



Exploring the mechanics of fundamental frequency variation during phonation onset

Mohamed A. Serry¹ · Cara E. Stepp² · Sean D. Peterson¹

Received: 25 March 2022 / Accepted: 20 October 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Fundamental frequency patterns during phonation onset have received renewed interest due to their promising application in objective classification of normal and pathological voices. However, the associated underlying mechanisms producing the wide array of patterns observed in different phonetic contexts are not yet fully understood. Herein, we employ theoretical and numerical analyses in an effort to elucidate the potential mechanisms driving opposing frequency patterns for initial/isolated vowels versus vowels preceded by voiceless consonants. Utilizing deterministic lumped-mass oscillator models of the vocal folds, we systematically explore the roles of collision and muscle activation in the dynamics of phonation onset. We find that an increasing trend in fundamental frequency, as observed for initial/isolated vowels, arises naturally through a progressive increase in system stiffness as collision intensifies as onset progresses, without the need for time-varying vocal fold tension or changes in aerodynamic loading. In contrast, reduction in cricothyroid muscle activation during onset is required to generate the decrease in fundamental frequency observed for vowels preceded by voiceless consonants. For such phonetic contexts, our analysis shows that the magnitude of reduction in the cricothyroid muscle activation and the activation level of the thyroarytenoid muscle are potential factors underlying observed differences in (relative) fundamental frequency between speakers with healthy and hyperfunctional voices. This work highlights the roles of sometimes competing laryngeal factors in producing the complex array of observed fundamental frequency patterns during phonation onset.

Keywords Phonation onset · Vocal folds · Collision · Laryngeal muscle activation · Relative fundamental frequency

1 Introduction

Phonation initiation is a highly complex phenomenon with laryngeal maneuvers that position and stiffen the vocal folds (VFs), leading to self-sustained oscillations driven by the lung pressure. The transient oscillatory dynamics of the VFs from the rest prephonatory position to sustained vibrations are referred to herein as phonation onset (Mergell et al. 1998; Lebacqz and DeJonckere 2019)¹. As a fundamental aspect of

voiced speech, phonation onset has been studied for decades (Lisker and Abramson 1967; Mohr 1971; Ohde 1984; Titze 1988; Löfqvist et al. 1989; Mergell et al. 1998; Hanson 2009; Zhang 2011; Sváček and Horáček 2018; DeJonckere and Lebacqz 2020; Azar and Chhetri 2022). In the past ten years or so, the fundamental frequency patterns during phonation onset have received renewed attention as they have been found to differ between healthy and pathological voices, enabling development of a practical and useful classification tool based upon relative fundamental frequency (Goberman and Blomgren 2008; Stepp et al. 2010a, 2011; Roy et al. 2016; Heller Murray et al. 2017). Fundamental frequency characteristics during transient periods of phonation, including phonation onset, correlate with kinematic vocal fold stiffness, a measure of laryngeal stiffness (Stepp et al. 2010b), and such a correlation can be a useful clinical

✉ Sean D. Peterson
peterson@uwaterloo.ca

Mohamed A. Serry
mserry@uwaterloo.ca

Cara E. Stepp
cstepp@bu.edu

¹ Department of Mechanical and Mechatronics Engineering,
University of Waterloo, Waterloo, ON N2L 3G1, Canada

² Department of Speech, Language and Hearing Sciences,
Boston University, Boston, MA 02215, USA

¹ We note that this definition differs from that of voice onset time, which considers onset as the initiation of VF oscillations, see, for example, Lisker and Abramson (1967).

indicator of laryngeal tension (McKenna et al. 2016; Park et al. 2021).

Phonation initiation exhibits a variety of fundamental frequency patterns depending on phonetic context and vocal health, as shown schematically in Fig. 1. In the case of isolated and initial vowels, fundamental frequency typically exhibits a gradual increase until a sustained phonation frequency is attained (Mohr 1971; Smith and Robb 2013). When the vowel is preceded by a voiceless consonant, as in /pa/, an initial spike in fundamental frequency followed by a gradual decay is observed (Ohde 1984; Löfqvist et al. 1989), where the onset period has been found to be dependent on the language and the acoustic nature of the voiceless consonant (unaspirated vs. aspirated) (Francis et al. 2006). With such gestures, speakers with healthy voices exhibit higher initial (relative) fundamental frequency values compared to speakers with vocal hyperfunction² (Stepp et al. 2010a). When the vowel is preceded by a voiced consonant, as in /ba/, there is less agreement in the literature regarding the temporal evolution of fundamental frequency, with some studies observing a gradual increase (Mohr 1971; Hombert et al. 1979), but others finding inconsistent patterns between and within speakers (Ohde 1984). Moreover, it has been found that fundamental frequency patterns in the case of vowels preceded by voiced consonants are

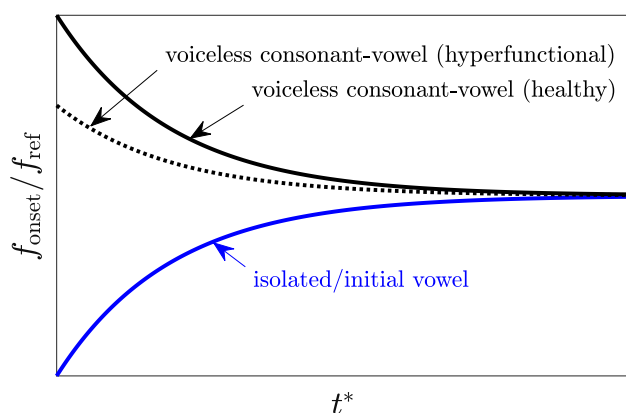


Fig. 1 (Color online) Schematic representation of onset fundamental frequency f_{onset} , normalized by reference fundamental frequency f_{ref} [e.g., fundamental frequency of the tenth onset cycle (Stepp et al. 2010a)] as a function of normalized time t^* [e.g., number of onset cycles (Stepp et al. 2010a)] for different phonetic contexts and vocal health. The depicted trends are consistent with the patterns observed experimentally in the literature (Stepp et al. 2010a; Ohde 1984; Löfqvist et al. 1989; Mohr 1971; Hombert et al. 1979; Stepp et al. 2011; Heller Murray et al. 2017)

² Vocal hyperfunction (VH) is a class of voice disorders characterized by the overuse/misuse of the vocal mechanism (Hillman et al. 1989, 2020). Based on the presence or lack of concurrent pathology, VH can be classified in two categories: phonotraumatic VH and non-phonotraumatic VH (Hillman et al. 2020).

context-dependent (Hanson 2009; Kirby and Ladd 2016). Regardless, empirical evidence suggests that onset frequency in vowels preceded by voiceless consonants is higher than that in vowels preceded by voiced consonants (Ohde 1984).

Several underlying factors have been hypothesized to drive the observed fundamental frequency patterns during phonation initiation, including laryngeal muscle tension, aerodynamics, and vocal fold contact³. Smith and Robb (2013) empirically investigated onset fundamental frequency patterns of vowels preceded by fricatives and stop consonants, in addition to isolated vowels. They speculated that the rise of onset fundamental frequency in the case of isolated vowels is due to a rise in VF tension. They further suggested that laryngeal muscle tension is a predominant factor in the case of vowels preceded by voiceless consonants. Löfqvist et al. (1989) investigated cricothyroid muscle activation during phonation onset using electromyography and found correlation between increased cricothyroid muscle activation and the higher fundamental frequency observed during the onset of vowels preceded by voiceless consonants. Löfqvist et al. (1995) estimated the glottal flow characteristics from oral flow measurements and found that peak glottal flow is higher in vowels preceded by voiceless consonants in comparison with voiced consonants, indicating a correlation with the observed higher initial fundamental frequency. Moreover, Löfqvist et al. (1995) found that the glottal flow characteristics differ between aspirated voiceless consonants and their unaspirated counterparts.

In addition to clinical studies, there have been several theoretical and numerical investigations attempting to elucidate the underlying mechanisms of fundamental frequency during transient phonation, including phonation onset. Ishizaka and Flanagan (1972) employed a two-mass numerical vocal fold model to explore the mechanics of voiced speech and noted the potential role of vocal fold contact in altering fundamental frequency during phonation onset. Titze (1988) studied theoretically the onset conditions of VF oscillations using a single-mass model, showing that aerodynamics change the equivalent oscillator stiffness, and consequently, the fundamental frequency of the VF system during phonation onset. Zhang (2009) extended this analysis using a continuum two-layered model, noting that under certain conditions, slight changes in VF geometry or stiffness can cause sudden changes in onset fundamental frequency. Serry et al. (2021), in a study of phonation offset using a simple impact oscillator model, demonstrated that increased

³ For a thorough discussion of some of the mechanisms through which aerodynamics and muscle tension are hypothesized to alter onset fundamental frequency, see the seminal work of Hombert et al. (1979).

collision duration results in higher fundamental frequency, which is expected to play a similar role during phonation onset.

As illustrated by Fig. 1, different fundamental frequency characteristics can arise through manipulation of the phonetic context, suggesting potentially complex interrelations between the contributing factors. It can be quite challenging to isolate and control individual factors, such as aerodynamics and laryngeal muscle activation, during phonation onset in studies with human participants. As such, in this paper we aim to investigate, by means of theoretical and numerical analyses, some of the underlying mechanisms leading to the disparate fundamental frequency behaviors depicted in Fig. 1.

In particular, we investigate the dynamic nature of fundamental frequency during phonation onset by extending the impact oscillator model introduced in Serry et al. (2021). This dynamic nature is due, in part, to elevating collision levels of the vocal folds during phonation onset. The theoretical analysis is then verified using the physiologically relevant three-mass body-cover model (Story and Titze 1995). Results from the aforementioned theoretical analysis are capable of predicting the fundamental frequency rise pattern displayed in Fig. 1, which implies collision as a potential underlying mechanism. Subsequently, we explore numerically how laryngeal muscle activation and its temporal variation can underlie the fundamental frequency drop patterns observed during voicing of vowels preceded by voiceless consonants, see Fig. 1. Finally, we attempt to investigate some of the laryngeal mechanisms that can potentially underlie the differences between healthy and hyperfunctional voices.

The organization of the paper is as follows: in Sect. 2, we introduce the employed phonation models; the role of VF collision is discussed in Sect. 3; the influence of the cricothyroid and thyroarytenoid muscles during phonation onset is explored in Sect. 4; and Sect. 5 concludes the manuscript.

2 Phonation models

In this section, we introduce the phonation models used in our analyses. The first is a hybrid phonation model that integrates the impact oscillator model introduced by Serry et al. (2021) and a linearized version of the Titze (1988) single-mass model, which is used to explore the role of collision during phonation onset. The second is a body-cover reduced-order model (Story and Titze 1995) used to explore the role of muscle activation and corroborate findings from the hybrid model with a more physiologically relevant VF description.

Similar to Mergell et al. (1998), it will be assumed, unless otherwise stated, that the neutral prephonatory gap between

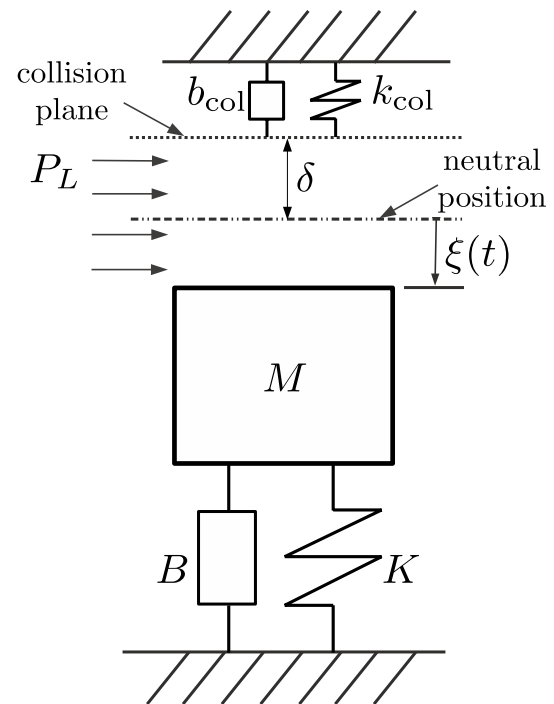


Fig. 2 Schematic diagram of the hybrid phonation model

the VFs is fixed during onset. In the case of isolated vowels, this assumption is supported by empirical data, showing that VF oscillations are initiated from a fixed prephonatory neutral position (Shiba and Chhetri 2016). In the case of vowels preceded by fricatives, onset has been found to start slightly before reaching the final prephonatory position (McKenna et al. 2016; Patel et al. 2017). Moreover, for simplicity we neglect temporal variations in aerodynamic and acoustic parameters, such as the acoustic impedance at the mouth. Variations in such parameters are believed to play roles in altering fundamental frequency during transient periods of phonation (Hombert et al. 1979); however, their significance in comparison with VF contact and laryngeal muscle tension is the subject of some debate. Smith and Robb (2013), for instance, found that onset fundamental frequency patterns of vowels preceded by fricatives and stop consonants are very similar despite their differing aerodynamic characteristics, implying a minor role of aerodynamics. On the contrary, the empirical data of relative fundamental frequency presented in Lien et al. (2014) and Park et al. (2021) suggests that these factors may be prevalent. Herein, we focus on collision and muscle activation, leaving a comprehensive exploration of aerodynamics and acoustics for future work.

