JSLHR

Article

Relation of Structural and Vibratory Kinematics of the Vocal Folds to Two Acoustic Measures of Breathy Voice Based on Computational Modeling

Robin A. Samlan[a] and Brad H. Story[a]

Purpose: To relate vocal fold structure and kinematics to 2 acoustic measures: cepstral peak prominence (CPP) and the amplitude of the first harmonic relative to the second (H1–H2).
Method: The authors used a computational, kinematic model of the medial surfaces of the vocal folds to specify features of vocal fold structure and vibration in a manner consistent with breathy voice. Four model parameters were altered: degree of vocal fold adduction, surface bulging, vibratory nodal point, and supraglottal constriction. CPP and H1–H2 were measured from simulated glottal area, glottal flow, and acoustic waveforms and were related to the underlying vocal fold kinematics.
Results: CPP decreased with increased separation of the vocal processes, whereas the nodal point location had little effect. H1–H2 increased as a function of separation of the vocal processes in the range of 1.0 mm to 1.5 mm and decreased with separation > 1.5 mm.
Conclusions: CPP is generally a function of vocal process separation. H1*–H2* (see paragraph 6 of article text for an explanation of the asterisks) will increase or decrease with vocal process separation on the basis of vocal fold shape, pivot point for the rotational mode, and supraglottal vocal tract shape, limiting its utility as an indicator of breathy voice. Future work will relate the perception of breathiness to vocal fold kinematics and acoustic measures.

Key Words: voice simulation, vocal folds, breathy voice, acoustics

The purpose of this study was to establish the relation of two acoustic measures used for characterizing breathy voice to vocal fold kinematics. The possibility of using the acoustic signal to both infer physiology and predict perception has long been a goal of acoustic voice analysis. Kreiman, Gerratt, and Antoñanzas-Barroso (2007) described this effort as follows: "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" (p. 608).

A breathy voice is typically produced by maintaining glottal space throughout the vocal fold vibratory cycle; this can occur as a consequence of several underlying factors (Colton, Casper, & Leonard, 2006). First and foremost is incomplete prephonatory vocal fold adduction; a typical prephonatory configuration for modal phonation involves visibly complete adduction of the vocal processes, although a small direct current (DC) flow is typically measured at maximum closure (Hertegård & Gauffin, 1995; Holmberg, Hillman, & Perkell, 1988; Södersten & Lindestad, 1990). Disorders with persistent posterior glottal incompetence include vocal fold paralysis, abductor spasmodic dysphonia, and an abducted pattern of muscle tension dysphonia. Breathy voice secondary to incomplete glottal closure can occur in the context of complete vocal process adduction. A structural (tissue) deficit, lack of mechanical tension (as in superior laryngeal nerve injury), or loss of muscle tone (as in recurrent laryngeal nerve injury) can lead to a bowed or concave vocal fold shape and cause a spindle-shaped space between the folds at maximum glottal closure. Vocal fold and vocal process lesions can cause a variety of gap shapes at maximum closure. Disruptions of left–right vibratory symmetry (i.e., the vocal folds are vibrating out of phase with one another) can also cause breathiness; when the vocal folds reach midline at slightly different

[a]Speech Acoustics Laboratory, University of Arizona, Tucson
Correspondence to Robin A. Samlan: rsamlan@email.arizona.edu

Editor: Anne Smith
Associate Editor: Jack Jiang
Received July 13, 2010
Accepted January 28, 2011
DOI: 10.1044/1092-4388(2011/10-0195)

time points, the amount of time in a vibratory cycle where the vocal folds are maximally together (*closed quotient*) is decreased. The closed quotient continues to decrease as the phase shift increases.

Acoustically, glottal airspace that is maintained throughout the vibration cycle results in reduced energy in the harmonic components of the fundamental frequency (F0) of vibration and increased aspiration noise due to large DC airflows. Two commonly used measurements are thought to reflect these changes in the acoustic signal and relate to perceived breathy voice quality: cepstral peak prominence and the amplitude of the first harmonic relative to the second (H1–H2). The specific aim of this study is to relate these measures to structural and vibratory parameters of vocal fold vibration.

Incomplete glottal closure creates turbulent airflow, resulting in noise above approximately 2–3 kHz (Hanson, 1997; Klatt & Klatt, 1990). Quantifying the relative amplitudes of harmonic and noise energy in this region has taken many forms over the years. One technique that appears to robustly capture the relative amplitude of harmonic energy in dysphonic voices is *cepstral peak prominence* (CPP; Hillenbrand, Cleveland, & Erickson, 1994; Hillenbrand & Houde, 1996). CPP is a technique used to assess the regularity of the harmonic peaks; regular, high-amplitude harmonics yield a higher CPP than do irregular, low-amplitude harmonics. The cepstra can also be averaged over time and quefrency (i.e., the equivalent of frequency in the cepstral domain) to generate what is known as the *smoothed cepstral peak prominence* (CPPS; Heman-Ackah, Michael, & Goding, 2002; Heman-Ackah et al., 2003; Hillenbrand & Houde, 1996). Cepstral measures demonstrated higher sensitivity and specificity than did other acoustic measures in identifying dysphonic samples (Heman-Ackah et al., 2003). CPP is lower in patients with unilateral vocal fold paralysis and vocal nodules, two disorders that typically cause breathiness (Balasubramanium, Bhat, Fahim, & Raju, 2010; Hartl, Hans, Vaissière, & Brasnu, 2003; Hartl, Hans, Vaissière, Riquet, & Brasnu, 2001; Kumar, Bhat, & Prasad, 2009). In patients with dysphonia secondary to vocal fold paralysis, paresis, or bowing, CPP accounted for more variance in breathiness ratings (r^2 = .751) than other acoustic measures, including jitter, shimmer, signal-to-noise ratio, and the amplitude of the first harmonic relative to the amplitude of the second harmonic, the first formant, or the third formant (Shrivastav & Sapienza, 2003).

The possibility that cepstral measures capture general ratings of dysphonia rather than "breathiness" specifically has been raised (Hartl et al., 2001). This is supported by findings that two cepstral measures, including CPP, decreased as aperiodicity related to noise and random jitter increased (Murphy, 2006) and that CPP predicted overall severity ratings of dysphonia in muscle tension dysphonia better than other spectral and cepstral measures, with r^2 = .66 (Awan, Roy, & Dromey, 2009). CPPS that was measured from both sustained /a/ and connected speech correlated most highly with perceptual ratings of "grade" (r^2 = .64 and r^2 = .74 for /a/ and connected speech, respectively), followed by "breathiness" (r^2 = .49 and r^2 = .50, respectively), and then "roughness" (r^2 = .18 and r^2 = .25, respectively; Heman-Ackah et al., 2002).

A different type of measure called *H1–H2* is defined as the difference in magnitude of the first and second harmonics, as measured from a power spectrum of the glottal flow waveform. When the pattern of vocal fold vibration is smoothly varying and lacks abrupt changes due to collision, the glottal flow tends toward the shape of a sinusoid. This results in a glottal flow spectrum in which the F0 component (H1) dominates over the amplitudes of the other harmonics and leads to a large value of H1–H2 based on the glottal flow spectrum. It is often more convenient, however, to estimate H1–H2 from the spectrum of the acoustic pressure signal (i.e., a microphone signal). In this case, a correction must be applied to account for the effect of the first resonance of the vocal tract (F1; Hanson, 1997). The resulting measure is denoted as *H1*–H2*,* where the asterisks indicate that the correction for the presence of F1 has been applied, thus allowing the H1*–H2* magnitude to be interpreted as an indicator of the spectral and temporal characteristics of the glottal flow waveform (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995; Kreiman et al., 2007) much like H1–H2. Thus, large values of H1*–H2* imply that the glottal flow lacks abrupt changes due to vocal fold collision (Holmberg et al., 1995). Such characteristics of vocal fold vibration and glottal flow are typical of those produced in breathy voices, as has been demonstrated by inverse filtered glottal flow waveforms that contain prolonged opening and closing time courses with rounded glottal flow (Fischer-Jorgensen, 1967; Hillenbrand et al., 1994; Huffman, 1987).

It has been hypothesized that H1*–H2* is an acoustic correlate for the perception of breathy voice in healthy speakers. It has been used to differentiate "clear" from "breathy" or "murmured" vowels in languages where this distinction is phonemic (Fischer-Jorgensen, 1967; Huffman, 1987; Kirk, Ladefoged, & Ladefoged, 1993; Wayland & Jongman, 2003). H1*–H2* has been found to be higher in adult females than in adult males (Hanson & Chuang, 1999; Iseli, Shue, & Alwan, 2007; Klatt & Klatt, 1990), a difference noted as early as 15 years of age (Iseli et al., 2007). The just noticeable H1–H2 difference (which, here, was based directly on the glottal flow spectrum) in synthetically altered samples averaged

2.72 dB for Mandarin speakers and 3.61 dB for English speakers, demonstrating it to be a perceivable acoustic cue within the range of differences exhibited by speakers (Kreiman & Gerratt, 2010). In healthy subjects producing typical, moderately breathy, and severely breathy phonation, Hillenbrand et al. (1994) reported that a correlation between an uncorrected H1–H2 measure—based on the spectrum of the microphone signal and perceived breathiness—was 0.66, demonstrating that although amplitude difference between the first two harmonics varies with breathiness, it is not the sole acoustic correlate of breathy voice quality.

In superior and recurrent laryngeal nerve paralyses simulated with a two-mass model of the vocal folds, H1–H2 (in this study, measured from the derivative of the glottal flow) was higher in the paralyzed conditions than in the healthy condition (Smith, Berke, Gerratt, & Kreiman, 1992). In clinical populations, the measurement has been less compelling. H1*–H2* did not decrease from pre- to posttherapy assessment for 10 patients with vocal nodules, even though overall dysphonia and stroboscopic findings improved (Holmberg, Doyle, Perkell, Hammarberg, & Hillman, 2003). The same pattern was found from pre- to posttherapy for 24 teachers; roughness and strain decreased, but uncorrected H1–H2 measured from the long-term average spectrum did not change (Chen, Hsiao, Hsiao, Chung, & Chiang, 2007). It should be noted that breathiness ratings did not decrease in either study. Although it might be thought that H1*–H2* is associated specifically with perceived breathiness rather than overall severity of dysphonia, this explanation is unlikely, given that uncorrected H1–H2 did not differentiate between patients with unilateral vocal fold paralysis and controls (Hartl et al., 2003).

