Article

Mapping Thyroarytenoid and Cricothyroid Activations to Postural and Acoustic Features in a Fiber-Gel Model of the Vocal Folds

Anil Palaparthi [1,2,]*, Simeon Smith [1] and Ingo R. Titze [1,2]

[1]   National Center for Voice and Speech, The University of Utah, 1901 S Campus Dr, Suite 2120, Salt Lake City, UT 84112, USA; simeon.smith@utah.edu (S.S.); ingo.titze@utah.edu (I.R.T.)

[2]   Department of Bioengineering, The University of Utah, Salt Lake City, UT 84112, USA

[*]   Correspondence: anil.palaparthi@utah.edu

check for updates

Featured Application: The goal of this research publication is to advance the current state of computer simulation of voice production. Many future questions about the physiology, biomechanics, and acoustics of vocalization are expected to be answered by simulation to avoid excessive use of animals and human subjects in experimentation. Here we feature the control of vocal output with muscle activation as inputs, a particularly difficult experimental procedure to conduct on animals and human volunteers.

Abstract: Any specific vowel sound that humans produce can be represented in terms of four perceptual features in addition to the vowel category. They are pitch, loudness, brightness, and roughness. Corresponding acoustic features chosen here are fundamental frequency (f_0), sound pressure level (SPL), normalized spectral centroid (NSC), and approximate entropy (ApEn). In this study, thyroarytenoid (TA) and cricothyroid (CT) activations were varied computationally to study their relationship with these four specific acoustic features. Additionally, postural and material property variables such as vocal fold length (L) and fiber stress (σ) in the three vocal fold tissue layers were also calculated. A fiber-gel finite element model developed at National Center for Voice and Speech was used for this purpose. Muscle activation plots were generated to obtain the dependency of postural and acoustic features on TA and CT muscle activations. These relationships were compared against data obtained from previous in vivo human larynx studies and from canine laryngeal studies. General trends are that f_0 and *SPL* increase with CT activation, while NSC decreases when CT activation is raised above 20%. With TA activation, acoustic features have no uniform trends, except SPL increases uniformly with TA if there is a co-variation with CT activation. Trends for postural variables and material properties are also discussed in terms of activation levels.

Keywords: intrinsic laryngeal muscle activations; voice acoustics; vocal fold posturing; voice production

1. Introduction

This study was motivated by a desire to eventually control a voice simulator with inputs related to perception. Aside from vowel perception, a sound can be represented in terms of pitch, loudness, and timbre [1]. Timbre is the quality of the sound that differentiates one sound from another when pitch and loudness are the same [2]. In a first attempt, it can be divided into two components, brightness and roughness [3]. Acoustically, the four perceptual features can be quantified with fundamental frequency, sound pressure level, spectral content, and aperiodicity. Neglecting additive noise from air turbulence (breathiness or aspiration in a vowel), it is believed that these four perceptual or acoustic features contain many of the characteristics of a given vowel sound [4]. Quantifying fundamental frequency

and sound intensity is well described in the literature. The fundamental frequency of a sound can be measured directly from the oral pressure signal using standard algorithms such as autocorrelation, zero-crossing rate or with more advanced algorithms such as the Yin algorithm [5] or SWIPE' [6]. Sound pressure level, the logarithmic expression of relative sound intensity, can be computed from the radiated pressure signal according to acoustic standards [7]. Brightness is much more abstract and has no agreed-upon standards for calculation on the acoustic signal. Some acoustic features that represent brightness are the total harmonic distortion [8] and the normalized spectral centroid [2,9]. Roughness also has no standard definition in the literature other than it being attributed to the component of the sound that is not regular (periodic) [10,11]. However, even periodic sounds can be perceived as rough if multiple harmonics are contained in a critical band [12]. Measures of aperiodicity roughness are jitter and shimmer [13], while a measure of lack of roughness is the cepstral peak prominence (CPP) [14–16]. Non-linear dynamic metrics such as correlation dimension [17,18] and various entropy measures [18,19] are also being studied to accurately quantify aperiodicity. In this study, roughness is considered to be related to the power in the sound that is not part of the fundamental and its harmonic series. Subharmonics and sideband frequencies are included in the quantification of roughness.

If the vocal tract configuration (vowel) is kept constant during phonation, all variations in the four perceptual features depend on changes in the vocal source, but the interaction with the vocal tract may vary with fundamental frequency [20]. The laryngeal configuration, and specifically the vocal fold posture, depend on the laryngeal muscle activations [21]. Vocal fold oscillation then depends on the amount of lung pressure available to set the vocal folds into motion. In this study, we focused on the interrelationship between the thyroarytenoid (TA) and cricothyroid (CT) muscle activations to achieve these four specific physical attributes of a sound as well as vocal fold posturing (setting) features. While our computational model includes postural mechanics derived from five intrinsic muscles (lateral cricoarytenoid (LCA), interarytenoid (IA), and posterior cricoarytenoid (PCA) in addition to CT and TA), preliminary results showed that random combinations of five muscle activations and lung pressure produced too complex relationships to comprehend easily. Hence, a heuristic approach was taken to constrain the LCA, IA, PCA, and lung pressure (P_L) for the purpose of targeting self-sustained oscillation. Future studies with computational learning strategies will address the more intricate relations with all the muscle activations.

The vocal tract configuration remained constant throughout the phonations (an /α/ vowel). The posturing variables included are vocal fold length (L) and tissue fiber stress in the superficial layer of the lamina propria (σ_{SLLP}), ligament (σ_{lig}), and TA muscle (σ_{mus}). These variables were chosen because they are essential for initiation and self-sustaining vocal fold oscillation at a desired fundamental frequency and intensity. Researchers have made attempts to find these relationships using electromyographic recordings in humans [22–26], in-vivo canine larynges experiments [21,27–31], and computational models [32–34]. However, majority of these attempts were made to identify the relationship with fundamental frequency only and to a lesser extent with vocal intensity [35,36]. In this study, we use computational tools to extend to all the four acoustic and two postural features. Additionally, we compare the results from this study with results published in some previous studies.

The earlier works of electromyography recordings in humans was done by Faaborg-Andersen (1957, 1965) [22,23], Hirano et al. (1969) [24], and Gay et al. (1972) [25]. Hirano et al. (1969) recorded the electromyography activity of three intrinsic laryngeal muscles in the regulation of fundamental frequency and vocal intensity in six subjects. The muscles studied were CT, IA, and the vocalis portion of the TA muscles. The results suggested that all three muscles were involved in regulating fundamental frequency and intensity, with f_o and intensity increasing with an increase in CT activation. Gay et al. (1972) also found that CT plays a major role among all the five intrinsic laryngeal muscles in the control of fundamental frequency, and subglottal pressure in the control of intensity. With electromyography conducted on four subjects, Titze et al., (1989) [32] found that f_o can both increase and decrease with increased TA activity, but only increases with CT activity. More recent studies in humans focused on

intrinsic laryngeal muscle coordination during normal speech, respiration, and swallowing [37–39]. Hillel, (2001) [38] found that TA and LCA activity decreased after initial burst preceding phonation onset, while the IA sustained glottis position during phonation.

