders, environmental effects on the voice, and optimal ways of training and rehabilitating voices.

Given this primary purpose, a secondary purpose for the RRC is to reach out to the public and inform everyone about vocal health. Because of its affiliation with the National Center for Voice and Speech, the RRC has positioned itself to become the focal point of such an outreach.

History

The RRC began as a dream shared by two friends, Wilbur J. Gould, M.D. and Donald L. Seawell, L.L.D. Gould is a world-renowned otolaryngologist practicing in New York City, and Seawell is Chairman of the Board of Trustees of the Denver Center for the Performing Arts. Both have an abiding love for the performing arts, having channeled their formal training in medicine and law, respectively, into service for those who earn their living on stage. Their dream was to build a world-class facility, not unlike an olympic training center for athletes, to study the behavior of vocal performers in action.

The opportunity to make the dream a reality came in 1983, when the Denver Center for the Performing Arts was undergoing a phase of expansion. An additional building was leased from the City of Denver. It was the former police building. On the fourth floor of this building were the jail cells, which began to look like laboratories and therapy rooms to the eyes of the visionaries, Gould and Seawell. A remarkable transformation took place in the period of about a year. Seawell often remarks jokingly: "Jim (Gould) wanted a voice laboratory, about 10 x 10. What I didn't realize is that he meant 10 city blocks by 10 city blocks".

When the basic floor plan was conceived, Gould realized that early technical assistance was the key to long-term success. He retained the services of Lawrence Kirkegaard, Ph.D., an architectural acoustician and Ingo R. Titze, Ph.D., a physicist working in the field of voice science. Kirkegaard designed and supervised construction of a recording studio and Titze designed the research laboratories. Although primary emphasis was on functionality and modern technology, the working spaces were given an artistic touch. This not only symbolized the merging between the arts and the sciences, but created a more friendly atmosphere for visiting performing artists.

After construction of the recording studio, the associated audio and video control rooms, and the research laboratories were completed, it was evident that some long-term technical direction was needed for audio and video recording and for voice research. Titze remained on board as a consultant, mapping out a strategy for research priorities and personnel that would position the RRC competitively for external funding in about five years.

For video and audio productions, Dirk Olsen was hired to develop similar long-range plans. Thus, the RRC functionally became a two-branch organization, with some overlap in support staff (engineering, administrative assistance, and clerical services). Ronald C. Scherer, Ph.D., a graduate from the University of Iowa in voice and speech science, assumed a new position as a full-time researcher. Dick Jenkins was appointed audio engineer, William Winholtz became video and general laboratory engineer, and Justin Dick served as administrative assistant.

As the two branches implemented their long-range plans, the marriage between video and audio production and research was not as natural as it once seemed. Difference sources of funding were pursued, different styles of management were observed, and different criteria for peer approval and national recognition emerged. A high level of national visibility and acceptability was obtained by both branches, as evidenced by multiple awards and publications.

In 1988, the video and audio production branch of the RRC assumed a new name, Denver Center Media, and became administratively a part of Denver Center Productions. The RRC expanded its research activities and added a public outreach and educational component. Titze became Director of Research and Scherer became Senior Scientist. Gould maintained his official title as Director of the RRC, but acted more in an advisory capacity. Scherer became the day-to-day manager of the operations and spearheaded the formal training of theatre students in voice science. He also supervised and took an active role in the Voice Workshops, the principal public outreach program of the RRC.

The final major historical development came in 1990, when researchers from the University of Iowa, the RRC, the University of Wisconsin, and the University of Utah received a major cross-institutional government grant to study voice and speech. This grant launched the National Center for Voice and Speech, a consortium of the four institutions named above. Titze became director of this National Center. The National Institute on Deafness and Other Communication Disorders, who provided the financial support, mandated that the new center provide (1) research in voice and speech communication and its disorders; (2) training of scientists at the pre-doctoral and post-doctoral level; (3) continuing education for practitioners in the field; and (4) dissemination of information to the public. Because the RRC was positioned in a major metropolitan area in the U.S. (Denver), it became the hub of the public outreach (dissemination) effort.

Facilities, Administration, and Personnel

The Denver Center for the Performing Arts was founded in 1972 and is dedicated to presenting and fostering excellence in the performing arts. The Denver Center Theatre Company is the largest Theatre Company in the Rocky Mountain region and performs in four distinctive theatres. **The Stage** seats 650 people with a thrust stage design which the audience surrounds on three sides. **The Space** is a theatre-in-the round and resembles an Elizabethan theatre seating 450. **The Source** is an intimate theatre with a thrust design of 157 seats and **The Ricketson Theatre** is a proscenium theatre and seats 195.