2.1 Hybrid phonation model

The hybrid model is shown schematically in Fig. 2. It enables analysis of fluid–structure interactions during phonation

onset, where the glottal flow is modeled using a linearized Bernoulli flow model, while incorporating the role of VF collision. The governing equations are

$$M\ddot{\xi} - \mathcal{B}_1\dot{\xi} + K\xi = 0, \quad \xi(t) \geq -\delta, \quad (1a)$$

$$M\ddot{\xi} + \mathcal{B}_2\dot{\xi} + K\xi = -k_{\text{col}}\delta, \quad \xi(t) < -\delta, \quad (1b)$$

where $\xi(t)$ is the VF mass displacement from its neutral position, M is its mass, K is the tissue stiffness, k_{col} is collision stiffness, $\delta \geq 0$ is the neutral gap, and $\mathcal{K} = K + k_{\text{col}}$. The damping terms are given by $\mathcal{B}_1 = 2\tau P_L/(k_t\delta) - B$ and $\mathcal{B}_2 = B + b_{\text{col}}$, where B is the structural viscous damping coefficient, b_{col} is an additional damping coefficient incorporated during collision, P_L is the subglottal lung pressure, τ is a time delay term associated with the propagation of the mucosal wave on the medial surface of the VFs, and k_t is a pressure recovery term (Titze 1988). The mass, stiffness, and damping coefficients are given per unit area. The neutral gap, δ , serves as a proxy for degree of VF adduction, such that $\delta = 0$ corresponds to complete VF closure. This model neglects acoustic effects and assumes negligible supraglottal pressure; hence, P_L is correspondent to the transglottal pressure. It is assumed that the dynamics of the hybrid phonation model are oscillatory in both collision and non-collision regimes, that is,

$$\omega_1^2 := \frac{K}{M} - \frac{\mathcal{B}_1^2}{4M^2} > 0, \quad (2a)$$

$$\omega_2^2 := \frac{\mathcal{K}}{M} - \frac{\mathcal{B}_2^2}{4M^2} > 0, \quad (2b)$$

where $\omega_1, \omega_2 \geq 0$ denote the angular frequencies in the non-collision and collision regimes, respectively.

The impact oscillator model of Serry et al. (2021), referred to herein as the S21 model, can be recovered from Eq. (1) by omitting the viscous forces (i.e., by setting $\mathcal{B}_1 = \mathcal{B}_2 = 0$), resulting in

$$M\ddot{\xi} + K\xi = 0, \quad \xi(t) \geq -\delta, \quad (3a)$$

$$M\ddot{\xi} + K\xi = -k_{\text{col}}\delta, \quad \xi(t) < -\delta. \quad (3b)$$

This model isolates the effects of collision and primitive parameters (e.g., mass, stiffness, and neutral gap) on fundamental frequency, providing an abstract, yet useful, insight into the role of VF contact during real phonation scenarios.

The linearized version of the Titze (1988) model can be recovered from the hybrid phonation model by assuming collision-free oscillations, that is, $\xi(t) > -\delta$, yielding⁴

$$M\ddot{\xi} - \mathcal{B}_1\dot{\xi} + K\xi = 0. \quad (4)$$

Equation (4) provides useful insights into the fluid–structure interaction between the VFs and the glottal flow during phonation onset and, in particular, the role of aerodynamics (in the form of negative damping that results from linearizing the Bernoulli flow model) in initiating VF oscillations. The onset conditions predicted from Eq. (4) [see Eq. (6)] agree reasonably with experimental measurements from a physical model of the VF mucosa; in particular, phonation threshold pressure⁵ is positively correlated with the neutral gap δ (for sufficiently large δ) and the VF viscous damping coefficient B (Titze et al. 1995).

2.2 Body-cover model

The reduced-order three-mass body-cover model (BCM) (Story and Titze 1995) is employed to verify and extend the findings from the simpler, more analytically tractable, hybrid phonation model⁶. This model, which embeds the essential physiological components of the VFs, consists of two cover masses and a body mass, all connected via springs and dampers to model the VFs viscoelastic tissues (see Fig. 13 in Appendix 1). The model assumes the motion of the VFs to be symmetric about the medial plane; hence, only one of the folds is needed in the model construction. Collision of the opposing folds is modeled by activating additional nonlinear spring forces applied to the cover masses, where the spring forces are proportional to the degree of overlap of the cover masses with the medial (collision) plane, see Equations (6a) and (6b) in Story and Titze (1995)⁷. The model implements

⁴ Equation (4) can be obtained by linearizing Equation (8) in Lucero and Koenig (2007) at the point of zero displacement (from the neutral position) and velocity.

⁵ Phonation threshold pressure is the minimum subglottal pressure needed to initiate phonation (Titze 1988). It can also refer to the minimum subglottal pressure required to maintain VF oscillations (Titze et al. 1995).

⁶ The literature is rich in various phonation models that can be used in lieu of the BCM, including refined lumped-element (Galindo et al. 2017; Alzamendi et al. 2020) and high-fidelity models (Geng et al. 2021; Movahhedi et al. 2021). The BCM has been selected herein for its reasonable computational requirements and demonstrated capability to capture the essential physics of phonation in various studies, see for example Zaňartu et al. (2014); Serry et al. (2021); Deng et al. (2019, 2022); Titze (2004); Lowell and Story (2006); Gómez-Vilda et al. (2007).

⁷ This is essentially the same collision model as for the hybrid phonation model. Both models neglect adhesive forces during collision, which can influence VF biomechanics (Bhattacharya and Siegmund 2015).

the muscle activation rules of Titze and Story (2002) to control the primitive model variables via three dimensionless muscle activation parameters, a_{CT} , a_{TA} , and a_{LCA} , which account for the relative activation of the cricothyroid (CT), thyroarytenoid (TA), and lateral/posterior cricoarytenoid (LCA/PCA) muscles, respectively. The neutral glottal gap in the BCM is modulated through activation of the LCA muscle, where the neutral glottal half width x_0 is given by $x_0 = 0.25L_0(1 - 2a_{LCA})$, where L_0 is the resting VF length (Titze and Story 2002). As seen from this relation, LCA activation is negatively correlated with the VF neutral gap, where increasing the activation of the LCA muscle leads to adducting the VFs.

Air flow through the glottis is modeled using a quasi-steady Bernoulli flow formulation with quasi-steady viscous correction for losses in the glottis (Pelorson et al. 1994; Lucero 1996; Lucero and Schoentgen 2015). The quasi-steady viscous model has shown good agreement with experimental observations of flow through a larynx model (Van den Berg et al. 1957). We note herein that our glottal flow model is similar, but not identical, to those presented in Pelorson et al. (1994); Lucero and Schoentgen (2015) as it incorporates flow separation and its formulation is suitable for modelling acoustic effects due to subglottal and supraglottal tracts. Viscous corrections are employed to account for non-negligible losses that occur during the initial stages of phonation when the flow speeds are relatively slow and during periods when the glottis is nearly closed (Fulcher et al. 2013). See Appendix 1 for further details on the employed flow model.

Acoustics are modeled using the wave reflection analog (WRA) method (Kelly and Lochbaum 1962; Liljencrants 1985; Story 2005). Similar to Galindo et al. (2014) and Zañartu et al. (2014), a subglottal tract area function is adapted from respiratory system measurements of human cadavers (Weibel et al. 1963), covering only the trachea and bronchi. A supraglottal tract is also included, which is configured to simulate the /i/ vowel (Takemoto et al. 2006). The BCM dynamics are driven by the lung pressure, P_1 , input to the inferior end of the subglottal tract. To mitigate numerical instabilities in the WRA implementation, P_1 is ramped up from zero to the desired value during phonation onset according to the relation $P_1(t) = P_{1,0}(1 - e^{-t/\sigma})$, where $P_{1,0}$ is the steady-state lung pressure and $\sigma = 0.2$ ms. The settling time for the ramp is less than 1 ms. The system dynamics are solved using an explicit version of Newmark's method (Newmark 1959; Galindo et al. 2014) with a sampling frequency of 140 kHz. Initial conditions in all BCM simulations are identical, with zero velocity for all masses and unstretched model springs. As in Serry et al. (2021), we consider the time-series of the glottal area, A_g , in our frequency analysis, where frequency is determined from the time duration between sequential signal peaks.

3 Relationship between collision and fundamental frequency

Serry et al. (2021) demonstrated a direct correlation between VF collision and fundamental frequency, wherein transitioning from a VF oscillation regime with collision to one without collision during phonation offset results in a decrease in fundamental frequency due to the net reduction in “system stiffness”. We posit that this mechanism is also a contributing factor underlying the temporal variation in fundamental frequency during phonation onset.

3.1 Insights from the S21 model

In this section, we summarize key findings from Serry et al. (2021), which investigated phonation offset, and expand the analysis therein to explore phonation onset. We note that despite the symmetries between phonation offset (decaying VF oscillations) and phonation onset (rising VF oscillations), there exist some notable differences between the two phenomena, including phonation threshold pressure values (Titze et al. 1995), and aerodynamic characteristics (depending on the phonetic context) (Löfqvist et al. 1995). Herein, we aim to utilize the symmetries between the two phenomena to elucidate the role of VF collision in altering fundamental frequency during phonation onset.

From Eq. (3a) (the S21 model without collision), we note that the natural frequency of the oscillator is $f_0 = \sqrt{K/M}/(2\pi)$. For convenience, we define the normalized neutral gap $\tilde{\delta} = \sqrt{K/(2E)}\delta$, where E is the energy of the VF system per unit area, which can loosely be considered as the energy originally imparted to the system via aerodynamics⁸. The system energy is constant-in-time owing to the lack of viscous losses, as can be seen from Eq. (3). We further define the stiffness ratio $\tilde{k} = K/\mathcal{K}$. The fundamental frequency of the model is (Serry et al. 2021)

$$f = \begin{cases} \frac{2f_0}{\frac{2}{\pi}\sqrt{\tilde{k}}\arctan\left(\sqrt{\left(\frac{1}{\tilde{\delta}^2}-1\right)/\tilde{k}}\right) + \frac{2}{\pi}\arcsin(\tilde{\delta})+1}, & \tilde{\delta} \leq 1, \\ f_0, & \tilde{\delta} > 1, \end{cases} \quad (5)$$

for which the behavior depends on whether or not the system has sufficient energy (vibration amplitude) to cause collision. Note that when $\tilde{\delta} > 1$ (no collision), frequency is independent of the oscillator energy. Utilizing a quasi-steady assumption with E as a parameter, we explore the effect of varying

⁸ As can be seen from Eq. (3), the S21 model does not incorporate aerodynamic effects. Stating that the system energy is imparted via aerodynamics is a crude assumption. In Sect. 3.2, we consider aerodynamic energy transfer during phonation onset more rigorously using the hybrid phonation model.

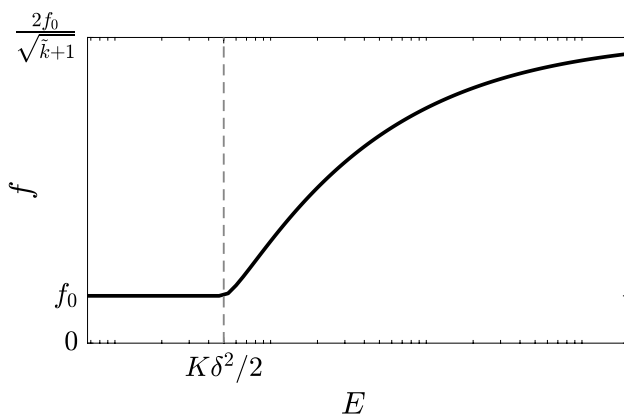


Fig. 3 Fundamental frequency of the S21 model as a function of the system energy. The dashed vertical line indicates the energy level at which collision initiates

the system energy on fundamental frequency. The role of energy, E , becomes evident when collision occurs ($\tilde{\delta} \leq 1$), wherein fundamental frequency increases as E increases, with an asymptotic value $2f_0/(\sqrt{k} + 1)$, corresponding to oscillations at zero neutral gap ($\delta = 0$) (see Fig. 3). The asymptotic behavior suggests that the collision-based mechanism is inefficient at changing frequency at high energy levels, as large energy increases result in modest gains in fundamental frequency.

