Nonlinear source-filter coupling in phonation: Theory

1 author:

Ingo R Titze
University of Utah
420 PUBLICATIONS   **19,731** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   singing voice View project

Project   Voice disorders and assessment View project

Nonlinear source–filter coupling in phonation: Theory[a)]

Ingo R. Titze[b)]

Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242 and
National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204

A theory of interaction between the source of sound in phonation and the vocal tract filter is developed. The degree of interaction is controlled by the cross-sectional area of the laryngeal vestibule (epilarynx tube), which raises the inertive reactance of the supraglottal vocal tract. Both subglottal and supraglottal reactances can enhance the driving pressures of the vocal folds and the glottal flow, thereby increasing the energy level at the source. The theory predicts that instabilities in vibration modes may occur when harmonics pass through formants during pitch or vowel changes. Unlike in most musical instruments (e.g., woodwinds and brasses), a stable harmonic source spectrum is not obtained by tuning harmonics to vocal tract resonances, but rather by placing harmonics into favorable reactance regions. This allows for positive reinforcement of the harmonics by supraglottal inertive reactance (and to a lesser degree by subglottal compliant reactance) without the risk of instability. The traditional linear source–filter theory is encumbered with possible inconsistencies in the glottal flow spectrum, which is shown to be influenced by interaction. In addition, the linear theory does not predict bifurcations in the dynamical behavior of vocal fold vibration due to acoustic loading by the vocal tract. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2832337]

I. INTRODUCTION

The acoustic features of all vowel productions, and many consonant productions, have generally been described by a linear source–filter theory (Chiba and Kajiyama, 1958; Fant, 1960; Flanagan, 1972; Stevens, 1999). This linear theory is based on the assumption that the source of sound for vowels and voiced consonants (pulsatile airflow in the larynx) is independent of the filter, an acoustic resonator formed by airways known as the vocal tract. It is traditionally assumed that the source–filter combination can be characterized by mathematical convolution of source and filter functions in the time domain or by multiplication of Fourier-transformed source and filter functions in the frequency domain. Time domain convolution and frequency domain multiplication are linear mathematical operations that carry with them the superposition assumption; that is, the output of any combination of inputs is the linear combination of all the individual outputs. Stated another way, the output of any and all input frequencies can at most be an amplitude and phase changed version of these input frequencies. The filter cannot influence the source to produce new frequencies or change the overall energy level of the source. We will show here that this assumption is generally not valid, but under certain conditions is an appropriate simplification.

Many aspects of speech production have been successfully described by a linear source–filter theory. In particular, linear prediction of speech (Markel and Gray, 1976; Atal and Schroeder, 1978) has been the flagship of speech analysis, synthesis, and processing for over 30 years. But it has been recognized all along, however, that the linear theory is more applicable to male speech than female and child speech (e.g., Klatt and Klatt, 1990). As long as the dominant source frequencies lie well below the formant frequencies of the vocal tract, the source is influenced only in simple ways by the filter, mainly in terms of glottal flow pulse skewing and pulse ripple. This mild interaction occurs for most male adult speech, but greater interaction occurs for female and child speech, and even more for singing, where the fundamental frequency range spans more than two octaves and the lower partials of the source cross the formants. In these more intense interactions, bifurcations in the dynamics of vocal fold vibration can occur that may generate sudden F_0 jumps, subharmonic frequencies, or changes in the overall energy level at the source. The earliest computer simulations of source–filter coupling (Flanagan, 1968; Flanagan and Landgraf, 1968; Ishizaka and Flanagan, 1972) showed the interactivity clearly. The one-mass vocal fold model did not self-sustain oscillation without a vocal tract (a highly exaggerated coupling effect), and the two-mass model showed sudden frequency discontinuities when F_0 passed through the first formant frequency F_1. In human phonation, as our companion paper shows, the frequency discontinuities are clearly observable, but to a lesser extent than in these earlier models.

The purpose of this paper is to elucidate the underlying mechanisms of source–filter interaction, both with simple analytical models and with a highly sophisticated computational model. Specific questions of interest are: (1) what is the primary parameter that regulates the degree of interaction, (2) can new frequencies and greater output power be produced with increased interaction, (3) are there regions of

---

harmonic stability and instability that can be exploited by a vocalist, (4) are sudden F_0 discontinuities (reported in a companion paper on human subjects) always triggered by F_0 interactions with F_1, the first formant frequency, or can higher harmonic-formant interactions contribute as well, and (5) how does the subglottal system contribute to nonlinear coupling differently from the supraglottal system?

It is hypothesized that humans (and perhaps many animals) have the ability to operate their source–filter system with either linear or nonlinear coupling. One way to express the degree of coupling is through the relative impedances of the source and filter. For linear source–filter coupling, the source impedance (transglottal pressure divided by glottal flow) is kept much higher than the input impedance to the vocal tract (vocal tract input pressure divided by the airflow into the vocal tract). This linear coupling is accomplished by adducting the vocal folds firmly and widening the epilarynx tube (a normally narrow region of the vocal tract above the vocal folds also known as the laryngeal vestibule). The glottal flow is then determined strictly by aerodynamics, while acoustic pressures above and below the glottis have little influence on either the transglottal pressure (which drives the glottal flow) or the intraglottal pressure (which drives the vocal folds). For nonlinear source–filter coupling, the glottal impedance is adjusted to be comparable to the vocal tract input impedance, making the glottal flow highly dependent on acoustic pressures in the vocal tracts (above and below the glottis). This is accomplished by setting specific adduction levels of the vocal folds that match a narrower epilarynx tube. Evidence of nonlinear coupling is the production of new frequencies in the form of distortion products, lowering of the oscillation threshold pressure (the Hopf bifurcation), production of subharmonics or modulation frequencies, sudden F_0 jumps, or chaotic vibrations, as either vowel or F_0 are changed (companion paper).

In many attempts to model the glottal airflow with explicit mathematical formulas (e.g., Rosenberg, 1971; Fant et al., 1985; Fant and Lin, 1987), nonlinear source–filter coupling in the form of distortion products had already been introduced implicitly by making the glottal flow pulse shape different from the glottal area pulse shape. Usually, the peak of the flow pulse was delayed with respect to the area pulse shape. But such a delay (referred to as flow pulse skewing) cannot be justified if only quasisteady aerodynamic calculations are carried out (i.e., linear source–filter coupling). Rothenberg (1981) showed that the peak delay of the glottal flow pulse is a result of low-frequency source–filter interaction. Zhao et al. (2002) make a case for intraglottal pressure skewing for different glottal shapes based on an aerodynamic treatment that involves flow separation and vortex shedding, but the effect on glottal flow is likely to be small in comparison to the vocal tract loading effect, which has recently been further explored by this author (Titze, 2001; 2004a; b; 2006a). There is now strong evidence that glottal flow pulse skewing always involves the vocal tract, explicitly or implicitly. Constructing a glottal flow pulse shape without consideration of the vocal tract load can lead to inconsistencies in combining the source with the filter.

Source–filter interactions that involve changes in vocal fold vibration have been demonstrated by several investigators. Sudden changes in vocal fold vibration can be triggered by vocal tract length changes (Hatzikirou et al., 2006), a good example of source–filter interaction. Further observations about source–filter coupling were reported by Švec et al. (1999) on human subjects and excised larynges, Mergell and Herzel (1997) on a female subject, Miller and Schutte (2005) on singers, Neumann et al. (2005) on male opera singers, Zhang et al. (2006a,b) on a physically constructed model, Zañartu et al., (2007) with a computer simulation model, and Jiang and Tao (2007) with analytical mathematic methods. Some of these studies will be referred to in more detail later. In singing, vowel modifications (e.g., changing /u/ to /U/ or /i/ to /I/) are used routinely to strengthen a vowel on a certain pitch (Appelman, 1967; Coffin, 1987). Entire singing styles (operatic, musical theatre, yodeling) are based on the concept that certain vowels and voice qualities work best with certain pitches, a concept that would have no explanation if the source–filter system were linear. The entire voice register terminology is based on observed phenomena related to interaction within the source–filter building blocks, which includes the subglottal system (Titze, 2000; Chap. 10). Vocal pedagogues who invented terms like chest voice and head voice were not so naive to suggest that the source of sound moves from location to location, but rather that interactions with certain parts of the airway are stronger with certain source–filter adjustments and lead to special sensations along the airway. The role of the subglottal system for chest voice was implicated years ago by Van Den Berg (1957) and Vennard (1967). Chest voice production has both a glottal feature (a relatively long closed phase) and significant acoustic coupling to the trachea. For head voice, there appears to be more of an interaction with the supraglottal tract.

In an accompanying paper (Titze et al., 2008), a primary objective is to differentiate purely source-generated bifurcations, including F_0 jumps, from vocal tract induced bifurcations. Three vocal exercises are designed in the accompanying paper to deliberately cross the fundamental frequency with the first formant frequency and to observe the resultant nonlinear effects in nine normal males and nine normal females. It is hypothesized that crossing F_0 with F_1 changes the acoustic load dramatically and that this crossing can destabilize vocal fold vibration. The main goal is to determine the proportion of irregularities that are due to nonlinear source–tract interactions. Expected manifestations of nonlinearity are sudden pitch jumps, subharmonic generation, or chaotic vocal fold vibration. Results indicate that the most frequent bifurcation is a sudden F_0 jump, as predicted by Fletcher (1993) for pressure-controlled valves in gas flows. The objective in this paper is to provide a theoretical framework for the bifurcation phenomena in vocal fold vibration with a nonlinear source–filter construct.

II. INTERACTION BASED ON VOCAL TRACT REACTANCE

Because the vocal tract is relatively short in comparison to a wavelength at typical speech fundamental frequencies
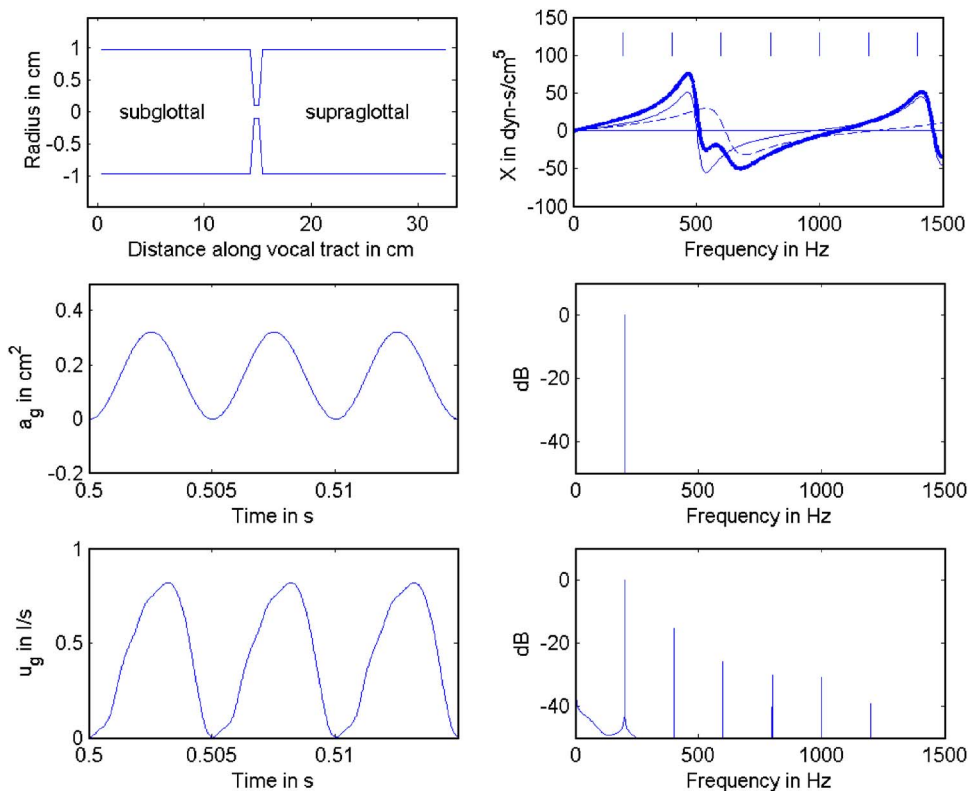
Ingo R. Titze: Nonlinear source–filter theory

FIG. 1. (Color online) Harmonic frequency generation by source–filter interaction; (top left) vocal tract shape; (top right) reactance curves, thin solid line for supraglottal, dashed line for subglottal, and thick solid line for combined; (middle left) sinusoidal glottal area function; (middle right) spectrum of glottal area; (bottom left) glottal flow; (bottom right) spectrum of glottal flow.