The relation between time domain measures of the glottal flow waveform and the resultant acoustic measure of H1*–H2* typically has been studied with models of the glottal flow pulse and corresponding spectral characteristics. On the basis of this approach, H1–H2 has been shown to increase when the open quotient (Qo) of the glottal waveform is large. The Qo is a ratio of (a) the time period in which the glottal airflow increases and decreases to (b) the total period of vocal fold oscillation (Hanson, 1997); Qo approaches a value of 1 when the glottal flow tends toward a sinusoid and becomes small when the flow becomes pulse-like (Murphy, 2007; Titze, Mapes, & Story, 1994). This relation has also been shown for natural speech. In a study of female voices, Holmberg et al. (1995) reported a correlation of –.69 between H1*–H2* and a "flow adduction quotient" measured from inverse-filtered flow. This adduction quotient was defined as 1–Qo, thus suggesting that an increase in H1*–H2* would be positively correlated (i.e., .69 in their study) with increasing Qo.

The studies summarized above were important for linking particular glottal flow patterns to the radiated acoustic signal and to voice quality perception. Using a model of the glottal flow and estimating a glottal flow waveform by inverse filtering, however, assume that the glottal flow reflects vocal fold kinematics and that it can be separated from the influence of the vocal tract resonances. Although these assumptions may be reasonable for a vocal production with low F0 and short glottal open phase, glottal flow typically results from nonlinear interaction of the vocal tract resonances with the aerodynamics at the glottis, thus eliminating (or at least reducing) the separability of the flow source and the vocal tract filter (Titze, 2008). The *linear source-filter theory* describes a periodic glottal flow source exciting the vocal tract resonances (Fant, 1960; Flanagan, 1972; Stevens, 2000; Stevens & House, 1961); glottal flow, then, is a reflection of vocal fold vibration and is not influenced by the acoustic pressures above or below the vocal tract (Titze, 2008). An assumption of this theory is that the filter can enhance or suppress spectral components of the source but cannot influence the source directly. The inherent simplification of the speech production system in the linear source-filter theory does not account for interactions between the source and the vocal tract that are known to occur during speech production. Titze (2008) described a *Level 1 interaction,* in which glottal flow is influenced by supraglottal and subglottal pressures, although the vocal fold vibration itself (i.e., the glottal area waveform Ag) is not. The interaction is observed in the time domain as skewing of the glottal flow relative to the glottal area, and harmonic frequencies that are not present in the glottal area waveform can appear in the glottal flow (Story, 2002; Titze, 2006b; Titze, 2008). Titze (2008) further described a *Level 2 interaction,* in which the vocal tract causes a change in the modes of vocal fold vibration. Including nonlinear interactions between vocal fold physiology and the shape of the glottal flow waveform provides a more realistic representation of the sound production system than either a glottal flow pulse model or inverse-filtered glottal flow waveforms and may facilitate understanding the relation of vocal fold kinematics to aerodynamic and acoustic quantities.

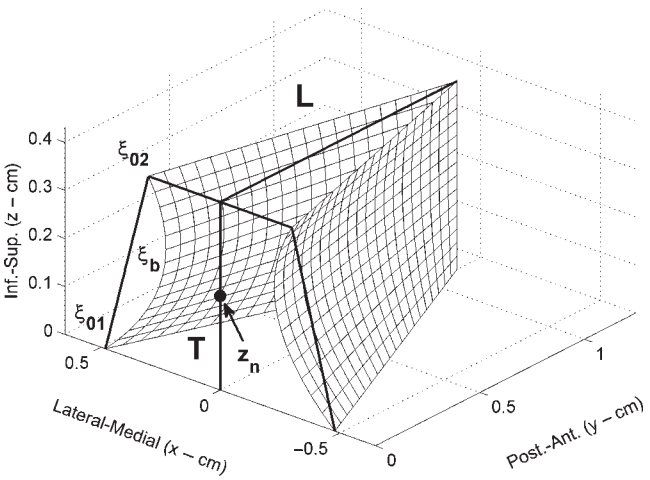
The specific aim of this study was to identify the kinematics of breathy voice quality—that is, to determine how specific structural and vibratory features that likely cause the perception of breathy voice are reflected in the acoustic measures of CPP, CPPS, H1–H2, and H1*–H2*. To determine these relations, systematic variation of individual aspects of vocal fold structure and function, combined with access to the glottal area, glottal flow, and high-quality acoustic waveforms, is required. Such manipulations are not possible in human participants but can be achieved using computer modeling.

Kinematic Model of Voice Production

A model utilizing inputs of vocal fold movement and vibration allows direct examination of the relation between vocal fold structure and vibration and acoustic measures. Titze (1984, 1989, 2006a) described a kinematic representation of the medial surface of the vocal folds in which vibration was specified by superimposing a time-varying component onto a postural component, as shown in Figure 1. In the postural component, the medial surface shape of each vocal fold is defined by a bulging parameter (ξ_b) that specifies the curvature of the vocal fold surface and a prephonatory adduction value (ξ_{02}) that sets the distance of the superior vocal process from the glottal midline. The inferior adduction value (ξ_{01}) is based on a rule that relates glottal convergence to a "nodal point ratio" (discussed below; Titze, 2006a, p. 209).

The time-varying component is a specification of vibration patterns based on normal independent movement patterns of the vocal folds, called *vibratory modes.* The modes facilitate energy transfer from the airflow to the tissue, contributing to vocal fold self-oscillation (Titze, 1988). The model is based on a *translational mode,* in which the vocal folds move together and apart in the horizontal plane, and a *rotational mode,* in which the superior and inferior vocal fold edges move to touch and separate 180° out of phase with one another (known as *vertical phase difference*). The *nodal point* (z_n) is the pivot point around which the rotational mode changes phase. A *ribbon mode* defines the anterior–posterior vibration pattern (Titze, 2006a).

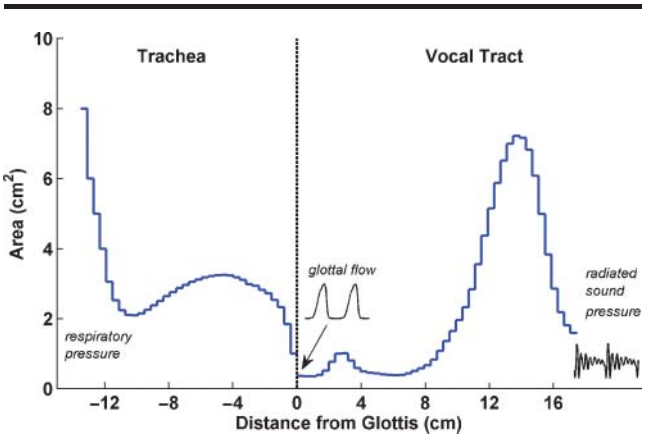
Figure 1. Kinematic model. ξ_{01} = inferior adduction; ξ_{02} = superior adduction; ξ_b = bulging; z_n = nodal point; L = vocal fold length; T = thickness; nodal point ratio (R_{zn}) = Zn/T; Inf.-Sup. = inferior–superior; Pos.-Ant. = posterior–anterior.



The kinematic vocal fold model is aerodynamically and acoustically coupled to the tracheal and vocal tract airway system (Liljencrants, 1985; Story, 1995) on the basis of findings by Titze (2002). This allows for Level 1 interactions between vocal fold vibration and the vocal tract. The vocal tract shape can be specified for various vowels using the kinematic model known as "TubeTalker" (Story, 2005), which generates an area function representation of the vocal tract. The complete model consists of the three-dimensional (3D) representation of the vocal fold medial surfaces, as shown previously in Figure 1, bounded on the upstream (subglottal) side by a tracheal area function (Story, 1995) and bronchial termination (Titze, 2006a) and on the downstream (supraglottal) side by a vocal tract area function. This system of airways is shown in Figure 2, where the tracheal section is on the left, the vocal tract is on the right, and the glottis is located at the origin of the coordinate system. The kinematic vocal fold model, together with TubeTalker, allows for specification of structural and vibratory features and subsequent observation of the simulated glottal area (Ag), glottal flow (Ug), and radiated acoustic signals.

The focus of this study was on the anatomic and kinematic parameters that likely lead to the perception of breathy voice quality. The parameter ξ_{02} describes the distance of each vocal process from midline at maximum adduction; larger ξ_{02} values increase the minimum glottal area and tend to cause rounding of the Ag waveform (see Figure 1). Bulging (ξ_b) provides a convex curvature to the vocal fold edge. A value of 0 indicates no curvature to the edge and, although the "typical" ξ_b during phonation is unknown, values of 0.103 cm during abduction and 0.016 cm during adduction were measured from

Figure 2. Speech production model showing the shape of the trachea and vocal tract (as configured for the vowel /a/). The dashed vertical line represents the location of the kinematic vocal fold model shown in Figure 1.



denervated canine larynges (Titze, 2006a). The curvature is thought to occur as the result of contraction of the thyroarytenoid muscle, and higher bulging values assist adduction, lessening the effects of high ξ_{02} (Alipour & Scherer, 2000). In addition to muscle effort, ξ_b can be increased through surgical intervention to a bowed or immobile fold.

The nodal point (z_n) is the pivot point of the rotational mode, as described previously, and is related to the point of mucosal upheaval (Yumoto, Kudota, & Kurokawa, 1993). The location of z_n is difficult to determine in human participants because current vibratory visualization techniques do not provide high-quality imaging in a coronal plane. Although some indication of the range of z_n values expected in human vocal fold vibration can be inferred from medial surface measurements (Döllinger & Berry, 2006; Yumoto et al., 1993) and modeling experiments (Alipour, Berry, & Titze, 2000), the actual range and role of z_n in vibratory kinematics is not well understood. For purposes of modeling, the nodal point location can be more efficiently represented as the ratio of the nodal point to the thickness of the vocal folds (see Figure 1) such that a *nodal point ratio* is defined as $R_{zn} = z_n/T$. For the simulation experiment in this study, R_{zn} was allowed to range from 0.2 cm to 0.8 cm, which is nearly the full extent of the vocal fold thickness.

