DCPA-RR 1/1985

The Recording And Research Center

RESEARCH REPORT

APRIL, 1985

The Denver Center For The Performing Arts
1245 Champa Street, Denver, Colorado 80204

DCPA-RR 1/1985

The Recording and Research Center

RESEARCH REPORT

April, 1985

THE DENVER CENTER FOR THE PERFORMING ARTS
The Recording and Research Center
1245 Champa Street
Denver, Colorado 80204
(303) 893-4000

The Denver Center For The Performing Arts

BOARD OF TRUSTEES

Donald R. Seawell, Chairman
Harold V. Cook, First Vice President
Walter Predovich, Second Vice President
Rev. John V. Anderson, Secretary
Barbara E. Mackay, Assistant Secretary
Earl R. Moore, Treasurer
Stan J. Boulier
Mrs. Barbara Davis
John Felt
Robert S. Garner
E. Atwill Gilman
Rex John
Dr. Frank McGlone
Frank H. Ricketson
Cleo Parker Robinson
Mary Roebling
William Scheitler
Dolphe Simic
Warren K. Young

HONORARY MEMBERS

Governor Richard D. Lamm
Mayor Federico Peña
Donald T. Carney
Lucien Wulsin, Chairman, Colorado Council on the Arts and Humanities

NATIONAL ADVISORY BOARD

Douglas Fairbanks, Jr.
Justice Arthur J. Goldberg
Wilbur James Gould, M.D.
Roger L. Stevens

DEPUTY CHAIRMAN

Barbara E. Mackay

MANAGEMENT COMMITTEE

Karyn Browne, Director of Administration
Dorothy V. Denny, Director of Development
Justin Dick, Director of Adminstration, Recording and Research Center
Robert S. Garner, Director, Center Attractions
Sarah Lawless, Executive Director, Denver Center Theatre Company
Henry E. Lowenstein, Producer, Lowenstein Theatre
Adam Reilly, Director, Denver Center Cinema
Brockman Seawell, Director, Denver Center Productions
Gully Stanford, Director of Public Affairs
Jan Steinhauser, Director of Public Relations

ACKNOWLEDGEMENTS

The research reported herein was supported in part by the following organizations:

Helen G. Bonfils Foundation

Hewlett Packard

Digital Equipment Corporation

National Institutes of Health - Grant No. NS 16320-5

CONTENTS

A. THEORY OF VOICE PRODUCTION

B. VOICE DIAGNOSTICS AND TREATMENT

C. VOICE PRODUCTION IN TRAINED VOCALISTS

D. EFFECTS OF AGING AND DISEASE ON VOICE PRODUCTION

Recording and Research Center
RESEARCH PERSONNEL

1. ADMINISTRATION

Wilbur J. Gould	Director of Recording and Research Center
Ingo R. Titze	Director of Research
Ronald C. Scherer	Research Scientist
Justin Dick	Director of Administration of the RRC
Raymond P. Wood	Associate Medical Director
Kim R. Sheplor	Administrative Assistant

2. INVESTIGATORS

Bonnie Raphael, Assistant Research Scientist and Voice Coach, affiliated with Denver Center Theatre Company

Lorraine Ramig, Research Associate, affiliated with the Department of Communicative Disorders and Speech Science, University of Colorado, Boulder

Florence Blager, Speech Pathologist, affiliated with the University of Colorado Health Sciences Center

Yoshiyuki Horii, Research Associate, affiliated with the Department of Communicative Disorders and Speech Science, University of Colorado, Boulder

David Kuehn, Research Associate, affiliated with the Department of Speech Pathology and Audiology, The University of Denver.

3. TECHNICAL STAFF

Willie Winholtz	Chief Engineer
Dick Jenkins	Audio Engineer
Mike Achten	Software Consultant
Jay Shaffer	Production Technician

4. INTERNS

Sandy Anderson	(Oct. 1, 1984-)
Julie Artigliere	(Sept. 1, 1984-)
Ken Barkin	(Jan. 3, 1985-)
Karen Bond	(Oct. 10, 1984-)
Judy Coe	(Oct. 1, 1984-)
Norris Kelly	(Jan. 3, 1985-)
Vernon Vail	(Aug. 1, 1984-)

FOREWORD

This is the first of a series of research reports that will be issued by the Recording and Research Center of the Denver Center for the Performing Arts. In a broad sense our research is aimed toward the betterment of the performing arts. This includes an improved understanding of how the human body functions (or dysfunctions) when highly refined and skilled motor tasks are executed during rehearsal and performance. It includes improved understanding of how performance is affected by the environment (climate, pollution, noise, architecture, etc.), and how performance is changed under conditions of fatigue, aging, food and drug intake and disease.

Our initial focus is on voice production. Investigations of the larynx and vocal tract from the basic science point of view, making models and predictions of its function under normal and exceptional conditions, are going in collaboration with the University of Iowa. In addition, we are studying actors, singers and professional speakers in natural and controlled situations, and recording their voices with state-of-the-art instrumentation. Modification (enhancement) of recorded voice characteristics is a further aim of this research group.

We welcome comments and suggestions by our readership and hope that these reports will be of value to the artistic and the scientific community.

Ingo R. Titze
Director of Research
Recording and Research Center

THE PHYSICS OF FLOW-INDUCED OSCILLATION OF THE VOCAL FOLDS

PART I. SMALL-AMPLITUDE OSCILLATIONS

Ingo R. Titze

University of Iowa
and
The Denver Center for the Performing Arts

ABSTRACT

A theory of vocal fold oscillation is developed on the basis of the body-cover hypothesis. The body (vocalis muscle) is represented by a single lumped mass-spring-damper, and the cover (mucosa) is represented by a distributed surface layer that can propagate a wave. Linearization of the surface-wave displacement along the glottis results in an autonomous differential equation written explicitly in terms of tissue displacement and velocity, and further small-amplitude approximations yield closed-form expressions for conditions of oscillation. It is shown that effective damping of the folds is reduced by reducing the mucosal wave velocity, by increasing the subglottal pressure, and by approximating the vocal folds. A three-dimensional region of oscillation is defined on the basis of these source parameters, and the effect of vocal tract acoustic loading is included by an additional negative damping term that is proportional to air inertance of the vocal tract. The treatment is harmonized with former treatments based on two-mass models and collapsible tubes, and a tutorial section is provided. A sequel paper (Part II) will cover limit cycles and large-amplitude oscillation.

INTRODUCTION

A detailed analytical treatment of vocal fold oscillation is found in Ishizaka and Matsudaira (1972). This treatment includes (1) a review of pressure-flow relationships in ducts and orifices under steady flow conditions, and (2) an application of these relationships to a soft-walled glottal duct (represented by two masses and three springs) in order to demonstrate small-amplitude oscillation conditions. It is shown that a nonlinear aerodynamic reacting pressure on the lower mass (resulting from a phase lag of the upper mass) leads to small-amplitude (linearized) negative stiffness and damping, and regions of stability and instability are bracketed on the basis of average flow, average glottal opening, and spring constants.

Although the treatment serves as a major milestone in quantifying vocal fold vibration, its application to general phonatory theory has been limited because (1) the key parameters, individual masses and stiffnesses of the coupled oscillators have been difficult to relate to any particular anatomical structure of the tissue (and hence have been difficult to measure), (2) the oscillation conditions are stated in rather complex mathematical terms because the characteristic equations are fourth-order, and

(3) some of the results, such as fundamental frequency dependence on subglottal pressure, are not in agreement with observation. The treatment has been clarified and tutorialized by Stevens (1977) and Broad (1979), but still needs more conceptual simplicity to be incorporated routinely into the speech and hearing science literature. The more recent collapsible-tube analogy by Conrad (1980) does offer a more appealing conceptualization, but it is oversimplified in terms of tissue movement and acoustic loading conditions. The collapsible-tube problem appears to be a special case of the coupled oscillator problem solved by Ishizaka and Matsudaira (1972), and will also be shown to be a special case in our formulation.

This paper, then, serves two purposes. The first is to present a framework of basic principles by which the mechanics of vocal fold vibration can be understood on limited mathematical terms, and the second is to rework the linearized (small-amplitude) analytical treatment with parameters that lend themselves to more direct experimental confirmation. The small-amplitude restriction means that the results apply only to growing oscillations that build around slightly abducted vocal folds, with no glottal closure. A by-product of this analytical treatment will be the unification of the two previous small-amplitude approaches by Conrad (1980) and Ishizaka and

Matsudaira (1972). The more general large-amplitude theory, which involves limit cycles and special nonlinear effects due to vocal fold collision, will be discussed in a follow-up paper. Because the large amplitude theory also lends itself better to experimental verification, we will restrict ourselves to theoretical issues in this first paper, leaving a sizable amount of data gathered on excized larynges as part of the follow-up discussion.

I. BASIC PRINCIPLES OF VOCAL FOLD VIBRATION

The first principle to establish is that sustained vocal fold oscillation is flow-induced. Without airflow in the glottis, only transient oscillatory responses can be evoked by muscle contractions or other excitations. A continual flow of energy from the glottal airstream to the tissue is necessary to overcome the frictional losses in vocal folds. In addition, of course, inertial and elastic properties (mass and stiffness) of the tissue provide the necesssary overshoot and restoring force, as needed in any mechanical oscillator.

A positive flow of energy from the airstream to the tissue can be realized if the net aerodynamic driving force has a component in phase with the tissue velocity. Consider the fundamental equation of motion for a mass-spring oscillator,

$$M\ddot{\xi} + B\dot{\xi} + K\xi = f(\xi,\dot{\xi}, t) \quad , \tag{1}$$

where M, B, and K are mass, stiffness, and damping, respectively, ξ , $\dot{\xi}$, and $\ddot{\xi}$ are displacement, velocity, and acceleration, respectively, f is the driving force, and t is time. If f is time-indepedent, the differential equation is said to be autonomous. This is the case of interest here, because it describes systems that self-oscillate rather than those that are driven into oscillation by an external source.

The dependence of f on $\dot{\xi}$ is crucial. Whenever f is in the direction (or in phase with) the velocity, energy is imparted to the mass-spring-damper system (Figure 1). Conversely, whenever f is opposite to the direction (or out of phase with) the velocity, energy is taken out of the mass-spring-damper system. This kind of forcing can lead to either self-oscillation or driven oscillation. The key is whether or not the moving hand is considered part of the system. If it is, the system will be governed by an autonomous differential equation. Otherwise the differential equation is non-autonomous.

A more interesting case is illustrated in Figure 2. The child on the swing must certainly be considered a part of the

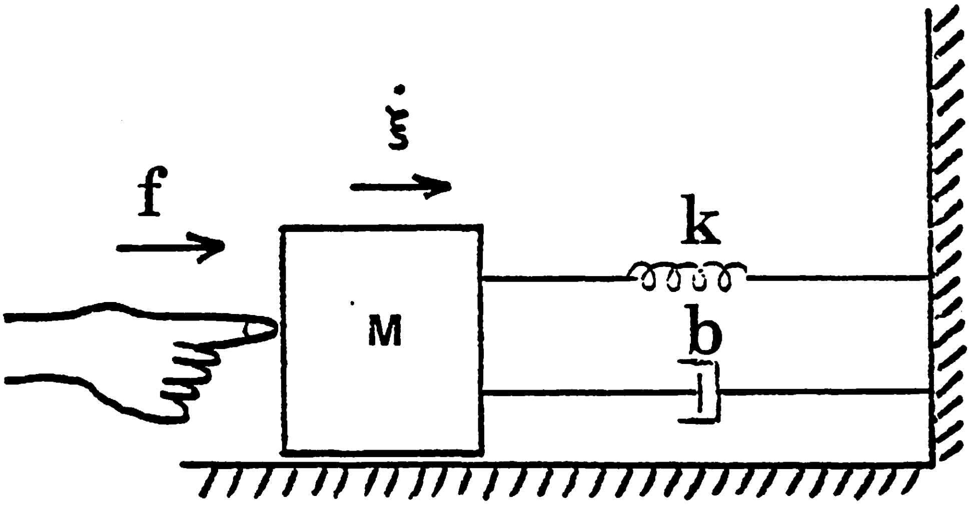
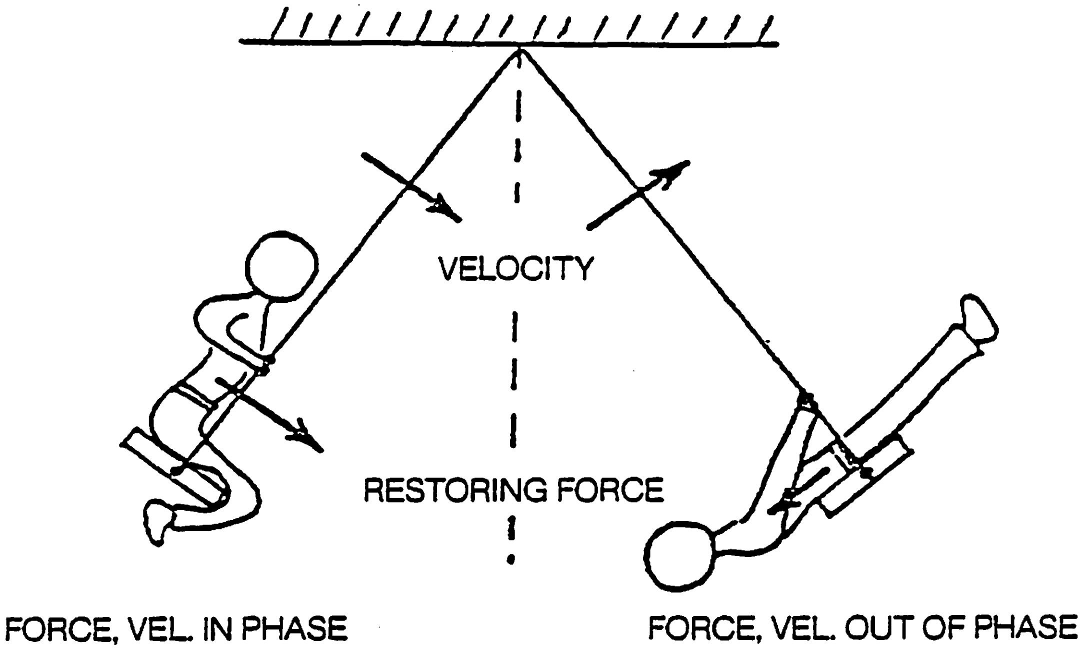
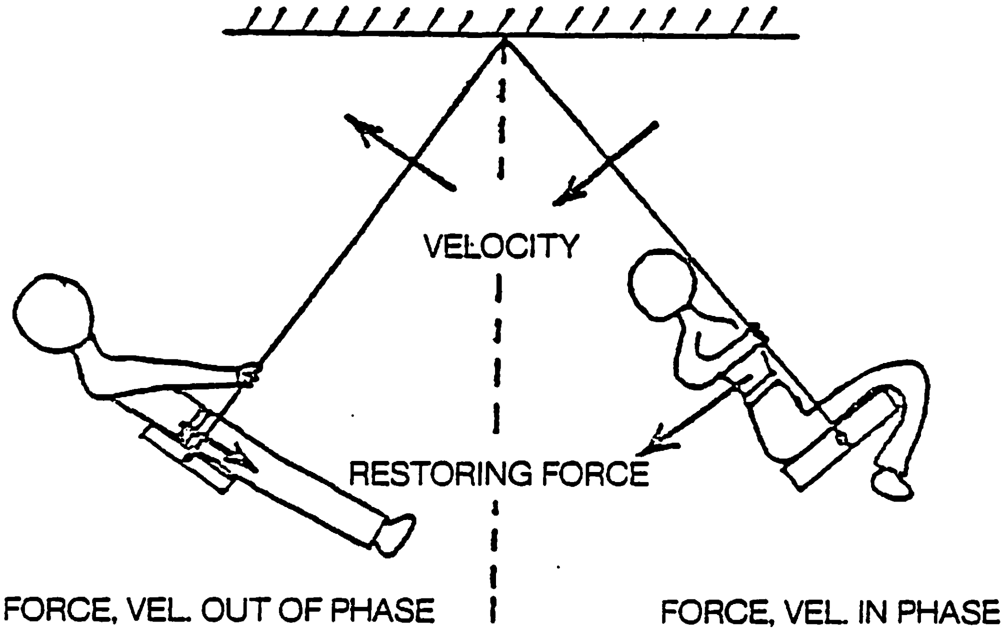


Figure 1. Simple mechanical oscillator. Driving force is in direction of velocity.



(a) Motion to the right.



(b) Motion to the left.

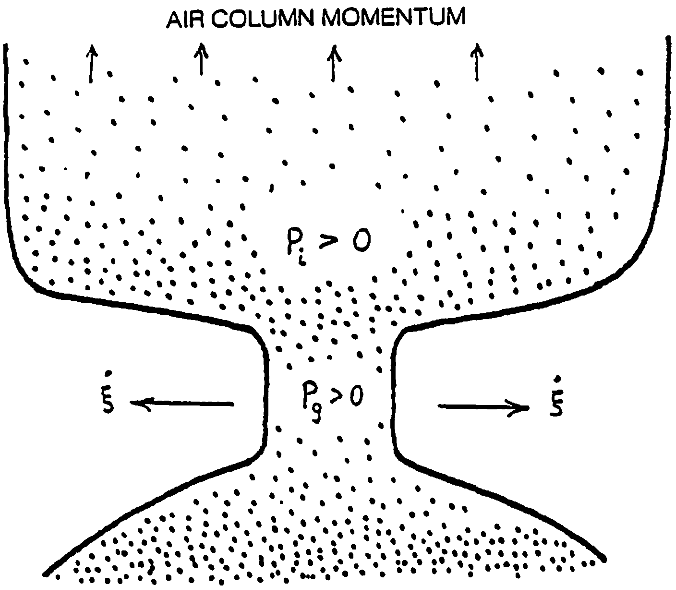
Figure 2. Self-sustaining oscillations by child on swing. Restoring force is greater when it is in direction of velocity.

pendulum oscillator. Its body supplies the energy for sustained oscillation. The child learns to raise and lower its center of gravity such that the effective restoring force is always greatest when it is in phase with the velocity (Figure 2a left and Figure 2b right), thus pumping up the energy of the oscillator. Rotational energy from leg curling and stretching may also be transformed to translational energy if the body motions are properly timed. The timing (or phasing) of forces with velocity is critical, and is often not fully appreciated by the inexperienced child.

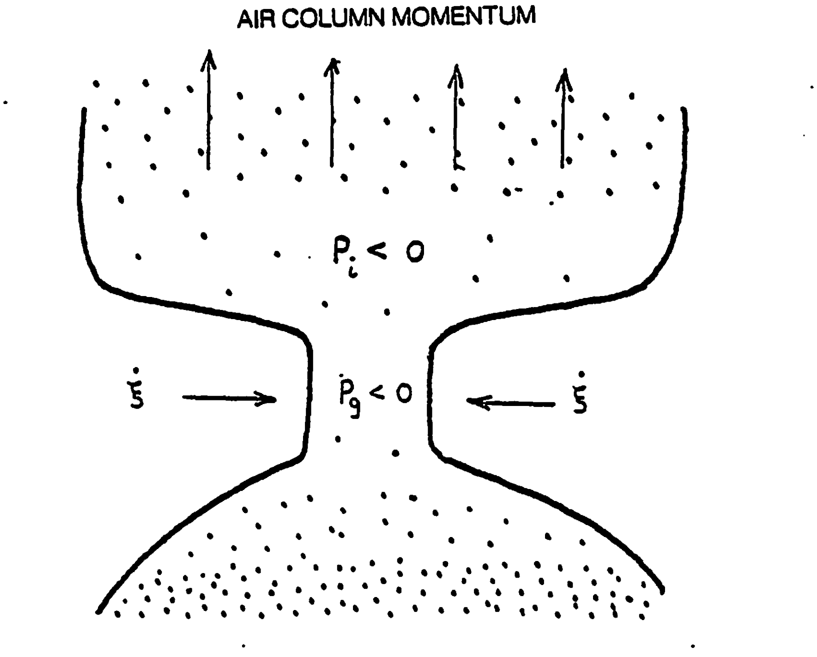
The question now becomes, how does the glottal airstream in the vocal fold system supply a velocity-dependent driving force? The explanation often given in elementary textbooks, (e.g., Lieberman, 1977) that a negative Bernoulli force sucks the vocal folds together prior to closure, works only if we pay no attention to what happens on the return path. The same force would be in the opposite direction to the velocity just after glottal opening, thus cancelling the impulse imparted prior to closure. In analogy with the child on the swing, somehow the system needs to change the effective driving force on alternate half-cycles.

A driving-force asymmetry in the vocal folds can be achieved in at least two fundamentally different ways, (1) by creating different supraglottal or subglottal "loading" pressures during different half-cycles, or (2) by varying the glottal geometry to create different intraglottal pressures in different half cycles. These two conditions are, of course, not mutually exclusive, so that in normal phonation both may occur simultaneously.

Consider the vocal tract loading pressures first. At fundamental frequencies below the first formant frequency (the usual case for speech), the air column in the vocal tract is primarily inertive, i.e., it acts like a mass of air that is accelerated and decelerated as a unit (Rothenberg, 1981). As the glottis is opening and the glottal flow is increasing, the air column in the vocal tract is being accelerated along with the glottal air. Since a positive acoustic pressure is needed for this acceleration, the vocal tract input pressure P_i rises above atmospheric pressure (Figure 3a). The average intra-glottal pressure P_g rises corresondingly, pushing the vocal folds apart. The driving force is thus in the direction of tissue velocity, and energy is supplied to the vocal folds. When the glottis is closing, on the other hand, the supraglottal air column tends to maintain its forward momentum. It creates a negative supraglottal pressure P_i in its wake (Figure 3b). This negative pressure (suction) lowers the average glottal pressure, P_g ,



(a) GLOTTAL OPENING



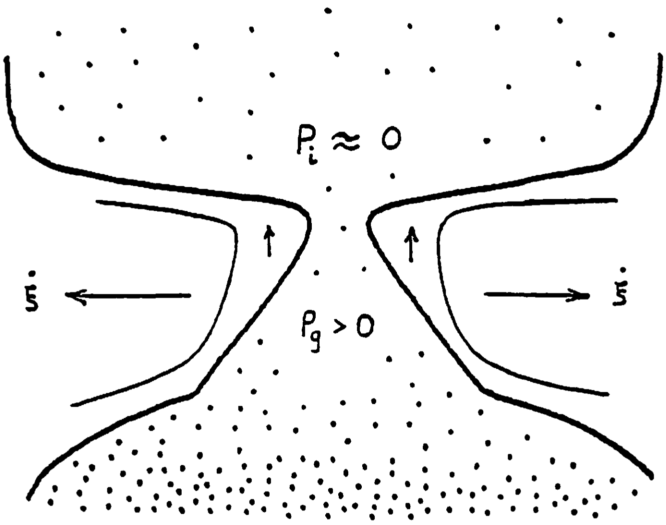
(b) GLOTTAL CLOSING

Figure 3. Glottal pressures during opening and closing of the glottis. Due to air inertia, largest glottal pressures occur during opening.

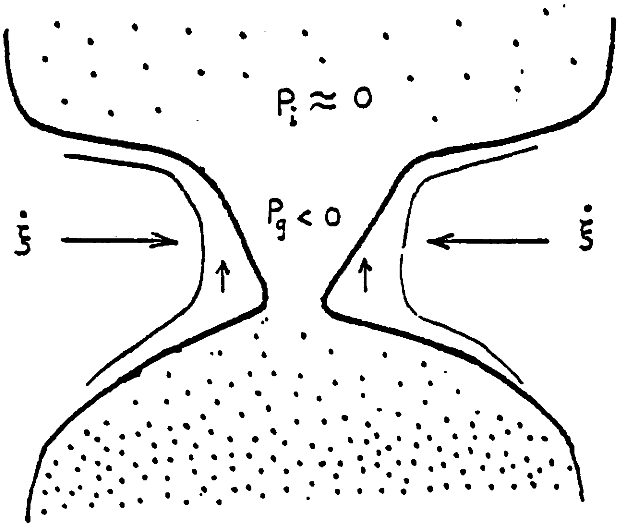
causing the driving force on the tissue to be diminished on this return path. At some point prior to closure, the average glottal pressure may actually become negative, in which case the driving force is again in phase with the tissue velocity. This need not happen, however. The only requirement is that the driving force is less positive during closing than opening.

It is important to recognize that this explanation differs markedly from traditional explanations involving only the (negative) Bernoulli force. The key here is the supraglottal inertive load. Without it, the argument about velocity-dependent driving forces fails.

Consider now the second, distinctly different, condition under which velocity-dependent driving forces can be achieved. If the tissue has a normal mode of vibration that results in different glottal shapes over different portions of the cycle, and if the different glottal shapes produce different glottal pressure profiles, a force asymmetry can also be established. This is illustrated in Figure 4. In order not to complicate the explanation, assume that in this case there is no vocal tract loading. This makes the supraglottal acoustic pressure P_i approximately zero (atmospheric). Furthermore the airstream emerges from the glottis in the form of a jet that experiences



(a) GLOTTAL OPENING



(b) GLOTTAL CLOSING

Figure 4. Glottal pressures during opening and closing of the glottis with upward propagating mucosal wave. Larger glottal pressures occur for the convergent shape in (a).

little or no aerodynamic pressure recovery upon exit (Ishizaka and Matsudaira, 1972; Scherer, 1981). Within the glottis, however, the pressure varies with glottal area. If we assume Bernoulli's energy conservation law to be applicable (at least approximately), the glottal pressures will rise above zero from top to bottom for a convergent glottis (Figure 4a), whereas the glottal pressures will fall below zero from top to bottom for a divergent glottis (Figure 4b). (Recall that, according to Bernoulli's energy principle and the flow continuity principle, an increase in duct area is accompanied by an increase in duct pressure, and visa versa). This establishes, at once, the driving pressure asymmetry. Since the average pressure in the glottis is greater during opening than during closing, a velocity-dependent driving force is realized, and energy flows from the airstream to the tissue.

An important question arises. Is the tissue capable of vibrating in this mode? Normal mode analyses (Titze and Strong, 1975; Titze, 1976) have shown that an entire series of modes with horizontal and vertical phase differences can exist if the tissue is treated as a viscoelastic medium bounded by rigid structures (e.g. cartilages). In the lowest mode (having the lowest natural frequency), the tissue moves in phase vertically, as in Figure 3. In a second mode, top and bottom move 180 degrees out of phase, and the center is not moving at all. The combination of these two modes, which result in the frequently observed pattern shown in Figure 4, exhibits a phase difference less than 180 degrees, typically about 60-90 degrees (Baer, 1975).

We will show in this paper that the vertical phase difference can also be derived by assuming a simple surface wave to be propagating on the vocal fold from the bottom to the top (Note upward arrows in Figures 4a and 4b). This surface wave can represent a disturbance of respiratory mucus external to the skin, or disturbance of moist, jelly-like mucosal tissue under the skin (or both), to be travelling upward in the glottis at a velocity c. Furthermore, a standing wave could describe a periodic re-distribution of mucus and soft tissue between the lower and upper margins. A millimeter or so into the tissue (laterally), the surface wave would be attenuated sufficiently to treat the remainder of the vocal fold (the muscle) as a single mass.

The results of our mathematical analysis will show that a region of growing oscillation can be defined in a three-dimensional space whose coordinates are mucosal wave velocity $\underline{c}$, the inverse of lung pressure P_L, and a pre-phonatory glottal width ξ_0 (Figure 5a). As in the Ishizaka and Matsudaira (1972) theory, small-amplitude oscillation is assumed around the pre-phonatory width ξ_o. It is clear from the figure that larger driving pressures, lower surface wave velocities, and tighter adduction generally improve the likelihood of oscillation. Lower surface velocities are congruent with larger vertical phase differences, as we will demonstrate in the more quantitative description of the conditions of oscillation.

With regard to vocal tract loading, we will show that the region of oscillation is defined by two primary variables, supraglottal inertance I_2 and lung pressure P_L , as shown in Figure 5(b). Increases of I_2 and P_s above a certain threshold will produce sustained (or growing) oscillation by effectively rendering the damping of the oscillator as zero (or negative). In this case, the vibration characteristics of the source are strongly dependent on the resonator, as will be discussed in greater detail in section III.

II. ANALYTICAL TREATMENT OF VOCAL FOLD OSCILLATION

Our model of the vocal folds will consist of a hybrid between lumped-element and distributed systems. The body of the folds (muscular and deep ligamental layer) will be represented by a single mass-spring-damper system. Superimposed on the motion of this body will be a simplified surface-wave motion in the cover (Figure 6a). This body-cover independance reduces the

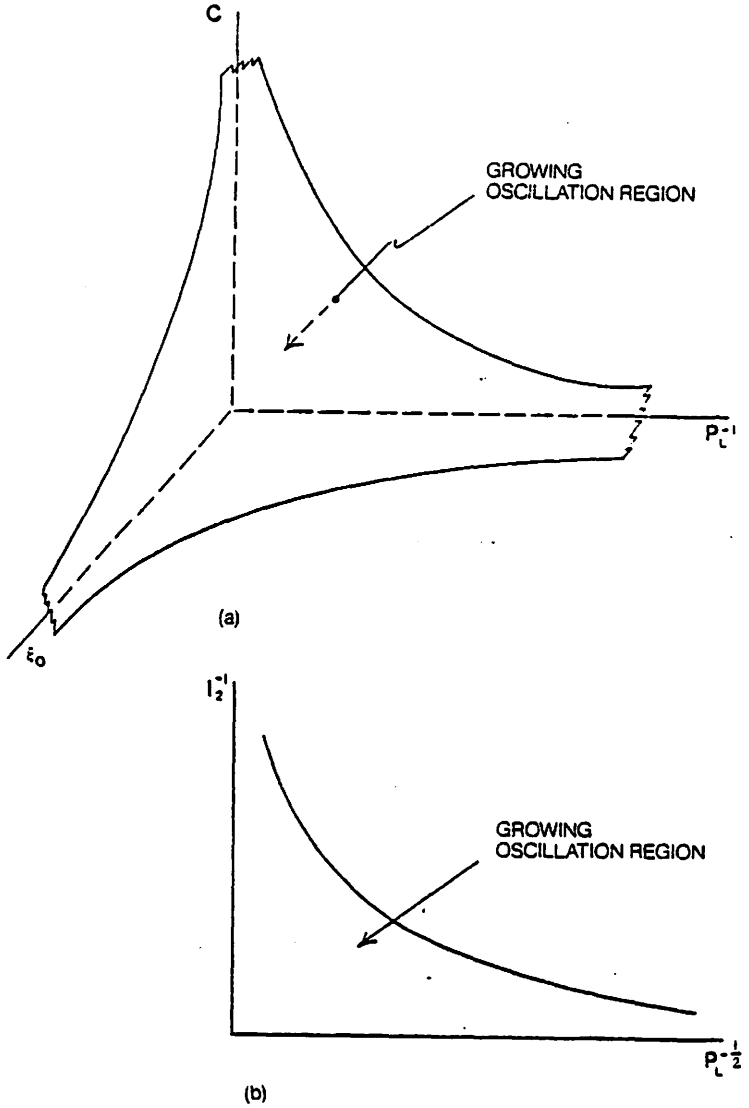


Figure 5. Regions of growing oscillation. (a) Three-dimensional region with parameters c, ξ_0, and P_L. (b) Two-dimensional region with parameters P_L and I_2.

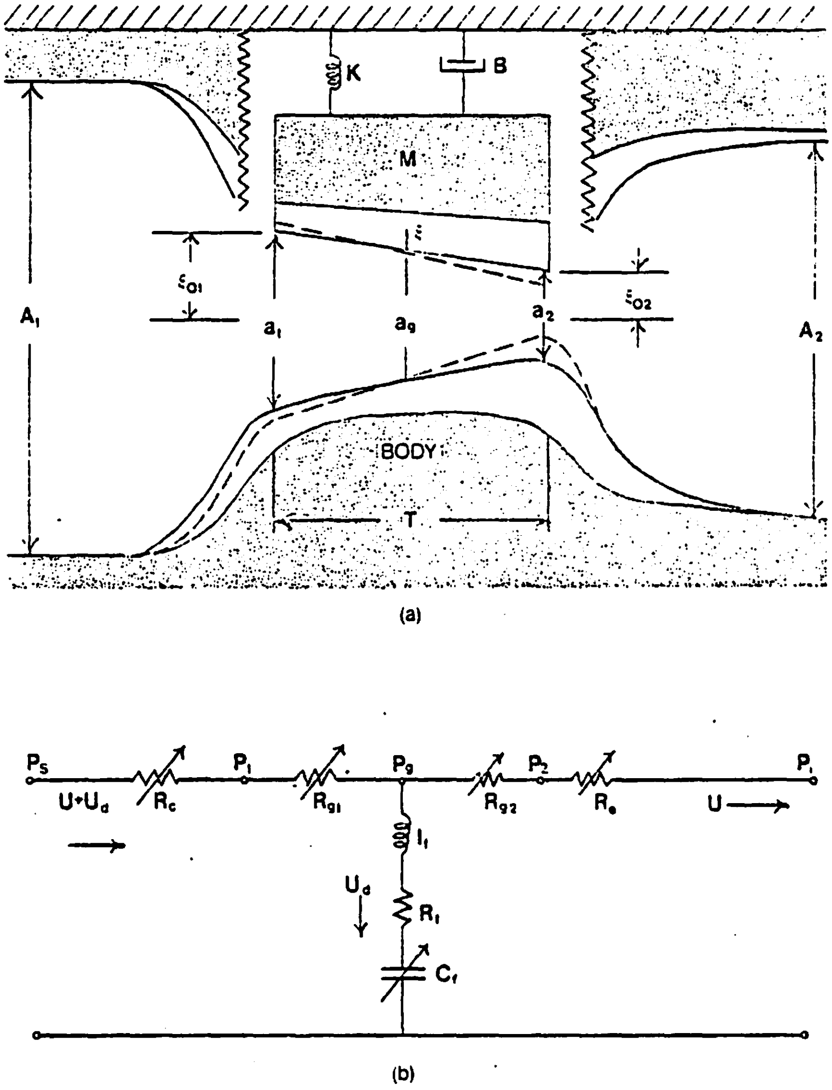


Figure 6. Simplified body-cover model of the vocal folds. (a) Configuration and (b) glottal flow circuit.

complexity of the analysis and allows for somewhat simpler statements about oscillatory conditions than in the Ishizaka and Matsudaira (1972) treatment that couples two masses with three springs. In the later large-amplitude approach, we will allow the mucosal wave velocity to be a function of longitudinal stiffness in the cover. Our purpose here is to demonstrate that, if mucosal waves are assumed, the driving force in equation (1) can be an explicit function of $\dot{\xi}$, and small-amplitude negative damping can be realized. This is a little less ambitious than the approach taken by Ishizaka and Matsudaira (1972), who essentially proved the existance of flow-coupled tissue modes on the basis of fundamental tissue properties.

A. A Linearized Glottal Area Function Based on Surface Waves

Consider a two-dimensional glottis as shown in Figure 6(a). Solid lines and dashed lines show glottal configurations in different portions of the vibratory cycle for small amplitude vibration. A_1 and A_2 are subglottal and supraglottal areas, respectively, and a_1, a_g, and a_2 are specific glottal areas at entry, mid-glottis, and exit. We assume that A_1 and A_2 are constant, but that glottal areas are time- and space-dependent. In particular, we assume that

$$a(z,t) = 2L[\xi_0(z) + \xi_1(z,t)] , \qquad (2)$$

where z is measured from the midpoint of the glottis in the direction of flow, L is the length of the vocal folds (normal to plane of paper), $\xi_0(z)$ is the pre-phonatory glottal half-width, and $\xi_1(z,t)$ is the time-dependent counterpart.

It is known that mucosal surface waves propagate along the glottis in the direction of airflow (Hirano, 1975). In the simplest way, these waves can be described by a one-dimensional wave equation with wave velocity c, such that

$$\partial^2\xi_1/\partial t^2 = c^2\partial^2\xi_1/\partial z^2 \ . \tag{3}$$

This has the general d'Alembert solution

$$\xi_1(z,t) = \xi_1(t - z/c) \ , \tag{4}$$

which can be verified by simple substitution.

It is seen from equation (4) that mucosal wave propagation causes a time-delay in movement from bottom to top of the vocal folds. We will demonstrate that this time-delay helps to provide some of the necessary instabilities for vocal fold oscillation.

In order to reduce the partial differential equation (3) for vocal fold tissue displacement to an ordinary differential equation that will eventually include coupling to the airflow, let us assume a linear z-dependence of the pre-phonatory glottis

$$\xi_0(z) = (\xi_{01} + \xi_{02})/2 - (\xi_{01} - \xi_{02})\ z/T \quad , \tag{5}$$

where ξ_{01} and ξ_{02} are inferior and superior glottal widths, respectively, and T is the vocal fold thickness. With regard to the time-dependent component, we expand the general solution (4) in a Taylor series around the midpoint (z=0) of the glottis and keep only the first order (linear) term,

$$\xi_1(t - z/c) = \xi_1(t,o) - (z/c)(\partial\xi_1/\partial t)|_{t,o} + \ldots \tag{6}$$

$$\cong \xi - z\dot{\xi}/c \quad . \tag{7}$$

The glottal areas at entry and exit can now be defined by letting z = T/2 and introducing a time delay $\tau = T/2c$. Then

$$a_1 = 2L(\xi_{01} + \xi + \tau\dot{\xi}) \tag{8}$$

$$a_2 = 2L(\xi_{02} + \xi - \tau\dot{\xi}) \quad . \tag{9}$$

Any arbitrary area along the glottis can be obtained by combining equations (5) and (7).

Although we will not restrict ourselves to $e^{i\omega t}$ time variation in the bulk of this discussion, it is interesting to relate the time delay τ to a phase delay θ for harmonic motion. In the complex plane, the addition of ξ and $\tau\dot{\xi} = i\omega\tau\xi$ (equation 8) yields a phase angle

$$\theta = \tan^{-1} \omega\tau = \tan^{-1} \omega T/2c \qquad (10)$$

between the center of the glottis and the upper margin. The phase angle between upper and lower margins is twice as large, about 115 degrees for a thickness of 0.5 cm, a 100 Hz frequency, and a 100 cm/s wave velocity (Baer, 1975).

B. Displacement Flow and Equation of Motion

Consider now the lumped-element flow circuit shown in Figure 6(b). A similar circuit was proposed by Conrad (1980). The subglottal pressure is P_s, the glottal entry pressure is P_1, the glottal exit pressure is P_2, and the supraglottal pressure (input pressure to vocal tract) is P_i. The nonlinear flow resistances in the contraction and expansion regions are represented by R_c and R_e, respectively, and the intraglottal resistances are R_{g1} and R_{g2}, also nonlinear. Flow at exit is u and flow at entry is $u + u_d$, where u_d is the displacement flow, resulting from the yielding glottal wall. Except for the mucosal wave motion, the tissue is lumped at the midpoint of the glottis by way of a net vocal fold compliance C_f, an inertance I_f, and a resistance R_f. As will be seen momentarily, these flow circuit elements are directly related to mass, damping, and stiffness of the body of the folds.

Since vocal fold displacement has been linearized along the glottis, the mean displacement flow is at the midpoint and can be written as

$$u_d = 2LT\dot{\xi} \quad . \tag{11}$$

The pressure variation within the glottis is not linear, however. A mean glottal pressure P_g, which serves as the net driving pressure for the body of the tissue, can be computed from an intraglottal pressure function $P(z)$,

$$P_g = \frac{1}{T}\int_{-T/2}^{+T/2} P(z)dz = 2LT(I_f\ddot{\xi} + R_f\dot{\xi} + \xi/c_f) \tag{12}$$

$$= M\ddot{\xi} + B\dot{\xi} + K\xi \quad , \tag{13}$$

where M, B, and K are the mass, damping, and stiffness <u>per unit area</u>. The area in question is the medial surface area LT over which the net driving pressure acts. We will show that the above equation of motion contains the necessary elements for oscillation, as proposed in equation (1). First, however, we focus our attention on the intraglottal pressure variation P(z). Since it is not a linear variation, it is incorrect to call the average pressure $(P_1+ P_2)/2$, nor is it quite correct to divide the total glottal resistance into two equal parts above and below the midpoint, as is done in linear circuit analysis with a T-section. Figure 1(b) therefore serves more for conceptualization of the problem than for exact loop equation analysis.

C. Transglottal Pressure and Mean Driving Pressure

We assume that the results of investigations by van den Berg et. al. (1957), Ishizaka and Matsudaira (1972), Scherer (1981), and Gauffin et. al. (1983) on steady pressure-flow profiles in static models of the glottal airway are applicable to oscillatory flow. Generalizing their results, the transglottal pressure for a non-divergent glottis $(a_1 \geq a_2)$ can be approximated by

$$\boxed{P_s - P_i = (k_c - k_e)(\rho/2)|u + u_d|(u + u_d)/a_2^2 = kP_{k2}}\ , \tag{14}$$

where k_c and k_e are pressure coefficients for contraction and expansion, respectively, $k = k_c - k_e$ is the total transglottal pressure coefficient, ρ is the air density, u_d is the mean displacement flow that adds to a constant flow u, a_2 is the exit (minimum) area, and P_{k2} is the kinetic pressure at exit. Air viscosity in the glottis is neglected here, which is justifiable if the minimum glottal diameter is restricted to be greater than a few tenths of a millimeter. The values of the pressure coefficients have been estimated by the above investigators to be

$$1.0 < k_c < 1.4 \ , \quad k_e = 2(a_2/A_2)(1 - a_2/A_2) \ , \qquad (15)$$

with the lower limit for k_c applying to the highly convergent glottis and the upper limit applying to the rectangular glottis.

The pressure drop for a divergent glottis is more difficult to quantify because flow detachment from the wall may occur prior to glottal exit, making the location of the minimum diameter a variable within the glottis. Fortunately, the divergent glottis ($a_1 < a_2$ in an overall static and dynamic sense) occurs over a smaller portion of the glottal cycle than the convergent glottis in normal male phonation, and may be non-existant in females and higher-pitched phonation in males (i.e., falsetto and mid-register). It is estimated, therefore, that the loss of precision in the transglottal pressure will not affect the basic principles of oscillation described here. Furthermore, it is not clear that flow detachment from the walls will occur the same way in oscillatory flow as in static flow. By including the displacement flow in equation (14), we have made the assumption that the static pressure-flow relationships apply to the oscillatory state. This has yet to be verified. The constant acceleration and deceleration of the air may make the moving medium less turbulent because there is less time for steady flow regimes to be established.

Since all the pressure losses in equation (14) have been attributed here to entry and exit losses, the intraglottal pressures can be derived from Bernoulli's energy conservation principle,

$$P(z) = P_2 + P_{k2}[1-a_2^2/a^2(z)] \quad , \tag{16}$$

where P_2 is the exit pressure that corresponds to the exit area a_2. The average of this intraglottal pressure, as indicated by integration on the left side of equation (12), is

$$P_g = P_2 + P_{k2}\left\{1 - [(a_2^2/T)(-a^{-1})(\partial a/\partial z)^{-1}]_{-T/2}^{+T/2}\right\} \quad . \tag{17}$$

In obtaining the explicit integration above it has been recognized that the glottal area gradiant $\partial a/\partial z = (a_2 - a_1)/T$ is independent of z because of the assumed linear glottis. In the strictest sense, u_d should also have been written as a function of z, but this would have made the integration rather formidable. As a first-order approximation, the use of the mean displacement flow as a constant over z is justifiable because u_d

is usually an order of magnitude less than u. By way of an example, consider a 0.5 cm vocal fold thickness, a 1.0 cm vocal fold length, a 100 Hz fundamental frequency, and an average vibrational amplitude of 0.025 cm (per fold) along the length. Since this amplitude is reached in about a quarter of a cycle (2.5 ms), the linear velocity of the tissue is 10 cm/s and the peak displacement flow is 10 cm^3 /s (equation 11). On the other hand, equation (14) predicts an average steady flow of 200 cm^3/s for an 8 cm H_2O transglottal pressure and a 0.05 cm glottal width ($k_c - k_e = 1.0$). This suggests that $u_d << u$ at low frequencies and when glottal closure is not under consideration.

Upon further evaluation of the limits in equation (17) and substitution of the linear area gradiant $\partial a/\partial z$ into the equation,

$$P_g = P_2 + P_{k2}(1 - a_2/a_1) \quad . \tag{18}$$

Van den Berg et. al. (1957) and Ishizaka and Matsudaira (1972) have shown that the exist pressure P_2 is slightly below the vocal tract input pressure P_1 . In other words, there is a slight pressure recovery in the existing jet, as indicated by $-k_e$ in equation (14). This exit pressure recovery can be written as

$$P_2 - P_i = k_e P_{k2} \quad . \tag{19}$$

If we substitute P_2 from (19) into (18) and replace the exit kinetic P_{k2} by $(P_s - P_i)/k$ from equation (14), we obtain an important relationship between average driving pressure of the vocal folds and transglottal pressure,

$$\boxed{P_g = P_i + (P_s - P_i)(1 - a_2/a_1 - k_e)/k} \quad . \qquad (20)$$

It is often tacitly assumed in phonetic science that the vocal folds are driven by the transglottal pressure. This is true only if the vocal tract input pressure is zero, if the glottis is highly convergent, and if there is no exit pressure recovery. For that trivial case, the transglottal pressure in fact becomes the subglottal pressure. In general, the notion of transglottal pressure driving the vocal folds should be abolished.

Note that the effective driving pressure is a function of glottal geometry, as stated in Section I. During glottal opening a_2/a_1 is smaller than during glottal closing, which sets up the asymmetry in the driving force discussed earlier. As a rectangular glottis is approached, the driving pressure can become zero if the vocal tract input pressure P_i and the exit area a_2 are small. This can happen in the face of a large transglottal pressure. Negative driving pressures can also occur, especially if P_i goes negative (as it does prior to closure when the supraglottal air inertia creates a suction above the folds) or if the glottis diverges (equation not experimentally verified, however).

We will now show how the net driving pressure in equation (20) can be used to establish conditions of oscillation. We begin with the simplest case in which there is no acoustic load on the glottis, i.e., when there is no vocal tract above or below the folds.

D. Oscillation Conditions for No Vocal Tract Coupling

When the subglottal pressure equals the lung pressure ($P_s = P_L$) and the vocal tract input pressure is atmospheric ($P_i=0$), the driving pressure becomes

$$P_g = P_L(1 - a_2/a_1 - k_e)/k \quad . \tag{21}$$

This is an interesting case because it can be simulated with relative ease in the laboratory on excised larynges. It is also interesting because the driving pressure is independent of flow. Only the lung pressure is relevant, due to the clamping of both subglottal and supraglottal pressures. We consider here the typical case where the supraglottal area is large ($a_2 \ll A_2$, or $k_e \cong 0$). Combining equations (8) and (9) with (20) and replacing the left side of equation (12) with the new result yields the equation of motion,

$$P_g = (P_L/k)(\xi_{01} - \xi_{02} + 2\tau\dot{\xi})(\xi_{01} + \xi + \tau\dot{\xi})^{-1} \tag{22}$$

$$= M\ddot{\xi} + B\dot{\xi} + K\xi \quad . \tag{23}$$

Note that the driving pressure has a component that varies directly with tissue velocity. Even though the velocity appears in both the numerator and the demomonator, its presence in the numerator dominate because of the factor of two. It is clear that the driving pressure is greater for positive $\dot{\xi}$ than for negative $\dot{\xi}$, which is in harmony with the more general discussion of Section I.

In order to get simple relationships for oscillatory conditions, we consider small-amplitude vibration by letting

$$\xi = \bar{\xi} + \tilde{\xi} \quad , \tag{24}$$

where $\bar{\xi}$ is an average (DC) displacement and $\tilde{\xi}$ is a small oscillatory component. the DC component satisfies the quadratic equation

$$\bar{P}_g = (P_L/k)(\xi_{01} - \xi_{02})/(\xi_{01} + \bar{\xi}) = K\bar{\xi} \tag{25}$$

which is obtained by letting $\ddot{\xi} = \dot{\xi} = 0$ in equations (22) and (23). Solution, if needed, can be obtained by quadratic formula.

If we now assume that $(\xi_{01}+\bar{\xi})^2 >> (\tilde{\xi} + \tau\dot{\xi})^2$, expand the third factor in equation (22) binomially around $\xi_{01}+\bar{\xi}$, multiply it by the second factor, and retain only first-order (linear) terms, we can show that only constant coefficients of $\tilde{\xi}$ and its derivatives remain in equation (22). This then allows definition of effective small-amplitude damping and stiffness per unit area by combining the coefficients of the resulting linear second order differential equation. The coefficients are

$$B^* = B - (P_L/k)\tau(\xi_{01} + \xi_{02} + 2\bar{\xi})/(\xi_{01} + \bar{\xi})^2 = B - B' \tag{26}$$

$$K^* = K + (P_L/k)(\xi_{01} - \xi_{02})/(\xi_{01} + \bar{\xi})^2 = K + K' \quad . \tag{27}$$

We will refer to B' and K' as aerodynamic damping and stiffness. The mass per unit area of the vocal fold tissue is unaltered (i.e., $M^*=M$).