Animal model studies on muscle activations predominantly focused on canines because their vocal fold physiology is close to that of humans. Earlier in vivo canine model studies used electrical stimulation to study the function of IA [27,28] and PCA [29] muscles during phonation. The IA was thought to mainly control the amount of subglottic pressure available during phonation. Recent studies using in vivo canine models also used electrical stimulation but with much finer control to investigate the effect of intrinsic laryngeal muscle activation on the fundamental frequency and glottal posture [30], as well as on fundamental frequency and intensity [31]. The muscles were activated independently from threshold to maximal contraction while airflow was increased to phonation onset and beyond. It was observed that onset frequency was primarily affected by CT activation, and TA increased or decreased f_o and sound pressure level (SPL) under other muscle activation conditions. The LCA/IA activity maintained vocal fold adduction at higher subglottic pressure levels. Sound pressure level was highly correlated with subglottal pressure in all conditions. They also found that the same f_o and SPL could be achieved with a variety of muscle activation combinations, as reported earlier in [32].

Computational models have been used to study the effects of intrinsic laryngeal muscle activations on fundamental frequency and intensity. Farley (1996) [33] used a simplified mathematical model of the larynx to find that f_o is simultaneously controlled by TA, cricothyroid pars oblique (CTO), and cricothyroid pars recta (CTR). Titze and Story (2002) [34] modeled CT, TA, LCA, and PCA muscle activations along with lung pressure to control a three-mass vocal fold model. The results showed that oscillation regions in muscle activation control spaces are similar to those measured in human subjects. Lowell and Story (2006) [40] also used a three-mass vocal fold model for voice simulation in adult males to study the role of individual CT and TA muscle activities on f_o. They found that largest f_o change with CT activation was observed at low TA levels.

Coming to the posturing features, Chhetri et al. (2010) [21] used an in vivo canine model to study the role of intrinsic laryngeal muscle activations on vocal fold length. At lower superior laryngeal nerve (SLN) stimulation levels, they observed linear change in length, and at higher levels, the strain reached a plateau. Chhetri et al. (2012) [30] also included glottal distance at the vocal processes as one of the posturing features. They observed that LCA/IA activation primarily closed the cartilaginous glottis and TA activation closed the mid-membranous glottis. Vahabzadeh-Hagh et al. (2017) [41] used an in vivo canine hemilarynx model to study the role of paired intrinsic laryngeal muscles on three-dimensional vocal fold postural changes. They found that combined TA and CT activation yields a rectangular glottal surface, and further addition of TA yields a divergent glottis.

The current study focuses on the role that TA and CT muscle activations play in controlling the four acoustic variables (f_o, SPL, spectral content, and aperiodicity) and two postural variables (length and tissue fiber stress in the three vocal fold layers). To the best of our knowledge, there is no previous study that focused on the combination of these variables for self-sustained vocal fold oscillation in a finite-element computational model. In this study, we used the fiber-gel finite element model, a muscle activation-based model that produces flow-induced oscillation. The model accepts five intrinsic laryngeal muscle pairs as inputs to set up prephonatory glottal shapes. Lung pressure then drives the vocal folds into self-sustained oscillation. The outputs of the model are many time-varying signals, including oral pressure, oral flow, glottal flow, glottal flow derivative, and transglottal pressure. The findings from the fiber-gel model are compared to the results from human electromyography (EMG) studies (Titze et al. 1989) [32], and in vivo canine model studies (Chhetri et al. (2012) and Chhetri and Park, 2016) [30,31].

2. Materials and Methods

2.1. Fiber-Gel Finite Element Model

The fiber-gel model combines a viscoelastic ground substance (a gel) with directional fibers under tension in all layers of the vocal fold (superficial layer of the lamina propria (SLLP), ligament, and muscle) [42]. At the physiologic level of input, the control parameters are lung pressure and activation levels of five intrinsic laryngeal muscle pairs (if left-right symmetry is assumed), CT, TA, LCA, PCA, and IA. With these inputs, and stored anatomical dimensions of the larynx ([43], Chapter. 1) and the airways [44], pre-phonatory vocal fold and vocal tract shapes are produced. Figure 1 shows how all five intrinsic muscles are contributing to the vocal fold length, L and the position (ξ_{02}, ψ_{02}) of the vocal process of the arytenoid cartilage in the horizontal plane. The plane is at the level of the vocal ligament. All muscles except the CT apply forces directly to the arytenoid cartilage. The CT has a forward and downward pull on the thyroid cartilage relative to the cricoid cartilage, therewith elongating the vocal folds. The dotted-line vector for LCA indicates that the fibers run below those of the TA, while the dotted line for the PCA vector indicate an effective force direction due to curvilinear fiber attachment. The combined mechanics of the cartilages is solved with the second-order Runge-Kutta differential method [43]. Only right-side cartilages and forces were shown in Figure 1 for clarity. The mechanics of the cartilages on the left side are solved similarly.
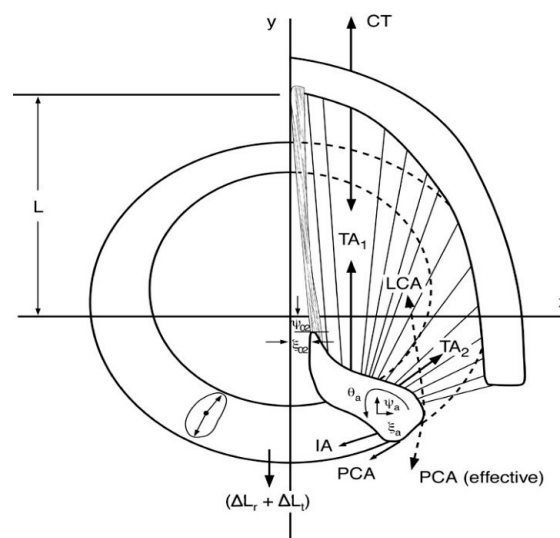
Figure 1. Forces acting on arytenoid cartilage for cricoarytenoid movements (Titze, 2006 [43]).

In the rostral-cranial (vertical) direction, the medial surface of the vocal folds is defined with three parameters, the vocal fold thickness T, the lower adduction of the arytenoid cartilage ξ_{01}, and a medial surface bulging factor ξ_b. With these parameters, the glottal half-width at any point (y, z) along the surface of each vocal fold is defined as

$$\xi_0(y,z) = (1 - y/L)[\xi_{02} + (\xi_{01} - \xi_{02} - 4\xi_b z/T)(1 - z/T)] \tag{1}$$

The parameters in these equations are determined by muscle activations with empirical rules

$$L = L_0(1 + \varepsilon) \tag{2}$$

$$T = T_0/(1 + \varepsilon) \tag{3}$$

$$\xi_{01} = \xi_{02} + \max(0, 0.1T(1 - 1.5a_{TA})) \tag{4}$$

$$\xi_b = 0.05T(1 - a_{TA}) \tag{5}$$

where ε is the vocal fold strain (determined by all muscles) and a_{TA} is the activation of the TA muscle. Here, L_0 was set to be 1.6 cm, and T_0 was set to be 0.7 cm for an average male vocal fold. Figure 2a shows the right vocal fold geometry as parameterized by the empirical rules given above. Also shown is the variable ξ_m, the amplitude of vibration when self-sustained oscillation takes place. The left vocal fold is also governed by the same parameters. The geometric properties of the layers are modeled according to the histological data available from the literature [45–47]. The full details of vocal fold posturing with intrinsic laryngeal muscle activations is provided in ([43], Chapter 3).