Although the parameters ξ_{02}, ξ_b, and R_{zn} influence the shape and timing of the vocal fold vibration and the glottal area (Ag) waveform, the resultant glottal flow (Ug) waveform is also influenced by the inertance of the supraglottal vocal tract (Ishizaka & Flanagan, 1972; Rothenberg, 1983) and generation of aspiration noise due to turbulence. It has been shown that increasing the supraglottal inertance by narrowing the epilaryngeal area leads to increased skew of the glottal flow and more rapid flow declination as well as decreased oscillation threshold pressure (Titze, 1988, 2008; Titze & Story, 1997). Aspiration noise is produced by turbulence when the airflow through the glottis is high. The aerodynamic mechanisms through which such turbulence is generated are complex (Krane, 2005; Krane, Barry, & Wei, 2007; Shadle, 1985; Sinder, 1999; Zhang & Mongeau, 2006) and consequently were not represented computationally in our model. Instead, the effect of turbulence was approximated by adding a noise component to the glottal flow when the Reynolds number (Re) exceeded a threshold value. This approach has often been used in speech production modeling for both aspiration and fricative-type sounds (e.g., Fant, 1960; Flanagan & Cherry, 1969), although it is clearly recognized that the physical realities of jet structure, vortex shedding, production of dipole and quadrupole noise sources, and potential multiple locations of such sources are not represented.

The specific formulation of a noise source added to the glottal flow used in this study is based on Titze (2006a, p. 263). At every time point, the Reynolds number is calculated as

$$\mathrm{Re} = \frac{U_g \rho}{L \mu}, \tag{1}$$

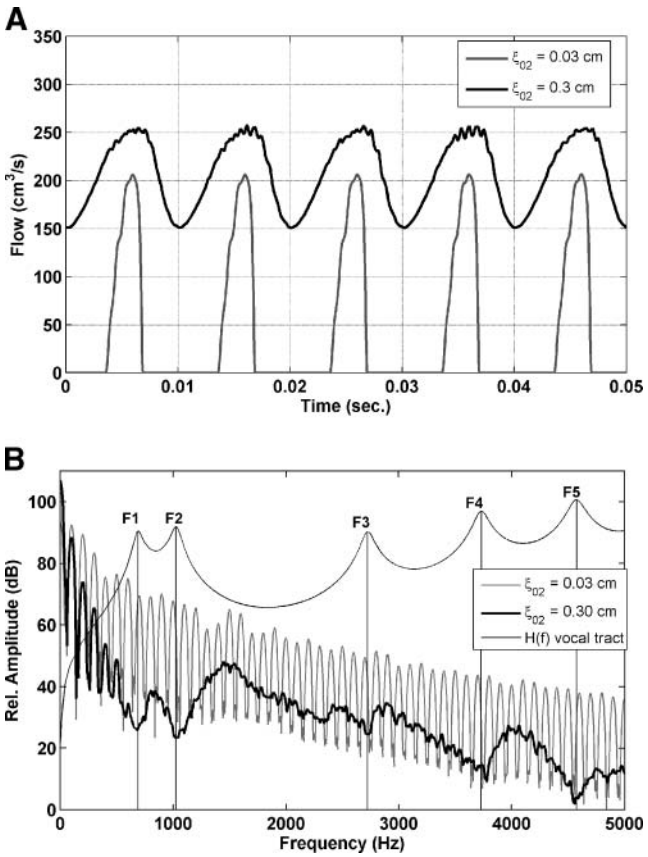
where U_g is the instantaneous glottal flow, L is the length of the glottis, ρ is the air density, and μ is the air viscosity (Titze, 2006a, p. 263). The noise component of the flow is then generated in the form proposed by Fant (1960), such that

$$U_{nois} = \begin{cases} N_f(\mathrm{Re}^2 - \mathrm{Re}_c^2)(4 \times 10^{-6}) & \text{for} \quad \mathrm{Re} > \mathrm{Re}_c \\ 0 & \text{for} \quad \mathrm{Re} \le \mathrm{Re}_c \end{cases}, \tag{2}$$

where N_f is a broadband noise signal (random noise generated with values ranging in amplitude from –0.5 to 0.5) that has been band-pass filtered between 300 Hz and 3000 Hz (second order Butterworth), Re is the calculated Reynolds number, and Re_c is a threshold value below which no noise is allowed to be generated. Fant (1960, p. 274) suggested that Re_c should be on the order of 1800 or less for fricative sounds. On the basis of spectral analysis of simulated vowels and preliminary listening experiments, a value of $Re_c = 1200$ was chosen, along with the scaling factor of 4×10^{-6}, for all subsequent simulations in this study. The result is a noise source whose amplitude is modulated by the periodic variation of the nonturbulent glottal flow.

Figure 3A shows two sample glottal flow waveforms that were simulated with the kinematic vocal fold model coupled to the vocal tract and tracheal airways. The waveform (which is shown as the gray line) resulted from a setting of $\xi_{02} = 0.03$ cm and has a long closed phase and small noise component. The black line is the glottal flow generated when $\xi_{02} = 0.3$ cm; in this case, there is no closed phase, an offset flow of about 150 cm^3/s, and a strong noise component that can be seen as the fine jagged structure in high-amplitude portions of the waveform. Spectra of each waveform are shown in Figure 3B. The strong harmonic structure of the $\xi_{02} = 0.03$ cm case (gray spectrum) can be clearly seen by the presence of harmonic peaks over the entire 5000-Hz range shown in the plot. In contrast, the black spectrum representing the $\xi_{02} = 0.30$ cm case has a large DC component, harmonic frequencies that extend only to 500 Hz, and a noise spectrum shaped by resonances of the vocal tract and trachea. In fact, the amplitude of this flow spectrum is suppressed near frequencies corresponding to the vocal tract resonances as shown by the transfer function also plotted in Figure 3B. A similar suppression of the flow amplitude at select frequencies was also noted by Titze (2008) and results from the nonlinear interaction

Figure 3. Waveforms and spectra produced by the kinematic model. **Panel A:** Example glottal flow (Ug) waveforms generated by the model. The gray line is the Ug for ξ_{02} = 0.03 cm, and the black line is the Ug for ξ_{02} = 0.3 cm. **Panel B:** Spectra of the waveforms shown in Figure 3A. sec. = seconds; Rel. = relative; H(f) = frequency response function.

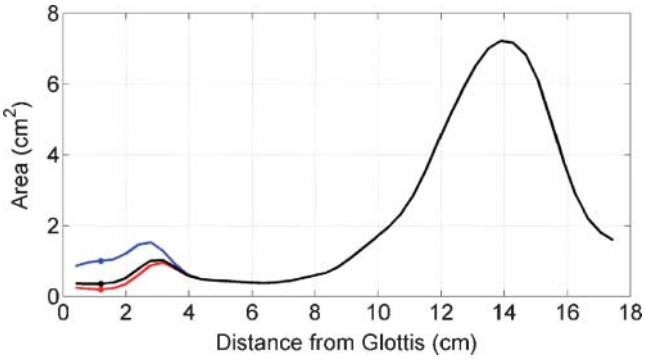


of the flow with the acoustic wave propagation (i.e., Level 1 interaction).

Signal Generation

The model explained in the previous section was used to generate vowel samples. In pilot work, three settings for each of five vocal fold parameters thought to relate to breathiness were used to generate samples of sustained /a/ vowels based on the vocal tract area function shown in Figure 4. The parameters that systematically varied were adduction (ξ_{02}), bulging (ξ_b), nodal point ratio (R_{zn}), fundamental frequency (F_0), and left/right phase difference. Four expert listeners—each with at least 15 years of experience rating impaired voices—rated the samples for roughness, breathiness, and asthenia. Experts judged the higher F_0 condition as asthenic rather than breathy, so the parameter was eliminated from future studies. Other parameters were narrowed to a range where listeners perceived breathiness. Further pilot testing with experts revealed that epilaryngeal area

Figure 4. Vocal tract model for the vowel /a/ showing the three configurations of the epilarynx at 0–4 cm from the glottis. The black line is the neutral setting, the red line represents the constriction to 0.2 cm^2, and the blue line demonstrates the expansion to 1.0 cm^2.



(A_{epi}) modification moderated breathiness ratings, and it was added as a test parameter.

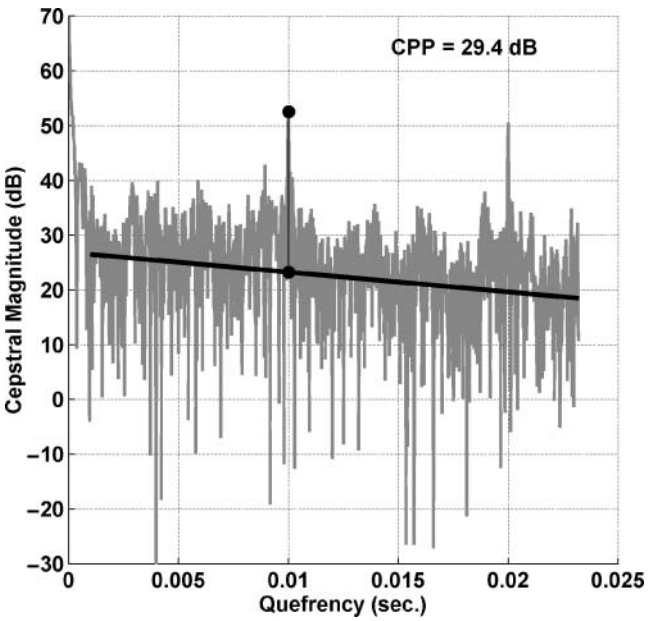
Based on analysis of the expert listeners' ratings, the following four physiological parameters were modified for the present study: ξ_{02}, ξ_b, R_{zn}, and A_{epi}. The data were generated in sets of 900 simulations in which all model parameter settings were identical for the left and right vocal folds. Each set included combinations of (a) 30 values of ξ_{02}, equally spaced between 0 cm (no gap between the vocal folds) and 0.3 cm and (b) 30 values of R_{zn}, equally spaced between 0.2 cm and 0.8 cm. The simulation sets differed by the values of the bulging parameter ξ_b and the configuration of the entry portion of the vocal tract (referred to as the *epilarynx*). For each of the 900 simulations in a given set, the bulging parameter ξ_b was maintained at one of four constant values: 0.01, 0.1, 0.15, and 0.2 cm. In addition, three configurations of the epilarynx were used where the defining parameter, called A_{epi}, was chosen to be the cross-sectional area at a location 1.2 cm from the glottis. In the first configuration, the epilarynx was maintained in what is called (throughout this article) the *neutral* configuration. This was the original measured /a/ vocal tract area function, where A_{epi} = 0.36 cm^2 (Story, 2008). The second and third configurations were generated by constricting and expanding the epilarynx relative to the neutral case (the A_{epi} was set to 0.2 cm^2 and 1.0 cm^2, respectively, for the constricted and expanded cases). The epilaryngeal area settings are depicted in Figure 4, demonstrating the differences in epilaryngeal area and the smooth integration of this area with the portions of the vocal tract upstream and downstream (Story, 2005).

