**Supplement**

# Modeling the Biomechanical Influence of Epilaryngeal Stricture on the Vocal Folds: A Low-Dimensional Model of Vocal–Ventricular Fold Coupling

## Scott R. Moisik[a] and John H. Esling[b]

**Purpose:** Physiological and phonetic studies suggest that, at moderate levels of epilaryngeal stricture, the ventricular folds impinge upon the vocal folds and influence their dynamical behavior, which is thought to be responsible for constricted laryngeal sounds. In this work, the authors examine this hypothesis through biomechanical modeling.
**Method:** The dynamical response of a low-dimensional, lumped-element model of the vocal folds under the influence of vocal–ventricular fold coupling was evaluated. The model was assessed for F0 and cover-mass phase difference. Case studies of simulations of different constricted phonation types and of glottal stop illustrate various additional aspects of model performance.
**Results:** Simulated vocal–ventricular fold coupling lowers F0 and perturbs the mucosal wave. It also appears to reinforce

irregular patterns of oscillation, and it can enhance laryngeal closure in glottal stop production.
**Conclusion:** The effects of simulated vocal–ventricular fold coupling are consistent with sounds, such as creaky voice, harsh voice, and glottal stop, that have been observed to involve epilaryngeal stricture and apparent contact between the vocal folds and ventricular folds. This supports the view that vocal–ventricular fold coupling is important in the vibratory dynamics of such sounds and, furthermore, suggests that these sounds may intrinsically require epilaryngeal stricture.

**Key Words:** vocal folds, ventricular folds, vocal-ventricular fold coupling, epilaryngeal stricture, constricted phonation, glottal stop, biomechanical model

Research on laryngeal behavior in normal and disordered voice production and in speech emphasizes the vocal folds as the cause of diverse phenomena. This is true for phonetic models of larynx function in speech (e.g., Gordon & Ladefoged, 2001; Stevens, 1998; Zemlin, 1998) and for the study of disordered voice (Bridger & Epstein, 1983; Eysholdt, Rosanowski, & Hoppe, 2003; Voigt et al., 2010). Far less attention has been given to the study of the speech-related functions of the epilarynx (the supraglottal part of the larynx, which comprises the ventricular folds, epiglottis, and aryepiglottic folds). That the epilarynx is implicated in many speech sounds, such as laryngealized phonation, glottal stop, and pharyngeal consonants, is supported by a sizable body of laryngoscopic, laminagraphic, and laryngeal ultrasound evidence (Allen & Hollien, 1973; Brunelle, Nguyễn, & Nguyễn, 2010;

Edmondson, Chang, Hsieh, & Huang, 2011; Edmondson & Esling, 2006; Esling, Fraser, & Harris, 2005; Esling & Harris, 2003; Iwata, Sawashima, Hirose, & Niimi, 1979; Lindqvist-Gauffin, 1972; Moisik, Esling, Bird, & Lin, 2013). The goal of the present work was to examine at the biomechanical level what the epilarynx contributes to the production of such sounds.

Here the focus is on two types of speech-related laryngeal behavior: (a) "constricted" (Esling & Harris, 2005) phonation types, which include creaky voice (or vocal fry/pulse register/laryngealized voice) and harsh voice (or tense/pressed voice) and (b) glottal stop (an arrest or prevention of vocal fold vibration resulting in silence). To study the epilaryngeal contribution to these behaviors, a low-dimension model capable of representing the mechanical influence of the epilarynx on the vocal folds was created. In this article, we demonstrate that, through vocal–ventricular fold coupling, epilaryngeal constriction perturbs vocal fold dynamics in ways characteristically associated with constricted phonation and glottal stop.

## Influences of the Epilarynx on the Vocal Folds

There are many ways that the epilarynx contributes to laryngeal behavior in speech. For example, van den Berg (1955) proposed that obliteration of the laryngeal ventricle by downward movement of the ventricular folds causes a loss of the low-pass filtering effect associated with this space. Sundberg (1974; also see Honda et al., 2010; Yanagisawa, Estill, Kmucha, & Leder, 1989) demonstrated that the uncoupled resonance of the epilarynx causes the so-called singing formant. Titze (2008) argued that the epilarynx can, through constricting, aero-acoustically modify source-filter coupling, which influences vocal fold vibration and acoustic power. Bailly, Pelorson, Henrich, and Ruty (2008) claimed that the ventricular folds impose aero-acoustic effects on the vocal folds, such as increasing the pressure recovery above the glottis, which can suppress or perturb phonation and which is important in the process of vocal–ventricular phonation (see Bailly, Henrich, & Pelorson, 2010). Several studies have shown that vibration of the epilarynx is responsible for "growly" voice used in singing styles and speech (Bailly et al., 2010; Borch, Sundberg, Lindestad, & Thalén, 2004; Crevier-Buchman et al., 2012; Lindestad, Södersten, Merker, & Granqvist, 2001; Moisik, Esling, & Crevier-Buchman, 2010; Sakakibara, Fuks, Imagawa, & Tayama, 2004; Tsai et al., 2010). This vibration typically causes aero-acoustic damping of the concomitant glottal source or, if the epilarynx vibrates alone, produces a quasiperiodic epilaryngeal source.

Several researchers have postulated another possible voice-production role of the epilarynx: mechanical interaction with the vocal folds. Allen and Hollien (1973; also Hollien, 1974) suggested that the ventricular folds mechanically load or couple with the vocal folds, resulting in damped vibration. Laver (1975, p. 224) proposed that the "ventricular folds become involved in the phonation of the true vocal folds by … pressing down on the true vocal folds … [and] … combine to vibrate as more massive, composite elements." Edmondson and Esling (2006) observed apparent impingement by the ventricular folds on the vocal folds in phonetic laryngoscopy data of several different languages. Furthermore, several others have hinted at the possibility of vocal–ventricular contact in association with "constricted" sounds (e.g., Agarwal, Scherer, & Hollien, 2003; Bailly et al., 2008; Gerratt & Kreiman, 2001; Imagawa, Sakakibara, Tayama, & Niimi, 2003).

Physiological research suggests that a complex combination of internal and external laryngeal mechanisms is responsible for epilarynx constriction (Esling, Zeroual, & Crevier-Buchman, 2007; Fink, 1974; Painter, 1986; Reidenbach, 1998a, 1998b, 1997; Sakakibara, Kimura, Imagawa, Niimi, & Tayama, 2004). In constricting, the epilarynx undergoes posteroanterior narrowing, as during swallow or effort closure, causing the soft laryngeal tissues to fold, buckle, and bulge together. Because part of the mechanism that drives epilaryngeal constriction also causes medial and downward displacement of the ventricular folds, it is possible for these structures to compact into the vocal folds during epilaryngeal constriction, provided the vocal folds have adducted first (Esling, 1996).

## The Vocal–Ventricular Fold Coupling Hypothesis: Outline and Issues

Based on the above considerations, we propose a vocal–ventricular fold coupling (VVFC) hypothesis. The purpose of the present work was to examine the plausibility and possible dynamical effects of the suspected mechanical interaction between the vocal folds and ventricular folds using a low-dimensional, lumped-element computational model of these structures. From the outset, the expectation is that contact between the vocal folds and the ventricular folds results in mechanical coupling of these structures such that the dynamical response of the vocal fold system becomes dependent on the behavior of the ventricular fold system. The dynamical effects expected to occur because of the coupling are as follows:

1.  The addition of the ventricular fold mass to that of the vocal folds will increase the effective mass of the oscillating system as a whole, which will lower its frequency response;

2.  The resistive impedance of the ventricular-fold loading on the vocal folds will result in greater damping, which will limit the likelihood of the system engaging in self-sustained vibration and reduce the amplitude of vibratory displacements;

3.  The ventricular fold coupling will introduce new degrees of freedom to the overall system, resulting in new modes of oscillation, increasing the oscillatory complexity, and inducing irregularity in the pattern of oscillation; and

4.  Contact between the ventricular mass and the superior surface of the vocal fold cover will interfere with the transmission of the mucosal wave.

Increased mass (1), damping (2), oscillatory irregularity (3), and perturbation to the normal mucosal wave pattern (4) are all generally consistent with constricted phonation types, and vibration inhibition is consistent with glottal stop. The general prediction is that, with VVFC, the vocal folds should vibrate at a lower frequency, exhibit more difficulty in engaging self-sustaining oscillation, and show oscillatory irregularity in terms of period and amplitude. Other factors, such as subglottal pressure and vocal fold configuration, determine the quality of the sound output (e.g., creaky voice would most probably result from lower rather than higher subglottal pressure).

These four predictions were examined in the context of three core questions about VVFC: (a) Does VVFC affect vocal fold dynamics and are the effects consistent with what is observed in constricted sound production? (b) What is the nature of the coupling in VVFC? (c) Can vocal fold activity alone (i.e., without VVFC) simulate constricted sounds, and, if so, what does this tell us about VVFC?

## Method

### Model Foundation: Representing the Vocal Folds, the Ventricular Folds, Their Coupling, and Vocal Tract Aero-Acoustics

The VVFC model has two components: a mechanical model of the vocal and ventricular folds and a one-dimensional, aero-acoustic model of the vocal tract. The mechanical model of the vocal folds is based upon the *ST95 model* (Story & Titze, 1995), which is a low-degree-of-freedom body-cover model of the vocal folds. The aero-acoustic simulation follows Birkholz's (2005) inhomogeneous transmission line model of the vocal tract. (See Appendices A and B for additional details on model implementation.)

To represent the ventricular fold, a mass $m_v$ was added to the ST95 model. Although $m_v$ primarily represents the ventricular fold, it also represents the lumped mass of the base of the epiglottis and the tissue of the wall of the aryepiglottic fold, which is continuous with the ventricular fold, as depicted in Figures 1, Panels A and B. This mass is anchored to the laryngeal wall by stiffness $k_v$ and damping $d_v$ (Figure 1, Panel C). Two springs, $k_{vu}$ and $k_{vb}$, respectively, connect $m_v$ to vocal fold masses $m_u$ and $m_b$.[1] Each of these springs is associated with its own damping, $d_{vu}$ and $d_{vb}$ respectively, which represent the dissipation of mechanical energy of the vocal folds into the ventricular folds. Although the main point of contact in VVFC is between the mucosal surfaces of the vocal and ventricular folds, the coupling spring $k_{vb}$ was included as an option to increase the coupling between the two sets of folds (hereafter called VB coupling). It is supposed to represent the transfer of energy between the ventricular fold and the more lateral portion of the vocal fold (see Figure 1, Panel B). No such option was included for coupling the ventricular fold with the lower cover mass of the vocal folds, on the assumption that this tissue is further away from the ventricular fold mass than are either the upper cover or body masses.

Lumped-element models are considerably abstract in their representation of structure. Because of this, the addition of $m_v$ may seem similar in nature to the additional masses used in models of laryngeal disease or abnormal growths on the vocal folds (such as polyps, cysts, and nodules; e.g., Koizumi & Taniguchi, 1990). There are, however, some key differences. First, models of vocal pathology employing an extra mass are typically bilaterally asymmetrical—the mass is added to one vocal fold only (Birkholz, 2011); in the VVFC model, the additional mass is bilaterally symmetrical. Second, simulated pathological masses are typically connected only to the vocal fold "cover" mass; by comparison, each ventricular fold mass in the VVFC model is connected to the ipsilateral laryngeal wall (and vocal fold during VVFC). Finally, unlike simulations of smallish growths on the vocal folds, the ventricular masses are large in vertical extent and in mass (relative to the masses representing the vocal folds); thus, the aero-acoustic and mechanical conditions are quite distinct from those in models of pathological growths on the vocal folds.