(a)　　　　　　　　　　　　　　　　　(b)
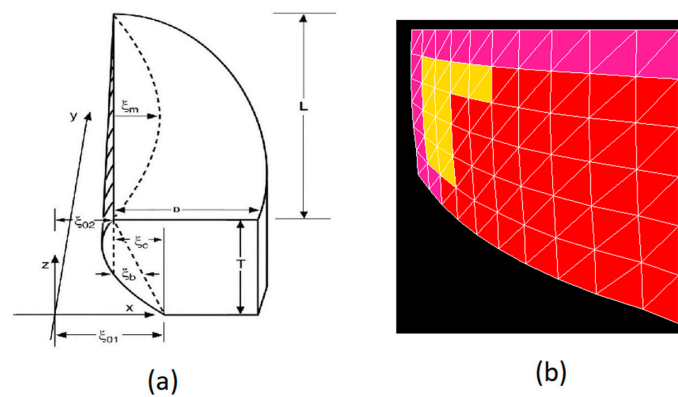
Figure 2. (**a**) Right vocal fold geometry in the Fiber-Gel Model, and (**b**) coronal section of the right vocal fold in finite element implementation. Purple is the superficial layer of the lamina propria (SLLP), yellow is the ligament, and red is the thyroarytenoid (TA) muscle.

As tissue is deformed with muscle contraction (e.g., length, thickness, and medial surface of the vocal folds), viscoelastic parameters are modified with basic constitutive equations. The passive fibrous tissue properties in the y-direction in the three vocal fold layers and the other intrinsic laryngeal muscles are characterized with a combined linear and exponential stress function ([43], p. 77):

$$\sigma_y \;=\; -\sigma_0(\varepsilon - \varepsilon_1) + \sigma_2\Big[e^{B(\varepsilon - \varepsilon_2)} - 1 - B(\varepsilon - \varepsilon_2)\Big] \tag{6}$$

In the above equation, σ_0 and σ_2 are scale factors, ε_1 is the strain where the linear function begins and ε_2 is the strain where the exponential term begins. The stress σ_y is converted into an equivalent shear modulus μ′, to which a gel shear modulus μ is added for transversally isotropic gel properties. In combination, the viscoelastic model becomes a fiber-gel model. The active properties of all intrinsic muscles are determined by a modified Kelvin model ([43], p. 76). The modified Kelvin model is governed by maximum isometric active stress, σ_m. Table 1 shows the parameters of Equation (6) and σ_m used in this study. The full details of the viscoelastic properties obtained from measurements on multiple species' biological tissues are given in ([43], Chapter 2).

Table 1. Parameter values used in fiber-gel finite element model for modeling the five intrinsic laryngeal muscles (CT, LCA, TA, IA, and PCA) along with vocal ligament (LIG) and SLLP layers.

	CT	LCA	TA	IA	PCA	LIG	SLLP
$\sigma_0\left(\text{dyn/cm}^2\right)$	2.2×10^4	3×10^4	2×10^4	2×10^4	5×10^4	2×10^4	0.2×10^4
$\sigma_2\left(\text{dyn/cm}^2\right)$	5×10^4	59×10^4	1.5×10^4	30×10^4	55×10^4	0.15×10^4	1.5×10^4
B	7	4	6.5	3.5	5.3	12.8	6.5
ε_1	−0.9	−0.9	−0.9	−0.9	−0.9	−0.9	−0.9
ε_2	−0.06	0.05	−0.5	0	0.1	−0.5	−0.2
$\sigma_m\left(\text{dyn/cm}^2\right)$	400×10^4	140×10^4	180×10^4	140×10^4	96×10^4		

CT-cricothyroid; LCA-lateral cricothyroid; TA-thyroarytenoid; IA-interarytenoid; PCA-posterior cricoarytenoid; SLLP-superficial layer of lamina propria.

Each vocal fold consisted of three layers: SLLP, vocal ligament, and TA muscle. A two-dimensional finite-element mesh as seen in Figure 2b was constructed for each of $M = 15$ coronal slices in the posterior-anterior direction of each vocal fold, following [48]. The elements were triangular, allowing the shape functions (interpolations within the elements) to be computed analytically from the nodal points. This choice was based on speed of computation. There were two triangular elements per cell, with $L = 10$ columns of cells and $N = 6$ rows of cells in the coronal plane ([43], pp. 214–231). In solving the matrix equations for $(L+1)*(N+1)$ nodal displacements, aerodynamic forces were applied at the open boundaries and string (fiber) forces were applied between the slices. The material in each slice was considered incompressible and transversally isotropic [48]. Thus, one elastic constant (the transverse shear modulus $\mu = 0.5$ kPa for all the layers) and a viscosity η (= 0.1 Pa-s for all the layers) defined the gel property. A second shear modulus μ' defined the fiber property (Equation (6) above). Boundaries were fixed at anterior, lateral, and posterior surfaces. Boundaries were free to move with pressure distributions on inferior, medial, and superior surfaces. The lateral boundary was curved as shown in Figure 2a.

The vibrational displacement was superimposed on the postural displacement. A Bernoulli approach was used to calculate glottal pressures up to the minimum glottal diameter, from which point jet flow was assumed in the remainder of the glottis ([43], Chapter 7). When lung pressure is turned on, airflow and surface pressures are computed in the glottis (air space between the vocal folds) and on all surfaces along the vocal tract using wave-reflection algorithm [43,49,50]. The vocal tract airways were spatially sampled with sections of 0.3968 cm length, half the distance sound travels in 1/44,100 s. There were 36 subglottal sections and 44 supraglottal sections. The choice of section length allowed forward and backward travelling waves to be computed at section boundaries at a sampling rate of 44.1 kHz, in synchrony with wave propagation to guarantee computational stability. Energy dissipation was included by incorporating viscous losses and yielding wall losses within each section, as well as kinetic losses at section boundaries [51]. Radiation from the mouth was computed with a low-frequency parallel inertance/resistance equivalent of the radiation impedance ([43], Chapter 6, [52]).

The vocal folds will self-sustain oscillation if vocal fold adduction, glottal geometry, and elastic and viscous properties are such that phonation threshold pressure (PTP) is exceeded with the applied lung pressure. The oscillation modulates the airflow, producing oscillatory flow and acoustic pressures throughout the airways. The basic vibrational and acoustic underpinnings for the fiber-gel finite element model are detailed in Titze (2006 [43], Chapters 6 and 7)). More recent applications and validations have been published in [51,53,54].

2.2. Data Collection

A brute force approach was used to vary TA and CT muscle activations. TA and CT muscle activations were varied between 0% and 100% of their maximum muscle activation levels in steps of 5%. Lung pressure, LCA, PCA, and IA activations were initially also varied randomly for the given CT and TA activations, but ultimately needed to be constrained to focus only on cases of self-sustained oscillation (a small percentage of the initial exploratory set). The constraint equations are given in the Results section. A total of 441 signals were generated using this approach. Each signal was 0.4 s in duration. The waveforms, oral pressure (p_o), vocal fold length (L), and fiber stress (σ_y, see Equation (6)) in the three vocal fold layers (SLLP, vocal ligament, and TA muscle) were recorded for every signal. Using these waveforms, acoustic and postural features were computed. Fundamental frequency was computed using the SWIPE′ algorithm [6]. Sound pressure level radiated at 30 cm from mouth was computed using the following equation [7]:

$$SPL \;=\; 10\log_{10}\frac{I}{I_0} \;=\; 10\log_{10}\frac{\left(\bar{p}_o\right)^2}{4\pi R^2 R_m I_0} \tag{7}$$

where the bar denotes a time-average of the oral pressure, p_o, R is the radius from mouth, R_m is radiation resistance [52,55], and I_0 is the standard reference intensity (10^{-12} watt/m^2).

