Journal of Phonetics (2001) **29,** 365–381
doi:10.006/jpho.2001.0149
Available online at http://www.idealibrary.com on IDEAL®

Toward a taxonomy of nonmodal phonation

Bruce R. Gerratt and Jody Kreiman*

Bureau of Glottal Affairs, Division of Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Center, Los Angeles, CA 90095-1794, U.S.A.

The study of nonmodal phonation, like the study of other aspects of voice quality, spans many disciplines. Descriptions of such phonation abound, but variations in scope, purpose, terminology, measurement technique, and level of description make it difficult to compare vocal phenomena across disciplines, or even across studies within a single discipline. We demonstrate how hypotheses about which kinds of nonmodal phonation types are the same and which are different can be tested by studies of listeners' perceptions. Evidence suggests that period-doubled phonation, amplitude modulations, and vocal fry form perceptually distinctive qualities, which also have consistent acoustic and physiological correlates. Evidence is much more ambiguous for qualities like breathiness and creak, which vary continuously from modal phonation. A common theoretical framework for the description of vocal quality may eventually eliminate many impediments to unified description.

1. Introduction

The study of a phenomenon by different disciplines often enriches our knowledge about it, but the process can also lead to confusion as researchers with varied backgrounds pursue their individual interests. A case in point is the study of nonmodal phonation, which has a long intellectual history spanning disciplines from linguistics to biomedicine, and from physics to music appreciation. For example, linguists have examined the way in which changes in voice quality signal changes in meaning or offer cues to the grammatical structure of utterances; otolaryngologists are interested in voice quality as a symptom of disease; and singing teachers are concerned with how voice quality changes across a singer's range. The variety of questions motivating studies of nonmodal phonation has resulted in a confusing literature that is spread across many journals, and that reflects the different priorities, methods, and terminological traditions of unrelated academic areas.

The term "modal" originated in the study of vocal register in singing, and designated phonation that includes the range of fundamental frequencies normally used for speaking or singing—that is, the mode of an $F0$ distribution for an individual (e.g., Hollien, 1974). In this use, it contrasts with falsetto, chest, loft, head, pulse or fry, and other phonatory registers. Outside the singing literature, "modal" also appears to mean the

*Address correspondence to J. Kreiman. E-mail: jkreiman@ucla.edu

0095-4470/01/040365 + 17 $35.00/0

usual or baseline kind of phonation, and "nonmodal" is used to describe any phonation that differs from or contrasts with the most "usual" variety. Many kinds of phonation (including breathy voice, creak, period doublings, vocal fry, diplophonia, and so on) may contrast with modal, depending on the particular intellectual tradition, and the particular way in which differences are defined varies across studies and disciplines. A physiological observation may lead to a term such as "ventricular phonation", which refers to the presumed involvement of the ventricular folds in sound production. Similarly, acoustic analysis may reveal a signal that differs from the kind usually observed; for example, the term "harmonic doubling" refers to the appearance of interharmonics in an acoustic spectrum. Finally, a voice may differ perceptually from modal phonation. For example, "creaky voice" and "diplophonia" refer to the perceptual attributes of the signals.

This variability in terminology, combined with the scattering of literature across various academic disciplines, has resulted in substantial confusion about the phenomena being described. Terms that are derived from different domains (for example, acoustic *vs.* physiological descriptions) may often characterize the same vocal phenomenon, and a single term may be applied to rather different kinds of phonation. Further, many terms imply a particular physiologic process or perceptual reality that has not withstood or even necessarily received experimental scrutiny. As a result, readers often have significant difficulty determining precisely what vocal phenomenon an author is referring to. Thus, the study of nonmodal phonation has stalled in a disorderly, descriptive stage, and the important question of which kinds of phonation are the same and which are different remains largely unaddressed.

In this paper, we review the literature on several types of nonmodal phonation, and attempt to unify descriptions across disciplines and descriptive domains, as far as this is possible. Because the study of nonmodal phonation has not proceeded systematically, the data necessary to motivate even a provisional taxonomy are not available. For this reason, we also provide an example of the kind of study we believe is needed to establish unique, valid categories for phonation types. In our view, the question of which nonmodal vocal phenomena are distinct should be initially addressed in the perceptual domain, because voice quality is an auditory-perceptual phenomenon. Although two voice samples may have been produced rather differently, and the acoustic waveforms may look rather different, these differences are important only if they result in a perceptually salient difference in vocal quality. Basing the primary level of description on perception ensures that we first determine the psychological reality of the categories, a process that we hope will lead to a parsimonious list of nonmodal categories. In addition, listeners can easily judge similarities and differences among auditory stimuli, and such judgments provide empirical evidence for the existence of distinct vocal phenomena. Once these are determined, acoustic and physiological correlates can be identified and used to confirm or even help identify the classification of an utterance. The following example illustrates this process.

2. Supraperiodic phonation

Many vowel segments, whether produced by normal speakers or by speakers with vocal pathology, are neither periodic nor aperiodic in the classic sense. Rather, these acoustic signals appear "supraperiodic" in nature. That is, they demonstrate a repeating pattern

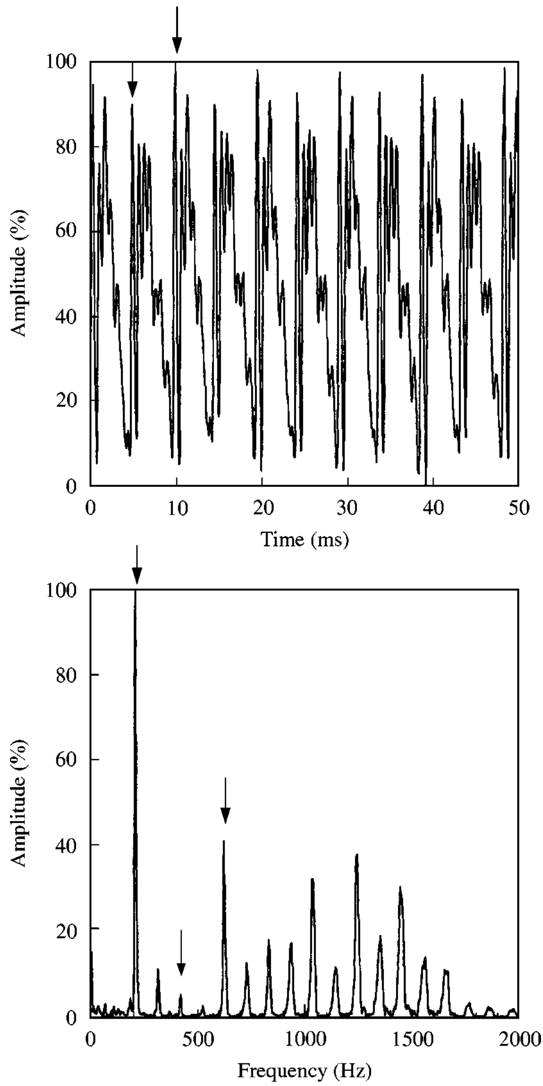

Figure 1. Acoustic waveform and linear FFT spectrum of a voice with alternating long–short/large–small cycles. Arrows in the top panel indicate the two repeating cycles; arrows in the lower panel indicate the harmonics of the "fundamental" frequency. Note interharmonics between the labeled harmonics.