Measurements of Simulated Signals

All acoustic measurements were completed automatically, as the 900 vowels within any given set were

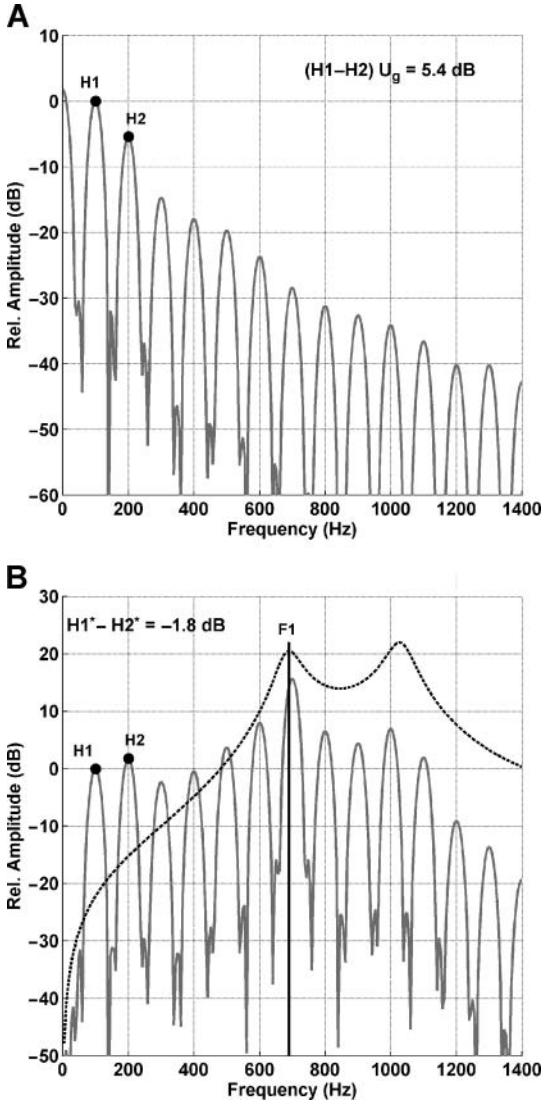
simulated. The CPP was measured using publicly available software (SpeechTool With CPP Scripts; Hillenbrand, 2008) that was executed from within a MATLAB script. This algorithm first calculates the cepstrum as the log power spectrum of the log power spectrum of the original signal. High-amplitude peaks in a cepstrum indicate that the signal has a well-defined harmonic structure; conversely, low-amplitude peaks (or the absence thereof) indicate that a signal may contain a significant noise component (Hillenbrand & Houde, 1996). The light gray line in Figure 5 shows an example of a cepstrum for a signal generated with ξ_{02} = 0.08 cm. The two distinct peaks located at quefrencies of 0.01 and 0.02 s result from the F_0 (100 Hz) and H2 (200 Hz) of the signal, respectively. The first step in determining the CPP is to fit a linear regression line relating the cepstral amplitude to quefrency. This provides an overall normalization of the cepstrum and is demonstrated by the thick black line in Figure 5. Note that the cepstral information below a quefrency of 0.001 s is not included in regression line calculation (Hillenbrand & Houde, 1996). The CPP value is calculated as the difference in amplitude between the highest amplitude cepstral peak and the amplitude of the regression line at the quefrency of that same peak; this is indicated in the figure by the two black dots whose difference is 29.4 dB. The smoothed version of the CPP (i.e., CPPS) is similarly calculated but is done so after the cepstrum has been smoothed with an averaging filter (Hillenbrand & Houde, 1996).

Figure 5. Calculation of the cepstral peak prominence (CPP). The cepstrum is shown as the gray line, a regression line is shown in black, and the two black dots indicate the height of the maximum cepstral peak and the amplitude at the same quefrency on the regression line, respectively. The distance between the two dots is the CPP.



The H1–H2 measures were made using custom MATLAB code. For H1–H2, spectra of the glottal flow were calculated every 0.0125 s and were averaged over the 0.4-s duration of a glottal flow signal. Each spectrum in the average was based on a 0.05-s sample that was multiplied by a Hanning window and zero-padded to 8,192 points. A peak-picking algorithm (Titze, Horii, & Scherer, 1987) was used to measure the amplitude of the first two harmonics from the mean glottal flow spectrum, and H2 was then subtracted from H1. An example

Figure 6. Calculation of amplitude difference between the first harmonic (H1) and the second harmonic (H2). **Panel A:** H1–H2 based on the glottal flow spectrum. **Panel B:** H1*–H2* based on the spectrum of the radiated acoustic pressure. F1 = first formant; F0 = fundamental frequency.



of a glottal flow spectrum is shown in Figure 6A where the overall amplitude has been adjusted so that H1 = 0 dB, and the H1 and H2 peaks are marked with the solid black dots. It is noted that the spectrum contains a strong DC component because the glottal flow signal was not preemphasized. To measure H1*–H2*, a spectrum was similarly generated for the pressure signal radiated at the lips, and the peak-picking algorithm was again used to find the amplitude of the first two harmonics. These values were, however, corrected for effect of the first formant (F1), and the corrected H2 was subtracted from the corrected H1 (Hanson, 1997). This is demonstrated in Figure 6B, where the output pressure spectrum is shown as the light gray line, and H1 and H2 are marked with black dots. The frequency response of the vocal tract shape (i.e., the /a/ vowel area function shown in Figure 2) is indicated by the dashed line; this was calculated with a frequency domain method (Sondhi & Schroeter, 1987; Story, Laukkanen, & Titze, 2000) that included energy losses due to yielding walls, viscosity, heat conduction, and radiation. The first formant frequency (F1), determined from the frequency response, was used in the H1 and H2 correction formula given by Hanson (1997) and is shown in the upper left portion of the plot. Typically, a linear predictive coding (LPC)–based analysis or other formant tracking approach would be needed to determine F1 on the basis of a speech signal, but because the vocal tract area function was available from the simulations, F1 could be determined directly.

Results

Figure 7 shows four 3D plots (surfaces) that demonstrate how each of the acoustic measurements varies as a function of parameters ξ_{02} and R_{zn}. The configuration is identical for each of the surfaces; the *x*-axis represents the 30 values of ξ_{02}, the *y*-axis represents the 30 values of R_{zn}, and the *z*-axis is always a particular acoustic measurement (i.e., H1–H2, H1*–H2*, CPP, and CPPS). For all surfaces in this figure, vocal fold bulging was 0.1 cm, and the epilaryngeal area was the neutral setting.

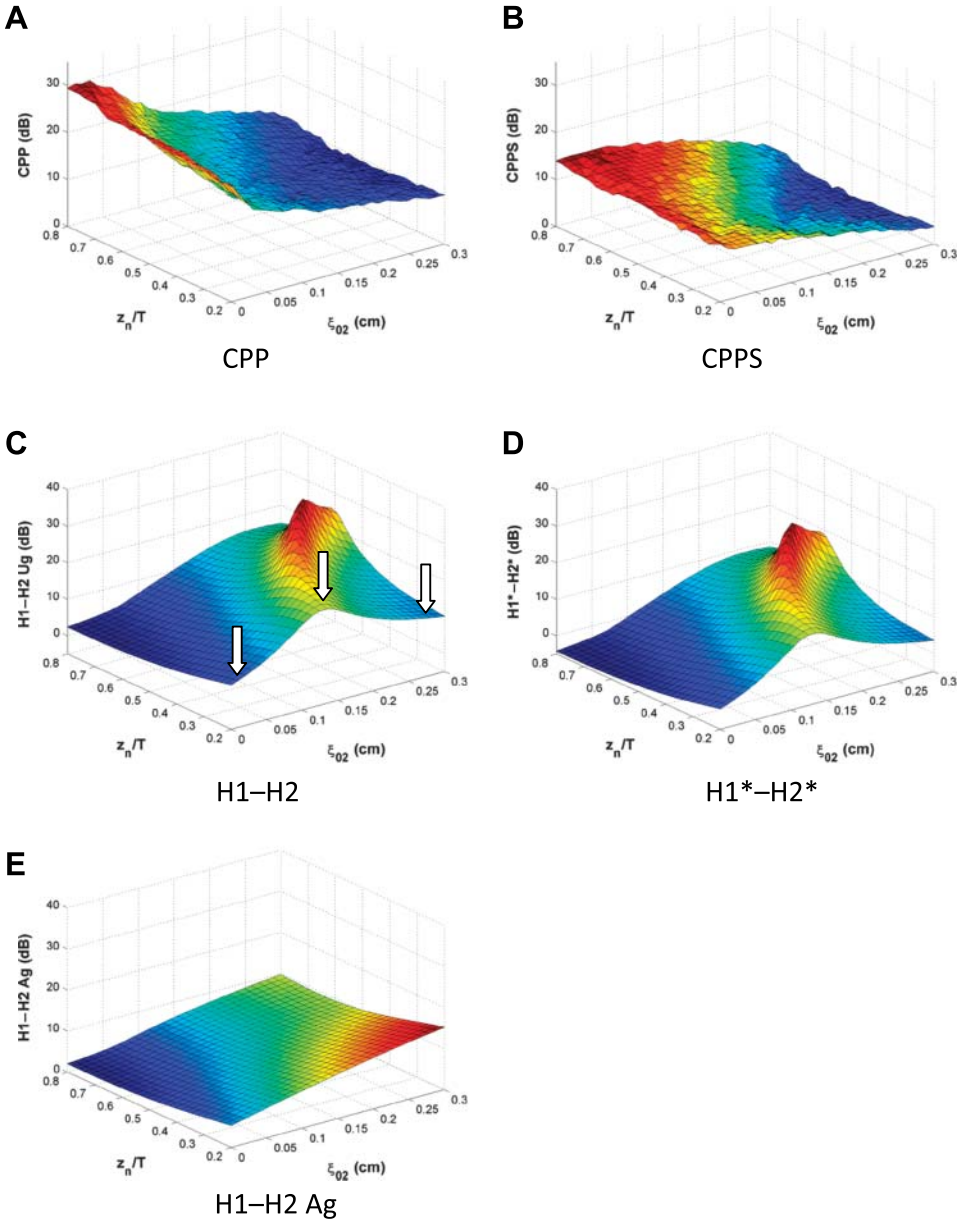
The results for CPP and CPPS are shown in Figures 7A and 7B, respectively. These are fairly simple surfaces from which it is clear that both measures decrease as ξ_{02} increases, whereas there is minimal change with an increase or decrease in R_{zn}. These figures indicate that increasing the distance between the vocal processes (e.g., increasing the size of glottal gap at maximum closure) decreases the regularity or strength of harmonics. It is noted that the range of the values for the CPP surface is about 20 dB, whereas, for the CPPS surface, the range is only 10 dB.