(e.g., at $F_0 = 200$ Hz the vocal tract contains less than $1/8$ of a wavelength), and because a speaker or singer wishes to convey all the phonetic variations of a spoken language, the length and shape of the vocal tract cannot be adjusted to resonate many of the source frequencies simultaneously. Thus, unlike in most musical instruments, for which the length and shape of the horn or bore is carefully designed to resonate the dominant source frequencies simultaneously, lining up source frequencies with vocal tract filter resonances is highly selective and rare in human phonation. Apparent "formant-harmonic tuning" occurs in high soprano singing (Sundberg, 1977; Joliveau *et al.*, 2004), but close inspection of the data reveals that F_0 is usually slightly less than the formant frequency F_1. In some cases, at very high F_0, oscillation occurs for $F_0 > F_1$. We will show that this is more likely for a falsetto-like vibration regime. Exact tuning of F_1 with a harmonic seems to occur only in so-called *overtone singing*, where a high frequency harmonic is reinforced by a formant that is tuned precisely to its frequency (Rachele, 1996).

For low-pitched speech or singing, the dominant source harmonics (typically F_0 through $3F_0$) are below the first resonance (formant) frequency F_1 of the vocal tract. For example, a bass or baritone speaking or singing a note G_2 (98 Hz) will reach the first formant of an /i/ or a /u/ vowel only with the third harmonic. For an /a/ vowel, source harmonics higher than the seventh are needed to reach the first formant. Because these higher harmonics often do not have a great influence on the nature of vocal fold vibration, their interaction with a formant is perceived as a vowel or voice quality characteristic, not a source change. But the flow pulse may nevertheless be influenced by higher harmonic source–filter interaction, as will be shown next.

A. Level 1 interaction: Flow pulse dependency on the subglottal and supraglottal vocal tract pressures

A first level of interaction is described in which vocal fold vibration is not significantly disturbed by oscillating pressures above or below the glottis, but glottal flow is. This level occurs in all speech and is therefore worthy of special consideration.

What is common about harmonics whose frequencies are less than F_1 is that they all experience positive (inertive) reactance from the vocal tract.[1] Figure 1 (top left) shows a uniform vocal tract (subglottal and supraglottal) separated by the glottis. The top right panel shows the reactance curves (subglottal as a dashed line, supraglottal as a thin solid line, and combined reactance as a thick solid line) up to 1500 Hz. The reactance curves were calculated with cascade transmission line matrices as originally outlined by Sondhi and Schroeter (1987) and further developed by Story *et al.* (2000). From 0 to 500 Hz, the supraglottal reactance is positive (inertive), whereas from 500 to 1000 Hz it is negative (compliant). The subglottal reactance stays inertive up to 600 Hz for this configuration.

Inertive reactance has been shown to skew the flow pulse (delay its peak relative to that of the glottal area), whether it is subglottal (X_1) or supraglottal (X_2). For a review of this flow pulse skewing, see Rothenberg (1981), Fant (1986), Fant and Lin (1987), or Titze (2006a). Note the delay in the peak of U_g in Fig. 1 (bottom left) in relation to a_g (middle left). Further discussion of these graphs will follow. For glottal flow calculation, the subglottal input impedance and the supraglottal input impedance add together algebraically (in complex form) to produce the effective load imped-

ance for the glottis. This can be shown by letting the input pressure to the vocal tract (epilarynx tube) be

$$P_e = Z_2 U_g, \tag{1}$$

where U_g is the complex (Fourier transformed) glottal flow and Z_2 is the complex supraglottal impedance. Similarly, the subglottal pressure is

$$P_s = Z_1(-U_g), \tag{2}$$

where Z_1 is the subglottal impedance. The transglottal pressure is then

$$P_s - P_e = -(Z_1 + Z_2)U_g = -(R_1 + R_2)U_g - i(X_1 + X_2)U_g, \tag{3}$$

where R_1 and R_2 are the subglottal and supraglottal resistances, $i = \sqrt{-1}$, and X_1 and X_2 are the corresponding reactances. Note that the combined reactance $X_1 + X_2$ shown in Fig. 1 (top right, thick line) is not necessarily symmetric about the horizontal axis because the subglottal and supraglottal peaks do not occur at identical frequencies.

For incompressible, quasisteady glottal flow, the transglottal pressure can also be expressed aerodynamically as

$$P_s - P_e = k_t \tfrac{1}{2} \rho u_g^2 / a_g^2, \tag{4}$$

where k_t is an empirically-determined transglottal pressure coefficient with an average value of about 1.1 (Scherer et al., 1983; Alipour and Scherer, 2007; Fulcher et al., 2006), ρ is the air density, u_g is the time-dependent flow, and a_g is the time-varying glottal area. Equation (4) is nonlinear in u_g and cannot easily be Fourier transformed, and since the resistances and reactances in Eq. (3) are in the frequency domain, Eqs. (3) and (4) cannot be equated. In previous work with a wave-reflection analog of the vocal tract (Titze, 1984), it was shown that the flow had the following closed-form solution:

$$u_g = \frac{a_g c}{k_t} \left\{ -\left(\frac{a_g}{A^*}\right) \pm \left[\left(\frac{a_g}{A^*}\right)^2 + \frac{4k_t}{\rho c^2}(p_s^+ - p_e^-) \right]^{1/2} \right\}, \tag{5}$$

where c is the sound velocity, and A^* is an equivalent vocal tract area defined as

$$A^* = A_s A_e / (A_s + A_e), \tag{6}$$

with A_s and A_e being the subglottal and supraglottal (epilaryngeal) entry areas, respectively. Further, in the wave-reflection algorithm (Kelly and Lochbaum, 1962; Liljencrants, 1985; Story, 1995; Titze, 2006b) p_s^+ is the incident partial wave pressure arriving from the subglottis and p_e^- is the incident partial wave pressure arriving from the supraglottis. It should be noted that the wave-reflection analogs do not include the exact near-field pressures (Zhao et al., 2002; Zhang et al., 2002), but capture the most important fields for wave propagation in the vocal tract.

Equation (5) is critically important for understanding source–filter interaction because it defines explicitly a coupling parameter, a_g / A^*. Note that when a_g / A^* is small, the flow reduces to

$$u_g = a_g \left[\frac{4(p_s^+ - p_e^-)}{k_t \rho} \right]^{1/2}, \tag{7}$$

which becomes a direct proportion to the glottal area a_g. If a constant subglottal pressure were to be applied to produce the glottal flow, the incident partial pressure wave p_s^+ in the wave-reflection algorithm would be replaced by half of the lung pressure P_L with a +1 reflection coefficient, so that the subglottal pressure would then be $P_s = p_s^+ (1+r) = 2p_s^+ = P_L$ (Titze, 1984; Story, 1995). Furthermore, if the incident supraglottal pressure p_e^- were set to zero, then

$$u_g = a_g \left[\frac{2P_L}{k_t \rho} \right]^{1/2}, \tag{8}$$

which is the asymptotic condition for linear (noninteractive) source–filter coupling. For linear coupling, the relative phase delays between p_s^+ and p_e^- due to wave propagation are responsible for the nonproportionality between u_g and a_g.

The skewing of the flow pulse as a result of an overall inertive reactance produces new harmonic frequencies in the glottal airflow that are not part of the glottal area waveform. This result is not new (Rothenberg, 1981; Fant and Lin, 1987; Koizumi et al., 1985), but most previous analyses have underestimated the amount of skewing because the exact geometry of the epilarynx tube was not known. The magnetic resonance images produced by Story (1995; 2005) have been the key data sets to validate strong source–filter interaction. The uniform tubes in Fig. 1 produce relatively weak interactions. In the middle panels, a sinusoidal a_g waveform is shown on the left and its corresponding single line spectrum is shown on the right. With this sinusoidal area (forced oscillation), glottal flow was produced (bottom left panel) with 0.8 kPa lung pressure in a voice simulator that could be switched between forced oscillation and flow-induced self-sustained oscillation. Equation (5) was combined with a 44 section uniform vocal tract with energy losses and the appropriate radiation impedance was used as previously described (Titze, 2006b). Note that the flow waveform is slightly skewed and has an entire spectrum of frequencies (bottom right). Wave propagation occurred in both subglottal and supraglottal tracts. The source–filter coupling is nonlinear because new frequencies (harmonic distortion frequencies) are created by the vocal tract.

The skewing of the flow pulse guarantees a dominant excitation near glottal closing and raises the energy in the harmonics (Fant, 1986). In the past it has been assumed that the harmonic spectrum of the source comes primarily from vocal fold collision. This may be true for many phonations, but this example shows that vocal fold collision is not essential to produce source harmonics. Nonlinear source–filter coupling can produce a spectrum of source frequencies, with a spectral slope of about −15 dB per octave in this case. Furthermore, the harmonic amplitudes are affected by the reactance curve. The short vertical lines in the top right panel above the reactance curves show the location of the harmonics in the bottom right figure. Note that harmonics 3 and 4 are in negative reactance territory and are depressed slightly in amplitude, relative to harmonics 2 and 5. The simplest explanation for this amplitude depression is that negative
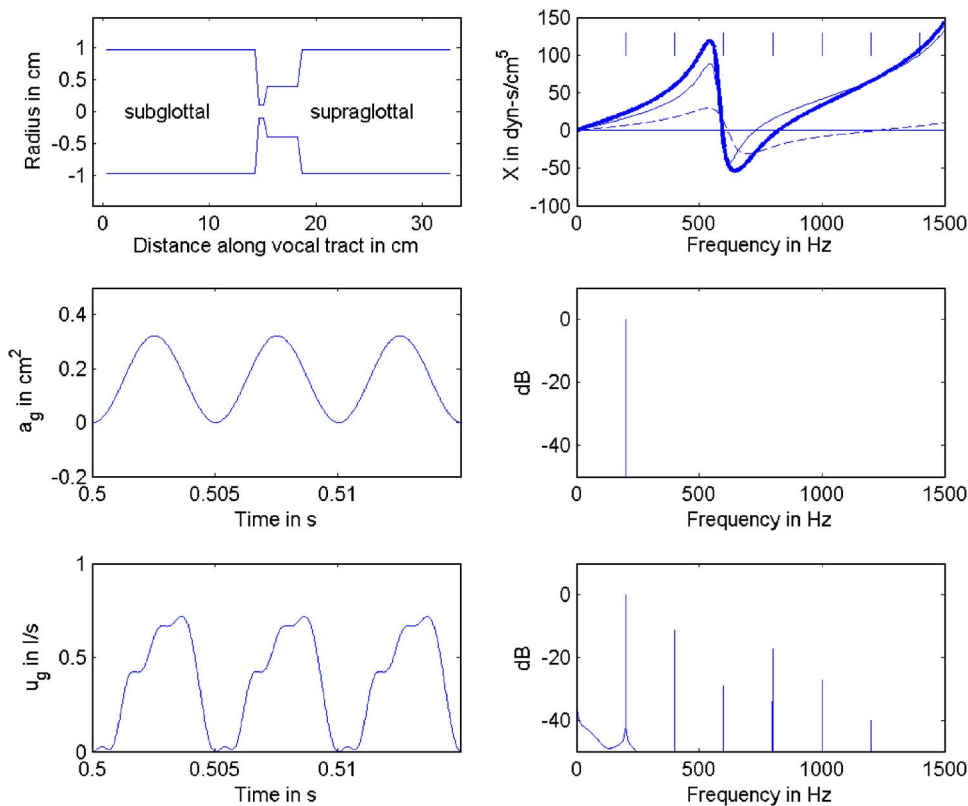
Ingo R. Titze: Nonlinear source–filter theory

FIG. 2. (Color online) Harmonic frequency generation by source–filter generation with stronger coupling (A_e=0.5 cm²); (top left) vocal tract shape; (top right) reactance curves, thin solid line for supraglottal, dashed line for subglottal, and thick solid line for combined; (middle left) sinusoidal glottal area function; (middle right) spectrum of glottal area; (bottom left) glottal flow; (bottom right) spectrum of glottal flow.

(compliant) reactance integrates the downstream flow and builds up an opposing pressure that reduces the flow at a given frequency. If negative reactance were present at all frequencies, the flow pulse would be skewed to the left because more flow would be accepted during glottal opening than glottal closing as the back pressure builds up. With frequency dependent reactance, selective components that would skew the pulse to the left are reduced in amplitude because, overall, the pulse is still skewed to the right. A more detailed discussion would require the consideration of all the phases of the components and how they are affected by variable reactance.