that extends over more than one apparent glottal cycle. Because there is more than one repeating pattern in these waveforms, it is difficult to define or identify the fundamental frequency of phonation, either acoustically or perceptually.

Figure 1 shows one kind of acoustic signal with this property.[1] The waveform at the top of the figure shows a characteristic pattern of different vocal periods alternating in duration, amplitude, or both. The spectrum at the bottom of the figure shows

[1] Audio samples are available at http://www.surgery.medsch.ucla.edu/glottalaffairs/.

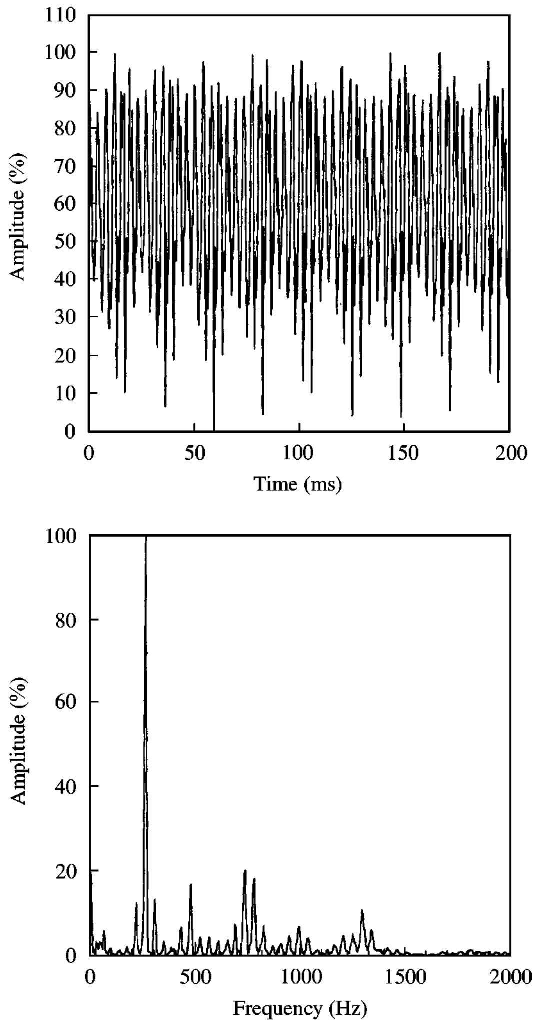

Figure 2. Acoustic waveform and linear FFT spectrum of a voice with amplitude modulations. The prominent harmonic at about 260 Hz corresponds to the rapidly-repeating pattern of the waveform. Harmonics are separated by about 44 Hz, corresponding to the frequency of the modulating envelope.

interharmonics between the expected harmonics of the apparent fundamental frequency (about 209 Hz). Many authors have described such phonation, which is fairly common. For example, Klatt & Klatt (1990) reported that they observed such patterns in 25% of the utterances from normal speakers they studied. Similar frequencies of occurrence have been reported in speakers with vocal pathology (Kelman, Gordon, Morton & Simpson, 1981; Dejonckere & Lebacq, 1983).

Figure 2 shows a second kind of supraperiodic voice signal. Instead of pairs of vocal cycles alternating in period and/or amplitude as in Fig. 1, this waveform resembles a relatively high-frequency wave (about 260 Hz) modulated by a much lower frequency envelope (about 44 Hz). An FFT of this waveform (shown in the lower portion of the figure) shows a complex harmonic structure.

These two kinds of phonation have been repeatedly described for centuries. In fact, the second kind of signal was apparently observed in an excised larynx model by Ferrein in 1741 (cited by Cooper, 1989; see also Herzel, 1993). Table I lists representative studies from linguistics, physics, otolaryngology, engineering, speech science, and musicology. As the table shows, many different terms have been used for each of these kinds of phonation, and identical terms have been used to designate both kinds of acoustic signals. Further, descriptions generally refer to the appearance of the acoustic signal or to the physiology that produced the signal. The quality of both kinds of signal is typically described as "rough" or "bitonal", but no perceptual evidence exists to support their classification as "same" or "different". To provide such evidence, we undertook the following study.

2.1. *Method*

2.1.1. *Stimuli*

Three sets of 15 voices were selected from an existing library of recordings. Each set included eight male and seven female speakers with vocal pathology, including functional, neurogenic, and mass lesion disorders. All voices were recorded using a microphone placed 5 cm away from the speaker's lips. Speakers were asked to sustain the vowel /a/ for as long as possible. Utterances were low-pass filtered at 8 kHz and sampled at 20 000 samples/s with 12-bit resolution. A 2-s sample was excerpted from the middle of each utterance and stored.

Voices were selected by examining waveforms and linear FFT spectra, without regard to perceptual quality. Voices in the first set ("period-doubled voices") demonstrated a clear pattern of cycles alternating in amplitude and/or frequency, as in Fig. 1. FFT spectra revealed a single fundamental frequency and interharmonics. Waveforms for the second set of voices ("amplitude-modulated voices") showed a pattern of amplitude modulation by a lower frequency signal, as in Fig. 2. The third set ("noisy voices") included voices that were highly aperiodic but not period doubled or amplitude modulated (Fig. 3). Vocal quality was consistent for the entire duration of each voice sample. Samples with significant variations in quality within an utterance were discarded.