In addition to system energy, collision is also modulated via VF adduction, which is embedded in the S21 model through the neutral gap, δ . Equation (5) shows that decreasing δ has a similar effect to increasing E (both lead to decreasing $\tilde{\delta}$). That is, for fixed system energy and stiffness, fundamental frequency can be increased purely through adduction. Similar to the energy rise, adduction only impacts fundamental frequency of the model when collision is present ($\tilde{\delta} \leq 1$). We note that the effect of increasing energy is mediated by the neutral gap, with a more muted response as the gap decreases. That is, the effectiveness of a rise in system energy at increasing frequency during phonation onset is dependent on the adduction level of the VFs. In reality, the exact relation is naturally expected to be complex due, in part, to the geometry of the glottis and the high degrees of freedom of the VFs.

Finally, the S21 model provides useful insights into the role of VF stiffness during phonation onset. From Eq. (3), we observe that K affects the dynamics of the VF system in both the collision and non-collision regimes, wherein increasing stiffness increases the (instantaneous) fundamental frequency. This indicates that changing VF stiffness during phonation onset through intrinsic muscle activation alters fundamental frequency even in the initial stage of onset in some phonetic contexts when VF oscillations are collision-free. Thus, we expect potentially competing factors of adduction,

aerodynamic energy transfer, and laryngeal tension to influence fundamental frequency during phonation onset.

3.2 Analysis using the hybrid phonation model

Analysis of the S21 model in the previous section relies on the quasi-steady assumption, wherein the dynamics of fluid–structure interaction during phonation onset and viscous friction losses are neglected. In this section, we consider an analytical treatment to the onset problem using the hybrid phonation model (see Sect. 2) and elucidate the dynamic nature of the collision-based mechanism. We consider the evolution of VF oscillations during onset while incorporating VF contact, which has been typically omitted in previous theoretical analyses of phonation onset (Titze 1988; Zhang 2009; Lucero and Koenig 2007).

Pre-collision, the hybrid model is equivalent to the linearized Titze (1988) model in Eq. (4), which predicts VF oscillations with exponential growth when [see, for example, Titze (1988)]

$$2\tau \frac{P_L}{k_i \delta} - B > 0. \quad (6)$$

However, realistic energy dynamics during phonation onset are complex and oscillatory due to several factors, including nonlinear fluid–structure interaction effects, VF collision, acoustics, and viscous losses. The primary energy transfer mechanism to the VF system is the temporal asymmetry of the average intraglottal pressure, where, loosely speaking, positive energy transfer from the glottal flow and energy dissipation to the flow take place when the VF configuration is convergent and divergent, respectively, with the total energy transferred from the flow being larger than that dissipated to the flow in order to sustain oscillations (Thomson et al. 2005). The hybrid model (Eq. 1) allows exploration of the general trends of the complex oscillatory energy dynamics beyond the initial onset of oscillations by incorporating simplified VF contact and aerodynamic effects.

We examine the energy evolution by considering the discrete system energy at the same phase in a sequence of oscillation cycles. Let τ_i , $i = 0, 1, 2, \dots$ be the time instances such that $\xi(\tau_i) = -\delta$ and $\dot{\xi}(\tau_i) < 0$, which correspond to the beginning of each collision. Let $\mathcal{V}_i = |\dot{\xi}(\tau_i)|$ be the oscillator velocity magnitude at time instance τ_i . The energy immediately prior to each collision (kinetic energy plus potential energy) is then $E(\tau_i) = M\mathcal{V}_i^2/2 + K\delta^2/2$. The velocity sequence $\{\mathcal{V}_i\}$ can be obtained approximately using the recurrence relation

$$\mathcal{V}_{i+1} = \mathcal{A}\mathcal{V}_i + \mathcal{W}, \quad i = 0, 1, 2, \dots, \quad (7)$$

where the initial velocity $\mathcal{V}_0 > 0$ is given. The parameter \mathcal{A} (a scaling term) is modulated by the energy losses and gains in the collision and collision-free regimes, respectively, and

the parameter \mathcal{W} (a drift term) is regulated by the neutral gap δ . Derivation of Eq. (7) and the exact definitions of \mathcal{A} and \mathcal{W} are provided in Appendix 2.

The dynamics of the recurrence relation given in Eq. (7) exhibit various behaviors depending on the numerical values of \mathcal{A} and \mathcal{W} (e.g., linear growth, exponential growth, and exponential decay). Herein, we are interested in cases where VF oscillations are bounded, thus corresponding to realistic phonation onset scenarios. On average, the aerodynamic energy transfer is larger than viscous losses after phonation initiation, which induces VF oscillations of growing amplitude. The (average) difference between aerodynamic energy transfer to the VF system and viscous dissipation gradually decays over time until the difference becomes zero, which corresponds to steady-state VF oscillations of constant amplitude (that is, sustained phonation).

The case of bounded energy growth can be determined from Eq. (7) when $\mathcal{W} \geq 0$ and $0 < \mathcal{A} < 1$, which is fulfilled when

$$0 < 2\tau \frac{P_L}{k_i \delta} - B < (B + b_{\text{col}}) \sqrt{k}. \quad (8)$$

This corresponds to the onset condition given in Eq. (6) under the additional constraint that the subglottal pressure is such that the damping ratio in the non-collision regime, $\mathcal{B}_1/(M\omega_1)$, is smaller than the damping ratio in the collision regime, $\mathcal{B}_2/(M\omega_2)$, to ensure VF oscillations of finite amplitude. As an example, if we set $k_i = 1.1$, $\delta = 10^{-3}$ m, $\tau = 1.5 \times 10^{-3}$ s, and $B = 2380$ Pa · s/m [similar to values used in Titze (1988) and Lucero (1996)] and additionally assume $b_{\text{col}} = 4B$ and $k_{\text{col}} = 3K$ [similar to assumptions in Steinecke and Herzel (1995)], Eq. (8) predicts that P_L should be within the approximate range [875, 2180] Pa, in order to have VF oscillations of bounded amplitude⁹. In this case ($\mathcal{W} \geq 0$ and $0 < \mathcal{A} < 1$), Eq. (7) can be rewritten as

$$\mathcal{V}_i = \frac{\mathcal{W}}{1 - \mathcal{A}} - \mathcal{A}^i \left(\frac{\mathcal{W}}{1 - \mathcal{A}} - \mathcal{V}_0 \right), \quad i = 1, 2, \dots \quad (9)$$

If

$$\mathcal{V}_0 < \frac{\mathcal{W}}{1 - \mathcal{A}}, \quad (10)$$

that is, the kinetic energy of the VFs is initially low, then the sequence $\{\mathcal{V}_i\}$ is monotonically increasing with an asymptotic upper bound $\mathcal{V}_\infty = \mathcal{W}/(1 - \mathcal{A})$ (see Fig. 4). This shows

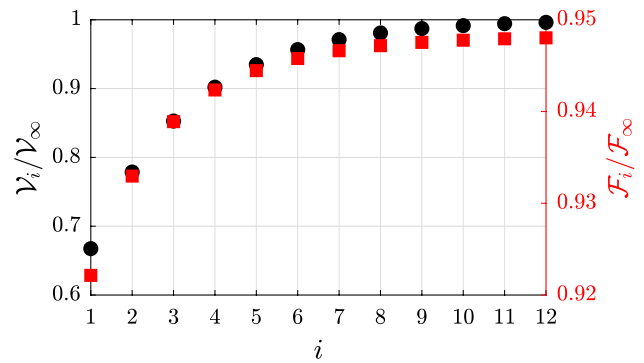


Fig. 4 (Color online) Illustrative example of the velocity (circles) and frequency (squares) sequences given in Eqs. (7) and (11), respectively, normalized with respect to the asymptotic upper bounds \mathcal{V}_∞ and \mathcal{F}_∞ , respectively, with fixed-in-time parameter values $M = \mathcal{B}_1 = K = 1$, $\mathcal{B}_2 = 2$, $\mathcal{K} = 3$, $\delta = 0.1$ and $\mathcal{V}_0 = 0.5\mathcal{W}/(1 - \mathcal{A})$. Note that, for the given parameter values, the conditions in Eqs. (8), (10), and (12) are fulfilled, which explains the monotonically increasing and bounded behaviors

that during phonation onset, the energy of the VF system increases gradually, on average, and achieves an asymptotic value, where aerodynamic energy transfer equals to viscous losses, which is associated with sustained phonation.

The frequency of the i th cycle, \mathcal{F}_i , is approximately given by

$$\mathcal{F}_i = \left[\frac{1}{\mathcal{F}_\infty} + \frac{\alpha_1}{\mathcal{V}_{i-1} + \beta} - \frac{\alpha_2}{\mathcal{V}_{i-1}} \right]^{-1}, \quad i = 1, 2, \dots \quad (11)$$

where $\mathcal{F}_\infty = ((1/\omega_1) + (1/\omega_2))^{-1}/\pi$ (see Appendix 3 for the derivation and the definitions of parameters α_1 , α_2 , and β). Note that if the sequence $\{\mathcal{V}_i\}$ is monotonically increasing, which is attained if the conditions in Eqs. (8) and (10) are satisfied, and that the initial velocity additionally satisfies the constraint¹⁰

$$\mathcal{V}_0 \geq \beta \frac{\alpha_2 + \sqrt{\alpha_1 \alpha_2}}{\alpha_1 - \alpha_2}, \quad (12)$$

then the sequence $\{\mathcal{F}_i\}$ is guaranteed to be monotonically increasing (see Fig. 4) with an asymptotic upper bound \mathcal{F}_∞ , which corresponds to the fundamental frequency in the case of zero neutral gap. In other words, the fundamental frequency exhibits a bounded increase during phonation onset due to the average increase in the kinetic energy of the VF system, which agrees in essence with the quasi-steady analysis in Sect. 3.1. This rising trend is also consistent with

⁹ The additional constraint in Eq. (8) is sensible as subglottal pressure during phonation is bounded by the physiological limitations of the vocal and respiratory systems. During normal speech, subglottal pressure values are within the approximate range of 200–800 Pa (Zhang 2016), whereas shouting can lead to subglottal pressures up to and beyond 10 kPa (Lagier et al. 2017).

¹⁰ Derivation of the lower bound is omitted for brevity. It can be obtained by imposing that the frequency function given in Eq. (11) be increasing with respect to its velocity argument (e.g., by setting the first derivative to be positive).

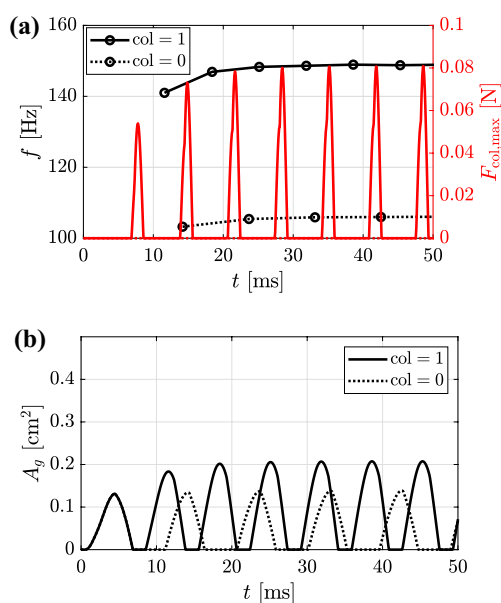


Fig. 5 Two exemplar simulations of phonation onset using the BCM with fixed subglottal pressure, $P_{l,0} = 800$ Pa, and muscle activation values $a_{CT} = 0.2$, $a_{TA} = 0.2$, and $a_{LCA} = 0.5$. Solid lines indicate the simulation with collision springs activated when contact occurs ($col = 1$), whereas dashed lines indicate the simulation with the collision springs deactivated, even when collision occurs ($col = 0$). **a** (Color online) Fundamental frequency (left), and maximum collision force (right) time-series. **b** Glottal area time-series

empirical studies of onset of initial and isolated vowels, see Fig. 1. The implications of matching with empirical data are discussed in Sect. 3.4.