   Figure 2 shows the same set of curves as in Fig. 1, but now the coupling between the source and the filter is increased to a more realistic value. By reducing the epilarynx tube cross-sectional area A_e from 3.0 cm² in Fig. 1 to 0.5 cm² in Fig. 2 (compare upper left graphs), the input impedance to the vocal tract has been increased. This input impedance is scaled by $\rho c / A_e$, the characteristic acoustic impedance of the first section of a tube, where ρ is the density of air and c is the sound velocity. The coupling parameter a_g / A^* in Eq. (5) was correspondingly increased from 0.1 to 0.35 because it contains A_e in the formula for A^* in Eq. (6). The mean glottal area a_g was held constant at 0.15 cm² and the subglottal area was held constant at 3.0 cm². Hence, the glottal source impedance remained the same. Note that the combined reactance curve (thick line, upper right in Fig. 2) now has a net upward shift toward positive (inertive) values. Specifically, over the entire 0–1500 Hz frequency range, negative (compliant) reactance occurs only between 600 and 800 Hz.

   The harmonic frequencies produced in the flow wave form (bottom left panel of Fig. 2) reflect this increase in coupling strength. An apparent "closed phase" is seen, even though there was no glottal closure. As in Fig. 1, the area function a_g was sinusoidal and always remained above zero (no truncation). This brings into question that whole enterprise of inferring vocal fold vibration patterns from inverse-filtered glottal flow, especially in terms of an open phase and a closed phase. Rothenberg and Zahorian (1977) showed that inverse filtering to obtain the glottal area from mouth flow is fundamentally a nonlinear process. Linear prediction cannot accomplish this task.

   Comparing the harmonic spectra (bottom right) across Figs. 1 and 2, we see that the strength of the harmonics is again related to the amount of inertive reactance present at the harmonic frequency. For the F_0 selected here (200 Hz), the third harmonic gained no strength with increased coupling because it still resides in negative (compliant) reactance territory, as in Fig. 1. The second and the fourth harmonics, however, both experience a slight amplitude increase due to higher reactance. In particular, the fourth harmonic, which was in negative territory, now experiences about zero reactance. This stronger fourth harmonic is responsible for the ripple seen in the flow waveform at the bottom left. (Only three ripple cycles are seen on the increasing flow; the fourth ripple cycle is hidden on the downward slope and on the flat portion near zero flow). The spectral slope is now only on the order of −10 dB per octave.

   Because the source spectrum can be affected by vocal tract reactance, and because this reactance is frequency dependent, a spectrogram with an F_0 glide at constant vocal tract shape (identical to the shape shown in Fig. 2, top left) was investigated. This type of a pitch–glide spectrogram was
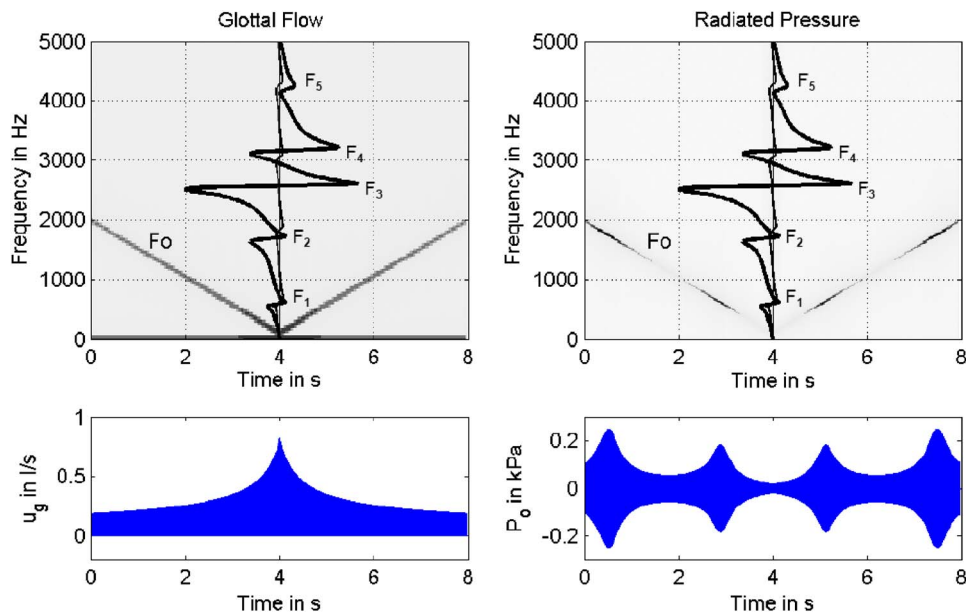
FIG. 3. (Color online) F_0 glide produced with a driven sinusoidal glottal area function with no source–filter interaction; (top left) spectrogram for glottal flow based on sinusoidal glottal area, with reactance curves overlaid vertically; (top right) spectrogram for radiated mouth pressure P_0; (bottom left) amplitude envelope of glottal flow; (bottom right) amplitude envelope of radiated mouth pressure.

analyzed for recordings of human subjects in the companion paper (Titze *et al.*, 2008). Figure 3 shows the noninteractive (linear) case as a control for later interactive cases. No interaction with tissue vibration was allowed because the vibration was still forced with a sinusoidal area rather than self-sustained. At the top left we see a spectrogram of the F_0 glide (straight sloping line), from 2000 to 100 Hz and back to 2000 Hz. The signal used in the spectrogram was the glottal flow u_g, a purely sinusoidal function derived from the sinusoidal area function a_g according to Eq. (8). The amplitude envelope of this sinusoidal flow is shown below the spectrogram on the left. Its peak at 4.0 s comes from the fact that the glottal area function (not shown) was programmed to vary inversely with the square root of frequency to approximate realistic amplitudes of vibration.

The spectrogram (top left panel) also shows the reactance curves of Fig. 2, now superimposed on the center of the spectrogram such that they are displayed vertically from 0 to 5000 Hz. Positive reactance is to the left and negative reactance is to the right of center. The reactance fluctuation in the third and fourth formant region (around 3000 Hz) is so large that the subglottal reactance is barely visible on the same scale. This large fluctuation is attributed to the epilarynx tube, which begins to establish its own characteristic quarter-wave resonance near the third formant at 2500–3000 Hz (Titze and Story, 1997).

Note again that no harmonic distortion frequencies are created by the vocal tract for this noninteractive F_0 glide, the control case. The single frequency component of the flow (F_0, sloping downward from 0.0 to 4.0 s and then upward from 4.0 to 8.0 s) is unaffected by the filter. What is affected by the filter is the oral radiated pressure P_o, shown on the right side of Fig. 3. As expected, P_o increases near 0.6 s when F_0 passes through F_2 and near 3.0 s when F_0 passes through F_1. These increases (and decreases on the other side of a formant) would be perfectly symmetric if the radiation at the mouth were independent of frequency. High frequencies radiate better than low frequencies, however, which pro-

duces a greater P_o at 2000 than at 100 Hz, even though the peak glottal flow was 0.2 vs 0.8 l/s at these extremes.

Figure 4 shows the flow-interactive case, with A_e remaining at 0.5 cm^2. Vocal fold vibration still remained forced (no interaction with tissue vibration). The confirming observation is that a series of extra harmonics ($2F_0$ through $4F_0$) is again created in u_g by vocal tract coupling. These harmonics are reinforced when there is positive reactance (positive reactance curve is left of the vertical center line) rather than at the center of the formants. This is especially noticeable in the second harmonic. [The dependence on reactance is also present in the fundamental, but the spectrogram gray scale was deliberately saturated for F_0 so that the higher harmonics could be seen.] At the formants, where the reactance changes suddenly from positive to slightly negative, the source harmonics are diminished in their amplitudes. This is further evidenced by the envelope of the flow wave form u_g, which is modulated by an uneven treatment of the harmonics by this reactance (note the valleys at 0.8 and 3.0 s in the bottom left panel). The peaks and valleys in the amplitude envelope are unmistakable evidence of nonlinear source–filter coupling because the glottal area and lung pressure were identical to the control case. As a result, the overall peak-to-valley ripple in the P_o waveform (bottom right) is less severe than in the linear case of Fig. 3. The dips in the u_g amplitude at the formants partially cancel the increase in energy transmission through the vocal tract at the formants (Rothenberg, 1987). Thus, the effect of nonlinear formant–harmonic coupling in the glottal flow (level 1 interaction) is to distribute the acoustic energy over the entire spectrum rather than to accentuate it at the center of a formant.

In summary, rather than attempting to resonate several source harmonics by tuning them to the tube resonances, as is done in many musical instrument designs, a vocalist may attempt to reinforce a cluster of several harmonics with favorable reactance. In most cases, as many harmonics as possible are placed on the lower frequency side of a formant (below the resonance frequency). For high-pitched singing,
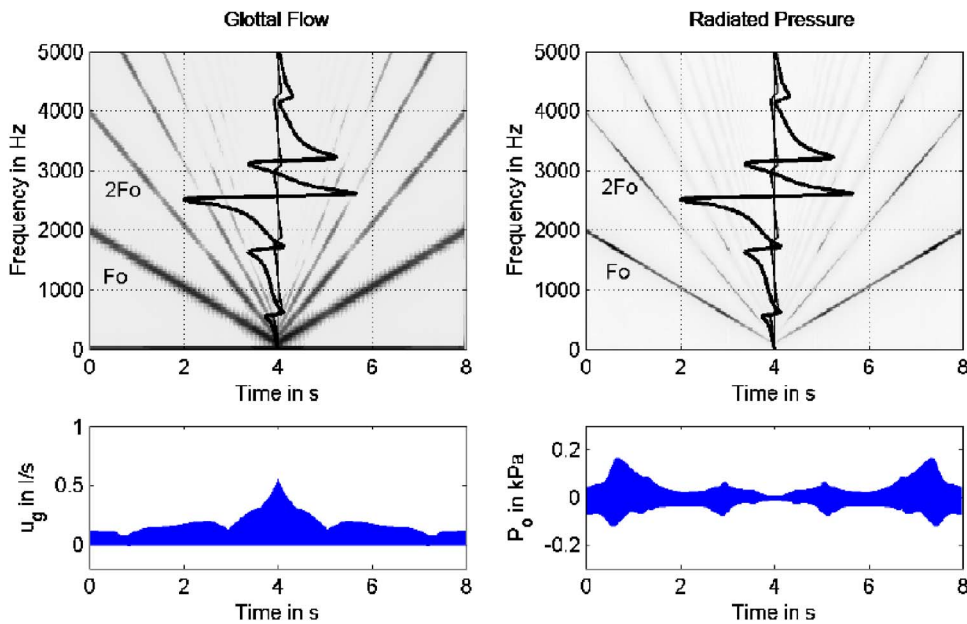
FIG. 4. (Color online) F_0 glide produced with a driven sinusoidal glottal area function with source–filter interaction ($A_e = 0.5$ cm^2); (top left) spectrogram for glottal flow based on sinusoidal glottal area, with reactance curves overlaid vertically; (top right) spectrogram for radiated mouth pressure P_0; (bottom left) amplitude envelope of glottal flow; (bottom right) amplitude envelope of radiated mouth pressure.

this often requires a special vowel, so that at least F_0, $2F_0$, and $3F_0$ can all be reinforced over a reasonable pitch range. An example is shown in Fig. 5, where subglottal reactance is now shown by the thin line and supraglottal reactance by the thick line. Note the ranges of F_0, $2F_0$, and $3F_0$ below the reactance curves at the bottom of Fig. 5. Only the subglottal reactance is negative in the $2F_0$ range, but in the next discussion it will be shown that the combination of a compliant (negative) subglottal reactance and an inertive (positive) reactance provides ideal reinforcement of vocal fold vibration. So, the loss in level 1 interaction may in part be overcome by a gain in level 2 interaction. This example suggests that certain vowels will be favored over other vowels. The utility of this type of interaction is that source and filter frequencies do not need to match exactly, leaving some degree of freedom for articulation and vowel migration with the general goal of strengthening the dominant low-frequency harmonics.
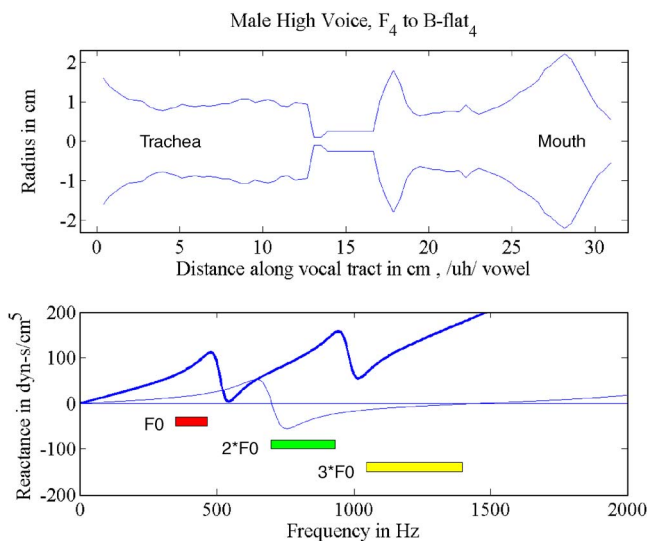

FIG. 5. (Color online) (Top) Vocal tract shape for the vowel /U/; (bottom) reactance curves, thick line supraglottal and thin line subglottal, and favorable ranges for F_0, $2F_0$, and $3F_0$ shown underneath.