1. Discussion of Damping Coefficients

It is clear from equation (26) that oscillation will grow (or be sustained) when there is negative (or zero) damping, i.e., when

$$B^* \leq 0 \quad . \tag{28}$$

This condition is enhanced by increasing the lung pressure P_L, by increasing the time delay τ between upper and lower portions of the vocal folds, or by decreasing the glottal opening (note squared term in denominator of B in equation 26). Most importantly, equation (26) predicts that no growing oscillation will occur with zero time delay τ or with infinitely rapid wave propagation in the mucosa (recall that $\tau = T/2c$).

On the other hand, reducing the wave velocity c in the mucosa (presumably by decreasing surface tension of the liquid on the folds or by decreasing the stiffness of the cover) will enhance oscillatory conditions. This prediction agrees with Ishizaka and Matsudaira's (1972) conclusion based on one-mass and two-mass models. Only the two-mass model can approximate the wave motion, whereas the one-mass model is equivalent to the $c \rightarrow \infty$ $(\tau \rightarrow 0)$ condition in our formulation.

Although it appears from equation (26) that the negative aerodynamic damping is directly proportional to P_L, this is not generally the case because the DC displacement $\bar{\xi}$ increases with lung pressure, which in turn reduces the negative resistance. This is illustrated in Figure 7, where the DC portion of the aerodynamic pressure (P_g from equation 22) is plotted together with the DC portion $K\xi$ of equation (23) as a function of ξ. The intersection of the two curves gives the DC displacement $\bar{\xi}$. As the aerodynamic pressure is raised from the solid line to the

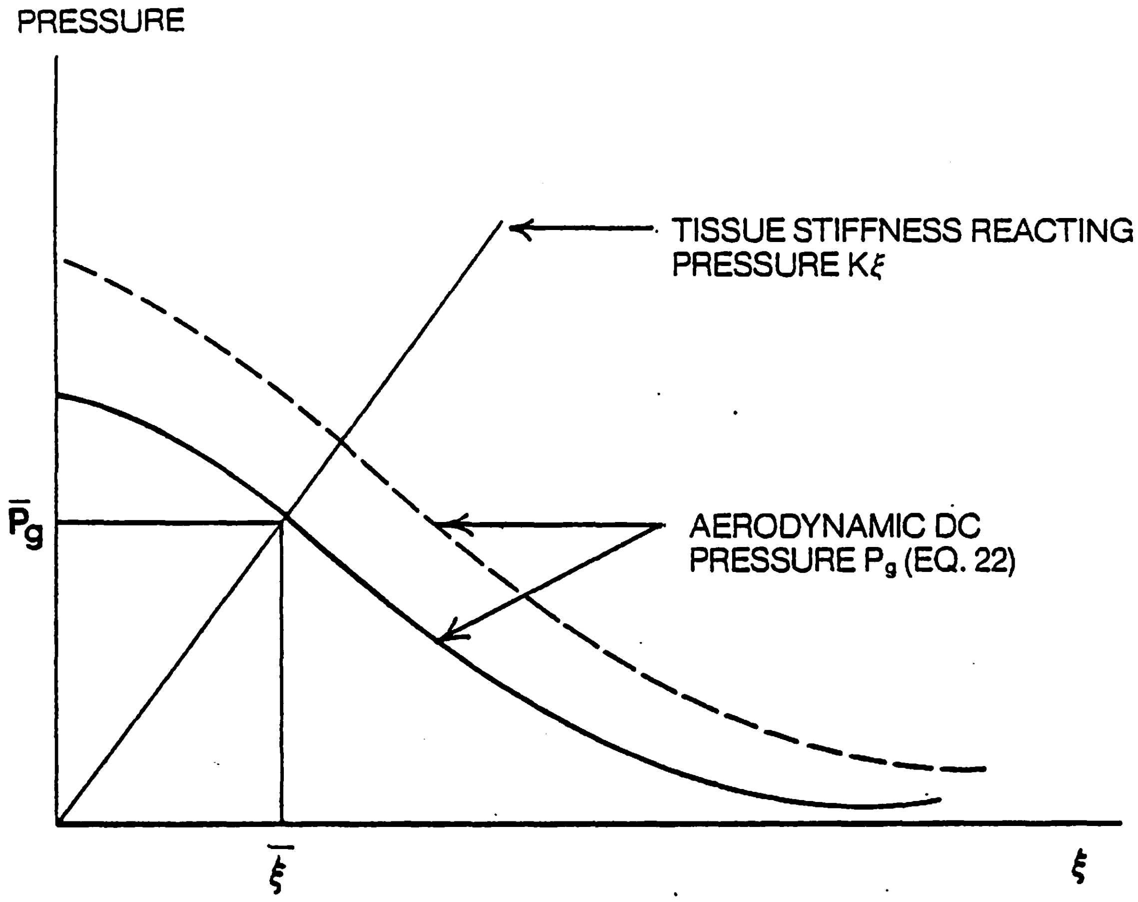


Figure 7. Equilibrium (DC) displacement of $\bar{\xi}$ the vocal folds and equilibrium pressure $\bar{P}_g$.

dashed line (by increasing P_L), the intersection moves to the right. The correspondingly larger value of $\bar{\xi}$ lowers the negative damping. This can further be illustrated mathematically by considering a simpler case. If $\bar{\xi} >> \xi_{01}$, i.e., if the vocal folds are blown apart with a DC displacement that is large in comparison with the inferior (and superior) pre-phonatory glottal half-width, equation (25) has the simple solution

$$\bar{\xi} = [(P_L/k)(\xi_{01} - \xi_{02})/K]^{\frac{1}{2}} \quad , \tag{29}$$

and the effective damping becomes

$$B^* = B - 2\tau[(P_L K/k)/(\xi_{01} - \xi_{02})]^{\frac{1}{2}} \quad (\bar{\xi} >> \xi_{01}) \quad . \tag{30}$$

This situation may be approached in the chest register, where the glottis is nearly square, the vocal folds are nearly touching, and the lung pressure is rather high to blow the folds apart during onset of phonation. Note that the aerodynamic damping is now proportional to the square-root of the lung pressure. The aerodynamic damping is also highly sensitive to glottal convergence ($\xi_{01}-\xi_{02}$) and to tissue stiffness.

At the opposite extreme, where the vocal folds are not blown apart much in relation to the pre-phonatory widths ($\bar{\xi} << \xi_{01}$), equation (26) reduces to

$$B^* = B - (P_L\tau/k)(\xi_{01} + \xi_{02})/\xi_{01}^2 \quad (\bar{\xi} << \xi_{01}) \quad . \tag{31}$$

This would be the case for stiffer folds with a more convergent glottis. Here the negative damping is directly proportional to P_L. We deduce form this that, in general, B' varies with the 0.5 to 1.0 power of subglottal pressure.

Equation (31) also shows that B' is essentially inversely proportional to ξ_{01}, ξ_{02} usually being smaller than ξ_{01}. Thus, abducted vocal folds are less likely to experience self-oscillation than closely approximated folds.

Another case of special interest is a rectangular glottis, where $\xi_{01} = \xi_{02} = \xi_0$. Here the small-amplitude damping assumes the even simpler form

$$B^* = B - (P_L/k)2\tau/\xi_0 \quad (\xi_{01} = \xi_{02} = \xi_0) \quad . \tag{32}$$

For this case $\bar{\xi} = 0$ because there is no DC driving pressure under our idealized Bernoulli conditions within a rectangular glottis (equation 25). This is one of the cases that Ishizaka and Matsudaira (1972) analyzed with a two-mass model. We can harmonized our theory with theirs if we adopt their equation of motion (on page 52 of their monograph) for the lower mass and eliminate their coupling stiffness s_c between upper and lower masses, leaving only the mass m_1, damping r_1, stiffness s_1, and a driving force that is proportional to the difference in displacement between upper and lower masses,

$$m_1\ddot{\xi}_1 + r_1\dot{\xi}_1 + s_1\xi_1 = 2\phi(\xi_1 - \xi_2) \quad . \tag{33}$$

Since in our treatment the coupling tissue stiffness is modeled through the mucosal wave velocity c, a separate s_c need not be included here. Ishizaka and Matsudaira labeled ϕ their aerodynamic coupling stiffness, which for equal upper and lower masses ($m_2 = m_1$) is

$$\phi = L(T/2)(1/2)\rho v^2/(2\xi_0) = L(T/2)(P_L/k)/(2\xi_0) . \tag{34}$$

In their treatment v is the average linear velocity in the square glottis. The last step in the above equation is our own derivation and is consistant with the previous assumptions of $k_e \cong 0$, $P_s = P_L$, and $P_i = 0$ for this particular case (recall equation 14).

By letting $\xi_2 = \xi_1 - 2\tau\dot{\xi}_1$, in a manner similar to the expansion (6) and (7) (but with the exception that the expansion is around the lower displacement), equation (33) yields

$$m_1\ddot{\xi}_1 + (r_1 - 4\tau\phi)\dot{\xi}_1 + s_1\xi_1 = 0 \quad . \tag{35}$$

In order to compare the damping term here with equation (30), we recall that B* is the damping per unit glottal area LT. Since m_1 assumes only half of the glottal area, division of equation (33) by LT/2 gives the Ishizaka and Matsudaira (1972) comparable result

$$B^* = 2r_1/LT - 2\tau(P_L/k)/\xi_0 \quad . \tag{36}$$

Note that the aerodynamic damping is identical to that in equation (32).

One of the conditions for oscillation in the Ishizaka and Matsudaira (1972) formulation was that the coupling stiffness s_c between the masses must be less than the aerodynamic stiffness ϕ . This is equivalent, in our formulation, to saying that the mucosal wave velocity c must be less than $2T\phi/r$. Otherwise, the delay time τ in equation (35) is too small and the aerodynamic resistance $4\tau\phi$ is insufficient to overcome r_1.

Conrad (1980) proposed that vocal fold oscillations are analogous to oscillations in a collapsible tube. These oscillations have been observed in arterial systems, where a downstream flow constriction is created by incomplete clamping of an artery. This condition can be approximated in our formulation by letting $a_2 = 2L\xi_{02}$ = constant, i.e., by clamping the upper lips of the vocal folds at a fixed value. From equation (9) it is clear that $\xi = \tau\dot{\xi}$ in this case, which simplifies equations (22) to (27) considerably. The negative damping term B' becomes $2(P_L/k)\tau\xi_{02}/\xi_{01}^2$, which agrees with equation (32) for the rectangular glottis ($\xi_{01}=\xi_{02}$).

Although the constraining (clamping) of a downstream portion of the glottal duct with yielding walls appears to produce oscillation, it is neither necessary nor a particularly appropriate condition for the vocal folds. Unless the false vocal folds were to be used to create the downstream constriction, nothing anatomical in the larynx exists that would maintain such a constraint. In fact, it is a small matter to demonstrate with an excized larynx that the false vocal folds are not needed to sustain *in vivo* like vibratory patterns. The collapsible tube analogy may be useful, therefore, on a conceptual basis, but offers no quantitative advantage.

The enclosed region of growing oscillation (Figure 5a) can be defined on the basis of equation (32). Consider a three-dimensional contour defined by $B^* = 0$, or

$$\xi_0 c/P_L = (T/Bk) = \text{constant} \;, \qquad (37)$$

where ξ_0, c, and $(1/P_L)$ are the three independent variables (τ has been replaced by T/2c), and thickness, tissue damping, and entry pressure coefficient are parameters. Any point between the origin and the three-dimensional hyperbolic surface is in the growing oscillation region. In the exterior region, oscillation is damped, and since physiologically all three of the variables are positive, the other octants need not be considered. In a follow-up paper we will test the validity of this theory by plotting experimentally determined values of P_L, ξ_0, and c on excised larynges.

2. Discussion of Stiffness Coefficients

As a final topic of discussion in this section, consider the effective stiffness K*. Equation (27) shows a positive aerodynamic stiffness K' for $\xi_{01} > \xi_{02}$, the convergent pre-phonatory glottis. For the square glottis, on the other hand, the natural stiffness is unaffected. This differs from the result obtained by Ishizaka and Matsudaira (1972), who found that the aerodynamic stiffness subtracted from the tissue stiffness of the lower mass. The net result was that the predicted oscillating frequency of the coupled system decreased with increased kinetic pressure (or transglottal pressure), which has been difficult to observe experimentally. Even the subsequent computer simulations (Ishizaka and Flanagan, 1972), which were based on fewer assumptions about pressures in the glottis, did not show this negative trend of frequency with P_L. One oversimplification in their analysis was the choice of zero driving pressure on the upper mass, which might account for the difference. The result in equation (35), based upon their equation of motion (but with s_c elimanated and replaced by our mucosal wave coupling), shows no negative stiffness.

The positive aerodynamic stiffness in equation (27) for a convergent glottis may, in part, be responsible for the observed increase in frequency with transglottal pressure in humans (Ladefoged, 1963) and excized larynges (Baer, 1975). By way of an example, if we choose $k = 1$, $\xi_{02} = \bar{\xi} = 0$ and $\xi_{01} = 0.3$ cm in equation (27) evaluate the fundamental frequency of the oscillator as $[(K + K')/M]^{\frac{1}{2}} / 2\pi$, where vocal fold mass is 0.25 g, tissue stiffness K is 100 g/cm (about 100 Hz natural resonance frequency), and vocal fold surface area is 1.0 cm^2 (1.5 cm x .7 cm), then

$$F_0 = 100(1 + P_L/30)^{\frac{1}{2}} \text{ Hz} \quad . \tag{38}$$

P_L is measured in cm H_2O to correspond with g/cm for the dimensions of stiffness. For pressures below 10 cm H_2O, the slope is slightly less than 2 Hz/(cm H_2O), which is a little on the low side in terms of the 2 - 5 Hz/(cm H_2O) reported in the studies above. We have argued previously, however (Titze, 1980), that nonlinearities in the tissue stiffness are likely to dominate in the relationship between F_o and P_L. We will devote a subsequent study to a more complete analysis of fundamental frequency control.

E. Oscillation Conditions For Lumped Inertive Vocal Tract Coupling

Whenever the fundamental frequency of oscillation is below the first formant frequency of the vocal tract, the reactive portion of the input impedance is inertive (Rothenberg, 1981). To a first order approximation, then, the vocal tract can be represented by a lumped inertance I_2 and a resistance R_2.

The vocal tract input pressure assumes the form

$$P_i = R_2 u + I_2 \dot{u} \quad . \tag{39}$$

To study solely the effects of vocal tract loading on the conditions of oscillation, let us eliminate vertical phase differences in tissue movement by setting $\tau = 0$, or $a_1 = a_2 = a$. This is the converse of eliminating the vocal tract load in Section D to study the effects phase differences in tissue movement. It provides a direct way of studying the oscillatory behavior of a one-mass model.

If we make the simplifying assumptions that the subglottal pressure drop in negligible ($P_s = P_L$) and that the vocal tract input pressure P_i is small in comparison with the subglottal pressure P_s, i.e., that most of the pressure drop is across the glottis rather than the vocal tracts, then the flow is easily computed from equation (14)

$$u = (2P_L/k\rho)^{\frac{1}{2}} a - u_d \quad , \quad P_L >> P_i \ . \tag{40}$$

We will refer to this as the low-load condition. Equation (20) reduces to

$$P_g = P_i - (k_e/k)P_L \quad , \tag{41}$$

where it has been recognized that $1 - a_2/a_1$ in equation (20) is zero for the restricted glottal geometry.

Note that, since (k_e/k) is a small quantity (on the order of 0.1), the two terms on the right side of equation (41) may be comparable, even though P_L is much larger than P_i. Substituting u from (40) into (39), P_i from (39) into (41), and letting $k_e \cong 2a/A$ from (15), the driving pressure becomes

$$P_g = (2P_L/k\rho)^{\frac{1}{2}}(R_2 a + I_2 \dot{a}) - R_2 u_d - I_2 \dot{u}_d - (P_L/k)(2a/A_2) \ . \tag{42}$$

Making the final substitutions for u_d from equation (11) and <u>a</u> from (8) and (9) with $\tau = 0$, the coefficients of ξ, $\dot{\xi}$, and $\ddot{\xi}$ can be combined with those on the right side of equation (13) to obtain the effective mass, damping, and stiffness per unit area,

$$M^* = M + 2LT\,I_2 \tag{43}$$

$$B^* = B - 2LI_2(2P_L/k\rho)^{\frac{1}{2}} + 2LT\,R_2 \tag{44}$$

$$K^* = K - 2LR_2(2P_L/k\rho)^{\frac{1}{2}} + (4L/A_2)(P_L/k) \tag{45}$$

A residual DC pressure

$$P_g^* = 2LR_2\xi_0(2P_L/k\rho)^{\frac{1}{2}} - (4L/A_2)\xi_0(P_L/k) \tag{46}$$

determines the equilibrium position, i.e., the degree to which the vocal folds are blown apart in addition to the pre-phonatory opening ξ_0.

Several observations can be made regarding these small-amplitude constants. First, note that an inertance in the effective mass of the system results from vocal tract inertia I_2. The mass of the air being accelerated adds to the mass of the tissue. This, in conjunction with some changes in effective stiffness, could lower the fundamental frequency appreciably. This "pulling" of the fundamental frequency was demonstrated by Ishizaka and Flanagan (1972) with various lengths of acoustic tubes attached to their two-mass and one-mass simulation models. The one-mass model, which relies solely on supraglottal inertance I_2 to produce negative damping (equation 44), was most susceptible to this frequency "pulling".

The threshold requirements for I_2 to realize sustained or growing oscillations are given in equation (44) and in Figure 5(b). Assuming that R_2 is small (a desired condition since R_2 produces positive damping), we see that the $B^* = 0$ condition results in a hyperbolic relationship between $P_L^{\frac{1}{2}}$ and I_2, all other quantities being treated as fixed parameters. Thus, increases in I_2 or P_L (or both) will enhance oscillation. Graphically, this corresponds to moving closer to the origin in Figure 5(b) if we plot the inverses of $P_L^{\frac{1}{2}}$ and I_2 to maintain a confined region of oscillation. The whole region of oscillation is expanded, of course, by decreasing B, the actual tissue damping.

With regard to tissue stiffness, we note that both a positive and a negative term are associated with this low-load condition. As mentioned, it is desirable to have R_2 small in order to minimize losses. But the third term in equation (45) is not particularly large either, unless A_2 is unusually small. For a typical $A_2 = 3\ \text{cm}^2$, $L = 1.5$ cm, $P_L = 8$ cm H_2O, and $k \cong 1.0$, the stiffness is about 16 Kdyn/cm per cm^2 of vocal fold area. This is about a tenth of the typical lumped tissue stiffness (Ishizaka and Flanagan, 1972). Thus, the low-load condition shows no clear indication of a strong source-vocal tract coupling stiffness, as would be expected.

III. DISCUSSION AND CONCLUSIONS

We have adopted the body-cover concept of vocal fold tissue morphology to capture the essential features of vocal fold vibration. By superimposing a vertically-propagating surface-wave displacement in the cover onto the horizontal diplacement of the body, a single autonomous differential equation described the oscillatory motion. By linearizing this equation for small-amplitide (growing) oscillations, closed-form solutions for effective damping, mass, and stiffness were obtained.

Our results are not markedly different from those obtained by Ishizaka and Matsudaira (1972). The primary mechanism for obtaining velocity-dependent driving forces seems to be time-delayed action, i.e., a delayed reacting pressure on the bottom of the folds due to delayed motion at the top. In our formulation, the delay results from a finite surface-wave velocity, whereas in the Ishizaka and Matsudaira formulation, the two coupled oscillators are locked into a flow-controlled mode that keeps upper and lower masses out of phase, as cogently illustrated by their state diagram (p. 63). It is interesting to point out that delayed action is often responsible for oscillation in feedback control systems.

In the collapsible-tube analogy offered by Conrad (1980), the mechanism is similar, except that the required supraglottal constriction would effectively be a pair of clamped upper lips of the true folds, or alternatively, a firmly positioned pair of false folds. Neither is physiologically very realizable, however.

Net surface-wave propagation in the mucosa (cover) must be from bottom to top. Reversal of this direction would not produce oscillation. Rather, energy would flow from the tissue to the airstream. This is equivalent to the statement that the bottom mass in the coupled oscillator system must always lead the top mass in phase. Negative effective damping will then be produced, which in our study was found to be inversely proportional to the surface-wave propagation velocity, or directly proportional to the time delay between upper and lower portions of the folds.

The delayed action hypothesis can also be used to conceptualize the effect of inertive loading by the vocal tract on the folds. Since the build-up airflow in the glottis and vocal tract is delayed with respect to the glottal area, and since the driving pressure on the folds varies inversely with airflow (Bernoulli's principle), a retarded flow means a greater driving pressure during opening than during closing. This tissue-velocity dependent driving force supplies energy to the vocal folds and helps build up oscillation. We have shown that negative damping is directly proportional to lumped vocal tract inertance. This was also shown by Ishizaka and Matsudaira (1972). Their results demonstrated, furthermore, that a net compliant vocal tract load would squelch oscillation.

We would offer a practical application of our analysis to speech and singing. Unlike most reed or brass instruments, whose source characteristics are strongly influenced by the resonator, the vocal instrument operates with varying degrees of coupling. When coupling is weak, the vocal folds create their own autonomous oscillating system by virtue of mucosal surface waves (or some other non-uniform tissue movement). This situation is approached in speech, where a large variety of articulatory gestures offers many different types of acoustic loads. To preserve constancy of phonation, uncoupling is desirable. The fact that a linear source-filter theory of speech production has been successful for many years attests to this.

When the fundamental frequency of phonation is raised above the normal speech range (as is often done in singing) two effects may take place which cause source-system independence to diminish. First, the vocal fold cover may stiffen, which would raise the mucosal wave velocity and thereby increases effective damping in the folds. Secondly, the reactive load of the vocal tract is greater at higher frequencies, particularly when the fundamental frequency approaches a formant frequency. Both of these factors would make the conditions of oscillation more dependent on the vocal tract, as is easily demonstrated by voice "breaks" that occur when an upward glissando is performed into an acoustic type of unfamiliar length. Some singers, who perhaps have a more "robust" tissue morphology in the larynx, may be able to maintain oscillation autonomous to the larynx, even in the face of vocal tract reactances that are compliant rather than inertive. Little or no vowel modification on specific pitches may then be necessary to "tune-up" the larynx with the vocal tract. On the other hand, those with less "robust" tissue morphology may find it necessary to adjust the vocal tract in specific ways to maintain tissue damping low through inertive loading.

To speculate a bit further, the subject of vocal hygiene may possibly be linked to the mechanical properties of the vocal folds and, in particular, to the viscoelasticity of the mucosa. Viscoelastic properties may change with dehydration, with infectious disease, with trauma, with toxic materials, with temperature, or any number of other internal or external conditions. We are presently investigating the effects of lubrication of vocal fold tissue on the oscillation region described in this paper. Exact comparisons between theory and experiment are more suitable for large-amplitude conditions (limit cycles), and will therefore be a topic of the follow-up paper.

ACKNOWLEDGEMENT

This research was supported by a grant from the National Institutes of Health, grant number NS 16320-05, for which we are grateful.

REFERENCES

Baer, T. (1975), "Investigation of phonation using excised larynxes," Ph.D. dissertation, MIT, Cambridge, Mass.

Broad, D. (1979), "The new theories of vocal fold vibration," In Speech and Language: Advances in Basic Research and Practice, Vol. 2, N. Lass (ed.), Academic Press, 203-256.

Conrad, W. (1980), "A new model of the vocal cords based on the collapsible tube analogy," Med. Res. Eng. 13, No. 2, 7-10.

Gauffin, J., N. Binh, T. Anathapadmanabha, and G. Fant (1983), "Glottal geometry and volume velocity waveform," Vocal Fold Physiology: Contemporary Research and Clinical Issues, D. Bless and J. Abbs (eds.) College Hill Press.

Hirano, M. (1975), "Phonosurgery: basic and clinical investigations," Official Report of the 78th Annual Convention of the Oto-Rhino-Laryngologicl Soc. of Japan.

Ishizaka, K., and M. Matsudaira (1972), "Fluid mechanical considerations of vocal cord vibration," SCRL Monograph 8, Santa Barbara, California.

Ladefoged, P. (1963), "Some physiological parameters in speech," Language and Speech 6, 109-119.

Lieberman, P. (1977), Speech Physiology and Acoustic Phonetics, McMillan, New York.

Rothenberg, M. (1981), "Acoustic interaction between the glottal source and the vocal tract," Vocal Fold Physiology, Stevens and Hirano, (eds.), University of tokyo Press, 304-323.

Scherer, R. (1981), "Laryngeal fluid mechanics: steady flow considerations using static models," Ph.D. Dissertation, University of Iowa.

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

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

Titze, I. (1976), "On the mechanics of vocal fold vibration," J. Acoustic. Soc. Am. 60, No. 6, 1366-1380.

Titze, I. (1980), "Comments on the myoelastic-aerodynamic theory of phonation," J. Speech Hear. Res. 23, No. 3, 495-510.

Titze, I. (1985), "Mechanisms of sustained oscillation of the vocal folds," Vocal Fold Physiology: Biomechanics, Aerodynamics, and Phonatory Control, I. Titze and R. Scherer (eds.), The Denver Center for the Performing Arts.

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

van den Berg, J. (1958), "Myoelastic-aerodynamic theory of voice production," J. Speech Hearing Res. 1, No. 3, 227-243.

SOME TECHNICAL CONSIDERATIONS IN VOICE PERTURBATION MEASUREMENTS

Ingo R. Titze

University of Iowa
and
The Denver Center for the Performing Arts

Yoshiyuki Horii
The University of Colorado, Boulder

and

Ronald C. Scherer
The Denver Center for the Performing Arts

ABSTRACT

Maximum theoretical bit shimmer and sampling jitter are computed and compared with normal vocal shimmer and jitter. 12 bits of resolution and 500 samples per cycle are needed to minimize the contaminating sampling noise without interpolation. Conventioanl tape recorders can add jitter and shimmer equal to that of a normal voice. Low-pass filtering is not harmful, unless peak picking strategies are used and the peaks are severely broadened by the low-pass filter. A window of 20 to 50 cycles should be used, if possible, and multiple tokens of an utterance are necessary to obtain a stable mean.

INTRODUCTION

It is well-known that voiced sounds in speech are not entirely periodic. Even the most serious attempt by a speaker to produce steady phonation with constant pitch, loudness, and quality results in perturbations in fundamental frequency, amplitude, and waveshape of the speech signal as recorded by a microphone. Fundamental frequency perturbations have been termed jitter and amplitude perturbations have been termed shimmer. No special name has yet been given to perturbations in waveform shape or the corresponding spectral variations.

The first major study on fundamental frequency perturbations was reported by Lieberman (1961). The primary focus of the study was on the statistical distribution of the magnitude of adjacent periods. It was noted that this distribution was dependent upon the mode of phonation (e.g. emotional content or linguistic structure), upon the mean fundamental frequency, and upon onset and release of voicing. The measurement of adjacent periods was made oscillographically with a resolution of ± 0.2 ms, the mean period (for six male speakers across all utterances) being about 10 ms (100 Hz). Nearly 50% of the differences between adjacent periods were at the 0.2 ms measurement noise level, however, indicating that better resolution would have been desirable. In a follow-up study, Lieberman (1963) used a larger number of speakers (32, of which 23 had pathologic larynges) and improved the temporal resolution to 0.05 ms by increasing the speed of the oscillographic record camera from 400 in/min to 1200 in/min. He defined a perturbation factor as the percent of the total number of perturbations that are greater than 0.5 ms. Lieberman found that this perturbation factor was roughly proportional to the mean period and tended to be less than 30% in all but a few pathologic cases.

Since then, a host of publications have emerged that have attempted to relate specific or general laryngeal diseases and the perception of vocal disorders to a variety of perturbation measures (Hecker and Kreul, 1971; Hollien, Michel, and Doherty, 1973; Koike, 1969, 1973; Emanuel et al., 1973; Horii, 1975, 1979; Kitajima and Gould, 1976; Crystal et al., 1970; Murry and Doherty, 1977, 1980; Ludlow et al., 1984; Davis, 1976; Hanson et al., 1983; Childers, 1977; Yumoto et al., 1982). Most of the recent investigations have employed digital computers for automatic or semi-automatic extraction of perturbation measures of amplitude, frequency, and additive noise.

It is hypothesized that aperiodicities in steady phonation arise from at least four distinct sources (1) randomness in the action potentials of laryngeal muscles, creating some fluctuations in the muscle forces and configuration of the larynx, (2) randomness in the distribution of mucus on the folds and asymmetries in vocal fold structure, (3) randomness in the flow emerging from the glottis (instability and turbulence), and (4) irregularity in source-vocal tract interactions that stem from non-stationary articulatory configurations. We will refer to these sources of aperiodicity as neurologic, biomechanic, aerodynamic, and acoustic sources, respectively.

It would appear that continued refinement of techniques for perturbation analysis would continue to yield useful information about normal and abnormal (as well as professionally trained) voice production, particularly when specific sources of randomness can be isolated. What is needed at this stage, however, is some better insight into the technical issues related to perturbation measurements. There is very little uniformity among laboratories in the procedures for recording and processing signals and in the definition and extraction of perturbation measures. This article is devoted to some of these issues. Questions we wish to address are (1) What sampling frequency and bit resolution are required for perturbation measurements of a desired accuracy? (2) How do low-pass filtering and tape recording affect the perturbation measures? (3) What can be gained by interpolating between samples? (4) What are the relative advantages of peak-picking and zero-crossing strategies for obtaining periods and amplitudes? (5) How many cycles are needed to make the measurements, and (6) How many tokens of the utterance are needed? We begin with some basic definitions of perturbation measures.

A. DEFINITIONS OF PERTURBATION MEASURES

Consider a_i to be any parameter measured on the i-th cycle. (Amplitude is suggested by the symbol, but it could be period, open quotient, or some other cyclic parameter). Let N be number of cycles in a given segment of phonation, and let $\bar{a}$ represent the mean value within the segment, i.e.,

$$\bar{a} = (1/N) \sum_{i=1}^{N} a_i \tag{1}$$

By letting δ^+ and δ^- represent forward and backward difference operators, respectively, for successive cycles, the following perturbation functions can be defined

$$p_{oi} = a_i - \bar{a} \tag{2}$$

$$p_{1i}^{+} = \delta^{+} p_{oi} = (a_{i+1} - \bar{a}) - (a_i - \bar{a}) = a_{i+1} - a_i \tag{3}$$

$$p_{1i}^{-} = \delta^{-} p_{oi} = (a_i - \bar{a}) - (a_{i-1} - \bar{a}) = a_i - a_{i-1} \tag{4}$$

$$p_{2i} = \tfrac{1}{2}\delta^{+}\delta^{-}p_{oi} = (a_{i+1} + a_{i-1})/2 - a_i \qquad (5)$$

Note that the zeroth-order perturbation function p_{oi} depends upon the mean value, the first-order perturbation functions p^{+}_{1i} and p^{-}_{1i} have the mean value removed, and the second-order perturbation function p_{2i} has the mean value and the linear trend removed. Higher order perturbation functions p_{mi} can, of course, be defined by repeated differencing if quadratic and cubic trends need to be removed also.

A fundamental definition of the effective value of a signal whose strength (power or intensity) is to be assessed is the root-mean-squared (RMS) value. Since the perturbation functions above are ordinary signals, an effective value of the perturbation in a given segment of N cycles can be written as

$$p_m = \left[(N-m)^{-1} \sum_{i=1+m/2}^{N-m/2} p^2_{mi} \right]^{\frac{1}{2}} \qquad (6)$$

for even values of the order m. For odd values of m the replacements i → (i-1/2) for forward differencing and i → (i+1/2) for backward differencing have to be made in the summation to make the formula meaningful.

Note that the zeroth-order RMS perturbation is simply the standard deviation from the mean, which has special statistical significance.

A second measure of the perturbation in a segment of N cycles is the mean-rectified (MR) value

$$P_m = (N-m)^{-1} \sum_{i=1+m/2}^{N-m/2} |p_{mi}| \qquad (7)$$

Again, the same procedure for shifting i by ± 1/2 in the the summation for odd values of m needs to be applied. It is the first-order MR perturbation that has generally been used as the mean jitter or shimmer (e.g., Horii, 1980). The zeroth-order MR perturbation has, to our knowledge, never been employed, but the second-order MR perturbation has been used occasionally when the linear trend needed to be removed (Ludlow, et al., 1983). A systematic study of the relative merits of these perturbations, and their theoretical relationships, is a topic that will be addressed in future investigations.

Perturbation ratios are obtained when the perturbation is normalized with respect to the mean value. This yields the coefficient of variation for the zeroth-order RMS perturbation, and the typical jitter and shimmer ratios for first-order MR perturbations. Usually these ratios are expressed in percent.

This kind of normalization with respect to mean value is not always useful, particularly when the mean value approaches zero. Fortunately this is not the case when fundamental frequency and amplitude are under consideration, but if the perturbation concept were to be extended to other glottal waveform parameters that may have positive and negative values, a zero mean (or near-zero mean) would cause instability in the measure.

In the discussions to follow, the first-order MR perturbation ratio will be used whenever % jitter or % shimmer are specified without further qualification. As stated, in a later publication we will deal with the relationships between various perturbation measures. For example, if there is no long-range trend, there is a predictable relationship between the RMS and the MR values. If there is a trend (linear, quadratic, sinusoidal), some measures may carry more meaningful information than others. In this study, only steady vowels with no intentional trend were considered.

B. SAMPLING FREQUENCY AND BIT RESOLUTION

If sampled data are used, the accuracy of the frequency perturbation measurement is limited by the sampling frequency, and the accuracy of the amplitude perturbation measurement is limited by the bit resolution. Without interpolation between samples and bit levels, the maximum error in the measurement due to digitizing is ± 1/2 bit level and ± 1/2 sample if simple peak detection or zero crossing strategies are employed. The fractional error in amplitude is then

$$\begin{aligned} e &= \pm(\tfrac{1}{2}\text{ bit level})/(\text{bit levels/amplitude}) \\ &= \pm(\tfrac{1}{2}\text{ bit level})(\text{No. amplitudes/full scale})/(\text{No. bit levels/full scale}) \\ &= \pm \tfrac{1}{2} R/2^N \end{aligned} \tag{8}$$

where R is the amplitude reduction factor relative to the full scale A/D converter voltage and N is the number of bits of the converter. Likewise, the fractional error in fundamental frequency (or period) is

$$\begin{aligned} e &= \pm(\tfrac{1}{2}\text{ sample})/(\text{samples/period}) \\ &= \pm(\tfrac{1}{2}\text{ sample})(\text{periods/sec})/(\text{samples/sec}) \\ &= \pm \tfrac{1}{2} F_0/F_s \quad , \end{aligned} \tag{9}$$

where F_0 is the fundamental frequency of phonation and F_s is the sampling frequency.

For most digitized speech waveforms, the number of bit levels per amplitude is greater than the number of samples per cycle, making the amplitude perturbation measurement more accurate than the frequency perturbation measurement. For example, if half of the range of a 12 bit A/D converter is used to record the amplitude (R=2, N=12), then the bit error is ± 2^{-12}

= 1/4096, or ±.025%. On the other hand, if a 200 Hz fundamental frquency is extracted with a 20 KHz sampling frequency, the sampling error is ± 1/200, or ± 0.5%. This example is rather typical and demonstrates the need for interpolation between samples for jitter ratios, which for normal voices can be in the range of 0.1% to 1%. The problem becomes particularly acute when fundamental frequencies as high as 500 Hz are involved. Here the jitter of the voice may be completely masked by the sampling noise.

The theoretical bit noise and sampling noise floors are shown by the dashed lines in Figure 1(a) and (b). Various values of bits/amplitude and samples/cycle are indicated on the abscissas of Figures 1(a) and (b), respectively, and the error in percent is shown on the ordinates. As the figures show, in order to stay below ± 0.1% error in both amplitude and frequency measurements, approximately 9 bits/amplitude and 500 samples/cycle are needed. The former is relatively easy to get, but the latter is often quite difficult. We will now show how actual measurements of jitter and shimmer were affected by these noise floors.

1. Measurements Without Interpolation

Steady-state [a] vowels of 2-3 seconds duration at a nominal pitch of F_3 (175 Hz) were produced by one subject (IT) in

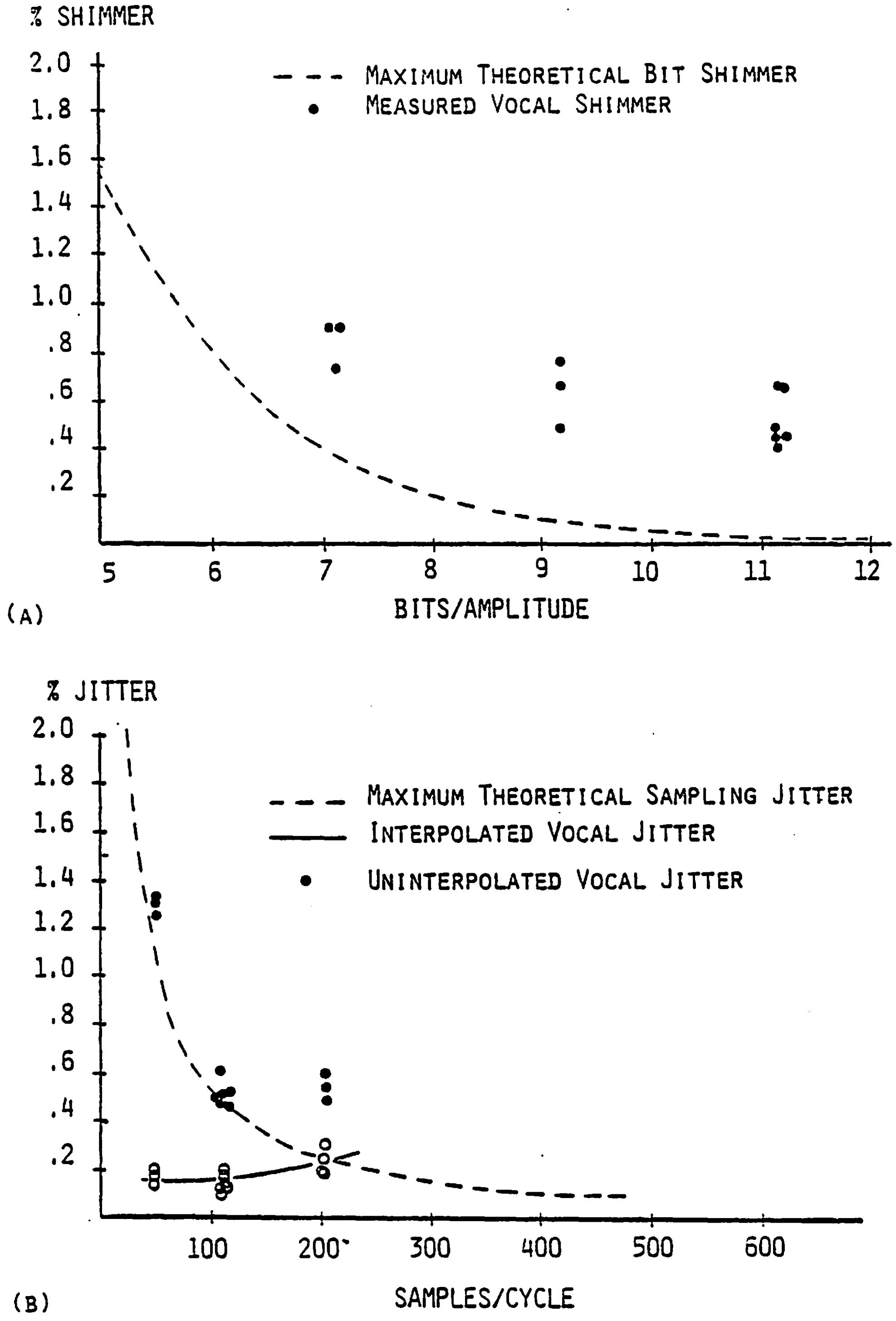


Figure 1. Jitter and shimmer ratio as a function of bits/amplitude and samples/cycle.

a sound-treated room. A bi-directional velocity ribbon microphone (RCA BK-11A) was placed approximately 2 inches from the mouth. An audio amplifier (Fisher CA-660) was used to amplify the signal to ± 5 volts peak to peak (maximum), and a low-pass filter (Wavetek-Rockland 432) was used for anti-aliasing at half of the sampling frequency used in digital recording (LPA-11 processor with PDP 11-44 host computer). A Fourcin Laryngograph signal was recorded on a second channel to assist in fundamental frequency extraction. A center portion of the vowel (.82 seconds) was digitized at 20 KHz/channel sampling frequency. This amounted to 128 computer blocks, or 32,768 samples for each token. The converter had a 12 bit resolution.

Several tokens of the utterance were recorded at three sampling frequencies (10 KHz, 20 KHz, and 35 KHz, which corresponded to 57, 114, and 200 samples/cycle, respectively) and again for three amplitude reduction factors, that corresponded to approximately 7, 9, and 11 bits/amplitude. The subject made every attempt to keep the phonatory adjustments the same across tokens, such that only the bit resolution and the sampling frequencies should be significantly different. Period and peak-to-peak amplitude extractions were performed by a software package called GLIMPES (Titze, 1984). For this particular application, a peak-picking strategy was employed without any interpolation. The results are shown by filled circles in Figure 1.

It is clear that jitter and shimmer in normal voice production can be low enough to be affected by sampling and bit noise. Only for $N \geq 12$ does the bit noise become essentially negligible and the sampling noise becomes negligible only when 500 or more samples per cycle are used. At the usual 100-200 samples per cycle, the sampling jitter is comparable to the vocal jitter, and at 50 samples per cycle, the sampling jitter completely masks the vocal jitter. Similar results were reported by Horii (1979), who suggested that a 40 KHz sampling frequency (200 samples per cycle at 200 Hz) would provide reasonable fidelity.

In some systems, high sampling rates present no problems for analog to digital converters and for digital storage media. But as more and more channels are being processed simultaneously in voice analysis, and micro-computer systems are being employed for data aquisition, the maximum data rate of the system can easily be exceeded. This suggests that some interpolation schemes may be advantageous. These will now be discussed.

2. Measurements with Interpolations Between Samples

The locations of zero crossings and extrema can be estimated between two samples separated in time by Δt. Consider Figure 2(a). For positive-going zero crossings that are located

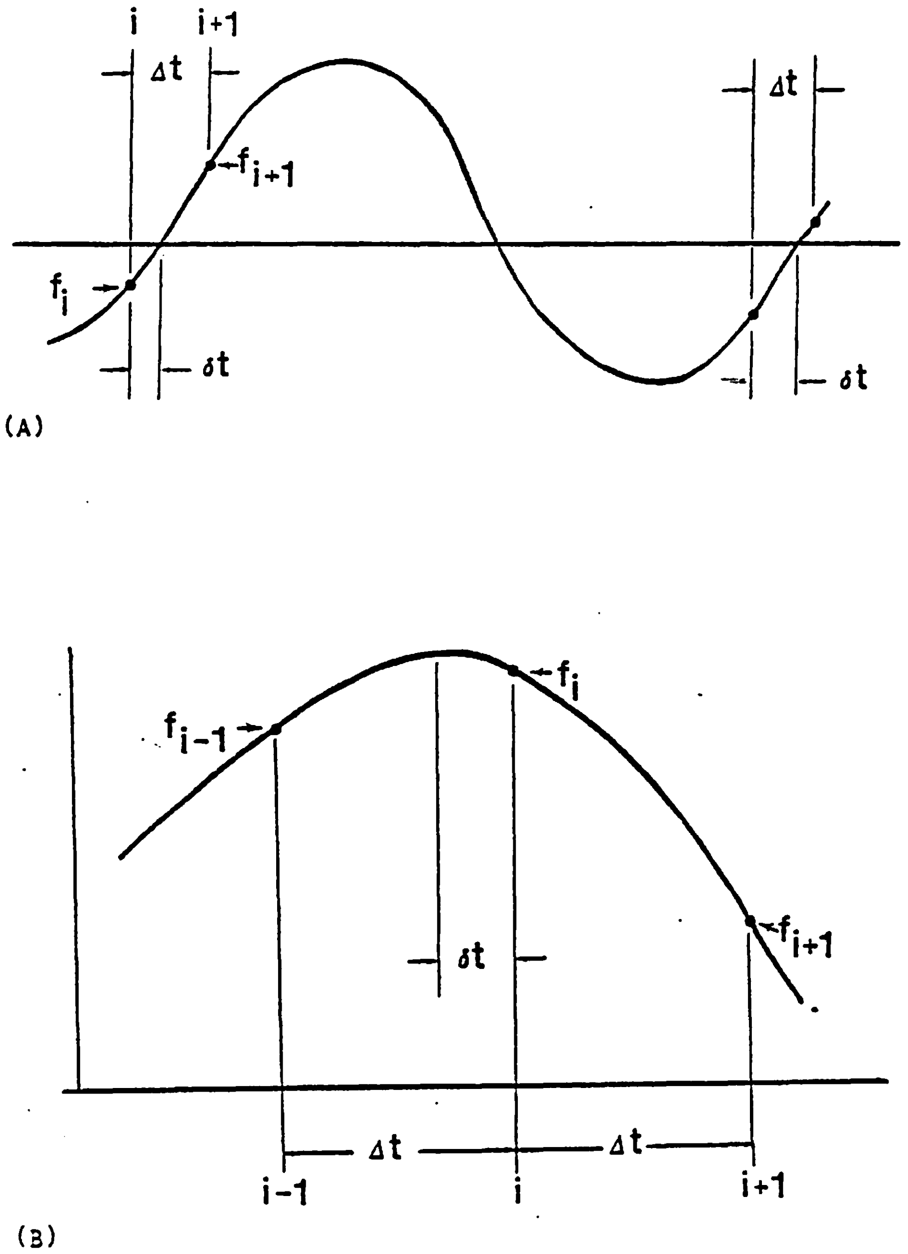


Figure 2. Interpolation of (a) zero crossings and (b) extrema.

between the ith sample and the i+1 sample, linear interpolation yields a fractional sampling interval δt as a correction to the location $i\Delta t$ of the sample just below the axis,

$$(\delta t/\Delta t) = -f_i/(f_{i+1} - f_i) \quad , \tag{10}$$

or

$$\delta t = -f_i/f_i' \quad . \tag{11}$$

This fractional sampling interval can be computed at two successive zero crossings, with the difference being the net correction to the period obtained without interpolation.

At the extrema, a second-order interpolation between three points is needed, as illustrated in Figure 2(b). If the uninterpolated extremum is located at sample i, the correction is

$$\delta t = -f_i'/f_i'' \quad , \tag{12}$$

where central derivatives

$$f_i' = (f_{i+1} - f_{i-1})/(2\Delta t) \tag{13}$$

$$f_i'' = (f_{i+1} - 2f_i + f_{i-1})/(\Delta t)^2 \tag{14}$$

are employed. Again, if the interpolation is used to correct fundamental period extraction via peak-picking or trough-picking algorithms, the difference between δt's at successive extrema is the net correction to the period obtained without interpolation.

As pointed out in the previous section, time resolution is more critical than amplitude resolution. Hence, if the locations of the maximum and minimum are accurately determined, the maximum and minimum values themselves need not be interpolated if 12 or more bits are used (recall Figure 1a).

In order to test the accuracy of the fundamental frequency extraction with interpolation, 20 cycles of a 100 Hz sinusoid were "jittered" by randomly choosing periods that varied between 198 and 202 samples (20 KHz sampling frequency). The peak-picking paradigm was then used to extract the frequencies. (It will be shown in the next section that peak-picking is less accurate for simusoids than zero-crossing.) This technique should have uncovered the greatest errors, therefore, and should serve as a good calibration. The results are shown in Table I.

TABLE I - JITTER CALIBRATION

FREQUENCY	EXTRACTED FREQ.	DIFFERENCE
100.0000	100.000	--
99.5025	99.502	--
100.5025	100.553	+.05
101.0101	100.908	-.10
99.0099	99.059	+.05
100.0000	100.000	--
100.5025	100.503	--
99.0099	99.010	--
99.5025	99.502	--
100.5025	100.553	+.05
101.0101	100.908	-.10
100.5025	100.553	+.05
100.0000	100.000	--
100.0000	100.000	--
99.0099	99.010	--
99.5025	99.502	--
100.5025	100.503	--
100.0000	100.000	--

Comparing the errors with the 200 sample per cycle value of the dashed line in Figure 1(b), we see that the maximum theoretical error of ±.25% is never obtained when interpolation is used. In only two cycles is the error as large as 0.1%, and we expect that for non-sinusoidal speech signals (where the peaks are more localized) the accuracy will be even better.