Normalized spectral centroid was computed to quantify brightness [9]. Spectral centroid (SC) measures the center of mass of the spectrum in Hz. It is computed using the following equation:

$$SC = \frac{\sum_{k=0}^{N} f_k x_k}{\sum_{k=0}^{N} x_k} \tag{8}$$

Here, N is total number of frequency bins, f_k is the center frequency of the bin in Hz, and x_k is the spectral magnitude in bin k. The normalized spectral centroid was then computed by dividing the spectral centroid by the fundamental frequency of the signal.

Approximate entropy (ApEn) quantifies the amount of irregularity and unpredictability present in a signal [56]. For a given time series signal $x(t)$, an m-dimensional delay-coordinate phase space $X = \{x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)\}$ was first constructed [18]. Here, m is the embedding dimension and τ is the time delay. Then, ApEn can be computed using the following equation:

$$ApEn(m, r) = \lim_{N \to \infty} \left[\Phi^m(r) - \Phi^{m+1}(r) \right] \tag{9}$$

where

$$\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log C_i^m(r) \tag{10}$$

Here, $C_i^m(r)$ is the correlation integral computed as suggested in [56], r is the radius of similarity (chosen as 0.2 *variance(x)), and N is the number of samples of the signal $x(t)$. In the current study, m was set to 2, and τ was set to 1 for all the signals.

3. Results

3.1. Lung Pressure

Only the cases for which phonation threshold pressure was exceeded were generated. To guarantee these conditions, lung pressure was increased with both CT activation and TA activation according to the following relation:

$$P_L = 0.8 + 0.025 a_{CT} + 0.01 a_{TA} \text{ kPa} \tag{11}$$

where a_{CT} and a_{TA} are normalized activation levels ranging from 0% to 100%.

The increase with CT activation needed to be greater than the increase with TA activation to remain above threshold for self-sustained oscillation. The equation results in a range of 0.8 kPa and 4.3 kPa for the lung pressure. Figure 3 shows a muscle activation plot with lung pressure as the contour parameter.
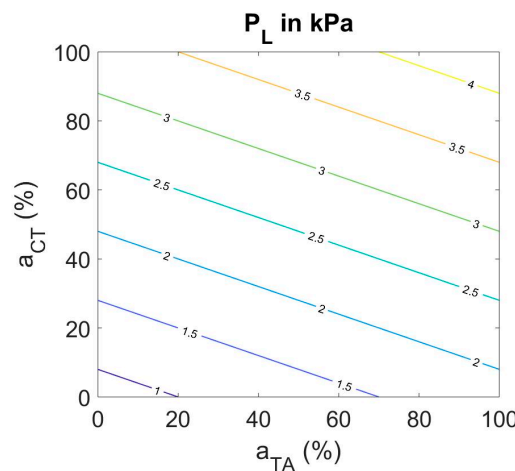

Figure 3. Muscle activation plot with lung pressure as the contour parameter.

Note that lung pressures up to 4.3 kPa are needed in this model to obtain self-sustained oscillation, but we do not claim that the simple linear relation in Equation (9) represents threshold pressure or an equal fraction above threshold.

3.2. Muscle Activation for Adduction

To guarantee self-sustained oscillation, LCA and IA activation needed to be kept in a relatively smaller range. Phonation threshold pressure is highly sensitive to vocal fold adduction [57]. The following relation was used:

$$a_{LC} = a_{IA} = 30 + 0.3a_{CT} - 0.1a_{TA} \tag{12}$$

where a_{LC} is the activation level of the LCA muscle normalized to its maximum value for forceful adduction and a_{IA} is the activation level of the IA muscle normalized to its maximum value. Figure 4 shows the muscle activation plot with the balanced LCA/IA combination as the contour parameter. A larger increase in LCA/IA activation with CT activation was needed than with TA activation because tissue incompressibility demands that the vocal folds abduct slightly when they are elongated. By contrast, TA activation shortens the vocal folds and adducts them slightly. Note that, overall, the LCA/IA activation range for self-sustained oscillation is between 30% and about 60% over the entire range of TA and CT activations.
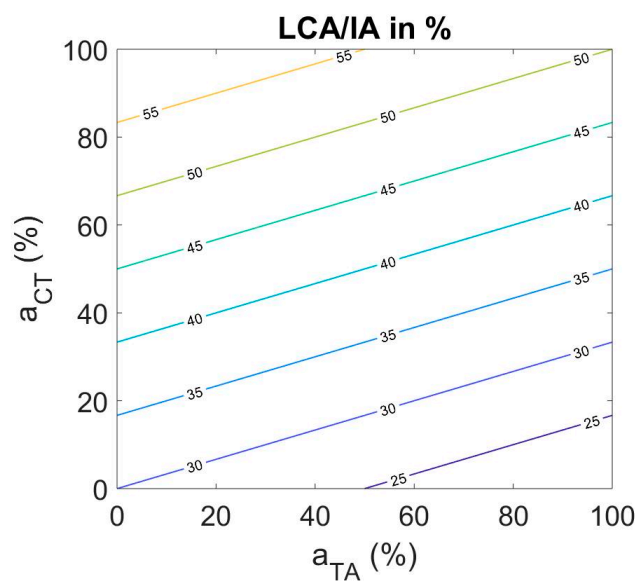
Figure 4. Muscle activation plot with LCA/IA activation as contour parameter.

Variations with posterior cricoarytenoid (PCA) activation were not included in this first attempt to address the mapping from physiologic input to acoustic output. PCA activation was kept constant at 0%. It is understood that the PCA muscle can be active in phonation. It is used for high-pitched phonation to resist the anterior pull of the arytenoid cartilage by the CT muscle. At this stage in the modeling of the laryngeal framework mechanics, we have included this resistance to anterior movement by increasing the stiffness of the cricoarytenoid joint in a nonlinear fashion, thereby limiting anterior movement of the cartilage.

3.3. Acoustic Features

Physiologic inputs and control parameters were mapped to the four acoustic features fundamental frequency, sound pressure level, normalized spectral centroid, and approximate entropy.

Muscle activation plots were generated for each of the four acoustic features with respect to CT and TA muscle activation levels.