Robert Garner, a successful presenter of national touring companies, joined the Denver Center in 1979. Center Attractions' productions light up the Denver Performing Arts Complex in three outstanding venues; the historic **Denver Auditorium Theatre**, a beautiful arch theatre, built in 1905 and seating 2,178; the exciting new **Temple Hoyne Buell Theatre**, with a seating capacity of 2,830; and the intimate 234-seat **Galleria Theatre** for cabaret productions. Robert Garner Center Attractions brings the best of Broadway to Denver including CATS, Les Miserables, The Phantom of the Opera and upcoming Miss Saigon.

The **National Theatre Conservatory** offers a Master of Fine Arts degree in acting, a certificate of completion in Acting and in Theatre Voice Coaching and training. The conservatory is an official candidate for accreditation with the North Central Association of College and Schools. The **Denver Center Media** to-date is an outstanding Broadcast quality recording facility and sound stage that has produced a variety of film and television specials including Legends and Land of Little Rain, which aired on Public Broadcasting Systems as part of the acclaimed American Playhouse season.

Within the **Administration Building**, the **RRC** is presently occupying space on four floors. The original (and primary) space is on the fourth floor. This floor was mentioned earlier as the former location of jail cells when the building served as police headquarters for the City of Denver. On this fourth floor are the reception area, the Acoustics Laboratory, the Electrophysiology Laboratory, the Aerodynamics Laboratory, a conference room, and office spaces for administrative personnel and researchers. Since the birth of the RRC in 1983, the fourth floor space has been vastly outgrown.

Administratively, the RRC is one of three producing divisions of the Denver Center for the Performing Arts, the other two being the Denver Center Theatre Company and Denver Center

Productions. The product of the RRC is *knowledge* about the human voice. But knowledge in and of itself is useless unless it is applied. The RRC makes a deliberate effort to translate highly technical information into consumer language. The following is an overview of the departmental structure and personnel of the RRC:

Personnel

Administration

Wilbur J. Gould, M.D., Director
Ingo R. Titze, Ph.D., Director of Research
Ronald C. Scherer, Ph.D., Senior Scientist
Barbara G. Bustillos, Program Assistant
Geron E. Coale, M.A., Secretary/Receptionist

Research

Standards in Voice Analysis and Recording
- Ingo R. Titze, Ph.D., Principal Investigator
- Darrell Wong, Ph.D., Communication Engineer
- Ronald C. Scherer, Ph.D., Senior Scientist
- William S. Winholtz, A.A.S., Research Engineer
- Lorraine O. Ramig, Ph.D., CCC-SP, Research Associate
- Chwen-geng Guo, M.S., Assistant Research Scientist

Neurologic Voice Disorders
- Lorraine O. Ramig, Ph.D., CCC-SP, Principal Investigator
- Stefanie M. Countryman, M.A., CCC-SP, Speech Pathologist
- Annette Pawlas, M.A., CCC-SP, Speech Pathologist

Voice of the Performer
- Ronald C. Scherer, Ph.D., Principal Investigator
- Lawrence R. Brown, D.M.A., Research Assistant
- Chwen-geng Guo, M.S., Assistant Research Scientist

Perception of Voice Qualities
- Yoshiyuki Horii, Ph.D., Principal Investigator
- Ronald C. Scherer, Ph.D., Senior Scientist

Aerodynamics of Phonation
- Ronald C. Scherer, Ph.D., Principal Investigator
- Chwen-geng Guo, M.S., Assistant Research Scientist

Modeling of Voice Disorders
- Marshall E. Smith, M.D., Principal Investigator
- Ingo R. Titze, Ph.D., Director of Research

Dissemination, Continuing Education, Public Outreach

SCFD Voice Workshops: *The Professional Voice: Use and Abuse*
Florence B. Blager, Ph.D., Director
Marilyn A. Hetzel, Ph.D.
Ronald C. Scherer, Ph.D.
Raymond P. Wood II, M.D.

RRC Clinical Voice Seminars
Marshall E. Smith, M.D., Coordinator

RRC Research Voice Seminars
Lorraine Olson Ramig, Ph.D., CCC-SP, Coordinator

NCVS Dissemination and Continuing Education
Barbara G. Bustillos, Coordinator
William S. Winholtz, A.A.S., Video Producer and Engineer

Audio and Video Recordings
William S. Winholtz, A.A.S., Audio Producer and Engineer

Technology Transfer
Darrell Wong, Ph.D., Coordinator
Ingo R. Titze, Ph.D.

Visitor Center
Barbara G. Bustillos, Museum
William S. Winholtz, A.A.S., Audio and Video Archive

Treatment Services

Voice and Speech in Parkinson's Disease
Lorraine O. Ramig, Ph.D., CCC-SP, Research Associate
Stefanie M. Countryman, M.A., CCC-SP, Speech Pathologist
Annette Pawlas, M.A., CCC-SP, Speech Pathologist

Medical Examinations
Raymond P. Wood II, M.D., Associate Medical Director
Marshall E. Smith, M.D., Research Associate

General Voice Therapy
Florence B. Blager, Ph.D., Speech Pathologist

Voice Coaching and Training
Marilyn A. Hetzel, Ph.D., Consulting Voice Coach
Kathy Maes, Ph.D., Consulting Voice Coach

Support Services

Computer Systems Management
Vern J. Vail, B.S.