The solid line and open circles in Figure 1(b) shows vocal jitter for a number of tokens of the previously described steady vowels when interpolation is used in a peak-and-trough-picking paradigm. Two things are immediately clear, (1) normal vocal jitter can be as low as about 0.1% and (2) the quantization errors in digital recording do not pose a serious limitation if the proper interpolation strategies are employed. As few as 50 samples per cycle were used here to resolve a jitter ratio in the 0.1% range. This would correspond to a 10 KHz sampling frequency at a fundamental frequency of 200 Hz. Without interpolation, this value of vocal jitter would be buried in the 1.2% sampling jitter.

Interpolation will be included as a nominal condition in the remainder of this discussion, which will now focus on the effects of tape recording and low-pass filtering prior to perturbation analysis.

C. TAPE RECORDING AND LOW-PASS FILTERING

It is common practice to record speech signals on conventional analog tape recorders prior to analysis. The noise introduced by these tape recorders can increase the shimmer and wow and flutter (variations in the speed of the tape) can increase the jitter. We purposely chose a tape recorder that had rather ordinary (non-professional) characteristics (Akai Gx 4000D reel-to-reel). Signals were recorded at about 0 dB input level (VU meter) and played back for digital recording as described previously. The results are shown in Figure 3. Both jitter and shimmer are nearly doubled when the data are taken from tape recordings.

The amount of degradation of the signal will, of course, depend upon the quality of the tape recording. Needless to say, high fidelity recording is essential, and we suspect that room conditions may also affect the measurement, particularly if the microphone is not located close to the mouth.

The speech signal is often low-pass filtered in order to facilitate fundamental frequency detection, particularly if zero-crossing strategies are employed. We investigated the effects of low-pass filtering in a systematic way by gradually lowering the cut-off frequency of the anti-aliasing filter

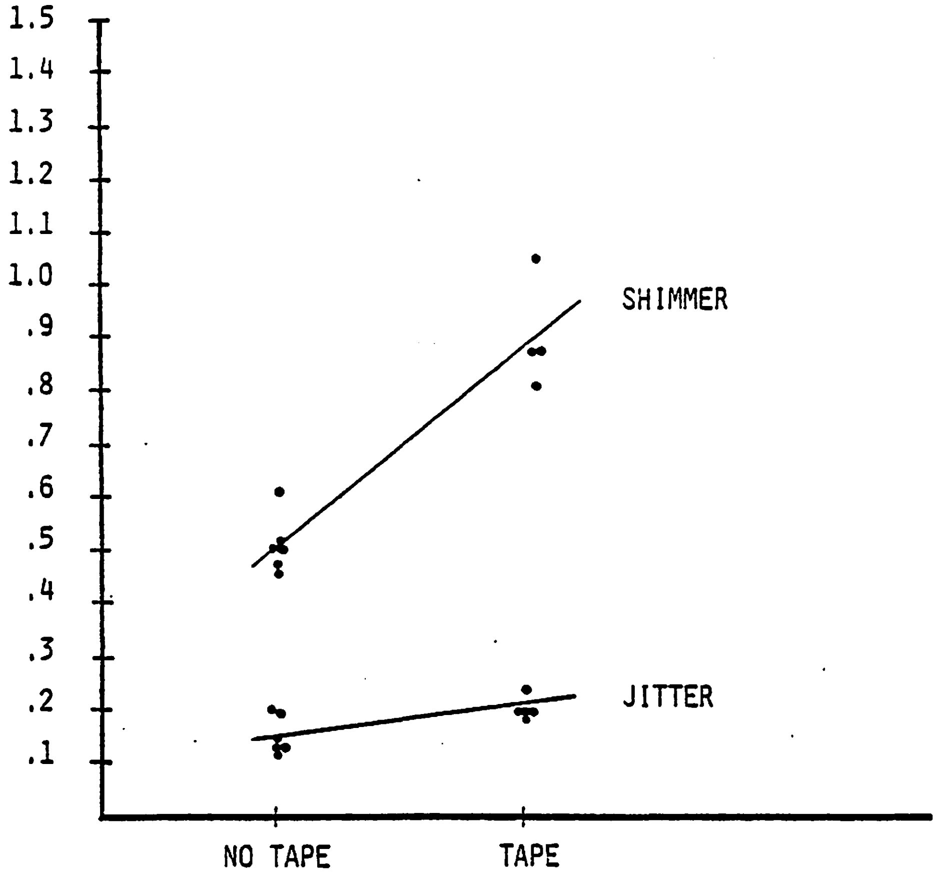


Figure 3. Jitter and shimmer ratio as a function of tape recording.

(Wavetek-Rockland 432) from 10 KHz to 200 Hz, slightly above the 175 Hz fundamental frequency. The sampling frequency was kept at 20 KHz and the signal was recorded directly into the computer (no tape recording).

The results are shown in Figure 4. Consider first the shimmer in Figure 4(a). Although some sizable fluctuations occur in the region from 2 KHz to 10 KHz, no definite trend is observed until the cutoff-frequency is below 2 KHz. In this lower cutoff-frequency region, the waveform peaks are smoothed considerably, resulting in a decrease in shimmer from about 0.55% to about 0.35%.

The jitter values are preserved over the entire 200 Hz to 10 KHz low-pass region if a switch in the fundamental frequency extraction scheme is made in the 200 Hz to 500 Hz region. As Figure 4(b) shows, peak picking is very error-prone below 500 Hz because severely rounded peaks do not allow for accurate determination of the extrema locations, even when interpolation is used. The slightest amount of noise in a broad peak can offset the location of the maximum by several samples, causing large jitter ratios (above 1.0%). Switching to a zero-crossing strategy when the waveforms become nearly sinusoidal can remedy the situation (Figure 4b).

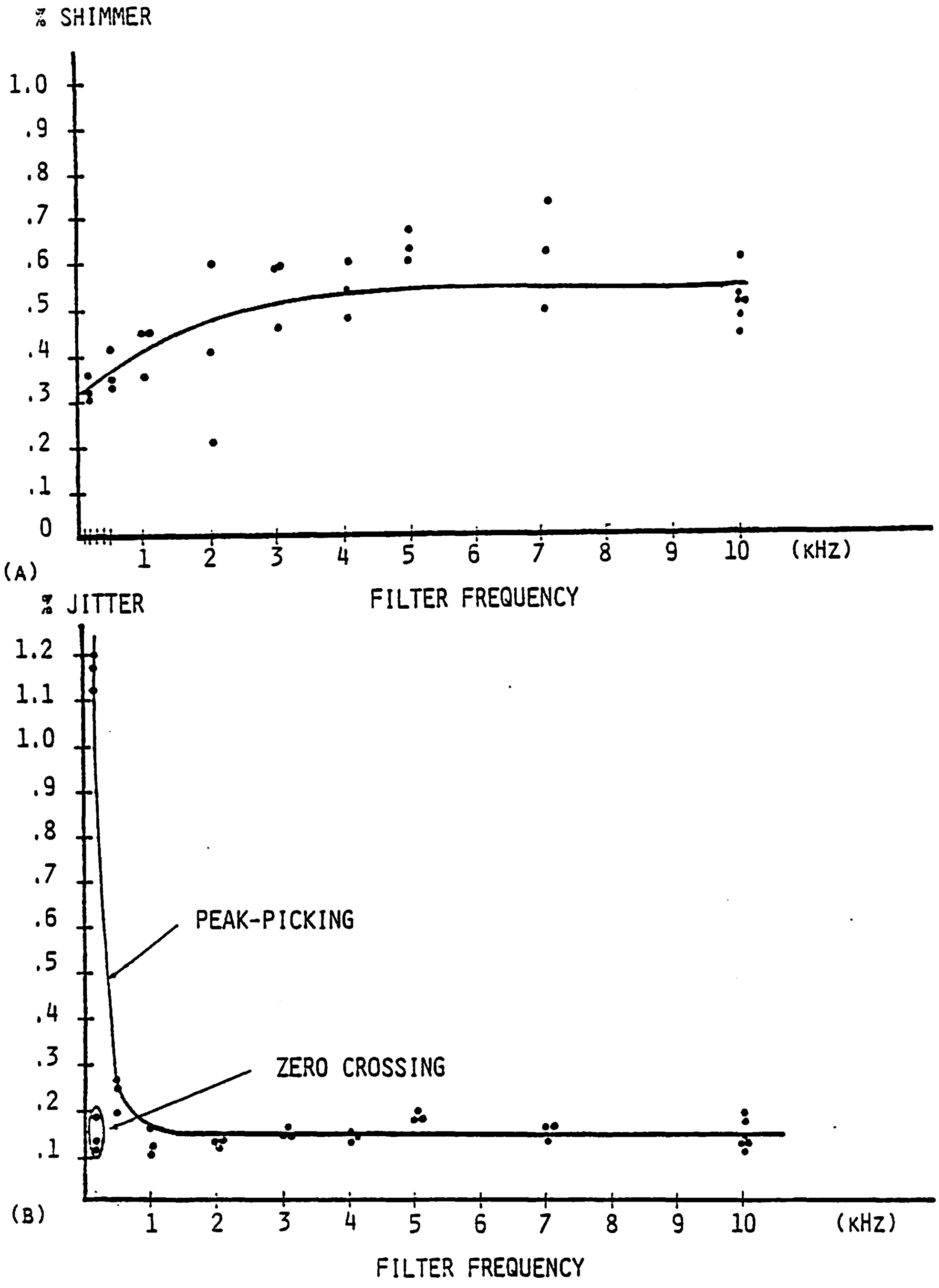


Figure 4. Jitter and shimmer ratio as a function of low-pass filter cutoff frequency.

It is also possible, however, to retain the peak-picking strategy if the peaks are first "cleaned-up". By computing a running sum (average) of N samples around the peak or trough, small amounts of additive noise can be removed and the true extremum can be established with greater accuracy. This is illustrated in Figure 5, where jitter ratios are computed for the same three tokens of a 200 Hz low-pass filtered vowel with different running sums. Note that by taking 11 samples in the sum, the jitter ratio is almost restored to its normal value.

It must be emphasized that these "cleaning-up" approaches should not be relied upon heavily. A ligitimate perturbation can easily be discarded in the process. We prefer to use two signals for perturbation measurements, one with predominantly low frequency content that allows us to mark the periods in a gross fashion (such as an electroglottographic signal, a neck accelerometer signal, or a low-pass filtered speech signal), and another unfiltered signal that can be used for accurate fundamental frequency extraction if the periods have been "roughed out".

All of the data presented thus far show that there is considerable token-to-token variability. We will now address the question of how many tokens are needed and how many cycles should be included within a token.

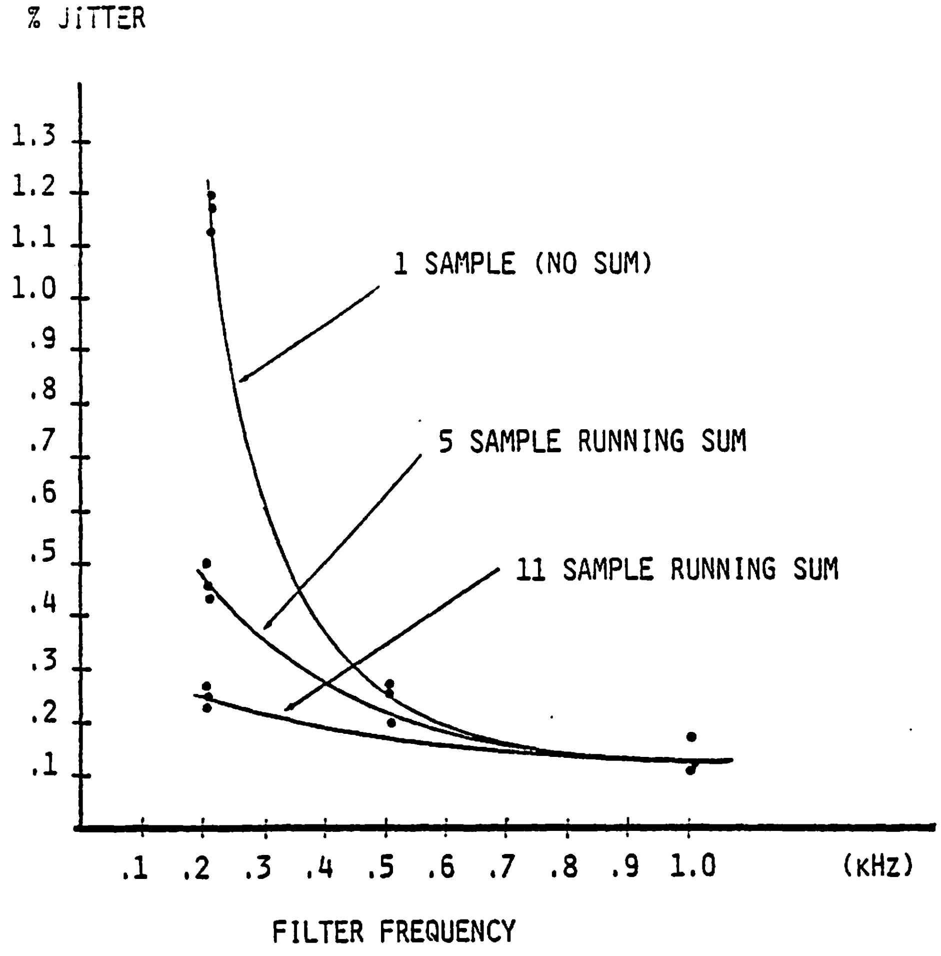


Figure 5. Jitter ratio as a function of low-pass filter cutoff frequency when peak-picking is modified by a running sum.

D. NUMBER OF CYCLES AND NUMBER OF TOKENS

An obvious consideration in the computation of perturbation factors is the number of cycles N (equations 1, 6, and 7) that are necessary for a meaningful average. In the data reported thus far, the entire length of the recorded utterance was used to compute the jitter and shimmer ratios. At 0.82 seconds and 175 cycles per second, this amounted to about 140 cycles (a few are discarded by GLIMPES at the beginning and end).

In order to determine the variability of the perturbation factors across shorter segments within a single token utterance, random groups of 10 cycles, 50 cycles, and 100 cycles were selected within the total 140 cycle utterance. The corresponding jitter and shimmer ratios are shown in Figure 6. The general trend is an increase in the variability of the perturbation ratios for smaller N, as expected. As a rule, 50 to 100 cycles are adequate to obtain a stable measure for steady vowels. For non-steady vowels, of course, smaller windows ($n < 20$) may purposely be selected to observe trends, but it must be remembered that a greater margin of error accompanies these measures.

Consider now a number of repeated tokens of the same utterance. Figure 7 shows histograms for jitter and shimmer

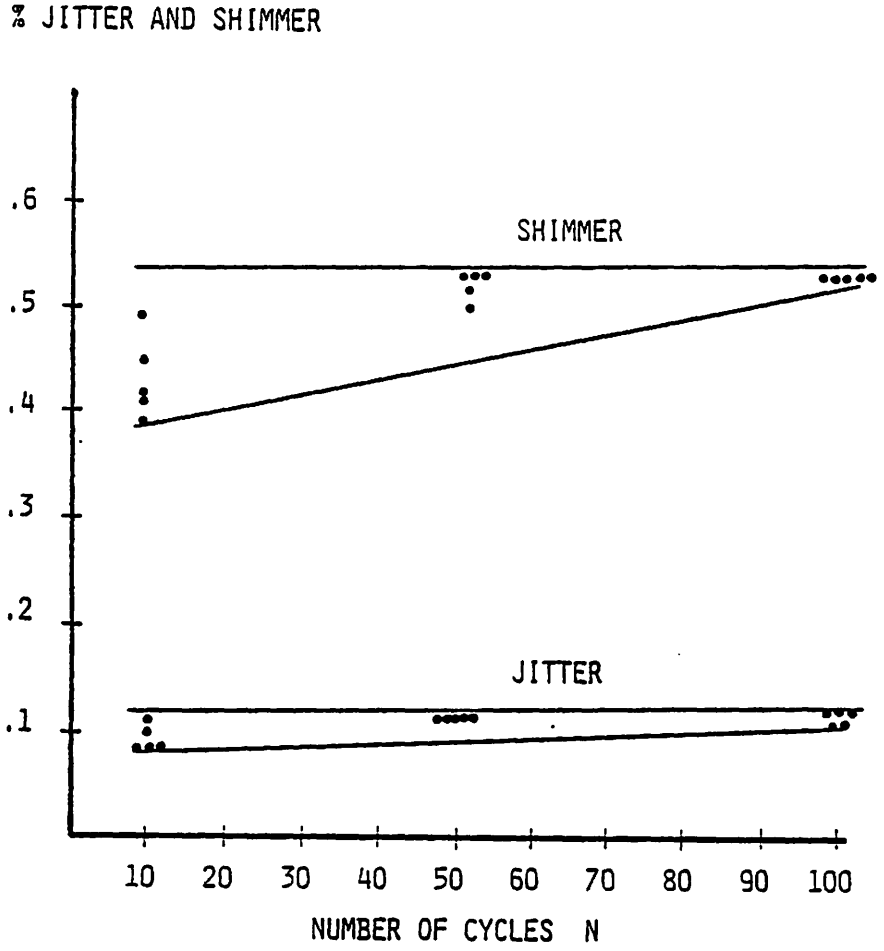


Figure 6. Jitter and shimmer as a function of the number of cycles in a given analysis window.

ratios obtained from 24 tokens. The mean shimmer ratio was 0.53%, with a standard deviation of 0.11% [Figure 7(a)]. Similarly, the mean jitter ratio was 0.15%, with a standard deviation of 0.024% [Figure 7(b)]. It is clear that repeated tokens are necessary if reliable jitter and shimmer ratios are to be obtained for specific voicing conditions.

CONCLUSION

We have raised some questions about the validity of perturbation measurements when insufficient attention is paid to technical and theoretical limitations. Interpolation between samples is necessary if jitter of normal voices is to be extracted with less than about 500 samples per cycle. 12 or more bits per amplitude is sufficient to extract normal shimmer. High quality tape recorders are needed to maintain normal perturbations (0.1 to 0.5%). Low pass filtering is acceptable, but special care must be taken when peak-picking strategies are used on sinusoidal or near-sinusoidal signals. A single token of a steady vowel is insufficient to establish a reliable perturbation measure. Rather, the usual statistical representation is needed. Within a given token, a larger number of cycles result in a more stable measure for steady phonation, but a smaller number of cycles may be desired to observe trends.

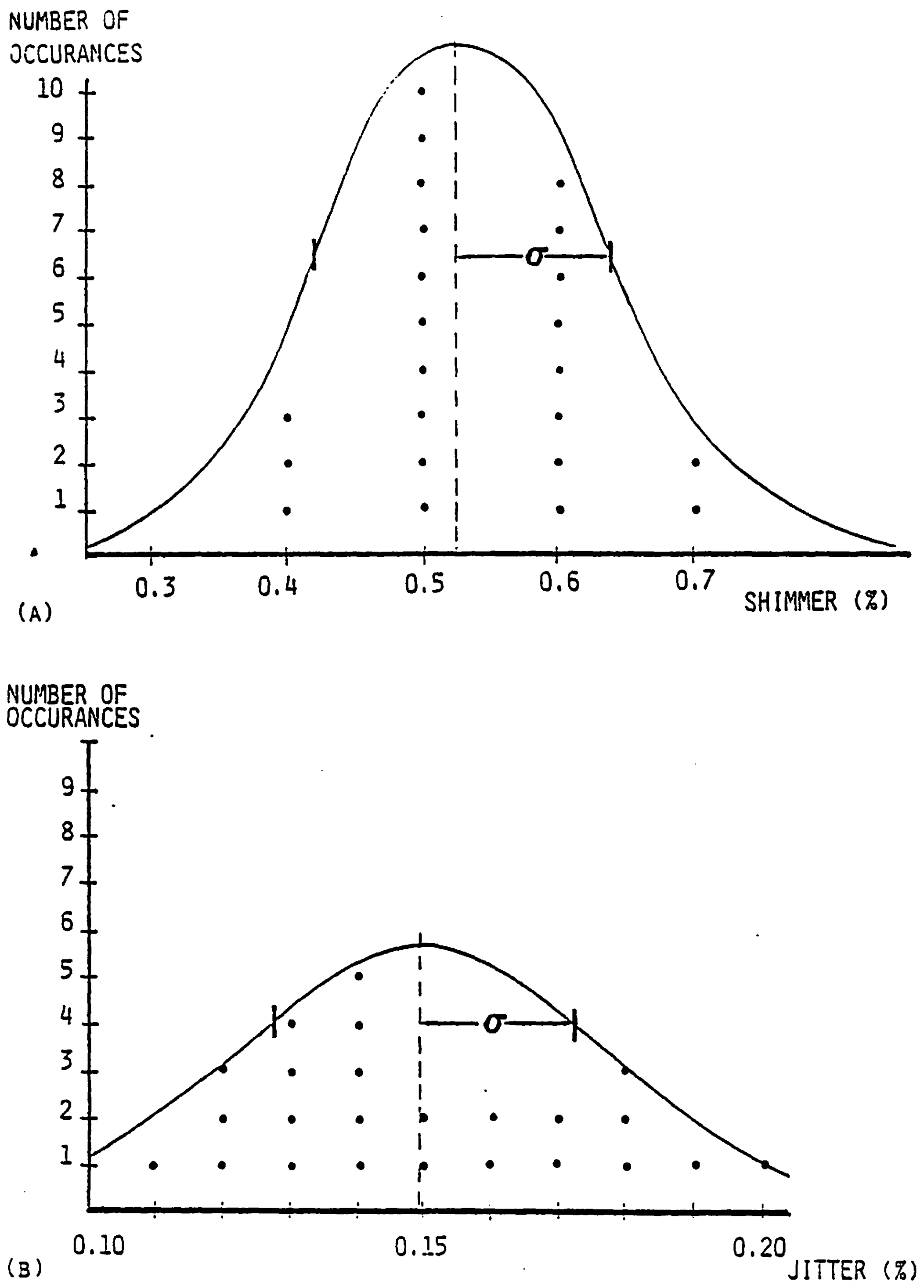


Figure 7. Histograms of jitter and shimmer for 24 tokens of the same utterance.

ACKNOWLEDGEMENT

This work has been supported by the National Institutes of Health, grant number R01 NS16320-05.

REFERENCES

Childers, D.G. (1977), "Laryngeal pathology detection," CRC Crit. Rev. Bioeng. 2, 375-425.

Crystal, T.H., W.W. Montgomery, C.L. Jackson and N. Johnson (1970), "Methodology and results on laryngeal disorder detection through speech analysis," Final report, Contract PH-86-68-192, Signatron, Inc., Lexington, Mass.

Davis, S.B. (1976), "Computer evaluation of laryngeal pathology based on inverse filtering of speech," SCRL Monograph No. 13, Speech Communications Research Laboratory, Inc.

Emanuel, F.W., M.A. Lively and J.F. McCoy (1973), "Spectral noise levels and roughness ratings for vowels produced by males and females," Folia phoniat. 25, 110-120.

Hanson, D.G., B.R. Gerratt and P.H. Ward (1983), "Glottographic measurement of vocal dysfunction, a preliminary report," Ann. Otol. 92, No. 5, 413-420.

Hecker, M.H.L. and E.J. Kreul (1971), "Descriptions of the speech of patients with cancer of the vocal folds. Part I: measures of fundamental frequency," J. Acoust. Soc. Am. 49, No. 4, 1275-1282.

Hollien, H., J. Michel, E.T. Doherty (1973), "A method for analyzing vocal jitter in sustained phonation," J. of Phonetics, 1, 85-91.

Horii, Y. (1975), "Some statistical characteristics of voice fundamental frequency," J. Speech Hear. Res. 18, 192-201.

Horii, Y. (1980), "Vocal shimmer in sustained phonation," J. Speech Hear. Res. 23, 202-209.

Kitajima, K. and W.J. Gould (1976), "Vocal shimmer in sustained phonation of normal and pathologic voice," Ann. Otol., 85, 377-381.

Koike, Y. (1969), "Vowel amplitude modulations in patients with laryngeal diseases," J. Acoust. Soc. Am. 45, 839-844.

Koike, Y. (1973), "Application of some acoustic measures for the evaluation of laryngeal dysfunction," Studia Phonologica, 1, 17-23.

Lieberman, P. (1961), "Perturbations in vocal pitch," J. Acoust. Soc. Am. 33, No. 5, 597-603.

Lieberman, P. (1963), "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges," J. Acoust. Soc. Am. 35, No. 3, 344-353.

Ludlow, C.L., D.C. Coulter and F.H. Gentges (1983), "The effects of change in vocal fold morphology on phonation," In V.L. Lawrence (ed.), Transcripts of the Eleventh Symposium Care of the Professional Voice, Part 1: Scientific Sessions: Papers, The Voice Foundation, New York, 77-89.

Ludlow, C.L., C.J. Bassick, Y.J. Lee, N.P. Conner, and D.C. Coulter. (1984), "The validity of using phonatory jitter to detect laryngeal pathology," J. Acoust. Soc. Am. 75, Suppl. 1, 58.

Murry, T. and E.T. Doherty (1977), "Frequency perturbation and duration characteristics of pathological and normal speakers," J. Acoust. Soc. Am. 62, 55.

Murry, T. and E.T. Doherty (1980), "Selected acoustic characteristics of pathologic and normal speakers," J. Speech Hear. Res. 23, 361-369.

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

Yumoto, E., W.J. Gould and T. Baer (1982), "Harmonic-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. 71, 1544-1550.

VOCAL EFFICIENCY AND AERODYNAMIC ASPECTS IN VOICE DISORDERS

Shinzo Tanaka
Kochi Medical School,
Kochi, Japan

Wilbur J. Gould
Vocal Dynamics Laboratory, Lenox Hill
Hospital, New York, and
The Recording and Research Center,
The Denver Center for the Performing Arts

ABSTRACT

The variation of vocal efficiency for the mean values for normal subjects was investigated in patients with laryngeal disease. The relative contributions of mean flow rate and intrapulmonic pressure to the variation of efficiency were explored to explain aerodynamic aspects in voice disorders. Vocal efficiency was determined by use of expressions involving simultaneous values of sound pressure level, mean flow rate, and intrapulmonic pressure. The intrapulmonic pressure was noninvasively obtained by plethysmographic and pneumotachographic methods. Values of vocal efficiency were generally abnormally low for the types of larynges studied. An aerodynamic-biomechanic classification of laryngeal disease was inferred from the data: 1) large chink of glottis, associated with high flow rate; 2) mass on vocal fold, associated with high values of both flow rate and intrapulmonic pressure; and 3) high stiffness of vocal fold, associated with high values of intrapulmonic pressure.

This paper has now been published in the Annals of Otology, Rhinology and Laryngology, Volume 94, January-February, 1985, p. 29-33.

INTRODUCTION

Usually vocal sound is generated at the glottis and radiated from the mouth opening after being transformed through the vocal tract. Vocal efficiency is defined as the ratio of sound power at the mouth opening (or the opening end of a tube if inserted in the mouth) to aerodynamic power calculated as mean flow rate times mean subglottal or intrapulmonic pressure. (1)

Vocal efficiency has been studied by a number of researchers since van den Berg measured it for one subject. Ladefoged and McKinney (2) reported differences of efficiency for different vowels. Isshiki (3) recommended the use of a mouthpiece to reduce the influence of mouth opening on vocal efficiency. Schutte and van den Berg (4) demonstrated the improvement in efficiency after treatment in one patient. However, there appears to be no study which reports measures of efficiency for a number of patients with voice disorders, probably because of difficulties in measuring subglottal or intrapulmonic pressure.

In general, subglottal pressure has been measured by means of an intraesophageal balloon, (1,2.4) a transdermal needle, (3) or a miniaturized pressure transducer. (5) In spite of their simplicity, these methods were too invasive to be used easily for clinical examination of patients with voice disorders.

A non-invasive method for measuring intrapulmonic pressure using body plethysmography has been described by Hixon. (6) With this method, the authors have studied the relationships between vocal intensity and aerodynamic parameters in ten normal subjects. (7) Using these results as normal values to be compared with those of patients, vocal efficiency in various laryngeal disease was investigated and is reported in this study. For each type of lesion, the relative contributions of mean flow rate and intrapulmonic pressure to efficiency were explored in order to explain aerodynamic aspects in voice disorder.

METHOD

Figure 1 shows the experimental equipment and interconnection among instruments.

Calculation of Intrapulmonic Pressure during Phonation.

When lung compression (intrapulmonic pressure) takes place in the airtight box, lung volume decreases according to Boyle's law, so that air pressure in the box (box pressure) accordingly decreases. Intrapulmonic pressure can be estimated from known values of lung volume, volume of air exhaled, box volume, box pressure, and box pressure change. (7) The lung volume is initially given by the functional residual capacity (FRC) which is calculated from pressure changes in the box and the mouth

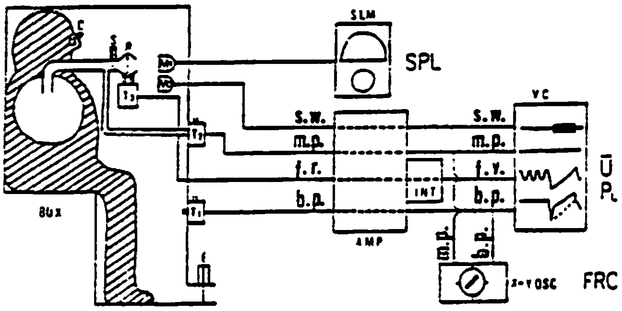


Figure 1. Block diagram of experimental equipment and interconnections among instruments. s.w. - sound wave of vocal signal; m.p. - mouth pressure; f.r. - flow rate; f.v. - respiratory flow volume; b.p. - box pressure; box - airtight box; S - shutter; E - equalizer for pressure; C - nose clip; R - resistance of pneumotachograph; M_1 and M_2 - microphones; T_1, T_2 and T_3 - pressure transducers; AMP - amplifiers; INT - integrator; OSC - oscilloscope; VC - visicorder; SLM - sound level meter.

while the airway is blocked by a shutter at the end of the resting expiration, as described by Dubois et al. (8) Further change of lung volume due to respiration or phonation causes a decrease in box pressure and affects the baseline of the box pressure record. When the flow rate during phonation is almost constant, compensation for this linear effect can be made directly on the record. The validity of this measuring of intrapulmonic pressure has been confirmed (within ± 0.5 cm H_2O) by comparing calculated intrapulmonic pressure and measured mouth pressure while a subject exhaled (without voicing) into a tube which simulated glottal resistance. (7)

Simultaneous Measurements of Sound Pressure Level, Mean Flow Rate, and Intrapulmonic Pressure. Each patient sat in the airtight box with a mouthpiece (attached to the pneumotachograph) placed in the mouth and a clip placed on the nose. The patient was instructed to sustain the vowel /a/ for a few seconds at a comfortable loudness and pitch level. Vocal intensity at the open end of the pneumotachograph tube was measured as sound pressure level (SPL) in decibels (C) at a distance of 18 cm between the end of the tube and the microphone. Background noise was less than 58 dBC. The sound level meter was used as a monitor for the patient to maintain a constant SPL, while the examiner recorded the reading of the meter.

Figure 2 shows typical simultaneous recordings of the sound wave (acoustic signal), respiratory flow volume, and box pressure for a patient. The interval when the amplitude of the sound wave on the recording paper was stable was chosen so that the measured SPL could be used as a simultaneous value with other parameters. Mean flow rate ($\bar{U}$) was measured as the slope of the respiratory volume trace. Intrapulmonic pressure (PL) was caculated as described above. For each patient, average values of the parameters were made over the three records showing the most stable recordings of the voice signal and flow rate.

Indexes of Comparison Between Pathological and Normal Results. The data for ten normal subjects have been obtained with the same method used in this investigation and are shown in Figure 3. (7) The linear averages in Figure 3 are given by the following relationships:

$$\mathrm{SPL} = 96.2 \log \bar{U} - 133.7 \qquad (1)$$

$$\mathrm{SPL} = 31.8 \log \mathrm{PL} + 56.4 \qquad (2)$$

$$\mathrm{SPL} = 2.44\mathrm{APL} + 8.9 \qquad (3)$$

where the aerodynamic power level (APL) equals 10 log $\bar{U}$PL.

Patients were examined only for phonation at comfortable loudness and pitch levels because of their limited abilities to

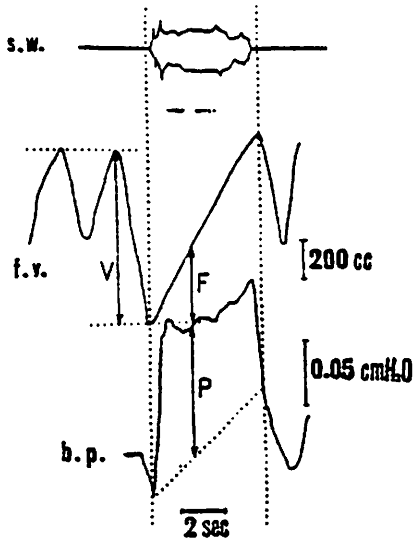


Figure 2. Simultaneous recordings of sound wave of vocal signal (s.w.), respiratory volume (f.v.), and box pressure (b.p.) in patient. P - change of box pressure, F - exhaled air volume during phonation, V - inhaled volume before phonation.

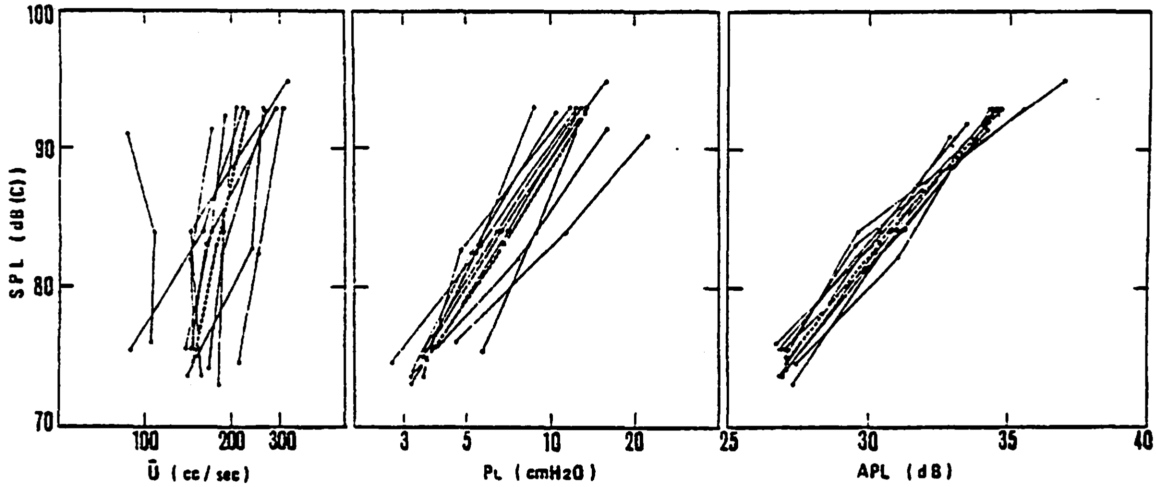


Figure 3. Relationships between sound pressure level (SPL) and mean flow rate ($\bar{U}$), intrapulmonic pressure (PL), and aerodynamic power level (APL) for ten normal subjects at comfortable pitches. APL is calculated from 10 log ($\bar{U}$PL). Solid lines represent the values for each subject. Dashed line expresses values obtained by averaging ten values at similar SPL (about 73, 83, and 93 dB). (Reprinted with permission.)

produce sustained vowels over wide ranges of intensity and fundamental frequency. Therefore, the comparison of a patient's values was given by variations or distances from the mean lines given by equations 1, 2, and 3 for normal subjects. For a given value of APL, DIM was defined as an index indicating the distance (or difference) of the patient's intensity level (SPL) from that given by equation 3.

$$\mathrm{DIM} = \mathrm{SPL} - 2.44\mathrm{APL} - 8.9 \tag{4}$$

When vocal efficiency is expressed as the ratio of sound power at the open end of the tube to aerodynamic power calculated from $\overline{U}$ times PL, the efficiency in decibels can be represented by

$$10 \log E = \mathrm{SPL} - \mathrm{APL} + c \qquad (c\text{:constant}) \tag{5}$$

where c is determined by scales in measures, the distance between the tube and the microphone, and reverberation characteristics of the box. (7) The linear equation for mean efficiency for the normal subjects can be expressed with 10 log E as a function of APL from equations 3 and 4.

$$10 \log E = 1.44\mathrm{APL} + 8.9 + c \tag{6}$$

When the SPL and APL of a patient are s and a, respectively, variation of the efficiency from the value given by equation 6 is

$$(s - a + c) - (1.44a + 8.9 + c) =$$
$$s - 2.44a - 8.9 = DIM \qquad (7)$$

Therefore, the index DIM indicates the patient's variation of efficiency in decibels from the mean line for normals.

In order to estimate the relative contributions of $\bar{U}$ and PL to the DIM for the patient, the distances of $\bar{U}$ and PL in decibels from the mean lines (equations 1 and 2) were also calculated and defined at the index for mean flow rate (DFM) and the index for intrapulmonic pressure (DPM), respectively.

$$DFM = 10 \log \bar{U} - 0.104SPL - 13.9 \qquad (8)$$
$$DPM = 10 \log PL - 0.314SPL + 17.7 \qquad (9)$$

From equations 4, 8, and 9, where APL = 10 log $\bar{U}$PL and $70<SPL<100$,

$$DIM = 2.44(DFM + DPM) - 1.3 \pm 0.3 \qquad (10)$$

Therefore, simple addition of DFM and DPM determines the DIM or the variation of vocal efficiency. Figure 4 offers a graphic interpretation of the three indexes DIM, DFM, and DPM.

The three indexes were obtained for each patient by substituting the values of SPL, $\bar{U}$, and PL into equations 4, 8, and 9.

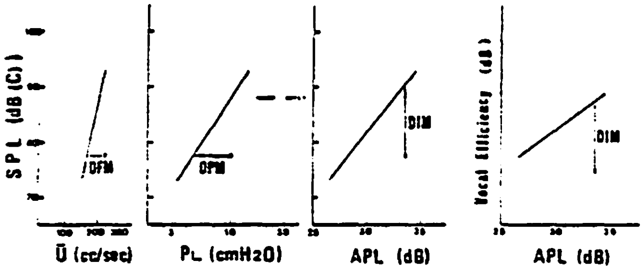


Figure 4. Interpretation of indexes DIM, DFM, and DPM (see Method section). Thick lines represent mean values for normal subjects.

RESULTS

Table 1 lists the lesions and the values of sound pressure level, mean flow rate, and intrapulmonic pressure for the patients. Hypertrophic lesions on the vocal folds were classified into nodes, polyps, and Reinke's edema (polypoid degeneration) according to shape, location, and stiffness. The size of nodes or polyps was roughly judged from the ratio of its length to that of the vocal fold. Among the patients with Reinke's edema, remarkable thick edema was revealed on the vocal folds of patient 12. Laryngoscopy revealed that patients 4 and 16 with unilateral recurrent nerve paralyses had large glottal chinks during phonation. Patients 5 and 15 had widespread glottal squamous cell carcinoma (T2) within a vocal fold, and had limited movement of the affected fold.

Table 2 gives the indexes DIM, DFM, and DPM for each patient. The patients were grouped according to type of lesions. The total ranges for the normal subjects are also shown in Table 2. The variation of vocal efficiency registered very low values compared to the normal ranges in cases of large nodes (patient 8), large polyp (patients 6, 14, 3 and 1), Reinke's edema (patient 12), recurrent nerve paralysis (patients 16 and 4), and T2

TABLE 1. LESIONS AND SIMULTANEOUS VALUES OF SOUND PRESSURE LEVEL, MEAN FLOW RATE, AND INTRAPULMONIC PRESSURE

Pt No.	*Sex*	*Age (yr)*	*Lesion*	*Sound Pressure Level (dBC)*	*Mean Flow Rate (cc/s)*	*Intrapulmonic Pressure (cmH_2O)*
1	M	29	B polyps (large)	87.0	270	14.1
2	M	58	U polyp (small)	85.0	171	11.5
			Postop	84.3	196	7.0
3	M	58	U polyp (large)	79.0	219	8.4
			Postop	81.0	127	7.0
4	F	40	U RNP	77.3	382	6.0
			Postop	76.7	165	6.1
5	M	49	Glottal cancer (T2)	75.0	207	15.0
6	M	42	U polyp (large)	76.0	249	6.6
			Postop	79.0	146	7.0
7	F	40	Reinke's edema	79.0	168	7.8
8	F	19	B nodes (large)	75.0	292	8.3
			Postop	75.6	149	5.9
9	M	35	U polyp (small)	81.7	225	5.7
			Postop	80.3	178	4.0
10	F	62	Reinke's edema	80.3	184	10.2
11	M	75	U polyp (small)	74.5	172	5.6
12	M	55	Reinke's edema	76.7	242	10.3
13	F	34	B nodes (small)	82.0	258	8.3
14	M	43	U polyp (large)	82.0	260	9.6
15	F	56	Glottal cancer (T2)	72.3	144	12.6
16	M	34	U RNP	74.3	1034	6.7

B — bilateral. U — unilateral. RNP — recurrent nerve paralysis.

TABLE 2. TYPES OF LESIONS AND INDEXES
DIM, DFM, AND DPM

Pt No.	*Lesion*	*DIM*	*DFM*		*DPM*	*DIM Postop*
8	B nodes (large)	−16.4	+3.0	⇆	+3.3	−5.1
13	B nodes (small)	− 8.2	+1.7	⇆	+1.1	
6	U polyp (large)	−11.3	+2.2	⇆	+2.0	−3.3
14	U polyp (large)	− 9.8	+1.7	⇆	+1.8	
3	U polyp (large)	− 9.6	+1.3	⇆	+2.1	+0.2
1	B polyps (large)	− 9.3	+1.4	⇆	+1.9	
11	U polyp (small)	− 7.3	+0.7	<	+1.8	
2	U polyp (small)	− 4.2	−0.4	<	+1.6	−1.1
9	U polyp (small)	− 3.0	+1.1	>	−0.4	+1.8
12	Reinke's edema	−15.0	+2.0	<	+3.7	
10	Reinke's edema	− 8.1	+0.4	<	+2.6	
7	Reinke's edema	− 6.0	+0.1	<	+1.8	
16	U RNP	−28.3	+8.5	▶	+2.6	
4	U RNP	−13.6	+3.9	▶	+1.2	−5.5
5	Glottal cancer	−19.1	+1.5	◀	+5.9	
15	Glottal cancer	−16.1	+0.2	◀	+6.0	
Actual range in normal subjects (n = 10)		− 3.0 + 3.0	+1.6 −4.1		+2.7 −2.0	

⇆ — within 1: < > — from 1 to 2.5: ◀ ▶ — more than 2.5.

glottal cancer (patients 5 and 15). Therefore, certain cases within each type of laryngeal disease examined in this study demonstrated ineffective conversion of aerodynamic to sound power compared to the normal subjects.

The results for six of the 16 patients indicated values of DIM lower than the normal range while both DFM and DPM registered values within the normal ranges (patients 3, 1, 11, 3, 10, and 7). This finding suggests that the measure of DIM might be more sensitive to those lesions than the simple measures of the DFM (index for mean flow rate) or DPM (index for intrapulmonic pressure).

The usefulness of the DIM measure in comparing before and after treatment conditions is inferred from the result that patients who underwent operations showed satisfactory improvements of the DIM measure (patients 8, 6, 3, 2, 9, and 4).

Comparison between the indexes DFM and DPM indicates that different types of lesions differed remarkably in the relative contributions of mean flow rate and intrapulmonic pressure to the DIM index, that is, to the variation of efficiency. Low efficiency was mainly associated with high flow rates in recurrent nerve paralysis with a glottal chink (patients 16 and 4), whereas relatively high lung pressure was associated with low efficiency in glottal cancer with limited movement of a vocal fold (patients 5 and 15). In hypertrophic lesions on the vocal folds, especially in large nodes and polyps (patients 8, 6, 14, 3, and 1), the differences between DFM and DPM were very small. Both values of flow rate and pressure tended to be relatively high in these cases.

DISCUSSION

The results of this study indicate that, compared with data for subjects with normal phonation, vocal efficiency is reduced when the larynx has pathological lesions such as nodes, polyps, edema, nerve paralysis, or cancer. In other words, less effective conversion of input aerodynamnic power to ouput sound power takes place for those with the types of laryngeal disease reported in this investigation. This finding is clinically significant because it offers a basis for an objective evaluation of voice disorders. The index DIM is equivalent to the variation of a patient's efficiency value (in decibels) from the mean values for normal subjects. Since the logarithm of vocal efficiency has approximately a linear relation to sound pressure level or aerodynamic power level in normal subjects, (1,4,7) differences in logarithmic expressions of the efficiency such as DIM may be practical. Furthermore, the index DIM is understandable to physicians because it can be explained simply

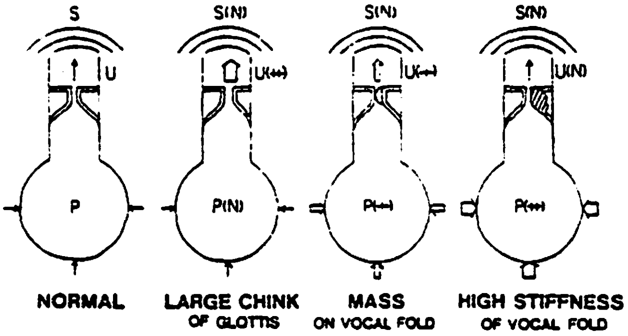


Figure 5. Classification of laryngeal disease. Each type of lesion is characterized by_relative quantities (N, +, and ++) of mean flow rate ($\bar{U}$) and intrapulmonic pressure (P [PL]) when vocal intensity (S) is roughly equal to normal intensity.

by the relationship between sound pressure level (output power) and aerodynamic power level (input power).

The cases of large nodes, large polyps, nerve paralyses, and glottal cancers examined in this study appeared to represent typical lesions of these types. Suzuki (9) found through ultrasound analyses of tissue resonance frequencies that glottal cancer and polyps were characterized by increased stiffness and mass, respectively. In accordance with the results of this present study and the findings of Suzuki, an aerodynamic-biomechanical classification of vocal fold lesion type associated with lwo vocal efficiency can be suggested:

1. Large chink of glottis; low efficiency of vocal production with high flow rate, as in recurrent nerve paralysis with a glottal chink.
2. Mass on vocal fold: low efficiency with high values of both flow rate and pressure, as in large hypertrophic lesions on the vocal fold.
3. High stiffness of vocal fold: low efficiency with high intrapulmonic pressure, as in glottal cancer with limited movement of a vocal fold.

Figure 5 illustrates this classification and shows the relative quantities of flow rate and pressure used in generating the same vocal intensity as the normal case. In the type "large chink of glottis," only a portion of the exhaled air appears to be available to produce sound. In the type "mass on vocal fold," an air leak though the glottal space adjacent to the mass and vibration difficulties due to the increased weight of the vocal fold are possible components. In the type "high stiffness of vocal fold," high subglottal pressure appears to be required to vibrate the vocal folds, even if only one side is affected.

This classification appears to be supported in part by the results of Hirano. (10) He reported in a statistical study that only 24% of polyp cases and 9% of cancer cases had mean flow rates higher than 300 cc/s, whereas 53% of recurrent nerve paralysis cases showed flow rates higher than 300 cc/s. Clinically, a number of patients complain of difficulties in phonating in spite of mean flow rates within the normal range. Therefore, a measure of vocal efficiency appears to be important in the examination of patients with voice disorders.

The mechanism of ineffective conversion of input aerodynamic power to output sound power for patients with laryngeal disease must be further investigated statistically and experimentally in order to give a satisfactory model for each type of laryngeal lesion. The classification above may help to group the aerodynamic aspects of voice disorders. In addition, the techniques for the estimation of intrapulmonic pressure and flow rate described in this investigation offer a promising noninvasive method for evaluating vocal efficiency.

REFERENCES

1. van den Berg, J.W. (1956) "Direct and indirect determination of the mean subglottic pressure," Folia phoniat. (Basel) 8, 1-14.

2. Ladefoged, P., and McKinney, N.P. (1963) "Loudness, sound pressure and subglottal pressure in speech," J. Acoust. Soc. Am. 35, 454-460.

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

4. Schutte, H.K. and van den Berg, J.W. (1976) "Determination of the subglottic pressure and the efficiency of sound production in patients with disturbed voice production," XVIth Int. Congr. Logopedics Phoniatrics. Interlaken: Karger Basel, 415-420.

5. Koike, Y. and Perkins, W. (1968) "Application of miniaturized pressure transducer for experimental speech research," Folia phoniat. (Basel) 20, 360-368.

6. Hixon, T.J. (1972) "Some new techniques for measuring the biomechanical events of speech production," ASHA 7, 68-103.

7. Tanaka, S. and Gould, W.J. (1983) "Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects," J. Acoust. Soc. Am. 73, 1316-1321.

8. Duboise, A.B., Botelho, S.Y., Bedell, G.N., Marshall, R. and Comroe, J.H. (1956) "A rapid plethysmorgraphic method for measuring thoracic gas volume, " J. Clin. Invest. 35, 322-326.

