Abstract:
PURPOSE: Many researchers have studied the acoustics, physiology, and perceptual characteristics of the voice source, but despite significant attention, it remains unclear which aspects of the source should be quantified and how measurements should be made. In this study, the authors examined the relationships among a number of existing measures of the glottal source spectrum, along with the association of these measures to overall spectral shapes and to glottal pulse shapes, to determine which measures of the source best capture information about the shapes of glottal pulses and glottal source spectra. METHOD: Seventy-eight different measures of source spectral shapes were made on the voices of 70 speakers. Principal components analysis was applied to measurement data, and the resulting factors were compared with factors similarly derived from oral speech spectra and glottal pulses. RESULTS: Results revealed high levels of duplication and overlap among existing measures of source spectral slope. Further, existing measures were not well aligned with patterns of spectral variability. In particular, existing spectral measures do not appear to model the higher frequency parts of the source spectrum adequately. CONCLUSION: The failure of existing measures to adequately quantify spectral variability may explain why results of studies examining the perceptual importance of spectral slope have not produced consistent results. Because variability in the speech signal is often perceptually salient, these results suggest that most existing measures of source spectral slope are unlikely to be good predictors of voice quality.

Measures of the Glottal Source Spectrum

Jody Kreiman
Bruce R. Gerratt
Norma Antoñanzas-Barroso
University of California, Los Angeles

Purpose: Many researchers have studied the acoustics, physiology, and perceptual characteristics of the voice source, but despite significant attention, it remains unclear which aspects of the source should be quantified and how measurements should be made. In this study, the authors examined the relationships among a number of existing measures of the glottal source spectrum, along with the association of these measures to overall spectral shapes and to glottal pulse shapes, to determine which measures of the source best capture information about the shapes of glottal pulses and glottal source spectra.
Method: Seventy-eight different measures of source spectral shapes were made on the voices of 70 speakers. Principal components analysis was applied to measurement data, and the resulting factors were compared with factors similarly derived from oral speech spectra and glottal pulses.
Results: Results revealed high levels of duplication and overlap among existing measures of source spectral slope. Further, existing measures were not well aligned with patterns of spectral variability. In particular, existing spectral measures do not appear to model the higher frequency parts of the source spectrum adequately.
Conclusion: The failure of existing measures to adequately quantify spectral variability may explain why results of studies examining the perceptual importance of spectral slope have not produced consistent results. Because variability in the speech signal is often perceptually salient, these results suggest that most existing measures of source spectral slope are unlikely to be good predictors of voice quality.

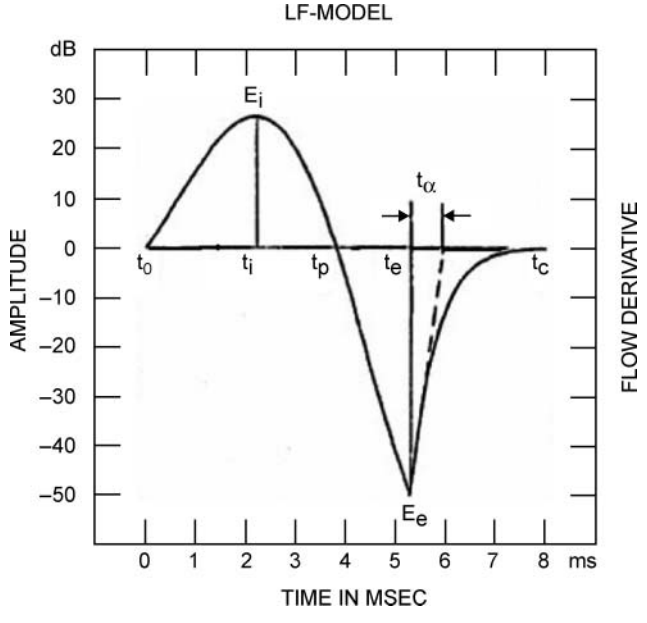
KEY WORDS: voice quality, acoustic measures, source spectrum

The voice source holds a central place in descriptions of speech production. Many investigators with a variety of research goals have studied the acoustics, physiology, and perceptual characteristics of the voice source. However, these studies have not settled the issues of which aspects of the source should be quantified, how measurements should be made, and how different measures relate to one another. This study examines the adequacy with which measures of the source spectrum quantify glottal pulse and source spectral shapes. Because vocal attributes that remain constant are unlikely to be perceptually salient to listeners, perceptually meaningful measures of source spectra should quantify those aspects of spectral shapes that actually vary from voice to voice. In addition, measures of source spectra should correspond to changes in patterns of vocal fold vibration because spectra are the result of vocal fold vibration. By examining relationships among a number of existing measures of the source spectrum and the association of these measures to spectral shapes and to glottal pulse shapes, we hope to provide data for motivating future hypotheses about which of the many possible features of the source spectrum are likely to contribute to listeners' perceptions of vocal quality. However, this study does not test any perceptual hypotheses directly.

Although this study focuses primarily on descriptions of the source in the spectral domain, source characteristics are also frequently described in terms of glottal pulse shapes in the time domain. The timing of glottal events is undeniably important for modeling movements of the vocal folds and patterns of airflow through the glottis, and most models of the voice source, including the popular Liljencrants-Fant (LF) model (see Figure 1; Fant, Liljencrants, & Lin, 1985), are implemented in the time domain (see Fujisaki & Ljungqvist, 1986, for review). However, such time-domain events also can be described in terms of their spectral effects in the frequency domain, and evidence suggests that the shape of the glottal source spectrum is an important determinant of vocal quality. For example, synthesis studies have shown that spectral features, including the difference in amplitude between the first and second harmonics (H1-H2), spectral tilt, and the bandwidth of the first formant, are associated with differences in voice quality (e.g., Bickley, 1982; Doval & d'Alessandro, 1999; Hanson, 1997; Klatt & Klatt, 1990). Some measures of the source spectrum (for example, H1-H2) also have well-established correspondences with linguistic features for voice quality (e.g., Huffman, 1987; Ladefoged, Maddieson, & Jackson, 1988; Wayland & Jongman, 2003).

Spectral measures also have a strong practical appeal as potential alternatives to time-domain measures of the source. First, time-domain measures of pulse shapes made on inverse filtered signals are accurate only if phase information is completely preserved during recording. This requires either the use of a pneumotachographic mask or a microphone with a low-frequency response near zero. However, not all listeners can hear changes in harmonic phase, even through headphones; and for those who can, the perceptual effect is small compared with changes in spectral slope or harmonic amplitudes (Plomp & Steeneken, 1969). Given this relative insensitivity to phase information in complex tones, spectral measures may adequately characterize listeners' perceptions of voice quality while sparing experimenters the burden of applying special phase-preserving recording techniques. Once the source excitation is separated from the effects of the vocal tract on the oral speech signal (usually by inverse filtering, which can be difficult in itself; see, e.g., Javkin, Antoñanzas-Barroso, & Maddieson, 1987, for review), residual formant ripple, bumps related to source/vocal tract resonance interactions, and the like often remain, making it difficult to determine major time-domain features of the voice source without significant ambiguities. For this reason, parameter extraction or model fitting in the time domain is necessarily a subjective process in which conflicts often arise between theoretical expectations and empirically derived pulse shapes. This difficulty generally does not occur in the spectral domain. Finally, features that are relatively easy to quantify in the spectral domain may be more difficult to extract and interpret in the time domain. For example, a single feature in the spectral domain (e.g., a change in H1-H2) may have a number of different possible multivariate causes in the time domain (Fant, 1995).

Figure 1. The LF model of the glottal voice source (Fant et al., 1985).



Parameter estimation in the spectral domain is not plagued with such technical difficulties, but spectral measures of the glottal source have their own limitations. Such measures usually window the speech signal, and thus average over time. Consequently, they do not effectively capture temporal details of quick changes in phonation of the kind that occur with consonant environment or prosody, which often happen over the course of one or two glottal cycles (e.g., Blankenship, 2002; Epstein, 2002, Redi & Shattuck-Hufnagel, 2001). Time-domain measures allow tracking of these kinds of rapid changes. Time-domain measures are also attractive because of their closer relationship to physiological events such as glottal opening, closing, and speed of vocal fold movement.

Because of the equivalence between time and frequency domain representations, it is possible to describe the theoretical relationship between time-domain variations in pulse shapes (often expressed in terms of LF model parameters) and the corresponding changes in the source spectrum. For example, Fant (1995, 1997; see also Fant & Lin, 1988, or Gobl, 1989) interpreted a variety of features of the glottal source pulse primarily by reference to spectral characteristics and the associated voice qualities on a continuum from "breathy" to "pressed." These so-called *R parameters* (see Table 1) are

Table 1. Definitions for the R parameters, following Fant (1995, 1997) and Ní Chasaide and Gobl (1997).