To ensure that acoustic criteria were applied consistently, the following pretest procedure was applied. The initial selection of voices was made independently by the second author, using the acoustic criteria described above. An unlabeled printed FFT spectrum and a waveform display were prepared for each voice. Spectra and waveforms were randomized and given to the first author, who was asked to sort the voices into three categories using the criteria described above. Only voices that were consistently categorized by both authors were retained.[2]

2.1.2. *Listeners and procedure*

Listeners included four clinicians and four linguists, all of whom were familiar with nonmodal phonation. They judged the dissimilarity of all possible pairs of the voices on

[2] Only one voice (that of a woman with an unusually low $F0$ but normal formant frequencies) was miscategorized.

TABLE I. Selected labels for supraperiodic phonation

Label	Description	References
Amplitude/frequency modulation	Simultaneous quasi-periodic changes in frequency and amplitude	Mazo, Erickson & Harvey (1995)
Bicyclic modulation	Alternating large–small cycles/sudden increase in the number of harmonics	Kiritani, Niimi, Imagawa & Hirose (1995)
Bicyclicity	Alternating long–short or large–small cycles	Kreiman, Gerratt, Precoda & Berke (1993*a*)
Biphasic vibrations	Contain a strong subharmonic at half the $F0$	Timcke, von Leden & Moore (1959)
Biphonation	Two independent $F0$s, appearing as low-frequency modulation in the corresponding time series	Sirvo & Michelsson (1976), Tigges, Mergell, Herzel, Wittenberg & Eysholdt (1997), Herzel & Reuter (1997)
Creak	Low-frequency, damped pulses; may be single or double	Laver (1980)
Dicrotic dysphonia	Double opening, then long closed phase	Moore & Von Leden (1958)
Diphthonia	Simultaneous production of two tones differing in pitch	Morgan (1882)
Diplophonia	Simultaneous production by the voice of two separate tones[3]	Ward, Sanders, Goldman & Moore (1969), Dejonckere & Lebacq (1983)
	Vocal folds oscillate out of phase	Dejonckere & Lebacq (1983)
	Two different cycle lengths in alternating pairs	Ludlow, Coulter & Gentges (1983)
	Quasi-periodic variations in vocal cord vibration; low-frequency modulation of the voice	Kiritani, Hirose & Imagawa (1993)
Diplophonic double pulsing	First of a pair of periods is delayed in time and reduced in amplitude	Klatt & Klatt (1990)
Diplothongia	Simultaneous production by the voice of two separate tones	Ward et al. (1969)
Double harmonic break	Extra set of harmonics parallel to $F0$ and its harmonics	Sirvio & Michelsson (1976)
Double voice	Simultaneous production by the voice of two separate tones	Morgan (1882), Jones (1935), Ward *et al.* (1969)
Dycrotic pattern	Small, short wave, followed by larger, longer wave, followed by significant damping	Cavallo, Baken & Shaiman (1984)
Frequency breaks	Abrupt one-octave change in $F0$	Kelman (1981)
Furcation	A split in $F0$	Sirvio & Michelsson (1976)

[3] Schreibweiss-Merin & Terrio (1986) cite additional terms used for this kind of phonation, including subglottal grumble tones, subtonal flutter tones, fluttering diplophonia, voiced masking, vicarious voice, and ventricular diplophonia, but do not provide acoustic descriptions.

TABLE I. (Continued)

Label	Description	References
Growl phonation	Alternating pulses of high and low peak-to-peak amplitude	Rose (1988)
Harmonic doubling	Extra harmonics between the harmonics of the fundamental	Keating (1980)
Laryngealization	Narrow glottal pulses, low $F0$, and optional diplophonic irregularities	Klatt & Klatt (1990)
Low-frequency modulations	Sidebands of the main spectral peaks	Herzel, Berry, Titze & Steinecke (1995)
Low-frequency segments	Abrupt 1-octave drop in $F0$	Ramig, Scherer, Titze & Ringel (1988)
Period-doubling bifurcation	Presence of $F0/2$ subharmonic frequency	Svec, Schutte & Miller (1996)
Pulse register phonation	Low-frequency, pulse-like vibratory pattern	Hollien (1974)
Subharmonic regime	Double period and large amplitudes	Herzel (1993)
Subharmonic vocalization	Diplophonia	Tigges *et al.* (1997)
Syncopated rhythm	Double opening, then long closed phase	Timcke *et al.* (1959)
Vocal (or glottal) fry	Low-frequency aperiodicity	Dejonckere & Lebacq (1983)
	The alternation of large and small glottal pulses	Herzel (1993)
	High-pitched phonation with intermittent subharmonics	Mazo *et al.* (1995)
Voice breaks	Alternating large–small cycles/sudden increase in the number of harmonics	Kiritani *et al.* (1995)

a 7-point equal-appearing interval scale, where "1" meant the two voices were extremely similar in quality, and "7" meant they were very different. This task was used because it required no *a priori* assumptions about the number or nature of perceptual categories in a stimulus set, their relative salience, the relationships among categories, or even whether categories of stimuli existed. Further, listeners needed no special training in identifying the different voice types or discriminating among categories. Therefore, the task itself did not focus attention on any particular stimulus attribute.

Male and female voices were presented at separate test sessions. Order of presentation was randomized across listeners. Stimuli within a voice set were rerandomized for each listener, with the constraint that a given voice never appeared in two consecutive pairs. For each voice set, listeners heard one order of each pair of voices (276 comparisons/listener for the male voice set, and 210 for the female voice set). Which voice within a pair occurred first varied at random, with the constraint that each voice occur first an equal number of times. An additional 20% of pairs (55 for the male voice set, and 42 for the female set), selected at random, were repeated so that test–retest reliability could be assessed. These repeated pairs were inserted at random into the total voice set,