9. Suzuki, M. (1982) "Mechanical properties of the human vocal fold in vivo in the case of vocal fold disorders," Jap. J. Otol. (Tokyo), 85, 654-666.

10. Hirano, M. (1975) "Basic and clinical investigation," Otologica (Fukuoka) 21, 239-440.

THE CLINICAL VOICE LABORATORY

Clinical Application of Voice Research

Wilbur J. Gould, M.D.
Vocal Dynamics Laboratory, Lenox Hill Hospital,
New York, and
The Recording and Research Center
The Denver Center for the Performing Arts

ABSTRACT

In the past decade the number of voice laboratories has increased dramatically. Their primary mission is to enhance patient care by the application of knowledge gained from basic research. They also are dedicated to further improvement of diagnostic and therapeutic resources. The strength of the voice laboratory lies in collaboration between the clinician and the scientist.

This paper has now been published in the Annals of Otology, Rhinology and Laryngology, Volume 93, July-August, 1984, p. 346-350.

In the last decade there has been a tremendous increase in the number of voice laboratories of various levels of sophistication throughout the world. Until 14 years ago, there were three laboratories in the United States and Canada; now there are approximately 35, and the number continues to increase rapidly, with the majority related to a residency training program. In some voice laboratories the resources are rudimentary; in others the staff includes voice scientists and laryngologists of extraordinary training. Most important is the current enthusiasm and interest and the need for collaborative work.

Research into voice function provides increasingly valuable means to evaluate problems of and objectively measure speech production in a clinical setting. The trained ear and the visualizing apparatus of the laryngologist are being supplemented by newer tools which could vastly increase the ability of the clinician to analyze the voice objectively as well as subjectively, and to diagnose, treat, and give follow-up care. (1,2,3) Combined and collaborative experiences of voice laboratories maximize the benefits of new methods by providing a broader base for diagnoses and analyses of presurgical and postsurgical treatment.

With improvement of therapeutic tools, including microlaryngology and laser surgery of an increasingly delicate nature, and with enhancement of diagnostic tools such as computerized tomographic CT scans, ultrasonic studies, nuclear magnetic resonance, etc., we are able to define status and report objectively to an extent heretofore impossible. Phonosurgery such as Teflon injection in the airway gap, thyroid cartilage surgical manipulation, recurrent nerve section, and partial selective section, cannot reach the full potential of usefulness unless the procedures are most adequately analyzed. Also, as we try to help patients gain a more useful voice - for example, after partial laryngeal surgery for cancer - the clinical laboratory offers studies (4) that can contribute to the development of new surgical techniques (Lawrence, V., Gould, W.J.; unpublished observations) and therapeutic approaches.

Before considering laboratory analytic studies, however, the laryngologist must examine the patient completely, taking an adequate history and listening carefully to the patient's voice. Special attention must be paid to details that might be of value in the functional analysis of a patient's phonation potential, e.g., possible causes of abuse, onset of complaint, vocational habits of vocal usage, previous pulmonary disease, smoking, hormonal disorders, etc.

The mirror is still an efficient method of evaluating the larynx. The tongue must be pulled, however, creating the problem of distortion of the laryngeal entry. The mirror can be used with magnification up to three times and with the operating microscope as well. The clinician also should take advantage of the stroboscope for laryngeal examination. Flexible fiberoptics allows examination the larynx and video recording of vocal fold function during running speech.

The analysis and study of phonated speech depends upon the understanding of three interrelated systems: the respiratory, laryngeal and articulatory. That is, phonated speech involves the conversion of subglottal aerodynamic power to laryngeal and acoustic dynamics, with adjustments and tuning of the vocal tract depending upon the kinesthetic as well as auditory senses. Of the studies developed from research into these areas, the most important to me as a practicing laryngologist are those which can be applied directly to patient care.

CLINICAL LABORATORY ASSESSMENT

Respiratory Analysis.

A number of studies by scientists and laryngologists/voice scientist teams indicate feasible clinical methods for respiratory analysis relevant to phonation. In the Vocal Dynamics Laboratory at Lenox Hill Hospital, we use a Collins 9-liter respirometer for screening purposes. This instrument has been most useful in measuring tidal volume, inspiratory capacity, and expiratory reserve volume. More recently, we are working with flow volume loops to determine active pulmonary function ability. The primary goal is to determine the usable air volumes available. We measure the residual volume by either a whole body plethysmograph or helium pulmonary studies.

In clinical work the assessment of phonatory potential is of obvious importance. Our studies suggest that the lower the functional residual capacity, - i.e., the less volume of air remaining in the lungs after the end of quiet expiration - the lower are the theoretical limits for further possible improvement in vocal production. Subsequent residual volume studies have helped to define the theoretical possibilties for further improvement of vocal production capability. If the residual volume is already as low as it would be in the trained singer, as determined in our studies, there is little room for improvement. If it is as high as for the untrained voice in the same study or higher still in the emphysematous patient with vocal problems, it may be possible to improve phonatory capability by implementing well-defined pulmonary exercises or prescribing medication to reduce the residual volume. The ability to reduce this volume to a minimum level seens to be a goal of therapy. (5,6,7)

Hixon (8) used a kinematic method in his studies of the respiratory element of vocal production. The respiratory unit called the chest wall is considered to be a two-part system of rib cage and abdomen. The volume of air the lung displaces is a result of the action of both parts. The anterior-posterior diameters of the rib cage and abdomen are "related linearly to their respective volume displacements". (8) The measurement of diameter is obtained by the use of magnetometers. The analog voltage output created by the electromagnetic coil-pairs gives an equivalent measure of distance, and, thereby, air volumes, after appropriate calculations. Vital capacity, total lung capacity, functional residual capacity and residual volume can then be measured. A great advantage is the ability, previously impossible, to study functions under many different conditions. The kinematic method is a very important non-invasive laboratory assessment aid which allows the investigator to evaluate the production apparatus of individuals with disorders related to functional misuse of the respiratory unit, spinal cord injury, spasmodic dysphonia, Friedrich's ataxia, motor neuron disease, etc. As Hixon stated, this research tool has become "an indispensible part of our daily clinical armamentarium for voice disorders". (8)

A number of other devices for pulmonary assessment are available. Many laboratories are in hospitals with excellent pulmonary laboratories readily accessible.

Analysis of laryngeal function.

The laryngeal sound source consists primarily of quasi-periodic pulses of air, created when the subglottal airflow and air pressure are sufficient to cause vibration of the adducted vocal folds. Basic studies of the dynamic airflow exiting the larynx and of the dynamic air pressures acting on the vocal folds during vibration are necessary in order to expand our theories of phonation as well as to implement improved patient care. (9,10,11,12) The study of the biomechanics, fluid mechanics, and acoustics of phonation using animal and computer models of the larynx will aid in determining limits of aerodynamic and physiological factors which define normal layrngeal function. (12) Larynx models are being used to examine pathological conditons such as asymetric tensions of the vocal folds (13) as well as hyperfunctional arytenoid adductions. (14) Studies of this nature are intended to eventually predict pathological human laryngeal function using noninvasive techniques. These areas of research are fertile ground for the laryngologist/voice scientist teams.

The capability of the sound generator can be assessed by visual analysis, acoustic analysis and aerodynamic studies.

The human phonatory mechanism is the source of sound in voiced speech production. The laryngeal sound source consists primarily of quasi-periodic pulses of air. These air pulses are created when the subglottal air flow and air pressure are sufficient to cause vibration of the vocal folds when in an adducted position. Basic studies of the dynamic air pressures acting on the vocal folds during vibration are necessary in order to expand our theories of phonation as well as to implement these theories for pathological populations. (10,11,12,13) The study of the biomechanics, fluid mechanics and acoustics of phonation using animal and computer models of the larynx will aid in the determining limits of aerodynamic and physiological factors which define normal laryngeal function. (13) Larynx models are being used to examine pathological conditions such as asymetric tensions of the vocal folds (14) as well as hyperfunctional arytenoid adductions. (15) Studies of this nature are intended to eventaully predict pathological human laryngeal function using noninvasive techniques. These areas of research are fertile grounds for the laryngologist/voice scientist teams.

Visual Analysis. Significant film research studies of voice production started in 1937 at the Bell Telephone Laboratories where the first laryngeal high-speed motion pictures were taken. Moore, von Leden, and Timcke advanced high-speed film research to a quantitative level in the late 1950's. (15,16,17) Although high-speed filming is beyond the reach of most clinical laboratories, the direction for proper medical and surgical care that this medium has granted us indirectly allows the clinician to do much improved work with better results. Studies of the motion of the mucosal surface, which have been based on stroboscopic as well as ultrahigh speed cinematography, demonstrate the complexity of the vibratory mechanism needed for efficient vocal production. (18,19,20) We have all seen patients in whom the gross mirror examination of the vocal fold is apparently normal after the so-called stripping operation for a polypoid vocal fold edge. Yet at the same time we have seen these patients presenting a hoarse voice quality. The stroboscope can show the stiffened characteristics of the operated vocal fold. The studies mentioned above clearly demonstrate the need for a mobile mucosa and the usefulness of the stroboscopic technique. (18,19,20) To guarantee a mobile mucosa, an effort must be made to save sufficient mucosa where possible during the laser or sharp dissection laryngoscopic surgical procedure. It would be ideal if the vocal fold were left covered with mucosa while the excess tissue was removed.

The practical clinical laboratory can also use the stroboscope to determine locations of pathological disturbance of the mucosal waves. As stroboscopic tools are improved, they will be used with fiberoptic laryngoscopes as well as mirrors and with simultaneous video recording for later clinical, research, pedagogical, and archival purposes.

The flexible fiberscopes now widely available, along with film or video recording, facilitate laryngeal examination, improved patient record-keeping, and objective following of a patient's care. Clinical research integrating the fiberscope has greatly advanced, as Blaugrund et al report. (21) The postsurgical vocalization in the case of partial laryngectomy is better understood. With further laboratory studies, even more efficient procedures will be devised.

Acoustic Analysis. Listening to the patient talk has been and is a clinician's primary tool for voice quality judgment. While much can be learned about vocal pathology through listening by the laryngologist and speech pathologist, this perceptual judgment cannot easily be quantified and is unreliable in many circumstances (unpublished observations).

Sound spectrogrpahic analysis has been the primary tool for the visual judgment of the degree of hoarseness in the research laboratory as well as in the clinical laboratory setting. It does allow for comparison of the degree of hoarseness, albeit on a subjective level. Digital spectrographic analysis has provided a further advancement in acoustic signal processing. However, with increased application of computer techniques to this problem, we have achieved far more accurate analysis. (22,23) Berke, et al (24) analyzed five acoustic measures to compare voice before and after surgery. Four of these seem to account for a large variance of judgment of the change of voice quality following surgery. The important point is that useful acoustic analysis techniques are being developed that are efficient and sensitive clinical aids for the measurement of changes in vocal quality. These techniques can also help in the evaluation and extension of our surgical technique.

An example of the collaborative potential among laboratories utilizing a set of acoustic measures of voice is being developed by Titze and Scherer at the University of Iowa and The Recording and Research Center at the Denver Center for the Performing Arts. Titze developed a program called Glottal

Imaging by Processing External Signals (GLIMPES) in which a large variety of acoustic measures and estimates of vocal fold motion will be obtained from the simultaneous recording of a number of signals. (25) These signals include the output from the microphone and electroglottograph, as well as the neck accelerometer, which measures vibrations at the neck surface. Also included are the photoglottograph which transduces light shone through the glottis, creating an output proportional to the dynamic glottal area, and the oral flow mask (11) which transduces the flow from the mouth and can be inverse filtered to obtain the dynamic laryngeal airflow signal.

A voice analysis network is being devised through which laboratories around the United States might tie into a central computer to obtain the analyses of signals generated at the home laboratory. This could lead to standardization of voice analysis and relatively easy access to computer requirements. In addition, efforts such as this will allow the formation of a large data base corresponding to the severity or condition of vocal diseases and provide a corresponding directory of participating clinicians and researchers, their institutions, projects, and pertinent resources. (4)

Aerodynamic Studies. Studies of air-flow efficiency during phonation offer valuable information to medical clinicians. Accepted limits are determined, of course, by the study of average population groups. After maximum inhalation, a relatively short period of maximally sustained phonation implies air usage greater than expected. A possible deficiency in glottal closure caused by either a lesion or a gap created by surgery or tumor is suspected under these circumstances. Sometimes a subglottic tumor can then be recognized that would otherwise not be diagnosed. Also, a short period of maximally sustained phonation offers an objective clue that a vocal fold paralysis might have an incomplete approximation. The air loss would be a clue to the amount of closure to be attained surgically.

Researchers are beginning to test a clinical method that relates pulmonary air source capabilities to utilization of the air flow at the larynx. This indirect and relatively noninvasive technique was discussed by Smitheran and Hixon (26) and was used by Homberg and Leanderson (27) to estimate both subglottal pressure and glottal volume airflow in order to calculate laryngeal airflow resistance for normal and disordered voices. In particular, Homberg and Leanderson obtained flow resistance calculations for chronic laryngitis, polyps in the postoperative stage, nodules and various paralyses as well as dysphonia. They found that the indirect method of estimating flow resistance was a useful tool for "clinical work on voice disorders as well as for basic research for speech and singing". (27)

Electromyography. Electromyographic studies help determine which laryngeal muscles in use during different phonatory and respiratory conditions. It is used in some clinical laboratories to follow a patient with laryngeal paralysis, and can indicate the optimal time for surgical correction. The major problem is that electromyogrpahy is invasive when maximially useful, and it is difficult to convince patients to submit to it.

Electroglottography. Although the electroglottographic output signal is typically assumed to be proportional to the dynamic contact area of the medial surface of the vocal folds, no data conclusively show this. Therefore, numerous studies are being done such as a combination of transillumination, acoustical studies, ultrahigh-speed photography and electroglottography to indicate when closure may be occurring relative to the electroglottographic signal. Once there is true identification of the laryngeal activity being measured, electroglottography offers a means of assessing vocal fold activity with little discomfort to the patient and little interference with vocal function. The clinical laboratory then is equipped with an easily handled and useful analytic tool. (1,28)

Ultrasound. Pulse-echo imaging of the larynx using rapid ultrasonic scanners offers a potential for the clinical laboratory which is not yet utilized. This technique may become suitable for viewing and recording dynamic behavior of laryngeal structures with minimal interference to voice and speech production. This noninvasive and essentially radiation free clinical technique, which is widely available for study in other areas of the human body, offers promise for the future once the image-processing is improved as expected. (28)

Supraglottic Studies.

Research laboratories devote a great deal of attention to the spatial, aerodynamic, and acoustical characteristics of the area above the larynx itself. While these evaluations and studies are of great importance to understanding vocal function, they are as yet not practical tools for the clinical voice laboratory. The extensive project started by Fujimura years ago in Tokyo is now being established in Madison, Wisconsin under J.H. Abbs. This technique utilizes radar x-ray devices to track minute lead pellets on the tongue, lips and palate, allowing the scientist to estimate configurations of the oral apparatus during speech. Many research studies related to palatal sensor devices as well as nasal airflow enhance the understanding of vocal production. Pitch and intensity characteristics of the voice and variations relating to facial muscle activities and other areas are being intensely studied. Other studies presently at the investigative stage should provide information useful to the clinician.

Auditory Efficiency.

The ability of the patient to monitor his/her voice on the basis of auditory perception is extremely important. Therefore, it is necessary to determine whether or not a patient has an auditory disability, and if there is any question, hearing tests should be administered during the evaluation stages.

DISCUSSION

Clinical laboratory testing can be of significant aid to the practicing laryngologist. However, care must be taken to select the most practical from many available resources. The clinically oriented laboratory can obtain maximally useful information from the evaluation of respiratory efficiency for those voice deficiencies which, at the very least, are secondary to pulmonary disorders such as emphysema, asthma and other disorders related to obstructive diseases. In addition, cases involving gross pulmonary misuse can be referred for proper therapy. The clinically oriented laboratory can recognize and record laryngeal function disorders as well. Visual means, such as rigid or flexible fiberoptic instruments, acoustical means, electroglottographic instruments and electromyographic instruments are available. Information obtained can be analyzed and interpreted, aiding in the diagnosis and care of the patients. At the very least, an acoustic record of all preoperative voices should be kept, and where possible, the picture of a lesion, equivalent to the preoperative picture taken by a plastic surgeon or an otologist, should be taken.

It must be emphasized that a vastly complex laboratory is not needed for recording the acoustic signal or for respiratory analysis. Good tape recorder instrumentation is available, and a pulmonary laboratory is usually accessible to the laryngologist who does not have pulmonary evaluation equipment in his office. Audiograms are easily obtained for the screening that is necessary where there is any question of hearing deficiency.

The future is even brighter with respect to practical clinical laboratory techniques that would aid the work of the laryngologist. A team including the voice scientist and the laryngologist can take advantage of the latest electronic and computer advances that expand the capability of the individual laboratory.

Most important, interrelations with other laboratories will enlarge our information base and help in diagnostic problems. Standardization of voice analysis techniques would help all of us. As recommended in the First National Conference on Research Goals and Methods in Otolaryngology, 1982, the idea of the research directory/data base has come of age. (4). The computerized bibliographic library of the Voice Foundation, which includes most of the literature on voice studies, adds to that capability.

Improving our understanding and treatment of voice disorders will depend on cooperation between laryngologists, voice scientists, and voice and speech therapists. No one group can handle this monumental task alone. Voice research as a team effort offers the laryngologist new and useful means for a more complete analysis of vocal problems. These analyses are positive aids for the diagnosis and therapy of our patients.

ACKNOWLEDGEMENT

The author thanks Ron Scherer for his valuable contributions to this presentation.

REFERENCES

1. Hanson, D.G., Ward, P.H. and Gerratt, B.R. (1983) "Glottographic measurement of vocal dysfunction," Ann. Otol. Rhinol. Laryngol. 92, 413-420.

2. Rontal, E., Rontal, M., Jacob, H.J. and Rolnick, M.I. (1983) "Quantitative and objective evaluation of vocal cord function" Ann Otol. Rhinol. Laryngol. 92, 421-423.

3. Hirano, M. (1981) Clinical Examination of Voice. New York: Springer-Verlag.

4. Gates, G.A. (1982) First National Conference on Research Goals and Methods in Otolaryngology. Recommendations of the writing groups," Ann. Otol. Rhinol. Laryngol. 91, (suppl. 100).

5. Hixon, T.J. and Putnam, A. (1983) "Voice disorders in relation to repiratory kinematics," Semin Speech Lang. 4, 217-231.

6. Gould, W.J. and Okamura, H. (1974) "Respiratory training of the singer," Folia phoniat. 26, 275-286.

7. Gould, W.J. and Okamura, H. (1973) "Lung volumes in singers," Ann. Otol. Rhinol. Laryngol., 82, 1-7.

8. Hixon, T.J. (1982) "Speech breathing kinematics and mechanism inferences therefrom," in: Gullner, S., Lindblom, B., Lubker, J. and Pearson, A. (Eds.) Speech Motor Control, Oxford, England: Pergamon Press, 75-93.

9. Kitzing, P. and Lofqvist, A. (1975) "Subglottal and oral air pressures during phonation - preliminary investigation using a miniature transducer system," Med. Biol. Eng. 13, 644-648.

10. Koike, Y. and Perkins, W. (1968) "Applications of a miniature transducer for experimental speech research," Folia phoniat. (Basel) 20, 360-368.

11. Rothenberg, M. (1981) "Some relations between glottal air flow and vocal fold contact area," in: Ludlow, C., O'Connell-Hart, M, (Eds.) Proceedings of the Conference on the Assessment of Vocal Pathology. ASHA 11, 48-55.

12. Titze, I.R. and Talkin, D.T. (1979) "A theoretical study of

the effects of various laryngeal configurations on the acoustics of phonation," J. Acoust. Soc. Am. 66, 60-74.

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

14. Scherer, R.C., Cooper, D., Alipour-Haghighi, F. and Titze, I.R. (1984) "Vocal process contact pressure," in: Lawrence, V. (Ed.) Transcripts of the Twelfth Symposium of the Care of the Professional Voice, Vol. 1, New York: The Voice Foundation, 61-68.

15. Timcke, R., von Leden, H. and Moore, P. (1958) "Laryngeal vibration: measurement of the glottic wave, Part 1. Normal vibratory cycle," Arch. Otolaryngol. 68, 1-9.

16. Timcke, R., von Leden, H. and Moore, P. (1959) "Laryngeal vibration: measurement of the glottic wave, Part 2. Physiological variation," Arch. Otolaryngol. 69, 438-444.

17. Moore, P., von Leden, H. and Timcke, R. (1960) "Laryngeal vibration: measurement of the glottic wave, Part 3. Pathologic larynx," Arch Otolaryngol. 71, 16-35.

18. Gould, W.J., Jako, G.J. and Tanabe, M. (1974) "Advances in high-speed motion picture photography of the larynx," Trans. Am. Acad. Opthalmol. and Otolaryngol. 78, 276-278.

19. Hirano, H., Gould, W.J., Lambiase, A. and Kakita, Y. (1980) "Movements of selected points on a vocal fold during vibration," Folia phoniat. 32, 39-50.

20. Schonharl, E. (1960) Die Stroboskopie in der praktischen Laryngologie, Stuttgart: Theime.

21. Blaugrund, S.M., Gould, W.J., Haji, T., Meltzer, J., Bloch, C. and Baer, T. (1984) "Voice analysis in the partially ablated larynx, a preliminary report" Ann. Otol. Rhinol. Laryngol. 93, 311-317.

22. Yumoto, E., Gould, W.J. and Baer, T. (1982) "Harmonics-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. 71, 1544-1550.

23. Yanagihara, N. (1967) "Significance of harmonic changes and noise components in hoarseness," J. Speech Hear. Res. 10, 531-541.

24. Berke, G.S., Gerratt, B.R. and Hanson, D.G. (1983) "An acoustical analysis of the effects of surgical therapy on voice quality," *Otolaryngol Head Neck Surg.* 91, 502-508.

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

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

27. Homberg, E. and Leanderson, R. (1983). "Laryngeal aerodynamics and voice quality," in: Lawrence, V. (Ed.) *Transcripts of the Eleventh Symposium, Care of the Professional Voice, Part II, Medical/Surgical Sessions: papers*. New York: The Voice Foundation, 124-129.

28. Baer, T., Titze, I.R. and Yoshioka, H. (1983) "Multiple simultaneous measures of vocal fold activity," in: Bless, D.M., Abbs, J.H. (Eds.) *Vocal Fold Physiology, Contemporary Research and Clinical Issues*, San Diego, College-Hill Press, 229-237.

29. Zagzebski, J.A., Bless, D.M. and Ewanowski, S.J. (1983) "Pulse echo imaging of the larynx using rapid ultrasonic scanners," in Bless, D.M., Abbs, J.H. (Eds.) *Vocal Fold Physiology, Contempory Research and Clinical Issues,* San Diego, College-Hill Press, 210-223.

VOCAL FATIGUE: PROLONGED LOUD TALKING BY TRAINED VOICES

Ronald C. Scherer, Ph.D.
The Denver Center for the Performing Arts

Ingo R. Titze, Ph.D.
University of Iowa
and
The Denver Center for the Performing Arts

Bonnie N. Raphael, Ph.D.
The Denver Center for the Performing Arts

Raymond P. Wood, M.D.
The University of Colorado Health Sciences Center

ABSTRACT

One female and one male vocalist were subjected to prolonged speaking tasks. Diagnostics on their vocal condition were performed in 15 minute intervals. These consisted of acoustic perturbation measures, perceptual evaluations by self and a panel of listeners, and medical (otolaryngologic) examinations. Results indicate that self-perception of fatiguing is more sensitive than all other measures, that current acoustic perturbation measures are least sensitive, and that medical assessment of vocal fold condition is in agreement with self and listener perceptual assessment.

INTRODUCTION

Vocal endurance is a requirement for many professional singers, actors and speakers. Lack of endurance can curtail some professional vocal careers and seriously hinder others. It is often a symptom of pathological conditions within the larynx.

It seems reasonable to assume that vocal fatigue, a perceptual term related to undesirable quality resulting from prolonged use, would manifest itself as changes in acoustic measures. If the voice begins to "break down" during use, vocal output may become more unstable, and manifest itself as changes in such perturbation measures as jitter and shimmer, where jitter refers to cycle-by-cycle changes in period or frequency, and shimmer refers to cycle-by-cycle changes in signal amplitude. Vocal fatigue may also be associated with throat discomfort and changes in tissue characteristics of the vocal folds.

A number of studies appear directly pertinent to the study to be reported here. Sander and Ripich reported a study in 1981 of 24 young males who spoke loudly or very loudly for a total of 10 phonation minutes. Only 10 of the 24 subjects felt that their voices were less clear after the reading task, although 17 of the 24 felt they were using more effort to produce voice at the end of the experiment. There was great variety among the subjects in how they felt their voices fluctuated with regard to becoming better and worse as the reading task progressed. 20 of the 24 subjects experienced some discomfort during the task, either in the throat, chest, ear or back. Minor pain was reported by 9 of the 24. The most popular discomfort was a "dry throat" (16 of the 24). There were no significant changes in the acoustic measures (pitch range, max vowel duration, max vowel intensity) between the beginning and end of the experiment. Also, auditors' perceptions of the voice quality did not correspond to the subjects' judgments of their voice clarity changes or discomforts. In addition, the auditors were not very successful at differentiating prespeech and postspeech samples (they were incorrect 36% of the time). The results of this study suggests that during a loud speaking task, 1) speakers' perceptions of vocal well-being vary among the speakers, 2) the subjects typically experience some throat discomfort, 3) there may be no significant change in acoustic measures throughout the task, and 4) vocal quality judgments by auditors may not correspond to speakers' self perceptions of voice changes. This study did not appear to vocally tax the speakers to a large degree, however.

Neils and Yairi (1984) reported a study in which six female subjects read aloud for 45 minutes (at an intensity level that is not specified in their report) within 3 levels of noise, followed by 45 minutes of vocal rest. Every 15 minutes a 32-word passage and 3 CVC syllables were recorded. Airflow measures were also obtained before and after the reading and after the rest period. The results from their study indicate (as in the previous study) that there was much diversity in the responses and inconsistent behavior among the subjects. The only measure of statistical significance [unspecified and unexplained, however] across time was the peak airflow measures. Fundamental frequency and vocal quality ratings did not change significantly over time.

Sherman and Jensen (1962) asked 15 normal males to read (presumably at normal loudness levels) for one and one-half hours, followed by a one-half hour of silence. The Rainbow passage was recorded before the task, after the first and second 45 minute interval, and after the one-half hour of silence. The recordings at the four time intervals were judged by auditors for degree of harshness on a 7-point equal-appearing interval scale. Results indicated a decrease in harshness ratings across the first three intervals, with a subsequent increase in harshness rating after the rest period. These results suggest that a "warm-up" effect of prolonged phonation, or an adaptation to the task, occurred (Sherman & Jensen, 1961, p.176). It is important to note that Sherman and Jensen reported that some subjects indicated increased vocal strain during the first half hour, after which "their voices seemed to get better and that they felt as though they could continue reading indefinitely" (p.175). For our purposes, these results suggest that 1) there may be an initial deviance from normal vocal quality, followed by 2) an improvement over time of vocal quality, with 3) a return toward a deviant voice quality after vocal rest. This suggestion would constitute an opposite hypothesis that would appear logical, viz., that vocal quality would worsen over time, with recovery to normal after vocal rest.

The studies reported above also suggest that "vocal fatigue" may be a rather elusive subject of investigation and is subject-specific in effect. We need straightforward measures, however, that correlate well with effects of vocal fatigue for diagnostic and treatment purposes. If vocal fatigue is closely associated with muscle fatigue (Bigland-Ritchie & Woods, 1984), and if muscle fatigue is associated with fluctuations in neuromuscular control, then perturbation measures such as jitter and shimmer may be relevant. These measures may also be relevant if there are random changes in mucosal or mucus characteristics, or increased airflow turbulence or instabilities through the larynx.

The study here is a pilot project to investigate vocal fatigue. The informal hypotheses are:

1. Talking loudly over a period of time should result in the perception of vocal changes by the subject and by listeners.
2. Acoustic perturbation measures should change over time and indicate greater instability of the voice.
3. Auditor ratings of voice quality should parallel both the subjects' self voice ratings and acoustic analyses.

EXPERIMENTAL DESIGN

Two subjects participated in this experiment. Both subjects have relatively well trained voices. The female subject has had voice training for the theater, and is presently a voice coach for a repertory theater. The male subject has had formal singing training, has sung professionally, and currently teaches singing.

The experiment required two series of tasks of the subjects. Each Fatiguing Task required the subjects to read from a text at a loudness level very close to the maximum loudness the subject could produce. The subject monitored the required sound pressure level on a Bruel Kjaer sound level meter. Each Fatiguing Task lasted 15 minutes, separated by the Diagnostic Task. During the Diagnostic Task, the subject recorded 3 phonations of the vowel /a/ and rated how he/she felt about his/her voice. The /a/ recordings were made at a predetermined comfortable loudness level (and SPL) and at a musical fifth above the lowest sustainable pitch. The self voice rating was made on a scale of 1 to 100, where 100 represented "the best shape your voice has ever been in" and 1 represented "the worst (shape your voice has ever been in)".

The first Diagnostic Task recordings were made prior to the first Fatiguing Task reading. The subjects were asked to continue the experiment of Fatiguing Task followed by Diagnostic Task until either the subject or the experimenter decided that the subject should stop (for whatever reason). Both subjects coincidentally decided to stop after 5 Fatiguing Task readings. Following the last Fatiguing Task, the subjects recorded during another Diagnostic Task, then had two rest periods of 15 minutes, after each of which the Diagnostic Task recordings were made again. These last two periods were the "rest periods".

In addition, prior to the experiment and after the last Fatiguing Task, fiberoptic examinations of the larynx of the subjects were made by a laryngologist.

The /a/ phonations of the subjects during the Diagnostic Tasks were recorded directly into a VAX 11/750 computer using a Digital Sound Corporation A/D system. The recordings were digitized for 2 seconds at 20,000 samples per second. These /a/ recordings were used both for obtaining acoustic measures and for auditor analysis. The program GLIMPES was used to obtain measures of jitter, shimmer, and harmonics-to-noise ratios. The GLIMPES program can resolve jitter and shimmer ratios to about 0.1%.

The /a/ recordings were randomly recorded onto magnetic tape using an Otari MX5050 tape recorder. Listeners heard the /a/ phonations binaurally over high-quality earphones. Each token was presented three times in quick succession prior to the auditor rating the vocal quality of the token. The auditors rated the voice quality of the phonations on a scale of 1 to 7, where 1 represented Good Voice Quality and 7 represented Poor Voice Quality. The lists were repeated for reliablity measures.

There were two basic experiments, distinguished by the type of Fatiguing Task. The permitted pitch at which the subject read during the Fatiguing Task was free to change in a normal way (FREE PITCH READING), or was held constant at the musical fifth above the lowest sustainable pitch (MONO PITCH READING), the same pitch used for the Diagnostic Task. It was felt that the MONO

PERTURBATION FACTORS WITH LINEAR TREND
(SHIMMER, JITTER):

$$PF = \frac{100\sum_{i=2}^{N}\left|x_i - x_{i-1}\right|}{(N-1)\bar{x}}$$

HARMONICS-TO-NOISE RATIO*:

$$H/N = N \int_0^T f_A^2(t)\,dt \Big/ \sum_{i=1}^{N} \int_0^{T_i} \left[f_i(t) - f_A(t)\right]^2 dt$$

$$\text{where} \quad f_A(t) = \sum_{i=1}^{N} \frac{f_i(t)}{N}$$

* YUMOTO, E., GOULD, W.J., AND BEAR, T. (1982). "HARMONICS-TO NOISE RATIO AS AN INDEX OF THE DEGREE OF HOARSENESS," J. ACOUST. SOC. AM. 71, 1544-1550.

Figure 1. Equations used in acoustic analyses (automatically performed within the program GLIMPES).

PITCH READING task would result in greater vocal fatigue since the range of physiological change during the Fatiguing Task perhaps would be less than in the FREE PITCH READING Fatiguing Task.

ACOUSTIC ANALYSIS CALCULATIONS

Figure 1 shows the equations used to calculate the frequency and amplitude perturbation factors (i.e., jitter and shimmer, respectively). Also the equation used to calculate the harmonics-to-noise ratios is given. The coefficient of variation was also calculated but not reported here. It is noted that greater perturbation in these measures refers to a decrease in the value for H/N and an increase in the value for jitter and shimmer.

RESULTS

Acoustic Analysis

The acoustic results appear to be fairly straightforward: The various measures (harmonic-to-noise ratio, jitter, shimmer) do not appear to change significantly over the 8 Diagnostic Task sessions.

Figure 2 shows the mean acoustic measures (averaged over the 3 tokens obtained during each Diagnostic session) across the

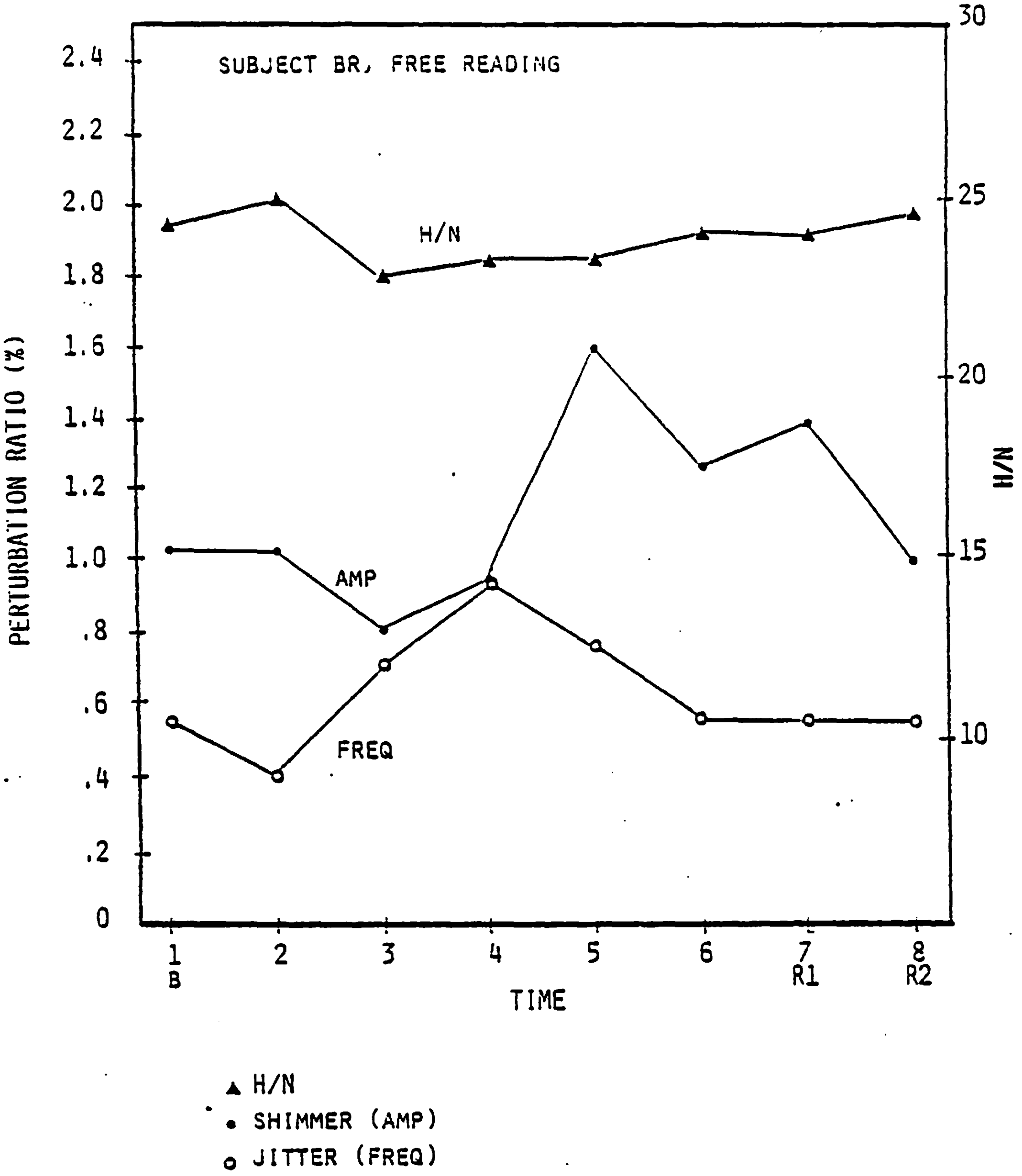


Figure 2. Mean values of the perturbation ratio measures for subject BR for the FREE PITCH READING task.

8 recording sessions for the female subject when she was permitted to freely change the pitch during the Fatiguing Task. The H/N measure is fairly constant over time. The shimmer (AMP) measures appear to increase toward the middle of the experiment, suggesting that the voice becomes worse after about 45 minutes of loud talking. Similarly, the jitter (FREQ) measure appears to rise afte 15 minutes of talking, also suggesting a worse voice. Both these measures then suggest improved voice toward the end of the experiment. Figures 3-5 show the three harmonics-to-noise, jitter, and shimmer values (and means) obtained at each Diagnostic Task recording interval. Figures 3-5 indicate that the variation of values of these acoustic measures is large at most of the recording times; with similar overlap regions, and suggest that the means for the trio of values are not significantly different across time.

Figure 6 shows the mean acoustic measures (averaged over the 3 tokens obtained during each Diagnostic Task session) across the 8 recording sessions for the female subject when she read at the monotone pitch. There is now greater change in the H/N ratio across time, suggesting that after about 45 minutes the voice began to deteriorate until the second rest period. A similar statement is suggested by the other two measures (jitter and shimmer). Again, however, the variation in values shown in

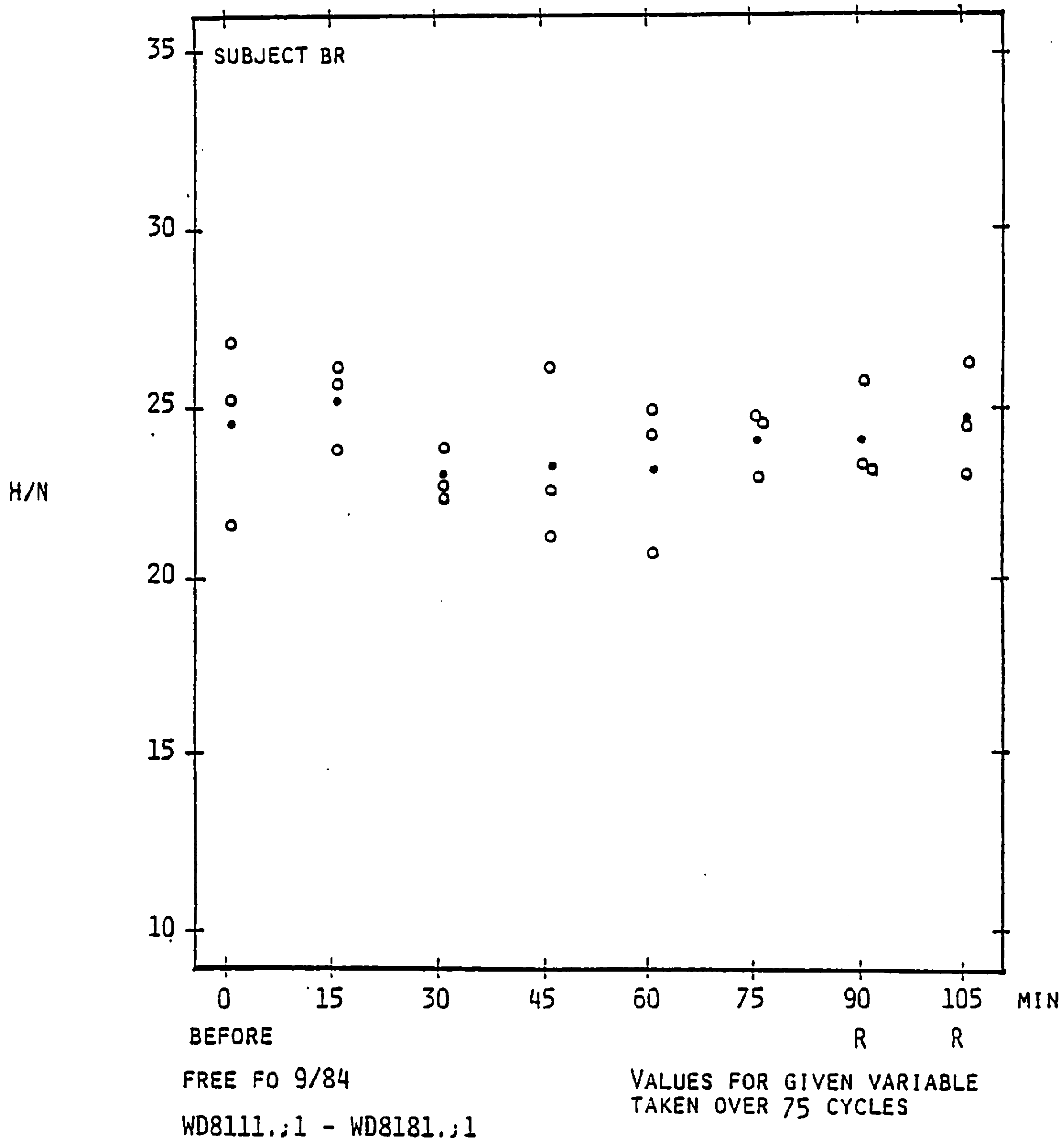


Figure 3. Harmonics-to-noise ratio (H/N) values for subject BR for the FREE PITCH READING task. The filled circles refer to the means of the three values (open circles) at each of the specific 15 minute intervals.

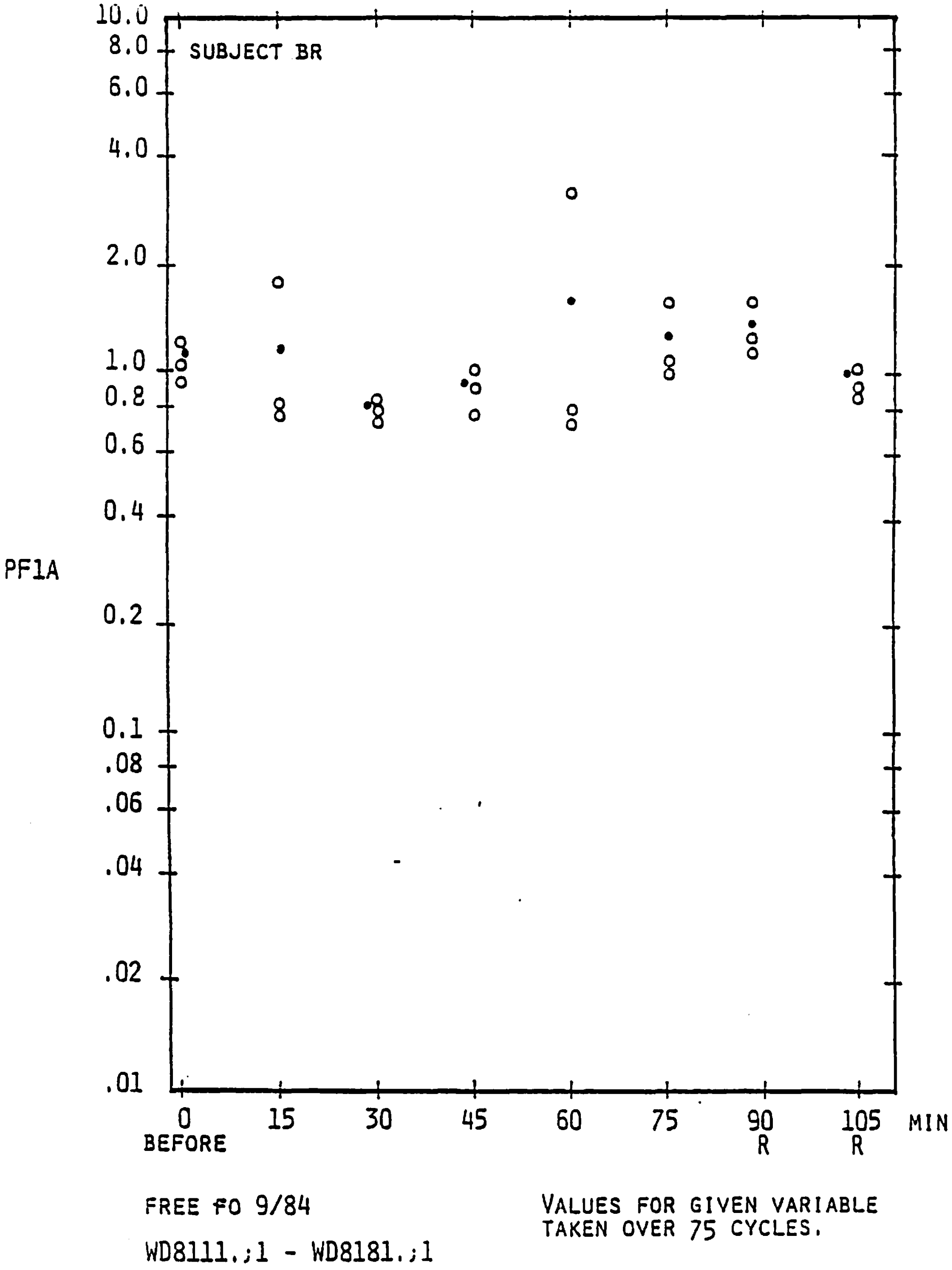
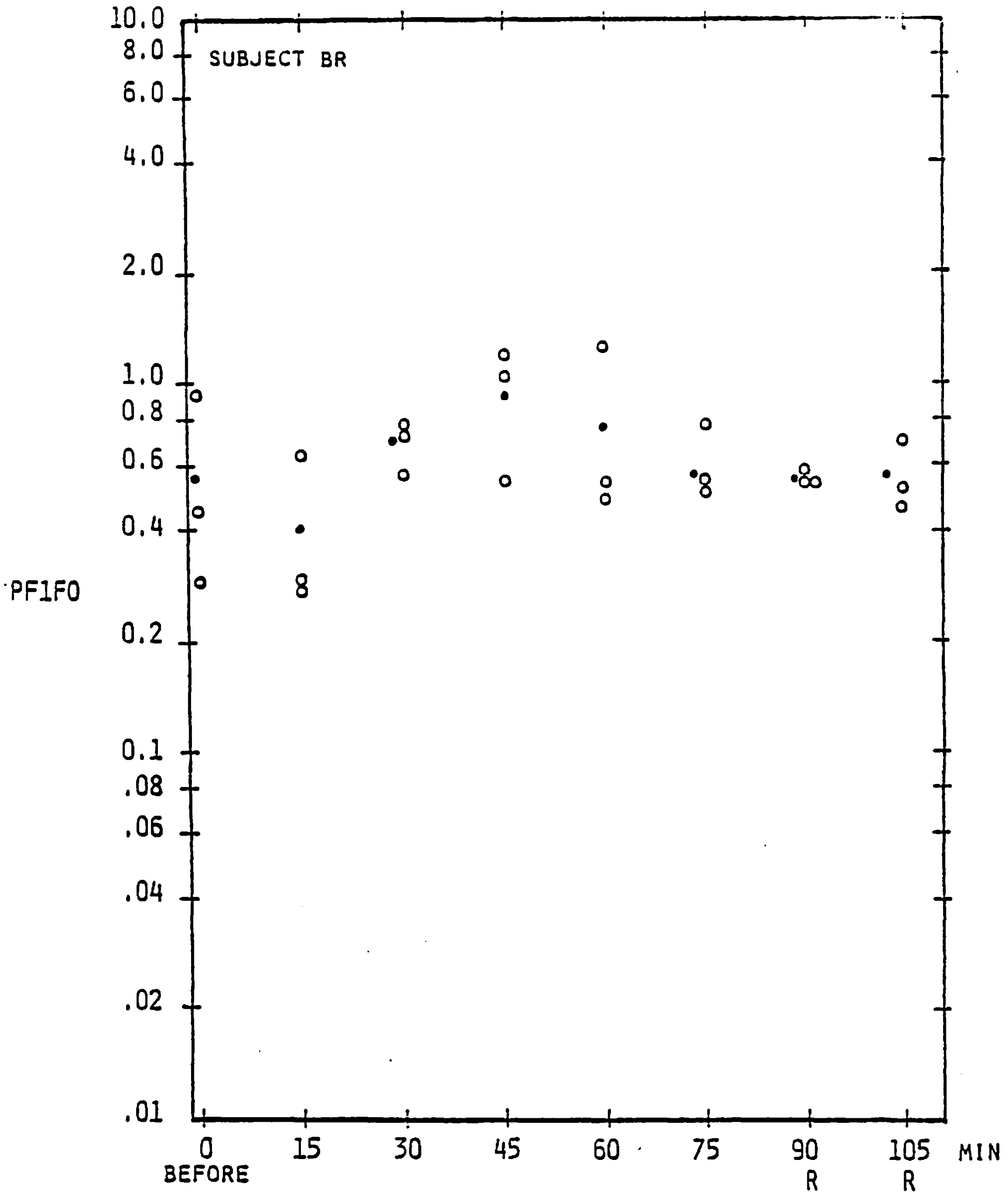


Figure 4. Shimmer (PF1A, amplitude perturbation) values for subject BR for the FREE PITCH READING task. The filled circles refer to the means of the three values (open circles) at each of the specific 15 minute intervals.