Equipment Maintenance
William S. Winholtz, A.A.S.

Materials Coordination
Geron E. Coale, M.A.

Recording and Research Center Advisory Committee

Arlen D. Meyers, M.D., Chair

Affiliations

National Center for Voice and Speech
University of Colorado-Boulder
University of Colorado Health Sciences Center
University of Iowa

Funding

The primary sponsor of research conducted at the RRC is the National Institute on Deafness and Other Communication Disorders, located in Washington, D.C. This federal agency funds several grants to individual investigators as well as a center grant to a larger group of investigators belonging to the National Center for Voice and Speech.

Synopses of Primary Research Projects

1. Toward Standards for Voice Analysis and Recording

The purpose of this eight-year project is to develop a scientific basis for standardization of acoustic voice analysis. Standardized measurements are essential for scientific research. Specific aims of the research are to develop test utterances of assessment of vocal function through acoustic analysis, to eliminate redundancies among existing measures of voice acoustic measures, to set criteria for recording voice utterances for acoustic testing, and to create archival collections of voice recording of vocal artists.

2. Mechanisms of Phonation: The Trained Voice

Well-trained actors and singers can use their voices over a wide range of qualities: loudness, pitch, time and contextual demands. This project examines how they do that. The results should add significantly to the diagnostic, training and therapy concepts and techniques for voice. Some subjects will come from the Denver Center Theatre Company and the National Theatre Conservatory.

3. Treatment Efficacy in Parkinson's Patients

One in every one hundred individuals over 60 has Parkinson's disease. At least 75 percent of these individuals have a breakdown in communication. This project determines the effectiveness of intensive voice therapy on speech improvement in this group of patients. The results will help increase understanding of the underlying physiology of the disease and its progression.

4. Speech Synthesis and Analysis

A system is developed to disseminate and retrieve information about the usefulness of computerized voice and speech analysis and synthesis. The public and communication professionals are informed about packages that help in analysis of voice and speech. This project helps to integrate and standardize clinical and research resources across the country for voice and speech diagnostics and research.

5. Perception and Acoustic Analysis of Vocal Qualities

This project determines the important and subtle acoustic changes of the voice that result in our perception of normal and abnormal voice quality. The results help us understand what it is about the function of the larynx that leads us to a wide variety of voice qualities. This enhances voice therapy training and techniques, as well as vocal training for the stage.

6. Dissemination of Findings and Information

Providing information to the public on the prevention, detection, and treatment of voice and speech disorders is the function of this project. Emphasis is placed on educating the public in voice disorders, stuttering, and velopharyngeal incompetency. Information is made available through pamphlets, video and audio tapes, workshops, and other projects. The public is informed about health care facilities and proper referral sources for voice and speech problems.

Refereed Journal Publications by RRC Investigators in the First Decade (1983-1993)

1984

Gould, W.J. (1984). The clinical voice laboratory. Annals of Otology, Rhinology and Laryngology, 93(3), 346.

Horii, Y. (1984). Some spectral characteristics of airflow signals associated with oral reading. Folia Phoniatrica, 36, 157-163.

Perlman, A.L., Titze, I.R., Cooper, D.S. (1984). Elasticity of canine vocal fold tissue. Journal of Speech and Hearing Research, 27, 212-219.

Sorensen, D., Horii, Y. (1984). Frequency characteristics of male and female speakers in the pulse register. Journal of Communication Disorders, 17, 65-73.

Sorensen, D., Horii, Y. (1984). Directional perturbation factors for jitter and for shimmer. Journal of Communication Disorders, 17, 143-151.

Titze, I.R. (1984). Parameterization of glottal area, glottal flow, and vocal fold contact area. Journal of the Acoustical Society of America, 75(2), 570-580.

1985

Alipour-Haghighi, F., Titze, I.R. (1985). Viscoelastic modeling of canine vocalis muscle in relaxation. Journal of the Acoustical Society of America, 78(6), 1939-1943.

Barcz, D.V., Wood, R.P., Stears, J., Jafek, B.W., Shelds, M. (1985). Subarachnoid space: Middle ear, pathways and recurrent meningitis. American Journal of Otolaryngology, 6, 157-163.

Cooper, D.S., Titze, I.R. (1985). Generation and dissipation of heat in vocal fold tissue. Journal of Speech and Hearing Research, 28, 207-215.

Horii, Y. (1985). Jitter and shimmer in sustained vocal fry phonation. Folia Phoniatrica, 37, 81-86.

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

Kitajima, K., Gould, W.J., Blaugrund, S. (1985). Voice onset in pathologic speakers. Acta Otolaryngologica, (Suppl. 419), 175-179.

1986

Fuller, B.F., Horii, Y. (1986). Differences in fundamental frequency, jitter and shimmer among four types of vocalization. Journal of Communication Disorders, 19, 441-447.