Parameter	Definition
EE	Value of negative peak of the differentiated flow pulse; point of maximum excitation of vocal tract. Associated with overall signal amplitude.
FA	1/(2ΠTa) = F0/(2ΠRA), where Ta is the time constant of the return phase of the pulse. A measure of spectral tilt.
FG	RG × F0. Measures a boost in the H1-H2 range related to the shape of the glottal pulse.
OQ	(1 + RK)/2RG. Alternatively given as Te/T0, where Te is the time of point EE and T0 is the duration of the pulse. OQ controls the amplitude of the lowest harmonics.
RA	Ta/T0, where Ta is the time constant of the return phase of the pulse and T0 is the duration of the pulse. A measure of spectral tilt that defines the frequency above which the spectrum acquires an additional falloff of −6 dB/octave.
RD	In terms of LF model parameters, RD = (UO/EE) × (F0/110), where UO is the peak value for the glottal pulse. Alternatively, in terms of the R parameters, RD = [(0.5 + 1.2 RK)(RK/4RG + RA)]/0.11. A shape parameter that measures the entire spectral shape, proposed to quantify the continuum from "pressed" to "breathy" phonation.
RG	T0/2Tp, where T0 is the duration of the pulse and Tp is the time from 0 to peak flow. Alternatively, FG/F0. Normalizes parameter FG for F0.
RK	(Te-Tp)/Tp, where Te is the time of point EE and Tp is the time from 0 to peak flow. Measures pulse symmetry.

often used in applications in which detailed cycle-by-cycle measurements of spectral changes are of interest (e.g., Gobl, 1988; Gobl & Karlsson, 1991). Application of these measures reflects the assumption that time-domain measures of the source are important mainly to the extent that they determine spectral features (e.g., Gobl & Ní Chasaide, 1992). For example, the LF parameter *RA* (defined as the effective duration of the return phase) is considered an index of spectral tilt, and the parameter *open quotient (OQ)* relates the relative timing of glottal opening and vocal tract excitation to the amplitudes of the lowest frequency harmonics (see Ní Chasaide & Gobl, 1997, for review). Further, Fant (1995) demonstrated a very strong and apparently linear relationship between the LF model parameter RD and H1-H2. Of course, such measures are subject to the limitations of the time-domain measurements on which they depend, but they do allow researchers to combine spectral information with temporal precision.

Existing Measures of Source Spectral Slope

Although experimenters broadly agree that the source spectral slope is an important vocal attribute, a similar degree of agreement has not been reached regarding the manner in which slope should be quantified, and many measures are in current use. These measures fall into several general categories. First, measurements may be made directly on source pulses, as recovered by inverse filtering (see Figure 2A). Some measures within this category are derived from a single glottal pulse. For example, the *parabolic spectral parameter* (*PSP;* Alku, Strik, & Vilkman, 1997) is defined as the steepness of a parabola fit to the spectrum of a single glottal flow pulse. Other measures derive from analyses of a single cycle that has been repeatedly concatenated (see Figure 2C) or from a sequence of adjacent cycles (see Figure 2E). These measures are often calculated from the spectrum of the first derivative of the source pulses (see Figure 2A). For example, a regression line can be fit to the harmonic peaks in the spectrum of the glottal pulses (Jackson, Ladefoged, Huffman, & Antoñanzas-Barroso, 1985; see Figure 3C). The *harmonic richness factor* (*HRF;* Childers & Lee, 1991) is the ratio of the amplitude of the fundamental to the sum of the amplitudes of the harmonics above the fundamental. Traditional measures such as differences in amplitudes of individual harmonics (typically H1-H2, but also H2-H4; see Figure 3a) are often made on spectra calculated from the output of the inverse filter. Finally, authors have measured the deviation of the empirical source slope from an "ideal" slope in different frequency bands (typically four bands, each 1 kHz wide, from 0 to 4 kHz; Ní Chasaide & Gobl, 1997, or Sundberg & Gauffin, 1979; see Figure 3B). The ideal slope assumed by these measures (–12 dB/octave) was originally derived from idealized source pulses that were triangular in shape (Carr & Trill, 1964). Spectra of natural voices (even normal ones) vary in slope, do not fall off evenly at the predicted rate, and bear little resemblance to these ideal spectra, limiting the theoretical appeal of these measures.

These measures of the spectrum reflect only the contributions of the harmonic part of the voice source to vocal tract excitation.[1] Inharmonic (noise) energy also contributes significant excitation (e.g., Hillenbrand & Houde, 1996), particularly in female voices in which persistent glottal gaps may be present (e.g., Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995; Linville & Fisher, 1992), in male or female "sexy" voice (Henton & Bladon, 1985), and in pathologic phonation. For example, Holmberg et al. found that for women with normal voice, most vowel productions displayed a mix of harmonic energy and noise in the F3 region; some showed mostly noise, and only a few tokens were produced with predominantly harmonic energy. Although measuring the spectrum of the combined harmonic and inharmonic excitations is relatively trivial in synthetic speech (where all parameters are known),

[1]Although if these measures are computed over a stretch of speech, they do incorporate some noise as a result of F0 instabilities or pitch changes (Alku et al., 1997).

Figure 2. Glottal flow derivatives and their spectra. Spectra have been normalized to equal peak amplitudes, and the *y*-axis shows the amplitude of each harmonic as a percent of this maximum. A: A single synthetic glottal source pulse (flow derivative). B: Fast Fourier transform (FFT) of the single glottal pulse in Panel A. The peak in this spectrum is sometimes called the *glottal formant* (Doval & d'Alessandro, 1999). C: The glottal pulse in Panel A, concatenated to produce a series of pulses. D: FFT of the sequence of glottal pulses in Panel C. E: A series of consecutive glottal pulses from the natural voice sample, recovered by inverse filtering. F: FFT of the glottal pulse train in Panel E.