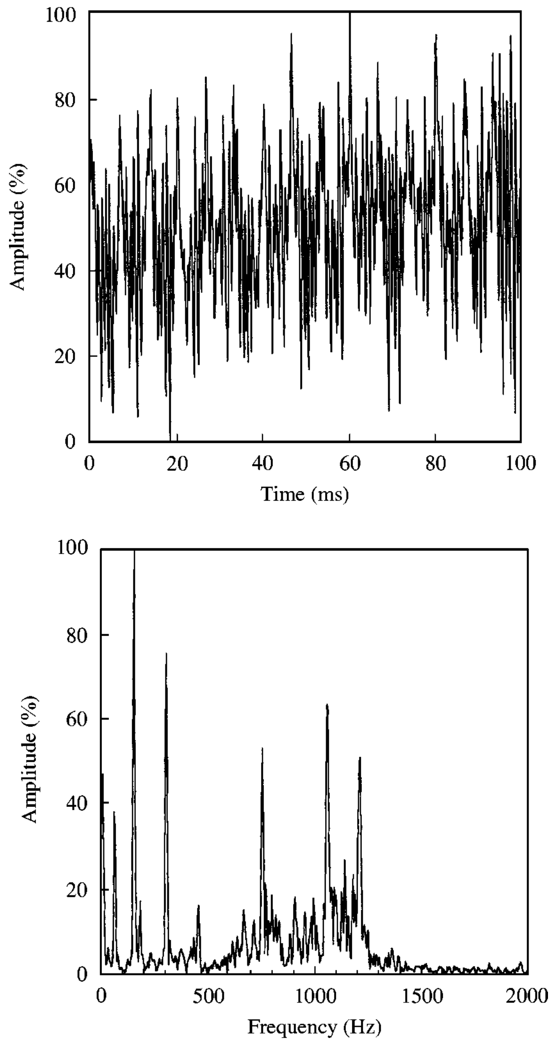

Figure 3. Acoustic waveform and linear FFT spectrum of a voice that is markedly aperiodic, but not period doubled or amplitude modulated. Individual cycles are difficult to identify in the waveform, and the spectrum shows significant amounts of interharmonic noise.

with the constraint that two identical pairs of voices were separated by at least 10 other pairs.

Voices within a pair were separated by 0.5 s. Listeners controlled the rate at which stimuli were presented. To minimize listener fatigue, testing took place in four sessions (two for each voice set) on different days. Test time for a single listener totaled about 2 h.

Listeners were instructed that they would hear pairs of moderately- to severely-disordered voices. They were asked to judge the dissimilarity of speakers in each pair with respect to voice quality only, and to ignore differences in severity of pathology, vowel quality, and $F0$ as much as possible. They were encouraged to simply listen for similarity or difference, and to avoid labeling the voices or categorizing them by apparent diagnosis. They were not told the purpose of the study, nor were they told that the stimulus voices had been selected in any particular way. To familiarize them with the range of voices to be judged, they heard the entire set of voices in a random order several times prior to each listening session.

All testing took place in free field. Listeners were seated 3 ft from a speaker in a soundtreated room. Stimuli were played through a 16-bit D/A converter at a constant comfortable listening level. Responses were recorded and stored by the computer.

2.1.3. *Multidimensional scaling analyses*

Listeners' judgements were multidimensionally scaled using a nonmetric individual differences model (Schiffman, Reynolds & Young, 1981). Separate solutions in 1–6 dimensions were found for the female and male voice sets.

2.2. *Results*

2.2.1. *Listener reliability*

The percentage of repeated ratings within $\pm$ one scale value was calculated for each listener and voice set.[4] For female voices, values ranged from 54.8 to 82.9% (mean = 70.2%; S.D. = 9.46%). For male voices, values ranged from 61.9% to 86.2% (mean = 72.3%; S.D. = 9.73%). (For a 7-point scale, chance for this measure = 38.8%.) These values are consistent with those typically found for the task used here (see Kreiman, Gerratt, Kempster, Erman & Berke, 1993*b*, for review).

2.2.2. *The scaling solutions*

Two-dimensional solutions were selected for both the male and female voice sets, based on stress, variance accounted for by the different solutions, and interpretability. The perceptual spaces for male and female voices are shown in Fig. 4. Each solution accounted for 86% of the variance in the underlying data.

K-means cluster analysis confirmed that the two dimensions separated the three voice sets into statistically significantly different groups, as shown in the figures (male voices: for D1, $F(2, 21) = 44.78$, $p < 0.05$; for D2, $F(2, 21) = 31.74$, $p < 0.05$. Female voices: for D1, $F(2, 18) = 38.66$, $p < 0.05$; for D2, $F(2, 18) = 41.13$, $p < 0.05$). Noisy voices (shown by triangles in the figures) and period-doubled voices (shown by stars) apparently overlap slightly in quality. Three voices—two noisy and one period doubled—were misclassified for the male voice set; one noisy voice was misclassified for the female voice set. Amplitude-modulated voices (shown by circles) are clearly separated from the other categories for both voice sets.

2.3. *Discussion*

As these findings indicate, differences among period-doubled, amplitude-modulated, and aperiodic voices are perceptually salient, even for untrained listeners performing an

[4]Other measures of within-listener agreement (e.g., Pearson's *r* for the first *vs.* second ratings of a voice) are not appropriate for these data, because many listeners used the low end of the scale (ratings of 1 and 2) very infrequently. The resulting limited range of ratings makes correlation values unreliable.

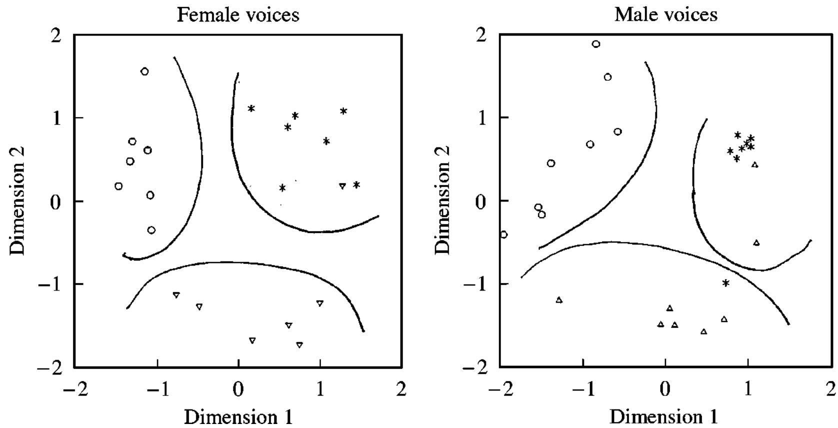

Figure 4. Multidimensional scaling results for the female and male voice sets. Period-doubled voices are shown as stars, amplitude-modulated voices as circles, and noisy voices as triangles. Curves indicate significantly different clusters of points, as described in the text.

unstructured task. Because a common perceptual quality made the voices in each group sound alike, we conclude they form distinct perceptual classes. Labeling systems for nonmodal phonation should thus include separate terms for each of these kinds of phonation.