Gondo, K., Horii, Y. (1986). Nasalization characteristics of speech produced by young and old female adults. Folia Phoniatrica, 38, 301-302.

Haji, T., Horiguchi, S., Baer, T., Gould, W.J. (1986). Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation. Journal of the Acoustical Society of America, July, 58-62.

Horii, Y. (1986). Correlation between fundamental frequency and intensity of vibrato. Folia Phoniatrica, 38, 308.

Horii, Y., Brandengurger, P. (1986). Predicting children's oral reading proficiency from selected acoustic and perceptual measures. Journal of Communication Disorders, 19, 299-309.

Horii, Y. (1986). Utilization of finger-tapping rate with microcomputers as an index of hemispheric specialization for language processing. Journal of Computer Users in Speech and Hearing, 2, 18-24

Ramig, L.A. (1986). Aging speech: Physiological and sociological aspects. Language and Communication, 6(1), 25-34.

Ramig, L.A. (1986). Acoustic analysis of phonation in patients with Huntington's disease: A preliminary report. Annals of Otology, Rhinology and Laryngology, 95(3), 288-293.

Titze, I.R. (1986). Mean intraglottal pressure in vocal fold oscillation. Journal of Phonetics, 14, 1-6.

Titze, I.R. (1986). Session 2. Glottal flow models, Chairman's summary. Journal of Phonetics, 14, 405-406.

1987

Acker, B.F., Miller, R., Raphael, B.N., Rubin, L.S., Scherer, R.C., Sundberg, J. (1987). Discussion: The speaking voice. Journal of Voice, 1(1), 88-91.

Alipour-Haghighi, F., Titze, I.R., Durham, P. (1987). Twitch response in canine vocalis muscle tissue. Journal of Speech and Hearing Research, 30, 290-294.

Gay, M., Blager, F.B., Bartsch, K.L., Emerey, C.F. (1987). Psychogenic habit cough: Review and case reports. Journal of Clinical Psychiatry, 48(2) 483-488.

Gould, W.J. (1987). Surgery in the professional singer. Ear, Nose, Throat Journal, 66(8), 35-42.

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

Horii, Y., Ramig, P. (1987). Pause and utterance durations and fundamental frequency characteristics of repeated oral readings by stutterers and nonstutterers. Journal of Fluency Disorders, 12, 257-260.

Keidar, A., Hurtig, R., Titze, I.R. (1987). Acoustic and perceptual similarities between phonation in different voice registers. Journal of Voice, 1(2), 223-233.

Nash, M., Lucente, F.E., Srinivasan, K., Gould, W.F. (1987). Condylomatous lesions of the upper aerodigestive tract. Laryngoscope, 97, 1410-1416.

Ramig, L.A., Shipp, T. (1987). Comparative measures of vocal tremor and vocal vibrato. Journal of Voice, 1(2), 162-167.

Raphael, B.N., Scherer, R.C. (1987). Voice modifications of stage actors: Acoustic analysis. Journal of Voice, 1, 83-87.

Scherer, R.C., Titze, I.R. (1987). The abduction quotient related to vocal quality. Journal of Voice, 1(2), 246-251.

Spiegel, J.R., Sataloff, R.T., Gould, W.J. (1987). The treatment of vocal fold paralysis with injectable collagen: Clinical concerns. Journal of Voice, 1/1, 119-121.

Titze, I.R., Horii, Y., Scherer, R.C. (1987). Some technical considerations in voice perturbation measurements. Journal of Speech and Hearing Research, 30, 252-260.

1988

Blager, F.B., Gay, M.L., Wood, R.P. (1984). Voice therapy techniques adapted to treatment of habit cough: A pilot study. Journal of Communication Disorders, 21(5), 393-400.

Finkelhor, B.K., Titze, I.R. (1988). The effect of viscosity changes in the vocal folds on the range of oscillation. Journal of Voice, 1(4), 320-325.

Fukuzawa, T., Blaugrund, S.M, El-Assuooty, A., Gould, W.J. (1988). Acoustic analysis of hoarse voice: A preliminary report. Journal of Voice, 2(2), 127-131.

Fuller, B.F., Horii, Y., (1988). Spectral energy distribution in four types of infant vocalizations. Journal of Communication Disorders, 21, 251-261.

Gray, S., Titze, I.R. (1988). Histologic investigation of hyper-phonated canine vocal cords. Annals of Otology, Rhinology, and Laryngology, 97(4), 381-388.

Leonard, R.J., Ringel, R., Horii, Y. Daniloff, R. (1988). Vocal shadowing in singers and nonsingers. Journal of Speech and Hearing Research, 31, 54-61.

Perlman, A.L., Titze, I.R. (1988). Development of an *in vitro* technique for measuring the elastic properties of vocal fold tissue. Journal of Speech and Hearing Research, 31, 288-298.