FREE FO 9/84

WD8111.;1 - WD8181.;1

VALUES FOR GIVEN VARIABLE
TAKEN OVER 75 CYCLES.

Figure 5. Jitter (PF1FO, frequency perturbation) values for subject BR for the FREE PITCH READING task. The filled circles refer to the means of the three values (open circles) at each of the specific 15 minute intervals.

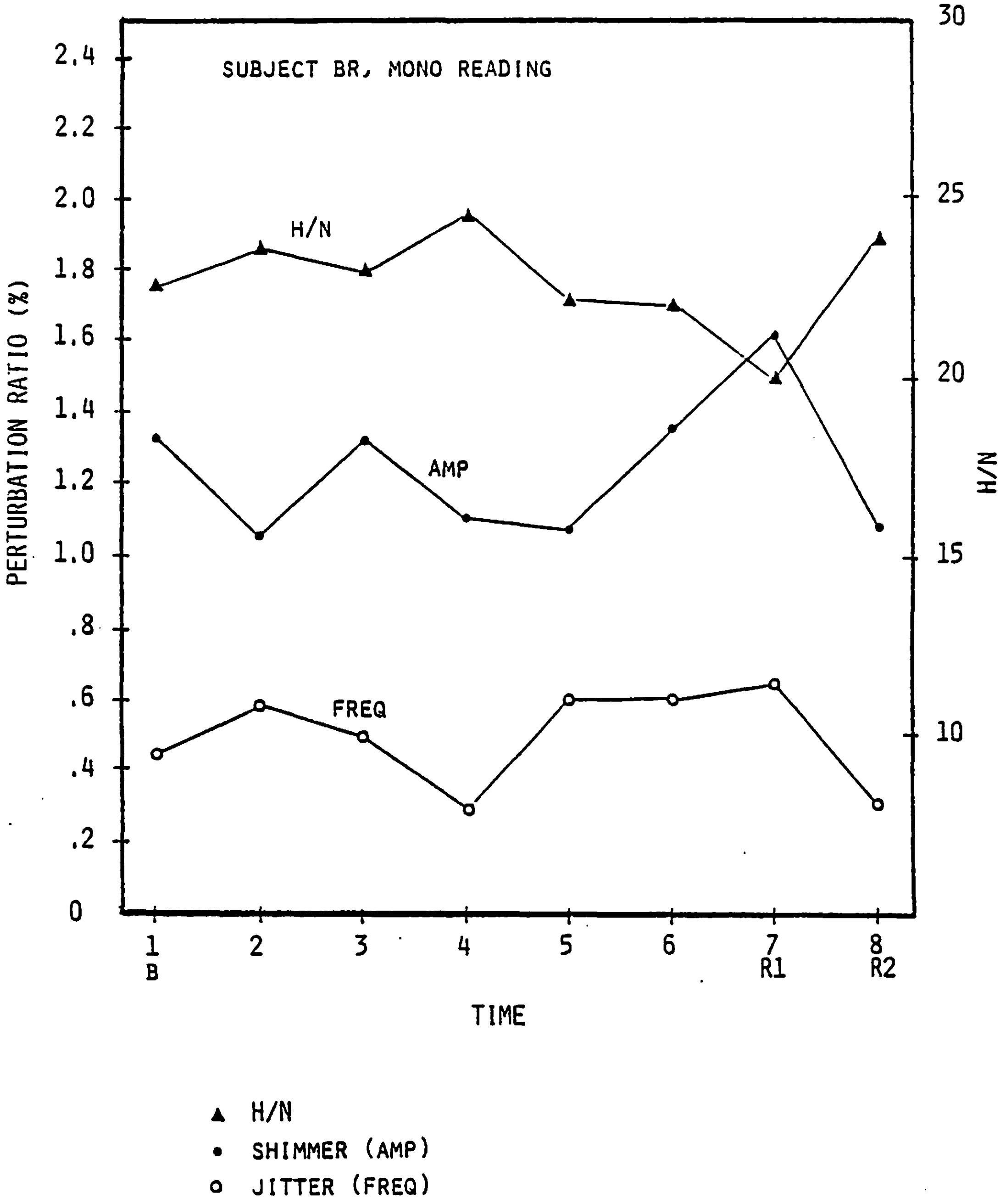


Figure 6. Mean values of the perturbation ratio measures for subject BR for the MONO PITCH READING task.

Figures 7-9 is too large, and the overlap ranges too similar, to conclude that there are any significant differences among the mean values across time.

Figure 10 shows the mean acoustic measures (averaged over the 3 tokens obtained during each Diagnostic Task session) across the 8 recording sessions for the male subject when he read with monotone pitch (he did not participate in an experiment with free pitch reading). The values for H/N and jitter do not change significantly whereas the shimmer values suggest a trend toward a worse voice condition from time 2 through time 7, with a large reduction after the second rest period. Figures 11-13 again show the mutual overlap of data across time characteristic of this experiment. The cluster of data in Figure 11 appear to suggest a significant increase in the H/N values (perhaps related to an improvement of the voice) after the first rest period, with a subsequent significant decrease in the H/N values (perhaps related to a poorer voice quality) after the second rest period. The shimmer data for this subject, shown in Figure 12, appear to suggest that there may be a significant improvement between the first and second rest periods. Both Figure 12 and 13 suggest, due to the data clusters, that the voice quality may have deteriorated following the first 15 minute vocal rest period. (This would suggest that the H/N higher values for this period

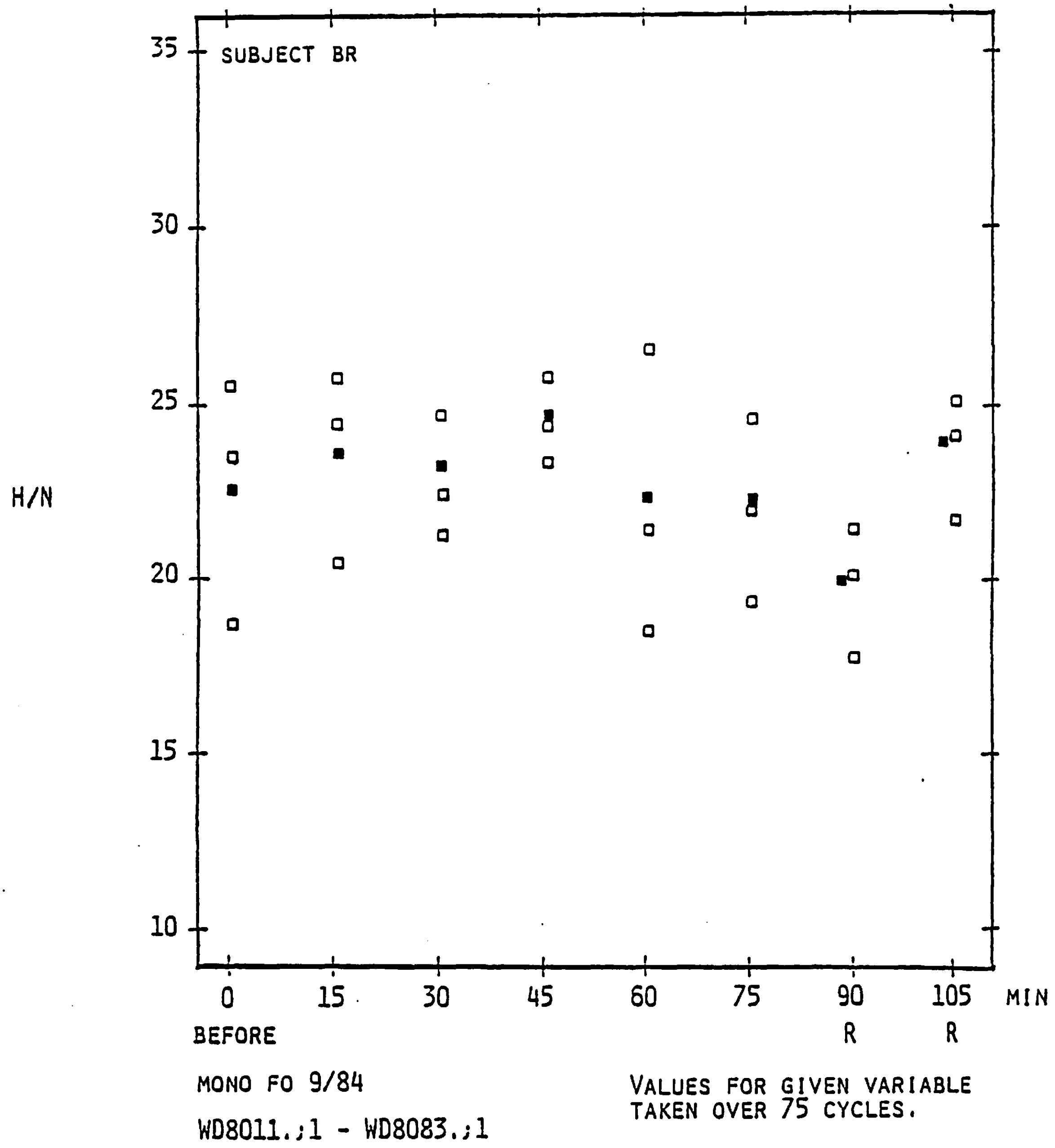


Figure 7. Harmonics-to-noise ratio (H/N) values for subject BR for the MONO PITCH READING task. The filled squares refer to the means of the three values (open squares) at each of the specific 15 minute intervals.

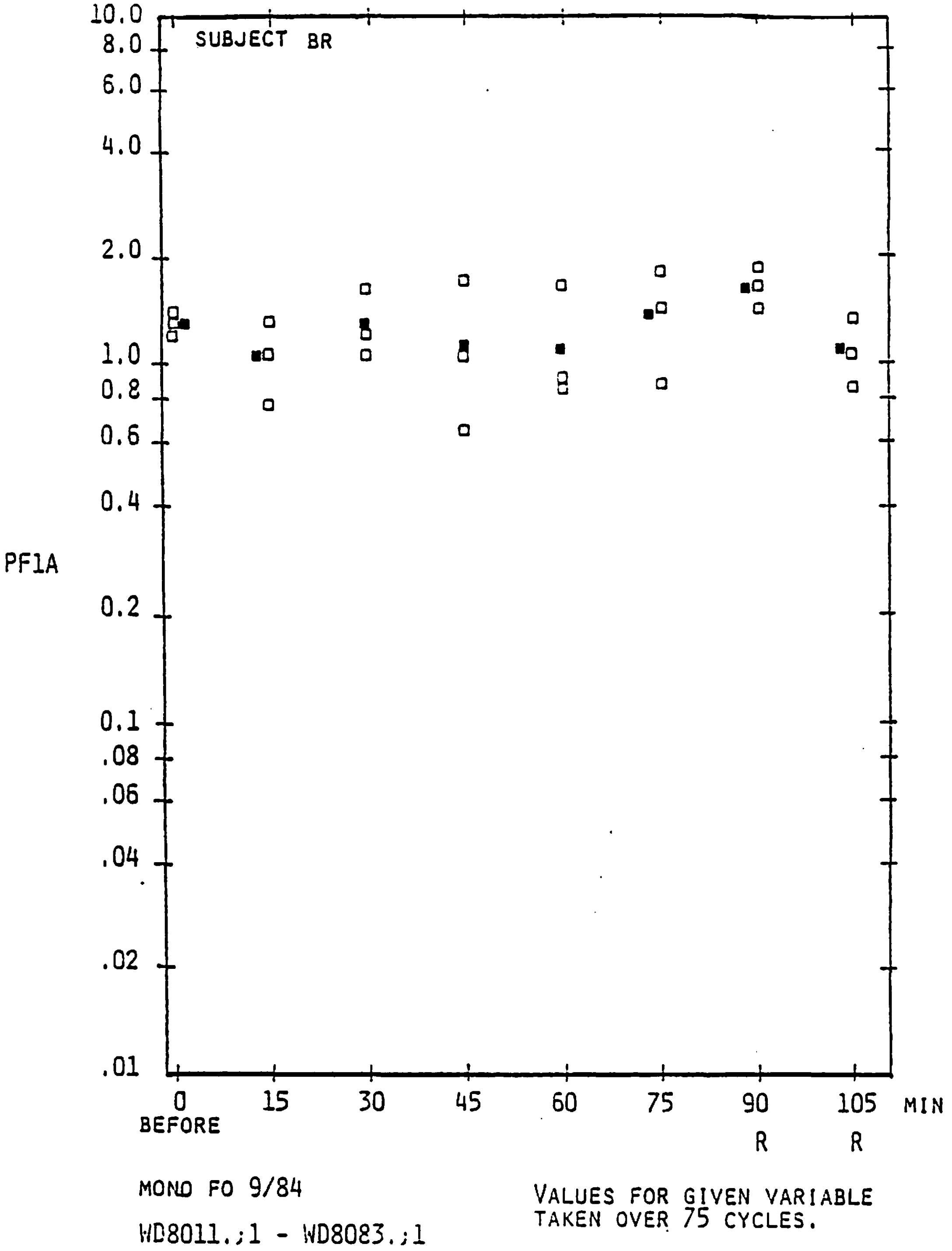


Figure 8. Shimmer (PF1A, amplitude perturbation) values for subject BR for the MONO PITCH READING task. The filled squares refer to the means of the three values (open squares) at each of the specific 15 minute intervals.

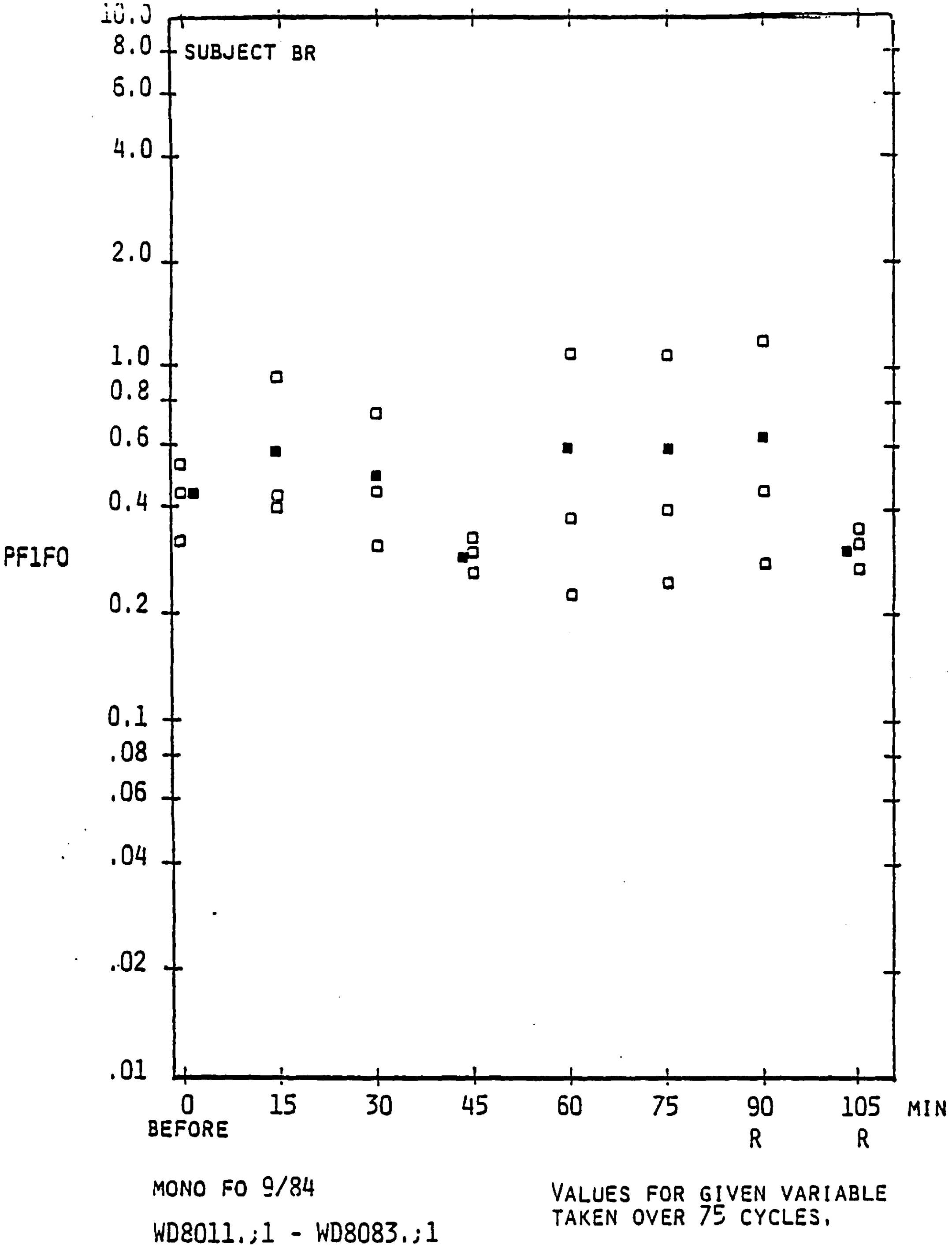


Figure 9. Jitter (PF1F0, frequency perturbation) values for subject BR for the MONO PITCH READING task. The filled squares refer to the means of the three values (open squares) at each of the specific 15 minute intervals.

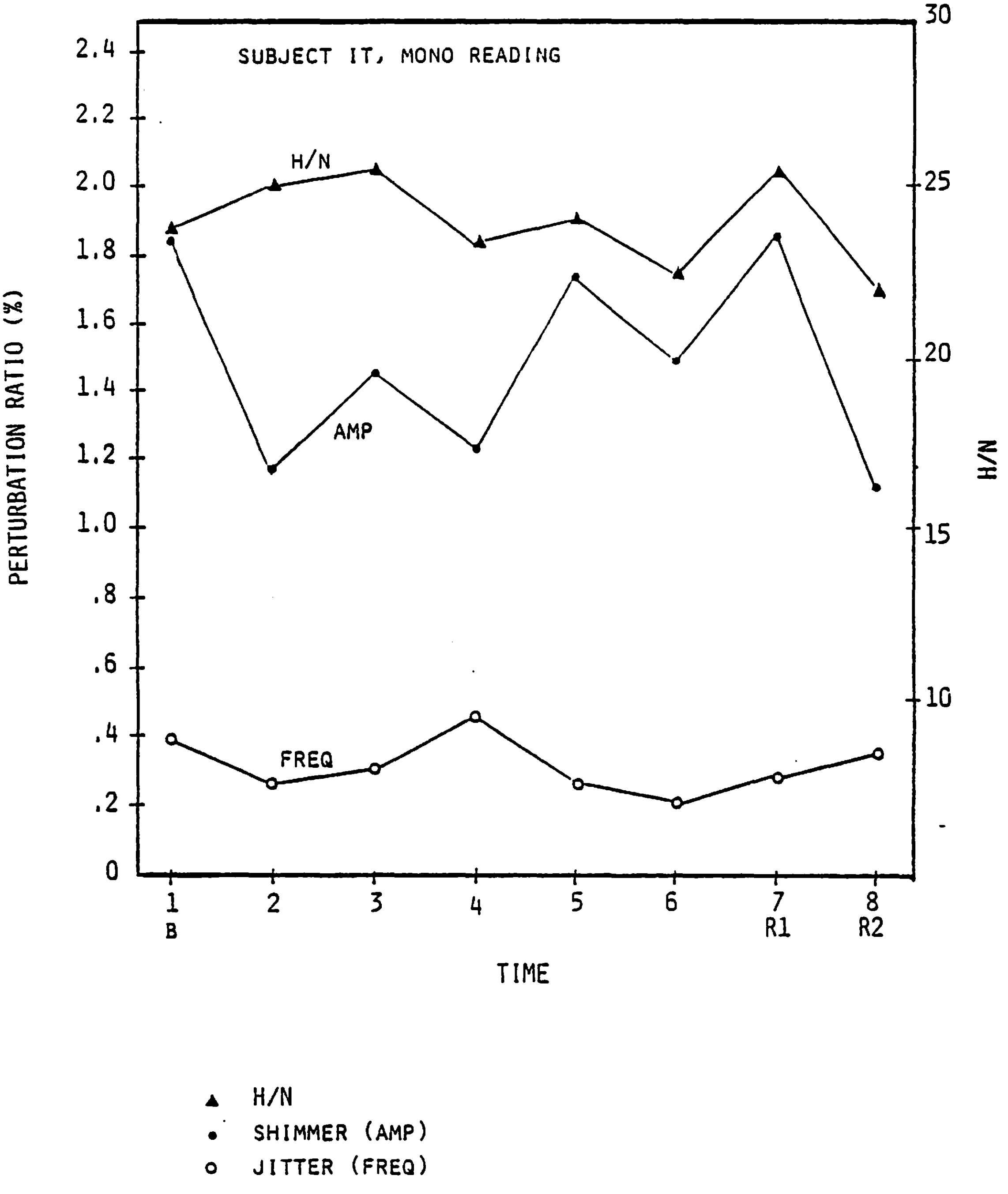


Figure 10. Mean values of the perturbation ratio measure for subject IT for the MONO PITCH READING task.

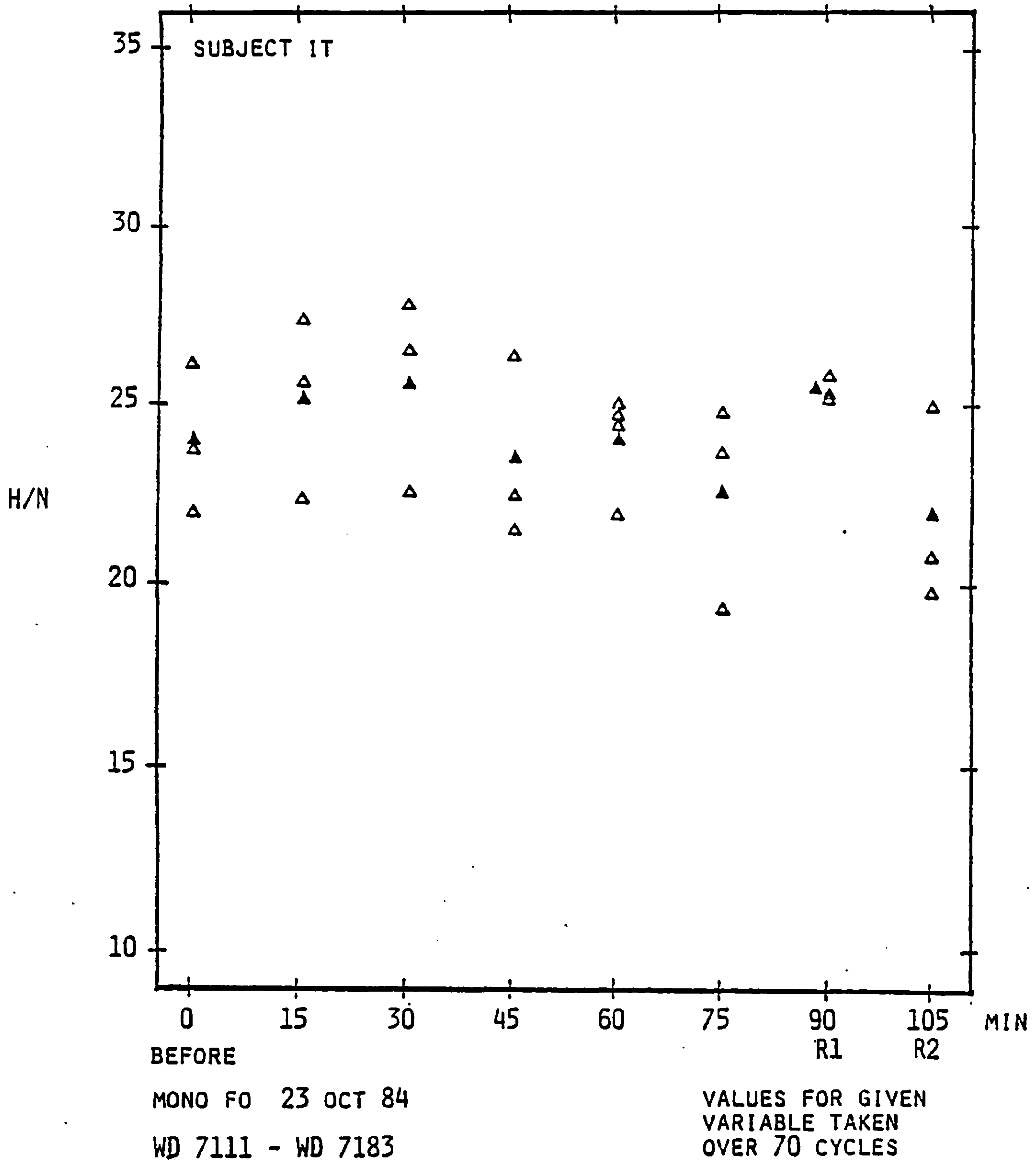


Figure 11. Harmonics-to-noise ratio (H/N) values for subject IT for the MONO PITCH READING task. The filled triangles refer to the means of the three values (open triangles) at each of the specific 15 minute intervals.

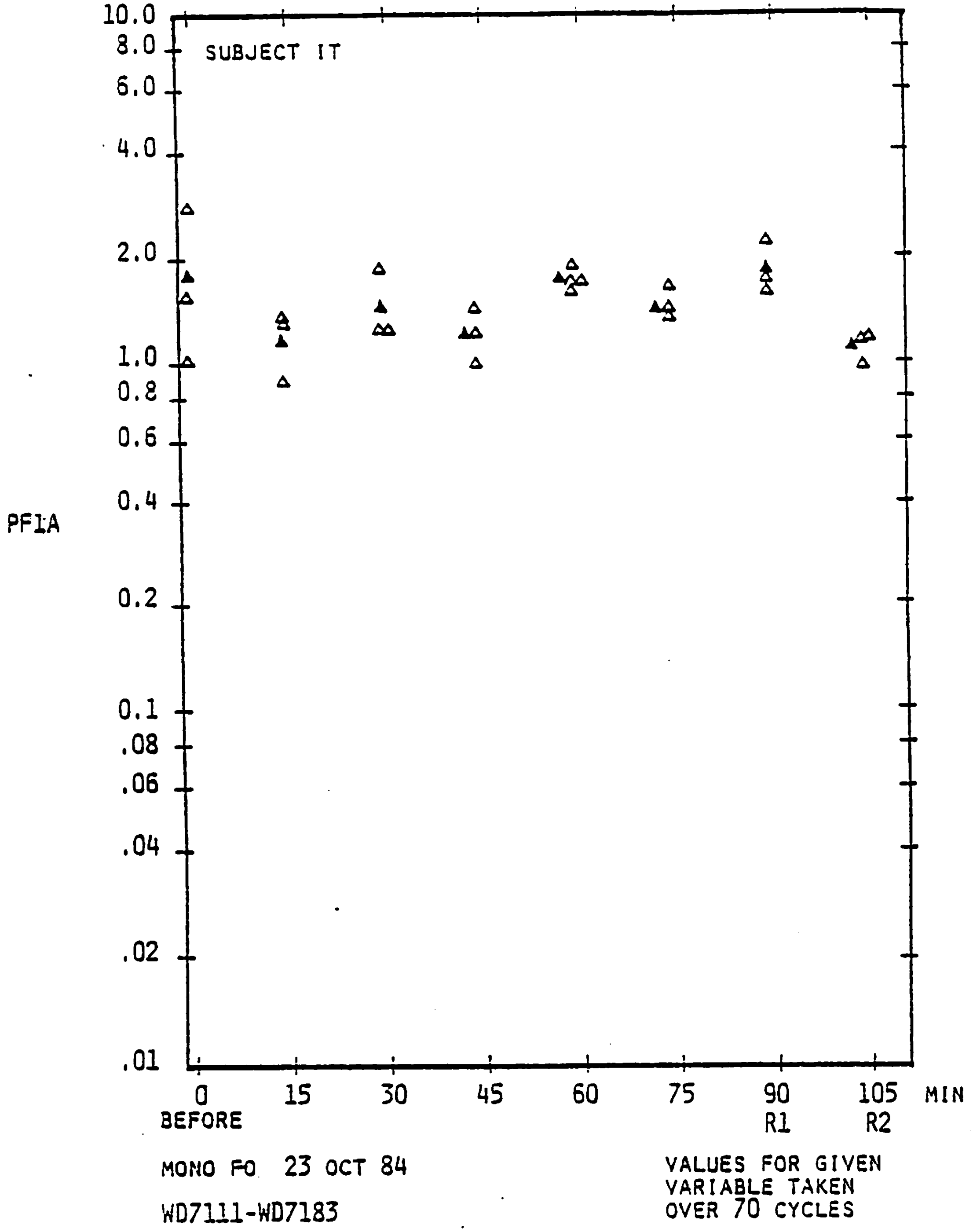


Figure 12. Shimmer (PF1A, amplitude perturbation) values for subject IT for the MONO PITCH READING task. The filled triangles refer to the means of the three values (open triangles) at each of the specific 15 minute intervals.

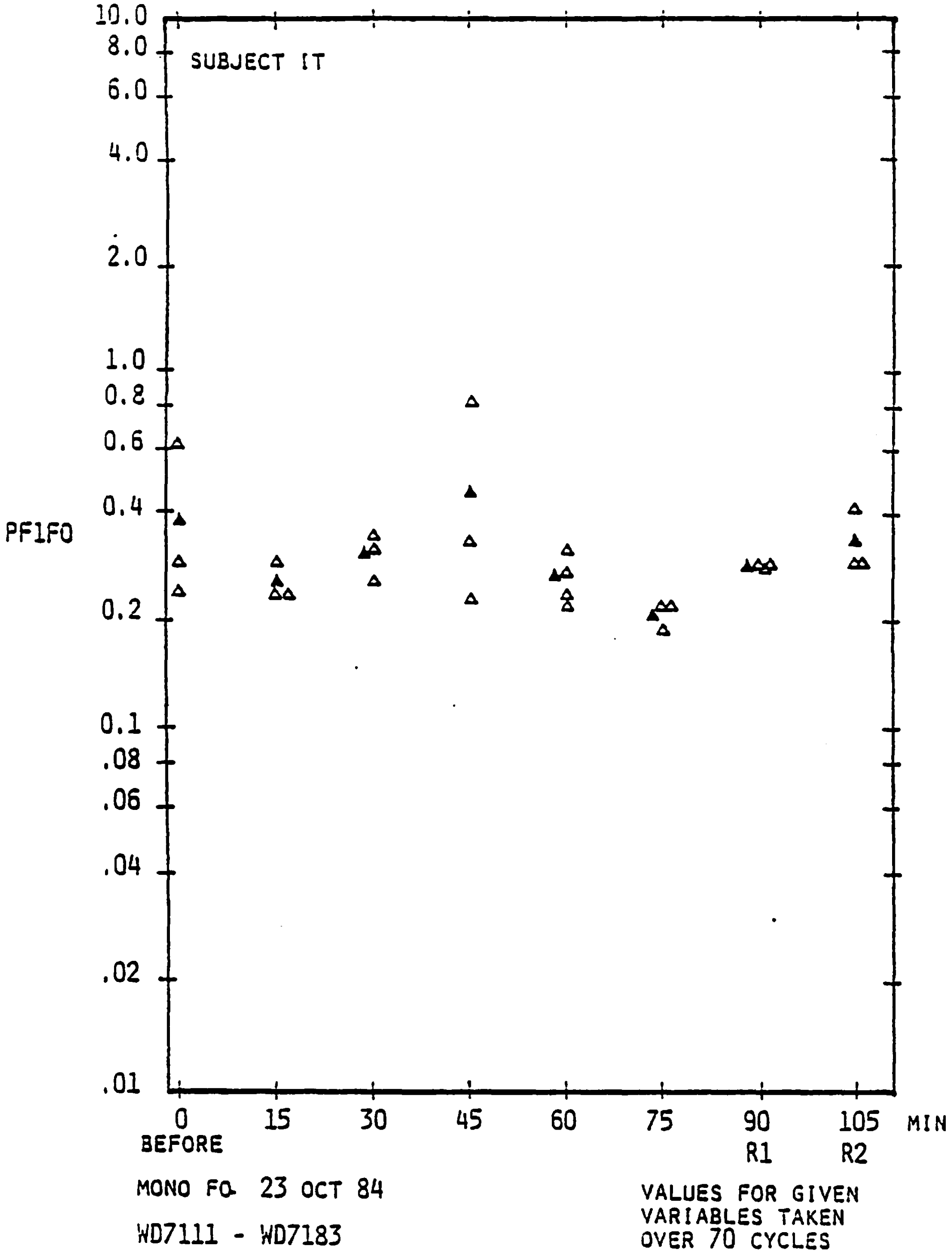


Figure 13. Jitter (PF1F0, frequency perturbation) values for subject IT for the MONO PITCH READING task. The filled triangles refer to the means of the three values (open triangles) at each of the specific 15 minute intervals.

are due to waveform changes of the acoustic cycles, rather than period or amplitude changes.)

Perceptual Ratings

Six experienced raters listened to the 2-second /a/ phonations of both subjects for the monotone reading experiment. The 6 raters were 2 speech pathologists who frequently work with voice cases, 1 theater voice coach (who also was subject BR), and 3 graduate students in speech pathology, 2 of whom are trained singers. The auditors heard each token twice. The ratings for the three tokens at each recording time were averaged across all raters so that one number represented the rating of voice quality at each interval of the experiment. The average intrajudge reliability (i.e., each listener against self) was $r = 0.62$ for the male subject and $r = 0.52$ for the female subject (using the Pearson product-moment correlation). The range of these reliabilities was 0.37 to 0.85 for the male and 0.39 to 0.61 for the female. All but two reliabilites were significant (refer Tables 1 and 2). The intraclass reliability (Ebel) for the "mean reliability for one rater" was near zero for both subjects. The Ebel reliability for mean ratings from the 6 raters (i.e., the reliability of the mean of 6 ratings for each recording time) was fairly good for the female subject (0.65) but also near zero for

Table 1.

SUBJECT BR

INTRAJUDGE RELIABILITY (LISTENER WITH SELF)

LISTENER	R*	T	SIG**
LR	.57	3.254	S
JA	.56	3.170	S
BR	.50	2.708	S
NK	.51	2.781	S
KB	.39	1.986	NS
FB	.61	3.611	S

* PEARSON PRODUCT - MOMENT CORRELATION

** T_{05} = 2.074, 22DF, TWO-TAILED

AVERAGE R = 0.52

Table 2.

SUBJECT IT

INTRAJUDGE RELIABILITY (LISTENER WITH SELF)

LISTENER	R*	T	SIG**
LR	.67	3.934	S
JA	.37	1.736	NS
BR	.54	2.736	S
NK	.85	7.033	S
KB	.49	2.450	S
FB	.80	5.812	S

* PEARSON PRODUCT - MOMENT CORRELATION

** T_{05} = 2.093, DF = 19, TWO-TAILED

AVERAGE R = 0.62

the male subject (refer Table 3). Thus, although each rater was typically fairly consistent with himself or herself, as a group the ratings do not appear strongly reliable.

The subject's ratings of their own voices and the mean auditors' voice quality ratings over time are shown in Figure 14. The subjects' voice self evaluations are different. Subject BR indicated that her voice became slightly worse after the first fatiguing reading task, then improved slightly over the next 2 reading tasks, and became considerably worse over the next two, and finally improved before the end of the second rest period to a level worse than at the beginning of the experiment. Subject IT indicated a much different perception of his change in voice. He indicated a continual worsening of his voice from the first through the fifth reading task, with an improvement after the rest periods to a level worse than before the experiment began.

There appear to be some parallels between the subjects' own voice ratings and the auditors' ratings of voice quality. Figure 14 shows that for subject BR, both sets of perceivers indicated some improvement in voice near the beginning third of the experiment, both indicated a worsening effect during the middle third of the experiment, and both indicated a slight improvement toward or at the end after vocal rest. For Subject IT, both subject and auditors tended to indicate a worsening

Table 3.

INTRACLASS RELIABILITY (EBEL)

(AN "AVERAGE INTERCORRELATION")

	MEAN RELIABILITY FOR ONE RATER	RELIABILITY FOR MEAN RATINGS FROM THE 6 RATERS
SUBJECT BR	-0.04	0.646
SUBJECT IT	-0.001	-0.008

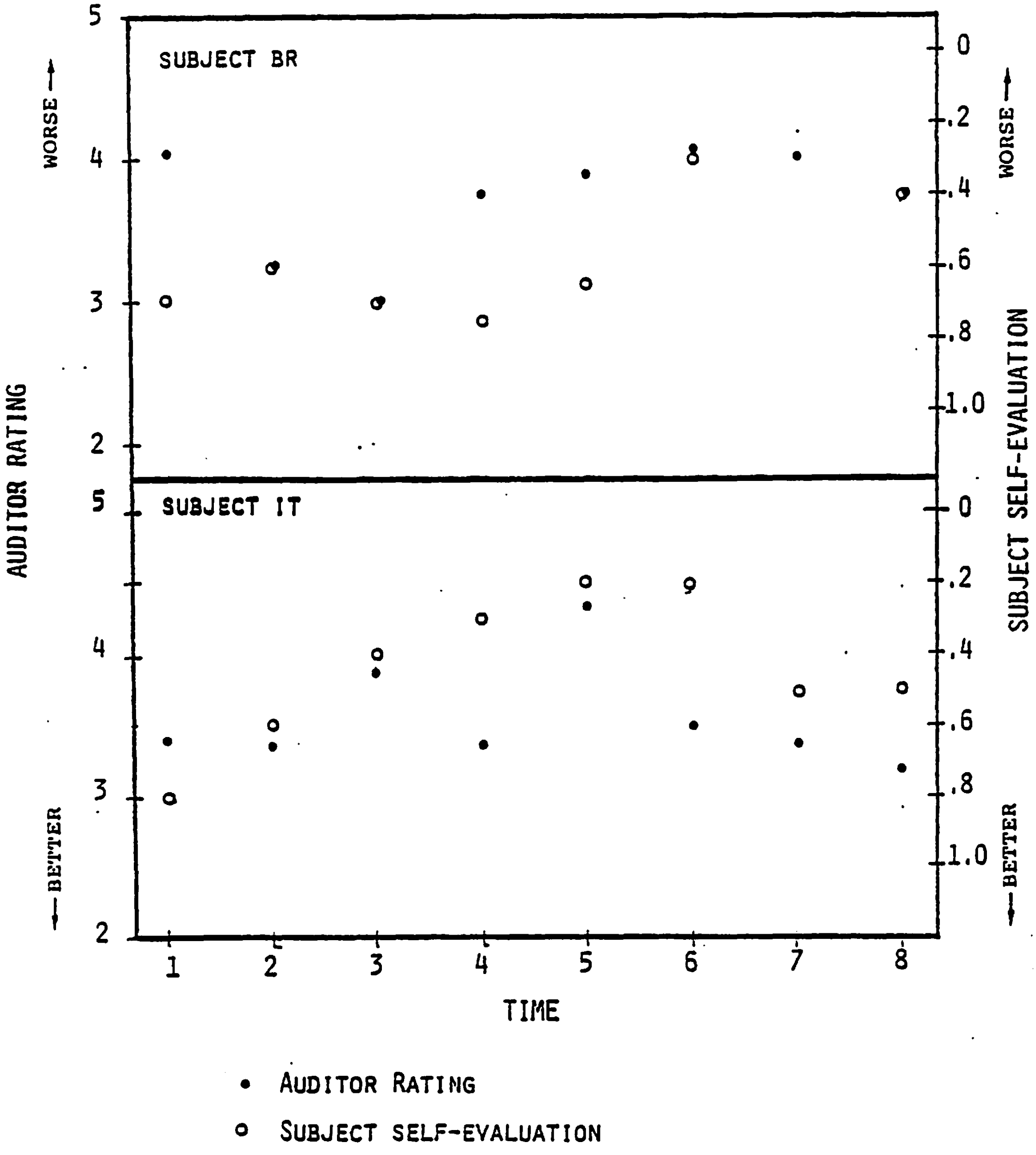


Figure 14. Mean auditor ratings of voice quality and subjects' voice self-evaluation ratings for the MONO PITCH READING tasks.

voice throughout the first half of the experiment, and both indicate an improvement in the second half through the rest periods.

Medical Findings

A fiberoptic examination of the larynx of the two subjects was performed prior to the first Diagnostic Task recordings and following the last Fatiguing Task (during the first rest period). The findings are shown in Table 4.

These results indicate that all three experiments were accompanied by changes in the appearance of the vocal folds between the beginning and the end of the experiment. Slight swelling of the laryngeal tissue was common. Slight bowing of the vocal folds was present at the end of both the MONO pitch experiments. For all three experiments, the conversational voice quality was also noted to have deteriorated. It is noted that the laryngologist making the medical remarks also made the video tapes, and thus was aware of the sequence of laryngeal images.

DISCUSSION

The acoustic analyses do not appear to show meaningful relationships across time for the present vocal fatiguing tasks. These measures are not sufficiently sensitive to vocal changes

Table 4. Medical remarks from fiberoptic examinations of the larynx.

SUBJECT BR, FREE PITCH

- PRETEST EXAMINATION:
  - PRESENT:
    - SLIGHT EDEMA LEFT TRUE CORD
  - SLIGHTLY ROUGH VOICE QUALITY
- POSTTEST EXAMINATION
  - PRESENT:
    - SLIGHT SWELLING OF INTERARYTENOIDS
    - INCREASED SWELLING OF LEFT TRUE CORD
    - MUCUS STRANDING PRESENT
  - SLIGHTLY HOARSE VOICE QUALITY

SUBJECT BR, MONO PITCH

- PRETEST EXAMINATION:
  - PRESENT:
    - MINIMAL PARADOXICAL MOVEMENT
    - LARYNX NORMAL
- POSTTEST EXAMINATION:
  - PRESENT:
    - SLIGHT PARADOXICAL MOVEMENT.
    - SLIGHT BOWING OF FOLDS
  - ROUGH AND BREATHY VOICE QUALITY

SUBJECT IT, MONO PITCH

- PRETEST EXAMINATION:
  - PRESENT:
    - POST AND MID 1/3 SWELLING OF TRUE FOLDS
    - SLIGHT EDEMA OF TRUE FOLDS
    - MUCUS STRANDING
    - ADHERENT MUCUS
  - GOOD VOICE QUALITY
- POSTTEST EXAMINATION:
  - PRESENT:
    - ADHERENT MUCUS ON RIGHT FALSE FOLD OR VENTRICLE WITH SWELLING
    - SLIGHT PARADOXICAL MOVEMENT ON INSPIRATION AND EXPIRATION
    - SLIGHT ANTERIOR CHINK
  - SLIGHTLY ROUGH VOICE QUALITY

produced in this way, or the procedure did not produce vocal instabilities that are manifested in these measures. Perhaps these measures are appropriate, but not obtained in an appropriate manner. For example, using 70-75 cycles to obtain the measures may be inadequate. Using many more cycles may reduce the variability of the measures from token to token. This is not likely, however, because we have found assymptotic behavior near 100 cycles for subject IT. Alternatively, a more sensitive measure may be to take a window of a smaller number of cycles, say 15, and obtain these measures for this window size across the entire 2-seconds of the /a/ phonation. This may reveal important low frequency perturbations and mean and standard deviations. This method has recently been used with some success for aging voice and neurological diseased voice studies by Ramig in our laboratory (see the three papers in Section C of the DCPA - Research Report).

Another severe limitation of the analysis is now understood to be the number of tokens obtained after each of the fatiguing tasks. The number of tokens needs to be larger to determine a strong central tendency and dispersion so that tests for significant differences across time can be obtained. We were basically following the lead from earlier research on perturbations which typically involved only one or two tokens per condition. (It should be noted, however, that Horii (1979) found mean jitter sizes (in msec) differing by a factor of approximately 2 across two tokens. We found jitter variations by factors of 4 to 5 for some specific conditions.)

Although we feel that the variability of perturbation values is generally too large at each time point to draw conclusions across time, we might draw attention to certain aspects relative to the prior studies cited in the Introduction section. The figures giving mean perturbation values (Figures 2, 6 and 10) consistently indicate a decrease in shimmer between the first two recordings, which were taken before the experiment began and following the first 15 minutes of loud talking. In line with this, both subjects reported a worsening of the voice between these two times (Figure 14). These data may reflect a deterioration of voice, which corresponds well to subjects' comments in the Sherman and Jensen (1962) study. It is noted, however, that the auditor ratings did not support this "worsening" trend at the beginning of the experiment (Figure 14).

Also, between the recording following the last Fatiguing Task reading (time 6 on Figures 2, 6 and 10) and the recording after the first rest period (time 7), the mean shimmer consistently increased, indicating a possible deterioration of voice. This may correspond strongly to the finding by Sherman and Jensen (1962) that the rating of harshness became greater following the vocal rest period. These corroborations do not, however, correspond to the perceptual ratings of this study (Figure 14), where both auditors and subjects indicated an improvement in voice or voice quality.

The inconsistency of how different subjects tracked their voice changes over time in the Sander and Ripich (1981) and Neils and Yairi (1984) studies was also a finding here. As shown in Figure 14, subject BR felt here voice deteriorated and then improved before again deteriorating, whereas subject IT felt his voice monotonically deteriorated until the end of the experiment.

The perceptual ratings by both the subjects of his or her voice and by the auditors indicate that there are factors in the production of the voice and in the sound of the voice that apparently were not being picked up by the acoustic measures. Additional or different measures are needed to better reflect the changing physiological function of the voice in these tasks.

Electroglottography may reveal significant changes in the glottographic waveforms corresponding to "vocal fatigue". Temporal trend analysis may be helpful if enough time intervals are obtained. This could be accomplished by incorporating frequent recordings of appropriate tokens perhaps even throughout the fatiguing task. Other relevant physiological measures might also be incorporated. Certainly electromyography should be considered.

The medical findings strongly indicated that specific changes had taken place between the beginning and end of the experiments. The increased swelling of tissue and slight anterior chink surprisingly did not result in important acoustical changes as derived in this analysis. The medical findings are supported by the perceptual ratings of the subjects in that the subjects reported their worse voice conditions at the end of the fifth Fatiguing Task, immediately after which the post medical test was performed.

CONCLUSIONS

This pilot study examined vocal fatigue in two subjects who were asked to read very loudly over an extended period of time. Recordings of /a/ productions and the subjects' perceptions of their voice were made every 15 minutes. Fiberoptic examinations were made before the experiment and following the last fatiguing period. The main conclusions are the following:

1. The results from this study parallel those from other studies regarding the large variability of subject response to

the fatiguing experiments and the lack of meaningfulness of the acoustic measures.

2.The subjects of this study were relatively well-trained vocally. This may have narrowed the range of vocal change compared with the range expected from subjects who are not as well trained.

3. The acoustic analyses (jitter, shimmer, harmonics-to-noise ratios) did not reveal significant vocal changes over time. There is a need for more sensitive measures and procedures to derive an acoustic/physiological formula to describe and predict consequences of prolonged vocal use.

4. Auditor vocal quality ratings bore some similarity to the changes in the subjects' voice self evaluations over time. This implies that there are voice changes perceived by the subjects that are indeed present in the acoustic signal.

5. Fiberoptic observations of the laryngeal tissue indicated that there were changes between the beginning and end of the fatiguing tasks, most noticeably a slight increase in tissue swelling and a slight bowing of the vocal folds.

For future projects, we suggest that greater attention be paid to paradigms that have greater control of the variables involved in vocal fatigue. For example, models of muscle fatigue (Bigland-Ritchie & Woods, 1984) may be applied for laryngeal function by choosing vocal tasks (such as very high pitch and very pressed phonation) that may tax isolated muscle groups. Also, there is a need to isolate fatiguing episodes to the larynx, while trying to avoid confounding influences of general body fatigue. Also, physiological factors including EMG should be considered in order to examine more closely neurophysiological function during vocal fatigue that may parallel acoustic and perceptual changes.

There is a need to determine the differences between vocal use, vocal overuse, vocal misuse, and vocal abuse over both short term and long term periods. The physiological reactions to these may be different, and thus also perhaps the relevant diagnostics and therapies. These studies need to be pertinent not only for normal and pathological voice conditions, but also relevant to the professional voice user such as the singer, actor and public speaker. The present pilot study has been an attempt to examine the voice under prolonged use under conditions that approach those required of the stage actor.

ACKNOWLEDGEMENTS

The authors would like to thank Vernon Vail, Julie Artilgliere, Sandra Anderson, Judy Coe and Kim Sheplor for their help in data

analysis, and Dr. Lorraine Ramig for her suggestions on the perceptual analysis. Portions of this paper were presented at the Eighth Midwinter Research Meeting of the Association for the Research in Otolaryngology, Clearwater Beach, FL, Feb 3-7, 1985.