Because period-doubled phonation and amplitude-modulated phonation differ perceptually from each other and from aperiodic phonation, examination of the underlying acoustics and physiology is a sensible next step in developing comprehensive descriptions of these kinds of nonmodal phonation. Acoustic correlates of perceived period doubling are examined in detail elsewhere (Kreiman *et al.*, 1993*a*; Gauffin, Granqvist, Hammarberg, Hertegård & Håkansson, 1995). Briefly, the greater the amplitude of the first interharmonic, the stronger the perception of period doubling becomes. The perception of amplitude-modulated voices and the production of both these kinds of phonation are only partly understood. Specific physiological correlates do appear to underlie perceived differences. For example, high-speed film evidence suggests that amplitude modulations occur when the two vocal folds vibrate at different (but close) frequencies (e.g., Ward *et al.*, 1969; Kiritani *et al.*, 1993). The production of period-doubled phonation is less well understood (see Svec *et al.*, 1996; Svec, Horacek, Sram & Vesely, 2000; or Berry, 2001), but computational modeling and studies of excised larynges suggest that this kind of phonation occurs in the presence of tension asymmetries and possibly a glottal gap (Timcke, 1956 [cited by Timcke *et al.*, 1959]; Ishizaka & Isshiki, 1976; Isshiki, Tanabe, Ishizaka & Broad, 1977; Wong, Ito, Cox & Titze, 1991; Smith, Berke, Gerratt & Kreiman, 1992).

3. Vocal fry

Studies of period doubling and amplitude modulations have mostly focused on speakers with vocal pathology, even though normal speakers also produce these kinds of phonation (Ward *et al.*, 1969; Klatt & Klatt, 1990; Svec *et al.*, 1996; Redi & Shattuck-Hufnagel, 2001). The studies reviewed above show how evidence can be obtained for the perceptual,

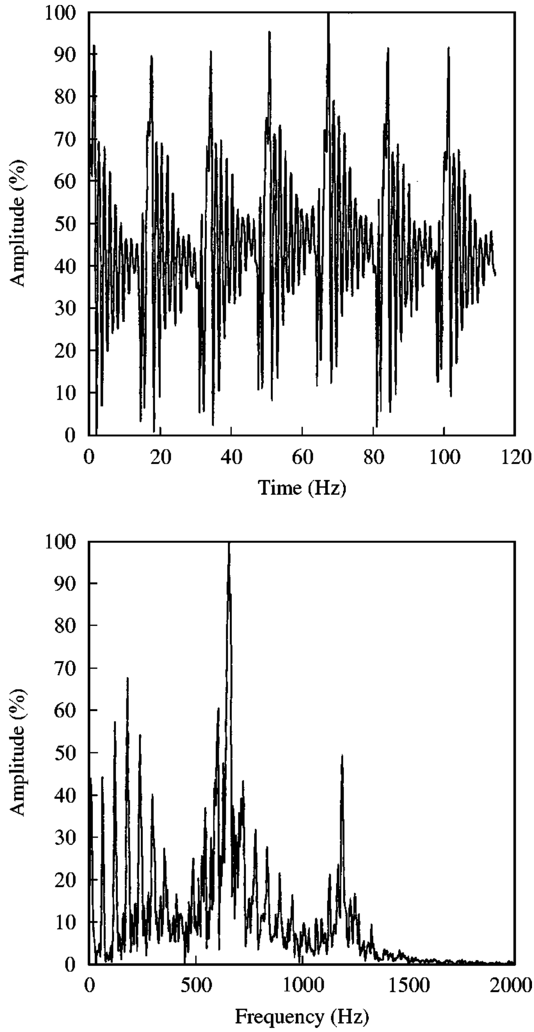

Figure 5. Acoustic waveform and linear FFT spectrum for a sample of vocal fry. Fundamental frequency is approximately 59 Hz. Note damping between pulses.

acoustic, and physiological distinctiveness of a nonmodal phonation type, and cumulative evidence across these different domains gives confidence that real differences exist in phonation and perception. Similarly comprehensive studies of normal phonation are unusual. However, one exception is the study of vocal fry (also called creak or creaky voice, laryngealization, glottalization, and pulse register phonation; see Henton & Bladon, 1988, for a review).[5]

As usually defined, vocal fry is a train of discrete laryngeal excitations, or "pulses", of extremely low frequency, with almost complete damping of the vocal tract between

[5]The study of vocal fry also focused originally on vocal pathology, but expanded to included normal speakers when investigators realized how common such phonation was (e.g., Hollien, Moore, Wendahl & Michel, 1966).

Figure 6. Position of the vocal folds during double-pulsed vocal fry phonation. Plots show two quick openings in succession, followed by a long closed phase. Adapted from Wendahl *et al.* (1963).

excitations (e.g., Hollien *et al.*, 1966). Fig. 5 shows a sample acoustic waveform produced by a normal speaker. Substantial evidence (reviewed by Hollien (1974), and Blomgren *et al.*, 1998) indicates that vocal fry is well defined physiologically, acoustically, and perceptually, and is distinct from modal phonation and from "harsh" voice. In contrast to modal voice, $F0$ in fry phonation ranges from about 7 to about 78 Hz, and does not differ for males and females (Hollien & Michel, 1968). The vocal tract damps almost completely between glottal pulses, which high-speed film analyses indicate are very short and impulse-like, with a very long closed phase and small vocal fold excursion (Moore & Von Leden, 1958; Wendahl, Moore & Hollien, 1963; Hollien, Girard & Coleman, 1977). X-ray data show the vocal folds during fry phonation are very thick and relatively short; the ventricular folds sometimes appear to come in contact with the true folds (Allen & Hollien, 1973). Airflow is much lower than during modal phonation, presumably because of the very long closed phase and short excursion (e.g., McGlone & Shipp, 1971).[6]

Finally, vocal fry is perceptually distinct from both modal and "harsh" nonfry phonation. Blomgren *et al.* (1998) reported that listeners could sort samples of vocal fry and modal phonation into categories with better than 95% accuracy; and Michel & Hollien (1968) found that vocal fry could be distinguished from "harsh" phonation with 95% accuracy. The role of pitch in listener judgments of the presence of vocal fry is not clear, because by definition fry is lower in $F0$ than modal phonation, and listeners may base their decisions largely on $F0$. However, perceptually important quality differences may also exist beyond simple pitch differences. For example, early studies using synthetic speech (Coleman, 1963; cf. Wendahl *et al.*, 1963) suggest that damping between cycles is critical for the perception of vocal fry. If the vocal tract response did not damp sufficiently between cycles, the phonation was judged to be modal. Whatever the basis for the perceptual distinction, these studies provide good cumulative evidence that vocal fry differs categorically from modal phonation.