Ramig, L.A., Scherer, R.C., Titze, I.R., Ringel, S.P. (1988). Acoustic analysis of voices of patients with neurologic disease: Rationale and preliminary data. Annals Otology, Rhinology and Laryngology, 97(2), 164-172.

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

Titze, I.R. (1988). A framework for the study of vocal registers. Journal of Voice, 2(3), 183-194.

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

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

Vaughan, C.W., Blaugrund, S.M., Gould W.J., Hirano, M., Osoff, R., Sataloff, R.T. (1988). Discussion: Surgical management of voice disorders. Journal of Voice, 2(2), 176-181.

1989
Alipour-Haghighi, F., Titze, I.R., Perlman, A. (1989). Tetanic contraction in vocal fold muscle. Journal of Speech and Hearing Research, 32, 226-231.

Horii, Y., Hata, K. (1989). A note on phase relationships between frequency and amplitude modulations in vocal vibrato. Folia Phoniatrica, 40(6), 303-311.

Horii, Y. (1989). Acoustic analysis of vocal vibrato: A theoretical interpretation of data. Journal of Voice, 3(1), 31-43.

Seymour, S.H., Titze, I.R. (1989). Audio and sound archives as a resource for voice science: A preliminary survey. Journal of Voice, 3(4), 277-282.

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

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

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

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

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

Wexler, D., Gray, S., Titze, I. (1989). Phonosurgical studies: Fat-graft reconstruction of injured canine vocal folds. Annals of Otology, Rhinology and Laryngology, 98, 668-673.
1990
Hil, D., Meyers, A., Scherer, R. (1990). A comparison of videolaryngostroboscopy and GLIMPSES in the analysis of vocal fold movement. Journal of Voice, 4(3), 185-197.

Pinto, N., Titze, I.R. (1990). Unification of perturbation measures in speech analysis. Journal of the Acoustical Society of America, 87(3), 1278-1289.

Ramig, L., Scherer, R., Klasner, E., Titze, I., Horii, Y. (1990). Acoustic analysis of voice in amyotrophic lateral sclerosis: A longitudinal case study. Journal of Speech and Hearing Disorders, 55(1), 2-14.

Smith, M.E., Kimball E.L. (1990). "I fell among cutthroats" - Spencer W. Kimball and cancer of the larynx. Journal of Collegium Aescupapium, 5, 15-25.

Smith, M.E., Berke, G.S. (1990). Effects of phonosurgery on laryngeal vibrations: part one - Theoretic considerations. Otolaryngology - Head and Neck Surgery, 103, 380-390.

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

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

1991

Alipour-Haghighi, F., Titze, I.R. (1991). Elastic models of vocal fold tissues. Journal of the Acoustical Society of America, 90(3), 1326-1331.

Alipour-Haghighi, F., Perlman, A., Titze, I. (1991). Tetanic response of the cricothyroid muscle. Annals of Otology, Rhinology & Laryngology, 100(8), 626-631.

Berke, G.S., Green, D.C., Smith, M.E., Arnstein, D.A., Honrubia, V., Natividad, M., Conrad, W. (1991). Experimental evidence in the "in vio" canine for a collapsible tube model of phonation. Journal of the Acoustical Society of America, 89, 1358-63.

Friduss, M.E., Mickel, R.A., Smith, M.E., Zimmerman M.C., Tartell P.B. (1991). Computer assisted instruction in otolaryngology: Use of a database for examination review. Otolaryngology - Head and Neck Surgery, 105, 723-726.

Karnell, M.P., Scherer, R.C., Fischer, L.B. (1991). Comparison of acoustic voice perturbation measures among three independent voice laboratories. Journal of Speech and Hearing Research, 34, 781-790.

Smith, M.E., Green D.C., Berke, G.S. (1991). Pressure-flow relationship in the canine larynx. Journal of Voice, 5, 10-17.

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

Wong, D., Ito, M., Cox, N., Titze, I. (1991). Observation of perturbations in a lumped-element model of the vocal folds with applications to some pathological cases. Journal of the Acoustical Society of America, 89(1), 383-394.

1992

Bredenkamp, J.M., Smith, M.E., Dudley, J.P., Williams, J.C., Crumley, R.L., Crockett, D.M. (1992). Otolaryngologic manifestations of the mucopolysaccharidoses. Annals of Otology, Rhinology and Laryngology, 101, 472-478.

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

Gray, S., Barkmeier, J., Jones, D., Titze, I., Druker, D. (1992). Vocal evaluation of thyroplastic surgery in the treatment of unilateral vocal fold paralysis. Laryngoscope, 102, 415-421.

Ramig, L. Olson, Winholtz, W. (1992). The vocal demodulator for analysis of vocal tremor. Journal of Speech and Hearing Research, 35, 562-573.

Ramig, L. Olson, Winholtz, W., Aronson, A., Sibler, S. (1992). Acoustic characteristics of vocal tremor or "flutter" in patients with amyotrophic lateral sclerosis. Annals of Otology, Rhinology and Laryngology, 101, 511-518.