Several previous models represent the ventricular folds as composed of two[2] coupled masses (Bailly et al., 2008, 2010; Imagawa et al., 2003). These models are usually intended to simulate vocal–ventricular phonation found in certain "throat singing" styles (see discussion above). Unlike these previous models, the goal here is to represent VVFC and not self-sustaining ventricular-fold vibration. Thus, to maintain simplicity, each ventricular fold was represented only by a single mass.

The overall anatomical distribution of the lumped masses is represented in Figure 1, Panel B, which illustrates the force vectors associated with the VVFC springs $k_{vu}$ and $k_{vb}$ (right side) and which also represent the compressive forces acting on the vocal–ventricular–epiglottal complex (left side). We assume that, in reality, there are three causes responsible for such compressive forces, which drive adduction and descent of the ventricular folds and their consequent compression into the vocal folds. The first is a downward force from the epiglottis, likely driven by the cricoepiglottic and thyroepiglottic muscles (Kimura et al., 2002). The second and third are a downward force from the anteromedial component of the craniolateral extension of the thyroarytenoid (TA) muscle and an adductory force from the posterolateral component of the same muscle system (Reidenbach, 1998a). In addition, posteroanterior action of the epilaryngeal stricture mechanism will cause longitudinal compression of the ventricular folds and consequent concentric bulging, effectively contributing to their medial and downward motion into the vocal folds.

Based on the above considerations, it seems very probable that the force exerted on the vocal folds by the ventricular folds has a vertical component. However, all masses in the VVFC model have one degree of freedom, which is in the lateromedial dimension. This means that no motion in the vertical (craniocaudal) dimension was considered. This abstraction stems from the simplifying choices made in representing the vocal folds with a lumped-element approach and the aero-acoustic simulation of the vocal tract.

This does not mean, however, that the model does not represent VVFC; it only means that the coupling here acts exclusively in the lateromedial dimension. The ST95 model itself is based on a similar abstraction to represent the body and cover masses of the vocal folds: In ST95, upper cover mass $m_u$ and body mass $m_b$ move in the same dimension, but in fact the spatial distribution of the "cover" mass follows medially and superiorly around the "body" (see Figure 1, Panel B). In this case, the mucosal wave travels across the vocal fold upwards and laterally, as depicted in

---

[1]No option was included for coupling the ventricular fold with $m_l$ on the assumption that this tissue is further from the ventricular fold mass than either the upper cover or body masses are.

[2]In general, it is conventional to model self-sustained oscillation of masses with at least two degrees of freedom to provide the delayed mechanical feedback required to support self-sustaining oscillation (McGowan, 1992; Titze, 2006).

**Figure 1.** Abstraction of vocal–ventricular fold coupling as a low-dimensional model. The dashed line in the midsagittal depiction of the larynx (Panel A) indicates the location of the anterior coronal plane shown in Panel B. Panel B shows the anatomical distribution of model masses (dashed outlines). On the left side of Panel B, a vector field (single-headed arrows) represents the compressive forces posited to exist within the epilarynx during constricted laryngeal states. The right side of Panel B illustrates the interpretation of these forces as coupling springs acting on the upper cover and body masses of the vocal folds (double-headed arrows). Schematic (Panel C) depicts the mechanical layout of the model. M = mass; k = spring; d = damping. Subscripts: b = vocal fold body; u = upper cover of vocal fold; l = lower cover of vocal fold; c = coupling between upper and lower masses; v = ventricular fold; vu = coupling between ventricular fold and upper cover of vocal fold; vb = coupling between ventricular fold and body of vocal fold.
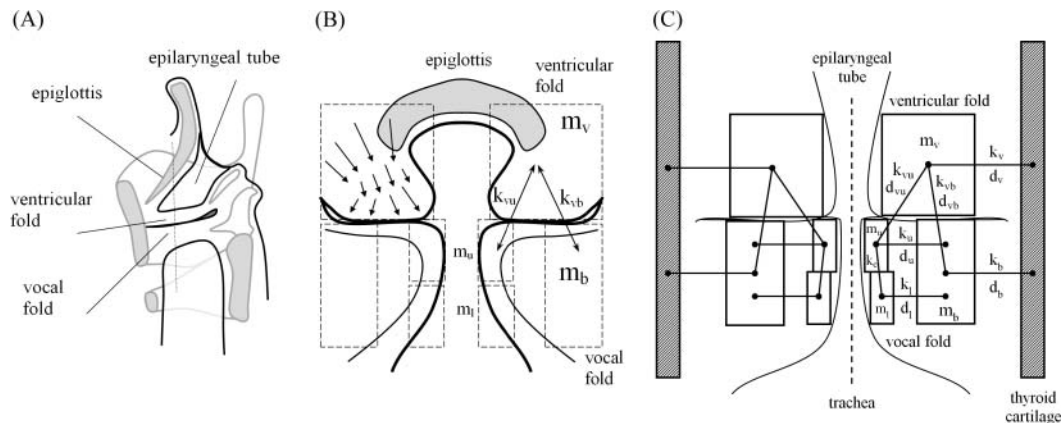


Figure 2, Panel A. Thus, the cover has vertical and latero-medial influences as the mass is redistributed throughout the glottal cycle (if we constrain our attention to two dimensions). Figure 2, Panel B shows the same cycle but adds VVFC and depicts the horizontal component of the coupling force arising from the collision (col) between the mass in the mucosal wave and the ventricular fold mass above (dotted arrow). It was therefore assumed that the model was a valid representation of VVFC.

A major consideration for design of the VVFC model was how to model the vocal–ventricular fold coupling. Two coupling forces are under consideration here: collision forces and mucous adhesion forces. It was assumed that contact between the vocal and ventricular folds constitutes collision, and therefore this type of force was assumed to always be active under VVFC. Mucous-based coupling was also considered as a special case. Thus, two basic types of coupling were examined: one with mucous adhesion and one without.

Because coupling was implemented using springs, two different VVFC springs were implemented: collision-and-mucous-type springs (push-pull; hereafter CM) and collision-only-type springs (push only; hereafter CO). In both cases (CM and CO), a compressive restoring force arises during collision between the vocal and ventricular folds (the horizontal component of the collision force arising between $m_u$ and $m_v$ is depicted in Figure 2, Panel B), and this was modeled by following the ST95 approach to vocal fold collision. In the CM case, however, the pulling force was assumed to arise from mucous adhesion, which inhibits tissue separation. This assumption was based on reports of mucous adhesion influencing vocal fold behavior. For example, Ayache, Ouaknine, Dejonckere, Prindere, and Giovanni

(2004) demonstrated that mucous-related adhesive forces that arise during the contact phase of the glottal cycle when the vocal folds press together can lower fundamental frequency and increase glottal contact. Because the ventricle is rich in laryngeal mucous glands (Agarwal et al., 2003; Reidenbach, 1998a), it seems plausible that mucus might play an important role in VVFC. Because the scale of these forces is unknown, the approach here was to model the adhesive forces in the form of the pulling influence of the CM spring when it extends past its equilibrium point. In contrast, the CO spring is made inactive (force and damping constants are nullified) when it is extended past its equilibrium point (representing separation of effective masses but not of the vocal and ventricular fold overall; see Figure 2, Panel B).

### Model Parameters

Three main biomechanical parameter sets are under consideration here: those for the vocal folds, ventricular folds, and VVFC. Details are in Appendix C. Although an effort was made to adhere closely to parameter values given by other researchers, it was found that considerable experimentation was required to locate sets of parameters that enabled self-sustaining oscillation (as the goal was to study what happens to vibration with VVFC). Thus, the parameter values reflect the choice to use self-sustained oscillation as a heuristic for parameter settings.

Two conditions were explored in the context of the body-cover model of the vocal folds (Hirano, 1975): a "normal" (N) one and a "stiff-slack" (SS) one. In the N condition, the vocal fold parameters fall within the range of parameter specifications provided by Story and Titze (1995) and Story (2002). N represents a neutral configuration

**Figure 2.** Spatial abstraction and mass movement during the glottal cycle without (Panel A) and with (Panel B) vocal–ventricular fold coupling (VVFC). Dashed vertical line = glottal midline; dashed outlines = model masses; horizontal arrows = velocities of the masses; $m_u$ = upper cover mass; $m_l$ = lower cover mass; $m_v$ = ventricular fold mass; $u_g$ = glottal flow; mw = mucosal wave; col = force due to "collision" between the ventricular fold mass and the upper cover mass.
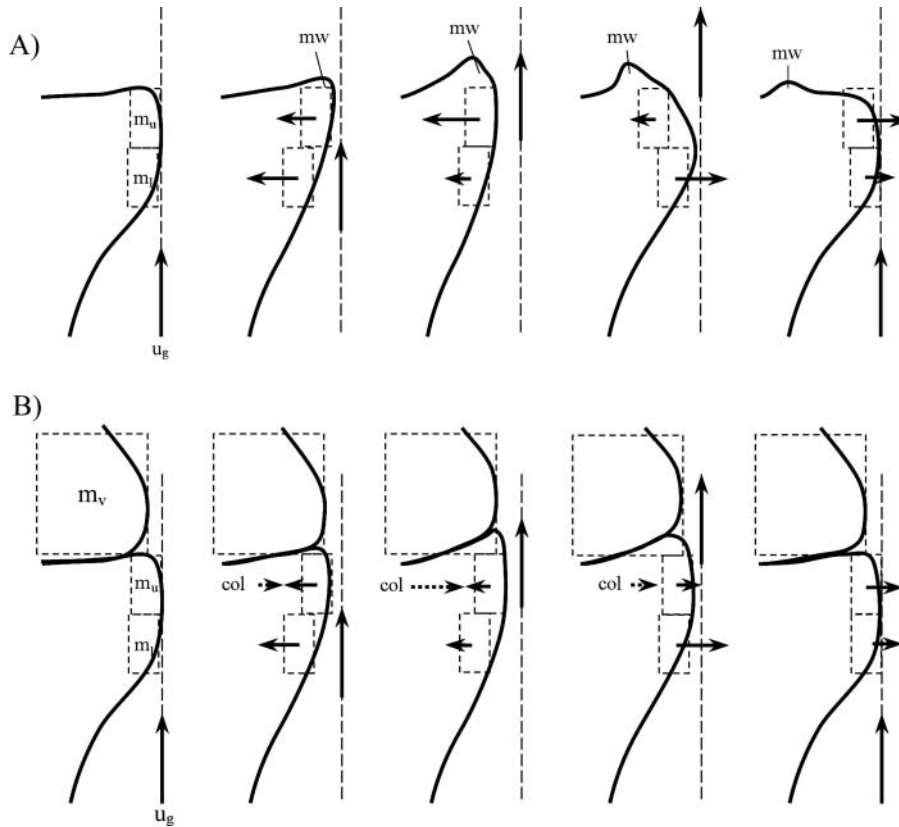


expected to produce modal-like phonation in the absence of VVFC. In contrast, the SS condition simulated the vocal-fold configuration suspected to occur during epilaryngeal stricture. SS is characterized by an increase in $k_b$ stiffness and concomitant decrease in the stiffness of $k_u$ and $k_l$, although more so for the upper cover spring (i.e., $k_u < k_l < k_b$). The upper-to-lower cover stiffness ratio was free to vary more widely than in the ST95 model, which uses a ratio of 0.7. As $k_l$ increases (relative to the upper cover), the lower cover begins to act as part of the vocal fold body (effectively reducing the vocal fold system toward a two-mass rather than three-mass configuration).

The SS condition is based on physiological observations (Fujimura, 1981; Hirano, 1975; Zemlin, 1998) and supported by recent finite-element modeling (Deguchi, Kawahara, & Takahashi, 2011) showing that unopposed TA contraction causes slackening of the vocal fold cover layer and a large increase in vocal fold body tension (as the muscle stiffens under contraction). Activation of the TA during constricted phonation types like creaky voice and harsh voice and during glottal stop production is consistent with the muscular requirements of epilaryngeal stricture (Esling et al., 2007). Given the divergence of parameter values between

the body and the cover, the SS setting should produce some effects associated with creaky and harsh phonation types, such as irregular oscillation. The point was to determine whether adding VVFC with this configuration would produce any additional effects that might reveal why the suspected VVFC occurs, if all that is required to produce constricted sounds is adjustment of the vocal fold configuration (as implicitly expressed by, e.g., Stevens, 1998).