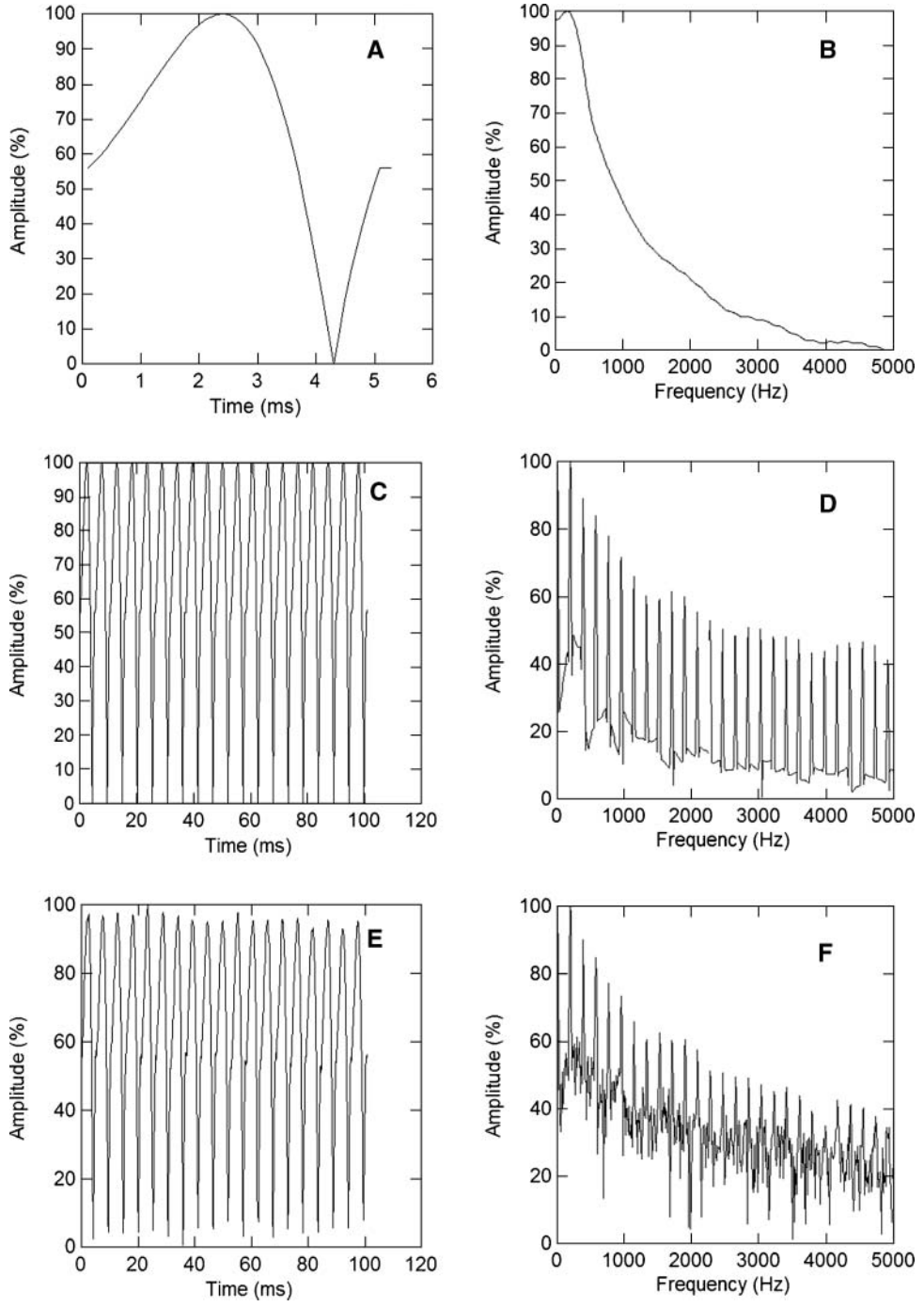
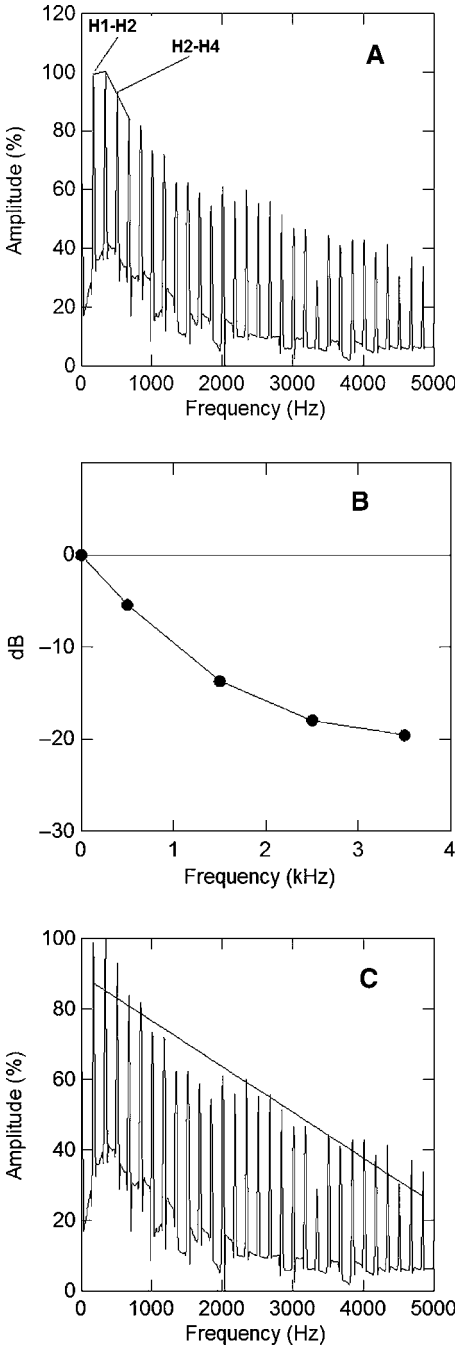


Figure 3. Acoustic analyses performed on one voice. See text for explanation. Spectra have been normalized to equal peak amplitudes, and the *y*-axis shows the amplitude of each harmonic as a percent of this maximum. A: H1-H2 and H2-H4. B: Schematic spectrum showing the average deviation from a constant −12 dB/octave slope in each of four frequency bands. C: A regression line fit to the peaks of all the harmonics in a dB spectrum.



the matter is problematic in natural speech, and such measures have not been described, to our knowledge.[2]

In principle, the previously mentioned measures reflect only the spectrum of the glottal source independent of vocal tract influences on the oral speech spectrum. However, in practice, separating the source from the vocal tract is technically difficult and fraught with ambiguities, as discussed above. A number of studies have sought to circumvent this difficulty by estimating the glottal source spectral slope directly from the complete oral speech signal. Two approaches have been taken. In the first approach, the long-term average spectrum (LTAS) of the voice is calculated over a long sample of connected speech—30 s or more—on the assumption that the influence of varying vocal tract resonances on spectral shape will average out across the sample, yielding a measure that approximates the overall source contribution. For example, Lofqvist and Mandersson (1987) measured the ratio of the spectral energy above and below 1 kHz and the ratio of the energy between 5 kHz and 8 kHz to that below 1 kHz, both from LTAS. More detailed spectral representations from LTAS were proposed by Linville (2002), who measured the spectral energy in 160-Hz-wide bands from 0 to 8 kHz, yielding 50 LTAS measures per utterance per speaker. It is also possible to measure the relative amplitudes of individual landmarks (H1-H2, H1-A1 [the amplitude of the first formant], etc.) from these long-term spectra, although such measures remain sensitive to variations in F0 and the vowel inventory of the sample. Hanson proposed a second approach for removing some of the influences of vocal tract resonances on the spectrum—a sort of virtual inverse filtering that does not require the use of specialized recording equipment (Hanson, 1997; Hanson & Chuang, 1999; see also Fant, 1995; Iseli & Alwan, 2004). This approach is desirable in circumstances where use of special equipment is impractical.

In the past, the choice among measures of the spectrum has usually been motivated by study-specific goals rather than by broader theoretical concerns. Acoustic measures are useful to the extent that they reflect the underlying voice production system or explain listeners' perceptions (Catford, 1977); however, no comprehensive theory presently exists describing correspondences among vocal physiology, acoustics, and perceived voice quality, so no theoretical basis exists for determining the approach that most usefully quantifies source spectral slopes. Further, the relationships among existing measures of the source spectrum remain unknown. Most studies rely on correlations between spectral measures and ratings of specific vocal qualities for validation of a proposed measure (for an exception, see Bickley, 1982), but given

[2]Measures that compare the harmonics-to-noise ratio in different frequency bands are conceptually related to measures of the spectral slope of the whole source but are difficult to interpret because of the influence of vocal tract resonances on the overall speech spectrum.

the confusion surrounding voice quality terminology and without knowledge of the intercorrelations among acoustic measures, interpretation of such correlations is difficult. For example, Hammarberg, Fritzell, Gauffin, Sundberg, and Wedin (1980) reported correlations between various LTAS measures and breathiness, creakiness, and hypo/hyperfunction; Klich (1982) found moderate correlations between rated breathiness and the relative spectral energy above 3500 Hz; and Huffman (1987) reported correlations between H1-H2 and phonemic breathiness in Hmong. In the face of such variability, drawing broad conclusions about the perceptual importance of various aspects of glottal source spectral slopes is difficult. Fant (1995) discusses hypothetical associations between the R parameters, events at the glottis, and a continuum of quality from "pressed" to "breathy" phonation, but validation of these associations has again relied on correlation between measurements and ratings of specific vocal qualities, and correlations reveal little about the psychophysical relationship between an acoustic feature and the voice quality perception that it evokes. These results are also ambiguous because of variations in the parameters and perceptual terminology used. For example, reports indicate that vocal strain (Karlsson, 1992) and creaky voice (Gobl & Ní Chasaide, 1992) are both characterized by decreases in the RK parameter and increases in the amplitude of the first formant relative to the first harmonic; strain also entailed increases in RG, whereas fluctuations in RA accompanied creaky voice. Thus, despite evidence that spectral slope is perceptually important, the precise manner in which listeners use this information remains obscure, making it impossible to assess the validity of the different acoustic measures.

Speech synthesis offers the opportunity to gather causal rather than correlational evidence about the perceptual importance of different aspects of glottal source spectral slopes. However, successful and efficient experimental use of synthesis requires knowledge about which variable among the many acoustic variables is likely to be perceptually salient. Redundancies among currently used slope measures make it difficult to select a parsimonious set of spectral variables for synthesis. Further, when generating hypotheses about perception of spectral slopes, it is helpful to consider how well the acoustic features correspond to the ways in which spectral slopes vary empirically. It is not clear how adequately existing measures actually assess variability in spectral shapes.