B. Level 2 interaction: Mode of vibration dependency on vocal tract reactance

There is another level of interaction, identified as level 2, which involves a change of the vibration pattern of the vocal folds as a result of vocal tract changes. For this level of interaction, subglottal reactance and supraglottal reactance affect vocal fold vibration differently. While supraglottal reactance is generally most favorable when it remains inertive (positive), the subglottal reactance is sometimes more favorable if it is compliant (negative). The complicating factor is the geometry of the vocal folds, or shape of the glottis, in direct analogy to Fletcher's (1993) differentiation of inward, outward, and lateral striking valves that can self-sustain oscillation in a pipe. An added complexity, not discussed by Fletcher, is that vocal folds can propagate a surface wave in their tissue, which changes the shape of the valve dynamically. Based on this surface wave, a relation for the mean intraglottal driving pressure on the vocal fold surface was derived previously (Titze, 1988),

$$P_g = P_s\left(1 - \frac{a_2}{a_1}\right) + P_e \frac{a_2}{a_1},\qquad(9)$$

where P_g is the mean (entry to exit) intraglottal driving pressure, P_s is the subglottal pressure, P_e is the supraglottal (epilarynx tube) input pressure, a_1 is the glottal entry area (lower margin of the vibrating portion of the vocal folds), and a_2 is the glottal exit area (upper margin of the vibrating portion of the vocal folds). In deriving Eq. (9), the pressure profile in the glottis was assumed to follow the Bernoulli energy conservation law, the glottal area was assumed to vary linearly from bottom to top, the acoustic pressure P_e was assumed to be the only pressure recovery at glottal exit, and the transglottal pressure coefficient was set to 1.0 for simplicity. (For details of the full derivation, see Titze (1988), p. 1542.)

As a valve, the vocal folds are closest to the $(+,+)$ case described by Fletcher (1993). In his notation, the valve is
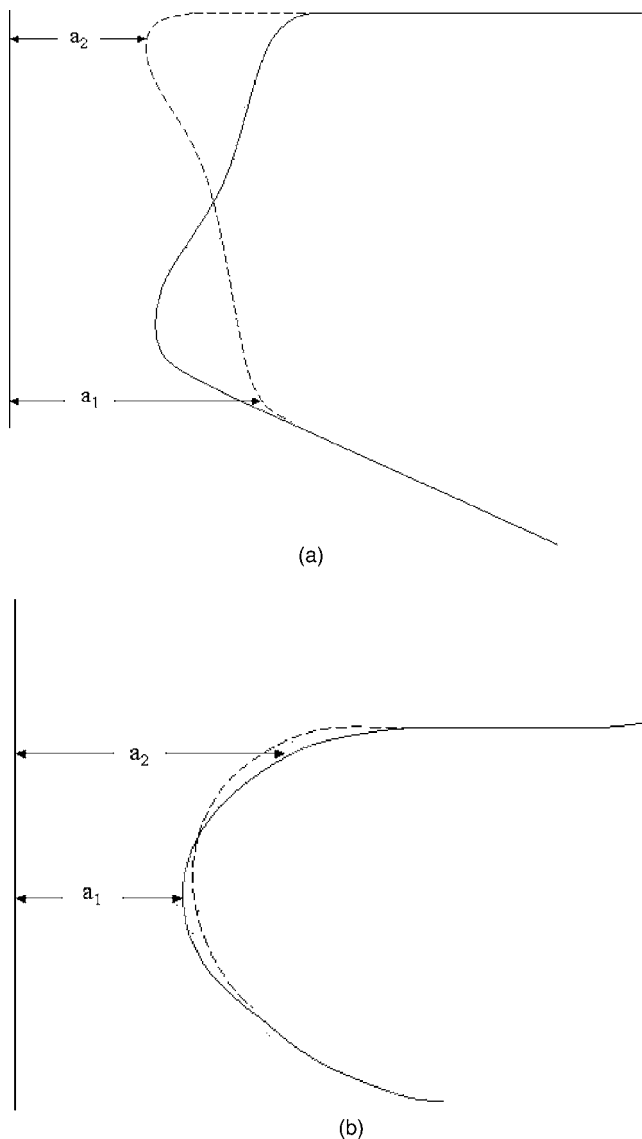
(a)

(b)

FIG. 6. Sketches of right vocal fold tissue displacement from the glottal midplane in coronal view of (a) modal register and (b) falsetto register. After Hirano, 1975.

$(+, +)$ if tissue moves laterally with both increasing subglottal pressure (first $+$) and increasing supraglottal pressure (second $+$). But, as Eq. (9) indicates, intraglottal pressure is greater for a convergent glottis than for a divergent glottis. Tissue surface waves on the vocal folds (i.e., standing waves, or modes of vibration) can be excited by the airflow to produce self-sustained oscillation, even without vocal tract interaction (Titze, 1988). Note that, even if P_s is a constant and P_e is zero in Eq. (9), an alternating push–pull pressure can be created by an a_2/a_1 ratio that is less than 1.0 for lateral movement and greater than 1.0 for medial movement. Vocal tract interaction is therefore an important, but not a necessary, condition for self-sustained vocal fold oscillation.

Figure 6 shows a sketch of how the medial surface of the vocal folds may change during vibration, both in terms of the static and the time varying configuration. The sketch is patterned after Hirano (1975). Two different registers of phonation are identified. In the so-called modal register [Fig. 6(a)], vibration is observed over much of the thickness of the vocal folds, so that the entry area a_1, where vibration begins, is low in the glottis vertically. To the contrary, in the so-called falsetto register [Fig. 6(b)], vibration is confined mainly to the upper portion of the vocal folds, with a_1 being much higher in the glottis. (The point where vibration effectively begins vertically in the glottis has been called the mucosal upheaval point; Yumoto and Kadota, 1998). If we approximate the vibrating portion of the medial surface with a straight line, then the prephonatory shape is both convergent $(a_2 < a_1)$ and divergent $(a_2 > a_1)$ for the modal register but mainly divergent $(a_2 > a_1)$ for the falsetto register. Physiologically, the thyroarytenoid (TA) muscle controls this medial surface shape. When the TA muscle contracts, it thickens and bulges out the lower part of the vocal fold, thereby "squaring up" the glottis and producing modal register. When the TA muscle is relaxed, the bottom of the vocal fold retracts and only the top remains engaged in vibration. The medial surface is more rounded. The vocal ligament is used for adductory positioning and tensing of the vocal fold tissues.

It will now be shown how vocal tract pressures can assist or hinder vibration in these basic two registers. Neglecting vocal tract resistance and steady pressures for the first part of the discussion, basic acoustic theory would predict the vocal tract pressures to be

$$P_s = -I_1 \frac{du}{dt} \quad \text{inertive subglottal tract} \tag{10}$$

$$= P_1 - \frac{1}{C_1} \int u\,dt \quad \text{compliant subglottal tract,} \tag{11}$$

$$P_e = +I_2 \frac{du}{dt} \quad \text{inertive supraglottal tract} \tag{12}$$

$$= P_2 + \frac{1}{C_2} \int u\,dt \quad \text{compliant supraglottal tract,} \tag{13}$$

where P_1 and P_2 are constants of integration. With these relations, the intraglottal pressure of Eq. (9) has four possible forms:

$$P_g = -I_1\left(1 - \frac{a_2}{a_1}\right)\frac{du}{dt} + I_2 \frac{a_2}{a_1}\frac{du}{dt} \quad \text{inertive – inertive} \tag{14}$$

$$= P_1 - \frac{1}{C_1}\left(1 - \frac{a_2}{a_1}\right) \int u\,dt$$
$$\quad + I_2 \frac{a_2}{a_1}\frac{du}{dt} \qquad \text{compliant – inertive,} \tag{15}$$

$$= -I_1\left(1 - \frac{a_2}{a_1}\right)\frac{du}{dt} + P_2$$
$$\quad + \frac{1}{C_2}\frac{a_2}{a_1} \int u\,dt \qquad \text{inertive – compliant,} \tag{16}$$
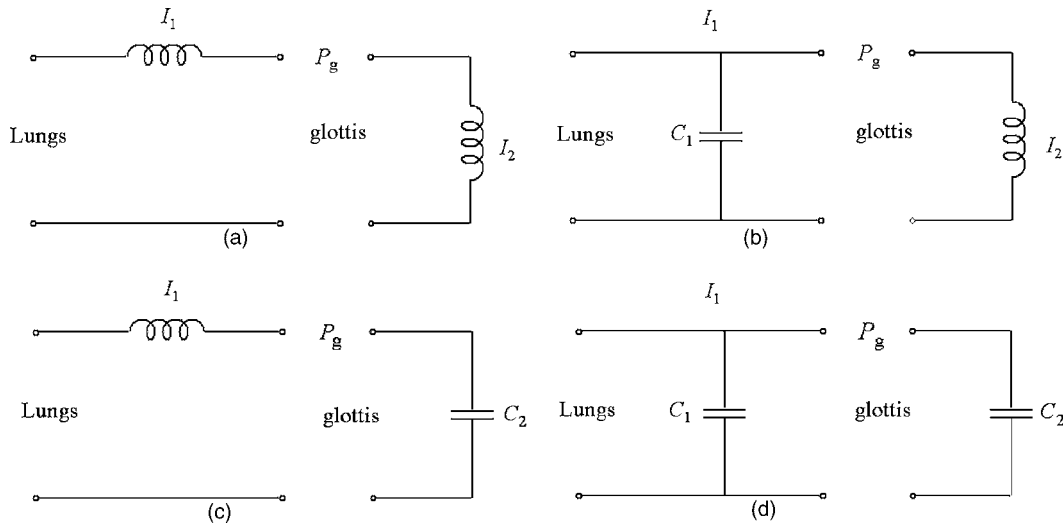
FIG. 7. Acoustic circuit diagrams for subglottal and supraglottal reactance, (a) inertive–inertive, (b) compliant–inertive, (c) inertive–compliant, and (d) compliant–compliant.

$$= P_1 - \frac{1}{C_2}\left(1 - \frac{a_2}{a_1}\right) \int u\,dt + P_2$$

$$+ \frac{1}{C_2}\frac{a_2}{a_1} \int u\,dt \qquad \text{compliant} - \text{compliant}. \qquad (17)$$

These cases are basically the same as the four cases described by Fletcher (1993). An acoustic "circuit" representation of these relations is shown in Fig. 7. Inertance is represented by coils I_1 and I_2 and compliance is represented by parallel plates C_1 and C_2, following the symbolism of electric circuitry for inductance and capacitance, respectively. The driving pressure P_g from the above-presented mathematical expressions is also labeled in Fig. 7. For maximum reinforcement of vocal fold vibration, the driving pressure P_g should provide an alternating push–pull on the vocal fold tissue, a push when the glottis is opening and a pull when the glottis is closing. Thus, when du/dt is positive (flow is increasing during glottal opening), P_g should be positive; when du/dt is negative (flow is decreasing during glottal closing), P_g should be negative.

Consider first the *inertive – inertive* case [Eq. (14) and Fig. 7(a)]. Both coefficients in front of du/dt in Eq. (14) should be positive for the push–pull condition. The only way this can occur is if $a_2/a_1 > 1.0$ over most of the open portion of the glottal cycle, which means the glottis must be mainly divergent. Falsetto register may provide this configuration quite readily, referring to Fig. 6(b). It is known that the top of the vocal folds spread apart slightly in falsetto register. A net divergent glottis at the top created by this spread would appear to be beneficial to both terms in Eq. (14). For modal register, the divergent configuration occurs over a smaller fraction of the glottal cycle, prior to closure. But maximum pressures occur in this fraction of the cycle, yielding strong excitation in a pulse-like manner. Thus, subglottal inertance I_1 generally provides both a help and a hindrance to vocal fold vibration in modal register. On the other hand, supraglottal inertance I_2 always provides the favorable push–pull condition, for both registrations and both glottal configura-

tions. The Flanagan and Landgraf (1968) and Ishizaka and Flanagan (1972) simulations had no subglottal tract. Hence, their vocal tract interaction effects were probably exaggerated. The Zhang et al. (2006a, b) investigations had only a subglottal tract, which could underestimate the overall interaction. The Zañartu et al. (2007) simulations included both a subglottal and a supraglottal tract. Given that their vocal fold model was composed of only a single mass, the results can be considered an excellent correction to the Flanagan and Landgraf (1968) model. In a later section, it will be shown that neglect of the subglottal reactance can be dramatic, even with greater degrees of freedom in tissue movement.