3.3.1. Fundamental Frequency

Fundamental frequency (f_o) increased with increase in CT activity for all TA activation levels (as seen in Figure 5). To the contrary, f_o increased with increase in TA activity at low CT activation levels and decreased with increase in TA activity at high CT activation levels. The range of f_o achieved with the fiber-gel model is between 70 and 400 Hz with anatomical dimensions adjusted for males.
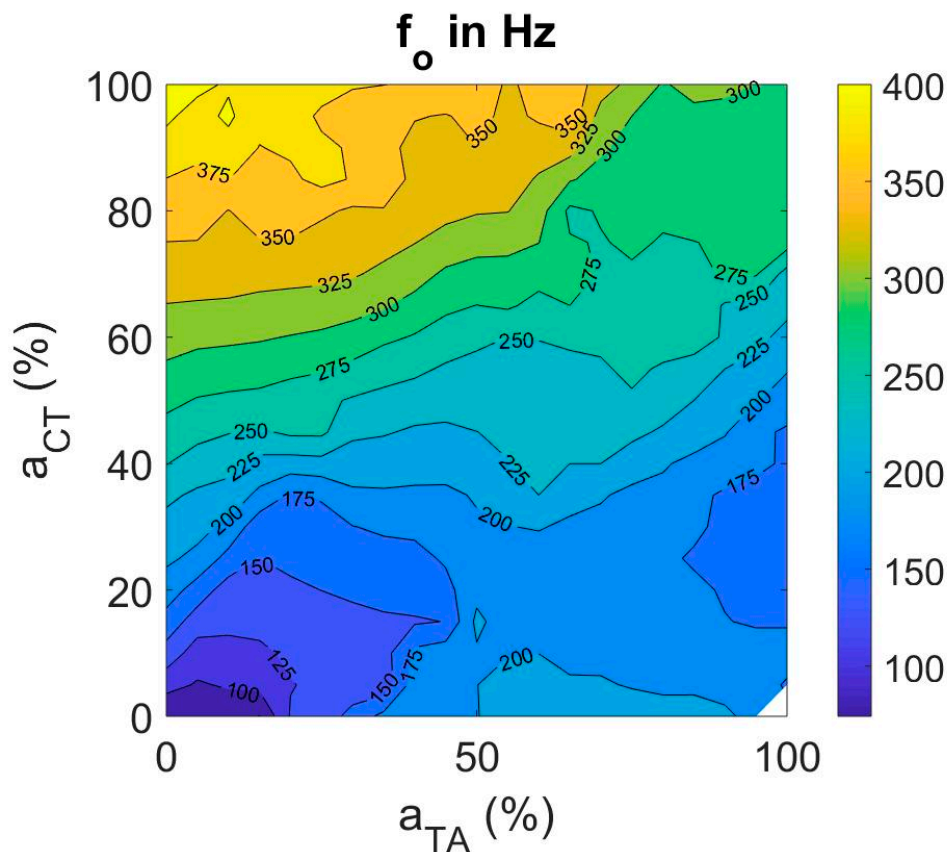
Figure 5. Muscle activation plot with fundamental frequency (f_o) as contour feature.

3.3.2. Sound Pressure Level (SPL)

SPL was found to increase steadily with CT activation (Figure 6). The increase in fundamental frequency with CT is the primary explanation. SPL increases on the order of 6–9 dB with every doubling of f_o [58]. With TA activation, SPL increased uniformly if there is a co-variation with CT activation. In the fiber-gel model described here, SPL varied between 40 and 90 dB across the CT and TA activation levels. It is interesting to note that the steepest assent in SPL is along the diagonal, increasing TA activation in proportion to CT activation.

As known from previous studies, SPL strongly depends on P_L [59]. Figure 7 shows a plot of SPL as a function of P_L as computed with the fiber-gel model. The gradual saturation in SPL with lung pressure increase results from a limitation in amplitude of vibration due to vocal fold collision. A quadratic relation between SPL and P_L is shown to be a reasonable approximation.
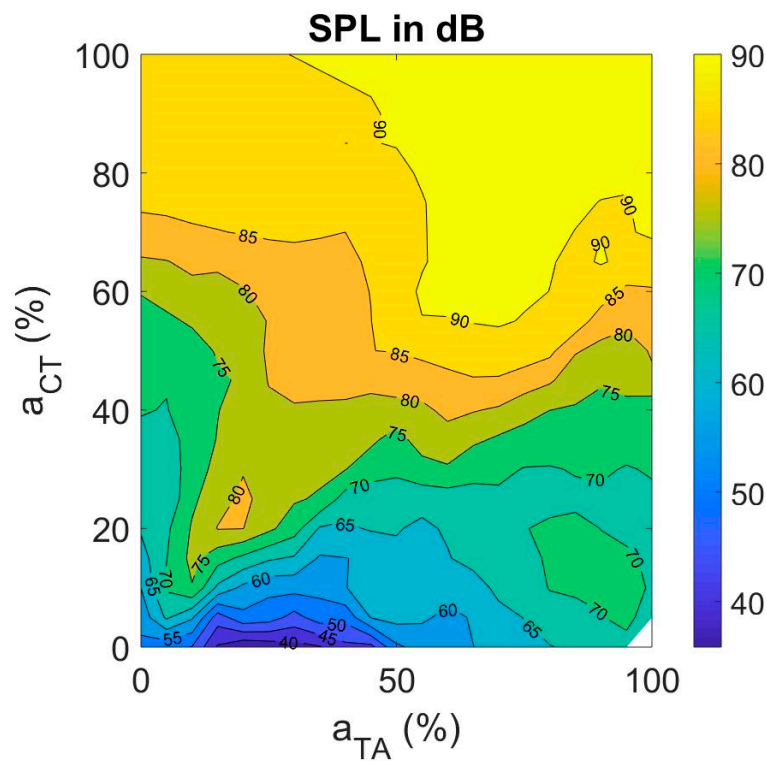
Figure 6. Muscle activation plots with sound pressure level (SPL) as the contour feature.
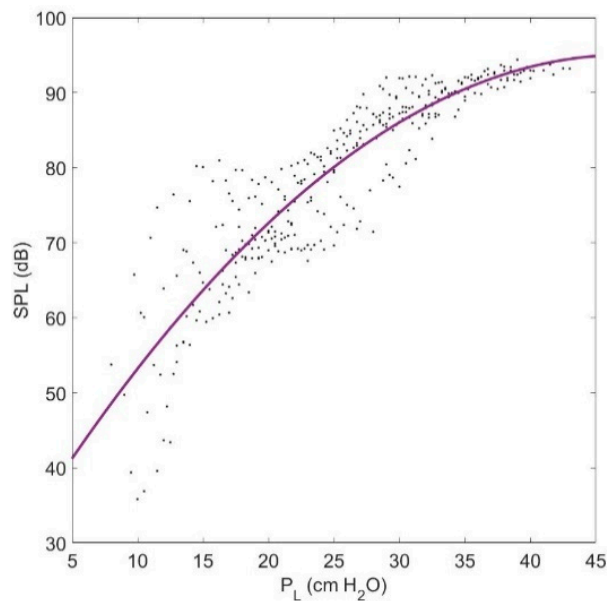

Figure 7. Scatter plot of SPL vs. P_L with a quadratic polynomial fit.

3.3.3. Brightness of a Sound

As suggested in the methods section, brightness of a sound can be represented with the normalized spectral centroid (NSC). Figure 8 shows the muscle activation plot with NSC as the contour feature. The NSC is increasing along the diagonal up to about 20% CT and TA activation levels and is decreasing thereafter. This suggests that about 20% CT and TA activation levels are optimal to obtain high NSC in the fiber-gel model. The NSC is steadily decreasing away from the diagonal with increase in CT activation level. The trend away from diagonal is similar for TA activation but not as strong as is observed with CT activation.
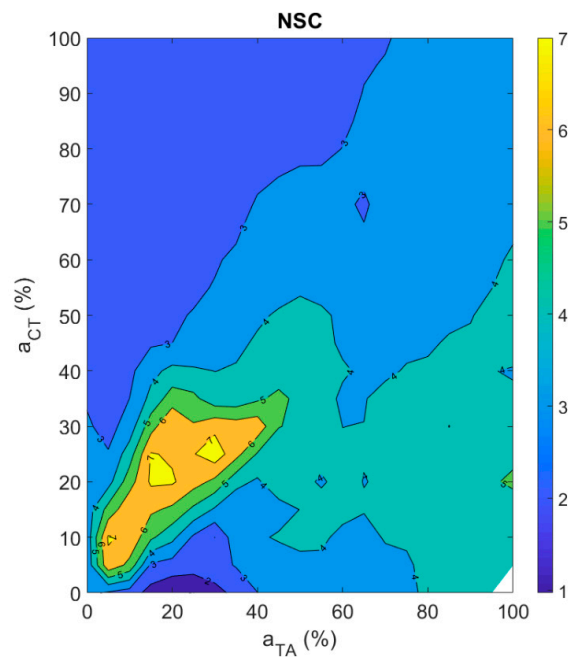
Figure 8. Muscle activation plot with normalized spectral centroid (NSC) as the contour feature.