To examine the extent to which existing measures of the source spectrum are both necessary and sufficient, we applied principal components analysis to sets of spectral measurements, to glottal pulse shapes, and to glottal source spectral envelopes for a large number of pathological and normal voices, and we compared the extent to which related factors emerged across domains. The goal of this research is examining the extent to which measures of the source spectrum quantify variations in spectral shape, relating changes in spectral shape to variations in glottal pulse shapes, and, ultimately, motivating hypotheses about which aspects of the source spectrum are perceptually salient by reference to acoustic variability, so that all aspects of voice can be understood in a single framework. To the extent that existing measures of source spectral slopes correspond to factors that describe variability in spectral slopes, they adequately quantify spectral slope variability, and, thus, are likely to be perceptually informative to listeners.

Methods

Voices

Voice samples recorded from 60 speakers with vocal pathology (23 men, 37 women) were used in these experiments. Samples were chosen at random, except that voices with period doubling or biphonation were excluded because of the difficulty of synthesizing these phonation types with existing glottal source models. To ensure that the range of naturally occurring source spectra was adequately sampled, the voices of 10 speakers without pathology (5 men, 5 women) were also recorded. Speakers without vocal pathology averaged 41.2 years of age (range = 30–56 years; *SD* = 9.1 years). Speakers with vocal pathology averaged 51.0 years of age (range = 22–89 years; *SD* = 16.4 years) and represented a variety of primary diagnoses, including vocal fold paralysis or paresis (n = 15), vocal fold mass lesion (n =15), chronic laryngitis (n = 12), adductory spasmodic dysphonia (n = 7), functional disorder (n = 4), vocal fold granuloma (n = 4), and Parkinson disease (n = 3). They ranged from mildly to severely dysphonic.

Each speaker was asked to sustain the vowel /a/ at comfortable levels of pitch and loudness. Voice samples were transduced with a Bruel & Kjaer ½" condenser microphone (Model 4193) and were directly digitized at 20 kHz with 16-bit resolution using an A/D system with linear phase sigma–delta technology to ensure signals were not aliased. A 1-s segment was excerpted from the middle of these utterances, high-pass filtered with a linear phase filter at 6 Hz to remove baseline drift caused by room air currents, and downsampled to 10 kHz prior to use in the following experiments.

Voice Synthesis

Each natural voice sample was copied using a custom formant synthesizer implemented in MATLAB (Mathworks, 2002). Analysis and synthesis procedures are described in detail elsewhere (Gerratt & Kreiman, 2001; Kreiman,

Gerratt, & Antoñanzas-Barroso, 2006).[3] Briefly, the synthesizer sampling rate was fixed at 10 kHz. Parameters describing the harmonic part of the glottal source were estimated by inverse filtering a representative cycle of phonation for each voice using the method described by Javkin et al. (1987). The extracted pulses were least-squares fit with a modified LF source model (Fant et al., 1985). Although the original LF model is well-suited to synthesizing normal voice quality, some constraints and modifications were needed to fit the model to experimental data derived from pathological voices. The present version of the LF model differs from the original in several respects in the modeling of the return phase. First, point t_c is not constrained to equal point t_0 for the following cycle (see Figure 1), so the closed phase is formally modeled in this implementation. Second, in many cases in our data, returning flow derivatives to 0 at the end of the cycle conflicted with the need to match the experimental data and conflicted with the requirement for equal areas under positive and negative curves in the flow derivative (the *equal area constraint*). Empirically, this constraint means that the ending flow level should be the same as the beginning flow level (although that value need not equal 0; Ní Chasaide & Gobl, 1997). This is probably not true for highly variable pathological voices. In many cases, the combination of this constraint with fitting the experimental pulses resulted in a flow derivative that stepped sharply to 0. This introduced significant high-frequency artifacts into the synthetic signals. These conflicts between model features, constraints, and the empirical data were handled by abandoning the equal area constraint and by modifying the second (exponential) segment of the model to produce the following equation:

$$U'(t) = Ee \times \exp[\acute{\varepsilon} \times (t - te)] + m(t - te), \qquad (1)$$

where $m = [Ee/(te - tc)] \times \{\exp[-\acute{\varepsilon}(tc - te)]\}$, $\acute{\varepsilon} = [1/(te - t_2)] \times \ln[(m/Ee) \times (te - t_2) + 0.5$, and t_2 equals the time increment to 50% decay in the return phase. This forces the flow to return to 0 and also has the effect of flattening out the return phase somewhat relative to the original LF model; but the modification improves the fit to many pathological voices for which this segment may be nearly a straight line. We used parameters of this revised LF model to specify the harmonic part of the glottal source in the synthesizer.

The spectral characteristics of the inharmonic part of the source (the noise excitation) were estimated using a cepstral-domain comb lifter similar to that described by de Krom (1993). Spectrally shaped noise was synthesized by passing white noise through a 100-tap finite impulse response filter fitted to that spectrum. F0 was tracked pulse by pulse on the time-domain waveform by an automatic algorithm. Location of cycle boundaries was verified by the first author. To overcome quantization limits on modeling F0, we synthesized the source time series pulse by pulse using an interpolation algorithm so that the pitch contour of the synthetic token matched that of the natural voice. A train of LF pulses with the appropriate periods was added to the noise time series to create a complete glottal source waveform. We initially set the ratio of noise to LF energy to match the value calculated from the original voice sample. Formant frequencies and bandwidths were estimated through use of autocorrelation linear predictive coding (LPC) analysis with a window of 25.6 ms (increased to 51.2 ms when stimulus F0 was near or below 100 Hz). We filtered the complete synthesized source through the vocal tract model to generate a preliminary version of the synthetic voice.

[3]A documented version of the synthesizer software is available for download at http://www.surgery.medsch.ucla.edu/glottalaffairs/index.htm. Documentation includes a detailed description of both software and analysis/synthesis procedures.

Perceptually Adjusted Stimuli and Listening Pretest

Measurement of many acoustic parameters is difficult and often inaccurate when phonation departs from periodicity (e.g., Bielamowicz, Kreiman, Gerratt, Dauer, & Berke, 1996; Titze, 1994). For this reason, all synthesizer parameters were adjusted from their measured values by the first author until the synthetic copies formed perfect perceptual matches to the natural target voices. Because investigator biases may add errors of their own during the process of perceptual adjustment, we used a listening pretest to verify the accuracy of these perceptual adjustments. In this pretest, we assessed listeners' ability to distinguish the natural voice tokens from the perceptually adjusted synthetic copies. Twelve listeners (UCLA students and staff; 21–55 years of age; M = 37.9 years; SD = 12.6 years) heard pairs of voices. On half of the trials, a synthetic voice sample (1 s in duration) was paired with its natural counterpart, and on the other half of the trials, stimuli were identical. Voices within a pair were separated by 350 ms. Each pair was repeated twice for a total of 280 trials per listener.

For each trial, listeners were asked to judge whether the two samples were the same or different and to rate their confidence in their response on a 5-point scale ranging from 1 (*positive*) to 5 (*wild guess*). Listeners were not allowed to replay the stimuli before responding. Order of voices in the "different" pairs was randomized, and the stimulus pairs were rerandomized for each listener. Listeners were tested individually in a double-walled sound suite. Stimuli were presented in free field at a comfortable constant listening level. Testing lasted approximately 30 min.

We pooled the responses across listeners to measure the overall discriminability of the synthetic and natural

tokens. Rates of correct and incorrect "same" responses (hits and false alarms) were calculated for each voice. Across voices, hit rates ranged from 79.2%–100%, with an average of 95.3% (*SD* = 5.1%). False alarm rates ranged from 41.7%–100%, with an average of 73.1% (*SD* = 14.5%).

Same/different responses for each voice were combined with confidence ratings to create a 10-point scale ranging from 1 (*positive voices are the same*) to 10 (*positive voices are different*). We constructed receiver operating characteristics (ROCs) consisting of 9 points each from these recoded data following the procedure described by Green and Swets (1966; see also MacMillan & Creelman, 2005). The area under the ROC (A_z) for each voice was calculated using SPSS software (Version 13.0; SPSS, Inc.), along with 99% confidence intervals around these values. For every voice, these confidence intervals included the chance value of 0.5. These data indicate that listeners were unable to distinguish any of the synthetic copies reliably from the corresponding natural samples. Therefore, we concluded that the perceptually adjusted synthetic tokens successfully modeled the quality of the natural target voice samples. However, to determine if these perceptual adjustments produce meaningful changes in measures of the source spectral slope and to ensure that investigator biases did not play a role in results, both original estimates and perceptually corrected values for glottal source parameters were retained in subsequent analyses.