REFERENCES

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

Horii, Y. (1979). "Fundamental frequency perturbation observed in sustained phonation". J. Speech Hear. Res. 22, 5-19.

Neils, L.R. and Yairi, E. (1984). "Effects of speaking in noise on vocal fatigue and recovery". ASHA 26, 182 (abstract).

Sander, E.K. and Ripich, D.E. (1981). "Vocal consequences of loud talking". ASHA 23, 745 (abstract).

Sander, E.K. and Ripich, D.E. (1983). "Vocal fatigue". Ann. Otol. Rhinol. Laryngol. 92, 141-145.

Sherman, D. and Jensen, P. (1962). "Harshness and oral reading time". J. Speech Hear. Dis. 27, 172-177.

ACTORS' PERCEPTION OF PERFORMANCE-RELATED VOCAL FACTORS

Bonnie N. Raphael, Ph.D.
The Denver Center for the Performing Arts

Ronald C. Scherer, Ph.D.
The Denver Center for the Performing Arts

ABSTRACT

There are a number of interdependent factors which are important to maintaining an actor's vocal health and to fulfilling management goals in professional theatre. These factors include the actor's perceptions and expectations of his/her voice under various conditions; the actor's history of training, medical treatment, and performance; and the demands being made on the voice by the nature of the work which he/she does on an ongoing basis.

This study explored the auditory and somatic perceptions which the actor has concerning his/her voice when the voice appears to be working optimally and when it is tired or otherwise deficient. In addition, the study elicited both working conditions and other factors which appear to affect the professional actor's voice.

The large number of descriptive terms employed by the actors participating in this study indicated their relatively high degree of awareness of the status of the voice. A closed-ended questionnaire, designed on the basis of the results obtained in this study and administered to the same subjects, will help to further document and quantify this information. These perceptual studies should create a meaningful vocabulary for use among a number of health and vocal training professions, and should serve to acquaint theatre management with conditions which both benefit and may prove detrimental to the vocal performance of the professional stage actor.

INTRODUCTION

Maintaining the vocal health of its actors is essential to the performance and management goals of the professional repertory theatre company. The many factors which may influence the actor's vocal condition at any given time include:

1) history of training, medical treatment and performance,

2) the particular demands necessitated by casting, rehearsals and performance,

3) the occurrence and severity of vocal problems and fatigue, and

4) the actor's perception and expectations when the voice is functioning either optimally or deficiently.

Documentation of the interrelationships among these factors does not appear to exist in the literature, and consequently neither do scientifically based recommendations which might serve to maximize vocal well-being in professional actors. A preliminary, long-range study has been undertaken at the Denver Center for the Performing Arts in order to explore these interrelationships. We anticipate that the results obtained in this study will help the actor to better understand and control conditions affecting the voice, will help voice therapists and medical personnel to better understand both the vocal vocabulary used by actors and the theatre demands related to vocal usage, and will help theatre management to understand a number of conditions related to increased vocal stress. In addition, the results will help to identify specific areas for subsequent research on factors affecting the professional voice user.

This progress report emphasizes a consideration of factor number four above, that is, the actor's perception and expectations of his or her own voice when it is functioning either optimally or deficiently.

QUESTIONNAIRES AND DATA COLLECTION

In the fall of 1983, a pilot questionnaire asking about the voice was distributed to thirty-one members of the resident acting company (see Appendix A). Nineteen questionnaires were returned. This preliminary questionnaire was designed with a twofold purpose:

1) to assess how professional actors in repertory theatre perceive and describe their voices when those voices are working optimally as well as when they are fatigued or not at their best, and

2) to extract a list of terms describing the voice to be used for a more intensive study of both optimal vocal performance

and vocal problems or fatigue in stage actors.

A continuation of the study was undertaken during the 1984-85 season. A number of written questionnaires were designed by the researchers in consultation with Dr. Wilbur James Gould and Dr. Raymond P. Wood, Jr. (see footnote 1), otolaryngological physicians whose respective practices include considerable consultation with professional voice users. These questionnaires included the following:

1) Subject/Performer History Form

An extensive history form was designed to elicit information regarding state of health, history of childhood and respiratory diseases, habitual use of drugs (tobacco, alcohol, recreational), vocal habits, extent and type of training, history of vocal problems or difficulties, etc. (see Appendix B). This questionnaire was given to most of the members of the 1984-85 acting company at the beginning of the season.

2) Consent Form and Open-Ended Questionnaire

Those members of the acting company who returned a completed history form and who indicated an interest in participating in this study were then asked to complete a Consent Form (Appendix C) and an Open-Ended Questionnaire dealing with their perception of both the feeling and the sound of the voice (Appendix D). Thirty-six completed sets of data were returned by early November, 1984.

Responses elicited by both the 1983 and the 1984 Open-Ended Questionnaires are summarized in the following tables. The terms contained in all the tables which follow were provided by the actors.

3) Closed-Ended Questionnaire

Based on the descriptive terms used by at least two actors in the 1983 questionnaire and at least three actors in the 1984 questionnaire, a closed-ended questionnaire was designed to identify more specifically and to attempt to quantify factors which might most accurately describe the way in which the actor perceives his or her voice when it is working both optimally and poorly. A report and discussion of those responses will be presented at a later time.

4) Research Journal and Related Production Data

In order to relate the above data with the ever-changing demands of a very heavy production schedule, an ongoing record is being compiled, of:

a) rehearsals, performances, and Conservatory class schedules for the entire season (Some of the Company actors also either teach classes or are advanced students in the Conservatory.), and

b) the presence, duration, and degree of severity of any voice-related difficulties occurring during the season. By using these basic forms, data for thirty-six Company members are being collected over a period of time ranging from five to eight months, depending on the length of contract for each individual actor. This progress report will concentrate specifically on data generated by the open-ended questionnaire. Later papers will be concerned with the closed-ended questionnaire, the journal, and related history and production information.

THE ACTOR'S PERCEPTION OF THE VOICE: DISCUSSION

The actors' responses to the open-ended questionnaire indicate that they appear to be fairly sensitive to and aware of the condition of the voice under various environmental, medical, and performance conditions. An analysis of the open-ended questionnaire responses yielded information in five different areas, which will be discussed here individually.

1) Vocal, Physical, and Respiratory Responses to Denver (Table 5)

Among the thirty-six actors who completed the 1984 questionnaire, nine had been living in the Denver area before the start of the 1984 season and therefore reported no particular responses to the relocation. Of the actors who did experience some reaction to Denver, negative occurrences far outweighed

TABLE 1: HOW THE VOICE FEELS WHEN IT IS AT ITS BEST

1983 (N=19)		1984 (N=36)	
DESCRIPTIVE TERM	Nbr. of Responses	DESCRIPTIVE TERM	Nbr. of Responses
Unaware of voice and mechanism	6	Effortless	13
Openness	5	Sympathetic vibration	11
Relaxed	5	Voice coming from the whole body	9
Power from the mid-section	5	Relaxed, loose	9
Vibration	4	Unaware of voice or throat	7
Energetic	4	No catching or tightness in the throat	7
Freedom	3	Free throughout the range, flexible	4
Centered	3	Connected to the emotions	4
Breathe more deeply	3	Rested	4
Effortless projection	3	Alert, energized	4
Released	2	Strong	3
Responsive	2	Centered	3
Strong	2	Moist	2
Generally healthy and fit	1	Breathing easy	2
Good projection	1	Easily warms up	2
Lively	1	Healthy	1
Communicative	1	Don't fatigue as easily	1
Calm	1	Warmth in the throat	1
Happy	1		
Confident	1		

TABLE 2: HOW THE VOICE SOUNDS WHEN IT IS AT ITS BEST

1983 (N=19)		1984 (N=36)	
DESCRIPTIVE TERM	Nbr. of Responses	DESCRIPTIVE TERM	Nbr. of Responses
Resonant throughout the range	5	Resonant	11
Clear	4	Clear	6
Easily controlled	4	Full, mellow	5
Clean	3	Free, open	4
Wide range	2	Clean, no cracking	3
Fluid ("a flowing river")	2	Wide range	3
Focussed sound	1	Registers blend easily	3
Smooth	1	Bright	3
Mellow	1	Round	2
		Deeper	2
		Loud	2
		Ping	1
		Tenorish	1
		Young	1
		Vulnerable	1
		Pitches very accurate	1

TABLE 3: HOW VOICE FEELS WHEN IT IS TIRED OR NOT WORKING PROPERLY

1983 (N=19)		1984 (N=36)	
DESCRIPTIVE TERM	Nbr. of Responses	DESCRIPTIVE TERM	Nbr. of Responses
Effort to produce	8	Throat fatigue	5
Physical fatigue	6	Scratch or tickle in throat	5
Strain	6	Burning in throat	5
Tightness	5	Awareness of throat and larynx	4
Trapped	4	Effort to project	4
More aware of voice and mechanism	3	Tense	4
Lethargic, unresponsive	3	Constricted	4
Tense	3	Back of throat catches	3
Soreness	3	Voice feels heavy	3
Pain	2	Rough throat	3
Difficulty swallowing	2	Have to clear a lot	3
Requires more air	2	Range reduction	3
Requires more concentration	2	Dry throat	3
Absence of vibration in body	1	Have to force	2
Loss of control	1	Neck and shoulder tension	2
Inflexible	1	Thickness	2
Large tongue	1	Out of breath	2
Tickle deep in throat	1	Inflexible, unresponsive	2
Gasping	1	Clogged	1
Cough	1	Unsociable	1
Burning in palate	1	Coated cords	1
Excessive phlegm	1	Poor concentration	1
Excessive dryness	1	Takes a while to warm up	1
Ache	1		
Compensation by rest of body	1		
"Like running an engine and constantly having to tune it at the same time"	1		

TABLE 4: HOW THE VOICE SOUNDS WHEN IT IS TIRED OR NOT WORKING PROPERLY

1983 (N=19)		1984 (N=36)	
DESCRIPTIVE TERM	Nbr. of Responses	DESCRIPTIVE TERM	Nbr. of Responses
Range reduced	5	Cracks	11
Less volume in lower register	4	Harder to produce and sustain	10
Breaks or catches	3	Raspy	9
Pinched	3	Habitual pitch lower	6
Rasping	5	Colorless, dull	5
Grey	2	Constricted	4
Dull, flat tone	2	Hoarse	3
Rough edge	2	Harder to control	3
Hoarse	2	Higher pitches hard	3
Gravelly	2	Flat	3
Less vibration	2	Gravelly	2
Scratchy	2	Breathy	2
Lower pitch	2	Throatiness	2
Harsh	1	Tense	1
Fog	1	Strained	1
High pitches forced	1	Low register husky	1
Thinner	1	Loss of upper range	1
Loss of high pitches	1	Loss of middle register	1
Frog	1	Weak	1
More nasal	1	Muffled	1
		Huskiness	1
		Nasal	1

TABLE 5: VOCAL, PHYSICAL, OR RESPIRATORY RESPONSES TO COMING TO DENVER (1984 Questionnaire)

N=36

DESCRIPTIVE TERM	Nbr. of Responses
Shortness of breath	16
Tire more easily	10
Slight dryness	9
Dry nasal passages at night	7
Dry throat in mornings	5
Voice is better	2
Upper range not as loose as usual	2
Nasal congestion	2
Chapped lips	2
Sore throat	2
Don't fatigue as much	1
No allergies	1
Low habitual pitch in mornings	1
Voice cracks occasionally	1
Congestion	1
Laryngitis	1
Headaches	1

TABLE 6: WORKING CONDITIONS OR VOCAL DEMANDS WHICH MAKE THE VOICE VULNERABLE TO FATIGUE OR STRAIN (1984 Questionnaire)

N=36

DESCRIPTIVE TERM	Nbr. of Responses
Dress rehearsals and preview performances	8
Long working hours of talking	4
First week of rehearsal	4
Sustained high-energy vocal production	3
Technical rehearsals	3
Large role	3
Working over a great deal of background noise	2
Getting off book	1
Doing two shows per day	1
Use of demanding character voice	1
Use of accent or dialect	1
First time onstage in each show	1
Striving for new emotional levels	1
Screaming, shouting	1
Having to sing in addition to acting	1
Outdoor performance	1

TABLE 7: OTHER FACTORS AFFECTING VOCAL CONDITION (1984 Questionnaire)

N=36

DESCRIPTIVE TERM	Nbr. of Responses
Inadequate sleep detrimental to voice	14
Lots of sleep beneficial to voice	11
Voice work or warm-up daily beneficial to voice	7
Winter cold and dryness necessitate more warm-up	7
Cutting down on dairy products good for sound.	6
Smoking or inhaling others' smoke is detrimental	5
Healthy, light diet is beneficial to voice	4
Lots of water beneficial to voice	4
Seasonal allergies affect voice	4
Alcohol makes voice worse or throat raw	3
Exercise is beneficial to voice	3
Menstrual cycle affects voice	3
Nerves and upset are detrimental to voice	3
Steam or humidity good for voice	3
Cold, damp air is detrimental	3
Nuts and chocolate adversely affect sound	2
Going from cold to hot or vice versa	2
One glass of wine with dinner beneficial	2
State of health directly related to voice	1
Matinees are easier on voice than evening performances	1
Late-night drinking is bad the next day	1
Not cooling down after strenuous performance is bad	1
Hot and warm weather makes voice loose	1
Night air after a demanding performance is bad	1
Fog is beneficial	1
Dry air is detrimental	1
Full stomach is bad for the voice	1
Hunger creates problems with the voice	1
Cold desserts with dinner are bad for voice	1

positive ones and fell into the following categories:

a) Sixteen of the actors responding experienced a shortness of breath when they first came to Denver. Five of them reported that this occurrence was temporary, lasting only a week or two. A similar number of them (5) reported the that the shortness of breath was noticeable only when they were exerting themselves physically.

b) Ten of the actors responding felt more fatigued in Denver than they had felt where they lived before. A number of them (4) reported that this occurrence was temporary and seemed to last only a week or two.

c) A number of the negative sensations experienced by newly-arrived actors appear to be associated with the relatively low humidity levels in Denver. Nine actors reported a very dry throat upon waking up in the morning. Seven actors reported that the nasal passages in particular felt very dry, and two reported chapped lips.

d) Medical symptoms which may be related to the Denver relocation included congestion (1), headaches (1), nasal congestion (2), sore throat (2), and laryngitis (1).

Only two of the actors reported a positive response to the Denver relocation. They felt that their voices sounded better in Denver than they had elsewhere. One reported that his voice was not fatiguing as much as it had been in California, and one reported that Denver seemed to be better for his particular allergies than Los Angeles had been.

In a future study, it would be of value to compare the above responses with those of a general population relocating into the Denver area.

2) Feelings associated with Optimal Voice Production (Table 1)

In both the pilot and the more extensive open-ended questionnaires, members of the acting company were asked how they feel when the voice is at its best. Vocabulary used to respond to this and other questions seems to be related directly to both the extent and the type of training which each actor had experienced in the past. For example, some of the actors interpreted this question to refer to mental feelings or emotional responses to their own voice; others dealt with physical sensations. Many actors offered both types of response.

a) General Sensations

Nine actors described themselves as alert or energized when the voice was at its best, and four said they felt generally rested. Two actors described themselves as feeling generally healthy and fit, one adding that the voice didn't fatigue as easily when it was in its best shape. Six actors described themselves as feeling "centered", one as confident, one as happy, and one as calm. Five more said that the voice felt strong when it was at its best.

b) Specific Sensations

Thirteen actors remarked that when the voice was at its best, they were unaware of it and of the vocal mechanism, seven more adding that there was no catch or tightness in the throat as they spoke. One felt a warmth in the throat itself, and two others reported feeling moist in the throat. Finally, two of the actors reported that the voice warmed up easily when it was at its best.

c) Feelings Associated with Voice Production

Sixteen of the actors remarked that when the voice was at its best, it felt released, open, flexible, responsive, or free throughout the range. Fifteen were aware of a great deal of vibration, with nine more remarking that the voice seemed to be coming from the entire body. Five were aware of a feeling of power from the midsection, one feeling that projection was really good. Five felt that they were able to breathe more deeply or easily, and sixteen of them sensed an effortlessness or an effortless projection. One remarked that he felt highly communicative, and four others felt a direct and unobstructed connection to their emotions when the voice is at its best.

d) Summary

For these actors, the best voice production is associated with feelings of two basic types: physical sensations, and mental or emotional responses to (or interpretations of) these sensations by the actor. In general, best voice production appears to be associated with a feeling of good health and increased energy. The actor reports being unaware of the voice or the mechanism. The voice feels released, open, flexible, responsive, free, or as if it were emanating from the whole body. And there seems to be no sense of excessive effort connected to the act of communication.

3. The Sound of the Voice at Its Best (Table 2)

Acting company members were also asked to describe the sound of the voice when it is at its best. Generally, their responses described the sound itself, the resonance of the voice, and the relationship between these and vocal range.

a) Characteristic Sound

Ten actors described their best voice as sounding clear. Six more said that the sound was clean, with no cracking. Four others described the sound as free or open. Three referred to the sound as fluid or smooth, one of them describing it as a "flowing river".

b) Resonance

Sixteen of the actors called their best sound resonant, five indicating that resonance extended throughout the range when the voice is at its best. Twelve described the best tone as full, mellow, round, or deeper than usual. Seven others described a sound which was focussed, or had a ping, or was bright, tenorish, or young.

c) Range

Five of the actors felt that their best voice functioned over a wide range. One observed that when he was in his best voice, his pitches were very accurate. Three said their registers blended very easily, and three others remarked that their high pitches were clearer and easier to sustain when the voice was at its best.

Four of the actors felt that the voice is easily controlled when it is working well. Two others remarked that the best voice was loud. One actress remarked that her best voice made her feel a bit vulnerable and that this was just a bit scary to her.

d) Summary

For these actors, the best voice seems to be characterized by the perception of a clear, clean tone, free of cracking and resonant throughout a larger-than-usual range. To the actor, the sound seems full, mellow, or deep.

4. Feelings Associated with Tired or Less-than-Optimal Voice (Table 3)

A large number of physical sensations and emotional responses were provided by the actors to describe the way the voice felt when it is tired out by strenuous use or is not working at its best. One actor very descriptively compared his experience to "running an engine and constantly having to tune it at the same time". The feelings described by the actors related both to physical sensations and to an awareness of effort when producing voice.

a) Physical Sensations

Six of the actors described themselves as generally physically fatigued when the voice was not working at its best, and five others described the fatigue as being located specifically in the throat. Seven of the actors remarked that they were more aware of the voice and of the mechanism when the voice was tired out by strenuous use.

Six actors felt a scratch or a tickle in the throat; three more described the back of the throat as "catching". Three described the throat as rough, and another six noticed a burning sensation in the throat or on the soft palate. Three described the sensation as a "sore throat" and another three noticed a pain or an ache in the throat. Four observed that the throat was very dry, and one was aware of a cough.

Sixteen of the actors described the feeling in their throat as constricted, tense or tight, and four more described the voice produced by the throat as trapped. One commented that his tongue felt really large and two more detected some difficulty swallowing. Three described a feeling of thickness or of the throat being clogged. Two observed excessive phlegm, another observed that his cords felt "coated" when his voice was very tired, and three actors noticed that they needed to clear the throat a lot more than usual.

Two of the actors observed accompanying neck and shoulder tension at these times, and one noticed that the rest of the body tended to compensate for the lack of ease in the throat area.

b) Vocal Production

Nine actors described the voice as feeling inflexible, unresponsive, lethargic, or heavy, one adding that at such times, it takes a while for the voice to finally warm up. Twenty actors remarked that it was an effort to produce voice or to project, that they felt strain and had to force to get the voice out.

Three noticed a reduction of range and one felt a loss of control when the voice was tired or not working at its best.

Three noticed that they needed to concentrate much more in order to produce voice. Two felt they needed more air at such times, and three felt as if they were out of breath or gasping.

Finally, three actors indicated a sharp reduction in vibration in the body itself, and one observed that at such times, he tends to be very unsociable.

c) Summary

For many of these actors, a feeling of fatigue, either general or confined to the throat area, seemed to accompany a tired voice. A number of the actors are more aware of the voice and the vocal mechanism at these times; specifically, they were aware of some negative sensations in the throat itself and a general sense of constriction or tightness in the area.

The voice emanating from this mechanism felt inflexible and unresponsive. Many of the actors found that it required a great deal of effort to produce sound or to project it effectively.

5) The Sound of the Tired Voice (Table 4)

Responses to this particular question reflected an awareness of effort and some difficulties in control, changes in the quality of the vocal tone produced and in the habitual pitch, and reduction in pitch and intensity ranges.

a) Effort and Unpredictability

Ten of the actors belived the voice was harder to produce and to sustain when it is tired or not working at its best. Three found the voice harder to control at this time. For nine of the actors, the sound seemed constricted, strained, tense, or pinched. Higher pitches are difficult or forced for four of them, and fourteen of the actors are aware of cracking, breaks, or catches in the tired voice.

b) Vocal Quality

Twelve of the actors reported that the sound of the voice is colorless, dull, flat, or grey. Five said they were hoarse, eleven said that they were raspy, and three found their voices to be harsh or to have a rough edge. Four of the actors described their tone as gravelly and two used the term scratchy. Two reported a huskiness, one particularly in the low register. One described a muffled tone, another described a persistent fog (sic) in the voice, and two called the sound throaty. One actor described the voice as thinner, another as weak, and two said the voice got breathy when they are tired. Finally, two actors described the voice as being more nasal when it is not working at its best.

c) Pitch and Range

Eight of the actors observed that their habitual pitch is lower when the voice is tired. Four found less volume in the lower register. Five found that their range is reduced. Two felt they lose high pitches or their upper range, and one described a loss of the middle register.

d) Summary

For a number of the actors, the voice when not working at its best is harder to produce and to sustain. The sound of the voice is either constricted and strained or dull, flat, and grey. For many, there is cracking, breaks, or catches; for others, the sound is raspy or rough.

Habitual pitch seems to be a bit lower, and for some the range is either reduced somewhat or contains some problem areas.

6) Rehearsal and Performance-Related Conditions Affecting the Voice (Table 6)

Actors identified certain times during the rehearsal and performance periods and certain vocal usages which they felt directly affected the voice.

a) Rehearsal and Performance Demands

Without identifying specific times, four actors felt that long working hours of talking, rehearsal and performance have a deleterious effect on the voice. Four others felt that the first week of rehearsal is a particularly vulnerable time for them because of the newness of the situation, the effort which they put into rehearsal at that time, and the mental or emotional confusion they experience when not yet comfortable with what they are doing.

Eight actors felt that dress rehearsals and previews are their worst times, because of the long working hours and tension build-up right before opening night. For three, the days of technical rehearsals are a problem because of the long hours, mechanical tasks, and repetition. It is interesting to note however, that not one actor participating in this study stated that opening night is a time of particular vulnerability for the voice.

For three of the actors, being cast in a large role produces a tendency to overwork and may cause some vocal difficulties. One actor mentioned that the first time onstage (as opposed to working in the rehearsal hall) is difficult. Another observed that getting off book (i.e. learning the lines by heart, putting the script down) is a particularly dangerous time for him. One actor said that two shows per day put stress on his voice, and one found outdoor performances particularly difficult.

b) Particular Vocal Demands

Seven of the actors observed that when they are forced to put forth sustained high-energy vocal production--because of the needs of the role being played or because they are either working over a great deal of background noise or screaming or shouting--their voices suffer consequences. Two actors observed that when they use a demanding "character voice" they tire more easily, and one felt that using a dialect or a foreign accent has the same effect. One observed that striving for new emotional release is taxing for him vocally. For one actress, singing provided more difficulties than acting. One had some vocal or breathing problems as a result of heavy physical exertion (i.e., a strenuous swordfight), one felt that an overuse of hard glottal initiation is invariably responsible for vocal fatigue, and one observed that when he is not breathing properly, the voice is more likely to fatigue both in rehearsal and in performance.

7. Additional Factors Influencing the Actor's Voice (Table 7)

a) Sleep

Twenty-five actors reported either that a lot of sleep is beneficial to the voice or that inadequate sleep is detrimental to it.

b) Diet and Social Habits

Four of the actors observed that drinking a great deal of water is beneficial to the voice, while four actors reported that a healthy, light diet is beneficial. Six actors believed that cutting down on dairy products is good for the sound of the voice, while two actors reported that nuts and chocolate have a detrimental effect, and one reported that having a cold dessert with dinner affects his voice negatively at evening performances.

One actor reported that working on a full stomach is very difficult, and another observed that performing when hungry has a negative effect on her voice.

Three actors reported that the use of alcohol makes the voice worse and one found that late-night drinking produces a bad effect on his vocal work the next day. On the other hand, two actors reported that, in their experience, a single glass of wine with dinner has a beneficial effect on the voice.

Five actors found that smoking or inhaling others' cigarette smoke has a deleterious effect on the voice.

c) Weather and Environment

Seven actors reported that winter cold and dryness is detrimental to the voice. One actor reported that dry air at any time of the year has a bad effect on his voice, and two others reported that going either from cold to hot or from hot to cold environment is deleterious. One actor found that going out into the night air right after a demanding performance affects the way he sounds the next day.

Four actors reported that steam or humidity or rain or the presence of fog is beneficial, although three others felt that cold, damp air is detrimental. And finally, one actor reported that hot and warm weather makes his voice loose.

d) Warm-Up and Conditioning

Seven actors reported that daily voice work or a pre-performance warm-up is beneficial to their vocal performance. Three actors found that a regimen of regular physical exercise is beneficial. One actor found that not warming down gently after a strenuous performance interferes with his vocal effectiveness on the following day.

e) Medical Factors

Two actors reported that their general state of health is always directly reflected in the sound of the voice. Four actors reported that a general state of emotional turmoil or nervousness will definitely affect the voice.

Four actors reported that spring was a problem time because of their allergies. Three women found that their voices are noticeably affected by either pre-menstrual tension or by their menses, and two diabetics reported that their general blood sugar level is directly related to dryness in the mouth and therefore affects the voice as well.

f) Chance

And finally, two actors reported that sometimes the voice sounds and feels good under the worst situations while feeling and sounding "positively awful" after they've done all of the right things!

CONCLUSION

This preliminary report is one part of a more extensive study dealing with professional actors' perceptions of the voice under a number of different conditions.

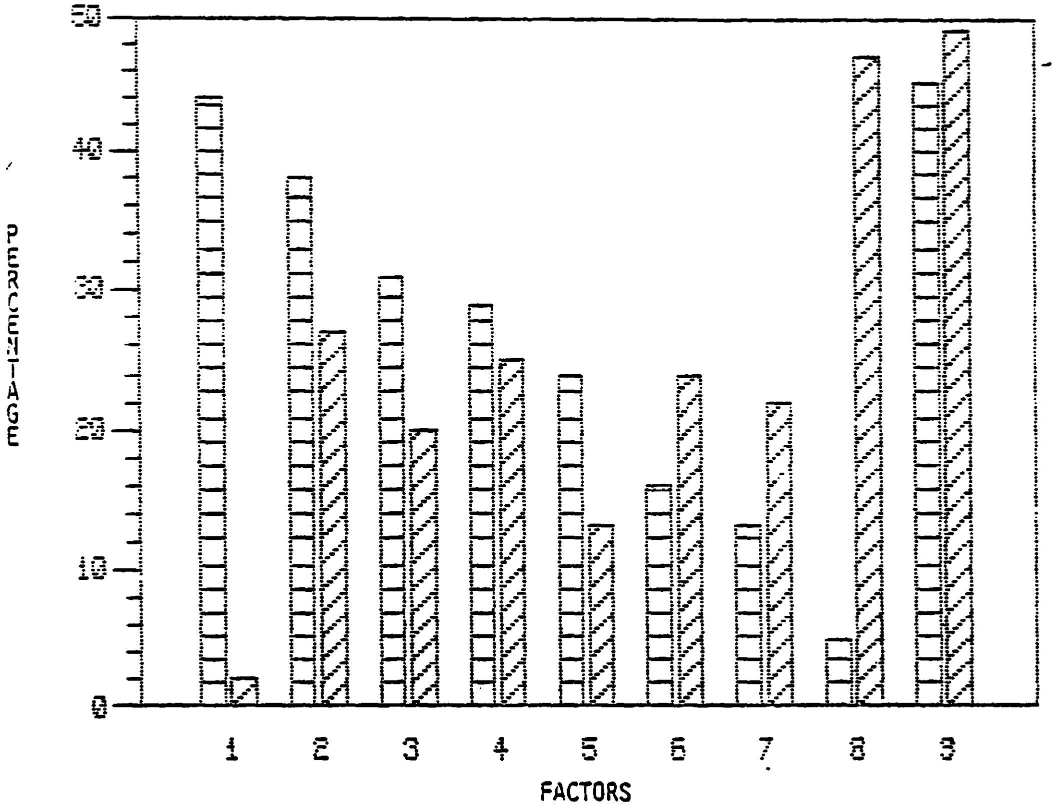
On the basis of actors' answers to the open-ended questionnaires completed during both the 1983-84 season and the 1984-85 season of the Denver Center Theatre Company, a number of conclusions may be formulated. On the basis of the number of actors who responded at length and the large number of descriptive terms they employed in those responses, it may be concluded that:

1. Professional actors working in repertory theatre seem to be relatively aware of the status of the voice, both when it is functioning at its best and when it is tired or deficient.

A. Bar Graph 1 indicates the percentage of actors who offered terms within subsequently deduced catagories

BAR GRAPH 1: THE FEEL OF THE VOICE AT ITS BEST AND WHEN TIRED OR NOT WORKING PROPERLY

[MOST FREQUENTLY-OCCURRING RESPONSES IN SAMPLE OF 55 ACTORS]



SHADING KEY: A. ▤ VOICE AT ITS BEST

B. ▨ VOICE WHEN TIRED OR NOT WORKING PROPERLY

FACTORS KEY:

1. VIBRATION
   - A. SYMPATHETIC VIBRATION, VOICE EMANATING FROM WHOLE BODY
   - B. ABSENCE OF VIBRATION IN BODY
2. RELAXATION
   - A. RELAXED, RELEASED, OPEN, LOOSE
   - B. STRAIN, TENSION
3. FATIGUE
   - A. ENERGIZED, ENERGETIC, STRONG, RESTED
   - B. PHYSICAL FATIGUE, THROAT FATIGUE
4. EFFORT
   - A. EFFORTLESS PRODUCTION AND/OR PROJECTION
   - B. EFFORT TO PRODUCE AND/OR TO PROJECT
5. AWARENESS
   - A. UNAWARE OF VOICE AND MECHNAISM
   - B. MORE AWARE OF VOICE AND MECHANISM
6. FLEXIBILITY
   - A. FREEDOM, RESPONSIVENESS, FLEXIBILITY
   - B. TRAPPED, UNRESPONSIVE, LETHARGIC, INFLEXIBLE, HEAVY
7. CONSTRICTION
   - A. NO CATCHING OR TIGHTNESS IN THROAT
   - B. TIGHTNESS, CATCHING, CONSTRICTION
8. THROAT SENSATION
   - A. MOIST, WARMTH IN THROAT
   - B. TICKLE, COUGH, ROUGHNESS, SCRATCH, DRY, BURNING IN THROAT, SORENESS, PAIN, ACHE
9. OTHER

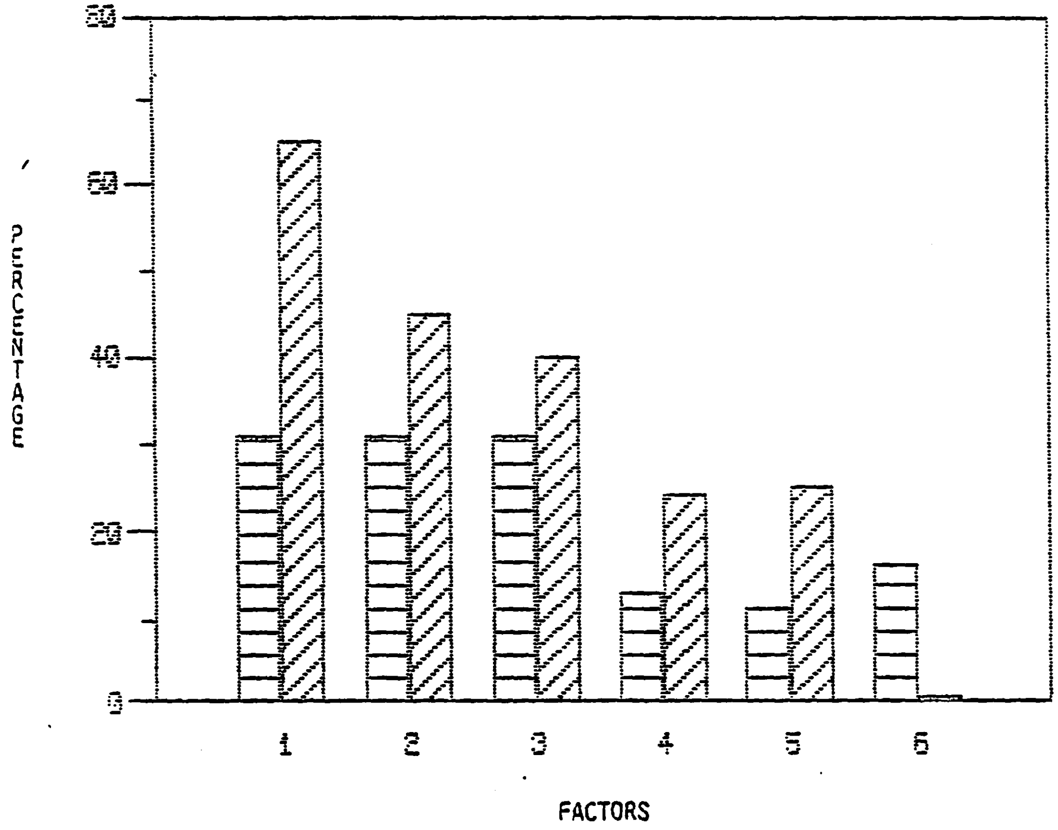
that describe "the feel of the voice at its best" or the feel of the voice "when tired or not working properly". The bars in the graph can be read in the following way: For example, for the first factor "vibration", approximately 44% of the 55 actors indicated that when they felt the voice was at its best, they felt sympathetic vibrations or felt that the voice emanating from their whole body. Furthermore, very few (about 2%) of the 55 actors used words that related to body vibrations to describe the feel of the voice when tired or not working properly, and the comments concerned the absence of such vibrations.

As Bar Graph 1 illustrates, the voice at its best seems to be felt by these actors as accompanied by a great deal of sympathetic vibration, and as relaxed or open as opposed to tense or strained. Many actors sense an effortless production and a feeling of strength and energy. When the voice is tired or deficient, on the other hand, a number of negative sensations are felt in the throat, and effort seems to be needed to produce or release the voice, which is felt as trapped or inflexible.

B. As Bar Graph 2 illustrates, the sound of the voice at its best seems to be perceived by many professional actors as clear, clean, resonant, mellow, or smooth. When the voice is tired or not working properly, on the other hand, a large number of noisy or aperiodic elements are noted, and the

BAR GRAPH 2: THE SOUND OF THE VOICE AT ITS BEST AND WHEN TIRED OR NOT WORKING PROPERLY

[Most Frequently-Occurring Responses in Sample of 55 Actors]



SHADING KEY:

A. ▤ VOICE AT ITS BEST

B. ▨ VOICE WHEN TIRED NOR NOT WORKING PROPERLY

FACTORS KEY:

1. Roughness
   - A. Clear, Clean, Smooth, No Cracking
   - B. Raspy, Rough Edge, Hoarse, Gravelly, Scratchy, Harsh Throatiness, Husky, Fog, Muffled, Frog
2. Resonance/Constriction
   - A. Resonant, Focussed
   - B. Pinched, Constricted, Tense, Strained, More Nasal, Breaks or Catches, Cracks
3. Resonance/Fullness
   - A. Mellow, Full, Bright, Round, Deeper, Ping, Loud, Tenorish
   - B. Grey, Dull, Flat Tone, Less Vibration, Colorless, Weak, Less Volume in the Lower Register, Thinner, Breathy
4. Control
   - A. Easily Controlled, Registers Blend Easily
   - B. Harder to Produce and/or to Sustain and/or To Control
5. Range
   - A. Wide Range, Pitches Very Accurate
   - B. Reduced Range, Lower Pitch, High Pitches Forced or Lost, Low Pitches Husky, Loss of Middle Register
6. Other

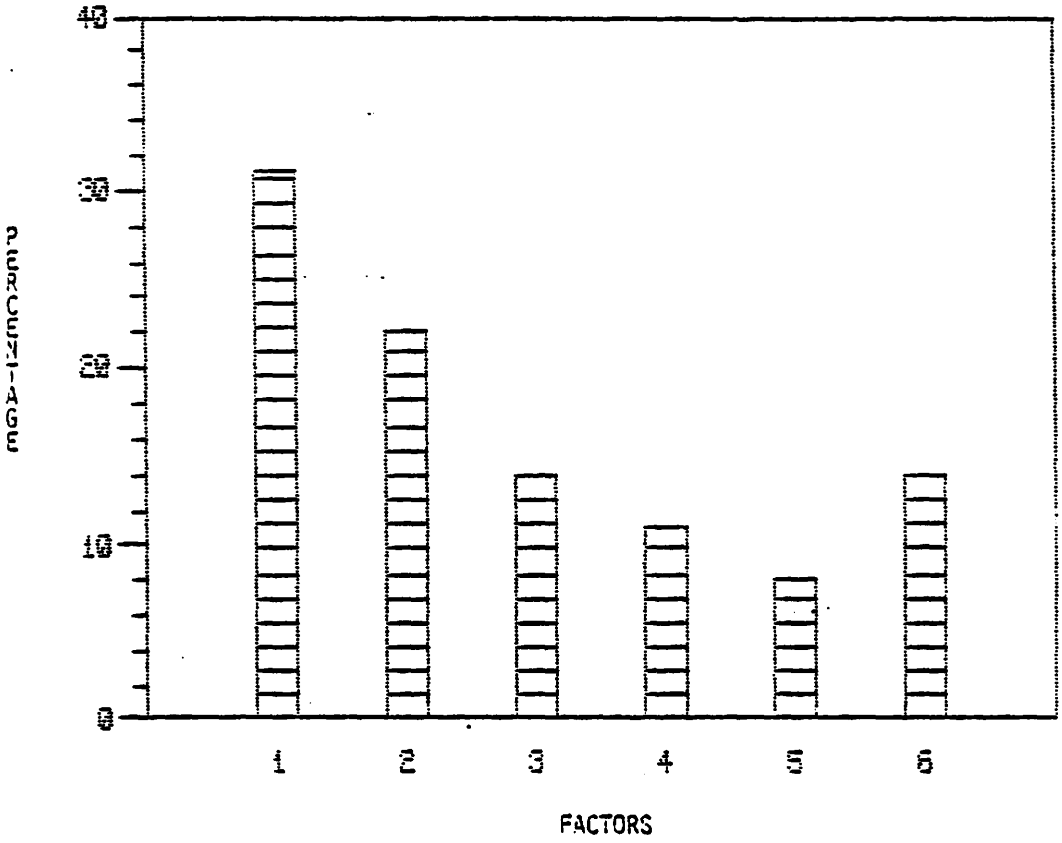
actors seem to hear raspiness, roughness, hoarseness, constriction, catches, or breaks. It is interesting to note, on the basis of a trend indicated by Bar Graph 2, that the professional actor seems to be more able to describe the negative aspects of the sound of his voice than the positive aspects.

2. In addition, as summarized in Bar Graph 3, some of the actors have identified certain working conditions or vocal demands which they feel have a direct effect on their vocal performance. A notable number of them indicated that sustained high-energy vocal production and/or having to work over a great deal of background noise are important factors. Others identified the period of dress rehearsals and preview performances as a time when the voice seems particularly vulnerable to problems. None of these actors identified opening night, however, as a time when the voice is particularly vulnerable to problems.

3. And finally, a number of the actors indicated that there are other factors which affect their vocal performance either positively or negatively. As illustrated by Bar Graph 4, these factors include amount of sleep, voice exercise, atmospheric environment and diet.

This study will be followed shortly by a paper reporting

BAR GRAPH 5:: VOCAL DEMANDS OR WORKING CONDITIONS WHICH
FATIGUE OR STRAIN THE VOICE
(MOST FREQUENTLY-OCCURRING RESPONSES IN SAMPLE OF 56 ACTORS)



FACTORS KEY: VOCAL DEMANDS

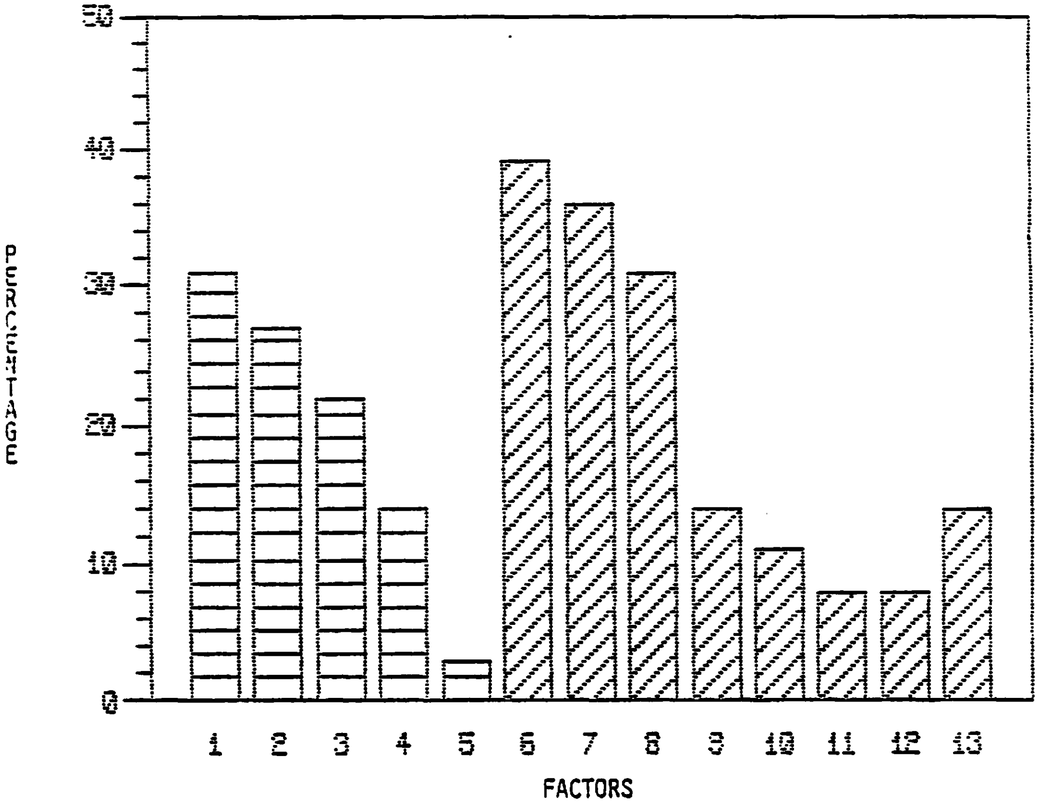
1. SUSTAINED HIGH-ENERGY VOCAL PRODUCTION, LARGE ROLE, WORKING OVER A GREAT DEAL OF BACKGROUND NOISE, USE OF DEMANDING "CHARACTER VOICE", SCREAMING, SHOUTING

WORKING CONDITIONS

2. DRESS REHEARSALS AND PREVIEW PERFORMANCES
3. LONG WORKING HOURS OF TALKING, TWO SHOWS A DAY
4. FIRST WEEK OF REHEARSAL
5. TECHNICAL REHEARSALS
6. OTHER

BAR GRAPH 4: OTHER FACTORS AFFECTING VOICE

(MOST FREQUENTLY-OCCURRING RESPONSES IN SAMPLE OF 36 ACTORS)



SHADING KEY: ▤ BENEFICIAL FACTORS

▨ DETRIMENTAL FACTORS

FACTORS KEY: BENEFICIAL FACTORS

1. A GREAT DEAL OF SLEEP
2. VOICE WORK, DAILY VOCAL WARM-UP, PHYSICAL EXERCISE
3. HEALTHY, LIGHT DIET; DRINKING A GREAT DEAL OF WATER
4. STEAM, HUMIDITY, WARM OR HOT WEATHER, FOG
5. OTHER

DETRIMENTAL FACTORS

6. INADEQUATE SLEEP
7. WINTER COLD AND DRYNESS; COLD, DAMP AIR; GOING FROM COLD TO HOT OR VICE VERSA
8. EXCESS OF DAIRY PRODUCTS, ALCOHOL, NUTS, CHOCOLATE
9. SMOKING, INHALING OTHERS' CIGARETTE SMOKE
10. SEASONAL ALLERGIES
11. MENSTRUAL CYCLE
12. NERVOUSNESS OR EMOTIONAL UPSET
13. OTHER

the results of the closed-ended questionnaire which was designed on the basis of the results reported here. These two studies, in combination, will have a very important twofold function:

1. They will serve to provide appropriate vocabulary for outside professionals--physicians, voice therapists, artistic directors, producers, and the like--working with actors on a regular basis, and
2. they will let professional actors know that problems do arise on a fairly regular basis and can be expected.

Later papers to be published from this same general study will consider relationships among factors in each actor's history; demands of casting, rehearsal, and performance through the 1984-85 season; and the occurrence and degree of severity of any vocal problems encountered by the actors.

The long-range study will provide us with a documented and relatively accurate picture of aspects of the professional actor's life and work which affect the voice. It will also prepare us to undertake a series of more specific studies which will provide much-needed documented and quantified information on the voice of the professional actor.

FOOTNOTES

1. Dr. Wilbur James Gould is the Director of the Recording and Research Center at the Denver Center for the Performing Arts. Dr. Raymond P. Wood, Jr. is a Research Associate at the Recording and Research Center.

APPENDICES

Appendix A: 1983 Open-Ended Questionnaire

Appendix B: 1984 Subject/Performer History Form

Appendix C: 1984 Subject Consent Form: Exploratory Study of Voice in Professional Actors

Appendix D: 1984 Open-Ended Questionnaire: Professional Actor Vocal Self-Assessment

DENVER
CENTER
THEATRE
COMPANY

November 10, 1983

Dear Acting Company Member,

The first major study undertaken by the Recording and Research Center of the DCPA will deal with vocal fatigue in actors. You can be of real assistance to us in our final planning stages for the methodology, if you would take just a few minutes to complete the following two descriptions:

1) Please describe the way you <u>feel</u> and <u>sound</u> when your voice is in its best shape. In other words, how do you know (by sensation, by sound) that your voice is as good as can be?

2) Please describe the way you <u>feel</u> and <u>sound</u> when your voice has been tired out by strenuous use. In other words, what symptoms or signals (aural, physical) are you aware of which let you know that something is not the way it should be?

Please write clearly and return to Bonnie Raphael as soon as possible. Thanks so much for your cooperation--more news of this study soon!

1050 13TH STREET DENVER COLORADO 80204 (303) 893-4200

Appendix B: 1984 Subject/Performer History Form

The Recording And Research Center

DENVER CENTER FOR THE PERFORMING ARTS
Voice Laboratory

PERFORMER/SUBJECT HISTORY FORM

This information sheet is designed to serve two purposes. First, during the season, if you wish to consult our company otolaryngologist, the information you provide will greatly facilitate the process. Second, if you choose to make yourself available for and are selected to participate in any of our research projects, this information can be most effectively incorporated. Please be assured that all this information will be strictly confidential and used only for clinical and research purposes. Thank you for filling this form out completely and promptly and returning it to Bonnie Raphael NO LATER THAN OCTOBER 10, 1984.

Today's Date ____________________

I. General Information

Name: ______________________________ Sex: __________ Age: ________

Local Address: ______________________________

Permanent Address: ______________________________

Local Telephone Number: (___) ______________ Answering Service: ______________

Social Security Number: ______________ Union Affiliation(s): ______________

Please list the major places in which you've lived since birth, in chronological order:

Location	Dates, or Length of Residence There

How long will you be in the Denver area? ______________________________

Please list shows and roles you've been cast in for this season: ______________

____

Which of these roles do you anticipate will be the most challenging for you?________

____

How are you usually cast? (Check all categories which apply.)

Juvenile_____ Ingenue_____ Character Roles_____ Musicals_____ Opera_____

Cabaret_____ Non-Speaking Roles_____ Television_____ Film_____ Outdoor Theatre_____

Other (please specify)____

Vocal Classification: Bass_____ Baritone_____ Tenor_____ Contralto_____ Soprano_____

Don't Know_____ Don't Sing_____

For how long have you been working as a performer?______________________________

Are you presently engaged in other means of employment as well? (please specify)

____

II. Family History

	Father	Mother	Sibling	Sibling	Current Spouse
Name					
Year of Birth or Approx. Age					
Nationality					
Education Completed					
Occupation					
State of Health					
Speech/Voice Problems					
Physical Handicaps, Respiratory problems					
Smoker?					
Marital Status					

III. Medical History

When were you born?____________ Where were you born?____________________________

Please check diseases, age, severity, and any associated complications:

(If you cannot remember, please indicate with a question mark.)