3.3 Numerical simulations with the body-cover model

To ground the analysis from the simplified models in a more physiologically-relevant framework, we resort to simulations using the BCM.

First, we consider onset simulations with fixed steady-state lung pressure, $P_{l,0} = 800$ Pa, and muscle activation values corresponding to low/normal CT and TA activation levels and fully adducted VFs, where $a_{CT} = 0.2$, $a_{TA} = 0.2$, and $a_{LCA} = 0.5$. Figure 5 displays the fundamental frequency, maximum collision force (among the cover masses), and glottal area profiles of two exemplar cases during onset. In one case, the collision springs are activated, resulting in nonzero collision forces when VF contact occurs ($col = 1$), and in the other case, the collision springs are deactivated through the entire simulation period, resulting in zero collision forces ($col = 0$).

Figure 5(a: left axis) shows that there is a gradual increase in fundamental frequency for both cases, though the increase is greater for the $col = 1$ case. Figure 5(a: right axis) illustrates the increase in collision forces during onset for the

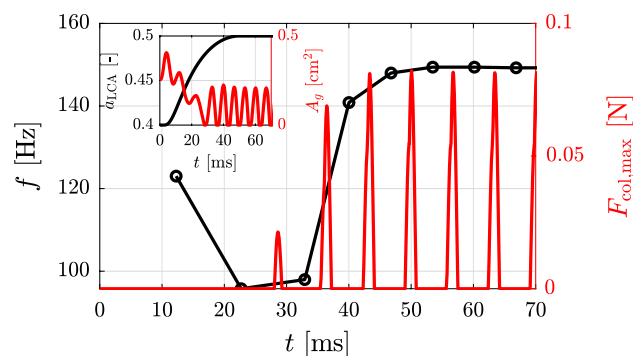


Fig. 6 (Color online) Frequency and maximum collision force versus time for an onset simulation with time-varying neutral glottal gap, where a_{LCA} varies from 0.4 to 0.5 over a 50 ms span (see the inset), $P_{l,0} = 800$ Pa, and $a_{CT} = a_{TA} = 0.2$. The inset of the figure also depicts the time-series of the glottal area waveform, A_g , showing its variation with changing LCA muscle activation

$col = 1$ case, which is attributed to the increased energy of the VF system. It can be inferred from Fig. 5a that the more rapid rise in fundamental frequency in the $col = 1$ case is correlated with the rise in collision forces, which is in agreement with our theoretical analysis in Sect. 3.2¹¹. The rise in fundamental frequency in the case of deactivated collision springs highlights the complex nature of the process, wherein nonlinear stiffness and aerodynamic contributions can also influence fundamental frequency. Moreover, the relatively larger increase in fundamental frequency in the case of activated collision springs indicates that collision plays a significant role in increasing frequency during onset when all other controlling factors (e.g., muscle activation) are fixed. Figure 5b shows that the amplitude of the glottal area waveform increases during onset also due to aerodynamic energy transfer, where a larger oscillation amplitude is noticed in the $col = 1$ case, which can be attributed to the (repulsive) contact forces during the contact periods.

To further highlight the influence of VF contact, we now consider an onset simulation with the same steady-state lung pressure, and CT and TA muscle activation values. However, we vary the LCA activation level from $a_{LCA} = 0.4$ to $a_{LCA} = 0.5$ over a period of 50 ms, which corresponds to the VFs being initially abducted then proceeding to the fully adducted state. This scenario simulates the glottal state during the onset of vowels preceded by voiceless consonants [see, for example, Diaz-Cadiz et al. (2019)]. Figure 6 displays the fundamental frequency, maximum collision force, and glottal area time-series for the simulation (LCA activation and glottal area time-series are shown in the inset). The figure shows that VF oscillations exhibit contact starting from $t \approx 30$ ms. Moreover,

¹¹ This correlation is also observed in other onset simulations with different fixed-in-time static subglottal pressure and muscle activation values.

the figure shows that prior to the initial contact, oscillations exhibit variations in fundamental frequency, potentially due to nonlinear and aerodynamic effects as stated in the discussion of Fig. 5. Furthermore, the figure displays that, starting from the initial VF contact instance, the oscillations exhibit a significant rise in fundamental frequency, which also correlates with the rise in collision forces, in agreement with the theoretical analysis in Sects. 3.1 and 3.2.

3.4 Comments on relations to empirical observations

The increasing frequency resulting from progressively greater degrees of collision during phonation onset predicted by the models in this study aligns with empirical observations for initial and isolated vowels (Smith and Robb 2013; Mohr 1971) (see Fig. 1). This also agrees with some reported observations for vowels preceded by voiced consonants (Hombert et al. 1979). Whereas variations in laryngeal muscle tension and/or aerodynamics are often proposed to be the underlying factors governing the rise in frequency for these conditions (Smith and Robb 2013), our study shows that these influences need not be present to generate the observed behavior, since the natural system dynamics tend to increase fundamental frequency during phonation onset. We emphasize that this does not mean these other factors are not playing a role during onset, only that they are not necessary to produce the observed frequency patterns.

As has been observed clinically (see Fig. 15), this gradually increasing effect of VF contact during phonation onset is also present in other phonetic contexts, including the onset of vowels preceded by voiceless consonants, implying the relevance of VF contact in various contexts. However, as shown in Fig. 1, fundamental frequency tends to decrease during onset for a vowel preceded by a voiceless consonant, indicating that the collision-based rise in frequency is overshadowed by other factors. In Sect. 4, we show that laryngeal muscle activation can induce the observed decreasing trends of fundamental frequency during the onset of vowels preceded by voiceless consonants.

4 Muscle tension and frequency regulation

In this section, we explore the influence of intrinsic laryngeal muscle tension and, in particular, the role of CT and TA muscles during phonation onset using the BCM. Intrinsic laryngeal muscles and their roles in phonation have been extensively investigated in several clinical (Chhetri and Neubauer 2015; Chhetri et al. 2012, 2014; Choi et al. 1995, 1993) and numerical (Geng et al. 2021; Alzamendi

et al. 2020; Movahhedi et al. 2021; Yin and Zhang 2013, 2014) studies, where it has been found that the CT and TA muscles are essential in regulating fundamental frequency. Increasing activation of the CT muscle has been found to increase phonation fundamental frequency (Löfqvist et al. 1989; Chhetri et al. 2014). On the other hand, the role of the TA muscle in modulating fundamental frequency is more complex as its activation can either increase or decrease fundamental frequency, with some conflicting results in the literature [see Movahhedi et al. (2021)]. Activation of the LCA and interarytenoid muscles have been found to be positively correlated with fundamental frequency (Choi et al. 1995), whereas PCA activation exhibits negative correlation with fundamental frequency (Choi et al. 1993).

To the best of our knowledge, there are few studies that have substantially investigated the temporal variations of laryngeal muscle activation during phonation onset and how these variations may underlie empirical observations of fundamental frequency [e.g., Löfqvist et al. (1989)]. In this study, we attempt to explore these temporal variations in order to elucidate some of the underlying mechanisms of phonation onset.

In real phonation scenarios, the intrinsic laryngeal muscles do not act in isolation and their effect on fundamental frequency depends on several factors, including the relative geometry and contraction levels of agonist/antagonist muscles (Alzamendi et al. 2022). To simplify our analysis in this section, we aim to isolate the effects of the CT and TA muscles and assume that the tension variation in other laryngeal muscles is negligible. In all simulations presented below, we set $P_{l,0} = 800$ Pa and $a_{LCA} = 0.5$, which corresponds to fully adducted vocal folds.

4.1 Cricothyroid muscle

The CT muscle plays a crucial role in regulating fundamental frequency by elongating and tensioning the VFs (Titze and Story 2002; Sonesson 1982; Löfqvist et al. 1989; Atkinson 1978; Chhetri et al. 2012). Electromyography has shown that activation of the CT muscle is higher in phonetic contexts wherein vowels are preceded by voiceless consonants in comparison with vowels preceded by voiced consonants, which correlates with the empirically observed higher onset fundamental frequency in such conditions (Löfqvist et al. 1989). It has been speculated that higher VF tension, which correlates with higher activation of the CT muscle, is required to mitigate VF vibrations during the production of voiceless consonants and that the higher tension carries over into the adjacent vowel (Hombert et al. 1979). Here, we aim to investigate this hypothesis numerically by varying CT muscle activation

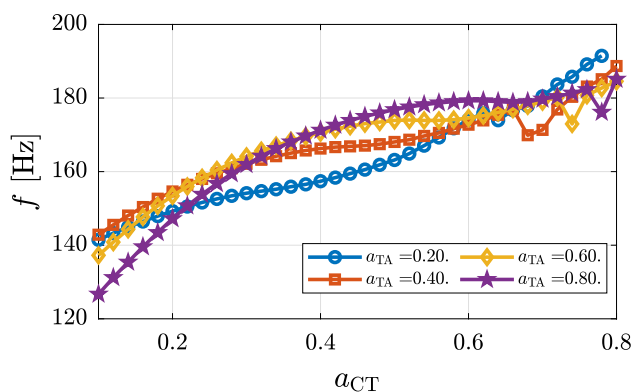


Fig. 7 (Color online) Sustained phonation fundamental frequency as a function of CT muscle activation for different TA activation levels

while keeping the activation levels of the TA and LCA muscles fixed.

We begin with a quasi-steady analysis wherein CT activation is fixed in time. Figure 7 presents sustained phonation fundamental frequency as a function of CT activation for different TA activation levels. In all cases, sustained phonation fundamental frequency is positively correlated with the CT muscle activation level (for fixed TA activation level), in agreement with previous numerical and clinical studies (Alzamendi et al. 2020; Titze and Story 2002; Chhetri et al. 2014). There are slight fluctuations observed in the fundamental frequency curves for large a_{CT} values, which can be attributed, in part, to the nonlinearity of the BCM. Assuming variations in the activation levels of other laryngeal muscles to be small, this suggests that, in contexts where vowels are preceded by voiceless consonants, CT muscle activation level may be decreasing in order to achieve the empirically observed decaying fundamental frequency patterns as seen in, for example, Stepp et al. (2010a)¹².

Figure 8 presents instances of the glottal area time-series during phonation onset for monotonically decaying CT activation, with initial value $a_{CT,i}$ and final value $a_{CT,f}$. Specifically, initial activation level $a_{CT,i}$ is varied between 0.2 and 0.6 across simulations and final activation value is set to be 0.2. The transition between the two a_{CT} levels occurs over a duration of 50 ms (see insets for the activation level temporal evolution), which is of the same order of magnitude as observed experimentally [see, for example, the electromyographic signals depicted in Figures 1–3 in Löfqvist et al. (1989)]. The figure shows that the amplitude of vibration grows the most rapidly when there is no change in CT activation ($a_{CT,i} = 0.2$), with the rate of amplitude growth

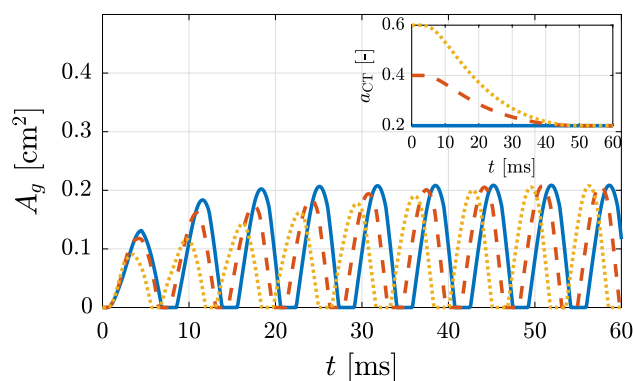


Fig. 8 (Color online) Time-series of glottal area for varying initial CT activation levels using the BCM at $a_{CT,f} = 0.2$ and $a_{TA} = 0.2$. Similar trends hold for $a_{TA} = 0.4$ (not shown). The time-series of CT activation are shown in the inset

decreasing with increasing $a_{CT,i}$. Higher CT activation levels result in stiffer folds, and thus, higher frequency and generally lower amplitude, which relax as the activation level decreases in time. This is in agreement with the claim in Hombert et al. (1979) that the increased VF tension in phonetic contexts with vowels preceded by voiceless consonants is required to inhibit VF oscillations during the production of the voiceless consonant.