Measurements of the Glottal Source Spectrum

A total of 78 measures of the slope of the source spectrum were made for each voice sample. Three different target signals were analyzed: the harmonic part of the glottal source (estimated via inverse filtering), the complete (harmonic plus inharmonic) glottal source (also estimated via inverse filtering), and the complete oral speech signal, including the acoustic effects of both the source and vocal tract. Measurements made on each of these target signals are described in turn in the following sections.

The harmonic part of the glottal source. For each voice, three different estimates of the harmonic part of the glottal source were studied: the unsmoothed output of the inverse filter, the synthetic LF-modeled glottal source pulses, and the perceptually corrected synthetic LF pulses. Each has advantages and disadvantages. The unsmoothed output of the inverse filter is closest to the original signal in that it involves the least experimenter intervention, but it generally includes some residual formant ripple and other distortions, and it is impossible to know whether the "correct" result has been achieved. Fitting the pulses with the LF (or any other) model eliminates noise and ripple but forces a theoretical model on the data, which adds its own artifacts. Finally, the perceptually corrected LF-fitted pulses are perceptually accurate (and thus "correct" in one sense), but the adjustment process is subject to experimenter bias, as discussed previously.

For each of these estimates of the harmonic part of the glottal source, the following measurements were made (see Table 2). The PSP was calculated as described by Alku et al. (1997) from the spectrum of a single extracted glottal cycle. The ratio of spectral energy above and below 1 kHz (e.g., Lofqvist & Mandersson, 1987) was calculated from the spectrum of a single pulse padded with zeros to a length of 1,024 points. We computed a variant of this measure by concatenating each glottal source pulse 20 times prior to computing the spectrum. The differences in amplitude between the first and second harmonics (H1-H2) and the second and fourth harmonics (H2-H4) were also calculated from power spectra (see Figure 3A). We calculated the harmonic richness factor (Childers & Lee, 1991) using the method described by Alku et al. (1997). The deviations of the empirical glottal source slope from an "ideal" spectral slope were measured in 1-kHz-wide bands from 0 to 4 kHz from

Table 2. Measures of the source spectrum for the three versions of the glottal source.

Acoustic measure	Glottal source version		
	Harmonic part of the glottal source	Complete (harmonic plus inharmonic) glottal source	Complete oral voice sample
Slope of line fit to N harmonics (dB)	X	X	
Deviations from ideal slope (dB)	X	X	
H1-H2 (dB)	X	X	X
H2-H4 (dB)	X	X	X
H1-A1 (dB)			X
H1-A3 (dB)			X
Harmonic richness factor	X		X
Parabolic spectral parameter	X		
Long-term average spectrum		X	X
Ratio of energy above/below 1 kHz	X	X	

1,024-point fast Fourier transform (FFT) analyses, as described by Sundberg and Gauffin (1979; see also Ní Chasaide & Gobl, 1997; see Figure 3B). Finally, we generated a set of spectral slope measures by computing the slope of a straight line fitted to the peaks of the first three, five, seven, and nine harmonics of a power spectrum and to the complete spectrum for each version of each glottal source (Jackson et al., 1985; see Figure 3C).

Combined harmonic and inharmonic excitation. We recovered the complete glottal source (harmonic plus inharmonic components) by performing pitch-synchronous inverse filtering on the perceptually corrected synthetic voice signals using the formant frequencies and bandwidths applied during synthesis. This completely canceled the vocal tract, leaving the whole glottal source time series. FFT power spectra were calculated for each glottal source time series, from which were derived measures of H1-H2 and H2-H4; the slopes of lines fit to the first three, five, seven, and nine harmonics of each spectrum and to the entire spectrum; the relative amount of spectral energy above and below 1 kHz; and the average spectral deviation from an ideal slope in 1-kHz-wide bands, as described above. Finally, we calculated long-term average spectra for each glottal source time series by averaging together pitch-synchronous FFT spectra across the entire 1-s sample and measuring the relative amounts of energy above and below 1 kHz in the resultant average spectrum.

Measures based on the oral speech signal (glottal source plus vocal tract). Additional measurements were made from the complete oral speech signal. Two versions of the entire 1-s sample were analyzed: the original sample and the perceptually corrected synthetic copy. For each of these, we measured H1-H2 and H2-H4 from FFT spectra and the relative amounts of energy below and above 1 kHz from an LTAS, calculated as described above but from averaged pitch-synchronous FFT spectra from the entire signal. Measures of H1-H2, H1-A1, and H1-A3 were also calculated as described by Hanson (1997), with the modifications described by Iseli and Alwan (2004).

R parameters. The parameters EE, FA, FG, OQ, RA, RD, RG, and RK, which describe the theoretical spectral effects of various aspects of glottal pulse shapes, were calculated from the LF-fitted source pulses, as described by Fant (1995; see also Ní Chasaide & Gobl, 1997). Table 1 gives definitions and formulae.

Correlational Analysis

The spectral analyses described above produced a total of 78 measures of the glottal source spectrum. However, many of these measures are closely related, forming fully matched "families." Univariate correlations among variables in families were examined as a preliminary means of reducing the number of variables.

The first correlation analysis examined matched measurements made on different versions of the glottal source pulses (the unsmoothed output of the inverse filter, the LF-fitted pulses, and the perceptually corrected pulses). Measures that assess the lower frequency part of the spectrum (H1-H2, the slope of a line fit to the first three and five harmonics, deviations from an ideal slope in bands from 0 to1 kHz and from 1 to 2 kHz, the HRF, and the PSP) were significantly correlated across all three glottal source versions ($r > .5$, $p < .01$ after Bonferroni correction). Measures of the shape of the higher part of the spectrum (H2-H4, deviations from ideal slope in bands from 2 to 3 kHz and from 3 to 4 kHz, the ratio of spectral energy above and below 1 kHz, and the slope of a line fit to the first seven or nine harmonics or to all the harmonics) were correlated for the LF-fitted and perceptually corrected pulses ($r > .5$, $p < .01$ after Bonferroni correction), but the same measures made on the unsmoothed output of the inverse filter were uncorrelated with the other glottal source versions ($r < .5, p > .01$ after Bonferroni correction). This suggests that the unsmoothed inverse filtered pulses include reliable information about the lowest harmonics but not about the details of the rest of the spectrum or its overall shape, presumably because of residual higher frequency distortions (ripples and bumps) in the unsmoothed pulses. As a result of these analyses, measures of the perceptually corrected glottal source pulses were retained in subsequent analyses, and measures of the unsmoothed and LF-fitted pulses were eliminated.

A second correlational analysis examined the relationship between measures made on the 1-s natural voice samples and on the synthetic copies. All measures of natural speech were significantly correlated with the corresponding measure of the synthetic signals ($r > .5$, $p < .01$ after Bonferroni correction), consistent with the previous observation that the synthetic copies were perceptually indistinguishable from the original voice samples. Measures of the synthetic voice samples were retained for subsequent analyses because all vocal tract and glottal source parameters are known for these samples.

Principal Components Analyses

Subsequent to the correlational analysis, principal components analysis (PCA) with varimax rotation was applied to identify redundancies among the 34 remaining measures, which are listed in Table 3. A second PCA was similarly used to derive factors that describe variability in the shapes of the glottal pulses. Glottal source pulses extracted from the perceptually adjusted synthetic tokens were analyzed because they are the most perceptually accurate of the available estimates of the glottal source. To normalize for differences in F0, a MATLAB interpolation algorithm was applied to select 70 equally spaced points along each pulse. Plots of the resampled

Table 3. Spectral measures used in the principal components analysis.

Spectral measure	Target signal		
	Perceptually corrected glottal pulses	Complete glottal source	Complete synthetic voice sample
Slope of a line fit to N harmonics (five measures)	X	X	
H1-H2	X	X	X
H2-H4	X	X	X
LTAS		X	X
Energy above/below 1 kHz	X	X	
Deviation from ideal slope in 1-kHz-wide bands (four measures)	X	X	
HRF	X		
PSP	X		
H1-H2, H1-A1, H1-A3, after Hanson (1997)			X
F0			X

Note. LTAS = long-term average spectrum; HRF = harmonic richness factor; PSP = parabolic spectrum parameter.

versus original pulses were examined to ensure that pulse shapes were preserved by this resampling procedure. The amplitude values for each perceptually corrected glottal pulse at each sampling instant served as input to the PCA.

A third PCA was used to derive factors describing patterns of variability in the shapes of glottal source spectra. FFT spectra (on a logarithmic scale) were calculated for each perceptually corrected glottal source pulse. Spectra were normalized to the amplitude of the first harmonic, and spectral envelopes were estimated by connecting the harmonic peaks. Spectra were resampled by selecting 70 equally spaced points along the envelope from 0 to 5 kHz. Input to the analysis consisted of the amplitude values for each of the 70 points in the resampled spectra.