According to Agarwal, Scherer, and Hollien (2003), the ventricular folds have higher tissue viscosity and lower mechanical stiffness than the vocal folds (see Bailly et al., 2010). Imagawa et al. (2003) used the following parameters for their two-mass ventricular fold model: $m_v$ = 260 mg ($2 \times 130$ mg), $k_v = 50\ \mathrm{Nm}^{-1}$, and $\zeta_v = 0.4$. The Imagawa et al. (2003) model values were used as baselines for experimentation within the VVFC model, but it should be noted that, unlike in the present study, the Imagawa et al. model does not represent VVFC. Thus, some divergence from these values was necessary because it was found that the different conditions for VVFC (CM and CO and the activity VB) and vocal fold configuration (N or SS) were highly sensitive to ventricular fold and vocal fold parameter values concerning whether or not self-sustaining oscillation would occur.

## Model Evaluation

Two response variables served in evaluating the model: F0 (glottal fundamental frequency) and φ (vertical phase difference of the vocal fold cover masses). Each measure was obtained over a 100-ms window located at least 100 ms into the simulation to avoid contamination by transient responses at the onset of simulation. F0 was calculated by using a peak-finding algorithm on the glottal flow pulse, calculating adjacent peak time differences to obtain glottal period and then inverting to obtain the fundamental frequency. Vertical phase difference was calculated as $\phi = 360\tau/T$, where $\tau$ is upper-lower mass time delay, and $T$ is fundamental period (Story & Titze, 1995).

The response of the model was first tested using combinations of the different conditions (no VVFC vs. VVFC, N vs. SS, [i] vs. [ɑ], and CM, CO, CM + VB, and CO + VB coupling). Correspondingly, 20 simulations were conducted (four without VVFC and 16 with VVFC). A second response evaluation involved two series of simulations (one for CM and one for CO). For each series, progressive adjustments were made to the coupling parameters (stiffness and damping) in increments from 10% to a maximum value for these parameters.

The second part consisted of five case study simulations made with a selection of parameter configurations. Four of these are for constricted phonation types (creaky-like and harsh-like) and one illustrates glottal stop simulation. The focus in these case studies is the comparison of model response with and without VVFC applied. Some methodological details unique to the individual cases, such as specific parameter settings, are presented alongside the simulation results.

## Results

### General Overview of Model Performance

Figure 3 illustrates model performance with all of the different coupling conditions (see Appendix C, Table A1, for specific parameters). Shapes and shading indicate the different conditions: without VVFC (black diamond); with

VVFC: CM (square), CO (circle), with VB coupling (red), and without VB coupling (white). Coupling strength between the ventricular and the upper cover masses was necessarily different for the N and SS contexts because the settings used for the former generally precluded phonation in the latter. Despite this divergence in parameter settings, the same basic trends hold between the two coupling conditions for F0 and φ.

VVFC had a lowering effect on F0, regardless of the coupling strength or the conditions of coupling, but the effect was more substantial with N than SS vocal fold parameters. In general, the explanation is that VVFC increased the effective mass and thereby lowered the natural frequency of the entire system (although the particular frequency depended on the specific parameters used for the VVFC, the $m_v$, and $k_v$). Vowel type did not have an appreciable effect on F0 in the N context, but it slightly lowered F0 in the SS context. The frequency response was more sensitive to whether there was VB coupling or not in the SS context.

The φ pattern was opposite to that of F0: VVFC increased the phase difference. In these examples, VVFC induced a phase shift of approximately 10° to 15°, but the effect was stronger in the SS context. Interference with the phase-difference pattern caused by VVFC appeared to inhibit full cover contact (i.e., closure at the upper and lower glottal sections) during the glottal closure phase. In such cases, the lower cover mass abducted before the upper cover mass had a chance to close its part of the glottis. A similar effect occurs if the upper and lower masses have widely divergent parameterization, and this is evident in the elevated phase-difference levels for the SS condition in general. Vowel context did not have an appreciable effect on φ.

Figure 4 shows the effects of incremental percentage-wise (10% increments starting at 10%; values at the maximum/100% are given in Appendix C, Table A1) increase of VVFC in the context of CM (squares) and CO (circles) coupling (VB coupling is used in both cases). Appendix C, Table A1 (see data for Figure 4) lists the parameters. The important point here is that, regardless of the type of VVFC (CM or CO), the increase of coupling strength

**Figure 3.** Overview data of model performance (see Appendix C, Table A1, for parameters). Black diamonds = no VVFC; squares = collision-and-mucous-type vocal–ventricular fold coupling; circles = collision-only-type VVFC; red shapes = coupling between ventricular fold mass and vocal fold body mass active; white shapes = coupling between ventricular fold mass and vocal fold body mass inactive.
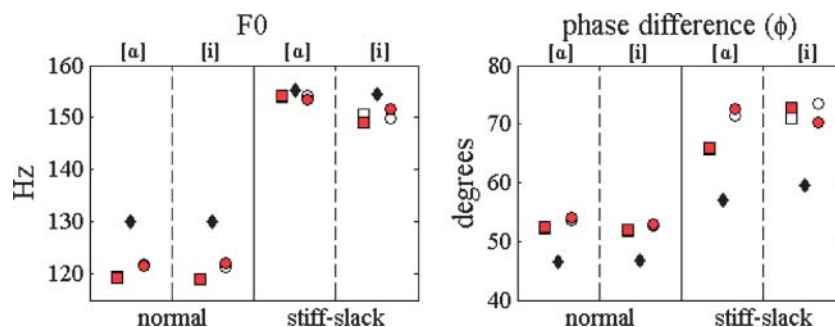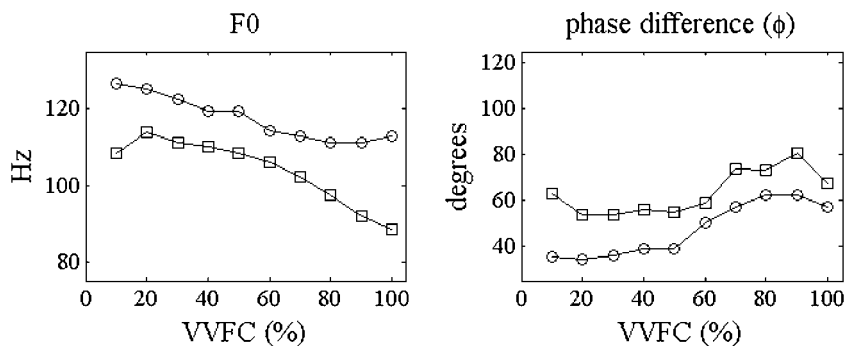
**Figure 4.** Model performance as a function of percentage-wise increases in VVFC (see Appendix C, Table A1, [α] context). Squares = collision-and-mucous-type VVFC; circles = collision-only-type VVFC.



generally coincides with the expected effects of a decrease in F0 and an increase in ϕ.

## Simulation Case Studies:Overview

The overarching goal of the VVFC model was to simulate a given vocal fold state with and without VVFC. The case studies presented below were selected from a large number of trial simulations. The selected cases represent the range of behavior observed in the process of experimental simulation and illustrate some of the more interesting effects associated with VVFC. These cases have been given names based on visual impression of the oscillatory behavior of the masses and/or the associated spectral output. Not shown here are cases in which the application of VVFC prevented phonation from occurring or in which VVFC did not have a discernible impact on the pattern of oscillation (not considering F0). Appendix C, Table A2, lists the different simulations illustrated here along with the parameters applied for the vocal fold masses, the ventricular fold masses, and for their coupling (symmetry was maintained in all simulations).

In the first four case studies, the figures show two sets of plots. The upper set of plots is for model response without VVFC and the lower set is for model response with VVFC. Each plot set contains the following plots: a spectrum (showing frequencies from 0 to 2500 Hz); a phase-plane plot for the upper cover mass (showing position vs. velocity); a plot of glottal flow, which provides an indication of glottal aperture (upper right); and a time-series plot of the right-side mass positions. The caption of Figure 5 provides further details.

*Case Study 1:Low harsh (N vs. N, CM, VB)*. Simulation using the N setting with CM-VVFC resulted in a train of alternating glottal flow pulses or "period alternation" (compare Arrow 1 and Arrow 2 in Figure 5)—essentially a type of amplitude modulation of the glottal source and classifiable as supraperiodic, nonmodal phonation (Gerratt & Kreiman, 2001). In the no-VVFC condition, F0 was 116 Hz and ϕ was 42°. By visual inspection, the quality of phonation appears similar to "modal" or "chest" voice.

In the VVFC condition, F0 dropped to 94 Hz (a difference of 22 Hz) and ϕ nearly doubled to 81°. Amplitude modulation of the glottal source caused subharmonics to appear in the spectrum (e.g., in Figure 5, compare Arrows 3 and 4, which are both placed between the first and second harmonics). This simulation is therefore consistent with a low, harsh voice quality (Esling & Harris, 2005; Laver, 1980). The phase plot for the upper cover mass in VVFC condition shows two orbits associated with the alternating glottal cycles, one with slightly greater displacement and velocity than the other. The displacement and velocity of the upper cover mass diminished by nearly a factor of 2 in comparison to the no-VVFC condition. This might have been due to increased damping acting on the upper cover mass. Inspection of the mass-displacement time series for the VVFC condition reveals that the alternating-pulse pattern was present in the motion of the ventricular fold mass and somewhat less so in the body mass of the vocal folds. The lower cover mass was least influenced in this way. The magnitude of the flow pulse diminished in the VVFC condition because of the diminished glottal aperture caused by the increased damping.

*Case Study 2: Irregularity stabilization (SS vs. SS, CM, VB)*. This case study illustrates "irregularity stabilization" or reinforcement (Figure 6; note that VVFC damping was nullified to achieve self-sustaining oscillation in this case). Without VVFC, the SS configuration exhibited a slight tendency toward irregular oscillatory behavior (see Bracket 1 in Figure 6). This initial irregularity may reflect the influence of the transient response of the aero-acoustic system, which was stronger in the SS condition than when the vocal fold body-cover system was more uniformly stiff, as in the N condition. However, stable oscillation suddenly resumed after 200 ms (corresponding to the limit orbit, the dark ring marked by the arrow in the phase plot). With VVFC, this irregular tendency was "stabilized" into a "period-alternating" type of irregular pattern (Bracket 2 demarcates one cycle in Figure 6).

F0 decreased with VVFC applied (from 129 Hz to 103 Hz) and ϕ increased (from 44° to 92°). The glottal flow pulse increased in intensity somewhat with VVFC, but this is probably attributable to the absence of damping.