The *compliant–inertive* tract appears to be the most favorable for vocal fold vibration in modal register [Eq. (15) and Fig. 7(b)]. If the glottis is mostly convergent, with divergence occurring only over a small fraction of the cycle prior to closure, the integration of the flow in Eq. (15) produces a steady decrease in intraglottal pressure over the open portion of the glottal cycle (because $a_2 < a_1$). This gradual decrease in pressure (stronger during opening and weaker during closing) adds to the dominant push–pull produced by the inertive supraglottal tract, as described before. Tongue-tip trills (McGowan, 1992) have also been shown to be sustained by an upstream compliant reactance. The dominant compliance in tongue trills is a wall compliance (rather than an air compliance), but the effect is similar.

The *inertive – compliant* tract [Eq. (16) and Fig. 7(c)] is the least favorable for modal register. Supraglottal integration of the flow u raises the intraglottal pressure throughout the open portion of the cycle, creating a greater push during closing than during opening. This is contrary to the desired push–pull condition. In addition, when the glottis is convergent ($a_2/a_1 < 1.0$), the inertive subglottal tract further hinders oscillation, as discussed earlier. But some assistance is possible from subglottal inertance in falsetto register, if $a_2 > a_1$, also as discussed earlier.

Finally, a *compliant – compliant* tract [Eq. (17) and Fig. 7(d)] is also not favorable to modal register, but a little more

J. Acoust. Soc. Am., Vol. 123, No. 5, May 2008

Ingo R. Titze: Nonlinear source–filter theory   2741

so than the inertive–compliant combination just discussed. For convergence, although the gradual pressure reduction from the first integration in Eq. (17) is favorable, the second integration is detrimental. The worst of all situations exists for divergence. The glottis is simply blown apart by a uniformly increasing intraglottal pressure. Thus a *compliant–compliant* vocal tract squelches phonation when the glottis diverges.

To determine how the above-noted driving pressures affect the frequency and amplitude of self-sustained oscillation, it is typical to develop an autonomous differential equation of motion for tissue displacement in terms of flow-dependent driving pressures. Fletcher's (1993) small amplitude analysis of autonomous vibration of simple pressure-controlled valves in gas flows is relevant to vocal fold vibrations, but the alternating convergent–divergent glottis created by surface waves on the tissues were not modeled. Hence, the analysis was basically for a one-mass model (or the x_{10} mode in the abbreviated mode nomenclature of Titze (2006b, Chap. 4). Adachi and Sato's (1996) treatment of two-dimensional lip vibration (transverse and longitudinal to the flow) captured a z_{10} mode in addition to the x_{10} mode, but also did not include the surface wave. Titze's earlier (1988) analysis did include the surface wave, but considered only downstream inertive reactance of the tract (the most typical for low F_0 speech). In Chan and Titze (2006) and in Titze (2006b, Chap. 7), upstream inertive reactance was also considered, but not compliant reactance. Thus, to date nobody has derived an autonomous differential equation that includes both inertive and compliant reactance, upstream and downstream, with the inclusion of a surface wave on the tissue.

Given that Fletcher's (1993) closed-form (analytical) solutions contain both subglottal and supraglottal reactance explicitly, his equations are utilized here to approximate frequency and amplitude changes for autonomous (self-sustained) oscillation. Equations (19) and (20), from Fletcher (1993) p. 2176, were solved with the following parameters (Fletcher's notation) that are relevant for human phonation: $\sigma_1 = \sigma_2 = +1$ (Fletcher's overpressure parameters that define the valve type); $W = 1.0$ cm (vocal fold length); $m = 0.05$ g (vocal fold mass in vibration); $S_1 = S_2 = S_3 = 0.5$ cm^2 (inferior, medial, superior surfaces of vocal folds, respectively); $x_0 = 0.03$ cm (neutral glottal half-width); $\bar{p}_2 = 10\,000$ dyn/cm^2 ≈ 1 kPa (mean subglottal pressure); $\theta = 30° =$ entry and exit angles into glottis; $k = (0.05)\,(2\pi F_0) =$ tissue resonance bandwidth (157 Hz at $F_0 = 500$ Hz); $X_1, X_2 =$ subglottal and supraglottal reactances, respectively (variables).

Both X_1 and X_2 were varied according to the reactance curves of Fig. 2 (top right). They are redrawn for comparison at the top of Fig. 8. Recall that these reactances apply to a uniform 3.0 cm^2 subglottal tract and a uniform 3.0 cm^2 supraglottal tract that has a narrowed 0.5 cm^2 epilarynx tube.

Fletcher's analytical equations allow for direct specification of the natural frequency F_0 of the vocal fold oscillator in a no-load condition. This frequency was varied from 0 to 1500 Hz as a pitch glide to correspond to human ranges of F_0 and to predict the F_0 jumps observed in the data of the companion paper (Titze *et al.*, 2008). F was defined to be the
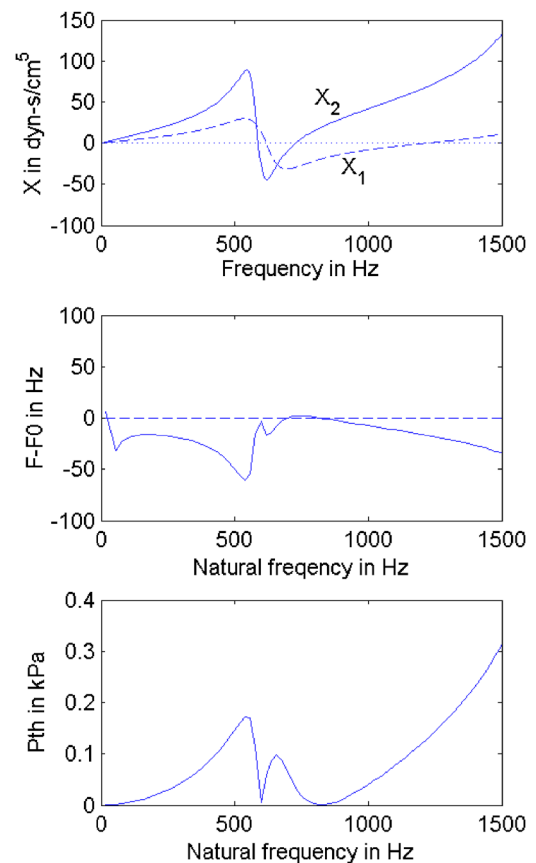

FIG. 8. (Color online) Fletcher's (1993) small oscillation analysis; (top) reactance curves; (middle) the difference between the oscillation frequency F and the no load frequency F_0; (bottom) oscillation threshold pressure P_{th}.

true oscillating frequency for any applied load. Figure 8 (middle graph) shows calculations for $F - F_0$, the difference in oscillation frequency from the no-interaction fundamental frequency. In the bottom graph we see the corresponding oscillation threshold pressure P_{th}.

Consider first some general trends. The oscillation frequency tends to be mostly below the no-interaction resonance frequency F_0. This is because inertive reactance dominates in airways that have a constricted region (the epilarynx tube in this case), creating effectively an increase in the mass of the oscillating system (tissue and air columns collectively). Note that there is a general inverse relation between $F - F_0$ and the supraglottal reactance X_2. The highest frequency is at 750 Hz, where $X_1 = -30$ dyn s/cm^5 (compliant) and $X_2 = 0$. This frequency is slightly greater than F_0. The most notable drop in frequency (-50 Hz) occurs just above 500 Hz, where both X_1 and X_2 are highly positive. The small peak around 600 Hz is for $X_1 = X_2 = 0$, the no-interaction condition for which $F = F_0$.

The threshold pressure P_{th} basically follows the sum of the reactance curves. Aside from a narrow dip near the tube resonances, where $X_1 + X_2 = 0$, the lowest threshold pressure is found in the 750–1000 Hz region, where X_1 is negative (compliant) and X_2 is positive (inertive). Thus, as stated earlier, the *compliant – inertive* acoustic load is the most favorable to vocal fold oscillation. Titze and Sundberg (1992) have shown that every doubling of lung pressure above threshold raises the source intensity by about 6 dB. In addi-
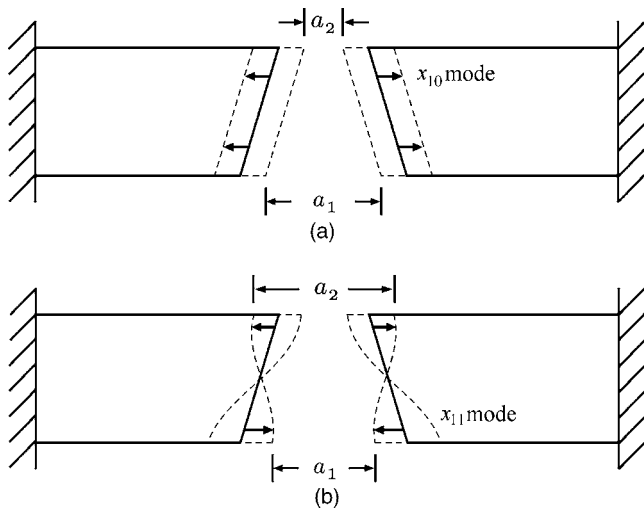
FIG. 9. Sketches of a convergent glottis with two vibrational modes, (a) the x_{10} mode and (b) the x_{11} mode.

tion, Alipour *et al.* (2001) showed that an approximate 6 dB increase in intensity was obtained in an excised larynx when a vocal tract was added by way of a physical tube that lowered the threshold pressure. Given that the threshold pressure in Fig. 8 varies from below 0.1 to above 0.3 kPa, and given that typical lung pressures for speech range between 0.5 and 1.0 kPa, it is possible that more than one doubling of P_{th} or 6–12dB in source intensity could be realized in the lower threshold regions with normal lung pressures. But a word of caution is in order. The Fletcher (1993) equations used for the Fig. 8 calculations did not contain vocal tract losses. Hence, the P_{th} fluctuations are probably overestimated. Nevertheless, significant changes in source energy are likely with only a few tenths of a kPa reduction in P_{th}.

In any vocal fold model other than the one-mass model, the a_2/a_1 ratio is variable throughout the glottal cycle. The amount of dynamic a_2/a_1 variation can be explained in terms of two dominant modes of vibration of vocal fold tissues (Fig. 9). For prephonatory (static) convergence, an x_{10} mode has no vertical variation in tissue displacement, as shown in Fig. 9(a). It produces less variation in the a_2/a_1 ratio than an x_{11} mode, which has a 180° phase difference between top and bottom displacement, as shown in Fig. 9(b). [For a complete description of tissue modes and their nomenclature, see Titze (2006b), Chap. 4]. The a_2/a_1 ratio for the x_{10} mode gradually increases and decreases over the open portion of the glottal cycle, but stays between 0.0 and 1.0. For the x_{11} mode, however, both convergence and divergence can be experienced over the cycle. Thus, the x_{11} mode is less dependent on vocal tract interaction than the x_{10} mode because the pushes and pulls from vocal tract pressures tend to cancel each other over the glottal cycle. This conclusion agrees with what Flanagan and Landgraf (1968) observed with a one-mass model, which only supports an x_{10} mode, and what Zhang *et al.* (2006a, b) observed on a physical model. It is also the reason why Zañartu *et al.* (2007) focused their recent analysis on the one-mass model. All models with a dominant x_{10} mode are highly affected by vocal tract reactance. In the two-mass model of Ishizaka and Flanagan (1972), the x_{11}

mode was present and vocal tract interaction was reduced. Thus, the percentage of x_{11} mode excitation (relative to x_{10}) serves as a decoupler of vocal tract interaction, in direct opposition to the narrowing of the epilarynx tube. It offers a vocal tract-independence mechanism of self-sustained oscillation (Titze, 1988; Lucero and Koenig, 2007; Chan and Titze, 2006; Jiang and Tao, 2007). Speakers and singers who adjust their source and filter for linear coupling rely heavily on this mechanism of self-sustained oscillation.

III. SOURCE–FILTER INTERACTIONS WITH COMPUTER SIMULATIONS

Because of the many simplifications and limitations imposed by the low-dimensional vocal fold models and analytical treatments, it is important to balance the above-presented results with computer simulation models that are capable of handling all levels of interaction and many degrees of freedom in tissue movement.