Results

PCAs

Results of the PCAs are summarized in Table 4. Analyses of spectral measures produced a four-factor solution accounting for 76.6% of the variance in the underlying measurements. The first factor accounted for 33.8% of the variance in the data. It was associated with measures of H1-H2 and with measures of the deviations from an ideal slope in the first two frequency bands (0–1 kHz and 1–2 kHz). This factor apparently represents the shape of the low-frequency harmonic excitation. The second factor accounted for 23.7% of the variance in the underlying data. It was associated with the slope of line fit to N harmonics, suggesting an interpretation in terms of the overall spectral slope. The third factor accounted for 10.8% of the variance. This factor was associated with measures of the slope of the entire glottal source (harmonic and inharmonic components) and appears to represent high-frequency noise excitation. Finally, the fourth factor was associated with measures of H2-H4. It accounted for an additional 8.3% variance in the data.

Analyses of glottal pulse shapes produced a five-factor solution that accounted for 88.9% of the variance in pulse shapes. The factors are shown in Figures 4 and 5. The first factor accounts for 31.6% of the variance and reflects the steepness of the opening phase and pulse asymmetry (see Figure 5). This factor is also related to

Table 4. Results of principal components analyses.

Pulse shape factors (88.9%)	Spectral shape factors (88.5%)	Spectral measure factors (76.6%)
Steepness of opening; asymmetry (31.6%)	Slope of spectrum above 4 kHz (28.4%)	H1-H2, low-frequency excitation (33.8%)
Duration of closed phase; open quotient (19.0%)	Slope of spectrum below 450 Hz (29.2%)	Overall spectral slope (23.7%)
Shape of return to 0 (14.4%)	Slope of spectrum from 1.5 to 2 kHz (17%)	High-frequency noise excitation (10.8%)
Shape and duration of closing A (13.6%)	Slope of spectrum from 2.3 to 2.7 kHz and from 3.5 to 4.0 kHz (13.9%)	H2-H4 (8.3%)
Shape and duration of closing B (10.3%)		

Note. Variances accounted for by each analysis and factor are given parenthetically.

Figure 4. Results of the principal components analysis (PCA) of glottal pulse shapes, shown in plots of the glottal flow derivatives. Plots have been normalized to the positive peak of the flow derivative. Each panel shows the mean, minimum, and maximum case for each factor. Points that weigh most heavily on each factor are plotted with stars. Pulses have been time warped to equal duration; amplitude units are arbitrary. A (Factor 1): Steepness of opening phase; pulse asymmetry. B (Factor 2): Duration of closed phase; open quotient. C (Factor 3): Shape of return to 0. D (Factor 4) and E (Factor 5): Shape and duration of closing.

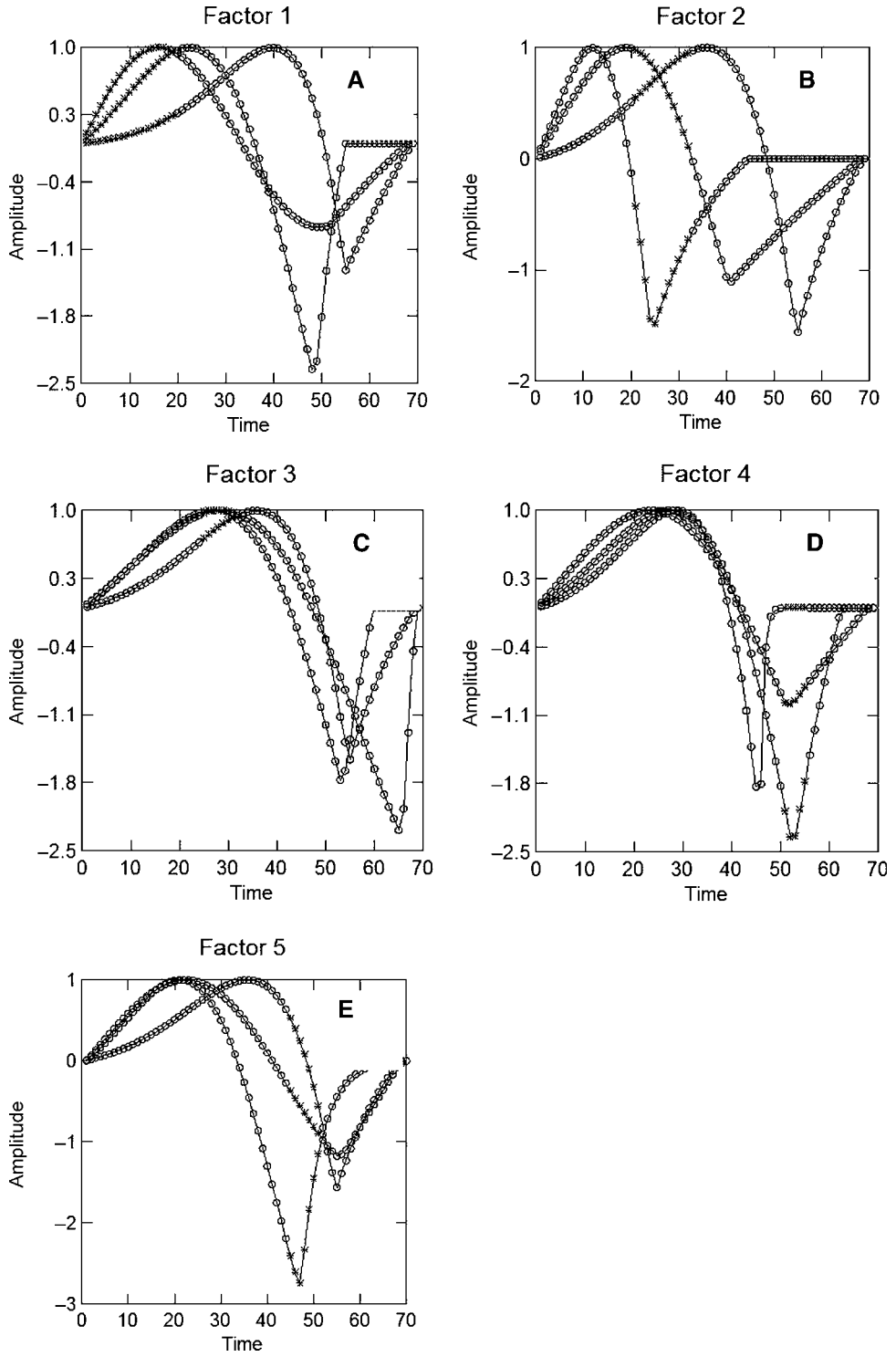
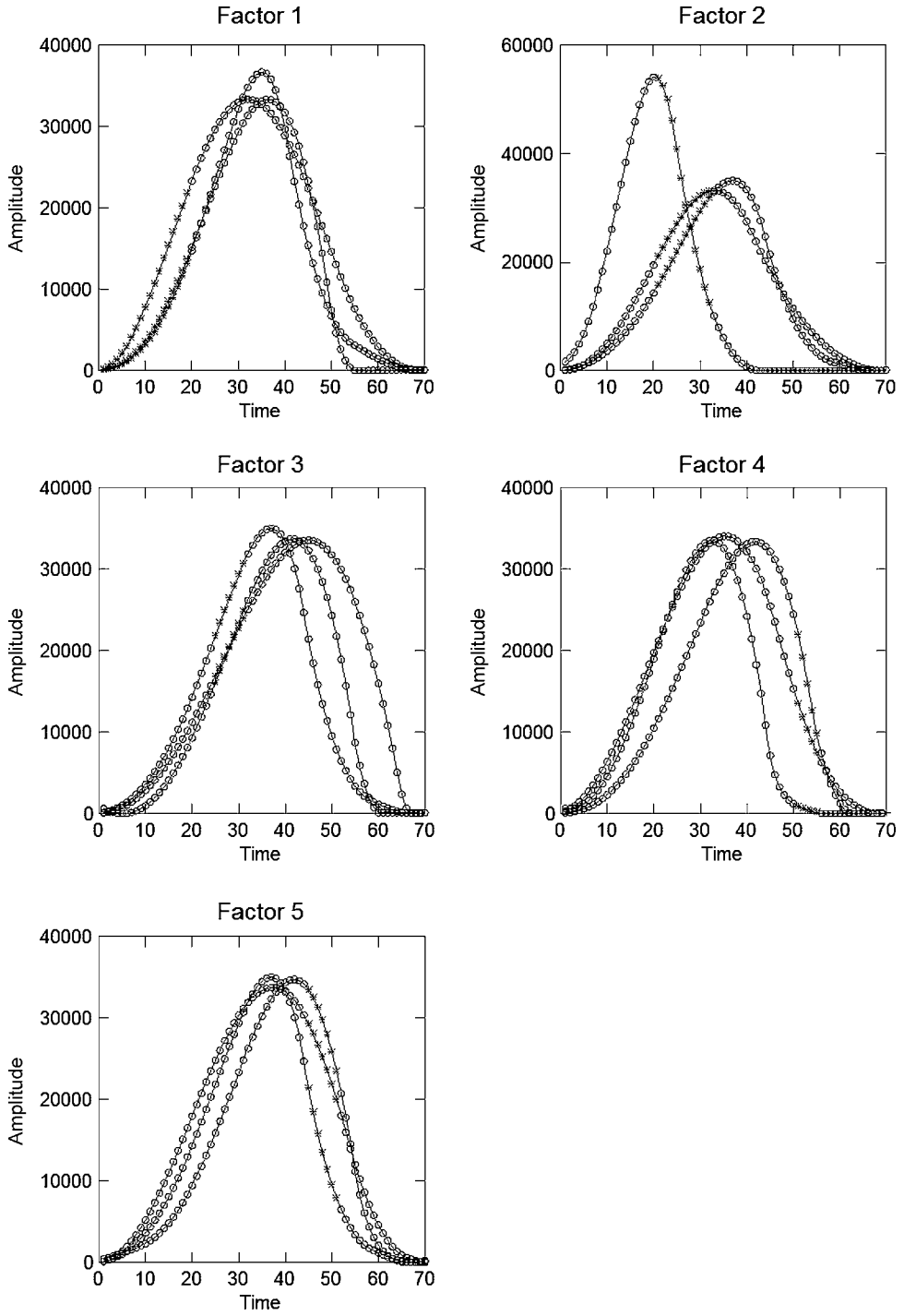