Figure 9 presents the temporal evolution of normalized fundamental frequency for the glottal area time-series shown in Fig. 8, as well as for analogous cases with $a_{TA} = 0.4$. Decreasing CT activation levels during onset, with sufficiently large initial values, generally results in a decaying fundamental frequency profile, which matches empirical observations of vowels preceded by voiceless consonants, see Fig. 1. Figure 9 displays non-monotonicity of the fundamental frequency profile in some cases, such as for $a_{TA} = 0.4$, $a_{CT,i} = 0.4$, wherein the frequency of the second cycle is larger than that of the first cycle despite the monotonic decay of CT activation levels. This has also been observed in empirical studies of onset fundamental frequency [see, for example, Löfqvist et al. (1989), Lien et al. (2015), and Fig. 15 in Appendix 4]. This non-monotonic behavior can be attributed, in part, to the collision-based mechanism, which is dominant in some cases, such as when $a_{CT,i} = 0.2$ (CT activation is constant-in-time). Collision onset causes fundamental frequency to increase, which opposes the effect of decreasing VF stiffness associated with the reduction in CT activation level. This demonstrates the complexity of phonation onset, where competing mechanisms are at play.

Similar fundamental frequency trends are observed when fixing initial CT activation and varying the final value, as shown in Fig. 10. This figure, in combination with Fig. 9, shows that when the drop in CT activation level is

¹² The extent of which laryngeal maneuvers, including the variation of the CT muscle activation, alter fundamental frequency is language-specific as highlighted in Francis et al. (2006).

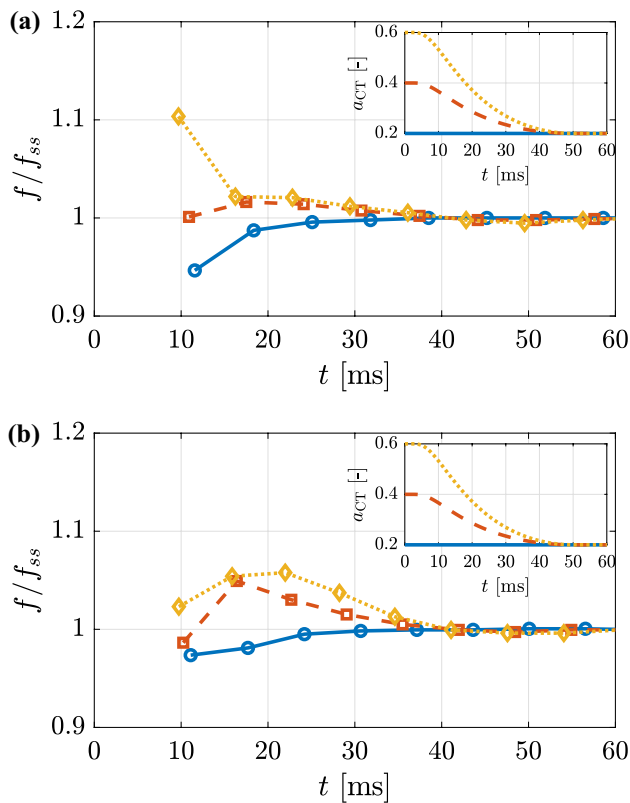


Fig. 9 (Color online) Time-series of fundamental frequency, normalized with respect to sustained phonation fundamental frequency f_{ss} , for varying initial CT activation levels using the BCM for **a** $a_{TA} = 0.20$, and **b** $a_{TA} = 0.40$. The time-series of CT activation are shown in the insets

sufficiently large, the magnitude of the reduction in CT activation is correlated with the drop in (relative) fundamental

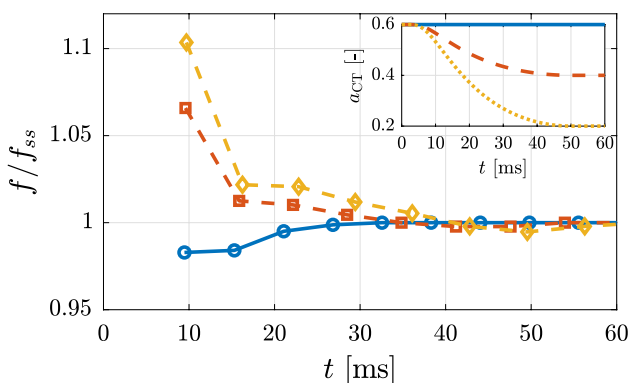


Fig. 10 (Color online) Time-series of fundamental frequency, normalized with respect to sustained phonation fundamental frequency f_{ss} , with $a_{TA} = 0.2$, $a_{CT,i} = 0.6$, and different final CT activation levels. The time-series of CT activation are shown in the inset

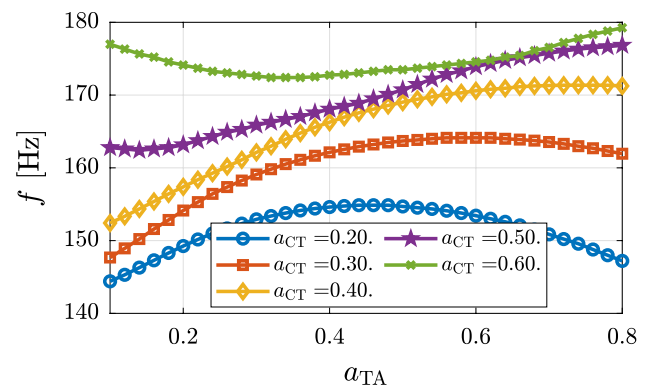


Fig. 11 (Color online) Sustained phonation fundamental frequency as a function of TA muscle activation level for different CT activation levels

frequency. This important finding will be used to explain some empirical observations in Sect. 4.3.

4.2 Thyroarytenoid muscle

Activation of the TA muscle contributes to adducting and shortening the VFs (Titze and Hunter 2007; Chhetri et al. 2012). TA activation during phonation initiation has been found to begin when the VFs start adducting and carries over into sustained phonation (Poletto et al. 2004). Studies with human subjects have shown that the relation between TA activation and sustained phonation fundamental frequency is proportional when fundamental frequency values are low, whereas the relation is inverse at high frequency levels (Titze et al. 1989). However, in vitro studies involving excised canine larynges (Chhetri et al. 2014) exhibit some deviations from human studies (Titze et al. 1989).

Figure 11 illustrates the relation between sustained phonation fundamental frequency and TA muscle activation for various CT activation levels. This figure shows that the relationship between steady-state fundamental frequency and TA activation level is relatively complex, with the influence of increasing TA depending on CT activation, in agreement with empirical data (Titze et al. 1989). For example, when $a_{CT} = 0.2$, corresponding to relatively low fundamental frequency, the relation is proportional when $a_{TA} \in [0.1, 0.3]$, but it is inverse when $a_{CT} = 0.6$. The steady phonation results suggest that in scenarios where CT activation follows a decaying profile and TA activation is fixed, higher TA activation values can result in smaller differences between initial and final fundamental frequencies (see, for example, the decrease in spread between the curves in Fig. 11 as a_{TA} increases from 0.1 to 0.3).

To test this, we perform onset simulations with decaying CT profiles similar to those shown in the inset of Fig. 10 at various fixed TA activation values. TA activation levels

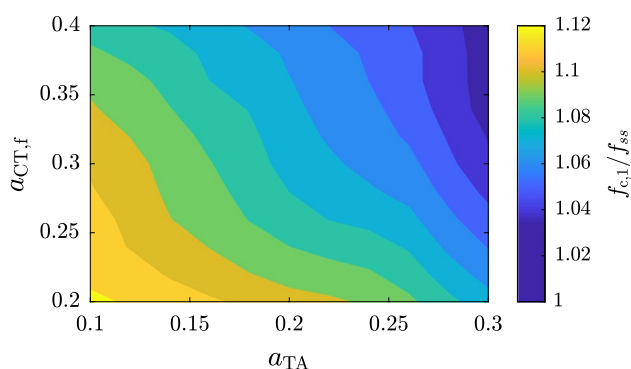


Fig. 12 (Color online) Contour plot of initial fundamental frequency normalized by steady state frequency, $f_{c,1}/f_{ss}$, as a function $a_{CT,f}$ and a_{TA} , where $a_{CT,i} = 0.6$

are selected such that states wherein CT and TA activation have agonistic and antagonistic influences on fundamental frequency in the steady-state analysis are both represented (i.e., the relation between fundamental frequency and a_{TA} is inverse at $a_{CT,i}$ and the relation between fundamental frequency and a_{TA} is proportional at $a_{CT,f}$, according to our quasi-steady analysis). We record the fundamental frequency of the first cycle, $f_{c,1}$ (the inverse of the time difference between the first two detected peaks of the glottal area waveform time-series) and normalize it with respect to the steady-state fundamental frequency during sustained phonation, f_{ss} . Figure 12 displays a contour plot of $f_{c,1}/f_{ss}$ as a function of $a_{CT,f}$ and a_{TA} . The figure shows that increasing a_{TA} from 0.1 to 0.3 results in the initial normalized fundamental frequency decreasing for decaying CT profiles with $a_{CT,i} = 0.6$ and $a_{CT,f} \in [0.2, 0.4]$, in agreement with our quasi-steady analysis.

4.3 Comments on relations to empirical observations

As seen in Fig. 1, empirical data from human studies indicate that fundamental frequency decays during the onset of vowels preceded by voiceless consonants, with higher onset relative fundamental frequency values in the case of adult speakers with healthy voices in comparison with adult speakers with hyperfunctional voices (Stepp et al. 2010a)¹³, where speakers with hyperfunctional voices often exhibit excessive and/or imbalanced muscular forces (Hillman et al. 1989, 2020). Figures 9 and 10 show that when the magnitude of the drop in CT activation during onset is sufficiently high there is a decrease in normalized fundamental frequency that is correlated with the magnitude of the reduction in CT

activation. This suggests that CT activation levels may be a factor underlying the differences between healthy speakers and speakers with vocal hyperfunction, wherein healthy speakers potentially produce relatively larger variations in the CT activation levels during phonation onset. Interestingly, the non-monotonic behavior (initial rise followed by a fall) in fundamental frequency present in some cases in Fig. 9 is also observed in empirical studies, see for example Löfqvist et al. (1989), Lien et al. (2015), and Fig. 15 in Appendix 4. Moreover, Fig. 12 indicates that higher TA activation levels result in lower normalized fundamental frequency values during phonation onset, at least initially. Thus, TA activation level is another potential factor that may underlie clinical observations, where our results suggest that speakers with vocal hyperfunction may produce higher TA activation levels, which also may explain the empirically observed lower initial relative fundamental frequency values (Stepp et al. 2010a).

5 Conclusion

In this paper, we aimed to uncover some of the potential underlying mechanisms driving the observed differences in onset fundamental frequency patterns in different phonetic contexts, see Fig. 1, where we resorted to theoretical and numerical analyses of single- and multi-mass models, respectively (see Sect. 2). We found that the increasing degree of VF collision during onset, associated with the rise in vibration amplitude, and/or the decrease in the neutral glottal gap, naturally gives rise to an increase in system fundamental frequency (see Sect. 3). Such an increase in fundamental frequency is experimentally observed during isolated/initial vowels, and in some instances of a vowel preceded by a voiced consonant (see Sect. 3.4). In these cases, laryngeal muscle tension may still play a role, but it is not a prerequisite for the observed behavior.

On the other hand, our analysis suggested that muscle activation is necessary to produce the observed decay in fundamental frequency evident in vowels preceded by voiceless consonants, since the system dynamics with all control factors fixed produce the opposite trend (see Fig. 9a). In particular, our analysis indicated that reduction in fundamental frequency is due, in part, to a concomitant decrease in cricothyroid muscle activation during onset (see Sect. 4.1). Interestingly, the competing mechanisms of muscle activation and collision can lead to a frequency pattern that initially rises then drops, which has been observed in experimental

¹³ Relative fundamental frequency is a normalized measure of fundamental frequency that is functionally similar to the frequency patterns presented in Sects. 4.1 and 4.2

studies of onset fundamental frequency (see Sect. 4.1). The magnitude of the reduction in cricothyroid muscle activation was found to be a potential factor underlying the differences in relative fundamental frequency between healthy and hyperfunctional voices during the onset of vowels preceded by voiceless consonants (see Sect. 4.3). Furthermore, our investigation suggested that increased thyroarytenoid muscle activation mitigates the drop in relative fundamental frequency caused by a decrease in cricothyroid muscle activation, which may also contribute to the experimentally observed differences between hyperfunctional and normal phonation (see Sects. 4.2, 4.3).