Ramig, L.O., Fred, C., Breeze, R., Rosenberg, N., Schneck, S., Driek, E., Qi, J., Lone, T., Zhang, Y., Snyder, J., Wells, T., Thompson, L., Mazziotta, J., Huang, S., Grafton, S., Brooks, D., Sawle, G., Schroter, G., Ansari, A. (1992). Survival of Implanted Fetal Dopamine Cells and Neurologic Improvement 12 to 46 months After Transplantation for Parkinson's Disease. Reprint from the New England Journal of Medicine, 327, 1549-1555 (November 26).

Smith, M.E., Berke, G.S., Gerratt, B.R., Kreimanm, J. (1992). Laryngeal paralyses: theoretical considerations and effects on laryngeal vibration. Journal of Speech and Hearing Research, 35, 545-554.

Smith, M.E., Mortelliti, A.J., Cotton R.T., Myer C.M. (1992). Phonation and swallowing considerations in pediatric laryngotracheal reconstruction. Annals of Otology, Rhinology and Laryngology, 101, 731-738.

Titze, I.R. (1992). Rationale and structure of a curriculum in vocology. Journal of Voice, 6(1), 1-9.

Titze, I.R. (1992). Acoustic interpretation of the voice range profile (phonetogram). Journal of Speech and Hearing Research, 35, 21-34.

Titze, I.R. (1992). Vocal efficiency. Journal of Voice, 6(2), 135-138.

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

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

In Press/Review

Brown, L.R., Scherer, R.C. Laryngeal reiteration in trillo. Journal of Voice.

Goetze, M., Horii, Y. A comparison of the effect of text and 'loo' on pitch accuracy in young children's singing. Psychology of Music.

Herzel, H., Berry, D., Titze, I.R., Saleh, M. Analysis of vocal disorders with methods from nonlinear dynamics. Journal of Speech and Hearing Research.

Horii, Y. Frequency and amplitude modulations in sustained vowel phonations produced by dysarthric patients. Journal of Speech and Hearing Disorders.

Hsiao, T.Y., Solomon, N.P., Luschei, E.S., Titze, I.R. Modulation of fundamental frequency by laryngeal muscles during vibrato. Journal of Voice.

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

Jiang, J.J., Titze, I.R. A methodological study of hemilaryngeal phonation. Laryngoscope.

Kotby, M.N., Titze, I.R., Saleh, M.M., Berry, D.A. Fundamental frequency stability in functional dysphonia. Acta Otolaryngologica (Stockholm).

Smith, M.E., Marsh J.H., Cotton R.T., Myer C.M. (1992). Voice problems after pediatric laryngotracheal reconstruction: vidolaryngostroboscopic, acoustic, and perceptual assessment. International Journal of Pediatric Otorhinolaryngol.

Sundberg, J., Scherer, R., Titze, I. Phonatory control in male singing. A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. Journal of Voice.

Thom, J., Horii, Y., Roush, J. Nasal-oral amplitude ratios of speech by hearing-impaired and normal hearing subjects. Journal of Speech and Hearing Disorders.

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

Titze, I.R., Jiang, J.J., Hsiao, T.Y. Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. Annals Otology, Rhinology & Laryngology.

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

Titze, I.R., Liang, H. Comparison of F_0 extraction methods for high precision voice perturbation measurements. Journal of Speech and Hearing Research.

Titze, I.R. Toward standards in acoustic analysis of voice. Journal of Voice.

Titze, I.R. Mechanical stress in phonation. Journal of Voice.

Titze, I.R., Solomon, N.P., Luschei, E.S., Hirano, M. Interface between normal vibrato and artificial stimulation of laryngeal muscles at near vibrato rates. Journal of Voice.

Titze, I.R. Current topics in voice production mechanisms. Acta Otolaryngologica (Stockholm).

Verdolini-Marston, K., Titze, I.R., Sandage, M. Effect of hydration treatments of laryngeal nodules and polyps and related measures. Journal of Voice.

NCVS Status and progress Report - 4
June 1993, 325-327

Public Outreach Update

Barbara Bustillos, Dissemination Coordinator
The Recording and Research Center, The Denver Center for the Performing Arts
Julie Ostrem, Continuing Education Coordinator
Department of Speech Pathology and Audiology, The University of Iowa

Dissemination

The purpose of dissemination is to distribute information to the general public on prevention, detection and treatment of voice and speech disorders. The methods of reaching the public vary from the use of media (television, radio and print), educational presentations (workshops, seminars and lectures) to networking with other professional organizations dedicated to voice care.

The key point for Dissemination is to identify a common language between research scientists and the general public, with the inclusion of underserved populations. One example of this was a recent voice workshop directed by Dr. Florence Blager.

The African-American and Latino Gospel/Folk choir groups were invited to a choir/chorus directors voice workshop. The interaction among large chorus and opera groups with the inner-city church choirs brought about joint questions of concern:

"I've never seen a scientific approach to voice before and I was intrigued," said one participant. "I have concerns about my voice because of my age and the fact that I fatigue easily and become hoarse. I want to learn how to care for my voice better."