An additional pattern intermediate between classic fry and period doubling has also been described by a number of authors (e.g., Moore & Von Leden, 1958; Wendahl *et al.*, 1963; Whitehead, Metz & Whitehead, 1984; Blomgren *et al.*, 1998; see also Table I). In this pattern, two or three pulses in rapid succession are followed by a period of significant vocal tract damping (Fig. 6). Multiply-pulsed vocal fry appears to be rather common. Cavallo *et al.* (1984) found the doubled pattern occurred in the acoustic waveforms of 88% of speakers producing vocal fry; and Blomgren *et al.* (1998) found double or triple pulsing in electroglottographic signals from 83% of male fry phonations (10 speakers, 120 utterances), and 38% of female vocal fry utterances.

Although this pattern appears to contrast with both the big–small–big–small pattern of period-doubled phonation and the sharp single pulses of classic vocal fry, it is unclear how it should be classified. Studies have shown that multiply-pulsed vocal fry is

[6] Data on subglottal pressure have varied, with studies reporting pressures higher (Murry, 1971) and lower (Blomgren *et al.*, 1998) than during modal phonation.

perceptually different from modal phonation. For example, Blomgren *et al.* (1998) reported that multiply-pulsed vocal fry was consistently identified as vocal fry, rather than modal phonation, in a forced choice classification task. However, studies comparing multiply-pulsed vocal fry to period-doubled phonation have not appeared. The term "vocal fry" is often used to indicate period-doubled phonation, possibly because both kinds of phonation have a rough, low-pitched sound, but the precise relationship among these different signals remains unknown.

4. Breathiness and other nonmodal phenomena

The patterns of nonmodal phonation described so far all differ categorically from modal phonation. That is, perceptual, acoustic, and physiologic evidence all indicate that period doubling, amplitude modulations, and vocal fry form coherent, distinct perceptual categories, and are different from modal phonation and from each other. These phonation types are characterized by a consistent change in the *kind* of vocal fold vibration, relative to modal phonation, and this change is accompanied by consistent changes in the acoustic signal and perceptual quality of the voice.

However, linguists and others have generalized the term "modal" to contrast with other kinds of phonation, including creakiness, laryngealization, and breathiness (e.g., Kirk, Ladefoged & Ladefoged, 1984; Huffman, 1987; Ladefoged, 1988; Dilley, Shattuck-Hufnagel & Ostendorf, 1996). We now turn to these other kinds of phonation—specifically, breathiness—that are often labeled "nonmodal", but that do not appear to differ in a distinctive, categorical way from modal phonation. Breathiness is used as an example here for convenience, but the same arguments can be made about roughness, hoarseness, harshness, and many other traditional vocal qualities.

Although many languages contrast breathy and modal phonation, breathiness is also used to characterize personal voice quality, and is generally treated as a continuous variable, so that we speak of voices that are extremely breathy, slightly breathy, and so on. In practice, in the absence of a phonological contrast, breathiness and modality form a continuum that is difficult to separate into "breathy" and "modal", whether sorting is based on perceived quality, acoustics, or the underlying glottal configuration (see Gordon & Ladefoged, 2001, for further discussion). First, listeners are singularly unable to agree in their judgments of how breathy a voice is (see Kreiman & Gerratt, 1999, for review). In fact, listeners do not even seem to be able to agree whether a voice is or is not breathy, except in cases where the voice is nearly aphonic. In a recent study (Kreiman & Gerratt, 2000) in which we asked 15 expert listeners to categorize each of 160 pathological voices as primarily breathy or not primarily breathy, only three were unanimously judged to be breathy. (Similar results occurred for roughness, and also for judgments of pitch in natural voices.) In contrast, it is relatively simple to label a voice as "amplitude modulated" or "not amplitude modulated" with even a little listening experience; and listeners in Blomgren *et al.* (1998) classified phonation as modal or vocal fry with better than 95% accuracy. Thus, breathiness seems to differ from the other kinds of phonation discussed in this paper, in that it apparently does not form a coherent perceptual category, and it varies continuously rather than categorically from modal phonation.

Efforts to specify the physiological substrates of breathiness have also met with little success. For example, breathiness is often defined as the impression of turbulent noise and audible escape of air through the glottis due to insufficient closure. However, voices with very large glottal gaps may not sound breathy; voices with very small noise components may sound extremely breathy; voices with large noise components may not sound particularly breathy; and so on. Similar difficulties arise with efforts to quantify breathiness acoustically (see Kreiman & Gerratt, 1999, for review). Thus, in the absence of a phonological contrast (i.e., a difference in meaning), categorically separating breathy from modal phonation is impossible, because there does not appear to be a single physiological or acoustic cue, or even a combination of cues, that consistently and reliably indicates breathiness is present.

The existence of nonmodal phonation types that do not appear to differ in consistent, specifiable ways from modal phonation raises the question of whether the distinction between modal and nonmodal phonation is actually valid and useful. The traditional definition of "modal" phonation is something like, "the kind of phonation that occurs most often". Given this definition of modal, the easiest way to define nonmodal phonation, and the definition that appears to have been informally adopted, is as anything that is not modal phonation. This negative kind of definition demonstrates how poorly developed our understanding of nonmodal phonation is. The term "nonmodal phonation" itself suggests that the distinction between modal and nonmodal is the primary distinction in a taxonomy of phonation types. The data reviewed here suggest this is not correct; instead, modal phonation (whatever that is) contrasts on the one hand with kinds of phonation that differ categorically from modal (like amplitude modulation, period doublings, and vocal fry), and on the other hand with qualities that vary continuously from modal (like breathiness). We suspect that these continuously varying qualities will provide significant impediments to the development of theory and the unification of description of the many varieties of phonation that occur in naturally-produced speech, as they have in the description of pathological voice quality.