A. Methods

An $L \times M \times N$ point-mass model of the vocal folds was used for simulation of flow-induced, self-sustained oscillation. The details of this model are beyond the scope of any single article, but are well documented in Titze (2006b, Chap. 4). L is the number of masses in the medio-lateral direction (7 in this simulation), M is the number of masses in the anterior–posterior direction (5 in this simulation), and N is the number of masses in the inferio–superior direction (5 in this simulation). Thus, the model had 175 point masses, each with two degrees of freedom (horizontal and vertical). Tissue properties were defined with a fiber-gel construct, where the fibers carried the nonlinear stress–strain characteristics of muscle, ligament, and mucosa, and the gel properties were defined with Young's moduli, shear moduli, and Poisson's ratios (Titze, 2006b; Chaps. 2–4). Aerodynamic pressures and glottal flow were calculated with a modified Bernoulli equation that included flow separation and jet formulation by rule (Titze, 2006b, Chap. 5). All vocal tract pressures were computed with the well-known wave-reflection analog (Liljencrants, 1985; Story *et al.*, 1996; Titze, 2006b, Chap. 6).

Based on the foregoing autonomous differential equation analyses, it was important to include both a subglottal and a supraglottal vocal tract to assess the individual and combined interaction effects. As in the previous section on level 1 interaction with forced oscillation, the subglottal system had 36 cylindrical sections, each 0.398 cm long, for a total length of 14.33 cm. The supraglottal system had 44 cylindrical sections, each of the same length, for a total of 17.51 cm. The first eight sections of the supraglottal tract again constituted the epilarynx tube, the diameter of which was kept uniform so that a single parameter (A_e) could be varied for coupling strength. The sampling frequency was 44.1 kHz. The medial surface of the vocal folds was nearly flat, with a slight convergence in the inferio–superior direction and a similar convergence (tapering) in the posterio–
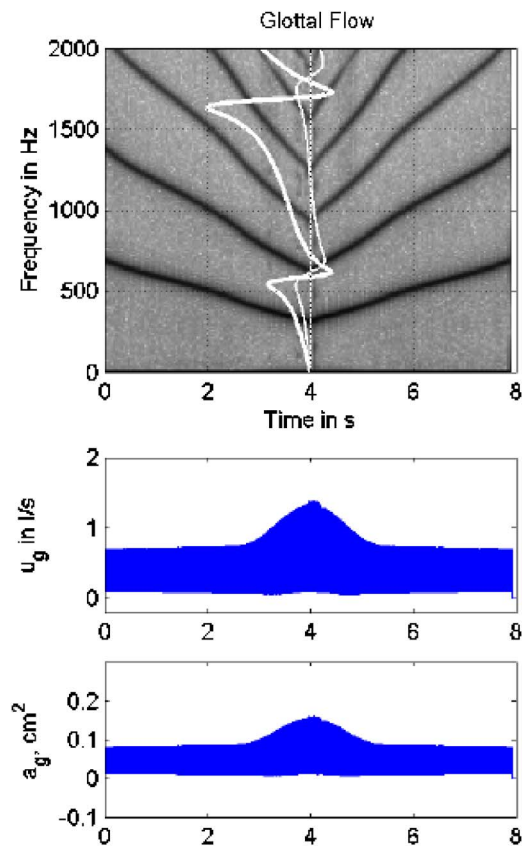
FIG. 10. (Color online) Simulation of downward and upward F_0 glide produced with a 175 point-mass self-oscillating biomechanical model of the vocal folds with no vocal tract interaction; (top) spectrogram, with reactance curves superimposed vertically, thick white line supraglottal and thin white line subglottal; (middle) glottal flow envelope; (bottom) glottal area envelope.

anterior direction. Dynamically, the shape of the glottis assumed a variety of mode configurations due to surface wave propagation.

Control parameters for this model were lung pressure P_L and simulated muscle activations of the intrinsic laryngeal muscles of the larynx: cricothyroid (CT), thyroarytenoid (TA), lateral cricoarytenoid (LCA), posterior cricoarytenoid (PCA), and interarytenoid (IA). These muscle activations (ranging from 0.0 for no activation to 1.0 for 100% activation) positioned the vocal folds (Titze, 2006b, Chap. 3; Titze and Hunter, 2007) and determined all of the tissue properties. Specific values will be given in the following sections.

B. Results

1. Simulation 1

A no-interaction case was first investigated as a control case. A high–low–high pitch glide was simulated under self-sustained oscillation, following the pattern produced earlier for level 1 interaction and produced by human subjects in the companion paper. Figure 10 shows a spectrogram for the glottal flow waveform, with both the subglottal reactance (thin white solid line) and the supraglottal reactance (thick white line) superimposed vertically. These reactances are the same as in previous figures (Figs. 2–4 and 8) for the uniform tubes with a 0.5 cm^2 epilarynx tube. The zero line for the

reactance curves is set at time=4 s, and the scaling of the reactance magnitude can be obtained from Fig. 2. Underneath the spectrogram, the amplitude envelope for the glottal flow u_g is shown in the middle panel and the amplitude envelope for the glottal area a_g is shown in the bottom panel. The following parameters of the model were held constant in the simulations: P_L=1.5 kPa, LCA activity=31%, IA activity =30%, PCA activity=0.0%, vowel=uniform tube with narrowed epilarynx tube as in Fig. 2, with no nasal coupling. CT activity was varied from 90-0-90% and TA activity from 0-10-0%. Details of how these activities affect laryngeal posturing in a speech-like gesture can be found in Titze and Hunter (2007).

To simulate no vocal tract interaction, the pressures in the glottis were programmed as described previously in the forced oscillation case (level 1 interaction). Waves propagated through the vocal tract and were radiated from the mouth, but this propagation produced no load on the glottis because the incident pressures (traveling waves) were nullified by programming for glottal flow and pressure calculation. Only aerodynamic pressures and flows were retained. Note that this caused glottal flow and glottal area envelopes to be proportional, as in the forced oscillation case. The glottal area spectrogram (not shown) was identical to the glottal flow spectrogram, indicating that there was no interaction. Both spectrograms showed harmonics; however, these resulted from vocal fold collision in this case instead of flow pulse skewing. (It is difficult to obtain a perfectly sinusoidal area with no collision with self-sustained oscillation.) The broad peak in the center of the u_g and a_g envelopes was due to greater laxness of the tissue, which caused greater vibrational amplitude when F_0 was low. The fundamental frequency (F_0) in the glide ranged from 700 to 330 Hz. Note that F_0 passed through the positive peak of the reactance curve at about 1.0 s and again at about 7.0 s, but this had no effect on the flow or area waveforms, indicating linear source–filter coupling for this control case.

2. Simulation 2

For the next simulation, subglottal interaction alone was investigated. To uncouple the supraglottal tract, the supraglottal pressure was set to zero in the program for the purpose of glottal flow and intraglottal pressure calculation, but not for wave propagation. Figure 11 shows the result. Note first that the overall signal strength is a little less than for no interaction, judged by height of the a_g and u_g envelopes, and there is not an exact proportionality between the glottal area and glottal flow. More important, new frequencies have been created due to bifurcations in tissue movement, as seen in the spectrogram on top. At 0.8 s, a period-3 subharmonic occurs, which changes to a period-5 subharmonic at 1.6 s. The timing of the period-3 bifurcation agrees with F_0 being in the maximum negative reactance dip and the period-5 bifurcation occurs when $2F_0$ enters the negative reactance region. Much needs to be understood about the nature and onset of these subharmonic bifurcations, but this is a topic for subsequent research. Suffice it to say here that individual harmonics passing through rapidly changing reactance regions can destabilize the vibration regimes, as the foregoing analysis
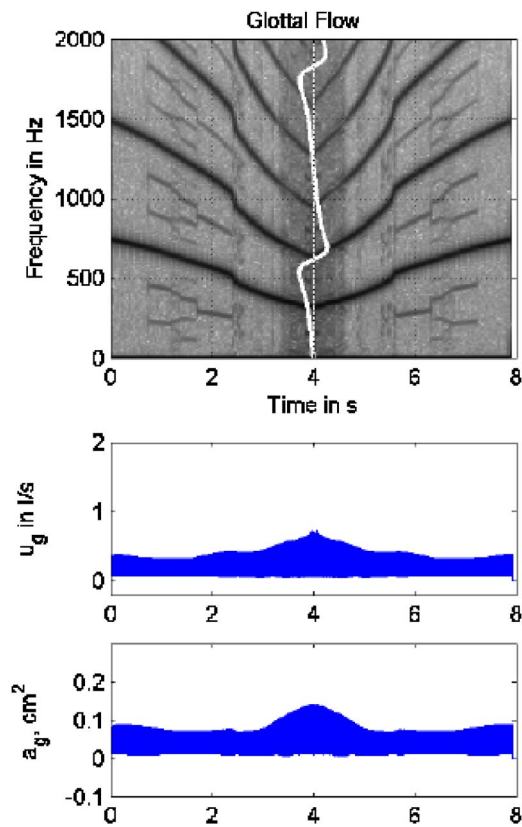
FIG. 11. (Color online) Simulation of downward and upward F_0 glide produced with a 175 point-mass self-oscillating biomechanical model of the vocal folds with subglottal interaction only; (top) spectrogram, with subglottal reactance superimposed vertically with white line; (middle) glottal flow envelope; (bottom) glottal area envelope.

FIG. 12. (Color online) Simulation of downward and upward F_0 glide produced with a 175 point-mass self-oscillating biomechanical model of the vocal folds with supraglottal interaction only; (top) spectrogram, with supraglottal reactance superimposed vertically with white line; (middle) glottal flow envelope; (bottom) glottal area envelope.

demonstrated. Our quantitative interest here is in F_0 drops, because they are the most predictable instabilities and the most prevalent in our companion paper on human phonation.

An F_0 drop of about 50 Hz occurs at 2.4 s. This F_0 drop, predicted well by the foregoing small oscillation analysis with Fletcher's (1993) model (Fig. 8) and by Zhang *et al.* (2006a,b) on a physical laboratory model that had only subglottal interaction, was a "correction" from a higher F_0 that prevailed in the compliant reactance region. The starting F_0 was 750 Hz rather than the no-interaction F_0 of 700 Hz seen in Fig. 10. Compliance basically adds stiffness to the interactive vibrating system, thereby raising F_0, whereas inertance adds mass, thereby lowering F_0. An interesting observation is that the 50 Hz pitch jump occurred below the positive peak of the reactance curve. This may be related to the fact that the second harmonic $2F_0$ entered the compliant region at about 2.4 s, which could have delayed the jump by adding some stiffness to the system before inertive reactance took over. At 3.2 s in Fig. 11, some perturbed vibration occurred, as indicated by the darker (noisier) background. This aperiodic vibration is related to an impedance mismatch instability, i.e., the source impedance became lower, on average, than the vocal tract input impedance, and highly variable due to oscillation. The F_0 upglide showed a slight asymmetry in the duration of another brief noisy regime at 5.6 s, indicating a small hysteresis effect.
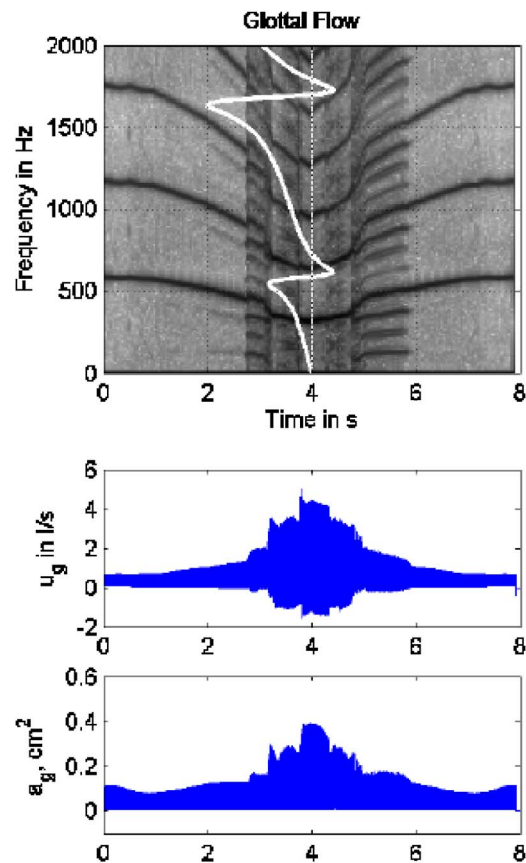
3. Simulation 3