The current study utilized the body-cover model integrated with muscle activation rules and a quasi-steady viscous flow model and implemented simplifying assumptions regarding the role of acoustics and aerodynamics during phonation onset. In future work, we aim to adopt more physiologically accurate phonation models [e.g., articulating triangular body-cover model incorporating all five intrinsic muscles (Alzamendi et al. 2022)] and potentially a more complex flow model in order to gain refined insights into the impact of aerodynamics, acoustics, and laryngeal muscle activation on fundamental frequency during phonation onset. We further aim to more robustly examine vocal hyperfunction in the context of relative fundamental frequency through our modeling framework.

Appendix 1: Quasi-steady viscous glottal flow model

Herein, we derive a viscous glottal flow model similar to that presented in Lucero and Schoentgen (2015), but that accounts for the convergent/divergent configurations arising in the body-cover vocal fold (VF) model. A schematic diagram of the body-cover model (Story and Titze 1995) is shown in Fig. 13, where y_1 , y_2 , and y_b denote the displacements of the inferior cover mass m_1 , superior cover mass m_2 , and body mass m_b , respectively. The thicknesses of m_1 and m_2 are given by T_1 and T_2 , respectively, and $T = T_1 + T_2$ yields the total VF thickness. The glottal areas associated with the displacements of the inferior and superior masses are given by $a_1 = 2L \max\{y_1, 0\}$ and $a_2 = 2L \max\{y_2, 0\}$, respectively, where L is the VF length. The subglottal and supraglottal areas are A_s and A_e , respectively, with associated pressures P_s and P_e .

The x -axis is aligned with the streamwise direction with $x = 0$ located at the inferior margin of m_1 , the junction between the lower and upper masses is located at $x = T_1$, and $x = T$ corresponds to the superior margin of m_2 . The density of air is ρ , the viscosity is μ , and the associated speed of sound is c_s . We assume the volumetric flow rate, q , to be quasi-steady. We assume Poiseuille flow throughout the glottis when the flow attachment criterion $0 < a_2 \leq 1.2a_1$ is satisfied (Pelorson et al.

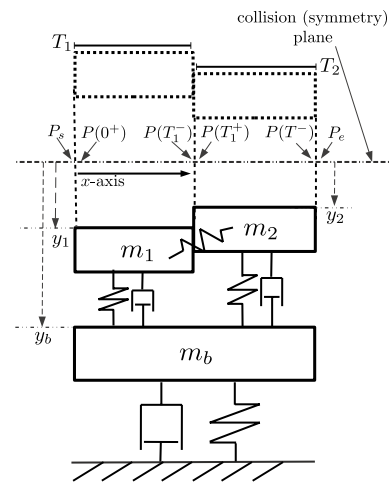


Fig. 13 Schematic of the body-cover model with locations of important pressure values identified

1994; Šidlof et al. 2011; Titze 2002). In the case of flow separation, wherein $a_2 > 1.2a_1 > 0$, Poiseuille flow is assumed over the inferior mass only and the supraglottal pressure ($P = P_e$) is applied to the superior mass. When the masses are in contact, the loading conditions described by Story and Titze (1995) are employed. In particular, if $a_1 > 0$ and $a_2 = 0$, then $P = P_s$ and $P = 0$ over the lower mass and upper mass, respectively. Moreover, if $a_1 = 0$ and $a_2 > 0$, then $P = 0$ and $P = P_e$ over the lower mass and upper mass, respectively. Finally, if $a_1 = a_2 = 0$, then $P = 0$ over the lower and upper masses. We note that in our derivation of the flow model, it is assumed that the glottal flow does not revert (i.e., $P_s - P_e \geq 0$).

Considering the attached flow case, $0 < a_2 \leq 1.2a_1$, let $P(0^+)$ denote the pressure at the inferior margin of the inferior mass, that is, $P(0^+) = \lim_{x \rightarrow 0^+} P(x)$ (see Fig. 13). We assume Bernoulli flow between the subglottal region and the entry to the glottis, with $A_s \gg a_1$, resulting in the equation

$$P(0^+) + \frac{1}{2}\rho\left(\frac{q}{a_1}\right)^2 = P_s. \quad (13)$$

Moreover, we assume Poiseuille flow along mass m_1 ($x \in (0, T_1)$), yielding

$$P(0^+) = P(x) + x \frac{12\mu L^2 q}{a_1^3}, \quad (14)$$

and, taking the limit as x approaches T_1 from the left side, yields

$$P(0^+) = P(T_1^-) + T_1 \frac{12\mu L^2 q}{a_1^3}. \quad (15)$$

Assuming a Bernoulli flow through the junction between the cover masses at $x = T_1$ yields

$$P(T_1^-) + \frac{1}{2}\rho\left(\frac{q}{a_1}\right)^2 = P(T_1^+) + \frac{1}{2}\rho\left(\frac{q}{a_2}\right)^2. \quad (16)$$

Assuming a Poiseuille flow again along mass m_2 ($x \in (T_1, T)$) results in

$$P(T_1^+) = P(x) + (x - T_1)\frac{12\mu L^2 q}{a_2^3}, \quad x \in (T_1, T). \quad (17)$$

Taking the limit as x approaches T from the left side yields

$$P(T_1^+) = P(T^-) + T_2\frac{12\mu L^2 q}{a_2^3}. \quad (18)$$

Finally, at the superior margin of the VFs, we assume flow separation with no pressure recovery, which gives

$$P(T^-) = P_e. \quad (19)$$

Combining Eqs. (13), (15), (16), (18), and (19) results in a quadratic equation for the flow rate given by

$$\frac{\rho}{2a_2^2}q^2 + 12\mu L^2\left(\frac{T_1}{a_1^3} + \frac{T_2}{a_2^3}\right)q - (P_s - P_e) = 0, \quad (20)$$

the solution of which is

$$q = \frac{2a_2^3\delta_p}{\gamma + \sqrt{\gamma^2 + 2\rho a_2^4\delta_p}}, \quad (21)$$

where $\delta_p = P_s - P_e$ and $\gamma = 12\mu L^2(T_1(a_2/a_1)^3 + T_2)$. Equation (21) can then be used to determine the pressures applied to each cover mass through substitution back into the Bernoulli/Poiseuille flow equations from which they were obtained. The aerodynamic forces over the lower and upper masses, F_1 and F_u , are then computed as

$$F_1 = L \int_0^{T_1} P(x)dx = LT_1 P(0^+) - L\frac{T_1^2}{2}\frac{12\mu L^2 q}{a_1^3} \quad (22)$$

and

$$F_u = L \int_{T_1}^T P(x)dx = LT_2 P(T_1^+) - L\frac{T_2^2}{2}\frac{12\mu L^2 q}{a_2^3}. \quad (23)$$

For the detached flow case, $a_2 > 1.2a_1 > 0$, we assume Bernoulli flow from the subglottal region to the inferior margin of the inferior mass, viscous flow over the inferior mass, and flow separation at the mass junction ($x = T_1$) leading to zero pressure recovery and uniform pressure P_e over the upper mass. Following a derivation similar to that of Eq. (20) yields a quadratic equation for flow as

$$\frac{\rho}{2a_1^2}q^2 + 12\mu L^2\frac{T_1}{a_1^3}q - (P_s - P_e) = 0, \quad (24)$$

which has the solution

$$q = \frac{2a_1^3\delta_p}{\bar{\gamma} + \sqrt{\bar{\gamma}^2 + 2\rho a_1^4\delta_p}}, \quad (25)$$

where $\bar{\gamma} = 12\mu L^2 T_1$. From this, we can obtain the aerodynamic force on the inferior mass, which is given by Eq. 22, and the superior mass, which is given by

$$F_u = L \int_{T_1}^T P(x)dx = LT_2 P_e. \quad (26)$$

When acoustics are modelled using the wave reflection analog, flow rate has to be given in terms of incident pressure waves. The subglottal pressure P_s can be written as $P_s = P_s^+ + P_s^-$, where P_s^+ denotes the incident subglottal pressure and P_s^- denotes the outward travelling subglottal pressure. Similarly, the supraglottal pressure P_e can be written as $P_e = P_e^+ + P_e^-$, where P_e^+ denotes the outward travelling supraglottal pressure and P_e^- denotes the incident supraglottal pressure. Continuity then yields $P_s^- = P_s^+ - \rho c_s q/A_s$ and $P_e^+ = P_e^- + \rho c_s q/A_e$. Consequently, the transglottal pressure $P_s - P_e$ can be written as

$$P_s - P_e = 2(P_s^+ - P_e^-) - \frac{\rho c_s}{A^*}q, \quad (27)$$

where $A^* = A_s A_e / (A_s + A_e)$. By plugging Eq. (27) into Eq. (20) and solving for q , we obtain, for fully attached flow,

$$q = \frac{4a_2^3\tilde{\delta}_p}{\chi + \sqrt{\chi^2 + 4\rho a_2^4\tilde{\delta}_p}}, \quad (28)$$

where $\tilde{\delta}_p = P_s^+ - P_e^-$ and $\chi = 12\mu L^2(T_1(a_2/a_1)^3 + T_2) + \rho c_s a_2^3/A_s^*$. Similarly, for the detached flow, we obtain

$$q = \frac{4a_1^3\tilde{\delta}_p}{\bar{\chi} + \sqrt{\bar{\chi}^2 + 4\rho a_1^4\tilde{\delta}_p}}, \quad (29)$$

where $\bar{\chi} = 12\mu L^2 T_1 + \rho c_s a_1^3/A^*$. Finally, if one of the cover masses is in contact, such that $a_1 = 0$ or $a_2 = 0$, then $q = 0$.

Appendix 2: Derivation of the approximate velocity recurrence relation

Herein, we derive the approximate recurrence relation in Eq. (7) for the hybrid model. Let $\eta : [0, \infty) \rightarrow \mathbb{R}$ be an oscillatory solution to Eq. (1) with oscillations of possibly varying amplitude and frequency. Consider an oscillation

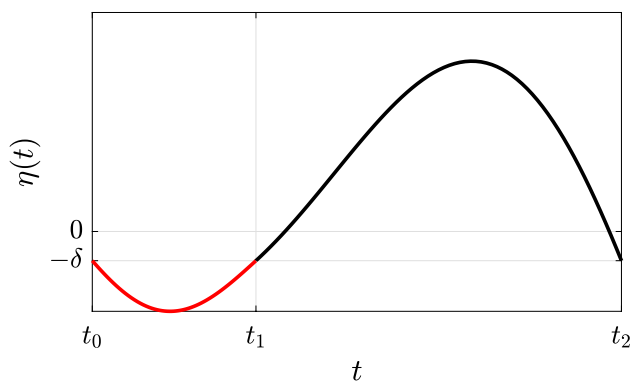


Fig. 14 (Color online) One oscillation period of the hybrid phonation model

period $\mathcal{I} = [t_0, t_2]$ depicted in Fig. 14, where it is assumed that $\eta(t_0) = -\delta$ and that the initial velocity $\dot{\eta}(t_0) = v_0 < 0$. To obtain the discrete system, we seek velocities v_1 and v_2 at time instances t_1 and t_2 , corresponding to the first and second times after t_0 such that $\eta(t) = -\delta$, respectively.

During the interval $\mathcal{I}_1 = [t_0, t_1]$, $\eta(t) \leq -\delta$ and the dynamics satisfy Eq. (1b). For convenience, define $\phi(\cdot) = -(\eta(\cdot) + \delta)$, then on the interval \mathcal{I}_1 , ϕ satisfies $M\ddot{\phi} + B_2\dot{\phi} + K\phi = c$, where $c = -K\delta$. Note that $\phi(t_0) = 0$ and $\bar{v}_0 = \dot{\phi}(t_0) = -v_0$. Therefore, ϕ is given explicitly as

$$\phi(t) = \frac{\bar{v}_0 - \frac{B_2 c}{2MK}}{\omega_2} e^{-\frac{B_2}{2M}(t-t_0)} \sin(\omega_2(t-t_0)) - \frac{c}{K} e^{-\frac{B_2}{2M}(t-t_0)} \cos(\omega_2(t-t_0)) + \frac{c}{K}, \quad t \in \mathcal{I}_1,$$

and $\phi(t) = \exp(-B_2(t-t_0)/(2M))(\lambda \sin(\omega_2(t-t_0)) + \bar{v}_0 \cos(\omega_2(t-t_0)))$, where $\lambda = \omega_2 c/K - B_2[\bar{v}_0 - B_2 c/(2MK)]/(2M\omega_2)$. In the case $c = 0$, $t_1 = t_0 + \pi/\omega_2$.