Another commented, "Coming to this workshop will enhance our singing for the workshop service...we want to give our best. These people here today will become the sparks and bring back to the community information and make the cycle of learning complete."

Dissemination activities reach a variety of age groups and professions with a wealth of information about voice and speech. An unofficial dissemination report has logged over 140 activities from 1992 to the present. NCVS investigators have not only attended to matters nationally, but internationally have addressed voice issues in France, Austria and Egypt.

Slide shows, video presentations and fiberoptic demonstrations of the vocal folds are taken on the road to share recent scientific findings to the public. A special target group has been youth.

Recently, at a jazz festival at the University of Northern Colorado, 16- to 22-year-old students got a first "inside" look at their vocal instrument.

"I never really knew how the vocal folds worked, and now every time I sing I'm going to think of that picture I saw on the video and remember," said one observer.

"I sing every day and I also teach children's choir. The new things I've learned will be lots of fun to teach children. I think children should start learning about voice when they are little so they can develop good habits and learn to sing correctly."

Networking between the NCVS and groups such as the VoiceCare Network and the Voice Foundation are only a few examples of a networking partnership spirit. VoiceCare Network Director Leon Thurman recently reported figures of over 700,000 in a two-year period of student and teacher outreach. Dr. Robert Sataloff was invited to the Denver Voice Care Clinics to address the issue of professional voice care and treatment to both a medical and community audience.

Dissemination investigators hope to create more partnerships in the future to address vocal care concerns. Museums, libraries, concerts and other public forums will be sought to reach the public. The leadership of concerned entertainers in the arts will be sought to place their mentor roles with the public and help give educational messages that concern both the entertainer and the scientist.

An example of this was a recent meeting between international jazz vocalist Bobby McFerrin and scientist Ingo Titze. Both expressed concerns of voice education and voice care. McFerrin summarized their feelings by saying:

"The only thing I care about...the only thing I'm concerned about is staying healthy vocally so I can continue to sing as long as possible."

Continuing Education

As stated in previous reports, the goal of the Continuing Education module of the NCVS is to hasten new information from research laboratories to clinicians. Groups targeted to receive this information are otolaryngologists, speech-language pathologists, and voice teachers/coaches. The NCVS has made progress toward this goal with implementation of seminars, videotapes, computer software and lectures. More than 7 million individuals--practioners and the general public--have been reached through the efforts of NCVS investigators over the past year. These investigators will continue to experiment with various media avenues to provide this information in ways that are effective and innovative, yet inexpensive and accessible to practioners.

A third Phonosurgery conference is being planned for Spring '94 in Madison, Wisconsin. Dr. Diane Bless coordinates the two-day seminars that feature presentations from NCVS investigators and other respected experts in the discipline. Approximately 200 otolaryngologists, speech-language pathologists, researchers and doctoral students attended each of the first two conferences.

John Laver, a professor from The University of Edinburgh, presented a week-long training course at The University of Wisconsin. The "each one teach one" workshop approach attracted clinicians who are known to be good teachers and have a willingness to teach material in different geographical regions in the U.S. and Canada. No registration fee was charged, although participants were encouraged to distribute the information to clinicians locally.

A software program for speech-language pathologists has been developed and is nearly ready for peer review. Users interactively update their skills on topics such as assessment and treatment of voice and speech disorders. Some of the advantages of providing continuing education using software are: it provides easy access to new information; it can be used in one's own home or office at the practitioner's convenience; it provides immediate feedback about the efficacy of knowledge gained; and it provides visual animation of concepts.

As the NCVS passes the halfway point in its first granting period, investigators involved in Continuing Education are evaluating current projects and discussing plans for future products. Under consideration are: videotapes detailing laboratory procedures for colleagues, video journals, video tutorials (some of which accompany textbooks), audio tapes, an electronic bulletin board service, as well as more traditional formats such as workshops. In addition to the phonosurgery conference described, Dr. Lorraine Ramig has suggested a series of workshops to illustrate effective methods for treatment of Parkinson patients.

NCVS Status and Progress Report - 4
June 1993, 329-331

Training Update

John Folkins, Ph.D., Training Coordinator
Department of Speech Pathology and Audiology, The University of Iowa

A thorough description of the NCVS training program has been described in previous issues of the status and progress report (See Volumes 1 and 2). In this summary, the NCVS trainees will be introduced.

Currently, two postdoctoral fellows, Drs. David Berry and Young Min, and six predoctoral trainees, Julie Barkmeier, Eileen Finnegan, Eileen Savelkoul, Alice Smith, Brad Story and Kenneth Tom, are funded by the NCVS Training program. In addition to the these predoctoral students, David Druker has participated as an NCVS predoctoral trainee. David has received a prestigious award from The University of Iowa. He has been designated as a University Scholar, and he receives stipend and tuition support from this award.