5. Conclusions

Because the study of voice is not at present a unified scientific endeavor, very little effort has been devoted to developing formal descriptions of the many different kinds of phonation that occur across speakers and languages. The studies reviewed here demonstrate how hypotheses can be formed about nonmodal vocal classification, and then be tested and confirmed by listener perception. Evidence suggests that period-doubled phonation, amplitude modulations, and vocal fry form perceptually coherent qualities, which also have consistent acoustic and physiological correlates. Evidence is much more ambiguous for qualities like breathiness and creak that vary continuously from modal phonation.

Significant challenges remain in this area, beyond the difficulties discussed here. A primary concern in developing a descriptive framework for nonmodal phonation is removing the boundaries that exist between the many different disciplines that are interested in nonmodal phonation. The different levels of description required by different disciplines may complicate this process. For example, a linguistic contrast between two phonation types may be implemented in rather different ways across languages or speakers within a language, but as long as a contrast is present, a common label can be used to represent various vocal phenomena. On the other hand, an otolaryngologist may consider them distinctly different entities. For example, "creaky voice" is ambiguous in the literature between vocal fry, period doubling, and something like "low pitched noise" (e.g., Hollien, 1974; Keating, 1980; Ladefoged, 1988), possibly because different speakers use all these kinds of phonation to mark a single phonemic contrast. Confusion may be minimized by defining terms, selecting terms that are minimally ambiguous, and specifying the intended level of description (for example, phonetic, phonological, biomechanical, and acoustic). A common theoretical framework for the description of vocal quality may eventually eliminate impediments to unified description.

This research was supported by grant DC 01797 from the National Institute on Deafness and Other Communication Disorders.

References

Allen, E. L. & Hollien, H. (1973) Vocal fold thickness in the pulse (vocal fry) register, *Folia Phoniatrica*, **25,** 241–250.

Berry, D. (2001) Mechanisms of modal and nonmodal phonation, *Journal of Phonetics*, **29,** 431–450. doi:10.1006/jpho.2001.0148.

Blomgren, M., Chen, Y., Ng, M. L. & Gilbert, H. R. (1998) Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers, *Journal of the Acoustical Society of America*, **103,** 2649–2658.

Cavallo, S. A., Baken, R. J. & Shaiman, S. (1984) Frequency perturbation characteristics of pulse register phonation. *Journal of Communication Disorders*, **17,** 231–243.

Coleman, R. F. (1963) Decay characteristics of vocal fry, *Folia Phoniatrica*, **15,** 256–263.

Cooper, D. S. (1989) Voice: a historical perspective, *Journal of Voice*, **3,** 187–203.

Dejonckere, P. H. & Lebacq, J. (1983) An analysis of the diplophonia phenomenon, *Speech Communication*, **2,** 47–56.

Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M. (1996) Glottalization of word-initial vowels as a function of prosodic structure, *Journal of Phonetics*, **24**, 423–444.

Ferrein, A. (1746) De la formation de la voix de l'homme. Suite des memoires de mathematique et de physique tires des registres de l'Academie royale des sciences de l'annee MDCCXLI., pp. 545–579. Amsterdam: Pierre Mortier (cited by Cooper, 1989).

Gauffin, J., Granqvist, S., Hammarberg, B., He̊rtegård, S. & Håkansson, A. (1995) Irregularities in the voice, some perceptual experiments using synthetic voices. In *Proceedings of the XIII international congress of phonetic sciences*, Stockholm, pp. 242–245.

Gordon, M. & Ladefoged, P. (2001) Phonation types: a cross-linguistic overview, *Journal of Phonetics*, **29,** 383–406. doi:10.1006/jpho.2001.0147.

Henton, C. G. & Bladon, A. (1988) Creak as a sociophonetic marker. In *Language, speech and mind: studies in honour of Victoria A. Fromkin* (L. Hyman & C. Li, editors), pp. 3–29. London: Routledge.

Herzel, H. (1993) Bifurcations and chaos in voice signals, *Applied Mechanics Review*, **46,** 399–413.

Herzel, H., Berry, D. A., Titze, I. R. & Steinecke, I. (1995) Nonlinear dynamics of the voice: signal analysis and biomechanical modeling, *Chaos*, **5,** 30–34.

Herzel, H. & Reuter, R. (1997) Whistle register and biphonation in a child's voice, *Folia Phoniatrica et Logopaedica*, **49,** 216–224.

Hollien, H. (1974) On vocal registers, *Journal of Phonetics*, **2,** 125–143.

Hollien, H., Girard, G. T. & Coleman, R. F. (1977) Vocal fold vibratory patterns of pulse register phonation, *Folia Phoniatrica*, **29,** 200–205.

Hollien, H. & Michel, J. (1968) Vocal fry as a phonational register, *Journal of Speech and Hearing Research*, **11,** 600–604.

Hollien, H., Moore, P., Wendahl, R. W. & Michel, J. (1966) On the nature of vocal fry, *Journal of Speech and Hearing Research*, **9,** 245–247.

Huffman, M. K. (1987) Measures of phonation type in Humans, *Journal of the Acoustical Society of America*, **81,** 495–504.

Ishizaka, K. & Isshiki, N. (1976) Computer simulation of pathological vocal-cord vibration, *Journal of the Acoustical Society of America*, **60,** 1193–1198.

Isshiki, N., Tanabe, M., Ishizaka, K. & Broad, D. (1977) Clinical significance of asymmetrical vocal cord tension, *Annals of Otology, Rhinology and Laryngology*, **86,** 58–66.

Jones, S. (1935) Observations on a case of 'double voice'. In *Proceedings of the second international congress of phonetic sciences* (D. Jones & D. B. Fry, editors), pp. 232–235. Cambridge: Cambridge University Press.

Keating, P. (1980) Patterns of fundamental frequency and vocal registers. In *Infant communication: cry and early speech* (T. Murry & J. Murry, editors), pp. 209–233. Boston: College Hill.

Kelman, A. W. (1981) Vibratory pattern of the vocal folds, *Folia Phoniatrica*, **33,** 73–99.

Kelman, A. W., Gordon, M. T., Morton, F. M. & Simpson, I. C. (1981) Comparison of methods for assessing vocal function, *Folia Phoniatrica*, **33,** 51–65.

Kiritani, S., Hirose, H. & Imagawa, H. (1993) High-speed digital image analysis of vocal cord vibration in diplophonia, *Speech Communication*, **13,** 23–32.