3.3.4. Aperiodicity Roughness in the Vocal Sound

Approximate entropy (ApEn) was used to quantify aperiodicity roughness in the vocal sound. The higher the ApEn value, the greater the aperiodicity roughness in a signal [56]. It can be observed from Figure 9 that the smallest ApEn value was obtained at about 20% TA and close to 0% CT activation levels. The maximum ApEn values were obtained on the diagonal at about 40% as well as at 90% CT and TA activation. Overall, periodic signals were obtained at low CT and TA activation levels and aperiodicity in the signals increased as CT or TA activation level was increased. However, the muscle activation plot is still very complex to read due to the presence of several islands of high or low aperiodicity.
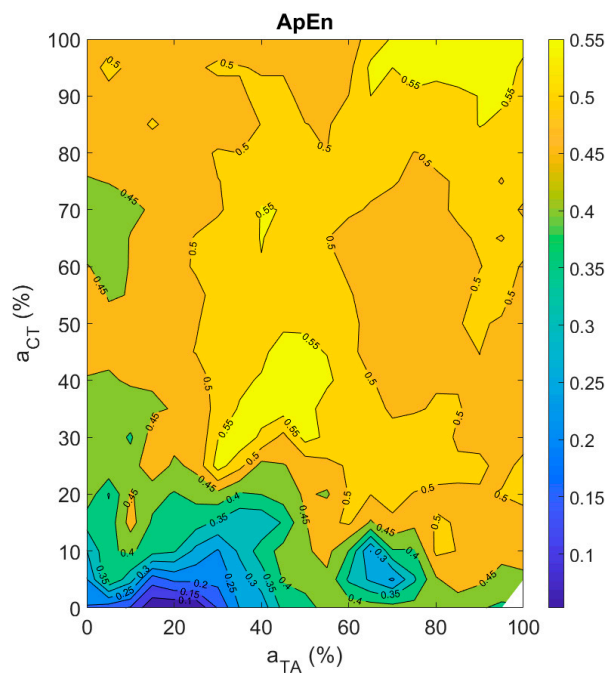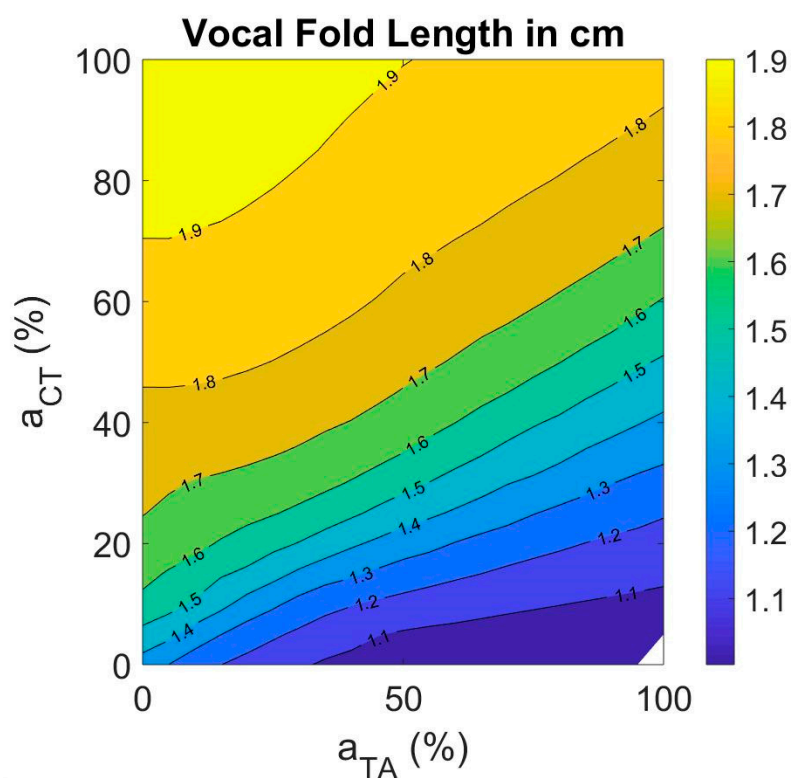

Figure 9. Muscle activation plot with approximate entropy as the contour feature.

3.4. Posturing Features

For the postural features, an average over the last 100 ms was computed to obtain the steady state values.

3.4.1. Vocal Fold Length

Range of vocal fold length is an important indicator of range of motion between the thyroid and cricoid cartilages. It is also predictive of the range of fundamental frequency [60]. The resting length of the vocal folds in the fiber-gel model was 1.6 cm. By varying TA and CT muscle activations between their limits, the vocal fold length varied between 1.0 cm and 1.9 cm (Figure 10). The length increased with increase in CT activation and decreased with increase in TA activation. It can be observed that the length did not increase uniformly with increase in CT activation. This is attributed to a non-linear increase in vocal fold fiber stress in all tissue layers as a function of CT activation.

Figure 10. Muscle activation plot with vocal fold length as the contour feature.