The surfaces for H1–H2 and H1*–H2* are shown in Figures 7C and 7D, respectively, and indicate a more complex relation to the underlying vocal fold parameters. At low nodal point ratio values, H1–H2 in Figure 7C increases with separation of the vocal processes for the smaller ξ_{02} values (up to 0.1 cm) and then decreases in value as ξ_{02} continues to increase. For nodal point ratios closer to the superior edge of the fold, the peak value of H1–H2 occurs at higher values of ξ_{02}, in effect producing a curved ridge on the 3D surface. For example, the maximum H1–H2 at R_{zn} = 0.6 occurs at ξ_{02} = 0.25 cm, compared to a similar maximum value that occurs when ξ_{02} = 0.1 cm and R_{zn} = 0.2. It is interesting to note that the sharp rise and fall is avoided at high nodal point ratio values such that H1–H2 gradually increases over most of the range of ξ_{02}. Although the H1*–H2* values in Figure 7D are different than those in Figure 7C, the pattern is the same: H1*–H2* first increases and then decreases as the size of the glottal gap and nodal point ratio increase. In either case, the presence of the ridge is a curious result because it indicates that H1–H2 can be an increasing function of vocal fold separation (and, presumably, breathiness) but only over a limited range of ξ_{02} values.

To determine whether the ridge in the H1–H2 surfaces occurs as the result of vocal fold vibration properties, H1–H2 was calculated from the glottal area signal, and the result is shown in Figure 7E. H1–H2 for glottal area increases as a function of ξ_{02}, indicating that the ridge is not created by some aspect of vocal fold vibration but by aerodynamic interaction with the vocal tract.

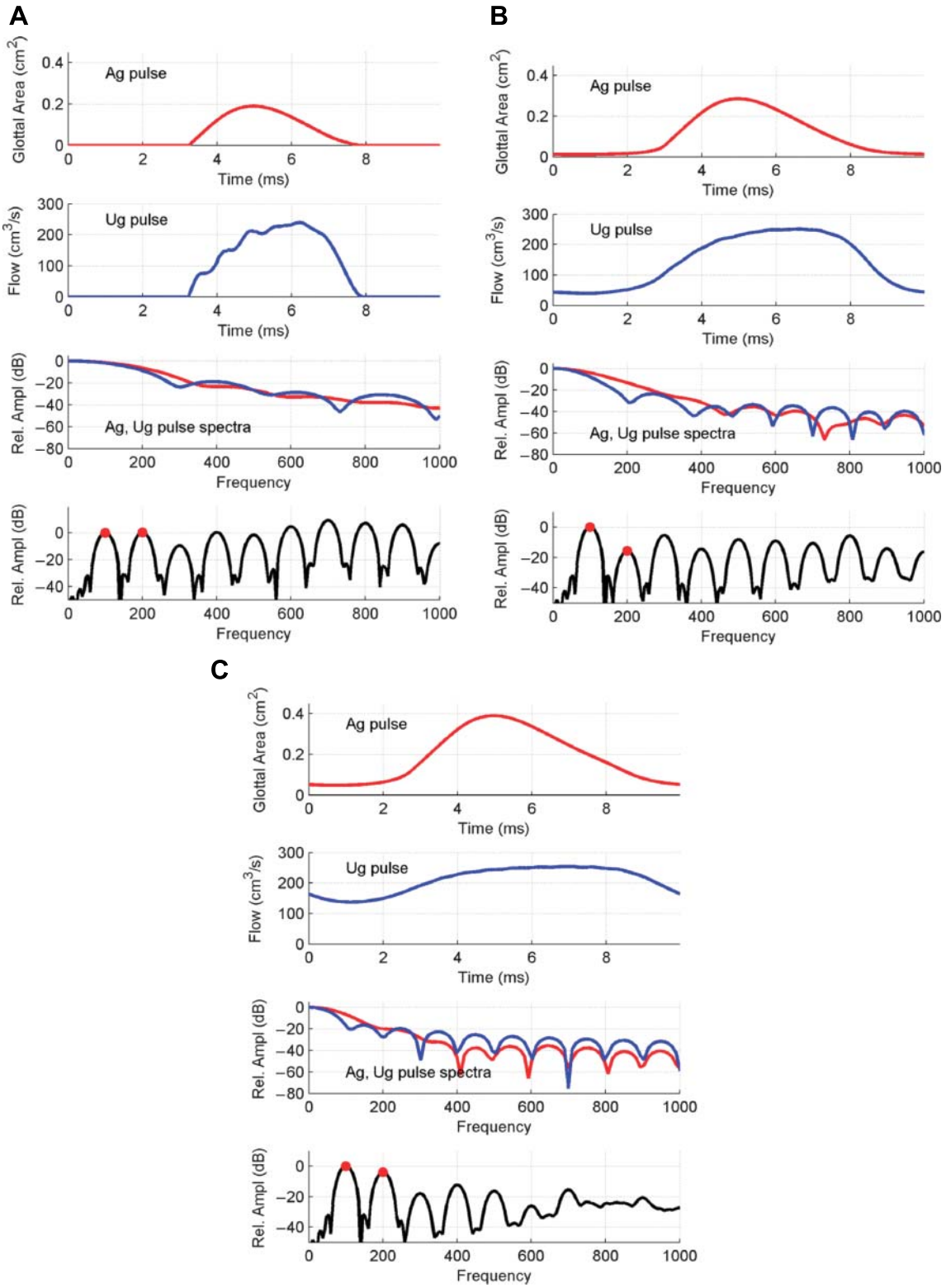
To further understand the contribution of the vocal tract and whether voice samples with the highest H1*–H2* demonstrate other acoustic features of breathiness or whether the ridge pattern is an artifact of the H1–H2 measurement, glottal area and flow pulses along with their corresponding spectra were generated for a series of points indicated by the arrows in Figure 7C. The points occur at the coordinates [R_{zn}, ξ_{02}] = [0.25, 0.02], [0.25, 0.12], and [0.25, 0.25]. Results are shown in Figure 8. At the most adducted point (see Figure 8A, first panel), the Ag pulse is skewed to the left, demonstrating that the glottal area increase (i.e., vocal fold separation) occurs faster than does the decrease in glottal area (i.e., vocal folds return to midline). The second panel in Figure 8A is the glottal flow pulse, and it is skewed to the right, demonstrating the influence of supraglottal inertance in a slower buildup to maximum flow and a more rapid-flow shutoff; formant ripple is also apparent. The third panel shows the spectra of the Ag (red) and Ug (blue) pulses. The Ug pulse spectrum indicates shallow zeroes, the first of which occurs at approximately 300 Hz and does not affect the amplitude of the fundamental or second harmonic. The fourth panel in Figure 8A is the spectrum of the output pressure signal, P_{out}. Note that the first zero observed in the third panel does not coincide with either of the first two harmonics, thus limiting its

Figure 7. The *x*- and *y*-axes describe the parameter settings of the kinematic vocal fold model for every sample; the *x*-axis displays 30 values of ξ_{02}, ranging from 0 cm to 0.3 cm, and the *y*-axis displays 30 values of z_n/T (i.e., R_{zn}), ranging from 0.2 cm to 0.8 cm. The *z*-axis displays the measured parameter: CPP (see Panel A); smoothed cepstral peak prominence (CPPS; see Panel B); H1–H2 (see Panel C); H1*–H2* (see Panel D); and H1–H2 for glottal area (Ag; see Panel E). All surfaces were generated using ξ_b = 0.1 cm and a neutral epilaryngeal area (A_{epi}) setting. The arrows in Figure 7C denote the locations of the points analyzed in Figure 8.



influence on the H1*–H2* value. The third harmonic, however, appears lowered in amplitude by the zero at approximately 300 Hz. As ξ_{02} is increased and the peak H1*–H2* value is reached (see Figure 8B), maximum Ag increases, and the pulse widens. The area pulse remains skewed to the left, and the flow pulse is skewed to the right. The flow pulse becomes wider (increased Qo), and there is a DC offset (see third panel of Figure 8B). The depth of the zero in the glottal flow pulse spectrum is similar to the previous example, except the location has shifted and the first zero occurs at 200 Hz. This causes a depression in the P_{out} spectrum coinciding with—and suppressing—H2. Additional zeroes occur at 400 Hz, 600 Hz, and 800 Hz, lowering H4 and H6 relative to the first point, although H8 is boosted by F1 at approximately 800 Hz. These two figures confirm that the

Figure 8. Ag pulse, Ug pulse, Ag and Ug spectra, and P_{out} spectra for three points [R_{zn}, ξ_{02}] at low nodal point ratio: [0.25, 0.02] (Panel A); [0.25, 0.12] (Panel B); and [0.25, 0.25] (Panel C). Rel. Ampl = relative amplitude.



Ug spectrum—and, therefore, interaction with the vocal tract—is responsible for the shape of the H1*–H2* surface rather than characteristics of vocal fold vibration itself, at least for a Level 1 type of interaction (i.e., flow interaction only).

For the point with the greatest degree of vocal process separation (see Figure 8C), a DC offset appears in the Ag pulse, and peak flow increases. The pulse retains its leftward skew. The DC offset of the Ug pulse increases, and the peak flow remains high. The flow pulse becomes quite wide, increasing the Qo. There are zeroes in the Ug spectrum at 100 Hz and 200 Hz, corresponding to both H1 and H2. Because both H1 and H2 are suppressed by the interaction with the vocal tract, their difference is smaller than in the previous example, in which only H2 was suppressed. There is a deeper zero at 300 Hz, and H3 shows relatively greater suppression than do H1 and H2. Zeroes also appear in the Ag spectrum but not until 400 Hz. In the P_{out} spectrum, limited harmonic energy is evident beyond H2. There is an increase in spectral amplitude for F1 at approximately 800 Hz, but this appears to be noise rather than harmonic energy. Thus, it appears that the combinations of R_{zn} and ξ_{02} along the ridge in Figures 7C and 7D produce a Qo that suppresses the second harmonic, resulting in a low-valued H2, even though the harmonics above 2F0 are readily apparent in the spectrum. It is noted that the CPP, being a more global measure of spectral regularity, is not as acutely affected by the Qo as the more spectrally focused H1–H2 measure.