Figure 5. Results of principal components analysis (PCA) of glottal pulse shapes, shown in plots of the glottal pulses. Pulse shapes were obtained by integrating the flow derivative pulses in Figure 4. Each panel shows the mean, minimum, and maximum case for each factor. Points that weigh most heavily on each factor are plotted with stars. Pulses have been time warped to equal duration; amplitude units are arbitrary. A (Factor 1): Steepness of opening phase; pulse asymmetry. B (Factor 2): Duration of closed phase; open quotient. C (Factor 3): Shape of return to 0. D (Factor 4) and E (Factor 5): Shape and duration of closing.



the R parameters RG and RK (rs = .81, p < .01) and can be seen in Figure 4A as the relative timing of the positive peak in the flow derivatives. The second factor (19.0% variance accounted for) is associated with the R parameters RG (duration of closed phase) and OQ (open quotient); R = .65, p < .01). The third factor (14.4% variance) reflects the manner in which the flow returns to 0. It is also associated with R parameters RG and OQ (R = .64, p < .01). Finally, the fourth and fifth factors are both associated with the shape and duration of the closing phase. Factor 4 is associated with R parameter FA (r = .50, p < .01), and Factor 5 is associated with LF parameter EE (r = .32, p < .01). Together, they account for 23.9% of the variance.

Analyses of glottal source spectral shapes produced a four-factor solution that accounted for 88.5% of the variance in the underlying data. These factors were interpreted (unambiguously) by reference to the weights for each factor on each of the 70 points that defined the spectral shape rather than by reference to visual inspection of contrastive spectra. The first spectral shape factor was related to the slope of the harmonic spectrum above 4 kHz and accounted for 28.4% of the variance. The second factor (29.2% variance) was related to the slope of the spectrum below 450 Hz. The third factor (17% variance) corresponded to the slope of the spectrum in the narrow region from 1.5 to 2 kHz, and the fourth factor corresponded to the slope of the spectrum in the regions from 2.3 to 2.7 kHz and from 3.5 to 4.0 kHz (13.9% variance).

Correspondences Across Domains

Because spectral attributes that remain constant across voices are unlikely to be perceptually salient to listeners, meaningful measures of the shapes of glottal source spectra should quantify aspects of spectral shape that actually vary from voice to voice. In addition, measures of glottal source spectra should correspond to changes in patterns of vocal fold vibration because spectra result from vocal fold vibration. To address the question of how well existing spectral measures meet these conditions, we examined the canonical correlations between the spectral measure factors derived via PCA, factors describing glottal pulse shapes, and factors describing the glottal source spectral shapes.

Well-chosen acoustic measures should link spectral shapes and perception (Klatt & Klatt, 1990) in the same way that R parameters model the theoretical relationships between physical pulse shapes and the spectral domain. The canonical correlation (shrunken R^2) between the five glottal pulse shape factors (see Table 4) and the R parameters (EE, RA, RK, RG, OQ, FG, FA, and RD) equaled .998 (p < .01). The correlations between the R parameters and the four spectral shape factors were also significant (shrunken R^2 = .57, p < .01) but smaller, presumably due to the vagaries of natural phonation, the use of pathological voices, and modifications to the LF model used in this study. These results indicate that when applied to natural voices, the R parameters, as a set, function in practice as they are intended to in theory.

The relationship between the spectral measure factors derived via PCA and the spectral shape factors was also significant but weaker than that between pulse shape factors and R parameters (shrunken R^2 = .45, p < .01). Examination of univariate correlations between individual factors and spectral measures confirmed that measures of the spectrum only partially capture differences among voices in glottal source spectral shape. Spectral shape Factor 1 (high-frequency slope) was moderately correlated with the slope of a line fit to all of the harmonics in a dB spectrum (r = .59, p < .01) but was not significantly correlated with any spectral measure factors derived via PCA. (All p values are adjusted for multiple comparisons.) Spectral shape Factor 2 (low-frequency slope) was correlated with spectral measure Factor 1 (H1-H2; r = –.52, p < .01) and with the slope of a line fit to the first three harmonics (r = .60, p < .01). Spectral shape Factors 3 and 4 (slope in the mid-frequency range) were not significantly correlated with any spectral measures or factors, and spectral measure Factors 2 (overall spectral slope), 3 (noise excitation), and 4 (H2-H4) were not significantly correlated with any spectral shape factors.

Discussion

These results indicate that high levels of duplication and overlap occur among existing measures of the glottal source spectrum. Correlation and PCAs reduced an original set of 78 measures to 4 independent factors: H1-H2, overall spectral slope, high-frequency noise excitation, and H2-H4. Further, with the exception of H1-H2 and the slope of a line fit to all harmonic peaks, existing measures are not particularly well aligned with patterns of spectral variability, casting doubt on their potential perceptual salience. Unlike other measures, H1-H2 is related to variability both in spectral shapes and in glottal pulse shapes. Because similar values resulted in whatever version of the glottal source was measured (and whatever measurement variant was applied), this parameter also appears robust in terms of measurement techniques (as does the lower part of the spectrum in general). The frequency with which H1-H2 emerges from studies of voice quality perception reflects both its robustness across the speech chain and its resistance to measurement artifacts.

In contrast, it appears that existing spectral measures do not model the higher parts of the glottal source spectrum (about 1.5–4 kHz) in detail. Understanding the perceptual importance of this spectral range is complicated

by the fact that substantial high-frequency excitation may be contributed by turbulent noise, which is not usually examined formally in studies of glottal source spectral shapes (but see Holmberg et al., 1995). Further, voices contain different amounts of inharmonic energy, so noise excitation is variably important across voices. Finally, harmonic energy and noise may mask one another (e.g., Gerratt & Kreiman, 2001), so the perceptual importance of differences in mid- to high-frequency spectral slope cannot be assessed without knowledge of patterns of both harmonic and inharmonic excitation. Nevertheless, these differences in spectral shape account for more than 30% of variability in overall spectral shapes, suggesting that further investigation of these issues is warranted.

Although the present study did not test any perceptual hypotheses, its results have implications for the study of voice quality perception. The failure of existing spectral measures to adequately quantify spectral variability may explain why the results of studies examining the perceptual importance of spectral slope have produced varying results. Because variability in speech signals is often perceptually salient, these results suggest that existing measures of the glottal source spectrum may not be good predictors of voice quality. H1-H2, the spectral slope in the mid-frequencies (roughly 1–3 kHz), and high-frequency excitation (harmonic and inharmonic) each account for more than 25% of variance in spectral shapes, suggesting that these features may be important determinants of voice quality. By using synthesis to manipulate these candidate features in a constant context, the precise relationship of spectral variability to perceived voice quality can be demonstrated. If the choice of candidate spectral measures is guided by their relationship to spectral variability and the validity of the measures is established with reference to their perceptual importance, then the spectral parameters may serve to map between the spectral and perceptual domains, as the R parameters map between physiology and acoustics.