3.4.2. Fiber Stress in Vocal Fold Layers

Figure 11 shows the muscle activation plots for vocal fold fiber stress with respect to CT–TA muscle activations in all the three vocal fold layers (left for SLLP, center for ligament and right for TA muscle). The fiber stress in the SLLP layer is decreasing with increase in TA activity and increasing with increase in CT activity. The pattern is similar in the case of fiber stress in the ligament layer, but at higher CT activity. At low CT activity, the fiber stress in the ligament was not dependent on TA activity. The fiber stress in the TA muscle layer increases with increase in both CT and TA activity. The vocal fold ligament has the highest fiber stress, followed by TA muscle, followed by SLLP layer. The maximum stress achievable for ligament, TA muscle, and SLLP are in the ratio 88.8:16.6:1.
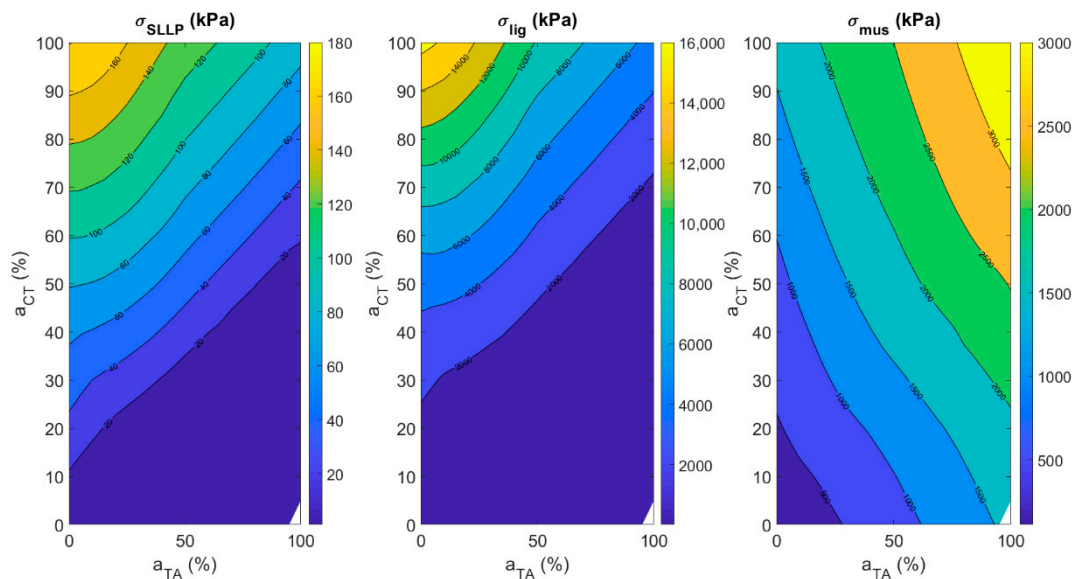
Figure 11. Muscle activation plots for fiber stress in (**left**) SLLP layer, (**center**) ligament layer, and (**right**) TA muscle with respect to TA–CT muscle activations.

4. Discussion and Conclusions

The current study focused on how TA and CT muscle activation levels control various acoustic and posturing features of voice production. Lung pressure, LCA, and IA activation levels were varied as a function of TA and CT activation levels such that the phonation threshold pressure was always exceeded. PCA muscle activation was set to 0%. The results suggested that f_o and SPL increase with CT activation, while NSC decreases when CT activation is raised above 20%. With TA activation, acoustic features have no uniform trends, except that SPL increases uniformly with TA if there is a co-variation with CT activation. The postural features L, σ_{SLLP}, σ_{lig}, and σ_{mus} increase with an increase in CT activation. Also, σ_{mus} increase whereas L, σ_{SLLP}, and σ_{lig} decrease with an increase in TA activation.

Aperiodicity roughness was found to be the most complex feature to quantify among all the acoustic and postural features that were studied in the current study. In the current study, signal-to-noise ratio, harmonic-to-noise ratio, smoothed cepstral peak prominence [15,16], correlation dimension [18], sample entropy [19], and approximate entropy were tried to appropriately quantify aperiodicity roughness in a signal. None of these acoustic features was completely accurate in measuring roughness in the signals generated by varying TA and CT muscle activations. Approximate entropy was found to be the best among them, as judged by visual inspection of signal aperiodicity and tests on sinusoidal signals with variable amounts of roughness. However, even ApEn does not give smooth contours in the muscle activation plots, suggesting that further research needs to be conducted to obtain an acoustic feature that can accurately quantify aperiodicity in a voice signal.

Comparing our results with findings from previous studies sheds more light on the relationship between input and output variables. Titze et al., (1989) [32] used electromyography on four subjects to study the role of CT and TA muscle activations in the regulation of fundamental frequency of phonation. Similar to the current findings, f_o increased with increase in CT activation at low TA values. There was also a decrease in f_o with increase in TA activation at high CT levels. No phonation was produced at extremely high TA and low CT by these subjects. They were not asked to produce anything higher than 440 Hz. We observed phonation above 440 Hz in our model, but the data were sparse with relatively large lung pressures and more irregularity.

It is well known that subglottal pressure plays a major role in the control of intensity. Our study confirmed this. Aside from P_L, CT activation had a major impact on SPL. This is probably due to the

increased radiation efficiency with higher f_o [52]. Every doubling of f_o adds at least 6 dB to the radiated sound pressure level.

No prior formal study has related a brightness feature like the normalized spectral centroid (NSC), or a roughness feature like approximate entropy (ApEn), to intrinsic laryngeal muscle activations. Hence, we could not compare our results against findings from other studies. Relations obtained for NSC and ApEn were much more complex than those obtained for f_o and SPL. Qualitatively, periodicity is related to the dominance of a single mode of vibration and entrainment of other modes to this dominant mode. When the ligament and the TA muscle fibers do not have similar natural mode frequencies, there is often not the required entrainment. This leads to aperiodic vibration. Such rough phonation was observed in many regions of the muscle activation plot.

With regard to the posturing features, vocal fold length change (strain) was found by Chhetri et al., (2012) [30] to change linearly with superior laryngeal nerve stimulation at low levels and reached a plateau at high stimulation levels. Vocal fold strain also decreased linearly with recurrent laryngeal nerve stimulation amplitude [21]. This result was also observed in our study. The vocal fold length increased linearly at low CT muscle activation levels and reached a plateau at high activation levels. The length decreased with increase in TA muscle activation at all CT activation levels. Vocal fold fiber stress was quantified in terms of muscle activations for the three layers of the vocal folds. All stresses increase exponentially with CT activation. For TA activation, fiber stresses decrease for SLLP and ligament, but increase for TA muscle fibers. This result correlates with vocal fold length change.

It should be noted that the conclusions from the current study are dependent on the accuracy of fiber-gel posturing model, which is not yet fully developed for three-dimensional vocal fold deformation with muscle activation. In particular, the medial surface bulging with TA contraction is currently being investigated in detail. Future investigations should be validated with further human and animal electromyography experiments, especially for the brightness and roughness features.

Author Contributions: Conceptualization, A.P. and I.R.T.; methodology, A.P. and I.R.T.; software, A.P. and S.S.; formal analysis, A.P.; resources, I.R.T.; data curation, A.P.; writing-original draft, A.P., S.S. and I.R.T.; writing-review and editing, A.P. and I.R.T.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. O'Callaghan, C. *Auditory Perception*, Winter 2016th ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2016.
2. Schubert, E.; Wolfe, J. Does Timbral Brightness Scale with Frequency and Spectral Centroid? *Acta Acust. United Acust.* **2006**, *92*, 820–825.
3. Daniel, P.; Weber, R. Psychoacoustical roughness: Implementation of an optimized model. *Acta Acust. United Acust.* **1997**, *83*, 113–123.
4. Bodden, M. Instrumentation for sound quality evaluation. *Acta Acust. United Acust.* **1997**, *83*, 775–783.
5. De Cheveigne, A.; Kawahara, H. YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* **2002**, *111*, 1917–1930. [CrossRef] [PubMed]
6. Camacho, A.; Harris, J.G. A sawtooth waveform inspired pitch estimator for speech and music. *J. Acoust. Soc. Am.* **2008**, *124*, 1638–1652. [CrossRef]
7. Titze, I.R. *Principles of Voice Production*; Prentice-Hall, Engle-Wood Cliffs: Bergen, NJ, USA, 1994; pp. 220–222.
8. Shmilovitz, D. On the definition of total harmonic distortion and its effect on measurement interpretation. *IEEE Trans. Power Deliv.* **2005**, *20*, 526–528.
9. Carral, S.; Vergez, C.; Nederveen, C.J. Toward a single reed mouthpiece for the oboe. *Arch. Acoust.* **2011**, *36*, 267–282. [CrossRef]
10. Zwicker, E.; Fastl, H. *Psychoacoustics: Facts and Models*; Springer: Berlin, Germany, 1990.