Kiritani, S., Niimi, S., Imagawa, H. & Hirose, H. (1995) Vocal fold vibrations associated with involuntary voice changes in certain pathological cases. In *Vocal fold physiology: voice quality control* (O. Fujimura & M. Hirano, editors), pp. 269–282. San Diego: Singular Publishing Group.

Kirk, P. L., Ladefoged, P. & Ladefoged, J. (1984) Using a spectrograph for measures of phonation types in natural languages, *UCLA Working Papers in Phonetics*, **59,** 102–113.

Klatt, D. H. & Klatt, L. C. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers, *Journal of the Acoustical Society of America*, **87,** 820–857.

Kreiman, J. & Gerratt, B. R. (1999) Measuring vocal quality. In *Handbook of voice quality measurement* (R. Kent & M. J. Ball, editors), pp. 73–102. San Diego: Singular.

Kreiman, J. & Gerratt, B. R. (2000) Sources of listener disagreement in voice quality assessment, *Journal of the Acoustical Society of America*, **108,** 1867–1879.

Kreiman, J., Gerratt, B. R., Precoda, K. & Berke, G. S. (1993*a*, May) Perception of supraperiodic voices. Presented at the 125th Meeting of the Acoustical Society of America, Ottawa.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A. & Berke, G. S. (1993*b*) Perceptual evaluation of voice quality: review, tutorial, and a framework for future research, *Journal of Speech and Hearing Research*, **36,** 21–40.

Ladefoged, P. (1988) Discussion of phonetics: a note on some terms for phonation types. In *Vocal physiology: voice production mechanisms and functions* (O. Fujimura, editor), pp. 373–375. New York: Raven Press.

Laver, J. (1980) *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Ludlow, C. L., Coulter, D. & Gentges, F. (1983) The differential sensitivity of frequency perturbation to laryngeal neoplasms and neuropathologies. In *Vocal fold physiology: contemporary research & clinical issues* (D. M. Bless & J. H. Abbs, editors), pp. 381–392). San Diego: College Hill.

Mazo, M., Erickson, D. & Harvey, T. (1995) Emotion and expression: temporal data on voice quality in Russian lament. In *Vocal fold physiology: voice quality control* (O. Fujimura & M. Hirano, editors), pp. 173–187. San Diego: Singular Press.

McGlone, R. E. & Shipp, T. (1971) Some physiologic correlates of vocal fry phonation, *Journal of Speech and Hearing Research*, **14,** 769–775.

Michel, J. & Hollien, H. (1968) Perceptual differentiation of vocal fry and harshness, *Journal of Speech and Hearing Research*, **11,** 439–443.

Moore, P. & Von Leden, H. (1958) Dynamic variations of the vibratory pattern in the normal larynx, *Folia Phoniatrica*, **10,** 205–238.

Morgan, E. C. (1882) Diphthonia or double voice, *Archives of Laryngology*, **3,** 48–57.

Murry, T. (1971) Subglottal pressure and airflow measures during vocal fry phonation, *Journal of Speech and Hearing Research*, **14,** 544–551.

Ramig, L., Scherer, R. C., Titze, I. R. & Ringel, S. (1988) Acoustic analysis of voices of patients with neurologic disease: a rationale and preliminary data, *Annals of Otology, Rhinology and Laryngology*, **97,** pp. 164–172.

Redi, L. & Shattuck-Hufnagel, S. (2001) Variation in the realization of glottalization in normal speakers, *Journal of Phonetics*, **29,** 407–429. doi:10.1006/jpho.2001.0145.

Rose, P. (1988) Phonetics and phonology of yang tone phonation types in Zhenhai. Paper presented at the First International Conference on Wu Dialects, Hong Kong.

Schiffman, S., Reynolds, M. & Young, F. (1981) *Introduction to multidimensional scaling: theory, method, and applications*. New York: Academic.

Schreibweiss-Merin, D. & Terrio, L. M. (1986) Acoustic analysis of diplophonia: a case study, *Perceptual and Motor Skills*, **63,** 755–765.

Sirvio, P. & Michelsson, K. (1976) Sound-spectrographic cry analysis of normal and abnormal newborn infants, *Folia Phoniatrica*, **28,** 161–173.

Smith, M. E., Berke, G. S., Gerratt, B. R. & Kreiman, J. (1992) Laryngeal paralyses: theoretical considerations and effects on laryngeal vibration, *Journal of Speech and Hearing Research*, **35,** 545–554.

Svec, J. G., Horacek, J., Sram, F. & Vesely, J. (2000). Resonance properties of the vocal folds: in vivo laryngoscopic investigation of the externally excited laryngeal vibrations, *Journal of the Acoustical Society of America*, **108,** 1397–1407.

Svec, J., Schutte, H. K. & Miller, D. G. (1996) A subharmonic vibratory pattern in normal vocal folds, *Journal of Speech and Hearing Research*, **39,** 135–143.

Tigges, M., Mergell, P., Herzel, H., Wittenberg, T. & Eysholdt, U. (1997) Observation and modelling of glottal biphonation, *Acustica*, **83,** 707–714.

Timcke, R. (1956) Die Synchron-stroboskopie von menschlichen stimmlipen bzw. ahnlichen schallquellen und messung der offnungszeit, *Zeitschrigt fur Laryngologie, Rhinologie, Otologie und ihre Grenzgebiete*, **35,** 331–335 [cited by Timcke *et al.*, 1959].

Timcke, R., Von Leden, H. & Moore, P. (1959) Laryngeal vibrations: measurement of the glottic wave, Part 2. Physiological variations, *Archives of Otolaryngology Head and Neck Surgery*, **69,** 438–444.
Ward, P. H., Sanders, J. W., Goldman, R. & Moore, G. P. (1969) Diplophonia, *Annals of Otology*, *Rhinology and Laryngology*, **78,** 771–777.
Wendahl, R. W., Moore, G. P. & Hollien, H. (1963) Comments on vocal fry, *Folia Phoniatrica*, **15,** 251–255.
Whitehead, R. L., Metz, D. E. & Whitehead, B. H. (1984) Vibratory patterns of the vocal folds during pulse register phonation, *Journal of the Acoustical Society of America*, **75,** 1293–1297.
Wong, D., Ito, M., Cox, N. B. & Titze, I. R. (1991) Observations of perturbation in lumped element model of the vocal cords with application to some pathological processes, *Journal of the Acoustical Society of America*, **89,** 383–394.