**Figure 5.** Plots for Case Study 1 (see Table 2 for parameters). Conditions: control simulation (upper plot set); simulation with VVFC (lower plot set). Plots (for each condition): spectrum (upper left); glottal flow (upper right); upper cover mass phase-plane plot (lower left); mass positions (lower right). Lines in mass position plots: red line = upper cover mass ($m_u$); blue line = lower cover mass ($m_l$); green line = body mass ($m_b$); thick black line = ventricular fold mass ($m_v$). N = normal; CM = collision-and-mucous-type spring; VB = $m_{vb}$ to $m_b$ coupling.
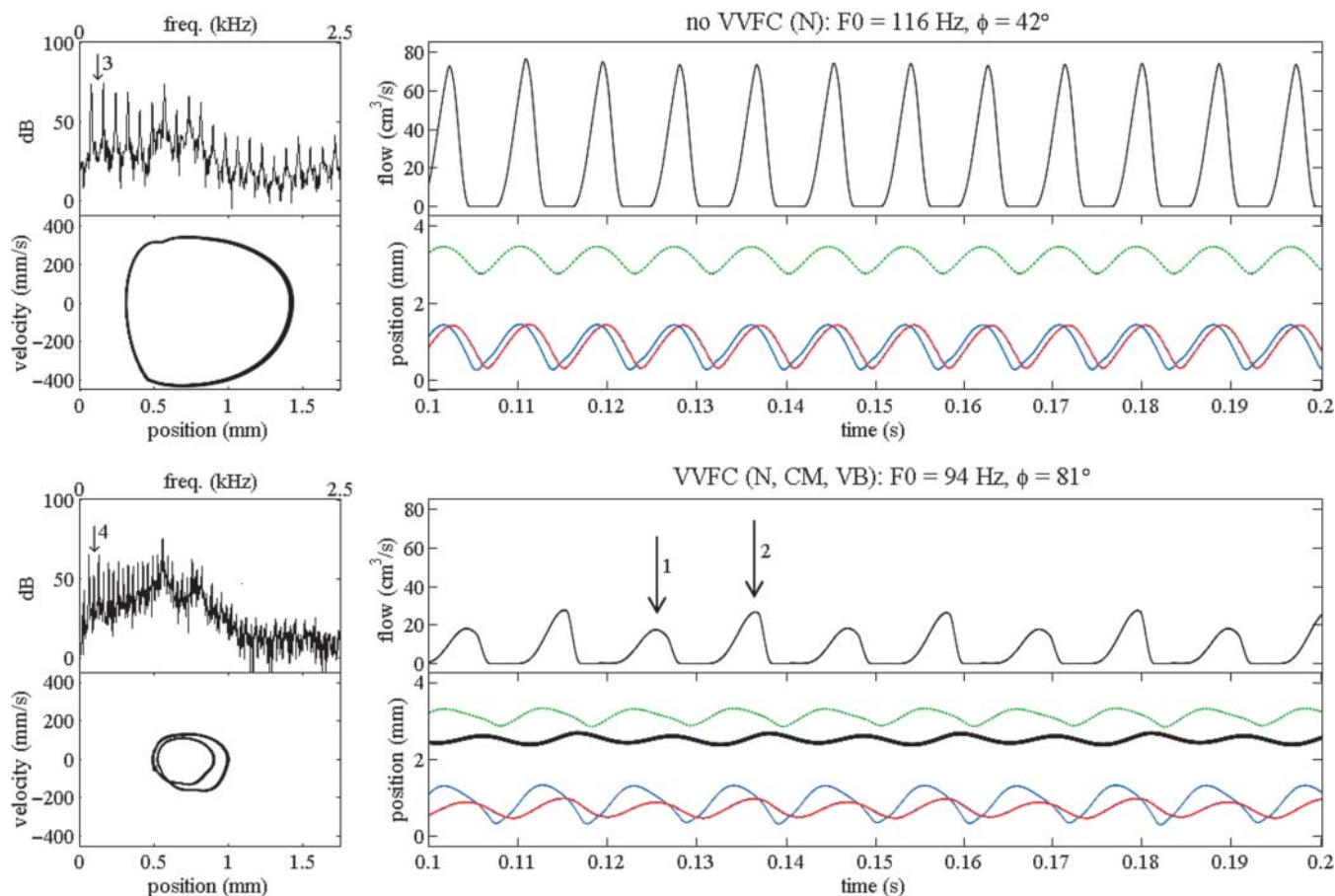
The original goal for this simulation was to attain a creaky-like behavior. Despite the low glottal flow, the irregularity in amplitude of the pulses, and the lengthened closed phase between the alternating pulses, the quality looks more like that in Case Study 1 (low harsh). This is because of the strong subharmonic structure (compare Arrows 1 and 2 in the spectra[3] of Figure 6, which show a subharmonic between the first and second harmonics in the lower plot (VVFC) and in the upper plot (no VVFC).

*Case Study 3: High harsh (SS vs. SS, CO).* This simulation (Figure 7) achieved a high F0 with SS parameters (but in this case, cover stiffness was increased, effective mass of the vocal folds was decreased, and VVFC damping was again nullified to achieve self-sustaining oscillation). Without VVFC, the vibration approached a falsetto-like mode, but the cover

displacement was large in comparison to a true falsetto (Story & Titze, 1995). Moreover, glottal contact still occurred. With VVFC applied, harsh voice at high pitch (Esling & Harris 2005) or "high harsh" was simulated. This type of harsh voice occurs under the physiologically conflicting combination of cricothyroid contraction (which is required to increase longitudinal tension of the vocal folds but consequently expands the posteroanterior dimension of the epilarynx) and epilaryngeal constriction (which is characterized by posteroanterior reduction of the epilaryngeal space). Harsh voice at high pitch is phonetically attested in the harsh register of Bai, a Tibeto-Burman language (see Edmondson & Esling, 2006).

VVFC in this configuration introduced instability into the glottal flow pulse. Compare the phase-plane plots of the two conditions: Unlike the no-VVFC condition, the plot for the VVFC condition shows considerable cycle-to-cycle variation. The difference in magnitude between these two cases reveals that oscillation was also much more restrained in amplitude of displacement with VVFC applied (even without damping), and this resulted in a substantial drop in

---

[3]Note that in the no-VVFC condition, the spectrum was calculated over the stable region (from 200 to 300 ms), not the transient one (under Bracket 1).

**Figure 6.** Plots for Case Study 2. SS = stiff-slack.
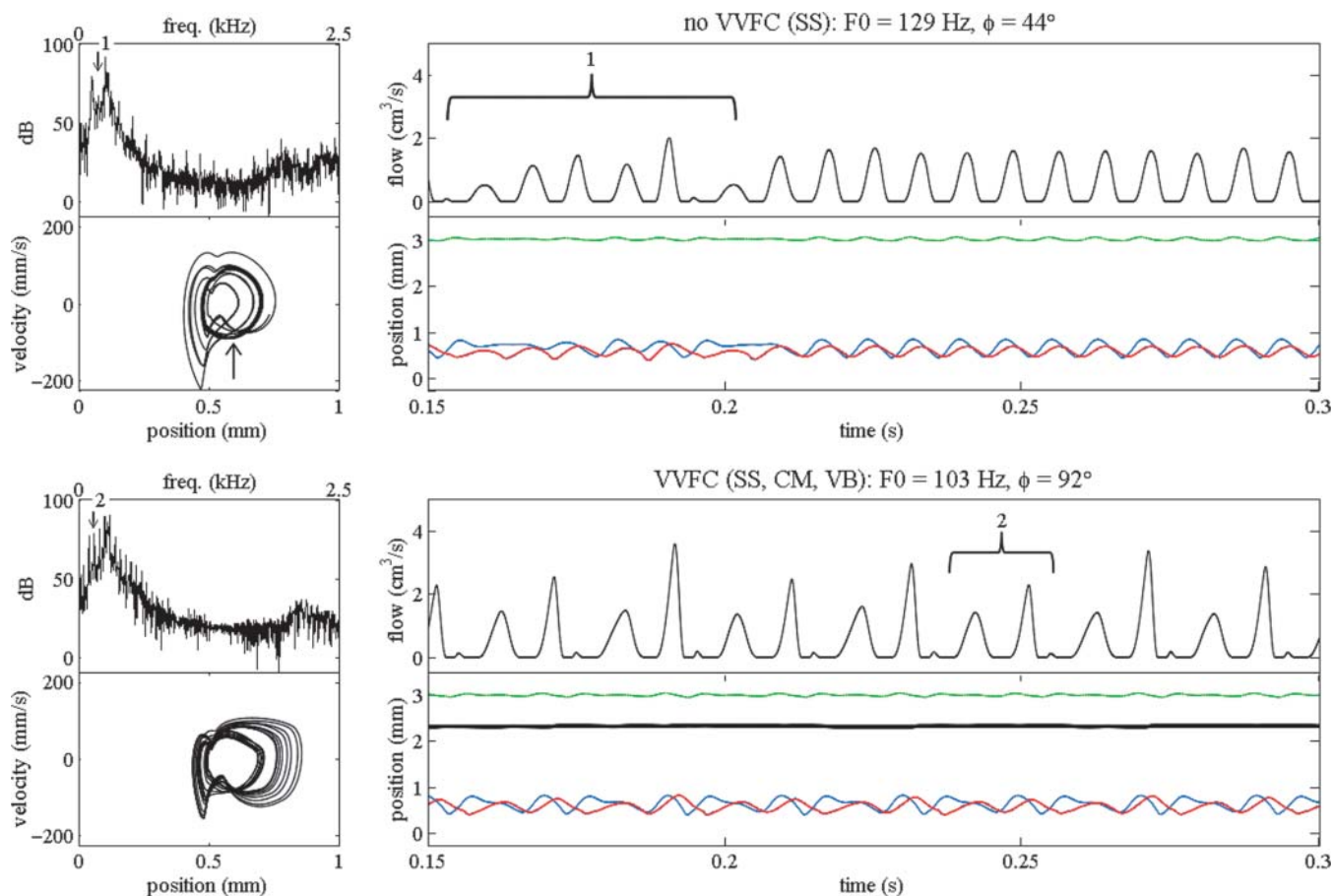


glottal flow. As the spectrum indicates, the VVFC case exhibited weak harmonic structure that was nearly lost in the interharmonic noise. This too contrasts with the no-VVFC condition, where the harmonics were very strong (~50 dB above the interharmonic noise). Because of the absence of strong subharmonic structure, the VVFC condition might qualify as a "noisy voice" nonmodal variant (Gerratt & Kreiman, 2001, p. 372), but this is still consistent with the concept of harsh voice (Laver, 1980).

*Case Study 4: Stabilized creaky (SS vs. SS, CO, VB)*. In this simulation (Figure 8), another attempt to simulate creaky voice was made, but this time at a very low F0. To represent the thickening up of the vocal folds in this condition, $m_u$ and $m_l$ were increased by a factor of 1.6, and $m_b$ was increased by a factor of 3 (relative to the N setting). The stiffness of $k_u$ was decreased to 4 Nm$^{-1}$. $k_l$ remained at a stiffness between that of $k_b$ and $k_u$.[4] Following Stevens

---

[4]Without a stiff enough lower cover, phonation with this parameter set is not possible, because the upper cover mass will never surpass the lower cover mass to change the glottal profile from convergent to divergent—essential for maintaining pressure conditions within the glottis that support oscillation rather than hinder it.

(1998), vocal fold adduction was increased by shifting the body masses $m_b$ medially by 0.5 mm. Subglottal pressure was set low (400 Pa).

With these somewhat extreme settings, it was found that a pattern consistent with descriptions of creaky voice—that is, one showing a substantial drop in F0, a lengthened closed phase, and "double pulsing" (Gerratt & Kreiman, 2001; Laver, 1980)—was possible to achieve in the no-VVFC condition.

Double pulsing is essentially a single, complex glottal cycle. To illustrate, we can refer to the glottal flow plot in the VVFC condition (Figure 8). In a single double-pulse cycle, first there is a large pulse (Arrow 1). A low-amplitude pulse (Arrow 2) then quickly follows this. Finally, there is a phase of closure (Arrow 3). Then the pattern repeats.

The VVFC condition exhibited irregularity stabilization/reinforcement, similar to that observed in Case Study 2. In this case, the irregular, double-pulse cycle (Bracket 1) stabilized into a repeating pattern. There were moments in the no-VVFC condition (e.g., between 0.16 and 0.24 s; Bracket 2 in Figure 8) when the vocal folds exhibited this pattern, but without the stabilization of the VVFC, their behavior was much more erratic.