The next simulation was for supraglottal interaction alone. Here subglottal pressure was set to the lung pressure, but supraglottal traveling waves were kept intact. Results are shown in Fig. 12. The waveform envelopes show that this case has the greatest overall signal energy. (The scale of u_g has been increased by a factor of 3 and the scale of a_g by a factor of 2 with respect to Figs. 10 and 11). A weak period-4 bifurcation occurred at 2.0 s, followed by a stronger one at 2.7 s. F_0 and $2F_0$ both went from higher to lower inertive reactance in this bifurcation region. With regard to fundamental frequency changes, note that F_0 started lower than in the noninteractive case, 570 Hz instead of 700 Hz. This "pulling down" of F_0 toward F_1, even when all muscle activities were the same as before, resulted from the inertive reactance of the vocal tract, which (as stated earlier) adds mass to the interactive oscillating system. A very small F_0 rise (10–20 Hz) is then seen at the beginning of the strong period-4 bifurcation, correlated with diminished reactance for both $2F_0$. This small rise, together with the subharmonic regime, delays the eventual larger drop in F_0 of about 100 Hz in positive reactance territory. Finally, the vocal fold vibrational amplitude became disproportionably large and slightly unstable in the 3.0–5.0 s region. A hysteresis effect was also seen in that the bifurcations were delayed on the upslope of the glide. Given that these bifurcations did not appear in the
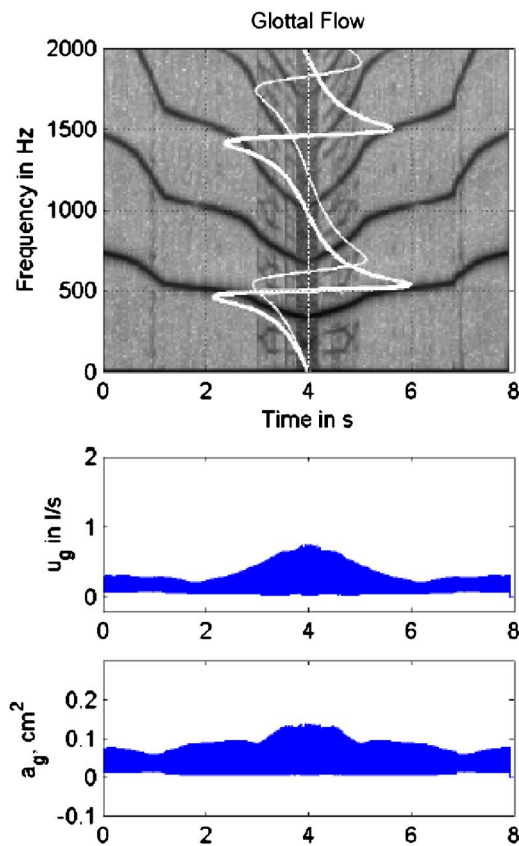
FIG. 13. (Color online) Simulation of downward and upward F_0 glide produced with a 175 point-mass self-oscillating biomechanical model of the vocal folds with both subglottal and supraglottal interaction, $A_e = 3.0$ cm^2; (top) spectrogram, with subglottal reactance (thin white line) and supraglottal reactance (thick white line) superimposed; (middle) glottal flow envelope; (bottom) glottal area envelope.
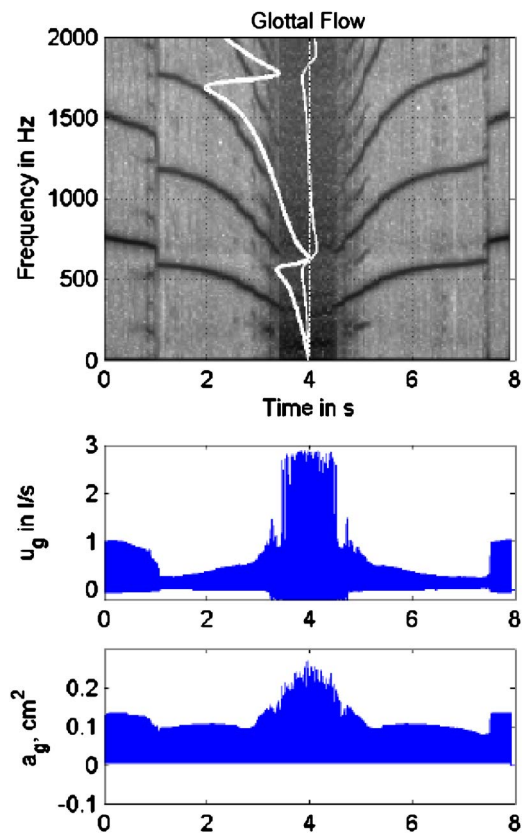

FIG. 14. (Color online) Simulation of downward and upward F_0 glide produced with a 175 point-mass self-oscillating biomechanical model of the vocal folds with both subglottal and supraglottal interaction, $A_e = 0.2$ cm^2; (top) spectrogram, with subglottal reactance (thin white line) and supraglottal reactance (thick white line) superimposed; (middle) glottal flow envelope; (bottom) glottal area envelope.

noninteractive case (Fig. 10), we conclude that they are the result of supraglottal source–filter interaction.

4. Simulation 4

The next simulation was with combined subglottal and supraglottal interactions. Thus, it represented a realistic (non-restricted) situation. The degree of supraglottal interaction was controlled by two separate values of the cross-sectional area of the epilarynx tube A_e. For the lowest degree of interaction ($A_e = 3.0$ cm^2) the epilarynx tube diameter was equal to that of the uniform tube (recall Fig. 1, top left). Figure 13 shows the simulation results for this configuration. Both subglottal and supraglottal reactance curves are shown in white on the spectrogram, supraglottal being the thicker line. Again the scales are arbitrary, but the relative magnitudes between subglottal and supraglottal reactance are correct. There were two F_0 drops, one at about 0.8 s, where F_0 entered a region of rapidly changing overall F_1 reactance and $2F_0$ entered a region of positive supraglottal reactance, and another at drop about 3.0 s, where F_0 entered the region of positive supraglottal F_1 reactance. Also, $2F_0$ entered the *compliant–compliant* region at the second drop, and a period-3 bifurcation occurred when $2F_0$ entered this region, similar to what was seen in Fig. 12. Some evidence of chaotic vibration was seen near 4.0 s, where the lowest F_0 and the highest vibration amplitude occurred. The overall signal strength was slightly

less than that of the noninteractive case (the scale for the signal envelopes was set back to that of Figs. 10 and 11).

5. Simulation 5

Figure 14 shows results for a severe case of epilaryngeal constriction ($A_e = 0.2$ cm^2). The overall signal strength was much higher (scale on the glottal flow envelope was changed to 3.0 l/s), and stronger bifurcations occurred. A 100 Hz F_0 jump occurred at 1.0 s, which was preceded by a period-4 subharmonic, again as F_0 entered the rapidly changing reactance region. When $2F_0$ entered the minimum reactance regions, around 3.0 s, destabilization occurred in the vibration, as evidenced by further bifurcations. Strong hysteresis effects were also evident in that reverse bifurcations occurred at higher frequencies. Of particular significance is the fact that vibrational amplitude increased and decreased sharply (and irregularly) near the lowest F_0 (middle graph), while the flow amplitude reached a plateau (bottom graph). This plateau is attributed to the fact that the input impedance to the vocal tract became higher than the glottal impedance for $A_e = 0.2$ cm^2, thus limiting the flow.

6. Changes in energy levels

Table I shows numerical results for a selected group of dynamical variables at all levels of interaction discussed earlier. (An intermediate case for $A_e = 0.5$ cm^2 was added to

TABLE I. Output quantities for various degrees of interaction.

	Mean glottal area (cm^2)	Mean glottal flow (l/s)	Aerodynamic power (w)	Radiated power (w)	Efficiency (%)
No interaction	0.0381	0.3284	0.4825	0.0117	2.42
Subglottal only	0.0336	0.1835	0.2264	0.0033	1.44
Supraglottal only (A_e=0.5 cm^2)	0.0579	0.5014	0.7408	0.0523	7.07
Subglottal and supraglottal (A_e=3.0 cm^2)	0.0347	0.1717	0.2226	0.0014	0.62
Subglottal and supraglottal (A_e=0.5 cm^2)	0.0405	0.1860	0.2228	0.0035	1.59
Subglottal and supraglottal (A_e=0.2 cm^2)	0.0534	0.3200	0.2604	0.0143	5.49

show the progression with increased nonlinear coupling.) The output quantities for comparison are mean glottal area, mean glottal flow, aerodynamic power at the glottis, acoustic power radiated at the mouth, and glottal efficiency (the ratio of radiated power to aerodynamic power). The numbers represent time averages over the entire 8.0 s pitch glide. The "no interaction case" (top row) produced self-sustained oscillation by excitation of the x_{11} mode with aerodynamic pressures only. It serves as the control case. Consider first the "subglottal only" interaction case (row 2). In comparison to "no interaction," every dynamic quantity was reduced in magnitude. This is consistent with the expectation that subglottal inertance generally hinders vocal fold vibration, and such inertance was present over most of the F_0 glide. To the contrary, for the supraglottal only interaction case (row 3), every dynamic variable was increased, again consistent with previous predictions and discussions. Most variables increased by a factor of about 2 with supraglottal interaction only, but acoustic radiated power increased by a factor of more than 4. Glottal efficiency (the ratio of acoustic radiated power to glottal aerodynamic power) increased by a factor of about 3. Thus, supraglottal interaction alone would be a highly favored condition. Unfortunately, the trachea is always present in human phonation. Its effect cannot be removed, but perhaps altered for better reinforcement of vocal fold vibration by larynx lowering or raising.

With both subglottal and supraglottal interaction (rows 4–6), the results were dependent on the degree of interaction, which was controlled by the parameter A_e, the epilarynx tube area. For A_e=3.0 cm^2, the same as the area of the rest of the uniform tube, the radiated acoustic power was only 0.0014 W, nearly an order of magnitude less than for no interaction. This suggests that interaction is not necessarily an advantage. If the impedances are not well matched, more power can be absorbed internally in the system (Titze, 2002). Efficiency was only 0.62% compared to 2.43% for no interaction. Mean glottal flow was about half, as was the aerodynamic power. Mean glottal area remained about the same.

As the degree of interaction increased with a decrease in the epilarynx tube area A_e (rows 5 and 6), all dynamical variables increased. For the greatest interaction, A_e =0.2 cm^2, all quantities except aerodynamic power were greater than those for the "no interaction" case. This suggests a double advantage, getting more radiated power for less

aerodynamic power used. Hence, the glottal efficiency had doubled, with the same lung pressure. Comparing mild interaction (A_e=3.0 cm^2) to strong interaction (A_e=0.2 cm^2), all variables increased categorically. The radiated acoustic power increased by a factor of 10, as did the glottal efficiency. We made no attempt to optimize the output variables by choosing the best value of A_e for the given glottal conditions, but we expect that such an optimization would yield an even higher efficiency.

IV. DISCUSSION AND CONCLUSIONS

Source–filter interaction has been divided into two levels. Level 1 is the interaction of glottal airflow with acoustic vocal tract pressures, even if vocal fold vibration is undisturbed. The interaction parameter is the mean glottal area divided by the effective (parallel combination) tube area of the subglottis and supraglottis. For constant adduction of the vocal folds and constant large tracheal diameters, the coupling parameter becomes the cross-sectional area of the epilarynx tube. Level 1 interaction produces harmonic distortion frequencies that contribute to the source spectrum. This interaction is present in all speech and singing, male and female. It has been described for nearly three decades, but generally underestimated in magnitude because details of the lower vocal tract were unknown. Level 1 interaction contributes to the spectral slope and the spectral ripple in the glottal sound source, even when the spectrum is purely harmonic and no bifurcations in vocal fold vibration occur. The supraglottal and subglottal impedances are additive for this interaction. If both impedances are inertive (positive), a maximum skewing of the flow pulse is achieved, which increases the maximum flow declination rate and thereby vocal intensity. Individual harmonics can be enhanced or suppressed by frequency-dependent reactances that change from positive to negative.

An interesting discovery was made in this investigation with regard to level 1 interaction. The entire spectrum of source frequencies can (theoretically) be produced without vocal fold collision. This finding could have an impact on voice therapy, particularly for vocal fold pathologies resulting from excessive tissue collision stress. With a sinusoidally varying glottal area and no vocal fold contact, a

−12 dB/octave spectral slope was shown here to be achievable with an epilarynx tube cross-sectional area of 0.5 cm².

Level 2 interaction is realized more in high F_0 productions for which the dominant harmonics ($F_0, 2F_0, 3F_0, \ldots$) are near the formants. Frequency jumps and a variety of new source frequencies or instabilities can be produced, including subharmonics and non-random noise. The instabilities occur mostly when one of the dominant harmonics encounters sudden changes in reactance, destabilizing the modes of vibration of the tissue that are affected differently by reactance. The control parameter for these phenomena is the same as for level 1 interaction. Computer simulations with a high-dimensional model showed that vocal efficiency, the ratio of radiated acoustic power to aerodynamic power, can increase by an order of magnitude when the epilarynx tube area is narrowed from 3.0 to 0.2 cm², but this narrowing also produced greater instabilities when dominant harmonics were in unfavorable reactance regions (i.e., near formants).