Childhood Diseases and Disorders	Age	Severity	Associated Complications
____ Measles	___	______	____________
____ Mumps	___	______	____________
____ Diphtheria	___	______	____________
____ Scarlet Fever	___	______	____________
____ Polio	___	______	____________
____ Stomach Trouble	___	______	____________
____ High Fevers	___	______	____________
____ Other (Congenital)	___	______	____________
____ Other (Hereditary)	___	______	____________
____ Other (specify)	___	______	____________

Respiratory Diseases	Age	Severity	Associated Complications
____ Tonsillitis	___	______	____________
____ Adenoiditis	___	______	____________
____ Sinus Trouble	___	______	____________
____ Allergies (specify)	___	______	____________
____ Ear Infections	___	______	____________
____ Strep Infections	___	______	____________
____ Post-Nasal Drip	___	______	____________
____ Bronchitis	___	______	____________
____ Pneumonia	___	______	____________
____ Asthma	___	______	____________
____ Others (specify)	___	______	____________

Any serious physical injuries (please indicate what and when)?____________

____

____

Any menstrual or pregnancy problems?______________________________

Any surgery? (Please indicate when, purpose, extent)____________________

____

Removal of tonsils and adenoids: Yes____ No____ When____________________

Complications?____________________________________

Medications taken regularly (name of drug, reason for taking, for how long, dosage)

____

Any history of dental problems, braces, prostheses, dentures? ____________________

____

Any problems yawning or opening the mouth wide (popping or locking jaw, etc.) or other jaw problems? ____________________

Any history of hearing or visual problems? (specify) ____________________

____

Any history of psychological or neurological problems? (specify) ____________________

____

Present Condition

Present state of health: Excellent____ Good____ Poor (specify) ____________________

Height__________ Weight______________

Do you or did you smoke? ______ What? ______________ How Much? ______________

For how long? ____________________

Do you or did you drink alcohol beverages? _______ What? ______________

How much? _________ For how long? ____________________

Are you presently using any recreational drugs? _____ What? ______________

Is use frequent, intermittent or occasional? ____________________

Do you work or live in an atmosphere of chemicals, dust, or other elements? ________

What? ____________________ How much? ____________________

Do you have any condition that may limit you in your career? ____________________

IV. Education, Training and Experience

	School	Location	Major Field	Dates	Degree Granted
High School					
College					
Graduate School					
Other					

Most Significant Acting Training:

Where? ____________________ With Whom? ____________

When? ____________________ For how long? ____________

Description ______________________________

Evaluation ______________________________

Any effects (positive or negative) on your present work? ____________

____

Most Significant Singing Training:

Where? ____________________ With Whom? ____________

When? ____________________ For how long? ____________

Description ______________________________

Evaluation ______________________________

Any effects (positive or negative) on your present work? ____________

____

Have you taken any courses in vocal pedagogy? (specify) ____________

Most Significant Voice and Speech Training:

Where? ____________________ With Whom? ____________

When? ____________________ For how long? ____________

Description ______________________________

Evaluation ______________________________

Any effects (positive or negative) on your present work? ____________

____

Have you taken any courses in speech science or speech pathology? (specify) ________

____

Most Significant Dance, Movement, and/or Physical Education Training:

Where? ____________________ With Whom? ____________

When? ____________________ For how long? ____________

DCPA/RFC
History—6

Description ____

Evaluation ____

Any effects (positive or negative) on your present work? ____________________

____

Does your present or past physical exercise regimen involve any lifting of heavy weights? (If yes, please specify)

Have you ever been taught, or are you now using any particular vocal, physical, relaxation, or mental exercises as a warm-up for rehearsals or performances? If so, please describe briefly the contents and approximate length of your regimen.

Performance Experience: (If you have a current resumé on file with the Denver Center for the Performing Arts, please indicate below, and we'll be happy to use that. If not, please summarize your performance experience below.)

Other vocational experience or training:

Hobbies and Interests:

V. Voice and Speech History

Breathing

Any problems breathing? ____

Any shortness of breath with physical activity? ____

Any problems breathing at night? ____

Can you breathe through your nose easily? ____

Any discomfort or dryness in the nose? ____ Any nosebleeds? ____

Does liquid ever go up your nose when you swallow or drink from a fountain? ____

Do you have trouble blowing up a balloon? ____ Any pains in chest? ____

Throat

Any excessive mucous in the throat? ____ Any post-nasal drip? ____

Bothered with a dry throat? ____ Coughing? ____ Throat Clearing? ____

Any trouble in swallowing? ____ Any pain, discomfort, or odd feelings in the throat? ____

Voice

How is your voice affected by:

time of day ____

certain days of the week ____

menstrual cycle ____

certain times of the year ____

general fatigue ____

nervousness ____

illness ____

performance ____

excessive amount of vocal use ____

other variables (please specify) ____

Do you notice any more problems after singing than after talking? (If yes, then please specify) ____

Do you notice any more problems after acting than after just talking? (If yes, then please specify.) ____

Do any particular performance locations or theatres provide vocal difficulties for you? ____

When you wake up, does it take a while for your throat to clear up? ____ If so, how long must you wait? ____ Do you do anything in particular to clear up your voice? ____

When is your voice usually the best? ____

Have you ever lost your voice completely? ____ What were the circumstances? ____

____

How long did the voice loss last?________What did you do to regain your voice?

____

Do you now, or did you in the past:

scream and/or shout____________how much?____________________
imitate sounds (cars, etc.)______________________________
talk more than average______________________________
cheerlead______________________________
participate in activities (e.g. sports) which involve loud or continuous talking______________________________
sing in a choir____________for how long____________________
sing solo__________type of music____________________

Have you ever had a breathing, throat, or laryngeal problem (such as a nodule, polyp, or contact ulcer)?________If yes, what was the problem?____________________

How old were you?______________________________

How long did you have the problem?______________________________

What was the cause?______________________________

Did you see a medical doctor?______, speech pathologist_____or other voice specialist?______________

What treatment (surgery, speech therapy, etc.) did you receive?______________

If you had speech therapy or voice retraining, for how long?______________

Was the problem alleviated?______________________________

Does the problem recur?__________If yes, under what circumstances?______________

____

Do you perform when you have a laryngeal problem?____________________

With what results?______________________________

Speech

Was English the only language spoken in your home when you were a child?____________

Do you presently speak any language besides English?______Which one(s)?__________

____

Are you aware of any regional characteristics in your own speech patterns at present? (If yes, please describe)______________________________

Is there any other voice and/or speech information you wish to provide which this questionaire has not considered?____

VI. Research Possibilities

Have you ever participated as a subject in a research project? (If yes, briefly describe what, where, and when.)____

Are you interested in serving as a subject for a research project on the performer's voice to be conducted here some time during the coming season? (You will be paid for your participation.)____

Would you like information on the kinds of research projects planned?____________

Thank you very much for your time and cooperation. Please be assured that all this information will be held in the strictest of confidence.

Revised (4)
BNR/RCS
Aug. 1984

Appendix C: 1984 Subject Consent Form: Exploratory study of the voice in professional actors

The Recording And Research Center

SUBJECT CONSENT FORM

EXPLORATORY STUDY OF VOICE IN PROFESSIONAL ACTORS

EXPERIMENTERS: Bonnie Raphael, Ph.D. and Ronald Scherer, Ph.D.

You are being asked to participate in an experiment that studies the voice of actors. The study will examine relationships between your responses to a vocal history questionaire, your own evaluations of your voice, and ongoing evaluations of your voice by the investigators during the theater season. You may be asked to fill out certain brief voice evaluation forms during the theater season, and your short interviews with the theater voice coach (Bonnie) during the theater season may be used to acquire further information for the study.

Your participation is voluntary, and there are no risks or direct benefits to you because of your participation in this study.

To insure confidentiality, information collected about you will be identified by a subject number and not by your name. Only the investigators will have access to your information.

You are free to ask questions about the study or to withdraw at any time.

I have read the above description of the study, and confirm that my participation is voluntary and that my cooperation is given freely and without coercion. I hereby consent to take part in this research project.

______________________________________ ______________________
SIGNATURE OF SUBJECT DATE

____
WITNESS

Please complete this form as soon as possible, and return to Bonnie Raphael, or drop it in her mailbox on the third floor of the Police Administration Building. If you have any questions, please feel free to ask! Thanks very much for your participation.

The Denver Center For The Performing Arts

The Recording and Research Center 1245 Champa Street Denver Colorado 80204 303/893-4000

Appendix D: 1984 Open-Ended Questionnaire: Professional Actor Vocal Self-Assessment

The Recording And Research Center

PROFESSIONAL ACTOR VOCAL SELF-ASSESSMENT

Dear Acting Company Member,

As part of our ongoing study of the voice of professional actors, please write clearly as you answer all of the following questions. Please return the completed form to Bonnie Raphael at your very earliest convenience, but certainly no later than Monday, November fifth. Thanks very much.

NAME: ______________________ DATE: ______________________

1) If you have just recently moved into the Denver area, please describe any vocal, physical, or respiratory changes of which you're aware which might be attributed to this relocation.

2) Please describe the way you feel and sound when your voice is in its best shape. In other words, how do you know (by sensation, by sound) that your voice is as good as can be?

3) Please describe the way you feel and sound when your voice has been tired out by strenuous use or is not working at its best. In other words, what symptoms or signals (heard, felt) are you aware of which let you know that the voice is not the way it should be?

(more room for answer on other side)

The Denver Center For The Performing Arts

4) Are there any particular times of the year, and/or times during a typical rehearsal period, and/or particular rehearsal or performance conditions which you feel make your voice especially vulnerable to fatigue or strain? (If yes, then please describe.)

5) Are there any other factors (e.g. amount of sleep, diet, weather, etc.) which you feel directly affect your voice?

6) What have other people said to you about your voice?

7) Is there any other information on these subjects which you would care to add to what you've said above?

Thanks very much for completing this form and promptly returning it to Bonnie Raphael. You can mail it to me c/o the Conservatory or can give it to me in person, but please do so by Monday, November 5th. Thanks!

"AT RISK" FOR HUNTINGTON'S DISEASE: ACOUSTIC ANALYSIS AND EARLY DIAGNOSIS

Lorraine A. Ramig, Ph.D.
University of Colorado, Boulder

ABSTRACT

Waveform and spectral characteristics were studied in the sustained vowel phonation of twenty individuals "at risk" for Huntington's disease. Results support subtle deviations in phonatory characteristics in one fourth of the "at risk" individuals; similar deviations were not observed in a control group. Implications for future research, including philosophical aspects of early diagnosis are discussed.

(Work supported by grants from the American Speech-Language-Hearing Foundation, Hereditary Disease Foundation and the Council on Research and Creative Work of the University of Colorado-Boulder).

INTRODUCTION

Huntington's disease (or HD) is a chronic, degenerative neurological disease characterized by progressively increasing choreiform movements and mental deterioration (Hayden, 1981). The spontaneous, inappropriate, purposeless movement in all parts of the body is the most striking feature of this disease and has been associated with extrapyramidal system dysfunction (Chusid, 1982). Darley et al., (1975) classified the characteristic speech of HD patients as quick hyperkinetic dysarthria of extrapyramidal system origin and speculated that the speech characteristics of these patients reflect "quick involuntary movements and variable tone" in musculature important for speech production.

Of particular significance to the study of HD is that 50% of the "at risk" offspring of these patients will inherit the disease. (Hayden, 1981). Furthermore, symptoms usually do not appear until after the childbearing years. Petajan et al., (1979) rationalized, "given the hereditary nature of the disease, the motor abnormalities must be present in nascent form long before the usual age of diagnosis in the twenties or thirties" (p. 163). These researchers measured abnormalities of single motor unit (SMU) control in the first dorsal interosseous (a finger muscle) of 88% of the HD patients they studied and 55% of the subjects "at risk". The HD patients could not achieve SMU control and bursts of chorea were frequent. In individuals "at risk", they observed small irregular, ballistic activation of motor units which they termed "microchorea" and excessive recruitment of SMUs associated with voluntary activity. They proposed that "microchorea" and the inability to sustain SMU firing may be physiological markers for HD in subjects "at risk" but without clinical signs.

It is probable that bursts of chorea observed in HD patient's first dorsal interosseous muscles also occur in their laryngeal musculature. Such activity is one reasonable physiologic correlate for the percepts of "irregular pitch fluctuations and voice arrests" Aronson (1980) observed in the phonation of HD patients. Similarly, it is possible that subtle motor abnormalities may occur in laryngeal musculature of individuals "at risk" and could be reflected in acoustic measures of voice.

It was based upon this rationale that Ramig (1984) began to study voices of HD patients and individuals "at risk" for HD. She observed the following disruptions in the sustained vowel phonations of HD patients: low frequency segments (abrupt drops in fundamental frequency of approximately one octave), vocal arrests and reduced maximum vowel duration. She suggested that these disruptions may be manifestations of a continuum of choreiform involvement in laryngeal musculature which may parallel progression of the disease (functional staging; Shoulson and Fahn, 1979). Low frequency segments and vocal arrests were observed in earlier stages of HD; significantly reduced maximum vowel duration was observed in the later stages of HD.

If this continuum concept is robust, it could be extended to individuals "at risk" for Huntington's disease. It could be predicted that certain of these individuals would demonstrate subclinical manifestations of phonatory instability as reflected in acoustic measures of phonation. Thus, this research was designed to study acoustic characteristics of sustained vowel phonation produced by individuals "at risk" for Huntington's disease.

METHOD

Twenty symptom-free individuals "at risk" for HD served as subjects for this study. Seven were male, thirteen were female; they ranged in age from 22 to 62 years. One hundred and fifteen individuals without neurological disease or family history of neurological disease made up a normative control group. Seventy-three were male, forty-two were female; they ranged in age from 20 to 80 years.

Subjects were seated in a sound treated booth (IAC-401-A) and fitted with a head band which suspended a condensor microphone (Sony ECM-21) 15 cm in front of the lips. Voice signals were recorded by a Sony (TC-277-4) tape recorder located outside the booth. Recording level settings on the tape recorder were constant throughout data collection. Among other tasks, subjects were asked to produce three sustained phonations of the vowel /a/ for maximum duration and three sustained phonations of the vowel /a/ for comfortable duration.

These recordings were analyzed in two ways. Initially, because the occurrence of phonatory disruptions (such as low frequency segments observed in HD patients) is unpredictable, a running analysis was made of oscillographic output of these voice signals (Seiman jet-spray oscillograph). All measures were made from hard copy waveform data (20 cm/second for male data; 40 cm/second for female data) to allow discrete visualization of waveforms. Examples of low-frequency segments in a Huntington's disease patient and an "at risk" individual are presented in Figure 1 (A and B). For purposes of this investigation and to avoid confounding with laryngeal onset and offset behaviors, low frequency segments were quantified further only if they occurred after at least 250 msec of modal phonation and 250 msec preceding

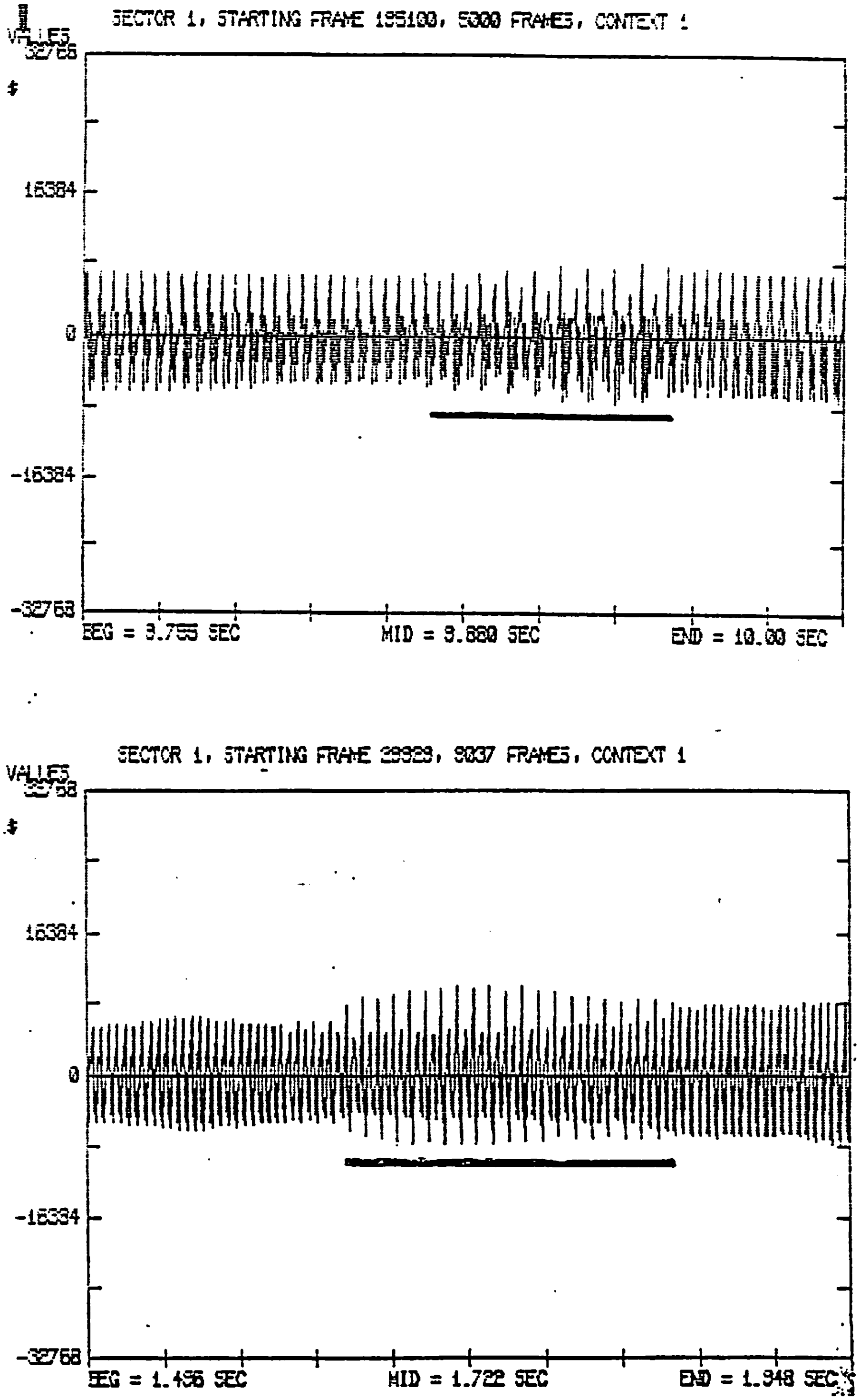


Figure 1-A

Figure 1-B

Figure 1. Waveform representation of 250 msec of sustained vowel phonation by a Huntington's disease patient (Figure 1-A) and 450 msec of sustained vowel phonation by an "at risk" individual (Figure 1-B).

termination of phonation.

A second form of analysis allowed measurement of more subtle, cycle-to-cycle waveform characteristics (jitter and shimmer). Because these acoustic measures have been reported to change with age (Ramig and Ringel, 1983), a subset of twenty-four speakers ranging in age from 25 to 75 years, was selected from the control group for this analysis. Each speaker's steadiest vowel was selected and digitized at 20,000 samples per second. From the mid-section of each vowel, two seconds of phonation were sampled and successive 20 cycle windows were analyzed for the two second duration by GLIMPES, a waveform analysis program by Titze (1984). This windowing is illustrated in Figure 2. For "at risk" individuals, measurements were made on three successive 20-cycle windows which did not include previously observed low frequency segments. Measures of the following acoustic variables were derived from this analysis: jitter, shimmer and harmonics-to-noise ratio (Figure 3). It should be pointed out that increases in jitter (Yumoto et al, 1984) and shimmer (Koike, 1968) and decreases in harmonics-to-noise ratio (Yumoto et al, 1984) have been related to increased phonatory instability.

RESULTS

Quantification of low frequency segments in individuals

FIGURE 2

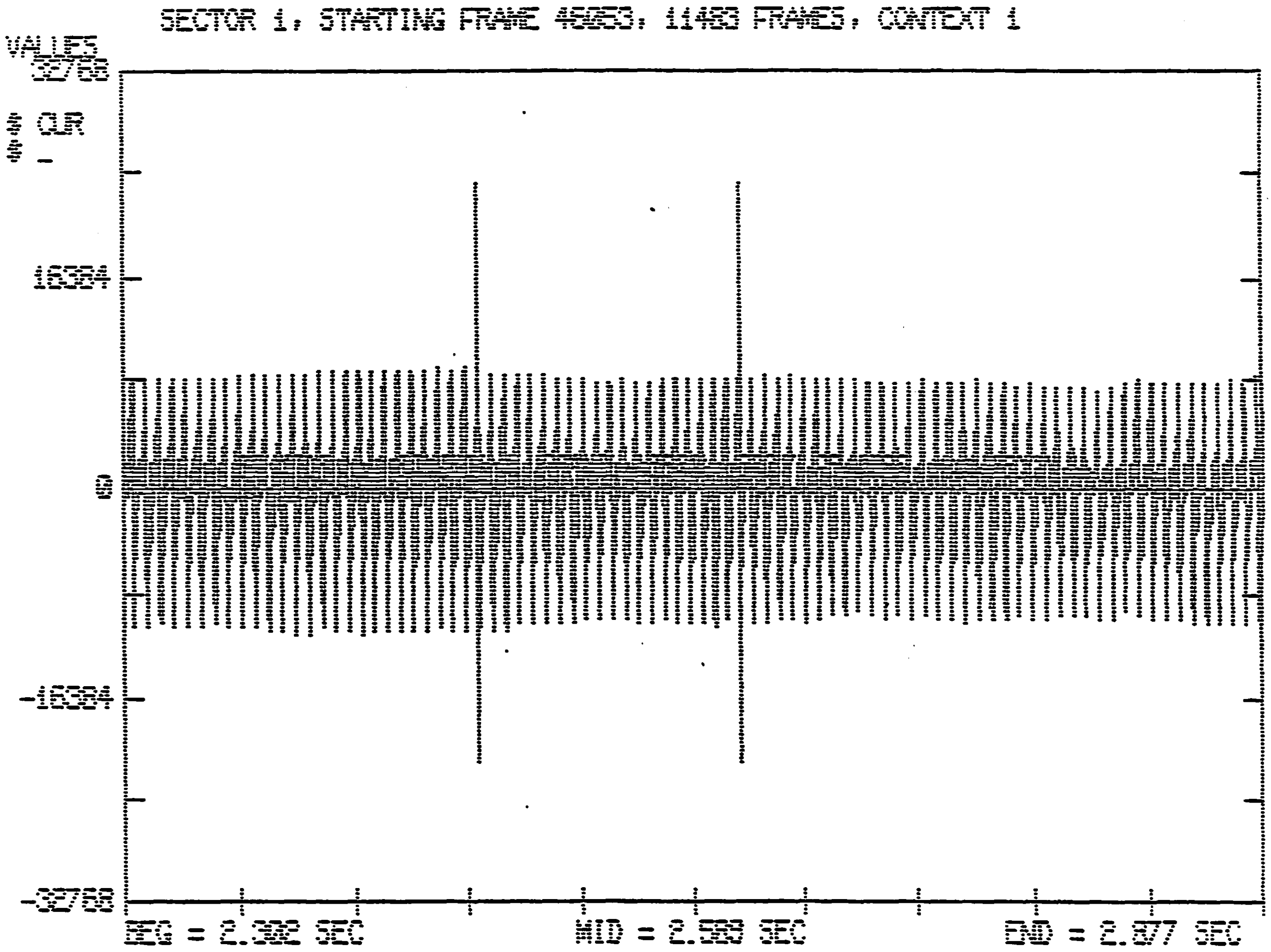

SECTOR 1, STARTING FRAME 46053, 11483 FRAMES, CONTEXT 1
VALUES
32768
CUR
-
16384
0
-16384
-32768
BEG = 2.302 SEC
MID = 2.589 SEC
END = 2.877 SEC

FIGURE 3

Perturbation Factors: (jitter and shimmer)

$$PF = \frac{100\sum_{i=2}^{N}\left|x_i - x_{i-1}\right|}{(N-1)\bar{x}}$$

Harmonics-to-Noise Ratio*:

$$H/N = N\int_0^T f_A^2(t)dt \bigg/ \sum_{i=1}^{N}\int_0^{T_i}\left[f_i(t) - f_A(t)\right]^2 dt$$

$$\text{where}\quad f_A(t) = \sum_{i=1}^{N}\frac{f_i(t)}{H}$$

*Yumoto, E., Gould, W.J., and Baer, T. (1982). "Harmonics-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. 71, 1544-1550.

"at risk" and control subjects is presented in Table 1. For presentation in this table, a subset of the one hundred and fifteen member control group was randomly selected to reflect the age and sex characteristics of the "at risk" group. Inspection of this table reveals that for the "at risk" subjects, the incidence of low frequency segments was approximately eight times as great as that observed in this normative control group. Incidence of low frequency segments in patients with Huntington's disease was over one hundred times as great as that observed in the "at risk" subjects.

The means and standard deviations of jitter, shimmer and harmonics-to-noise ratio for "at risk" individuals and the normative control group are presented in Table 2. For presentation in this table, "at risk" data is compared to the normative control group, classified by chronological age. Inspection of these data reveals that while means and standard deviations of shimmer and harmonics-to-noise ratio overlap between the "at risk" and control groups, the mean and standard deviation of jitter in the "at risk" group are almost double that observed in the control group. This is illustrated in Figure 4.

Inspection of the ranges of these measures (Table 3) reveals that for shimmer and harmonics-to-noise ratio there is

TABLE 1

INCIDENCE OF LOW FREQUENCY SEGMENTS EXPRESSED IN PERCENT (RATIO OF NUMBER OF LFS PER TOTAL NUMBER OF VOWELS; LFS/VOWELS)

CONTROL (N = 34)	"AT RISK" (N = 20)	HD PATIENTS (N = 8)
9/204 (4%)	18/54 (33%)	102/27 (378%)

TABLE 2

	JITTER		SHIMMER		H:N	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
HD "AT RISK"						
N = 20	0.81	0.52	2.02	0.95	24.53	2.76
CONTROL						
N = 24						
YOUNG (25-35)	0.45	0.26	1.71	0.91	25.97	2.53
MIDDLE (45-55)	0.57	0.19	2.03	0.76	24.72	1.80
OLD (65-75)	0.51	0.13	2.23	0.94	25.29	1.49

FIGURE 4

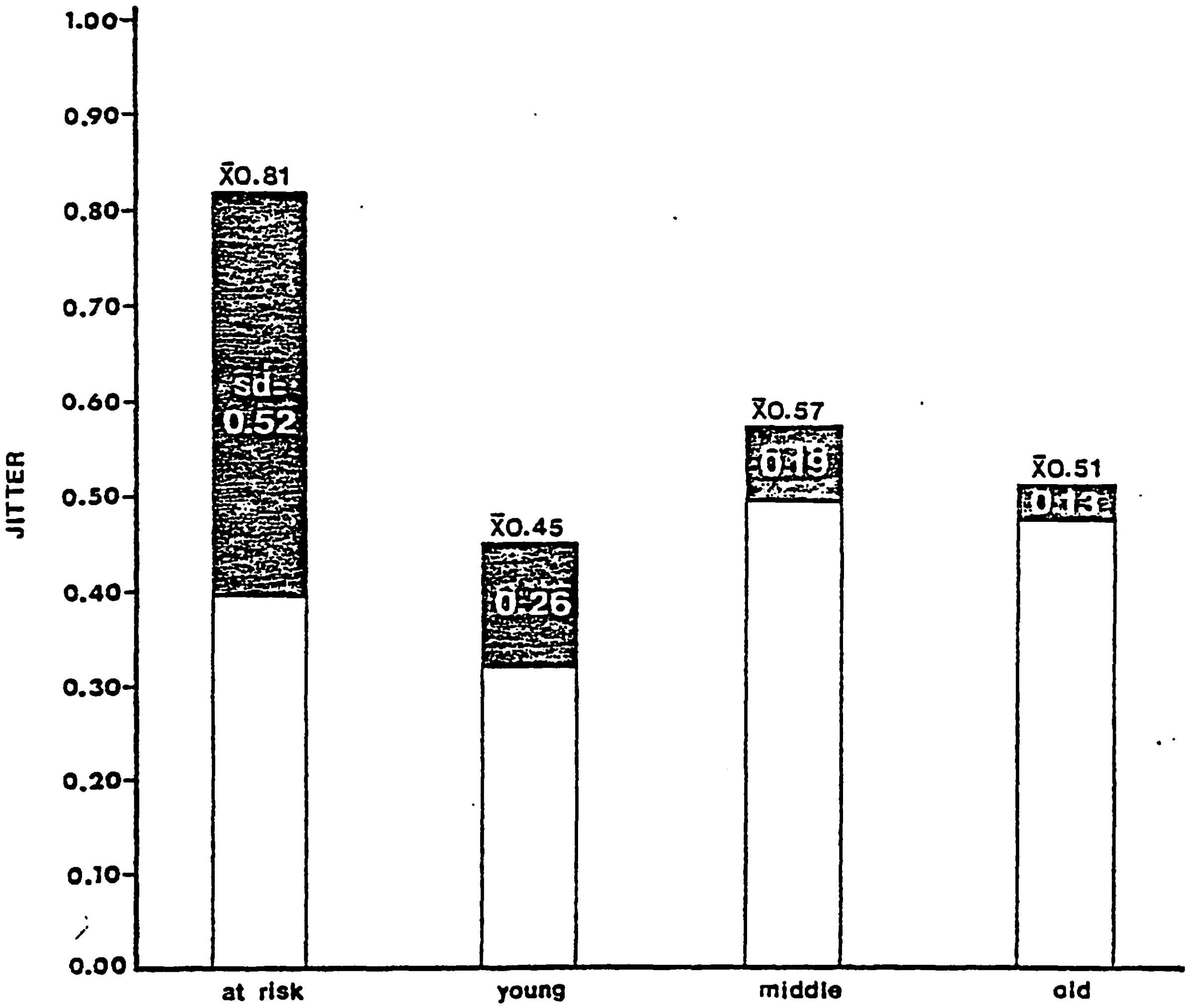

JITTER
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
X̄0.81
sd=
0.52
X̄0.45
0.26
X̄0.57
0.19
X̄0.51
0.13
at risk
young
middle
old

TABLE 3

RANGES

	JITTER	SHIMMER	H:N
HD "AR"	0.20 - 2.21	0.46 - 4.67	30.68 - 18.57
CONTROL	0.18 - 1.14	0.99 - 4.07	29.75 - 20.32

overlap between the "at risk" individuals and the control group. However, jitter values from five of the "at risk" individuals were beyond the jitter ranges of the control group. Two of these individuals had previously been identified as having low frequency segments (Ramig, 1984).

CONCLUSIONS

These preliminary findings support the hypothesis of increased phonatory instability in certain individuals "at risk" for Huntington's disease. The observation of low frequency segments in certain of these individuals extends the continuum of phonatory disruption observed in Huntington's disease (Ramig, 1984) to include certain "at risk" individuals. While the laryngeal pathophysiology responsible for these observations could be "microchorea", it is premature to speculate until more extensive analyses are carried out.

Furthermore, the significance of such vocal abnormalities in individuals "at risk" for Huntington's disease can only be determined relative to their carrier status. The recent genetic marker findings by Guscella et al., (1983) soon will allow identification of carrier status in these "at risk" individuals. Analysis of phonation in individuals with known carrier status will offer a unique potential to evaluate the significance of such vocal abnormalities.

In conclusion, we must keep in mind that the potential of early diagnosis of neurological disease is a tremendous responsibility. The recent localization of the HD marker (Guscella et al., 1983) is a prototype for similar approaches to other inherited neurological diseases and as such raises questions of widespread significance. Harper (1984) recently questioned, "what proportion of individuals will opt for a predictive test, how will they be told if they carry the marker, should unborn children be tested, how can confidentiality be maintained, when is it the family rather than the individual that is the unit to be tested?" If voice analysis has the potential to make a contribution to early diagnosis of neurological diseases, voice researchers must begin to think about the answers to these questions as well.

ACKNOWLEDGMENTS

I would like to express my appreciation to the individuals who participated as subjects in this research. Muriel "Chauncey" Rhyne of the Rocky Mountain Huntington's disease Foundation and Thomas Caldwell of the Chicago area Huntington's disease Foundation who provided important recruiting assistance. I wish to acknowledge the Department of

Communicative Disorders, Northwestern University, for allowing me to collect a portion of these data at their facility. The Recording and Research Center of the Denver Center for the Performing Arts, Ronald Scherer, Ph.D. and Ingo Titze, Ph.D. are gratefully acknowledged for facilitating analysis of these data.

REFERENCES

Aronson, A. Clinical Voice Disorder. New York: Thieme-Stratton, Inc. (1980).

Chusid, J.C. Correlative Neuroanatomy and Functional Neurology. Los Altos: Lange Medical Publications (1982).

Darley, F.L., Aronson, A.E. and Brown, J.R. Motor Speech Disorders. Philadelphia: W.B. Saunders (1975).

Guscella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E., Watkins, P.C., Ottina, K., Wallace, M.R., Sakagucki, A.Y., Young, A.B., Shoulson, I., Bonilla, E. and Martin, J.B. Nature, 306, 234-238 (1983).

Hayden, M.H. Huntington's chorea. New York: Springer-Verlag (1981).

Koike, Y. Vowel amplitude modulations in patients with laryngeal diseases. J. Acoust. Soc. Am. 45: 839-844, 1968.

Petajan, J.H., Jarcho, L.W. and Thurman, D.J. Motor unit control in Huntington's disease: a possible presymptomatic test, in Chase, T.N., Wexler, N.S. and Barbeau, T. (Eds.) Advances in Neurology, Vol. 23, New York: Raven Press, 1979.

Ramig, L.A. Acoustic analysis of voices of patients with Huntington's disease and individuals "at risk": implications for early diagnosis, Paper presented to the Clinical Dysarthria Conference, Tucson, 1984.

Ramig, L.A. and Ringel, R.L., Effect of physiological aging as selected acoustic characteristics of voice. Journal of Speech and Hearing Research. 26, 22-30 (1983).

Shoulson, I. and Fahn, S. Huntington's disease: clinical care and evaluation. Neurology 29, 1-3, 1979.

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

Yumoto, E., Sasaki, Y. and Okamura, H. Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. J. Speech Hear. Res. 27, 2-6. 1984.

ACOUSTIC ANALYSIS OF VOICES OF PATIENTS WITH NEUROLOGICAL DISEASE: A RATIONALE AND PRELIMINARY DATA

Lorraine A. Ramig, Ph.D.
University of Colorado - Boulder

Ronald C. Scherer, Ph.D.
The Denver Center for the Performing Arts

Ingo R. Titze, Ph.D.
University of Iowa
and
The Denver Center for the Performing Arts

ABSTRACT

Acoustic analysis of voices of patients with neurological disease may make significant contributions to early and differential diagnosis and measurement of disease progression. This paper presents a rationale for research in this area, as well as preliminary data from patients with myotonic muscular dystrophy, Huntington's disease, Parkinson's disease and amyotrophic lateral sclerosis. Results support the hypothesis that noninvasive acoustic analysis may offer information of clinical value to the otolaryngologist, neurologist and speech pathologist.

INTRODUCTION

Voice analysis of neurologically diseased patients may be an important tool in early diagnosis, differential diagnosis, and measurement of disease progression and treatment success. While the perceptual skills of the voice pathologist frequently contribute to these goals (1, 2, 3), recently developed acoustic and physiologic analyses may provide more sensitive and reliable data on phonatory functioning (4, 5, 6).

Only recently have researchers begun to apply acoustic and physiologic analysis to phonation of patients with neurological diseases (6, 7, 8, 9, 10) These few quantitative studies support the notion that acoustic and physiological analysis may exceed the perceptual level of inquiry and provide valuable additional information to quantify phonatory characteristics relative to site or progression of neurological disease, or effects of surgical or neuropharmacological treatment. Establishing the relationship between phonatory analysis and these variables would be of great clinical as well as academic value to the fields of voice pathology, otolaryngology and neurology.

The data presented here have been sampled from ongoing research projects designed to address the following question: can acoustic analysis of voice contribute to the diagnosis of 1) diseases of different neural subsystems? 2) systemic degeneration in neurological disease? and 3) subclinical manifestations of neurological disease?

METHODS

Subjects were seated in a sound-treated booth (IAC-401-A) and fitted with a head band which suspended a condensor microphone (Sony EMC-21) 15 cm in front of the lips. Voice signals were recorded by a tape recorder (Sony TC-277-4) located outside the booth. Recording level settings on the tape recorder were constant throughout data collection.

Among other tasks, each subject was asked to sustain the vowel /a/ for "as long as possible" at a comfortable frequency and intensity level. The same procedure was repeated using the vowels /i/ and /u/. All vowels were repeated three times. The entire task was repeated with subjects sustaining vowel phonation for a comfortable length of time.

These recordings were analyzed in two ways. Initially, a running analysis was made of oscillographic output of these voice signals (Seimans jet-spray oscillograph). To allow discrete visualization of waveforms, hard copy waveform data was run at 20 cm/second male data and 40 cm/second female data. In a second form of analysis, phonation was sampled from the mid-section of these vowels and digitized at 20,000 samples per second. Acoustic analyses (GLIMPES; 11) were carried out to derive the following variables, previously reported to be sensitive to phonatory instability: fundamental frequency, shimmer, jitter, and harmonics-to-noise ratio (13,14). Fundamental frequency is the number of vocal fold vibratory cycles per second. Jitter and shimmer are the cycle-to-cycle difference in time and amplitude respectively. Harmonics-to-noise ratio (12) reflects the relationship between the amount of periodic energy and noise per unit of time (Figure 1). It should be pointed out that phonatory instability has been associated with increases in jitter and shimmer (13) and decreases in harmonics-to-noise ratio (14).

Acoustic characteristics of phonation in patients with three distinct neural subsystem diseases.

Acoustic characteristics of phonation were studied in patients with three distinct neural subsystem diseases: Myotonic muscular dystrophy, a lower motor neuron disease characterized by flaccid dysarthria; Huntington's disease, an extrapyramidal system disease characterized by hyperkinetic dysarthria; and

FIGURE 1

Perturbation Factors:

$$PF = \frac{100\sum_{i=2}^{N}\left|x_i - x_{i-1}\right|}{(N-1)\bar{x}}$$

Harmonics-to-Noise Ratio*:

$$H/N = N\int_0^T f_A^2(t)\,dt \Big/ \sum_{i=1}^{N}\int_0^{T_i}\left[f_i(t) - f_A(t)\right]^2 dt$$

$$\text{where}\quad f_A(t) = \sum_{i=1}^{N}\frac{f_i(t)}{N}$$

*Yumoto, E., Gould, W.J., and Baer, T. (1982). "Harmonics-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. 71, 1544-1550.

Parkinson's disease, an extrapyramidal system disease characterized by hypokinetic dysarthria.

Myotonic muscular dystrophy patients present with myotonia, muscular weakness and wasting (15). Changes in muscles of these patients include hypertrophy and fragmentation of muscle fibers, followed by atrophy and replacement by fat and connective tissue (16). Electromyographic recordings from limb musculature reveal reduced size of muscle action potentials, indicating loss of fibers within a motor unit (16). While weakness in bulbar musculature is one of the earliest signs of this disease, to date there appears to be only one published paper focusing on speech characteristics of these patients (17). Clinical comments substantiate laryngeal involvement and typically describe these patients' voices as "weak and monotonous" (15), consistent with flaccid dysarthria.

Patients with Huntington's disease present with "chorea" - random, inappropriate bursts of motor unit activity that result in twitches or jerks and excessive recruitment of motor units, or overflow of motor unit activity during voluntary movements (18). Atrophy has been observed in caudate nucleus, putamen and striopallidal nerve fiber bundles, with subsequent release of pallidal and thalamic activity (19). While the etiology and pathogenesis of this disease remain unknown, recent findings reveal selective depletion of neurotransmitter agents, such as GABBA, in the striatum (19). Clinical perceptual observations report dysarthria as an early symptom which gradually increases. However, only within the past few years have researchers begun to study the speech of these patients (21, 22). Voices of individuals with Huntington's disease have been described as "intermittently harsh and strain-strangled" (2).

Parkinson's disease patients present with tremor, muscular rigidity and akinesia (19). These characteristics reportedly result from deficiency in dopamine due to disease, injury or dysfunction of the dopaminergic neuronal system (19). Studies of the extremities suggest that these patients have difficulty initiating, sustaining or shutting down individual motor units (18). It has been reported that units recruited with more than minimal muscle contraction often tend to fire in small groups at tremor frequency (18). Dysarthria, a common symptom of Parkinson's disease, is attributed in large part to marked limitation of range of movement in speech musculature (1). Voices of these patients have been described as "reduced in loudness, breathy and tremulous" (2).

There is little data on the laryngeal pathophysiology responsible for the phonatory disruptions observed in these diseases. However, based on the traditional approach (1, 2) which generalizes characteristics of limb pathophysiology to laryngeal function, one could predict distinct vocal characteristics for these diseases. For example, the observed limb pathophysiology in myotonic muscular dystrophy is weakness and wasting (15). These characteristics could be generalized to laryngeal muscles to explain the "weak, monotonous" (15) voices observed in these patients. The observed pathophysiology in limbs of patients with Huntington's disease includes random twitches and jerks (18). These characteristics could be generalized to laryngeal muscles to explain the "intermittently harsh, strain-strangled" (2) voices observed in these patients. Tremor, rigidity and akinesia characterize the limb pathophysiology in Parkinson's disease patients (19). If generalized to laryngeal musculature these characteristics could explain the "tremulous, reduced in loudness and breathy" (2) voices observed in these patients. This approach is summarized in Table 1.

Sustained vowel phonation data from one patient representing each disease group are presented in Figure 2 (B, C, D). Visual inspection of these data, in comparison with the data from a control subject (Figure 2-A), reveal distinct vocal patterns for each patient. Relatively large cycle-to-cycle disruptions in amplitude (shimmer) are apparent in the myotonic

TABLE 1

NEUROLOGICAL DISEASE	OBSERVED LIMB PATHOPHYSIOLOGY AND PREDICTED LARYNGEAL PATHOPHYSIOLOGY	PERCEPTUAL CHARACTERISTICS OF VOICE
MYOTONIC MUSCUALR DYSTROPHY -FLACCID	MUSCULAR WEAKNESS, WASTING	"WEAK, MONOTONOUS" (15)
HUNTINGTON'S DISEASE -HYPERKINETIC	RANDOM TWITCHES, JERKS	"INTERMITTENLY HARSH, STRAIN-STRANGLED" (2)
PARKINSON'S DISEASE -HYPOKINETIC	TREMOR, RIGIDITY, AKINESIA	"TREMULOUS, REDUCED IN LOUDNESS, BREATHY" (2)

FIGURE 2

Fig. 2-A

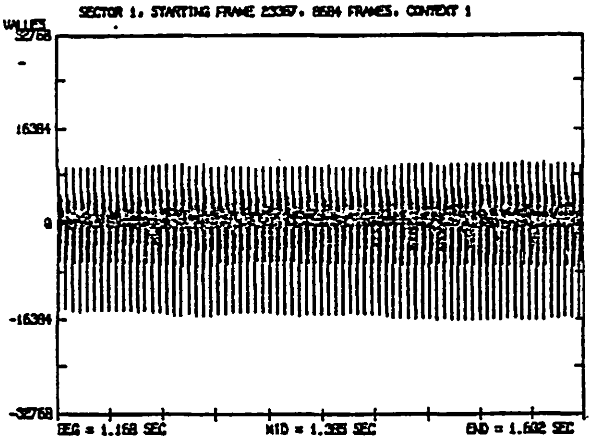


Fig. 2-B

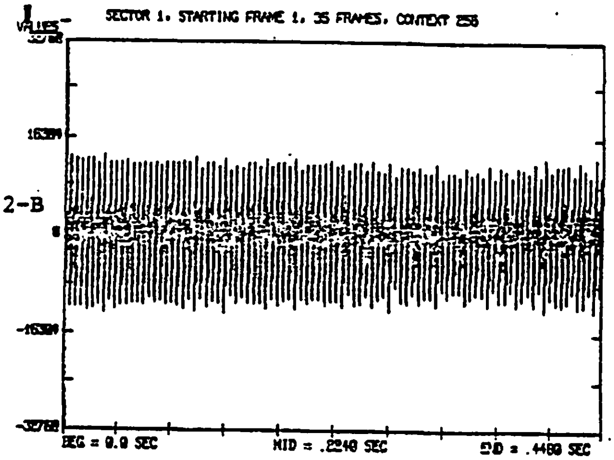


Fig. 2-C

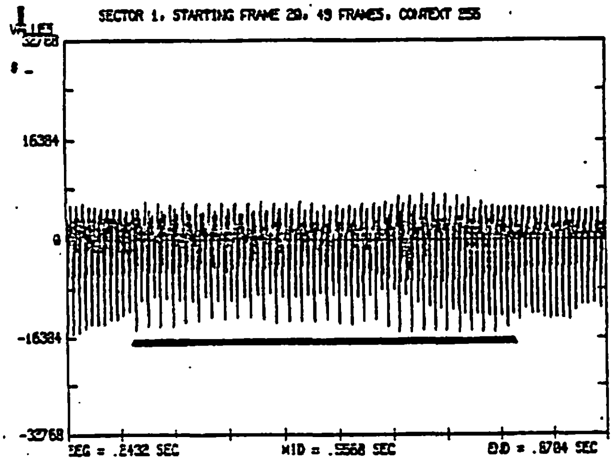
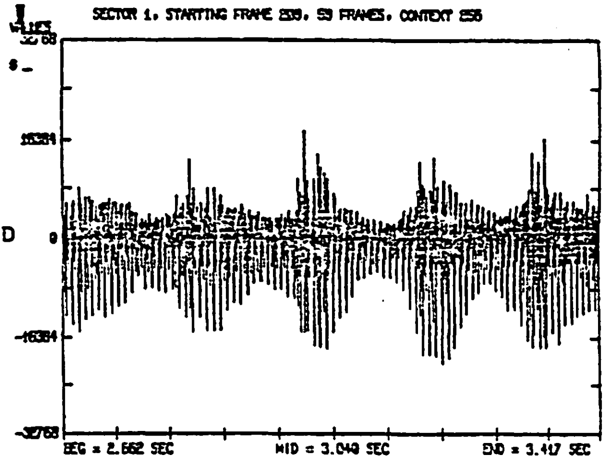


Fig. 2-D



voice (Fig. 2-B). In contrast, the data from the HD patient (Figure 2-C) reveal an abrupt, irregular phonatory disruption rather than cycle-to-cycle deviancies. This is shown as an abrupt decrease in fundamental frequency over 30 cycles in Figure 2-C. The voice of the Parkinson's disease patient (Figure 2-D) reveals rhythmic amplitude tremor, imposed upon cycle-to-cycle deviancies.

Quantification of these observed differences is presented in Table 2. The means represent data from four myotonic patiens, seven Huntington's disease patients and one Parkinson's disease patient. These data reflect a minimum of three fifty cycle windows sampled from each patients' steadiest comfortable /a/ phonation. Inspection of the quantified data supports the previous visual observations. For example, the patients with myotonic muscular dystrophy had shimmer values two times greater and jitter values five times greater than were observed in the control subjects. Relatively large shimmer and jitter values reflect large cycle-to-cycle variations in amplitude and period,respectively. Amplitude variations may depend largely on the adductory/abductory (A-AB) muscular system integrity because of the corresponding control over glottal airflow waveform characteristics (and therefore the associated intensity generation at the laryngeal level). Period variations would

TABLE 2

DATA IN PARENTHESES HAVE LINEAR TREND REMOVED

	SHIMMER		JITTER		HARMONICS-TO-NOISE	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
NORMAL CONTROL	1.79 (1.38)	0.73 (0.54)	0.39 (0.32)	0.16 (0.15)	25.97	1.69
MYOTONIC DYSTROPHY	3.08 (2.47)	1.65 (1.49)	2.02 (1.76)	1.89 (1.66)	17.48	2.48
HUNTINGTON'S DISEASE	2.27 (1.58)	0.87 (0.78)	0.68 (0.34)	0.52 (0.22)	16.35	1.88
PARKINSON'S DISEASE	8.76 (5.57)	3.07 (1.07)	1.15 (0.85)	0.14 (0.15)	13.65	3.65

depend largely upon the integrity of the muscles governing vocal fold elasticity (VFE). We conjecture that the relatively large values of jitter and shimmer for the myotonic dystrophy patients result from a large cycle-to-cycle instability of both the A-AB and VFE systems as a consequence of general laryngeal muscular weakness. Polyphasic motor unit action potentials of reduced duration (1 to 3 msec) have been reported in dystrophic limb muscles (20). These observations were attributed to loss of fibers which resulted in a decrease in both voltage amplitude and the contribution of distant fibers, as well as slight asynchrony. We hypothesize that such activity in laryngeal A-AB and VFE systems may be the pathophysiologic correlate for our acoustic observations. These acoustic observations support earlier perceptual work (2) which classified the voice disorder of myotonic patients as a 'relatively constant' disorder (p. 89).