**Figure 7.** Plots for Case Study 3. CO = collision-only-type springs.
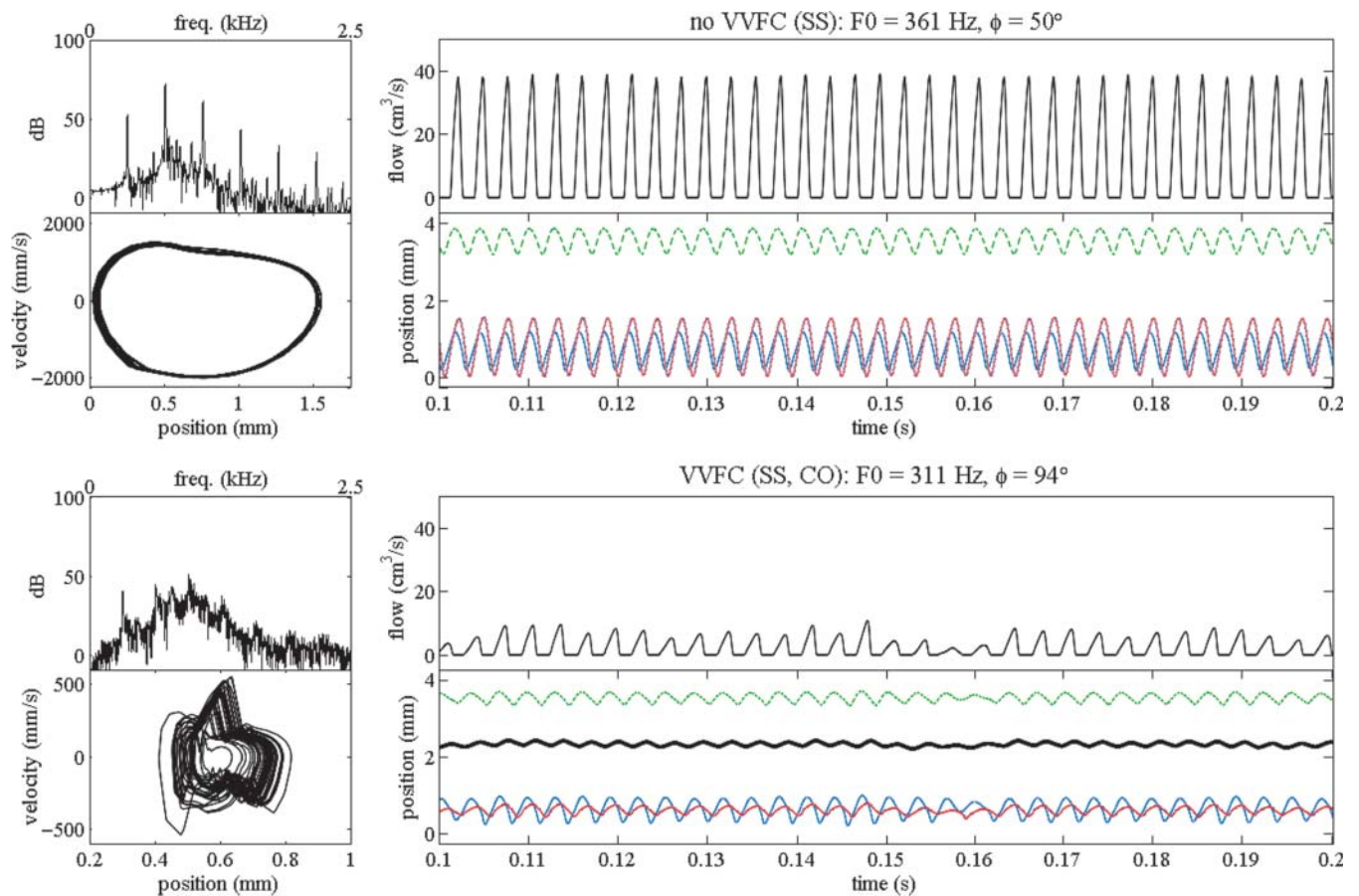


Besides the stabilizing effect, VVFC also lowered F0, this time by 9 Hz. The very high φ in both cases probably reflects the divergence in parameters between the upper and lower cover masses (the peaks in lateral displacement for each mass were nearly 180 degrees out of phase). The spectral profile of each condition was similar, but, in the no-VVFC condition, the noise amplitude was very high, whereas in the VVFC condition, the harmonics were much stronger overall. No obvious subharmonics were produced because the small pulse (Arrow 2) of the double-pulse sequence is very low in amplitude, and its timing is not in harmonic alignment with the large pulse (Arrow 1). This distinguishes the double-pulsing pattern from the period-alternation one seen in Case Studies 1 and 2.

*Case Study 5: Enhanced glottal stop.* Several studies (e.g., Edmondson et al., 2011; Esling et al., 2005; Garellek, 2013; Iwata et al., 1979; Moisik et al., 2011) claim that glottal closure often occurs with some apparent ventricular fold impingement upon the vocal folds. It has been suggested that this plays a role in arresting the vocal folds (e.g., Laver, 1980). This case study examines what, if anything, VVFC might contribute to glottal stop.

To simulate glottal stop, a ramp function controlled model parameters. This function linearly increased the equilibrium length of the vocal fold body spring, which pushed the body and, consequently, the cover masses medially (from 100 and 200 ms). Then the reverse applied (from 200 to 300 ms). Thus, the adduction of the vocal folds peaked at 200 ms, and the duration of the simulated glottal-stop gesture was 200 ms. It is assumed that, as the vocal folds adduct, the body-cover stiffness becomes more stiff-slack because of TA contraction, which is supported by EMG observations of glottal closure gestures (Hillel, 2001; Hirano & Ohala, 1969). To simulate this effect, the stiffness parameters of the vocal folds were modified with the same ramp function used for simulating increased adductory force. These values gradually changed from the N to SS setting and then back to N again by means of the ramp function (see Appendix C, Table A2, for parameters). Through experimentation, it was found that 1 mm of body-mass adduction (i.e., each $m_b$ moves medially by 1 mm) was sufficient to stop vocal fold vibration entirely, with a 20-ms response time following the start of the ramp function. If the extent of medial displacement was reduced to 0.5 mm, however,

**Figure 8.** Plots for Case Study 4.
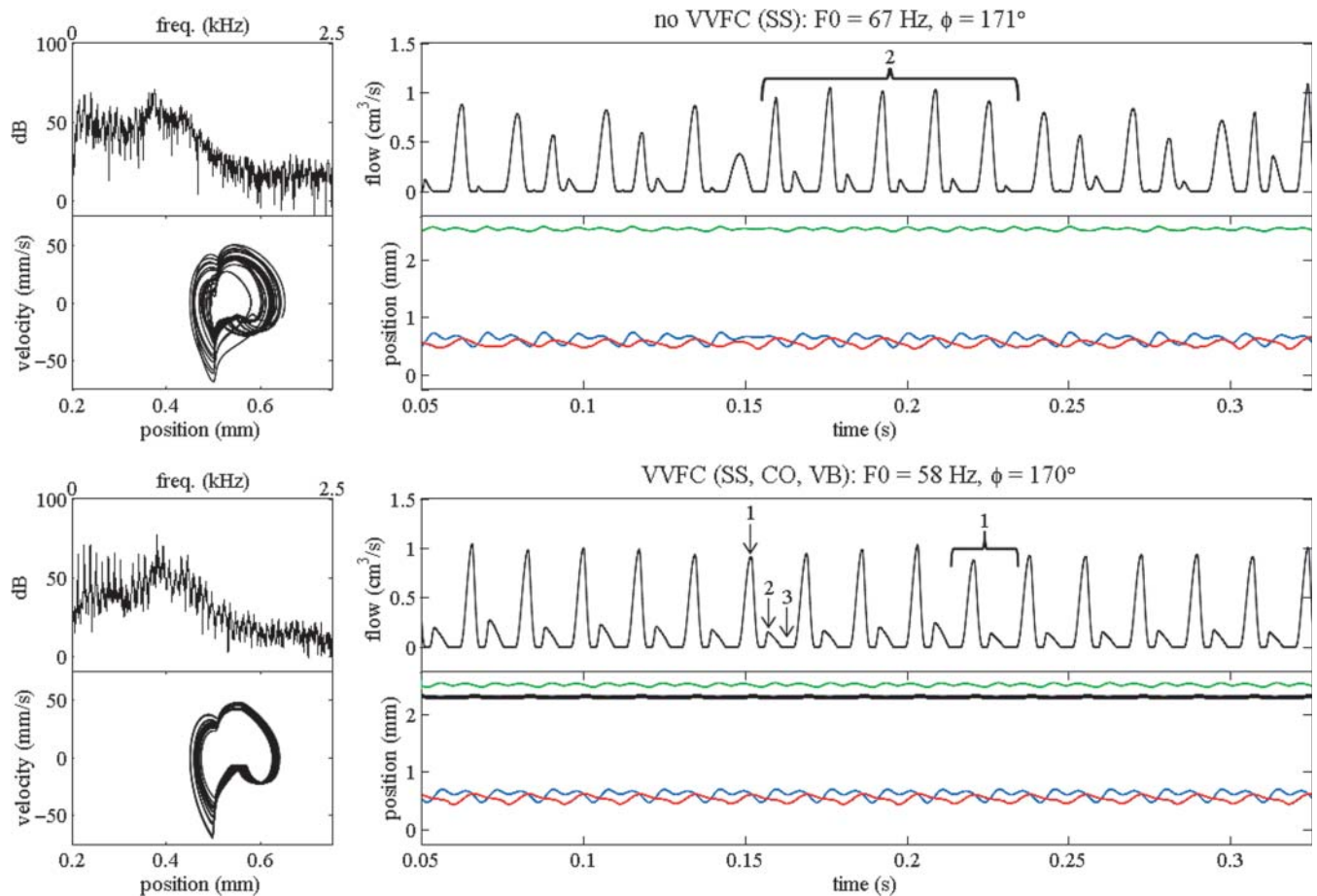


vibration did not entirely halt. Thus, to illustrate the effect of VVFC more clearly, this value for body medialization was used in the simulations presented here.

To simulate glottal stop with VVFC applied, the same ramp function-based vocal fold parameter sweeping described above was used, but, in addition, CO-VVFC parameters (see Appendix C, Table A2) were likewise increased and decreased by means of the ramp function. At the start and stop of the ramp function, the VVFC parameters were null. Unlike VVFC in the other simulations, here the ventricular folds were medialized to the point of being in contact at the peak of the glottal stop gesture. This was timed such that contact occurred at 150 ms, was held constant for 100 ms, and then was released at 250 ms, returning to its original configuration at 300 ms. The VVFC applied at approximately the point where the medial edges of the $m_v$ masses exceeded the medial edges of the $m_b$ masses.

Figure 9 shows the results of these two simulations. Without VVFC, the closure phase exhibited glottal leakage characterized by irregular pulses (Bracket 1). At the peak of the gesture (Bracket 2), the lower cover was even slightly abducted and relatively stationary (although the upper mass

continued to vibrate). With VVFC applied, vocal fold oscillation was rapidly halted, and there was no glottal leakage during the closure period (although the lower cover masses slightly abducted).

The simulation above demonstrates that VVFC did have the effect of aiding in arresting vibration when the vocal folds do not adduct as strongly (0.5 vs. 1 mm of body mass adduction) and/or their biomechanical parameters (SS setting) do not favor oscillatory arrest.

## Discussion

### Question A: Does the VVFC Have the Expected Effects?

The low-dimensional, lumped-element simulation of VVFC reproduced the hypothesized effects (see the introduction): (1) effective mass increased, (2) damping increased, (3) new modes of oscillation and oscillatory irregularity emerged, and (4) mucosal wave interference occurred. Allen and Hollien (1973) and Laver (1975, 1980) provided speculations that (1) and (2) should be consequences of VVFC.