A thick and pliable mucosal layer on the vocal folds can lead to self-sustained oscillation without much reliance on vocal tract reactance. A parameter for this is the strength of the x_{11} mode of vibration (characterized by large vertical phase differences) relative to the x_{10} mode. Some vocalists may have the choice to operate in either a nearly linear region, with maximum harmonic stability, or in a nonlinear region with greater output power and greater efficiency but at the expense of less harmonic stability.

In the companion paper (Titze *et al.*, 2008), the most frequently occurring instabilities in human subjects were F_0 jumps, in magnitude on the order of 30–40 Hz. But not all subjects exhibited these instabilities, suggesting that nonlinear interaction varies across subjects, and perhaps even within subjects for repeated vocalizations. The theoretical analyses here predicted the F_0 jumps, both with analytical treatments and with simulation. They are mostly triggered when F_0 passes through F_1, but occasionally when $2F_0$ passes through F_1, as both theory and measurement showed. Larger F_0 jumps tend to occur with greater coupling (i.e., a narrower epilarynx tube).

For the modal voice register, used largely in speech at relatively low F_0, the ideal vocal tract load for self-sustained oscillation would be subglottal compliance and supraglottal inertance. The x_{10} mode would get maximum reinforcement. Unfortunately, this combination does not exist at low F_0 in the human voice. Both reactances tend to be inertive because the trachea and the supraglottal tract are roughly of equal length. Although this *inertive–inertive* combination has been shown to be less favorable for self-sustained oscillation than the *compliant–inertive* combination, flow pulse skewing (level 1 interaction) benefits from dual inertive reactance. Hence the two effects are offsetting. The worst combination for self-sustained oscillation in the modal register is an inertive subglottal tract and a compliant supraglottal tract. This combination can occur in speech for low F_1 vowels such as /i/ and /u/. For example, if $F_0 = 300$ Hz, $F_1 = 250$ Hz, and $F_1^1 = 600$ Hz (first subglottal formant), then the *inertive–compliant* condition exists. If interaction is high (narrow epilarynx tube), the register can flip from modal to falsetto, which is more sustainable with this acoustic load. Females

may have cultivated a mixed register for speech to avoid this instability, given that a 300 Hz fundamental frequency is well within their speaking range. The companion paper shows the females exhibit fewer instabilities on a pitch glide than males, even though the likelihood of $F_0 - F_1$ crossing is greater.

A great amount of future work is needed to determine the extent to which combinations of subglottal and supraglottal reactance can be exploited, especially in high-pitched speech, loud speech, and singing on a variety of vowels. In this paper, all discussions pertained to a neutral-shaped vocal tract. Detailed interaction effects need to be developed for all vowels and consonants. Finally, although humans do not have much control over tracheal diameters, tracheal lengths can be changed with larynx raising and lowering. Perhaps the subglottal entry configuration can also be changed for more compliance, and the soft wall construct of the trachea (the portion in contact with the esophagus) may be useful for introducing subglottal compliance.

[1]Reactance is the energy-storing part of impedance, in contrast to resistance, which is the energy-dissipating part. These two parts of impedance are written as a complex number (real and imaginary part), with reactance being the imaginary part of the impedance. Positive reactance is labeled inertive (because the acoustic flow lags behind the pressure in phase) and negative reactance is labeled compliant (because the acoustic flow leads the pressure in phase).

Adachi, S., and Sato, M. (**1996**). "Trumpet sound simulation using a two-dimensional lip vibration model," J. Acoust. Soc. Am. **99**, 1200–1209.

Alipour, F., Montequin, D., and Tayama, N. (**2001**). "Aerodynamic profiles of a hemilarynx with a vocal tract," Ann. Otol. Rhinol. Laryngol. **110**, 550–555.

Alipour, F., and Scherer, R. C. (**2007**). "On pressure-frequency relations in the excised larynx," J. Acoust. Soc. Am. **122**, 2296–2305.

Appelman, D. R. (**1967**). *The Science of Vocal Pedagogy: Theory and Application* (Indiana University Press, Bloomington, IN).

Atal, B. S., and Schroeder, M. R. (**1978**). "Linear prediction analysis of speech based on a pole-zero representation," J. Acoust. Soc. Am. **64**, 1310–1318.

Chan, R., and Titze, I. R. (**2006**). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362.

Chiba, T., and Kajiyama, M. (**1958**). *The Vowel and Its Nature and Structure* (Tokyo-Kaisenikan, Tokyo).

Coffin, B. (**1987**). *Coffin's Sounds of Singing: Principles and Applications of Vocal Techniques with Chromatic Vowel Chart*, 2nd ed. (The Scarecrow Press, Metuchen, NJ).

Fant, G. (**1960**). *The Acoustic Theory of Speech Production* (Moulton, The Hague).

Fant, G. (**1986**). "Glottal flow: Models and interaction," J. Phonetics **14**, 393–399.

Fant, G., and Lin, Q. (**1987**). "Glottal voice source—Vocal tract acoustic interaction," Q. Prog. Status Rep. STL-QPSR **4**, 13–27.

Fant, G., Linjencrants, J., and Lin, Q. (**1985**). "A four-parameter model of glottal flow," STL Q. Prog. Status Rep. **4**, 1–13.

Flanagan, J. L. (**1968**). "Source-system interaction in the vocal tract," Ann. N.Y. Acad. Sci. **155**, 9–17.

Flanagan, J. L. (**1972**). *Speech Analysis, Synthesis, and Perception* (Springer, New York).

Flanagan, J., and Landgraf, L. L. (**1968**). "Self-oscillating source for vocal tract synthesizers," IEEE Trans. Audio. Electroacoust. **AU-16**, 57–64.

Fletcher, N. H. (**1993**). "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180.

Fulcher, L. P., Scherer, R. C., Zhai, G., and Zhu, Z. (**2006**). "Analytic representation of volume flow as a function of geometry and pressure in a static physical model of the glottis," J. Voice **20**, 489–512.

Hatzikirou, H., Fitch, W. T. S., and Herzel, H. (**2006**). "Voice instabilities due to source-tract interactions," Acta. Acust. Acust. **92**, 468–475.

Hirano, M. (**1975**). "Phonosurgery: Basic and clinical investigations," Otol. (Fukuoka) **21**, 239–440.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced source sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268.

Jiang, J., and Tao, C. (**2007**). "The minimum glottal airflow to initiate vocal fold oscillation," J. Acoust. Soc. Am. **121**, 2873–2881.

Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**, 2234–2439.

Kelly, J. L., and Lochbaum, C. (**1962**). "Speech synthesis," *Proceedings of the Fourth International Congress on Acoustics*, Paper G42, pp. 1–4.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Koizumi, T., Taniguchi, S., and Hiromitsu, S. (**1985**). "Glottal source-vocal tract interaction," J. Acoust. Soc. Am. **78**, 1541–1547.

Liljencrants, J. (**1985**). "Speech synthesis with a reflection-type line analog," Doctoral dissertation, Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, Sweden.

Lucero, J. C., and Koenig, L. L. (**2007**). "On the relation between the phonation threshold lung pressure and the oscillation frequency of the vocal folds," J. Acoust. Soc. Am. **121**, 3280–3283.

Markel, J. D., and Gray, A. H. J. (**1976**). *Linear Prediction of Speech* (Springer, New York).

McGowan, R. S. (**1992**). "Tongue-tip trills and vocal-tract wall compliance," J. Acoust. Soc. Am. **91**, 2903–2910.

Mergell, P., and Herzel, H. (**1997**). "Modeling biphonation—The role of the vocal tract," Speech Commun. **22**, 141–154.

Miller, D. G., and Schutte, H. K. (**2005**). " 'Mixing' the registers: Glottal source or vocal tract?," Folia Phoniatr Logop **57**, 278–291.

Neumann, K., Schunda, P., Hoth, S., and Euler, H. A. (**2005**). "The interplay between glottis and vocal tract during male passaggio," Folia Phoniatr Logop **57**, 308–327.

Rachele, R. (**1996**). *Overtone Singing Study Guide*, 2nd ed. (Cryptic Voices Productions, Amsterdam, The Netherlands).

Rosenberg, A. (**1971**). "Effect of the glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **49**, 583–590.

Rothenberg, M. (**1981**). "Acoustic interaction between the glottal source and the vocal tract," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hinano (University of Tokyo Press, Tokyo), 305–328.

Rothenberg, M. (**1987**). "Cosi fan tutte and what it means or nonlinear source-tract acoustic interaction in the soprano voice and some implications for the definition of vocal efficiency," in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaki, K. S. Harris, . (College-Hill Press, Little, Brown and Company, Boston), pp. 254–269.

Rothenberg, M., and Zahorian, S. (**1977**). "Nonlinear inverse filtering technique for estimating the glottal-area waveform," J. Acoust. Soc. Am. **61**, 1063–1071.

Scherer, R. C., Titze, I. R., and Curtis, J. F. (**1983**). "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," J. Acoust. Soc. Am. **73**, 668–676.

Sondhi, M. M., and Schroeter, J. (**1987**). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Acoust., Speech, Signal Process. **35**, 955–967.

Stevens, K. (**1999**). "Current studies in linguistics," *Acoustic Phonetics* (MIT, Cambridge, MA).

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

Story, B. H. (**1995**). "Speech simulation with an enhanced wave-reflection model of the vocal tract." Ph.D. dissertation, University of Iowa, Iowa City, IA.

Story, B. H. (**2005**). "Synergistic modes of vocal tract articulation for American English vowels," J. Acoust. Soc. Am. **118**, 3834–3859.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**, 537–554.

Sundberg, J. (**1977**). "The acoustics of the singing voice," Sci. Am. **236**, 82–91.

Svec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," J. Acoust. Soc. Am. **106**, 1523–1531.

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

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.

Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Denver, CO).

Titze, I. R. (**2001**). "Acoustic interpretation of resonant voice," J. Voice **15**, 519–528.

Titze, I. R. (**2002**). "Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model," J. Acoust. Soc. Am. **111**, 367–376.

Titze, I. R. (**2004a**). "A theoretical study of F_0-F_1 interaction with application to resonant speaking and singing voice," J. Voice **18**, 292–298.

Titze, I. R. (**2004b**). "Theory of glottal airflow and source-filter interaction in speaking and singing," Acta. Acust. Acust. **90**, 641–648.

Titze, I. R. (**2006a**). "Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation," J. Speech Lang. Hear. Res. **49**, 439–447.

Titze, I. R. (**2006b**). *The Myoelastic-Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Denver, CO).

Titze, I. R., and Hunter, E. J. (**2007**). "A two-dimensional biomechanical model of vocal fold posturing," J. Acoust. Soc. Am. **121**, 2254–2260.

Titze, I. R., Riede, T., and Popolo, P. (**2008**). "Nonlinear source-filter coupling in phonation: Vocal exercises," J. Acoust. Soc. Am. **123**, 1902–1915.

Titze, I. R., and Story, B. H. (**1997**). "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.

Titze, I. R., and Sundberg, J. (**1992**). "Vocal intensity in speakers and singers," J. Acoust. Soc. Am. **91**, 2936–2946.

Van Den Berg, J. (**1957**). "Subglottal pressures and vibration of vocal folds," Folia Phoniatr. **9**, 65–71.

Vennard, W. (**1967**). *Singing: Mechanism and Technique* (Fisher, New York).

Yumoto, E., and Kadota, Y. (**1998**). "Pliability of the vocal fold mucosa in relation to the mucosal upheaval during phonation," Arch. Otolaryngol. Head Neck Surg. **124**, 897–902.

Zañartu, M., Mongeau, L., and Wodlicka, G. R. (**2007**). "Influence of acoustic loading on an effective single mass model of the vocal folds," J. Acoust. Soc. Am. **121**, 1119–1129.

Zhang, C., Zhao, W., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, 2. Effects of flow parameters and ventricular folds," J. Acoust. Soc. Am. **112**, 2147–2154.

Zhang, Z., Neubauer, J., and Berry, D. A. (**2006a**). "The influence of subglottal acoustics on laboratory models of phonation," J. Acoust. Soc. Am. **120**, 1558–1569.

Zhang, Z., Neubauer, J., and Berry, D. A. (**2006b**). "Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds," J. Acoust. Soc. Am. **120**, 2841–2849.

Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation. 1. Computational methods and sound generation mechanisms," J. Acoust. Soc. Am. **112**, 2134–2146.

J. Acoust. Soc. Am., Vol. 123, No. 5, May 2008

Ingo R. Titze: Nonlinear source–filter theory   2749