If we assume c to be sufficiently small, utilizing a small neutral gap approximation, then t_1 can be obtained approximately as follows: let us write $t_1, \phi, \dot{\phi}$ as functions of c ($t_1 = t_1(c)$, $\phi = \phi(t, c)$, $\dot{\phi} = \dot{\phi}(t, c)$). Then, by implicitly differentiating the equation $\phi(t_1(c), c) = 0$ with respect to c and solving for dt_1/dc , we get $t'_1 := dt_1/dc = -D_2\phi(t_1(c), c)/\dot{\phi}(t_1(c), c)$, where D_{ij} denotes the partial derivative of a multivariable function f with respect to its i^{th} argument. By direct substitution, and using the fact that $t_1(0) = t_0 + \pi/\omega_2$, we get that $t'_1(0) = [1 + \exp(B_2\pi/(2M\omega_2))]/(K\bar{v}_0)$. Then, using first-order Taylor expansion, we get $t_1(c) \approx t_1(0) + t'_1(0)c$, which results in the approximate formula $t_1 = t_0 + \pi/\omega_2 + \bar{t}$, where

$$\bar{t} = c \frac{1 + e^{\frac{B_2\pi}{2M\omega_2}}}{K\bar{v}_0}. \quad (30)$$

By plugging the approximate formula of t_1 into the formula for $\dot{\phi}$ and using the trigonometric identities $\sin(\pi + x) = -\sin(x)$, $\cos(\pi + x) = -\cos(x)$, we obtain $\phi(t_1) \approx \exp(-B_2(\pi/\omega_2 + \bar{t})/(2M))(-\lambda \sin(\omega_2\bar{t}) - \bar{v}_0 \cos(\omega_2\bar{t}))$. Note that, by definition, \bar{t} is proportional to δ . By further utilizing the small neutral gap assumption, we implement the approximations $\exp(-B_2\bar{t}/(2M)) \approx 1$, $\sin(\omega_2\bar{t}) \approx \omega_2\bar{t}$, and $\cos(\omega_2\bar{t}) \approx 1$ in the above expression, in addition to the definition of λ , which yields

$$\dot{\phi}(t_1) \approx - \left(-c \frac{B_2}{2M} \frac{1 + e^{-\frac{B_2\pi}{2M\omega_2}}}{K\bar{v}_0} + e^{-\frac{B_2\pi}{2M\omega_2}} \right) \bar{v}_0 - \omega_2 c^2 \frac{1 + e^{-\frac{B_2\pi}{2M\omega_2}}}{K\bar{v}_0} \left[\frac{\omega_2}{K} + \frac{B_2}{2M} \frac{B_2}{\omega_2} \right].$$

By neglecting the second-order term (proportional to c^2), we obtain the final approximate formula $\dot{\phi}(t_1) \approx -\exp(-B_2\pi/(2M\omega_2))\bar{v}_0 + cB_2[1 + \exp(-B_2\pi/(2M\omega_2))]/(2MK)$, and in terms of v_1 and v_0 , where $c = -K\delta$ is substituted back, we get the approximate relation

$$v_1 = -e^{-\frac{B_2\pi}{2M\omega_2}} v_0 + \tilde{k}\delta \frac{B_2}{2M} \left(1 + e^{-\frac{B_2\pi}{2M\omega_2}} \right). \quad (31)$$

By applying a similar analysis on the interval $\mathcal{I}_2 = [t_1, t_2]$, where the small neutral gap approximation is utilized¹⁴, we obtain the approximate formulas $t_2 = t_1 + \pi/\omega_1 + \tilde{t}$, where

$$\tilde{t} = \delta \frac{1 + e^{-\frac{B_1\pi}{2M\omega_1}}}{v_1}, \quad (32)$$

and

$$v_2 = -e^{\frac{B_1\pi}{2M\omega_1}} v_1 - \delta \frac{B_1}{2M} (1 + e^{\frac{B_1\pi}{2M\omega_1}}). \quad (33)$$

Combining Eqs. (31) and (33) yields $|v_2| = \mathcal{A}|v_0| + \mathcal{W}$, where

$$\mathcal{A} = e^{\frac{\pi}{2M} \left[\frac{B_1}{\omega_1} - \frac{B_2}{\omega_2} \right]}, \quad \mathcal{W} = \frac{\delta}{2M} \left[\tilde{k}B_2 \left(\mathcal{A} + e^{\frac{B_1\pi}{2M\omega_1}} \right) + B_1 \left(1 + e^{\frac{B_1\pi}{2M\omega_1}} \right) \right].$$

By repeating the approximate solution procedure recursively over all oscillation periods, where collision is assumed to occur in each period, we obtain Eq. (7).

¹⁴ The damping term B_1 grows unboundedly as δ approaches zero ($B_1 = 2\tau P_L/(k_i\delta) - B$). Therefore, we additionally assume in the subsequent derivation that the subglottal lung pressure P_L is sufficiently small (depending on the value of δ), in order to have the damping term B_1 bounded.

Appendix 3: Derivation of the approximate frequency relation

The frequency of the oscillation period given in Fig. 14 is $\mathcal{F} = [t_2 - t_0]^{-1} = [(t_2 - t_1) + (t_1 - t_0)]^{-1}$. From the analysis in Appendix 2, we have $t_2 - t_1 = \pi/\omega_1 + \tilde{t}$, and $t_1 - t_0 = \pi/\omega_2 + \tilde{t}$, where \tilde{t} and \tilde{t} are given approximately by Eqs. (30) and (32), respectively. Consequently, we obtain $\mathcal{F} = [1/\mathcal{F}_\infty + \tilde{t} + \tilde{t}]^{-1}$. By substituting Eq. (31) into Eq. (32), \mathcal{F} can be written in terms of $|v_0|$ as:

$$\mathcal{F} = \left[\frac{1}{\mathcal{F}_\infty} + \frac{\alpha_1}{|v_0| + \beta} - \frac{\alpha_2}{|v_0|} \right]^{-1},$$

where

$$\alpha_1 = \delta \left(e^{\frac{B_2 \pi}{2M\omega_2}} + \frac{1}{A} \right),$$

$$\alpha_2 = \tilde{k} \delta \left(e^{\frac{B_2 \pi}{2M\omega_2}} + 1 \right),$$

$$\beta = \tilde{k} \delta \frac{B_2}{2M} \left(e^{\frac{B_2 \pi}{2M\omega_2}} + 1 \right).$$

By repeating the above derivation over each oscillation period, we obtain Eq. (11).

Appendix 4: Empirical data and observations

Figure 15a presents fundamental frequency data from two healthy male participants during the onset portion of repeated /ifi/ and /iti/ utterances¹⁵. For these utterances, the VFs transition from an abducted to an adducted state prior to and during the onset of the second vowel. Each data point corresponds to an average over four utterances, with error bars showing one standard deviation.

Figure 15b presents the duration of contact for the first several oscillation cycles incorporating collision. Contact time is estimated from the glottal angle, θ_G , extracted from endoscopic video recordings during the utterances. The contact time per cycle, t_c , corresponds to the interval when $\theta_G = 0$. The figure shows that, in general, VF contact time per cycle increases gradually during phonation onset for both participants. Based upon the analysis in Sect. 3.1, the increasing role of collision in the dynamics should result in an increase in fundamental frequency. As per the discussion in Sect. 4.1, decaying CT activation leads to a decrease in fundamental frequency. Comparing empirical

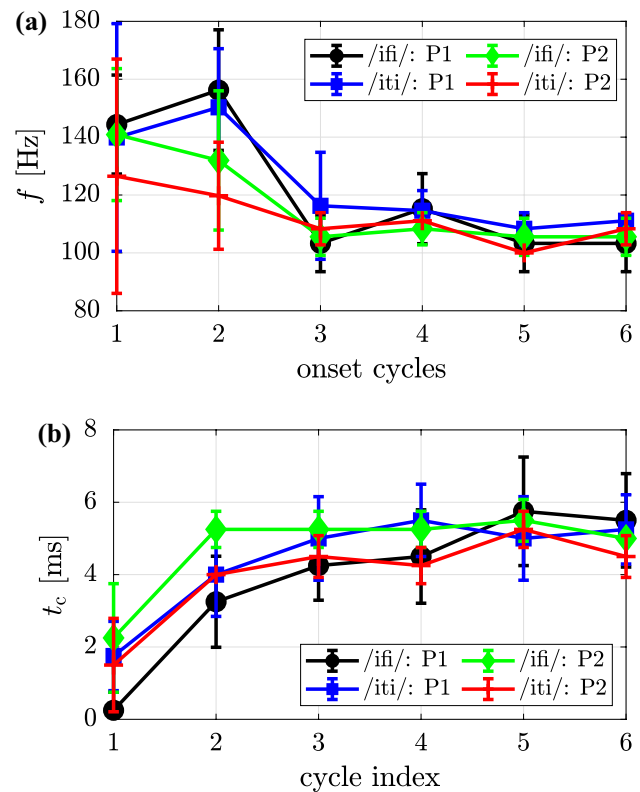


Fig. 15 (Color online) **a** Fundamental frequency and **b** contact duration during the onset portion of /ifi/ and /iti/ utterances collected from two participants (P1 and P2): points indicate the average across 4 utterances and error bars indicate one standard deviation

data in Fig. 15a with the simulation results in Fig. 9, we note qualitative similarities between the frequency patterns of participant 1 (P1) and the numerical case with $a_{TA} = 0.4$ and $a_{CT,i} = 0.6$, and the frequency patterns of participant 2 (P2) and the numerical case with $a_{TA} = 0.2$ and $a_{CT,i} = 0.6$. That is, the modeling exercise is successful in qualitatively replicating several experimentally observed patterns during phonation onset.

Acknowledgements The authors thank Jonathan Deng for running finite element simulations to validate some of the modeling assumptions employed in this work and Dr. Matías Zañartu for insightful discussions on the intrinsic musculature of the larynx. Research reported in this work was supported by the NIDCD of the NIH under awards P50DC015446 and R01DC015570. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Alzamendi GA, Manríquez R, Hadwin PJ, Deng JJ, Peterson SD, Erath BD, Mehta DD, Hillman RE, Zañartu M (2020) Bayesian estimation of vocal function measures using laryngeal high-speed

¹⁵ Clinical data provided by the STEPP Lab, Boston University; these data were utilized in several previous studies on the kinematics of VFs (McKenna et al. 2016; Diaz-Cadiz et al. 2019; Park et al. 2021).