**Figure 9.** Simulation of glottal stop (Case Study 5) with (top) and without (bottom) VVCF. Each plot set shows glottal flow (upper plot) and mass position (lower plot).
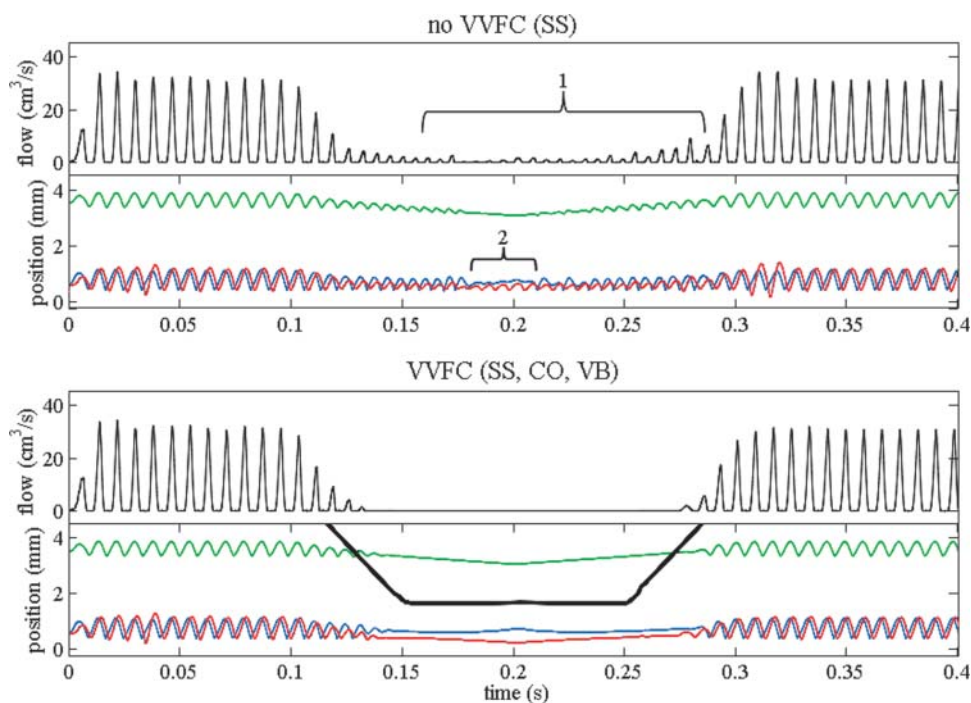


In the model, increased effective mass (1) was indicated by a decrease in the fundamental frequency of the oscillating system (judged from the glottal flow). This decline happened despite the fact that VVFC actually introduces additional stiffness into the system. This additional stiffness, however, is accompanied by increased damping (2). In the model, the increased damping was probably responsible for the difficulty in achieving self-sustaining oscillation when VVFC was applied (and in some simulations, it was necessary to turn off the VVFC damping to achieve vibration). When self-sustaining oscillation did occur, increased damping was indicated by reduced magnitude of displacement of the upper cover mass and, consequently, the glottal flow. It is possible that concomitant narrowing of the airway immediately above the glottis at the ventricular level was also a factor in reducing the glottal flow (van den Berg, 1955; van den Berg, Zantema, & Doornenbal, 1957).

Hypothesized Effect 3, corresponding with Laver's (1975) notion of VVFC constituting a "composite" oscillator, was evident from the occurrence of irregular vocal fold vibration when VVFC is applied (e.g., Case Study 1). Although the effect can be described as irregular, it is not entirely aperiodic; rather, it can be generally described as a "period-alternation" pattern, which corresponds with 2:1 amplitude modulation of the glottal pulse (i.e., every second pulse is diminished in intensity). Surprisingly, in the case studies featuring SS vocal fold parameters (Case Studies 2, 4, and 5), irregular vibrational behavior occurred even without

VVFC (especially Case Studies 2 and 4). However, the application of VVFC in such cases caused a stabilization of the oscillation toward the semi-irregular, period-alternating, or double-pulsing pattern (depending on the vocal fold parameters; compare Case Studies 2 and 4). This suggests that VVFC is intrinsically associated with such irregular patterns of vibration: It functions to reinforce or "stabilize" them.

Some comment should be made on what distinguishes the two types of irregular patterns observed here and in natural speech (Gerratt & Kreiman, 2001). Although similar in appearance to the period-alternation pattern, the double-pulsing pattern is different because of the timing of the pulses: The two pulses are close together in phase (e.g., see Arrow 1 and Arrow 2 in Figure 8) and then followed by a longer closure phase (Arrow 3 in Figure 8). In the double-pulsing case, it could be said that each individual pulse does not constitute a complete glottal cycle; rather, they form a single, complex cycle. In the period-alternation pattern, the spacing between the pulses is roughly equal (e.g., the pulses identified by Arrow 1 and Arrow 2 in Figure 5). Each individual pulse constitutes a complete glottal cycle, although the movement of mass in each cycle is different. These differences are manifest in the spectral profile: Because period alternation is periodic in nature (i.e., the pulses alternate in intensity but arrive at regular intervals), it creates subharmonic structure—it constitutes amplitude modulation. The double-pulsing pattern is not as periodic in nature and, thus,

has no clear subharmonic structure but rather interharmonic noise associated with the inconsistent timing and shape of each pulse.

The period-alternation effect is very similar to what is observed to happen to the glottal source in cases of vocal–ventricular (e.g., Lindestad et al., 2001) and vocal–aryepiglottic phonation (e.g., Moisik et al., 2010). In such cases of independent and simultaneous vibration of the vocal folds and some portion of the epilaryngeal structures above, the effect on the glottal source is more likely to be aero-acoustic in nature rather than mechanical. In the aero-acoustic case, the glottal airflow is damped by the opening and closing of the epilaryngeal airway. The lower natural frequency of the ventricular folds or aryepiglottic folds relative to the vocal folds means there is potential amplitude modulation of the glottal source (depending on the regularity of the epilaryngeal pulsing). Changes in the pressures acting on the vocal folds arising from the interruptions to downstream flow cyclically perturb their vibration. In the specific case of vocal–ventricular vibration, during every second glottal period, when the ventricular folds come into complete contact, there is increased pressure recovery above the glottis, and this reduces the transglottal pressure drop, momentarily inhibiting self-sustained oscillation of the vocal folds (Bailly et al., 2008, 2010). The ventricular space must be open to allow these effects to occur, but in the case of real VVFC, there is no ventricle (because vocal–ventricular fold contact mostly obliterates the ventricle space).

In the mechanical case, the perturbation to vocal fold vibration is primarily attributable to an increase in the number of natural modes of vibration of the overall system resulting from coupling the ventricular fold mass to the vocal fold mass. It is well established that coupled oscillators will tend to lock into frequency regimes characterized by the superposition of these modes (Tokuda, Horáček, Švec, & Herzel, 2007), and this would appear to be true for the VVFC model. This does not rule out simultaneous and supporting aero-acoustic influence in triggering the period-alternating mode, but it is unlikely that such effects were solely responsible for the behavior of the model in these cases. The damping of the system should also be considered, because it will likely cause vibrations associated with new modes of oscillation introduced by VVFC to decay rapidly—perhaps too quickly to have any appreciable effect on the vocal folds.

On the basis of the simulations, there is evidence that VVFC causes (4) interference with the mucosal wave. For natural phonation, Story and Titze (1995) suggested that the phase difference ($\phi$) is typically 27° to 61° (however, these data are for excised canine larynges). In the VVFC model, the VVFC condition was usually associated with an increase in $\phi$ toward values exceeding 61° (although the extent varies with the specific parameters used). Large shifts in cover-mass phasing can result in glottal leaks during the glottal closure phase if one mass starts to abduct too early or one adducts too late. Examples of glottal leaks because of skewed cover-mass phasing are in Case Studies 2 and 4. Such leaks might further destabilize vocal fold vibration if they grow with every glottal cycle (e.g., see the no-VVFC case in Case Study 4). Ironically, in such cases, VVFC might help to stabilize the extent of irregularity by partially correcting the phasing of the upper cover mass.

Overall, the focus here is on the mechanical role of VVFC. Aero-acoustic effects were not directly considered, although it is acknowledged that these likely play a significant role. Aero-acoustic considerations in the model followed the assumptions made in Birkholz (2005), one of which is that there is no pressure recovery at the exit of the glottis (or any cross-sectional expansion). In conventional models of glottal flow, a sudden expansion is assumed to occur at the glottal exit. In such cases, pressure recovery is assumed to follow the dynamic pressure loss (Titze, 2006), but under conditions of a wide epilaryngeal duct area, the pressure recovery can be neglected and jet flow can be applied to sections above the minimum glottal diameter. If, however, the epilaryngeal duct area is very narrow, which is true for constricted phonation types, the pressure recovery is no longer negligible. The approach taken for the present model was to adhere closely to the aero-acoustic model in Birkholz (2005), which neglects pressure recovery throughout the entire vocal tract. Because this may have had a significant impact on the behavior of the model, future research should address this issue.

### Question B: What Are the Effects of the Different Types of VVFC?

The original goal was to compare, on equal footing, the different types of VVFC: collision-and-mucous (CM) type and collision-only (CO) type and with or without coupling between $m_v$ and $m_b$ (VB coupling). However, consistent parameter settings were very difficult to achieve, because, depending on the specific conditions, the addition of VVFC very often resulted in failure of the vocal fold masses to engage in self-sustaining oscillation. Nonetheless, self-sustaining oscillation was achievable with both types of coupling, and, regardless of the coupling type, its effects were generally consistent with the VVFC hypothesis.

Notably, the presence or absence of VB coupling (represented by $k_{vb}$ and $d_{vb}$) did not have a strong effect on F0 or $\phi$, but note that values were kept low to allow self-sustaining oscillation to occur (Appendix C, Table A1). Furthermore, the geometry of the model was such that the ventricular mass $m_v$ was centered closer to $m_b$ than $m_u$. Because all masses in the model were constrained to just the lateromedial degree of freedom, the active, lateromedial component of VVFC force acting on $m_b$ by means of $k_{vb}$ and $d_{vb}$ was relatively small in relation to other forces acting on $m_b$. The motivation for modeling VB coupling was that the ventricular fold footprint on the vocal folds is not just over the medial part of the cover, but probably also over its more lateral part (as depicted in Figure 1b). In the ST95 model, the vocal fold cover is purely medial. There is no lateral section: The cover mass is lumped toward the medial edge of the vocal folds. To better understand what effect the more lateral region of mass distribution in VVFC might

be having, it will be necessary to modify the model in future work to include the lateral portion of the vocal fold cover or employ higher dimensional modeling, such as the finite element method.

### Question C: Can "Constricted" Sounds Be Simulated Without VVFC?

It was demonstrated that with some of the SS settings, it was possible to simulate creaky-like phonation without VVFC. It was also possible to simulate glottal stop without VVFC simply by increasing vocal fold adduction. In the model, VVFC appears to enhance glottal stop by supressing glottal leakage during the stop phase when SS settings apply. As noted above, however, there is laryngoscopic evidence for the occurrence of VVFC in natural instances of constricted sounds (e.g., Brunelle et al., 2010; Edmondson & Esling, 2006). For example, according to Garellek (2013), 95% of careful productions of glottal stop exhibit some degree of ventricular fold medialization (although only 17% of such productions have complete midline contact of the ventricular folds, and vocal fold vibration tends to stop before the extremum of ventricular medialization is reached). Furthermore, there is good reason to believe that vocal–ventricular fold contact (which cannot be evaluated directly from laryngoscopic evidence) does indeed occur based on laminagraphic (Allen & Hollien, 1973) and laryngeal ultrasound evidence (Moisik et al., 2011). What, then, is VVFC doing in the production of "constricted" sounds?