In contrast, the phonatory disruption observed in the HD patients did not reflect such a constant laryngeal involvement. While the shimmer and jitter values of the HD patients were within normal limits during undisrupted phonation, voices of these patients were characterized by abrupt drops in frequency of approximately one octave (low frequency segments). Drops in frequency such as these would depend upon muscles governing vocal fold elasticity (VFE). Random, inappropriate bursts of motor unit activity one-tenth to one second in duration have been observed in choreic limb muscles. We conjecture that such choreiform involvement in laryngeal muscles is sufficient to effect the vibratory characteristics of the vocal folds, causing a fundamental frequency drop of one octave. These acoustic observations support earlier perceptual work (2) which classified the voice disorder of patients with HD as an 'arythmically fluctuating' disorder (p. 107).

Compared to the control group, the representative patient with Parkinson's disease had greater shimmer values (eight times) and jitter values (three times) and a smaller harmonics-to-noise ratio (almost half). The decreased cycle-to-cycle phonatory instability reflected in these measures appears to be related in part to the apparent phonatory tremor (Figure 1-D). In fact, when the data was analyzed with linear trend removed, shimmer decreased from 8.70 to 5.57 and jitter decreased from 1.15 to 0.85. The lower frequency phonatory tremor suggests a breakdown in the internal oscillation of motor control. The frequency of tremor in limbs of Parkinsonian patients has been reported to be 4 to 7 per second (19). The vocal tremor observed here is within that range. These findings are consistent with the earlier perceptual work (2) which classified Parkinsonian patients as having a 'rhythmically flucuating neurologic voice disorder' (p.

76) (2).

It should be pointed out that the frequency of vocal tremor in the Parkinson's patient was approximately 5-1/2 times per second, which is near the frequency of vibrato observed in the singing voice as well as in patients with spastic dysphonia (24). This is consistent with the hypothesis that the mechanism used to drive the system during vocal tremor is also responsible for production of normal vibrato (24). It also has been suggested that singers are able to inhibit periodic neural signals coming through the recurrent laryngeal nerve to both the adductors and abductors of the larynx, while at the same time, allowing pulses that are transmitted to the superior laryngeal nerve to the cricothyroid to manifest themselves as frequency modulation (24).

These preliminary data from patients with myotonic dystrophy, Huntington's disease and Parkinson's disease support the hypothesis that patients with distinct neural subsystem diseases present distinct acoustic characteristics of voice. Future research must evaluate the representativeness of these preliminary data in order to determine if distinct acoustic characteristics of phonation can be established in general for specific neural pathologies.

Acoustic characteristics of phonation in a patient with amyotrophic lateral sclerosis: a longitudinal study

There are at least two ways to address the question of whether a continuum of vocal disruption parallels systemic degeneration in neurological disease. Voices of a group of patients at different stages of a disease could be compared, or voices of individual patients could be studied longitudinally. The data presented here was collected over a 6 month period from a patient diagnosed with amyotrophic lateral sclerosis (ALS), a disease of upper and lower motor neurons, which is progressive without remission (19).

ALS is characterized by muscular atrophy and weakness, as well as fibrillations and fasciculations (19). The dysarthria accompanying this disease is classed as mixed (spastic and flaccid) (1), and voices of these patients have been described as "wet hoarse" (2). The observed limb pathophysiology in amyotrophic lateral sclerosis is atrophy, weakness, fibrillations fasiculations (19). These characteristics could be generalized to laryngeal muscles to explain the "wet hoarse" (2) voices observed in the patients (Table 3). Figure 3A is a sample of sustained vowel phonation from an ALS patient at the beginning of the 6 month period and Figure 3B is sustained vowel phonation at the

FIGURE 3

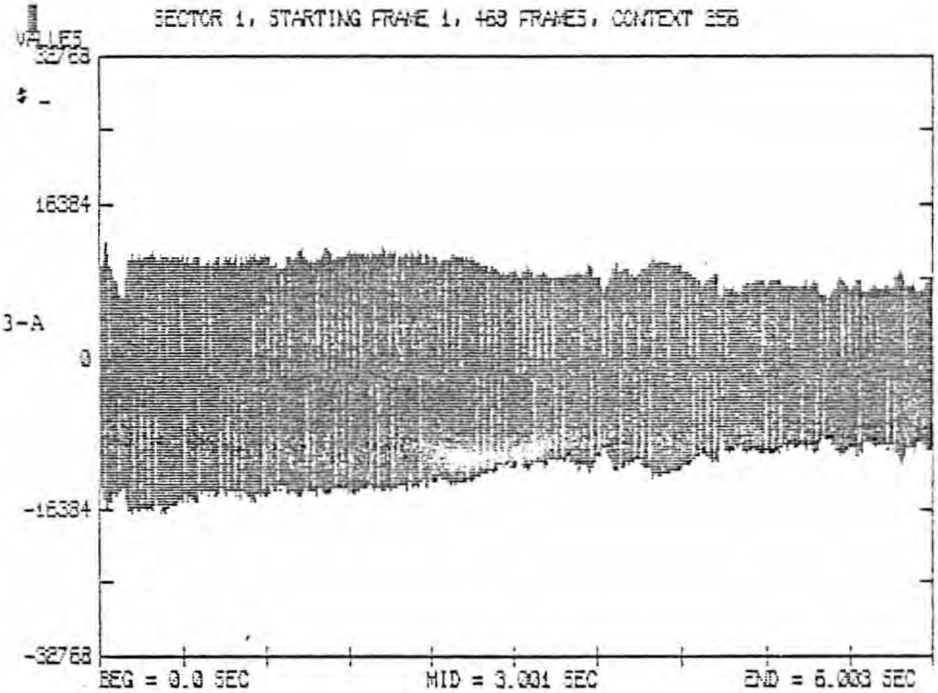


Fig. 3-A

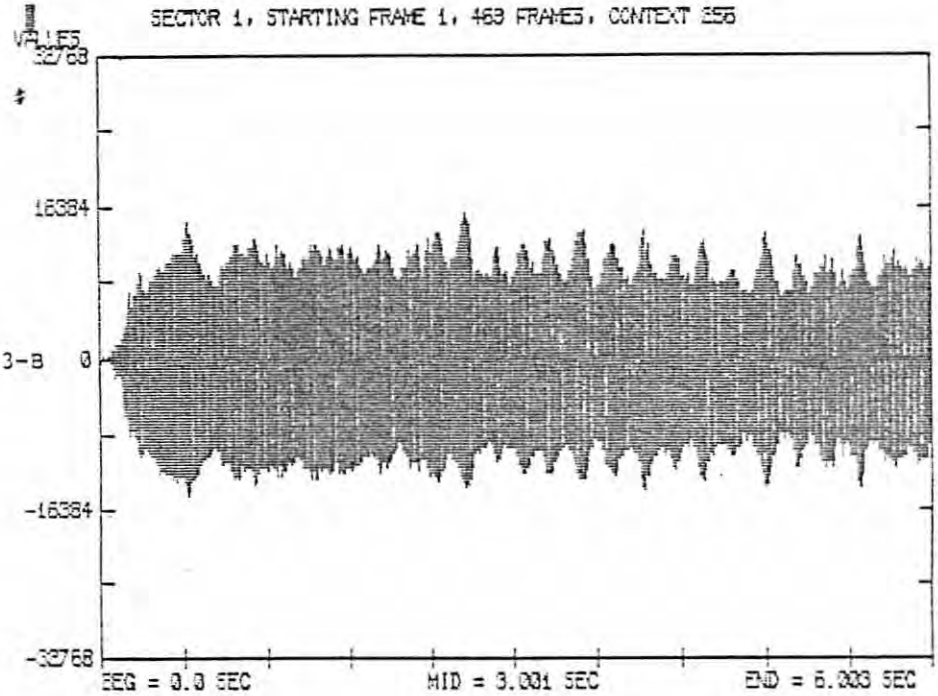


Fig. 3-B

TABLE 3

NEUROLOGICAL DISEASE	OBSERVED LIMB PATHOPHYSIOLOGY AND PREDICTED LARYNGEAL PATHOPHYSIOLOGY	PERCEPTUAL CHARACTERISTICS OF VOICE
AMYOTROPHIC LATERAL SCHLEROSIS -MIXED	ATROPHY, WEAKNESS, FIBRILLATION FASCIULATION	"WET HOARSE" (2)

end of this period. An increase in phonatory instability is apparent at the end of the 6 month period.

Quantification of these differences is presented in Table 4 and is illustrated graphically in Figure 4. Shimmer and harmonics-to-noise ratio were most sensitive to deterioration in phonatory function in this patient. Shimmer with linear trend removed increased from 1.67 (sd = 0.42) in trial 1 to 2.13 (sd = 1.23) in trial 4. Harmonics-to-noise ratio decreased from 21.74 (sd = 1.74) in trial 1 to 16.59 (sd = 3.53) in trial 4. One could speculate that this deterioration in phonatory stability was a reflection of degenerative changes in motor nuclei of the nucleus ambiguus.

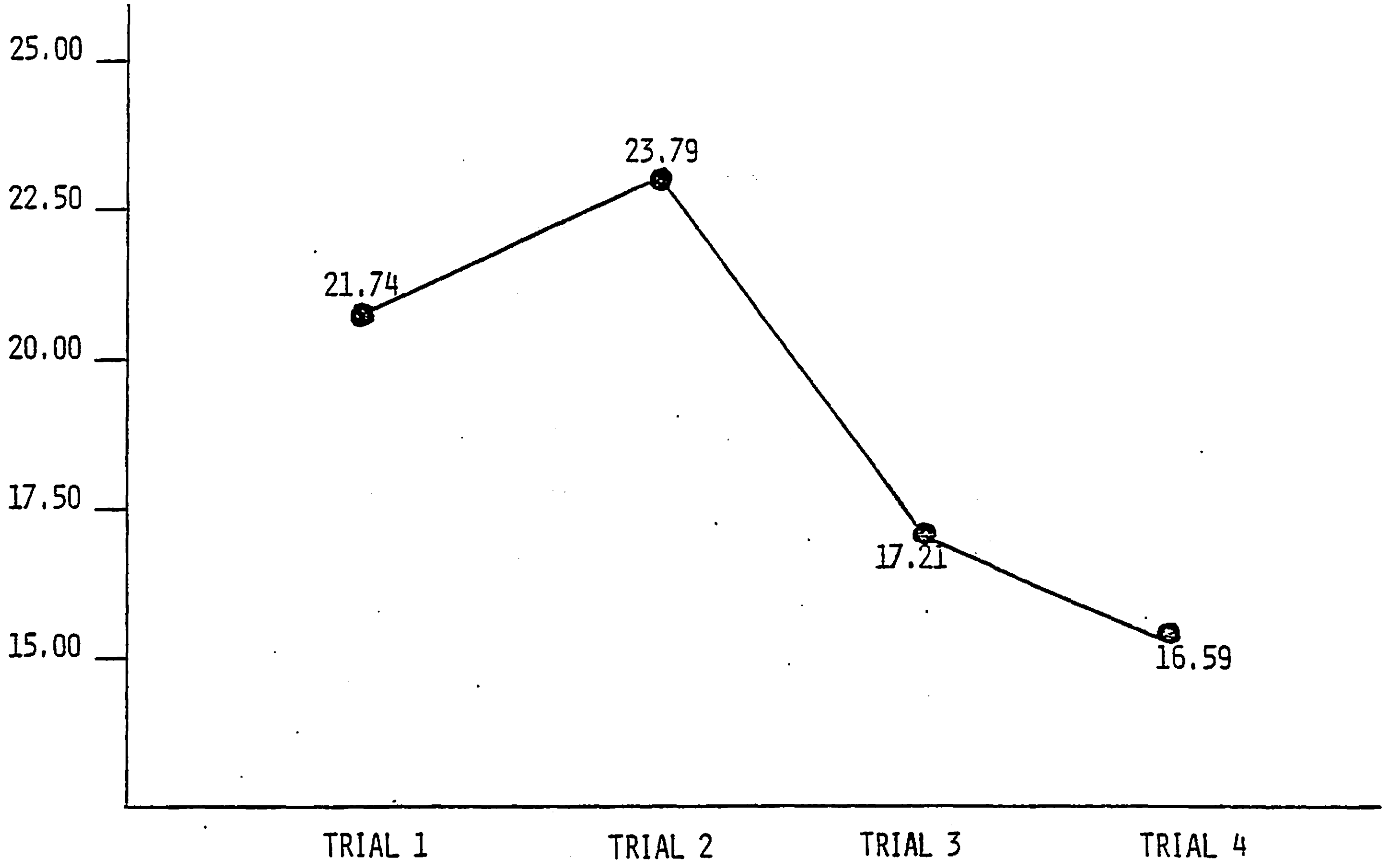
While this individual had not shown a marked decline in limb motor function during this six month interval, as measured by neurological examination, a decline in phonatory stability was apparent perceptually and is reflected in these acoustic measures. One could speculate that during this time period, degenerative changes in the nucleus ambiguus exceeded those in other motor nuclei. It also is possible that degenerative changes had a greater impact on the more finely tuned, complicated motor task of phonation. It is likely that even the slightest change in laryngeal muscle tone or control would be apparent immediately in phonatory measures, thus making them

TABLE 4

DATA IN PARENTHESES HAVE LINEAR TREND REMOVED

	TRIAL 1		TRIAL 2		TRIAL 3		TRIAL 4	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
FUNDAMENTAL FREQUENCY	145	3.68	153	3.47	174	2.48	162	2.83
SHIMMER	2.18 (1.67)	0.48 (0.42)	1.34 (1.03)	0.11 (0.07)	3.33 (1.39)	1.43 (0.48)	3.33 (2.13)	1.77 (1.23)
JITTER	0.44 (0.34)	0.04 (0.04)	0.32 (0.27)	0.04 (0.05)	0.27 (0.23)	0.06 (0.05)	0.49 (0.36)	0.22 (0.18)
H:N	21.74	1.74	23.29	1.32	17.21	2.52	16.59	3.53

FIGURE 4


25.00
22.50
20.00
17.50
15.00
21.74
23.79
17.21
16.59
TRIAL 1
TRIAL 2
TRIAL 3
TRIAL 4

extraordinarily sensitive to degenerative changes.

Acoustic characteristics of phonation in patients with Huntington's disease and individuals "at risk": implications for early diagnosis.

If degeneration in phonatory characteristics parallels disease progression, one could speculate that subclinical manifestations of disease may be reflected in phonation. This is of particular importance to the study of Huntington's disease (HD) because 50% of the "at risk" offspring of these patients will inherit the disease (25). Furthermore, symptoms usually do not appear until after the child bearing years and to date there is no early diagnostic indicator. (25). It has been rationalized "given the heredity nature of the disease, the motor abnormalities must be present in nascent form long before the usual age of diagnosis in the twenties or thirties" (26, p. 163). Abnormalities of single motor unit (SMU) control have been measured in the first dorsal interosseous (a finger muscle) of 88% of HD patients and 55% of subjects "at risk". HD patients could not achieve SMU control and bursts of chorea were frequent. In individuals "at risk", small irregular, ballistic activation of motor units, termed "microchorea", were observed, as well as excessive recruitment of SMUs associated with voluntary activity. It was proposed that "microchorea" and the inability to sustain

SMU firing may be physiological markers for HD in subjects "at risk" but without clinical signs (26).

It is probable that bursts of chorea observed in HD patients' first dorsal interosseous muscles also occur in their laryngeal musculature. Such activity is one reasonable physiologic correlate for the percepts of "irregular pitch fluctuations and voice arrests" observed in the phonation of HD patients. Similarly, it is possible that subtle motor abnormalities may occur in laryngeal musculature of individuals "at risk" and could be reflected in acoustic measures of voice.

To address this issue, sustained vowel phonation was studied in three groups of speakers: eight patients with Huntington's disease (HD), twenty symptom-free individuals "at risk" for Huntington's disease, and thirty-four control subjects without family history of neurological disease. Figure 5 displays low frequency segments observed in patients at the earliest stages of HD (27) (Fig. 5-A) and in one-fourth of the individuals "at risk" for HD (Fig. 5-B). Table 5 presents the incidence of these low frequency segments in HD patients, "at risk" individuals and a normative control group.

Inspection of the low frequency segment data reveals that for the "at risk" subjects, the incidence of low frequency

FIGURE 5

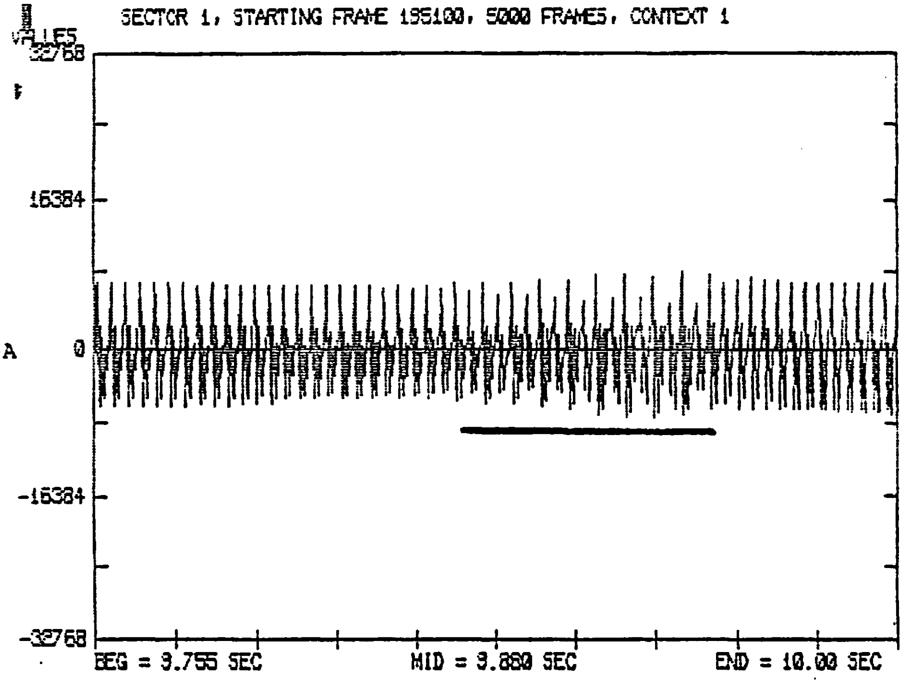


FIG. 5-A

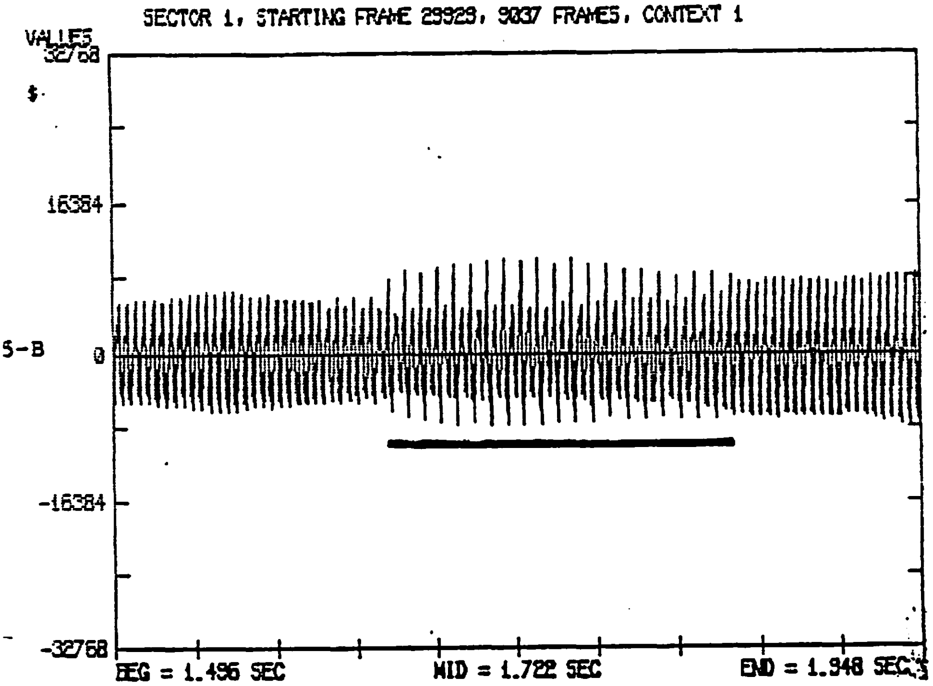


FIG. 5-B

TABLE 5

INCIDENCE OF LOW FREQUENCY SEGMENTS EXPRESSED IN PERCENT (RATIO OF NUMBER OF LFS PER TOTAL NUMBER OF VOWELS; LFS/VOWELS)

CONTROL (N = 34)	"AT RISK" (N = 20)	HD PATIENTS (N = 8)
9/204 (4%)	18/54 (33%)	102/27 (378%)

segments was approximately eight times as great as that observed in a normative control group.

These preliminary findings support the hypothesis of increased phonatory instability in certain individuals "at risk" for Huntington's disease. The observation of low frequency segments in certain of these individuals extends the continuum of phonatory disruption in Huntington's disease (21) to include certain "at risk" individuals. While the laryngeal pathophysiology responsible for these observations could be "microchorea", it is premature to speculate until more extensive analyses are carried out.

Furthermore, the significance of such vocal abnormalities in individuals "at risk" for Huntington's disease only can be determined relative to their carrier status. The recent genetic marker findings (28) soon will allow identification of carrier status in these "at risk" individuals. Analysis of phonation in individuals with known carrier status offers a unique potential to evaluate the significance of such vocal abnormalities.

CONCLUSION

These preliminary data support the ongoing research on phonation in patients with neurological disease. These data are consistent with the hypothesis that voice analysis of neurologically diseased patients may make important contributions to early and differential diagnosis, as well as in the measurement of disease progression.

Future research on the neurologically diseased voice must take into consideration a number of theoretical as well as practical issues. The establishment of relationships among phonatory characteristics and site or progression of neurological disease is dependent upon precise definition of neuropathology. Complete medical and pharmacological information is a necessity for valid data interpretation. Futhermore, categorization within disease group by stage of disease, etiology, etc. may reduce within group heterogeneity and facilitate more meaningful observations. For the same reason, the control group of individuals, having no personal or family history of neurological disease, should be categorized by chronological age and vocal use.

Tasks for study must also be carefully selected. Generalizing from the phonatory disorder observed in non-meaningful sustained vowel phonation to phonatory characteristics in meaningful speech must be questioned. Recent observations suggest that laryngeal behavior may differ significantly between such tasks (29; 30). Future research must incorporate tasks which assess a wide range of laryngeal behavior. Certain tasks, such as sustained vowel phonation, may be useful in revealing early breakdowns or differential symptoms. Other more dynamic tasks may provide useful information about the impact of certain neurological diseases on laryngeal pathophysiology.

Ongoing research should establish protocols and expand analysis techniques to maximally evaluate the contribution of acoustic information to early and differential diagnosis, as well as measurement of disease progression. Because acoustic data can be obtained from patients with movement disorders simply, noninvasively and relatively inexpensively, the potential contribution to these important clinical and academic areas must be evaluated. A goal for the future will be to understand the relationship between acoustic measures and laryngeal pathophysiology. To investigate that relationship, additional levels of measurement such as EGG, aerodynamic and EMG may be useful. We must keep in mind, however, that while these measurement techniques have been used to study the normal voice, they may not offer the most appropriate or practical way to quantify the neurologically diseased voice. In fact, we have been cautioned, "We should not allow our study of neurologically diseased voices to be directed by existing measures and technology" (31). Future research must be interactive and expansive both in terms of methods of study, as well as specific measurements, to allow the most valid observations of phonation in these patients.

Our interpretation of data should be expansive as well. Traditionally, it has been suggested (1, 2) that the laryngeal pathophysiology responsible for phonatory involvement in patients with neurological disease was consistent with limb pathophysiology. However, recent work (32) suggests that "assessments of motor speech disorders that assume a common pathophysiology for the limbs and the speech motor subsystems are likely to be in error" (p. 48). Such findings must be considered when developing hypotheses and interpreting phonatory data.

Future research will allow consideration of the preceeding theoretical and practical issues in the study of phonation in patients with neurological disease.

ACKNOWLEDGEMENTS

Deep appreciation is extended to the individuals who participated as subjects in this research.

BIBLIOGRAPHY

1) Darley, F.L., Aronson, A.E. and Brown, J.R.: Motor Speech Disorders. W.B. Saunders, Philadelphia, PA, 1975.

2) Aronson, A.E.: Clinical Voice Disorders. Thieme-Stratton, New York, NY 1980.

3) Hartman, D.E.: Neurogenic dysphonia. Ann. of Otol. Rhinol. and Laryngol. 93: 57-64, 1984.

4) Weismer, G.: Acoustic descriptions of dysarthric speech: perceptual correlates and physiological inferences. Seminars in Speech and Language, Thieme-Stratton, New York, NY, 1984.

5) Titze, I., Baer, T., Cooper, D. and Scherer, R.: Automated extraction of glottographic waveform parameters and regression in acoustic and physiologic variables. In Bless, D.M. and Abbs, J.H. (Eds.) Vocal Fold Physiology - Contemporary Research and Clinical Issues. College-Hill Press, San Diego (1983).

6) Hansen, D.G., Gerratt, B.R. and Ward, P.H.: Glottographic measurement of vocal dysfunction: a preliminary report. Annals of Otol. Rhininol. and Laryngol. 92: 413-420, 1983.

7) Hansen, D.G., Gerratt, B.R., and Ward, P.H.: Cinegraphic observations on laryngeal dysfunction in Parkinson disease. Laryngoscope (in press).

8) Ludlow, C., Coulter, D. and Gentges, F.: The differential sensitivity of frequency perturbations to laryngeal neoplasms and neuropathologies. In Bless, D.M. and Abbs, J.H. (Eds.) Vocal Fold Physiology - Contempory Research and Clinical Issues. College-Hill Press, San Diego (1983).

9) Neiman, R.F., Mountjoy, J.R. and Allen, E.L.: Myasthenia gravis focal to the larynx. Arch. Otolaryngol. 101: 569-570, 1975.

10) Rontal, M., Rontal, E., Leuchter, W. and Rolnick, M.: Voice spectrography in the evaluation of myasthenia gravis of the larynx. Ann. Otol. Rhinol. Laryngol. 87(5): 722-728, 1978.

11) Titze, I.R.: Parameterization of the glottal area, glottal flow and vocal fold contact area. Journal of Acoustical Society of America, 75: 570-580, 1984.

12) Yumoto, E., Gould, W.J. and Baer, T.: Harmonics-to-noise ratio as an index of degree of hoarseness. Journal of the

Acoustical Society of America, 71: 1544-1550, 1982.

13) Koike, Y.: Vowel amplitude modulatins in patients with laryngeal diseases. Journal of the Acoustical Society of America, 45: 839-844, 1968.

14) Yumoto, E., Sasaki, Y. and Okamura, H.: Harmonics-to-Noise ratio and psychophysical measurement of the degree of hoarseness. Journal of Speech and Hearing Research, 27: 2-6, 1984.

15) Ringel, S.: Diseases of muscle. In Handbook of Clinical Neurology. P.J. Vinken and G.W. Bruyn. North Holland Publishing, Amsterdam, 1979.

16) Harper, P.S.: Myotonic Dystrophy. W.B. Saunders, Philadelphia, PA, 1979.

17) Weinberg, B., Bosma, J.F., Shanks, J.C. and DeMyer, W.: Myotonic dystrophy initially manifested by speech disability. J. Speech Hear. Dis. 33: 51-59,

18) McDowell, H.F., Lee, J.E. and Sweet, R.D.: Extrapyramidal disease. In: Clinical Neurology. A.B. Baker and L.H. Baker. Harper and Row, Philadelphia, PA, 1982.

19) Merritt, H.H.: A Textbook of Neurology. Lea and Febiger, 1979.

20) Goodgold, J. and Eberstein, A.: Chapter 8 Myopathy. In: Electrodiagnosis of neuromuscular diseases. Williams and Wilkins, Baltimore, 1983.

21) Ramig, L.A. Acoustic analysis of voice of patients with Huntington's disease and individuals "at risk": implications for early diagnosis. A paper presented to the Clinical Dysarthria Conference, Tucson, Arizona, 1984.

22) Murry, T., Monroe, N., Caligiuri, M. and Grote, C.: Vowel formants and formant transitions in dysarthria associated with Huntington's disease. A paper presented to the Clinical Dysarthria Conference, Tucson, Arizona, 1984.

23) Murry, T. and Caligiuri, M.: Vowel formant characteristics in Huntington's disease. A paper presented at 106th meeting of Acoustical Society of America, San Diego, November, 1983.

24) Izdebski, K. and Dedo, H.H.: Characteristics of vocal tremor

in spastic dysphonia: a prelimary study. Presented at the Eighth Symposium: Care of the Professional Voice, The Voice Foundation: New York City, June, 1979.

25) Hayden, M.R.: <u>Huntington's Chreas</u>. Berlin: Springer-Verlag (1981).

26) Petajan, J.H., Jarcho, L.W. and Thurman, D.J.: Motor unit control in Huntington's disease: a possible presymptomatic test. In: <u>Advances in Neurology</u>. T.N. Chase, N.S. Wexler, T. Barbeau. Raven Press: New York, 1979.

27) Shoulson, I. and Fahn, S.: Huntington's disease: clinical care and evaluation. <u>Neurology</u>, 29: 1-3, 1979.

28) Guscella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E., Watkins, P.C., Ottina, K., Wallace, M.r., Sakaguchi, A.Y., Young, A.B., Shoulson, I., Bonilla, E. and Martin, J.C.: <u>Nature</u>, 306: 234-238, 1983.

29) Netsell, R.: Speech motor control: theoretical issues with clinical impact. In: <u>Clinical Dysantria</u>. W.R. Berry. College-Hill Press: San Diego, CA 1983.

30) Ramig, L.A. and Wood, R.: Fiberoptic observation of phonation in a patient with Huntington's disease. Personal communication, 1984.

31) Aronson, A.E.: Personal Communication, 1983.

32) Abbs, J.H., Hunker, C.J. and Barlow, S.M.: Differential speech motor subsystem impairments with suprabulbar lesions: neurophysiological framework and supporting data. In: W.R. Berry. College-Hill Press: San Diego, CA, 1983.

ACOUSTIC CORRELATES OF AGING

Lorraine A. Ramig, Ph.D.
University of Colorado, Boulder

Ronald C. Scherer, Ph.D.
The Denver Center for the Performing Arts

Ingo R. Titze, Ph.D.
The University of Iowa
and
The Denver Center for the Performing Arts

ABSTRACT

Speaker age may be an important variable in the assessment of normal and pathological voices as well as in voice identification and recognition. It is therefore necessary to determine acoustic variables which predict not only chronological age of speakers but also listeners' perception of speaker age. Auditory and acoustic studies were carried out on vowel productions of 24 subjects, aged 25-75 years, and representing good and poor levels of physiological condition (L.A. Ramig and R.L. Ringel, J. Speech Hear. Res. 26, 22-30, 1983). Results indicate that age could be predicted only in the poor physiological condition group. Acoustical analyses, including GLIMPES (I.R. Titze, J. Acoust. Soc. Am. 75, 570-580, 1984), derived mean fundamental frequency, harmonics-to-noise ratio (E. Yumoto, W.J. Gould, and T. Baer, J. Acoust. Soc. Am. 71, 1544-1550, 1982), amplitude perturbation and period perturbation. Multiple regressions revealed that this set of variables was a significant predictor ($p<0.01$) of age ratings. Implications regarding acoustic measures related to listeners' perception of aging and associated laryngeal characteristics will be discussed. (Work supported in part by NIA).

INTRODUCTION

Ptacek and Sander (1966) and Shipp and Hollien (1969) are among those who have reported that listeners are able to judge accurately the chronological age of a speaker by listening to recorded speech samples.

Ptacek and Sander (1966) reported listener accuracy of 78% in age ratings made from sustained vowel phonations. These findings, together with those of Ryan and Burk (1974), Linville and Fisher (1980), and others, suggest that acoustic characteristics of voice may change with advancing age and that listeners may use specific acoustic characteristics in their judgments of age.

Age-related changes in voice may be associated with degeneration in phonatory function attributed to chronological aging. Mean fundamental frequency, jitter and shimmer are among the acoustic variables researchers such as Wilcox and Horii (1980), Shipp and Hollien (1969), and others have reported to change with advancing age and to accompany possible laryngeal degeneration. Kahane (1982), for example, has suggested that age-related changes in the vocal ligament may introduce structural instability into the vocal mechanism that may cause the increased jitter associated with chronological aging.

In addition, Ramig and Ringel (1983) hypothesized that because changes in the physiological condition of the body may not parallel chronological aging, physiological condition may be an important variable in the study of the aging voice. They observed significant differences in selected vocal acoustic characteristics between speakers of the same chronological ages but of different levels of physiological condition.

At present, the role of physiological condition in the study of vocal aging has not been established. Furthermore, the relationship between acoustic characteristics associated with vocal aging and listeners' perception of age is unresolved. This

relationship between acoustic characteristics associated with vocal aging and listeners' perception of age is unresolved. This

study was designed to investigate the relationships among selected acoustic measures of voice, speakers' chronological ages, listeners' ratings of age, and speakers' physiological condition as exressed in Figure 1.

METHODS

Speakers

Thirty males in each of three chronological age groups (25-35, 45-55, and 65-75) participated in tasks to assess overall physiological condition. Measurements were made of the following variables, traditionally used to estimate physiological

FIGURE 1

RELATIONSHIPS OF INTEREST

ACOUSTIC MEASURES AND CHRONOLOGICAL AGES OF SPEAKERS

ACOUSTIC MEASURES AND LISTENERS' RATINGS OF AGE

LISTENERS' RATINGS OF AGE AND CHRONOLOGICAL AGES OF SPEAKERS

(GOOD AND POOR PHYSIOLOGICAL CONDITION)

condition: resting heart rate, resting systolic and diastolic blood pressure, percent fat, and forced vital capacity (Astrand, 1969; Brozek, 1952; Damon, 1974). From each of the three age groups, the eight subjects who performed the best and the eight who performed the poorest on these measures were selected. These 48 subjects comprise the experimental group (Ramig and Ringel, 1983). Measures of their body physiology are presented in Table 1.

Speech Samples

Voice recordings (Ampex model 351) were made in a sound treated booth while subjects sustained phonation of the vowel /a/ ("ah") at comfortable fundamental frequency and intensity levels. Voice samples, two seconds in duration, were taken from the mid-section of each vowel for further acoustic and perceptual analyses.

Acoustic Analysis

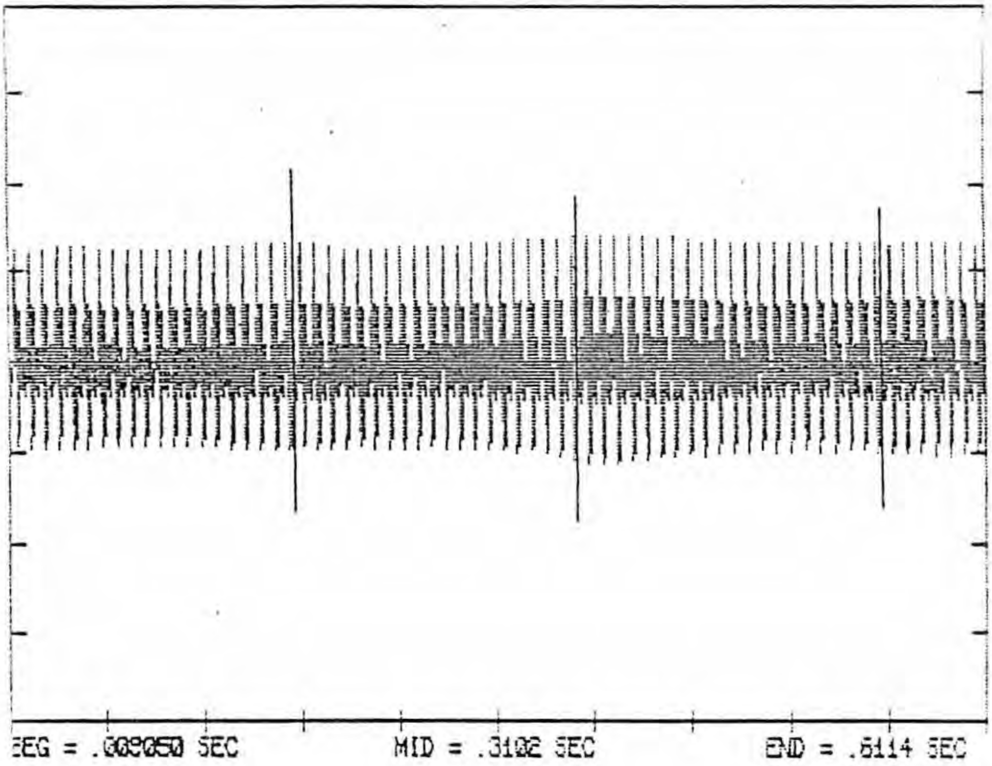
A subset of 24 of these recordings, four samples randomly selected from each age and physiological condition subgroup, were digitized at 20 thousand samples per second. Acoustic analyses (GLIMPES, Titze, 1984) were carried out on consecutive 20 cycle windows for the duration of the two seconds. This is illustrated in Figure 2. From these analyses, measures of the following acoustic variables, previously associated with vocal aging or

TABLE 1. Mean values and associated ranges of measures of body physiology.

Groups		*Age*	*Resting heart rate (bpm)*	*Systolic blood pressure (mmHg)*	*Diastolic blood pressure (mmHg)*	*% Fat*	*Forced vital capacity (L)*
Young age	mean	29.5	55.9	124.8	76.3	16.1	5.16
Good condition	range	26–35	44–64	115–140	70–80	14–20	4.52–6.47
Young age	mean	32.3	85.5	154.6	98.0	30.5	5.04
Poor condition	range	25–38	68–110	136–170	88–100	28–34	3.77–6.25
Middle age	mean	53.0	59.9	121.5	76.8	18.0	5.02
Good condition	range	46–56	52–68	116–128	70–84	15–22	4.16–7.30
Middle age	mean	52.6	73.5	161.3	95.8	28.1	4.26
Poor condition	range	42–59	50–88	138–180	82–100	23–35	3.23–6.00
Old age	mean	67.5	62.5	131.8	76.4	18.7	3.97
Good condition	range	62–75	56–76	118–142	68–80	15–21	2.68–5.05
Old age	mean	69.1	82.8	163.0	92.0	26.8	3.41
Poor condition	range	64–74	72–94	154–174	80–100	22–37	2.03–4.26

(Ramig and Ringel, 1983)

FIGURE 2



pathology, were derived: fundamental frequency, jitter, shimmer and harmonics-to-noise ratio (Yumoto, Gould and Baer, 1982). The means were obtained by averaging each measure across the windows within each 2-second phonation. The standard deviations of the measures across the windows were also obtained. We reasoned that the standard deviation of a measure across the windowed sections would reveal laryngeal instabilities that might relate to the chronological age or the perceived age of a speaker. The formula for the perturbation factors (jitter and shimmer), the coefficient of variation of fundamental frequency and harmonics-to-noise ratio are presented in Figure 3.

RESULTS

Acoustic variables vs. chronological age.

The relationship between the acoustic measures and chronological age of speakers are presented in Table 2. Stepwise multiple regression analyses with these acoustic measures as predictors and chronological age as the criterion, revealed that, when all 24 speakers were considered, 64% of the variance in chronological age was explained by this combination of acoustic measures. However, when the same multiple regression analyses were carried out for the good and poor physiological condition subgroups, these relationships changed as expressed in Table 3.

FIGURE 3

Calculations

Coefficient of Variation:

$$CV: 100\sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}}\Big/\bar{x}$$

Perturbation Factors:

with linear trend:

$$PF: \frac{100\sum_{i=2}^{N}\left|x_i-x_{i-1}\right|}{(N-1)\bar{x}}$$

without linear trend:

$$PF: \frac{100\sum_{i=2}^{N-1}\left|.5\,(x_{i+1}+x_{i-1})-x_i\right|}{(N-2)\bar{x}}$$

Harmonics-to-Noise Ratio[*]:

$$H/N: N\int_0^T f_A^2(t)\,dt\Big/\sum_{i=1}^{N}\int_0^{T_i}\left[f_i(t)-f_A(t)\right]^2 dt$$

$$\text{where}\quad f_A(t)=\sum_{i=1}^{N}\frac{f_i(t)}{N}$$

---

[*] Yumoto, E., Gould, W.J., and Baer, T. (1982). "Harmonics-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. 71, 1544-1550.

TABLE 2

MULTIPLE REGRESSION

ACOUSTIC MEASURES WITH CHRONOLOGICAL AGE

	R^2
JITTER SD	.15
H:N SD	.33
FO CV	.42
JITTER $\bar{X}$	.45
SHIMMER $\bar{X}$	.55
H:N $\bar{X}$	.62
FO $\bar{X}$	.63
SHIMMER SD	.64*

*$p \angle .05$

TABLE 3

MULTIPLE REGRESSION

Acoustic measures with chronological age

Good condition	R^2	Poor condition	R^2
Fo CV	.36*	Fo $\bar{X}$	.50
		H:N $\bar{X}$	.60
		Jitter SD	.65
		H:N SD	.77
		Shimmer SD	.80*

*p ∠.05

For speakers in good physiological condition, only 36% of the variance in chronological age was explained, and was explained by only one acoustic variable - the coefficient of variation of fundamental frequency, whereas 80% of the variance in chronological age was explained for poor condition speakers by the combination of mean fundamental frequency, mean harmonics-to-noise ratio, jitter standard deviation, harmonics-to-noise ratio standard deviation and shimmer standard deviation.

Acoustic variables vs. perceptual ratings of age.

Inasmuch as these acoustic measures explained 80% of the variance in chronological age for poor condition speakers, it is of interest to determine if such a strong relationship existed between the acoustic measures and listeners' ratings of age.

To address this issue, 30 listeners rated these 2 second speech samples on a five point equal appearing interval scale from which mean scale values for each speech sample were derived. Within and between task intrajudge reliability for these ratings were .93 and .74 (Person; Bruming and Kintz, 1963). Reliablility for group mean ratings was .93 (Ebel; Guilford, 1954).

Multiple regression analyses using these acoustic measures as predictors revealed that over all speakers, 68% of the variance in listeners' ratings of age could be explained by this combination of acoustic measures. These findings are summarized in Table 4. The same analyses carried out for good and poor condition speakers summarized in Table 5, revealed that 97% and 86% of the variance in listeners' ratings respectively, could be explained by this combination of acoustic measures.

Listener ratings of age vs. chronological age.

The relationship between listeners' ratings of age and actual chronological age of speakers is presented in Table 6. The amount of variance in listeners age ratings that could be explained by chronological age was significant only for speakers in poor physiological condition.

SUMMARY

Table 7 summarizes regression analyses for all speakers and conditions. The following observations are of importance: the acoustic measures studied here, which were selected for their previous association with vocal aging or pathology, explained 64% of the variance in speakers' actual chronolgical ages and 68% of the variance in listeners' ratings of age.

Of particular importance however, is how the relationship between the acoustic measures and chronological age differed

TABLE 4

MULTIPLE REGRESSION

ACOUSTIC MEASURES WITH LISTENERS' RATINGS OF AGE

ALL SPEAKERS

	R^2
JITTER $\bar{X}$	.51
H:N SD	.57
JITTER SD	.61
FO CV	.63
SHIMMER SD	.65
FO $\bar{X}$	.66
H:N $\bar{X}$	.67
SHIMMER $\bar{X}$	.68*

*$P \angle .05$

TABLE 5

MULTIPLE REGRESSION

Acoustic measures with listeners' ratings of age

Good condition

	R^2
Fo CV	.65
Jitter $\bar{X}$	.80
H:N SD	.89
Fo $\bar{X}$	.92
Shimmer SD	.93
Shimmer $\bar{X}$	.93
Jitter SD	.97
H:N $\bar{X}$	.97*

Poor condition

	R^2
Jitter $\bar{X}$	.32
Fo $\bar{X}$	.60
H:N SD	.69
Shimmer SD	.74
Jitter SD	.86*

*p ∠.05

TABLE 6

LISTENERS' RATINGS OF AGE AND CHRONOLOGICAL AGE OF SPEAKERS

	R^2
ALL SPEAKERS	.21*
GOOD CONDITION	.07
POOR CONDITION	.40*

*$P < .05$

TABLE 7

SUMMARY

			R^2
ACOUSTIC MEASURES	AND	CHRONOLOGICAL AGE OF SPEAKERS	
		ALL SPEAKERS	.64*
		GOOD CONDITION	.36*
		POOR CONDITION	.80*
ACOUSTIC MEASURES	AND	LISTENERS' RATINGS OF AGE	
		ALL SPEAKERS	.68*
		GOOD CONDITION	.97*
		POOR CONDITION	.86*
LISTENERS' RATINGS OF AGE	AND	CHRONOLOGICAL AGE OF SPEAKERS	
		ALL SPEAKERS	.21*
		GOOD CONDITION	.07
		POOR CONDITION	.40*

$^{*}P < .05$

between speakers in good and poor physiological condition.

One could speculate that old speakers in good physiological condition have not experienced degenerative changes characteristically associated with vocal aging. If the bodies of these individuals have not experienced typical age-related degeneration, it could be hypothesized that neither have their laryngeal mechanisms. Thus, the acoustic measures would be unable to explain a large amount of the variance in these speakers' chronological ages. On the other hand, the laryngeal mechanisms of speakers in poor physiological condition may have experienced degenerative changes in the neurological, muscular, cartilaginous or other laryngeal subsystems, and thus the acoustic analysis would be able to predict a significant amount of the variance in their chronological ages.

Inspection of the relationship between chronological age of speakers and listeners' ratings of age, presented in Table 7 supports this interpretation as well. Chronological age explained a significant amount of the variance in listeners' ratings of age only for speakers in poor condition.

It is likely that listeners were able to use the age related acoustic information available in the voices of poor condition speakers to acheive some degree of accuracy in their age ratings. On the other hand, the acoustic information available in voices of good condition speakers did not allow accurate ratings of age.

It is apparent from inspection of the relationship between acoustic measures and listeners' ratings of age that the measures used in this study have within them the information listeners used consistently in their age ratings. For speakers in poor condition, this acoustic information was also related to their chronological age. For good condition speakers however, this acoustic information was not.

Our findings, using these acoustic measures, may help to guide us in constructing more specifically relevant measures to predict speakers' chronological age. From this subset of acoustic variables, it appears that measures including frequency and waveform information were of most importance, with amplitude measures contributing least. In addition, these findings suggest that the standard deviation measure may be an important contribution to a set of significant acoustic predictors.

Continued research on the aging voice is necessary. Findings from such work, as reported here, may be of value in a number of important research areas including the relationship between general body aging and laryngeal aging, differential diagnosis between the normal aging voice and diseased voices, the relationship between degenerative laryngeal tissue changes and acoustic/biomechanical measures, speaker recognition and speaker identification.

ACKNOWLEDGEMENTS

The authors wish to thank Nancy Molzon and Vern Vail for their assistance.

REFERENCES

Astrand, I. "Electrocardiaographic changes in relation to the type of exercise, the work load, age and sex," I.H. Blackburn (Ed.), Measurement in Exercise Electrocardiography, Springfield, IL: Charles C. Thomas, 1969.

Brozek, J. "Changes of body composition in man during maturity and their nutritional implication," Federal Proceedings, 1952, II, 784-793.

Bruning, J.L. and Kintz, B.L. Computational Handbook of Statistics, Glenview, IL: Scott, Foresman, 1968.

Damon, E.L. "An experimental investigation of the relationship of age to various parameters of muscle strength," In H.A. deVries (Ed.), Physical Education and Athletics. Dubuque: William C. Brown, 1974.

Guilford, J.P. Psychometric Methods, New York: McGraw Hill, 1954.

Kahane, J. "Age related changes in the elastic fibers of the adult male vocal ligament," in V. Lawrence (Ed.), Transcripts of the Eleventh Symposium Care of the Professional Voice, New York: The Voice Foundation, 1982.

Ptacek, P. and Sander, E. "Age recognition from voice," J.

Speech and Hear. Res., 1966, 9, 273-277.

Ramig, L. and Ringel, R. "Effects of physiological aging on selected acoustic characteristics of voice," J. Speech and Hear. REs., 1983, 26, 22-30.

Ryan, W. and Burk, K. "Perceptual and acoustic correlates of aging in the speech of males," J. Comm. Dis., 1974, 7, 181-192.

Shipp, T. and Hollien, H. "Perception of the aging male voice," J. Speech and Hear. Res., 1969, 12, 704-710.

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

Wilcox, K.A. and Horii, Y. "Age and changes in vocal jitter," J. Gerontol., 1980, 35, 194-198.

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