Because similar trends were observed for both the CPP and CPPS surfaces and for both the H1–H2 and H1*–H2* surfaces, only the CPP and H1*–H2* surfaces are presented with respect to the additional parameter variations. These choices were made based on the wider range of values exhibited by the CPP measure and on the fact that H1*–H2* is clinically accessible, whereas its counterpart from the glottal flow signal typically is not.

CPP

Figure 9 shows a matrix of surfaces that demonstrates the variation in measured CPP values as a function of ξ_{02}, nodal point ratio (R_{zn}), medial surface bulging, and epilaryngeal constriction. Each surface is configured as shown previously in Figure 7, and they are arranged such that each consecutive column represents an increase in epilaryngeal area and each row represents an increase in bulging. In all cases, the CPP decreased monotonically from about 30 dB when the vocal processes were approximated (ξ_{02} = 0 cm) down to around 10 dB when ξ_{02} was set to its maximum value (i.e., largest glottal gap). There was a slight influence of R_{zn} on CPP value, whereby the CPP was slightly ($\leq$ 5 dB) higher at R_{zn} = 0.8 than at R_{zn} = 0.2 for all cases when the difference was more than 1 dB. The difference was typically more evident at higher bulging and higher ξ_{02} settings.

The surface with the neutral epilaryngeal area and ξ_b = 0.1 cm was previously shown in Figure 7A and is used here as a reference case (see Figure 9E) to which the other cases are compared. Although it is not suggested that this condition necessarily represents a "typical" case, choosing it as a reference is based on a measurement of ξ_b = 0.1 cm in a canine larynx (Titze, 2006a) and the measurement of epilaryngeal area in a human subject (Story, 2008).

Relative to the reference case, the effect of decreased bulging can be seen in Figure 9B; when the R_{zn} is high and ξ_{02} is low, as in the region in the upper left part of the CPP surface, CPP drops rapidly with an increase in ξ_{02}. This suggests that muscle atrophy or soft tissue deficit—both of which would contribute to a reduction in medial surface bulging—may lead to low CPP values with any separation of the vocal processes. The effect of increased bulging relative to the reference can be seen in the third and fourth rows of Figure 9B, where relatively high CPP values are a maintained over a larger range of ξ_{02} settings, especially when the R_{zn} value is greater than about 0.5. For example, in the case with a neutral epilaryngeal area and ξ_b = 0.15 cm (see Figure 9H), the CPP drops rapidly from about 30 dB with increasing ξ_{02} when R_{zn} is 0.2, but when R_{zn} is at 0.8, the CPP is nearly constant at around 30 dB for ξ_{02} values that range from 0 cm to 0.1 cm (although the CPP does slightly rise above 30 dB at about ξ_{02} = 0.05 cm). As the bulging increases to ξ_b = 0.2 cm, this same effect is enhanced such that high CPP is maintained for an even greater range of ξ_{02} values (until ξ_{02} = 0.15 cm) when the R_{zn} is high. These results suggest that increased bulging and a high R_{zn} can be used to maintain high CPP (large harmonic energy) even as separation of the vocal processes (ξ_{02}) is increased. Constricting the A_{epi} to 0.2 cm^2 (see column 1, "Constricted A_{epi}") also had the effect of maintaining higher CPP for a greater range of ξ_{02} values, whereas there was no appreciable effect of expanding the A_{epi} to 1.0 cm^2 (see column 3, "Expanded A_{epi}").

H1*–H2*

The surfaces in Figure 10 are arranged in the same manner as those in Figure 9; the epilaryngeal area, A_{epi}, increases across the columns, and ξ_b increases with each row. The reference surface based on the neutral epilaryngeal area and ξ_b = 0.1 cm is shown in Figure 10E and was previously shown in Figure 7D. Even though this measure is thought to increase with breathiness, H1*–H2* almost always increased with ξ_{02} for part of the adductory range and then decreased as ξ_{02} continued

Figure 9. CPP. Each plot is a three-dimensional surface representing the CPP results for 900 simulations of the vowel /a/. The *x*-axis is 30 values of ξ_{02}, ranging from 0 cm to 0.3 cm, and the *y*-axis is 30 values of z_n/T (i.e., R_{zn}), ranging from 0.2 cm to 0.8 cm. The *z*-axis shows the measured parameter CPP. In each row, A_{epi} increases from left to right. In each column, ξ_b increases from top to bottom.

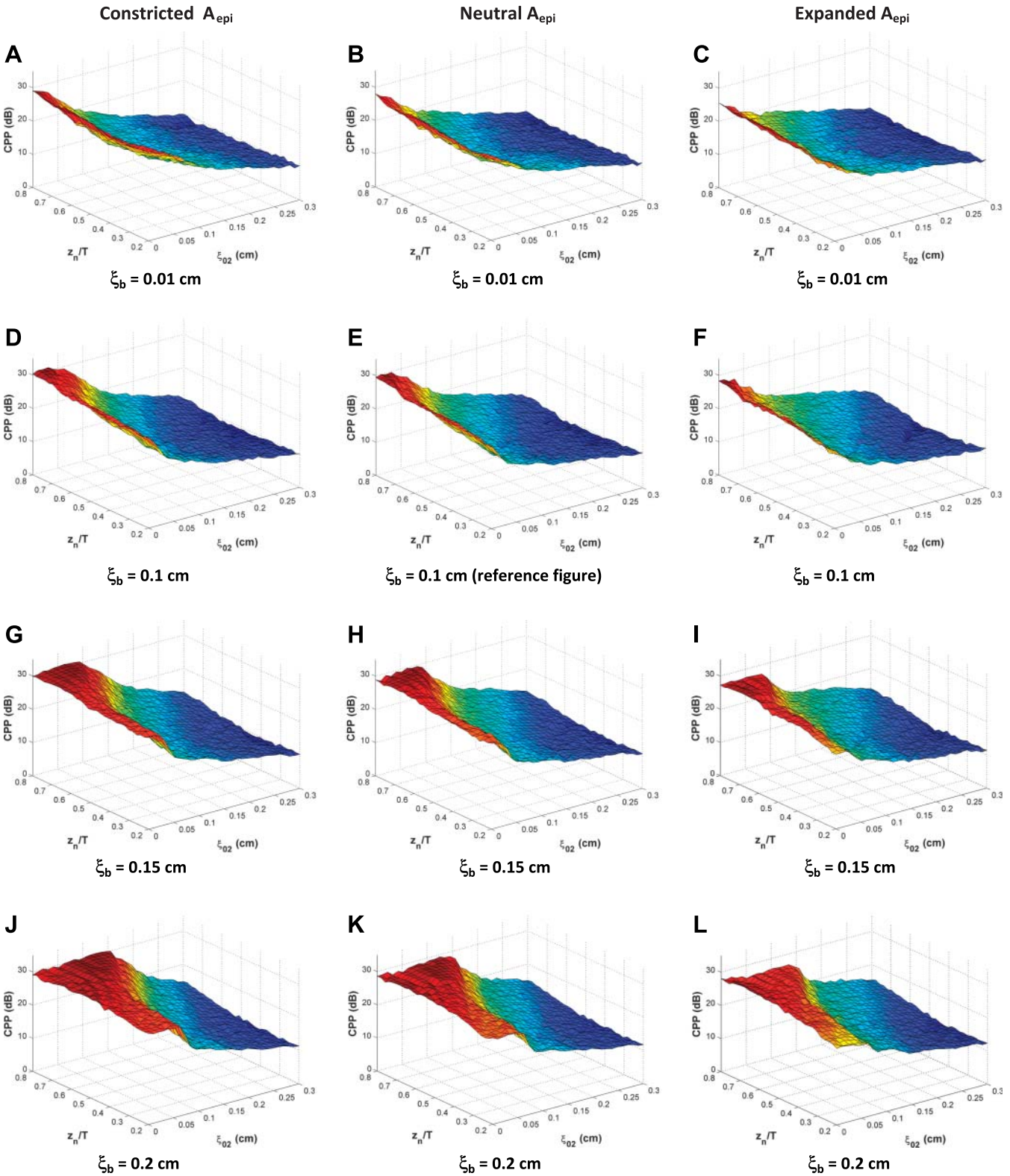
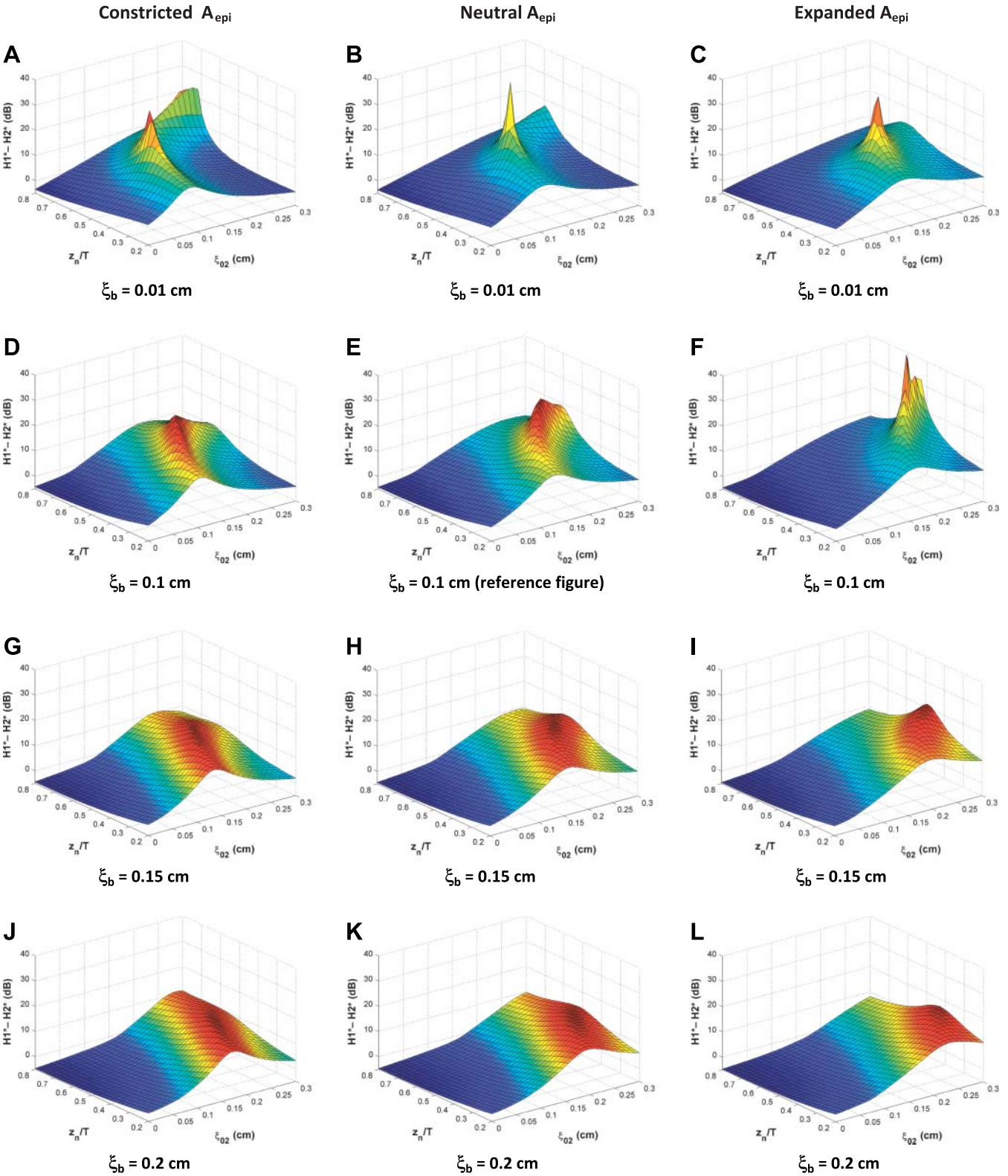