One possibility is that constricted gestures produced by the laryngeal system form a "cascading" series of strictures or closures, each increasing in degree of stability, but each "costing" more time to execute. There are at least three larynx-internal components in this cascade: the vocal folds, the ventricular folds, and the aryepiglottic folds (which work in concert with the epiglottis, "the AE level"). Stopping phonation with vocal fold adduction alone is quick but less stable than engaging vocal fold adduction in combination with ventricular fold engagement. Engagement of the AE level as a third stage in the cascade—although phonetically classified as epiglottal stop (Esling, 1996)—can indeed occur in the production of glottal stop as Lindqvist-Gauffin (1972) observed. This is presumably the most stable oscillation-inhibiting degree of closure, but it is also the most costly in terms of physiological effort and execution time.

Factors such as inertia, the impossibility of instantaneously nullifying subglottal pressure, and the concave downward or "inlet valve" shape of the subglottal space (Edmondson et al., 2011; Fletcher, 1993) make the vocal folds prone to continue vibrating once they have been set in motion. As suggested by many in the literature and as demonstrated here, increased effective mass, damping, and stiffness from VVFC will rapidly diminish displacement of the folds, hinder the oscillation, and thereby limit leakage in attempted glottal stop gestures. If ventricular fold adduction completely seals off the laryngeal airway, the (theoretically) infinitely high resistance to airflow should also greatly inhibit phonation.

As Lindqvist-Gauffin (1972) suggested, laryngeal speech functions in general may have been exapted from phylogenetically basic functions, one of which is protective laryngeal closure. In this view, glottal stop is inherently epilaryngeal in nature,[5] being closely related to gestures such as effort closure and swallowing closure (Fink, 1974), but refined for the purpose of speech under pressure to involve minimal effort and rapid production. Phonation types such as creaky and harsh voice would thus represent similar exaptations of the closure mechanism—but modified to support vocal fold vibration. In this context, vocal–ventricular fold contact may be biomechanically important for the reliable production of constricted phonation types by guiding the laryngeal dynamics toward the irregular regimes of vibration, which characterize creaky and harsh voice qualities.

## Conclusions

This research has proposed and examined a vocal–ventricular fold coupling (VVFC) hypothesis using a low-dimensional, lumped-element computational model of the vocal and ventricular folds. We have shown that such a model could reproduce effects expected to occur with VVFC, such as a decrease in F0, an increase in vocal fold damping, and interference with the vocal fold mucosal wave. Furthermore, VVFC appears to reinforce irregular vibratory patterns and enhances vocal fold arrest in glottal stop articulation.

Given that low-dimensional models of the vocal folds are highly limited in spatial resolution and subject to many simplifying assumptions, it is not possible to state conclusively what exactly VVFC does in natural speech. For example, many effects, such as longitudinal changes to vocal fold dynamics resulting from VVFC, could not be represented. Finite element modeling of VVFC would offer more detailed and robust results and stands as a logical next step for research into VVFC.

The present study has relevance for several fields of research. In phonetic and phonological theory, sounds such as glottal stop and creaky voice, long considered to be a function of vocal fold adduction alone, are probably enhanced by epilaryngeal stricture—if not intrinsically dependent on it. This study also has significance for speech simulation, where it is desirable to use computationally inexpensive models to simulate natural sounding voices (Birkholz, 2005, 2011). The VVFC model may serve as a means to synthesize constricted phonation types and could improve the synthesis of glottal stop. This study does not directly examine disordered speech, but it is certainly relevant to this topic. For example, hyperfunctioning of the epilarynx in functional speech disorders might explain some

---

[5]Our own laryngoscopic observations of rapid glottal-stop–vowel sequences (such as [iʔiʔiʔi]) indicate that the dominant motion visually correlated with [ʔ] is posteroanterior epilaryngeal stricture rather than lateromedial vocal fold motion. In fact, the vocal folds do not appreciably increase in their adductory compression but instead seem to consistently maintain the "pre-phonation" configuration.

of the characteristic constricted qualities of these voices, such as excessive glottal attack (Hess, Verdolini, Bierhals, Mansmann, & Gross, 1998). In general, this work constitutes a preliminary basis for future, more elaborate modeling studies that examine interaction between the epilarynx and the vocal folds.

## Acknowledgments

## References

Agarwal, M., Scherer, R. C., & Hollien, H. (2003). The false vocal folds: Shape and size in frontal view during phonation based on laminagraphic tracings. *Journal of Voice, 17,* 97–113. doi:10.1016/S0892-1997(03)00012-2

Allen, E. L., & Hollien, H. (1973). A laminagraphic study of pulse phonation. *Folia Phoniatrica, 25,* 241–250.

Ayache, S., Ouaknine, M., Dejonkere, P., Prindere, P., & Giovanni, A. (2004). Experimental study of the effects of surface mucus viscosity on the glottic cycle. *Journal of Voice, 18,* 107–115. doi:10.1016/j.jvoice.2003.07.004

Bailly, L., Henrich, N., & Pelorson, X. (2010). Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling. *The Journal of the Acoustical Society of America, 127,* 3212–3222. doi:10.1121/1.3365220

Bailly, L., Pelorson, X., Henrich, N., & Ruty, N. (2008). Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation. *The Journal of the Acoustical Society of America, 124,* 3296–3308. doi:10.1121/1.2977740

Birkholz, P. (2005). *3D-Artikulatorische Sprachsynthese* [3D Articulatory Speech Synthesis] (Unpublished doctoral dissertation). Universität Rostock, Rostock, Germany.

Birkholz, P. (2011). A survey of self-oscillating lumped-element models of the vocal folds. In B. J. Kröger & P. Birkholz (Eds.), *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2011* (pp. 47–58). Dresden, Germany: TUDPress.

Borch, D. Z., Sundberg, J., Lindestad, P.-Å., & Thalén, M. (2004). Vocal fold vibration and voice source aperiodicity in "dist" tones: A study of a timbral ornament in rock singing. *Logopedics Phoniatrics Vocology, 29,* 147–153. doi:10.1080/14015430410016073

Bridger, M. W. M., & Epstein, R. (1983). Functional voice disorders. *Journal of Laryngology and Otology, 97,* 1145–1148.

Brunelle, M., Nguyễn, D. D., & Nguyễn, K. H. (2010). A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica, 67,* 147–169. doi:10.1159/000321053

Edmondson, J. A., Chang, Y., Hsieh, F., & Huang, H. J. (2011). Reinforcing voiceless finals in Taiwanese and Hakka: Laryngoscopic case studies. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 627–630). Hong Kong, People's Republic of China: ICPhS Organizing Committee.

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology, 23,* 157–191. doi:10.1017/S095267570600087X

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association, 26,* 65–88.

Esling, J. H., Fraser, K. E., & Harris, J. G. (2005). Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics, 33,* 383–410. doi:10.1016/j.wocn.2005.01.003

Esling, J. H., & Harris, J. G. (2003). An expanded taxonomy of states of the glottis. In J. Romero (Ed.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1049–1052). Barcelona, Spain: ICPhS Organizing Committee.

Esling, J. H., & Harris, J. G. (2005). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. J. Hardcastle & J. M. Beck (Eds.), *A figure of speech: A Festschrift for John Laver* (pp. 347–383). Mahwah, NJ: Erlbaum.

Esling, J. H., Zeroual, C., & Crevier-Buchman, L. (2007). A study of muscular synergies at the glottal, ventricular and aryepiglottic levels. *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 585–588). Saarbrücken, Germany: ICPhS.

Eysholdt, U., Rosanowski, F., & Hoppe, U. (2003). Vocal fold vibration irregularities caused by different types of laryngeal asymmetry. *European Archives of Oto-Rhino-Laryngology, 260,* 412–417. doi:10.1007/s00405-003-0606-y

Fink, B. R. (1974). Folding mechanism of the human larynx. *Acta Oto-Laryngologica, 78,* 124–128.

Fletcher, N. H. (1993). Autonomous vibration of simple pressure-controlled valves in gas flows. *The Journal of the Acoustical Society of America, 93,* 2172–2180.

Fujimura, O. (1981). Body-cover theory of the vocal fold and its phonetic implications. In K. N. Stevens & M. Hirano (Eds.), *Vocal fold physiology* (pp. 271–281). Tokyo, Japan: University of Tokyo.

Garellek, M. (2013). *Production and perception of glottal stops* (Unpublished doctoral dissertation). University of California, Los Angeles.

Gerratt, B. R., & Kreiman, J. (2001). Toward a taxonomy of nonmodal phonation. *Journal of Phonetics, 29,* 365–381.

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics, 29,* 383–406.

Hess, M. M., Verdolini, K., Bierhals, W., Mansmann, U., & Gross, M. (1998). Endolaryngeal contact pressures. *Journal of Voice, 12,* 50–67.

Hillel, A. D. (2001). The study of laryngeal muscle activity in normal human subjects and in patients with laryngeal dystonia using multiple fine-wire electromyography. *The Laryngoscope, 111* (s97), 17–47.

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

Hirano, M., & Ohala, J. (1969). Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. *Journal of Speech, Language, and Hearing Research, 12,* 362–373.

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

Honda, K., Kitamura, T., Takemoto, H., Adachi, S., Mokhtari, P., Takano, S., … Dang, J. (2010). Visualisation of hypopharyngeal cavities and vocal-tract acoustic modelling. *Computer Methods in Biomechanics and Biomedical Engineering, 13,* 443–453. doi:10.1080/10255842.2010.490528

Imagawa, H., Sakakibara, K.-I., Tayama, N., & Niimi, S. (2003). The effect of the hypopharyngeal and supra-glottic shapes on the

singing voice. In R. Bresin & A. Askenfelt (Eds.), *Proceedings of the Stockholm Music Acoustics Conference* (Vol. 3, pp. 471–474). Stockholm, Sweden: Logos Verlag.

Ishizaka, K., & Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal, 51,* 1233–1268.

Iwata, R., Sawashima, M., Hirose, H., & Niimi, S. (1979). Laryngeal adjustment of Fukienese stops: Initial plosives and final applosives. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, 13,* 61–81.

Kimura, M., Sakakibara, K.-I., Imagawa, H., Chan, R., Niimi, S., & Tayama, N. (2002). Histological investigation of the supraglottal structures in humans for understanding abnormal phonation. *The Journal of the Acoustical Society of America, 112,* 2446.

Koizumi, T., & Taniguchi, S. (1990). A novel model of pathological vocal cords and its application to the diagnosis of vocal cord polyp. *Proceedings of the International Conference on Spoken Language Processing* (pp. 73–76). Kobe: Acoustical Society of Japan.

Laver, J. (1975). *Individual features in voice quality* (Unpublished doctoral dissertation). University of Edinburgh, Scotland.

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

Lindestad, P.-Å., Södersten, M., Merker, B., & Granqvist, S. (2001). Voice source characteristics in Mongolian "throat singing" studied with high-speed imaging technique, acoustic spectra, and inverse filtering. *Journal of Voice, 15,* 78–85.

Lindqvist-Gauffin, J. (1972). A descriptive model of laryngeal articulation in speech. *Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology, 13*(2–3), 1–9.

McGowan, R. S. (1992). Tongue-tip trills and vocal-tract wall compliance. *The Journal of the Acoustical Society of America, 91,* 2903–2910.

Moisik, S. R., Esling, J. H., Bird, S., & Lin, H. (2011). Evaluating laryngeal ultrasound to study larynx state and height. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 136–139). Hong Kong: People's Republic of China: ICPhS Organizing Committee.

Moisik, S. R., Esling, J. H., & Crevier-Buchman, L. (2010). A high-speed laryngoscopic investigation of aryepiglottic trilling. *The Journal of the Acoustical Society of America, 127,* 1548–1558. doi:10.1121/1.3299203

Painter, C. (1986). The laryngeal vestibule and voice quality. *Archives of Oto-Rhino-Laryngology, 243,* 329–337.

Reidenbach, M. M. (1997). Anatomical considerations of closure of the laryngeal vestibule during swallowing. *European Archives of Oto-Rhino-Laryngology, 254,* 410–412.