- videoendoscopy and glottal airflow estimates: an in vivo case study. *J Acoust Soc Am* 147(5):EL434–EL439
- Alzamendi GA, Peterson SD, Erath BD, Hillman RE, Zañartu M (2022) Triangular body-cover model of the vocal folds with coordinated activation of the five intrinsic laryngeal muscles. *J Acoust Soc Am* 151(1):17–30
- Atkinson JE (1978) Correlation analysis of the physiological factors controlling fundamental voice frequency. *J Acoust Soc Am* 63(1):211–222
- Azar SS, Chhetri DK (2022) Phonation threshold pressure revisited: effects of intrinsic laryngeal muscle activation. *Laryngoscope* 132(7):1427–1432
- Bhattacharya P, Siegmund T (2015) The role of glottal surface adhesion on vocal folds biomechanics. *Biomech Model Mechanobiol* 14(2):283–295
- Chhetri DK, Neubauer J (2015) Differential roles for the thyroarytenoid and lateral cricoarytenoid muscles in phonation. *Laryngoscope* 125(12):2772–2777
- Chhetri DK, Neubauer J, Berry DA (2012) Neuromuscular control of fundamental frequency and glottal posture at phonation onset. *J Acoust Soc Am* 131(2):1401–1412
- Chhetri DK, Neubauer J, Sofer E, Berry DA (2014) Influence and interactions of laryngeal adductors and cricothyroid muscles on fundamental frequency and glottal posture control. *J Acoust Soc Am* 135(4):2052–2064
- Choi H-S, Berke GS, Ye M, Kreiman J (1993) Function of the posterior cricoarytenoid muscle in phonation: in vivo laryngeal model. *Otolaryngol Head Neck Surg* 109(6):1043–1051
- Choi HS, Ye M, Berke GS (1995) Function of the interarytenoid (ia) muscle in phonation: in vivo laryngeal model. *Yonsei Med J* 36(1):58–67
- DeJonckere PH, Lebacqz J (2020) In vivo quantification of the intra-glottal pressure: modal phonation and voice onset. *J Voice* 34(4):645–e19
- Deng JJ, Hadwin PJ, Peterson SD (2019) The effect of high-speed videoendoscopy configuration on reduced-order model parameter estimates by Bayesian inference. *J Acoust Soc Am* 146(2):1492–1502
- Deng JJ, Serry MA, Zañartu M, Erath BD, Peterson SD (2022) Modeling the influence of covid-19 protective measures on the mechanics of phonation. *J Acoust Soc Am* 151(5):2987–2998
- Diaz-Cadiz M, McKenna VS, Vojtech JM, Stepp CE (2019) Adductory vocal fold kinematic trajectories during conventional versus high-speed videoendoscopy. *J Speech Lang Hear Res* 62(6):1685–1706
- Francis AL, Ciocca V, Wong VKM, Chan JKL (2006) Is fundamental frequency a cue to aspiration in initial stops? *J Acoust Soc Am* 120(5):2884–2895
- Fulcher LP, Scherer RC, Powell T (2013) Viscous effects in a static physical model of the uniform glottis. *J Acoust Soc Am* 134(2):1253–1260
- Galindo GE, Peterson SD, Erath BD, Castro C, Hillman RE, Zañartu M (2017) Modeling the pathophysiology of phonotraumatic vocal hyperfunction with a triangular glottal model of the vocal folds. *J Speech Lang Hear Res* 60(9):2452–2471
- Galindo GE, Zañartu M, Yuz JI (2014) A discrete-time model for the vocal folds. In: *IEEE EMBS international student conference*, pp 74–77
- Geng B, Movahhedi M, Xue Q, Zheng X (2021) Vocal fold vibration mode changes due to cricothyroid and thyroarytenoid muscle interaction in a three-dimensional model of the canine larynx. *J Acoust Soc Am* 150(2):1176–1187
- Goberman AM, Blomgren M (2008) Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease. *J Voice* 22(2):178–191
- Gómez-Vilda P, Fernández-Baillo R, Nieto A, Díaz F, Fernández-Camacho FJ, Rodellar V, Martínez R (2007) Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. *J Voice* 21(4):450–476
- Hanson HM (2009) Effects of obstruent consonants on fundamental frequency at vowel onset in English. *J Acoust Soc Am* 125(1):425–441
- Heller Murray ES, Lien Y-AS, Van Stan JH, Mehta DD, Hillman RE, Pieter Noordzij J, Stepp CE (2017) Relative fundamental frequency distinguishes between phonotraumatic and non-phonotraumatic vocal hyperfunction. *J Speech Lang Hear Res* 60(6):1507–1515
- Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C (1989) Objective assessment of vocal hyperfunction: an experimental framework and initial results. *J Speech Lang Hear Res* 32(2):373–392
- Hillman RE, Stepp CE, Van Stan JH, Zañartu M, Mehta DD (2020) An updated theoretical framework for vocal hyperfunction. *Am J Speech Lang Pathol* 29(4):2254–2260
- Hombert J-M, Ohala JJ, Ewan WG (1979) Phonetic explanations for the development of tones. *Language* 55(1):37–58
- Ishizaka K, Flanagan JL (1972) Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst Tech J* 51(6):1233–1268
- Kelly JL, Lochbaum CC (1962) Speech synthesis. In: *Proceedings of the fourth international congress on acoustics*
- Kirby JP, Ladd DR (2016) Effects of obstruent voicing on vowel f0: Evidence from “true voicing” languages. *J Acoust Soc Am* 140(4):2400–2411
- Lagier A, Legou T, Galant C, de La Amy, Bret'eque B, Meynadier Y, Giovanni A (2017) The shouted voice: a pilot study of laryngeal physiology under extreme aerodynamic pressure. *Logop Phoniatr Vocol* 42(4):141–145
- Lebacqz J, DeJonckere PH (2019) The dynamics of vocal onset. *Biomed Signal Process Control* 49:528–539
- Lien Y-AS, Gattuccio CI, Stepp CE (2014) Effects of phonetic context on relative fundamental frequency. *J Speech Lang Hear Res* 57(4):1259–1267
- Lien Y-AS, Michener CM, Eadie TL, Stepp CE (2015) Individual monitoring of vocal effort with relative fundamental frequency: relationships with aerodynamics and listener perception. *J Speech Lang Hear Res* 58(3):566–575
- Liljencrants J (1985) Speech synthesis with a reflection-type line analog (Unpublished doctoral dissertation). Royal Institute of Technology
- Lisker L, Abramson AS (1967) Some effects of context on voice onset time in English stops. *Lang Speech* 10(1):1–28
- Löfqvist A, Baer T, McGarr NS, Story RS (1989) The cricothyroid muscle in voicing control. *J Acoust Soc Am* 85(3):1314–1321
- Löfqvist A, Koenig LL, McGowan RS (1995) Vocal tract aerodynamics in/aca/utterances: measurements. *Speech Commun* 16(1):49–66
- Lowell SY, Story BH (2006) Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration. *J Acoust Soc Am* 120(1):386–397
- Lucero JC (1996) Relation between the phonation threshold pressure and the prephonatory glottal width in a rectangular glottis. *J Acoust Soc Am* 100(4):2551–2554
- Lucero JC, Koenig LL (2007) On the relation between the phonation threshold lung pressure and the oscillation frequency of the vocal folds. *J Acoust Soc Am* 121(6):3280–3283
- Lucero JC, Schoentgen J (2015) Smoothness of an equation for the glottal flow rate versus the glottal area. *J Acoust Soc Am* 137(5):2970–2973
- McKenna VS, Heller Murray ES, Lien Y-AS, Stepp CE (2016) The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *J Speech Lang Hear Res* 59(6):1283–1294

- Mergell P, Herzel H, Wittenberg T, Tigges M, Eysholdt U (1998) Phonation onset: vocal fold modeling and high-speed glottography. *J Acoust Soc Am* 104(1):464–470
- Mohr B (1971) Intrinsic variations in the speech signal. *Phonetica* 23(2):65–93
- Movahhedi M, Geng B, Xue Q, Zheng X (2021) Effects of cricothyroid and thyroarytenoid interaction on voice control: muscle activity, vocal fold biomechanics, flow, and acoustics. *J Acoust Soc Am* 150(1):29–42
- Newmark NM (1959) A method of computation for structural dynamics. *J Eng Mech Div* 85(3):67–94
- Ohde RN (1984) Fundamental frequency as an acoustic correlate of stop consonant voicing. *J Acoust Soc Am* 75(1):224–230
- Park Y, Wang F, Díaz-Cádiz M, Vojtech JM, Groll MD, Stepp CE (2021) Vocal fold kinematics and relative fundamental frequency as a function of obstruent type and speaker age. *J Acoust Soc Am* 149(4):2189–2199
- Patel RR, Forrest K, Hedges D (2017) Relationship between acoustic voice onset and offset and selected instances of oscillatory onset and offset in young healthy men and women. *J Voice* 31(3):389. e9-389.e17
- Pelorson X, Hirschberg A, Van Hassel R, Wijnands A, Auregan Y (1994) Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *J Acoust Soc Am* 96(6):3416–3431
- Poletto CJ, Verdun LP, Strominger R, Ludlow CL (2004) Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures. *J Appl Physiol* 97(3):858–866
- Roy N, Fetrow RA, Merrill RA, Dromey C (2016) Exploring the clinical utility of relative fundamental frequency as an objective measure of vocal hyperfunction. *J Speech Lang Hear Res* 59(5):1002–1017
- Serry MA, Stepp CE, Peterson SD (2021) Physics of phonation offset: towards understanding relative fundamental frequency observations. *J Acoust Soc Am* 149(5):3654–3664
- Shiba TL, Chhetri DK (2016) Dynamics of phonatory posturing at phonation onset. *Laryngoscope* 126(8):1837–1843
- Šidlof P, Doaré O, Cadot O, Chaigne A (2011) Measurement of flow separation in a human vocal folds model. *Exp Fluids* 51(1):123–136
- Smith AB, Robb MP (2013) Factors underlying short-term fundamental frequency variation during vocal onset and offset. *Speech Lang Hear* 16(4):208–214
- Sonesson B (1982) Vocal fold kinesiology. In: Grillner S, Lindblom B, Lubker J, Persson A (eds) *Speech motor control*. Pergamon, Oxford, pp 113–117
- Steinecke I, Herzel H (1995) Bifurcations in an asymmetric vocal-fold model. *J Acoust Soc Am* 97(3):1874–1884
- Stepp CE, Hillman RE, Heaton JT (2010) The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *J Speech Lang Hear Res* 53(5):1220–1226
- Stepp CE, Hillman RE, Heaton JT (2010) A virtual trajectory model predicts differences in vocal fold kinematics in individuals with vocal hyperfunction. *J Acoust Soc Am* 127(5):3166–3176
- Stepp CE, Merchant GR, Heaton JT, Hillman RE (2011) Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *J Speech Lang Hear Res* 54(5):1260–1266
- Story BH (2005) A parametric model of the vocal tract area function for vowel and consonant simulation. *J Acoust Soc Am* 117(5):3231–3254
- Story BH, Titze IR (1995) Voice simulation with a body-cover model of the vocal folds. *J Acoust Soc Am* 97(2):1249–1260
- Sváček P, Horáček J (2018) Finite element approximation of flow induced vibrations of human vocal folds model: effects of inflow boundary conditions and the length of subglottal and supraglottal channel on phonation onset. *Appl Math Comput* 319:178–194
- Takemoto H, Honda K, Masaki S, Shimada Y, Fujimoto I (2006) Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *J Acoust Soc Am* 119(2):1037–1049
- Thomson SL, Mongeau L, Frankel SH (2005) Aerodynamic transfer of energy to the vocal folds. *J Acoust Soc Am* 118(3):1689–1700
- Titze IR (1988) The physics of small-amplitude oscillation of the vocal folds. *J Acoust Soc Am* 83(4):1536–1552
- Titze IR (2002) Regulating glottal airflow in phonation: application of the maximum power transfer theorem to a low dimensional phonation model. *J Acoust Soc Am* 111(1):367–376
- Titze IR (2004) A theoretical study of f0–f1 interaction with application to resonant speaking and singing voice. *J Voice* 18(3):292–298
- Titze IR, Hunter EJ (2007) A two-dimensional biomechanical model of vocal fold posturing. *J Acoust Soc Am* 121(4):2254–2260
- Titze IR, Luschei ES, Hirano M (1989) Role of the thyroarytenoid muscle in regulation of fundamental frequency. *J Voice* 3(3):213–224
- Titze IR, Schmidt SS, Titze MR (1995) Phonation threshold pressure in a physical model of the vocal fold mucosa. *J Acoust Soc Am* 97(5):3080–3084
- Titze IR, Story BH (2002) Rules for controlling low-dimensional vocal fold models with muscle activation. *J Acoust Soc Am* 112(3):1064–1076
- Van den Berg J, Zantema J, Doornenbal P Jr (1957) On the air resistance and the bernoulli effect of the human larynx. *J Acoust Soc Am* 29(5):626–631
- Weibel ER, Cournand AF, Richards DW (1963) *Morphometry of the human lung*, vol 1. Springer, New York
- Yin J, Zhang Z (2013) The influence of thyroarytenoid and cricothyroid muscle activation on vocal fold stiffness and eigenfrequencies. *J Acoust Soc Am* 133(5):2972–2983
- Yin J, Zhang Z (2014) Interaction between the thyroarytenoid and lateral cricoarytenoid muscles in the control of vocal fold adduction and eigenfrequencies. *J Biomech Eng* 136(11):111006
- Zañartu M, Galindo GE, Erath BD, Peterson SD, Wodicka GR, Hillman RE (2014) Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction. *J Acoust Soc Am* 136(6):3262–3271
- Zhang Z (2009) Characteristics of phonation onset in a two-layer vocal fold model. *J Acoust Soc Am* 125(2):1091–1102
- Zhang Z (2011) On the difference between negative damping and eigenmode synchronization as two phonation onset mechanisms. *J Acoust Soc Am* 129(4):2163–2167
- Zhang Z (2016) Respiratory laryngeal coordination in airflow conservation and reduction of respiratory effort of phonation. *J Voice* 30(6):760–e7

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.