Figure 10. H1*–H2.* Each plot is a three-dimensional surface representing the H1*–H2* results for 900 simulations of the vowel /a/. The *x*-axis is 30 values of ξ_{02}, ranging from 0 cm to 0.3 cm, and the *y*-axis is 30 values of z_n/T (i.e., R_{zn}), ranging from 0.2 cm to 0.8 cm. The *z*-axis shows the measured parameter H1*–H2*. In each row, A_{epi} increases from left to right. In each column, ξ_b increases from top to bottom.



to increase. The value was negative at small ξ_{02} values in each figure, indicating that the corrected amplitude of the second harmonic was greater than that of the first harmonic when glottal gap was small.

The effect of decreased bulging can be observed by comparing the reference case (see Figure 10E) to the case in Figure 10B. A higher maximum H1*–H2* value occurred when bulging was decreased (43.9 dB vs. 28.1 dB), and the "peak" occurred with smaller ξ_{02} and lower R_{zn} settings. At the lowest R_{zn} value, decreased bulging minimally altered the location of maximum H1*–H2* with respect to degree of adduction. The H1*–H2* values were slightly higher with lower bulging (ξ_b) when the folds were fully adducted, suggesting that decreasing the bulging causes a slight increase in breathiness with small gap size, yet the maximum H1*–H2* value and value at ξ_{02} = 0.3 cm were similar in both cases. At the highest R_{zn} value, maximum H1*–H2* and H1*–H2* at ξ_{02} = 0.3 cm were slightly lower in the condition of decreased bulging than in the reference condition. The same patterns were observed for the constricted and expanded epilaryngeal area. In comparing Figure 10D with Figure 10A and Figure 10F with Figure 10C, areas of highest H1*–H2* (the peak and the ridge) are narrower, particularly when R_{zn} is high.

Consistent with decreasing the degree of breathiness, increasing the surface bulging, ξ_b, led to lower maximum H1*–H2* values as can be seen in Figures 10G through 10L. The maximum occurred at similar ξ_{02} settings as in the reference case, and at lower R_{zn} settings. At the lowest R_{zn}, H1*–H2* was slightly lower for the surfaces with higher ξ_b at minimum ξ_{02} and slightly higher at maximum ξ_{02}. This was also the case at the highest R_{zn}, where the H1*–H2* was negative or 0–1 dB for increasingly larger ranges of the ξ_{02} continuum as ξ_b increased, and higher at maximum ξ_{02}. At a bulging of 0.2 cm, for example, H1*–H2* was negative or 0–1 dB until ξ_{02} reached 0.2 cm, compared with an ξ_{02} of 0.12 dB for an ξ_b = 0.1 cm. The ξ_{02} location of the peak H1*–H2* increased as ξ_b increased for both R_{zn} values. The consequent "smoothing" of the peaks and increase in the range of ξ_{02} for which H1*–H2* increased was reflected across epilaryngeal area conditions. This can be observed by comparing Figure 10J with Figures 10G and 10J as well as comparing Figure 10F with Figures 10I and 10L.

Constricting the epilaryngeal area to a value of 0.2 cm^2 resulted in a lower maximum H1*–H2*, consistent with decreasing the degree of breathiness. It also decreased the H1*–H2* measured at ξ_{02} = 0.3 cm, more so at the lowest R_{zn} than at the highest. Expanding the epilaryngeal area to a value of 1.0 cm^2 resulted in a higher maximum H1*–H2 and shifted the nodal point ratio at which it occurred from 0.53 to 0.45.

Discussion

Cepstral peak prominence decreased as the separation between the vocal processes increased, regardless of other parameter settings. These findings are consistent with the inverse relation of CPP to perceived breathiness reported by Shrivastav and Sapienza (2003; see their Figure 1). The range of CPP values reported for their participants with breathiness (≈9–18 dB) is compatible with the CPP range measured from the simulated signals in this study (≈10–30 dB). In the present study, CPP was minimally influenced by R_{zn}, and the CPP at a particular adduction setting generally increased as the vocal fold edge bulging increased. These results lend credibility to the hypothesis that CPP generally reflects the size of the glottal gap during maximum glottal closure.

Certainly, CPP is responsive to vocal fold structural and kinematic changes that lead to increased space between the folds at maximum closure; however, the measure is also influenced by the shape of the supraglottal vocal tract. Decreasing epilaryngeal area while maintaining constant bulging, adduction, and R_{zn} typically increased the CPP value. It has been shown that constricting the epilarynx has caused clustering of F3, F4, and F5 (Story, 2004; Sundberg, 1974; Titze & Story, 1997), which should lift the cepstral peak by increasing the harmonic energy in the spectrum. In addition, epilaryngeal constriction causes increased inertance, which tends to increase the rightward skew of the glottal flow, also potentially increasing overall harmonic energy (cf. Rothenberg, 1983; Titze, 2008). Previous reports that CPP decreases as signal aperiodicity, overall dysphonia ratings, and breathiness increase (Awan et al., 2009; Heman-Ackah et al., 2002; Murphy, 2006) are consistent with the findings of this study—that is, increasing the separation between the vocal processes at maximum closure generally led to decreased harmonic energy and increased random (noise) energy in the higher harmonics of the sound pressure wave, which resulted in decreased CPP. The consistent reports of high correlations between CPP and breathiness, compared with inconsistent correlations between H1–H2 and breathiness, are likely related to the more global aspect of this measure—that it reflects aspects of both the harmonic and noise components of a breathy voice.

The surprising finding was that although the H1–H2 of the glottal area consistently increased with increasing separation of vocal processes, H1–H2 of the glottal flow and radiated acoustic signal did not. When observed as a function of ξ_b, R_{zn}, and A_{epi}, H1*–H2* increased up to an ξ_{02} value ranging from 0.10 cm to 0.30 cm and then decreased. H1*–H2*, then, effectively has a different relation to increasing glottal gap depending on the vocal fold shape, pivot point for the rotational mode, and supraglottal vocal

tract shape. None of these features can be fully known in human subjects, so it is not possible to determine how the measure can be applied to a specific group of subjects. This is true in subjects who exhibit abnormal vocal fold structure or vibration and also for healthy subjects producing different degrees of breathiness. It is likely that different vocal fold shapes, pivot points, and supraglottal shapes are being used from speaker to speaker, placing them at different portions of the surfaces. Therefore, the H1*–H2* values might increase with glottal gap size for one group and not the other.

In a study of systematic manipulation of H1 amplitude in synthesized copies of disordered voices, the just noticeable difference for H1–H2 was lower when spectral slope was steeper (Kreiman & Gerratt, 2010). These findings indicate that although H1–H2 might provide some information about the source spectrum, features that dictate spectral slope influence perception. Certainly, in the present study, variability in direction of change in H1*–H2* with increasing glottal gap appears related to the shape of the glottal flow waveform. Increased open quotient of the glottal flow waveform is responsible for the location of zeroes in the glottal flow spectrum, and both Qo and the skew are responsible for the depth of the zeroes (Titze et al., 1994). Alignment of the zeroes with the first or second harmonic frequency has a substantial impact on the resultant measure. It is interesting to note that even though glottal area also demonstrated larger Qo with increasing ξ_{02}, the larger Qo did not produce zeroes in the relevant range of the glottal area spectrum. Therefore, the characteristics of Ug that create the variable relation of H1*–H2* to ξ_{02} occur because of the nonlinear source–tract interaction rather than the increased duration of glottal opening during each vibratory cycle. As source–tract coupling increases through larger glottal area, the glottal flow becomes increasingly dependent on pressures in the subglottal and supraglottal systems and, therefore, is less predictable based on the Ag waveform (Titze, 2008).

It was further unexpected that the measure based on glottal flow (H1–H2) appeared identical in shape to the measure from the radiated acoustic signal (H1*–H2*). These findings are in contrast to those of Castillo-Guerra and Ruíz (2009), who compared acoustic measures to perceptual judgments of breathy voice. Two measures were used to assess the ratio between H1 and H2: *harmonic energy* (HE), which was based on the radiated acoustic signal, and *harmonic energy of residue* (HE_{RES}), which was calculated after passage through an inverse-filter model. The measure HE_{RES} better accounted for variance in the perceptual judgments across multiple types of perturbation, leading the authors to conclude that filtering effects of the vocal tract from the signal improves the accuracy of the information conveyed about breathy voice (Castillo-Guerra & Ruíz, 2009). The most likely explanation for this is that the authors did not appear to correct the harmonic amplitudes for the influences of F1 in their calculation. It is also possible that the inverse filtering approach that they used in calculating HE_{RES} did not capture the nonlinear source–tract interaction.