Reidenbach, M. M. (1998a). The muscular tissue of the vestibular folds of the larynx. *European Archives of Oto-Rhino-Laryngology, 255,* 365–367.

Reidenbach, M. M. (1998b). Aryepiglottic fold: Normal topography and clinical implications. *Clinical Anatomy, 11,* 223–235.

Sakakibara, K.-I., Fuks, L., Imagawa, H., & Tayama, N. (2004). Growl voice in ethnic and pop styles. In D. Naganuma (Ed.), *Proceedings of the International Symposium on Musical Acoustics.* Nara, Japan: University of Hokkaido. Available online at www.hoku-iryo-u.ac.jp/~kis/paper/isma04.pdf

Sakakibara, K.-I., Kimura, M., Imagawa, H., Niimi, S., & Tayama, N. (2004, August). *Physiological study of the supraglottal structure.* Paper presented at the International Conference on Voice Physiology and Biomechanics, Marseille, France.

Stevens, K. N. (1998). *Acoustic phonetics.* Cambridge, MA: MIT Press.

Story, B. H. (1995). *Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract* (Unpublished doctoral dissertation). The University of Iowa, Iowa City.

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

Story, B. H., & Titze, I. R. (1995). Voice simulation with a body-cover model of the vocal folds. *The Journal of the Acoustical Society of America, 97,* 1249–1260.

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

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

Titze, I. R. (2008). Nonlinear source–filter coupling in phonation: Theory. *The Journal of the Acoustical Society of America, 123,* 2733. doi:10.1121/1.2832337

Tokuda, T. I., Horáček, J., Švec, J. G., & Herzel, H. (2007). Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments. *The Journal of the Acoustical Society of America, 122,* 519–531.

Tsai, C.-G., Wang, L.-C., Wang, S.-F., Shau, Y.-W., Hsiao, T.-Y., & Auhagen, W. (2010). Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception, 27,* 209–222. doi:10.1525/mp.2010.27.3.209

van den Berg, J. W. (1955). On the role of the laryngeal ventricle in voice production. *Folia Phoniatrica, 7,* 57–69.

van den Berg, J. W., Zantema, J. T., & Doornenbal, P. J. (1957). On the air resistance and the Bernoulli effect of the human larynx. *The Journal of the Acoustical Society of America, 29,* 626–631.

Voigt, D., Döllinger, M., Braunschweig, T., Yang, A., Eysholdt, U., & Lohscheller, J. (2010). Classification of functional voice disorders based on phonovibrograms. *Artificial Intelligence in Medicine, 49,* 51–59.

Yanagisawa, E., Estill, J., Kmucha, S. T., & Leder, S. (1989). The contribution of aryepiglottic constriction to "ringing" voice quality: A videolaryngoscopic study with acoustic analysis. *Journal of Voice, 3,* 342–350. doi:10.1016/S0892-1997(89)80057-8

Zemlin, W. R. (1998). *Speech and hearing science: Anatomy and physiology* (4th ed.). Boston, MA: Allyn & Bacon.

Equations governing the mechanical motion of the system and the aero-acoustic state are sufficiently documented in other sources (Birkholz, 2005; Ishizaka & Flanagan, 1972; Story & Titze, 1995) and have not been repeated here. Numerical solution to the equations of motion for the mechanical model was obtained using the 4th order Runge-Kutta (RK4) method with a time step of $10^{-4}$ s. Test simulations were conducted at $10^{-4}$ s and $5 \times 10^{-5}$ s. The signal corresponding to the smaller time step had a slightly diminished amplitude and a slight phase delay, but the overall behavior of vibration remained consistent, and the glottal flow signals were 94.06% correlated. Furthermore, a previous version of this model constructed using an entirely different aero-acoustic simulation (Titze, 2006) produced very similar results to those presented here. Following Birkholz (2005), solving the system of equations for the aero-acoustic simulation was achieved via successive overrelaxation ($\omega = 1.25$). A trapezoid-rule-based scheme was used to approximate the derivative and integral functions of the pressure and flow state variables (with $\theta = 0.53$; see Birkholz, 2005). The mechanical model is bilateral, but only symmetrical behavior was simulated. As a convention, all signals were obtained from the centroid locations of the right-hand side masses.

Parameters for VVFC springs and damping ($k_{vb}$, $k_{vu}$, $d_{vb}$, and $d_{vu}$) were manipulated according to the needs of the particular simulation. Stiffness ranged from 0 $Nm^{-1}$ up to 3 times the value of the primary spring controlling the mass to which the representative spring was coupled (following the definition of collision springs in the ST95 model). For the vocal folds, the damping ratios were held constant ($\zeta_u$ was 0.4, $\zeta_l$ was 0.4, and $\zeta_b$ was 0.2, following the specifications of the ST95 model[6]), and $k_c$ (coupling between $m_u$ and $m_l$) was always 2 $Nm^{-1}$. Other parameters were manipulated according to various simulation objectives. In practice, values in between this range were used which allowed self-sustaining oscillation to occur (see Appendix C for specific settings). Damping was typically determined by setting $\zeta_{vb} = 0.2$ and $\zeta_{vb} = 0.4$, although values as low as 0 and as high as 1.6 were used (again, based on whether self-sustaining oscillation could be achieved or not). For no VVFC, or to "turn off" coupling, stiffness and damping were both nullified.

Ventricular fold positioning during modal conditions was set to match the Story (1995) data for the first supraglottal tube section. Under normal conditions (nonsinging and nonpathological), the lateromedial distance separating the medial edges of the ventricular folds (the ventricular fold "gap") is on the order of 0.5 cm (Agarwal et al., 2003). The area-function values presented by Story (1995) and used in the VVFC model are consistent with this. For constricted conditions, the center of $m_v$ was set to be 2.3 mm from the glottal midline, which yields an initial ventricular fold separation distance of 0.6 mm (corresponding to an area of 0.28 $mm^2$ for the tube section defined by the ventricular masses in the aero-acoustic model). This narrow initial cross-sectional area reflects laryngoscopic data of creaky and harsh phonation (Esling & Harris, 2005; Esling & Moisik, 2012), where ventricular fold medialization is so strong at the onset of phonation that the folds appear to be touching at their medial edges. In these data, the lateromedial distance between the ventricular folds increases somewhat during the course of phonation, but often the folds remain partially in contact, especially along their anterior extent.

The aero-acoustic model used in the simulations closely follows the formulation put forth by Birkholz (2005), but the exact configuration was modified for the VVFC model. Two vowel configurations were used, [i] and [ɑ]. The area functions for these vowels were obtained from MRI-based data presented in Story (1995). The supraglottal vocal tract consisted of 35 tube sections and the subglottal tract had 20 sections. All tube sections in the model were 0.5 cm long, except for the two sections constituting the lower and upper glottis, which were 0.2 cm in length. The supraglottal vocal tract was therefore 17.5 cm long. The laryngeal ventricle, piriform fossae, and nasal cavity were not modeled. Lung pressure was set according to the target of the simulation: Generally, low values (200–400 Pa) were used to simulate creaky voice and higher values were used to simulate harsh voice (1000 Pa); in other cases, a default of 800 Pa was used.

The impedance of the basic tube sections in the transmission line model was calculated using the specifications outlined in Birkholz (2005): Each tube section had components for air and wall impedance. However, the glottal and ventricular sections were treated differently. Because the mechanical models of the vocal folds and ventricular folds explicitly model the wall impedance in the glottal and ventricular tube sections, respectively, no wall impedance components were used in the transmission line for these sections. Basic tube sections implement a frictional resistance based on the calculation of airflow resistance in circular tubes under steady laminar flow. A "slit resistance" was used in the glottal and ventricular sections. This modeled the resistance to steady, laminar flow through a narrow slit (modeled as an ellipse with high eccentricity). It was intended to represent the more triangular rather than circular cross-sectional shape of these sections (Ishizaka & Flanagan, 1972; van den Berg, Zantema, & Doornenbal, 1957; see Birkholz, 2005).

---

[6] $\zeta$ is related to damping coefficients by the equation $d = 2\zeta(mk)^{1/2}$.

## Appendix C

Parameters for Trials and Case Studies

Parameters used for particular simulation trials, and the case studies are presented in Tables A1 and A2.

**Table A1.** Biomechanical parameter ranges.

| Figure | Test condition | Vocal fold | | | | | | Ventricular fold | | Vocal–ventricular fold coupling | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $k_u$ | $k_l$ | $k_b$ | $m_u$ | $m_l$ | $m_b$ | $k_v$ | $m_v$ | $k_{vb}$ | $k_{vu}$ | $\zeta_{vb}$ | $\zeta_{vu}$ |
| 3 | N | 30 | 80 | 100 | 50 | 50 | 100 | - | - | - | - | - | - |
| | SS | 20 | 100 | 400 | 50 | 50 | 100 | - | - | - | - | - | - |
| | N, CM, ±VB | 30 | 80 | 100 | 50 | 50 | 100 | 50 | 260 | 0/0.5 | 30 | 0/0.2 | 0.4 |
| | N, CO, ±VB | 30 | 80 | 100 | 50 | 50 | 100 | 50 | 260 | 0/0.5 | 30 | 0/0.2 | 0.4 |
| | SS, CM, ±VB | 20 | 100 | 400 | 50 | 50 | 100 | 50 | 260 | 0/0.5 | 0.5 | 0/0.2 | 0.4 |
| | SS, CO, ±VB | 20 | 100 | 400 | 50 | 50 | 100 | 50 | 260 | 0/0.5 | 0.5 | 0/0.2 | 0.4 |
| 4 | N, CM sweep | 30 | 100 | 100 | 50 | 50 | 100 | 0.5 | 260 | max 0.5 | max 30 | max 0.2 | max 0.4 |
| | N, CO sweep | 30 | 100 | 100 | 50 | 50 | 100 | 2.5 | 260 | max 0.5 | max 60 | max 0.2 | max 1.6 |

*Note.* Units: mg for mass, Nm$^{-1}$ for stiffness. N = normal; SS = stiff-slack; CM = collision-and-mucous-type springs; CO = collision-only-type springs; VB = $m_{vb}$ to $m_b$ coupling.

**Table A2.** Specific biomechanical parameters used in the case studies.

| # | Test condition | Vowel context | $p_s$ | Vocal fold | | | | | | Ventricular fold | | Vocal–ventricular fold coupling | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $k_u$ | $k_l$ | $k_b$ | $m_u$ | $m_l$ | $m_b$ | $k_v$ | $m_v$ | $k_{vb}$ | $k_{vu}$ | $\zeta_{vb}$ | $\zeta_{vu}$ |
| 1 | N, CM, VB | [ɑ] | 800 | 15 | 50 | 100 | 50 | 50 | 100 | 0.5 | 260 | 0.5 | 24 | 0.2 | 0.4 |
| 2 | SS, CM, VB | [i] | 200 | 4 | 30 | 400 | 20 | 25 | 200 | 0.5 | 260 | 0.5 | 1.5 | 0 | 0 |
| 3 | SS, CO | [ɑ] | 1000 | 20 | 150 | 400 | 6 | 15 | 50 | 50 | 260 | 0 | 60 | 0 | 0 |
| 4 | SS, CO, VB | [ɑ] | 400 | 4 | 100 | 400 | 80 | 80 | 300 | 50 | 260 | 0.5 | 5 | 0.2 | 0.4 |
| 5 | N → SS, CO, VB | [ɑ] | 800 | 15 → 4 | 100 → 25 | 100 → 400 | 50 | 50 | 100 | 50 | 260 | 0 → 2 | 0 → 24 | 0 → 0.2 | 0 → 0.4 |

*Note.* Units: mg for mass, Nm$^{-1}$ for stiffness, Pa for pressure. # = case study number; → = "transitions to."

**S704** *Journal of Speech, Language, and Hearing Research* • Vol. 57 • S687–S704 • April 2014