Postdoctoral Fellows

David Berry

After earning a doctorate in physics from Brigham Young University, David Berry accepted a postdoctoral position with the NCVS in September 1991. His major research interest is voice analysis and voice synthesis. Specifically, he is analyzing voice disorders with methods of nonlinear dynamics, modal analysis of finite element models, and separation of trends and subharmonics from random variations in the fundamental frequency and amplitude of the voice. He works closely with Dr. Ingo Titze and has a number of papers co-authored with Dr. Titze in review.

Young Min

Young Min has nearly completed the first of a two-year postdoctoral traineeship with the NCVS. After earning her medical degree from Stanford University, she elected to pursue her interest in research laryngology before beginning a residency in Otolaryngology. At The University of Iowa, she works closely with Drs. Harry Hoffman, Ingo Titze, Erich Luschei and Katherine Verdolini. During the past year, she has been involved in a number of research projects, including quantitative analysis of EMG waveforms in laryngeal paralysis, clinical applications of telemetry of laryngeal

EMG signals, stress-strain responses of the vocal ligament, hydration effects on phonation threshold pressure, role of magnetic resonance imaging in the evaluation of laryngeal paralysis, and botulinum toxin distribution within muscle after injection.

Predoctoral Trainees

Julie Barkmeier

Julie Barkmeier earned a bachelor's degree in psychology and a master's degree in speech pathology and audiology, both from The University of Iowa. She began her Ph.D. program in Fall 1989. Her research interests include neuroanatomy and physiology of the recurrent laryngeal nerve and clinical research pertinent to neurolaryngological populations, such as spasmodic dysphonia and vocal fold paralysis. Her primary mentors are Drs. Erich Luschei and Harry Hoffman. She is in the final stages of completing her dissertation, an electromicroscopic study of the epineurium of laryngeal nerves. She has begun the application process for a postdoctoral fellowship.

David Druker

David Druker began his career as a researcher in the area of voice and speech, not as a student, but as a research assistant for Dr. Ingo Titze. During the four years he worked in the Voice Acoustics and Biomechanics Lab at Iowa, he honed his interest to the area of articulatory speech synthesis. His previous work and educational experience had been in the areas of industrial education and electronics.

He began studies as a full-time doctoral student in Fall 1991. He plans to complete his NCVS program at The University of Utah under the direction of Dr. Steven Gray.

Eileen Finnegan

Eileen Finnegan earned a bachelor's degree in communicative disorders in 1980 and a master's degree in speech pathology and audiology from the University of Iowa in 1982. During her nearly ten years of work as a clinical speech-language pathologist, she focused on voice as her major area of interest. She enrolled in the doctoral program at The U of I in January of 1992.

Her key research interests are voice disorders and studies of laryngeal physiology and anatomy, under the direction of Dr. Erich Luschei.

Eileen Savelkoul

Eileen Savelkoul received a master's degree in speech-language pathology from Texas Christian University. She recently completed her first year of coursework at The University of Iowa in areas of language development, statistics, counseling, anatomy and physiology of the speech mechanism, speech physiology, and stuttering. Her clinical experience and coursework have revolved around her interest in stuttering and treatment efficacies.

She works closely with Dr. Patricia Zebrowski and is presently planning a research program in stuttering.

Alice Smith

Alice Smith has a bachelor's degree in German from Mary Baldwin College, a master's degree in therapeutic recreation from The University of North Carolina, and a master's degree in communication sciences and disorders from The University of Montana. Before pursuing doctoral studies at The University of Iowa, she worked as a hospital speech pathologist and as an instructor at The University of Montana.

At the University of Iowa, her research interests have been in measuring electromyographic activity from the velopharyngeal muscles. She works closely with Drs. John Folkins and Jerry Moon.

Brad Story

A lifelong interest in the sounds made by musical instruments led Brad Story to a doctoral program at The University of Iowa. In 1987, he earned a bachelor's degree in applied physics from The University of Northern Iowa. For the next three and one-half years, he worked in the private sector as an acoustical engineer developing mathematical and computer models for mufflers for heavy equipment. In his doctoral studies, he has completed coursework in areas such as acoustics and biomechanics of speech, speech physiology, and fundamentals of vibration. His primary research interest is the modeling of vocal tract acoustics.

He co-authored a manuscript entitled, "Acoustics of the Tenor High Voice," currently submitted to *The Journal of the Acoustical Society of America.*

Kenneth Tom

Kenneth Tom received a bachelor's degree in vocal performance from The New England Conservatory in 1980, and a master's degree in communication disorders from San Diego State University. He worked in a hospital as a speech pathologist before entering the doctoral program at The University of Iowa in Fall 1988. After completion of required coursework, he began work on his dissertation, entitled "Intensity Control in Male Falsetto Phonation." His advisor is Dr. Ingo Titze.

Kenneth has accepted a position as assistant professor at Emerson College in Boston. He will begin his duties there Fall 1993.