As with any model, there are limitations of the approach used in this study. One limitation is that the pattern of vocal fold vibration was generated by the superposition of only two basic vibratory modes, and the left and right folds vibrated in perfect symmetry. It is certainly possible that a component of breathy voice originates from more complex and, perhaps, asymmetric vibratory patterns that were not modeled in this study; these additional complexities could be implemented in future work. A second, and perhaps more severe, limitation is the means by which the turbulence-based noise component is added to the model. As explained in a previous section, modulating a broadband noise generator with the glottal flow waveform can provide what appears to be reasonable nonperiodic spectral energy, which is based on both acoustic analysis and listening. The limitation, however, is that the physical mechanisms of noise generation are not included in the model. Although quite complicated to implement, inclusion of aeroacoustic calculations would automatically initiate turbulence without a preset threshold and provide appropriate spectral shaping and amplitude scaling.

Conclusions

CPP is generally a function of vocal process separation and decreases with increasing vocal process separation across multiple levels of surface bulging and pivot point locations for the rotational mode. In general, the prominence of the cepstral peak is preserved at a higher value for the same glottal gap when surface bulging is increased or when epilaryngeal area is mildly decreased. H1*–H2* will increase or decrease with vocal process separation based on vocal fold bulging, pivot point for the rotational mode, and supraglottal vocal tract shape, limiting its utility as an indicator of breathy voice. Future work is underway to relate the perception of breathiness to vocal fold kinematics and acoustic measures.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant R01-DC04789.

References

Alipour, F., Berry, D., & Titze, I. (2000). A finite-element model of vocal-fold vibration. *The Journal of the Acoustical Society of America, 108,* 3003–3012.

Alipour, F., & Scherer, R. (2000). Vocal fold bulging effects on phonation using a biophysical computer model. *Journal of Voice, 14,* 470–483.

Awan, S., Roy, N., & Dromey, C. (2009). Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clinical Linguistics & Phonetics, 23,* 825–841.

Balasubramanium, R., Bhat, J., Fahim, S., & Raju, R. (2010). Cepstral analysis of voice in unilateral adductor vocal fold palsy. *Journal of Voice, 25,* 326–329. doi:10.1016/j.jvoice.2009.12.010.

Castillo-Guerra, E., & Ruíz, A. (2009). Automatic modeling of acoustic perception of breathiness in pathological voices. *IEEE Transactions in Biomedical Engineering, 56,* 932–940.

Chen, S., Hsiao, T., Hsiao, L., Chung, Y., & Chiang, S. (2007). Outcome of resonant voice therapy for female teachers with voice disorders: Perceptual, physiological, acoustic, aerodynamic, and functional measurements. *Journal of Voice, 21,* 415–425.

Colton, R. H., Casper, J. K., & Leonard, R. (2006). *Understanding voice problems: A physiological perspective for diagnosis and treatment* (3rd ed.). Baltimore, MD: Lippincott Williams & Wilkins.

Döllinger, M., & Berry, D. (2006). Computation of the three-dimensional medial surface dynamics of the vocal folds. *Journal of Biomechanics, 39,* 369–374.

Fant, G. (1960). *The acoustic theory of speech production*. The Hague, the Netherlands: Mouton.

Fischer-Jorgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics, 28,* 71–139.

Flanagan, J. L. (1972). *Speech analysis, synthesis, and perception*. New York, NY: Springer.

Flanagan, J. L., & Cherry, L. (1969). Excitation of vocal-tract synthesizers. *The Journal of the Acoustical Society of America, 45,* 764–769.

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

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

Hartl, D., Hans, S., Vaissière, J., & Brasnu, D. (2003). Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia. *European Archives of Oto-Rhino-Laryngology, 260,* 175–182.

Hartl, D., Hans, S., Vaissière, J., Riquet, M., & Brasnu, D. (2001). Objective voice quality analysis before and after onset of unilateral vocal fold paralysis. *Journal of Voice, 15,* 351–361.

Heman-Ackah, Y., Heuer, R., Michael, D., Ostrowski, R., Horman, M., Baroody, M., … Sataloff, R. T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otology, Rhinology, and Laryngology, 112,* 324–333.

Heman-Ackah, Y., Michael, D., & Goding, G. J. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16,* 20–27.

Hertegård, S., & Gauffin, J. (1995). Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography, and electroglottography. *Journal of Speech and Hearing Research, 38,* 85–100.

Hillenbrand, J. (2008). SpeechTool With CPP Scripts [Computer program]. Retrieved from http://homepages.wmich.edu/~hillenbr/.

Hillenbrand, J., Cleveland, R., & Erickson, R. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37,* 769–778.

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

Holmberg, E., Doyle, P., Perkell, J., Hammarberg, B., & Hillman, R. (2003). Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy. *Journal of Voice, 17,* 269–282.

Holmberg, E., Hillman, R., & Perkell, J. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America, 84,* 511–529.

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

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

Iseli, M., Shue, Y., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America, 121,* 2283–2295.

Ishizaka, K., & Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *The Bell Systems Technical Journal, 512,* 1233–1268.

Kirk, P. L., Ladefoged, J., & Ladefoged, P. (1993). Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec. In A. Mattina & T. Moutler (Eds.), *American Indian linguistics and ethnography in honor of Laurence C. Thompson* (pp. 435–450). Missoula, MT: University of Montana Press.

Klatt, D., & Klatt, L. (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.

Krane, M. H. (2005). Aeroacoustic production of low-frequency unvoiced speech sounds. *The Journal of the Acoustical Society of America, 118,* 410–427.

Krane, M. H., Barry, M., & Wei, T. (2007). Unsteady behavior of flow in a scaled-up vocal folds model. *The Journal of the Acoustical Society of America, 122,* 3659–3670.

Kreiman, J., & Gerratt, B. R. (2010). Perceptual sensitivity to the first harmonic amplitude in the voice source. *The Journal of the Acoustical Society of America, 128*(4), 2085–2089.

Kreiman, J., Gerratt, B., & Antoñanzas-Barroso, N. (2007). Measures of the glottal source spectrum. *Journal of Speech, Language, and Hearing Research, 50,* 595–610.

Kumar, B. R., Bhat, J. S., & Prasad, N. (2009). Cepstral analysis of voice in persons with vocal nodules. *Journal of Voice, 24,* 651–653. doi:10.1016/j.jvoice.2009.07.008.

Liljencrants, J. (1985). *Speech synthesis with a reflection-type line analog* (Unpublished doctoral dissertation). Royal Institute of Technology, Stockholm, Sweden.

Murphy, P. (2006). Periodicity estimation in synthesized phonation signals using cepstral harmonic peaks. *Speech Communication, 48,* 1704–1713.

Murphy, P. J. (2007, August). *Relationship between harmonic amplitudes and spectral zeros and glottal open quotient*. Paper presented at the International Conference on Phonetic Science (ICPhS), Saarbrucken, Saarland, Germany.

Rothenberg, M. (1983, June). Source–tract acoustic interaction and voice quality. In V. L. Lawrence (Ed.), *Transcripts of the Twelfth Symposium: Care of the Professional Voice* (pp. 15–31). New York, NY: The Voice Foundation. Retrieved from www.rothenberg.org/Sourcetractvq/Sourcetractvq.htm.

Shadle, C. H. (1985). *The acoustics of fricative consonants* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Shrivastav, R., & Sapienza, C. (2003). Objective measures of breathy voice quality obtained using an auditory model. *The Journal of the Acoustical Society of America, 114*(4, Pt. 1), 2217–2224.

Sinder, D. J. (1999). *Speech synthesis using and aeroacoustic fricative model* (Unpublished doctoral dissertation). Rutgers: The State University of New Jersey, New Brunswick.

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

Södersten, M., & Lindestad, P. (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech and Hearing Research, 33,* 601–611.

Sondhi, M. M., & Schroeter, J. (1987). A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 35,* 955–967.

Stevens, K. N. (2000). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, K., & House, A. (1961). An acoustical theory of vowel production and some of its implications. *Journal of Speech and Hearing Research, 4,* 303–320.

Story, B. (1995). *Speech simulation with an enhanced wave-reflection model of the vocal tract* (Unpublished doctoral dissertation). University of Iowa, Iowa City.

Story, B. (2002). An overview of the physiology, physics, and modeling of the sound source for vowels. *Acoustical Science and Technology, 23,* 195–206.

Story, B. (2004). On the ability of a physiologically constrained area function model of the vocal tract to produce normal formant patterns under perturbed conditions. *The Journal of the Acoustical Society of America, 115,* 1760–1770.

Story, B. (2005). A parametric model of the vocal tract area function for vowel and consonant simulation. *The Journal of the Acoustical Society of America, 117,* 3231–3254.

Story, B. (2008). Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. *The Journal of the Acoustical Society of America, 123,* 327–335.

Story, B. H., Laukkanen, A. M., & Titze, I. R. (2000). Acoustic impedance of an artificially lengthened and constricted vocal tract. *Journal of Voice, 14,* 455–469.

Sundberg, J. (1974). Articulatory interpretation of the singing formant. *The Journal of the Acoustical Society of America, 55,* 838–844.

Titze, I. (1984). Parameterization of the glottal area, glottal flow, and vocal fold contact area. *The Journal of the Acoustical Society of America, 75,* 570–580.

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

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

Titze, I. (2002). Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model. *The Journal of the Acoustical Society of America, 111,* 367–376.

Titze, I. (2006a). *The myoelastic aerodynamic theory of phonation*. Iowa City, IA: National Center for Voice and Speech.

Titze, I. (2006b). Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. *Journal of Speech, Language, and Hearing Research, 49,* 439–447.

Titze, I. (2008). Nonlinear source-filter coupling in phonation: Theory. *The Journal of the Acoustical Society of America, 123,* 2733–2749.

Titze, I. R., Horii, Y., & Scherer, R. C. (1987). Some technical considerations in voice perturbation measurements. *Journal of Speech, Language, and Hearing Research, 30,* 252–260.

Titze, I., Mapes, S., & Story, B. (1994). Acoustics of the tenor high voice. *The Journal of the Acoustical Society of America, 95,* 1133–1142.

Titze, I., & Story, B. (1997). Acoustic interactions of the voice source with the lower vocal tract. *The Journal of the Acoustical Society of America, 101,* 2234–2243.

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

Yumoto, E., Kadota, Y., & Kurokawa, H. (1993). Tracheal view of vocal fold vibration in excised canine larynxes. *Archives of Otolaryngology–Head and Neck Surgery, 119,* 73–78.

Zhang, Z., & Mongeau, L. G. (2006). Broadband sound generation by confined pulsating jets in a mechanical model of the human larynx. *The Journal of the Acoustical Society of America, 119,* 3995–4005.