These results also have practical implications. First, measurements of the lower part of the spectrum (for example, H1-H2) made on different versions of the glottal source pulse (unsmoothed, LF fitted, or perceptually corrected) are highly correlated, so it does not appear to matter which technique is used for estimating the spectral slope below about 450 Hz. Measures of the higher part of the spectrum do depend on measurement technique because residual ripple and error in the inverse-filtered pulses affect mainly the higher part of the spectrum. Smoothing the pulses by fitting them with an appropriate model can remove this high-frequency error, and doing so does not appear to introduce new distortions to any meaningful extent. Finally, measurements of H1-H2 and LTAS measures derived from continuous speech are highly correlated with those from extracted glottal pulses. However, our "continuous speech" samples are prolonged vowels, not text, and this result may not generalize to cases in which more formant variability is present.

In conclusion, hypotheses about the importance of a parameter in one speech domain (e.g., perception) can be generated by considering variability in other domains (e.g., acoustics and physiology). Currently, researchers and clinicians generate hypotheses, in large part, by using their intuition about how behavioral or medical interventions may affect voice quality. By attempting to understand each aspect of voice in the context of other aspects, we hope to better understand, and some day predict, how changes in laryngeal physiology result in acoustic patterns that are perceptually salient.

Acknowledgments

This research was supported by Grant DC01797 from the National Institute on Deafness and Other Communication Disorders. Portions of these results were presented at the 147th and 148th meetings of the Acoustical Society of America in New York City and San Diego, CA, respectively. We thank Paavo Alku, Helen Hanson, and Michael Döllinger for advice regarding implementation of the spectral measures and Sumiko Takayanagi for statistical advice. We also thank Gunnar Fant and two anonymous reviewers for many helpful comments.

References

Alku, P., Strik, H., & Vilkman, E. (1997). Parabolic spectral parameter: A new method for quantification of the glottal flow. *Speech Communication, 22,* 67–79.

Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Massachusetts Institute of Technology Research Laboratory of Electronics Speech Communications Group Working Papers, 1,* 71–82.

Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., & Berke, G. S. (1996). A comparison of voice analysis systems for perturbation measurement. *Journal of Speech and Hearing Research, 39,* 126–134.

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics, 30,* 163–191.

Carr, P. B., & Trill, D. (1964). Long-term larynx-excitation spectra. *The Journal of the Acoustical Society of America, 36,* 2033–2040.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington, IN: Indiana University Press.

Childers, D. G., & Lee, C. K. (1991). Vocal quality factors: Analysis, synthesis, and perception. *The Journal of the Acoustical Society of America, 90,* 2394–2410.

de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research, 36,* 254–266.

Doval, B., & d'Alessandro, C. (1999). The spectrum of glottal flow models. *Notes et Documents LIMSI–CNRS* [Notes and Documents of the Laboratory for Mechanics and Engineering Sciences—National Center for Scientific Research], *99–07,* 1–22.

Epstein, M. (2002). *Voice quality and prosody in English*. Unpublished doctoral dissertation, University of California, Los Angeles.

Fant, G. (1995). The LF model revisited. Transformations and frequency domain analysis. *Speech Transmission Laboratory–Quarterly Progress and Status Report (Stockholm), 2–3,* 119–156.

Fant, G. (1997). The voice source in connected speech. *Speech Communication, 13,* 125–139.

Fant, G., Liljencrants, J., & Lin, Q. (1985). A four-parameter model of glottal flow. *Speech Transmission Laboratory–Quarterly Progress and Status Report (Stockholm), 4,* 1–13.

Fant, G., & Lin, Q. (1988). Frequency domain interpretation and derivation of glottal flow parameters. *Speech Transmission Laboratory–Quarterly Progress and Status Report (Stockholm), 2–3,* 1–21.

Fujisaki, H., & Ljungqvist, M. (1986). Proposal and evaluation of models for the glottal source waveform. *Proceedings of the Institute of Electrical and Electronics Engineers International Conference on Acoustics, Speech, and Signal Processing, 11,* 1605–1608.

Gerratt, B. R., & Kreiman, J. (2001). Measuring vocal quality with speech synthesis. *The Journal of the Acoustical Society of America, 110,* 2560–2566.

Gobl, C. (1988). Voice source dynamics in connected speech. *Speech Transmission Laboratory–Quarterly Progress and Status Report (Stockholm), 1,* 123–159.

Gobl, C. (1989). A preliminary study of acoustic voice quality correlates. *Speech Transmission Laboratory–Quarterly Progress and Status Report (Stockholm), 4,* 9–22.

Gobl, C., & Karlsson, I. (1991). Male and female voice source dynamics. In J. Gauffin & B. Hammarberg (Eds.), *Vocal fold physiology: Acoustic, perceptual, and physiological aspects of voice mechanisms* (pp. 121–128). San Diego, CA: Singular.

Gobl, C., & Ní Chasaide, A. (1992). Acoustic characteristics of voice quality. *Speech Communication, 11,* 481–490.

Green, D. M., & Swets, J. A. (1966). *The theory of signal detection*. New York: Wiley.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngologica (Stockholm), 90,* 441–451.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America, 101,* 466–481.

Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America, 106,* 1064–1077.

Henton, C. G., & Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language & Communication, 5,* 221–227.

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research, 39,* 311–321.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. I. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic measures of female voice. *Journal of Speech and Hearing Research, 38,* 1212–1223.

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

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. *Proceedings of the Institute of Electrical and Electronics Engineers International Conference on Acoustics, Speech, and Signal Processing, 1,* 669–672.

Jackson, M., Ladefoged, P., Huffman, M. K., & Antoñanzas-Barroso, N. (1985). Measures of spectral tilt. *UCLA Working Papers in Phonetics, 62,* 77–88.

Javkin, H., Antoñanzas-Barroso, N., & Maddieson, I. (1987). Digital inverse filtering for linguistic research. *Journal of Speech and Hearing Research, 30,* 122–129.

Karlsson, I. (1992). Modeling voice variations in female speech synthesis. *Speech Communication, 11,* 491–495.

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

Klich, R. (1982). Relationships of vowel characteristics to listener ratings of breathiness. *Journal of Speech and Hearing Research, 25,* 574–580.

Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. (2006). *Analysis and synthesis of pathological voice quality*. Unpublished users' manual, retrieved February 22, 2006, from http://www.surgery.medsch.ucla.edu/glottalaffairs/index.htm

Ladefoged, P., Maddieson, I., & Jackson, M. (1988). Investigating phonation types in different languages. In O. Fujimura (Ed.), *Vocal fold physiology: Voice production, mechanisms and functions* (pp. 297–317). New York: Raven Press.

Linville, S. E. (2002). Source characteristics of aged voice assessed from long-term average spectra. *Journal of Voice, 16,* 472–479.

Linville, S. E., & Fisher, H. B. (1992). Glottal gap configurations in two age groups of women. *Journal of Speech and Hearing Research, 35,* 1209–1215.

Lofqvist, A., & Mandersson, B. (1987). Long-time average spectrum of speech and voice analysis. *Folia Phoniatrica, 39,* 221–229.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Mathworks. (2002). MATLAB (Version 6.5 R13) [Computer software]. Natick, MA: Author.

Ní Chasaide, A., & Gobl, C. (1997). Voice source variation. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 427–461). Oxford, England: Blackwell.

Plomp, R., & Steeneken, H. J. M. (1969). Effect of phase on the timbre of complex tones. *The Journal of the Acoustical Society of America, 46,* 409–421.

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics, 29,* 407–430.

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

Titze, I. R. (1994). *Workshop on acoustic analysis summary statement*. Denver, CO: National Center for Voice and Speech.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics, 31,* 181–201.

Received January 12, 2005

Revision received February 22, 2006

Accepted September 11, 2006

DOI: 10.1044/1092-4388(2007/042)

Contact author: Jody Kreiman, Division of Head and Neck Surgery, University of California, Los Angeles, School of Medicine, 31-24 Rehab Center, 1000 Veteran Avenue, Los Angeles, CA 90095-1794. E-mail: jkreiman@ucla.edu.