11. Eddinsa, D.A.; Kopf, L.M.; Shrivastav, R. The psychophysics of roughness applied to dysphonic voice. *J. Acoust. Soc. Am.* **2015**, *138*, 3820–3825. [CrossRef]

12. Bergan, C.C.; Titze, I.R. Perception of pitch and roughness in vocal signals with subharmonics. *J. Voice* **2001**, *15*, 165–175. [CrossRef]

13. Horii, Y. Jitter and Shimmer differences among sustained vowel phonations. *J. Speech Lang. Hear. Res.* **1982**, *25*, 12–14. [CrossRef]

14. Fraile, R.; Godino-Llorente, J.I. Cepstral peak prominence: A comprehensive analysis. *Biomed. Signal Process. Control* **2014**, *14*, 42–54. [CrossRef]

15. Heman-Ackah, Y.D.; Heuer, R.J.; Michael, D.D.; Ostrowski, R.; Horman, M.; Baroody, M.M.; Hillenbrand, J.; Sataloff, R.T. Cepstral peak prominence: A more reliable measure of dysphonia. *Ann. Otol. Rhinol. Laryngol.* **2003**, *112*, 324–333. [CrossRef] [PubMed]

16. Latoszek, B.B.; Maryn, Y.; Gerrits, E.; De Bodt, M. A meta-analysis: Acoustic measurement of roughness and breathiness. *J. Speech Lang. Hear. Res.* **2018**, *61*, 298–323. [CrossRef] [PubMed]

17. Liu, B.; Polce, E.; Sprott, J.C.; Jiang, J.J. Applied chaos level test for validation of signal conditions underlying optimal performance of voice classification methods. *J. Speech Lang. Hear. Res.* **2018**, *61*, 1130–1139. [CrossRef]

18. MacCallum, J.K.; Cai, L.; Zhang, Y.; Jiang, J.J. Acoustic analysis of aperiodic voice: Perturbation and nonlinear dynamic properties in esophageal phonation. *J. Voice* **2009**, *23*, 283–290. [CrossRef] [PubMed]

19. Fabris, C.; De Colle, W.; Sparacino, G. Voice disorders assessed by (cross-) sample entropy of electroglottogram and microphone signals. *Biomed. Signal Process. Control* **2013**, *8*, 920–926. [CrossRef]

20. Titze, I.R.; Palaparthi, A. Sensitivity of Source-Filter Interaction to specific vocal tract shapes. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 2507–2515. [CrossRef]

21. Chhetri, D.K.; Neubauer, J.; Berry, D.A. Graded activation of the intrinsic laryngeal muscles for vocal fold posturing. *J. Acoust. Soc. Am.* **2010**, *127*, EL127–EL133. [CrossRef]

22. Faaborg-Andersen, K. Electromyographic investigation of intrinsic laryngeal muscles in humans. *Acta Physiol. Scand.* **1957**, *41*, 1–149.

23. Faaborg-Andersen, K. Electromyography of laryngeal muscles in humans. Technics and results. *Aktuel Probl. Phoniatr. Logop.* **1965**, *12*, 1–72.

24. Hirano, M.; Ohala, J.; Vennard, W. The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. *J. Speech Hear. Res.* **1969**, *12*, 616–628. [CrossRef] [PubMed]

25. Gay, T.; Hirose, H.; Strome, M.; Sawashima, M. Electromyography of the intrinsic laryngeal muscles during phonation. *Ann. Otol. Rhinol. Laryngol.* **1972**, *81*, 401–409. [CrossRef] [PubMed]

26. Kochis-Jennings, K.A.; Finnegan, E.M.; Hoffman, H.T.; Jaiswal, S. Laryngeal muscle activity and vocal fold adduction during chest, chestmix, headmix, and head registers in females. *J. Voice* **2012**, *26*, 182–193. [CrossRef] [PubMed]

27. Nasri, S.; Beizai, P.; Sercarz, J.A.; Kreiman, J.; Graves, M.C.; Berke, G.S. Function of the Interarytenoid muscle in a canine laryngeal model. *Ann. Otol. Rhinol. Laryngol.* **1994**, *103*, 975–982. [CrossRef]

28. Choi, H.S.; Ye, M.; Berke, G.S. Function of the interarytenoid (IA) muscle in phonation: In vivo laryngeal model. *Yonsei Med. J.* **1995**, *36*, 58–67. [CrossRef]

29. Choi, H.S.; Berke, G.S.; Ye, M.; Kreiman, J. Function of the posterior cricoarytenoid muscle in phonation: In vivo laryngeal model. *Otolaryngol. Head Neck Surg.* **1993**, *109*, 1043–1051. [CrossRef]

30. Chhetri, D.K.; Neubauer, J.; Berry, D.A. Neuromuscular control of fundamental frequency and glottal posture at phonation onset. *J. Acoust. Soc. Am.* **2012**, *131*, 1401–1412. [CrossRef]

31. Chhetri, D.K.; Park, S.J. Interactions of subglottal pressure and neuromuscular activation on fundamental frequency and intensity. *Laryngoscope* **2016**, *126*, 1123–1130. [CrossRef]

32. Titze, I.R.; Luschei, E.S.; Hirano, M. Role of the thyroarytenoid muscle in regulation of fundamental frequency. *J. Voice* **1989**, *3*, 213–224. [CrossRef]

33. Farley, G.R. A biomechanical laryngeal model of voice F_0 and glottal width control. *J. Acoust. Soc. Am.* **1996**, *100*, 3794–3812. [CrossRef]

34. Titze, I.R.; Story, B.H. Rules for controlling low-dimensional vocal fold models with muscle activation. *J. Acoust. Soc. Am.* **2002**, *112*, 1064–1076. [CrossRef] [PubMed]

35. Finnegan, E.M.; Luschei, E.S.; Hoffman, H.T. Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech. *J. Speech Lang. Hear. Res.* **2000**, *43*, 934–950. [CrossRef] [PubMed]

36. Baker, K.K.; Ramig, L.O.; Sapir, S.; Luschei, E.S.; Smith, M.E. Control of vocal loudness in young and old adults. *J. Speech Lang. Hear. Res.* **2001**, *44*, 297–305. [CrossRef]

37. Perlman, A.L. Electromyography and the study of Oropharyngeal Swallowing. *Dysphagia* **1993**, *8*, 351–355. [CrossRef] [PubMed]

38. Hillel, A.D. The study of laryngeal muscle activity in normal human subjects and in patients with laryngeal dystonia using multiple fine-wire electromyography. *Laryngoscope* **2001**, *111*, 1–47. [CrossRef] [PubMed]

39. Poletto, C.J.; Verdun, L.P.; Strominger, R.; Ludlow, C.L. Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures. *J. Appl. Physiol.* **2004**, *97*, 858–866. [CrossRef]

40. Lowell, S.Y.; Story, B.H. Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration. *J. Acoust. Soc. Am.* **2006**, *120*, 386–397. [CrossRef]

41. Vahabzadeh-Hagh, A.M.; Zhang, Z.; Chhetri, D.K. Quantitative evaluation of the in vivo vocal fold medial surface shape. *J. Voice* **2017**, *31*, 513.e15–513.e23. [CrossRef]

42. Titze, I.R.; Alipour, F.; Blake, D.; Palaparthi, A. Comparison of a fiber-gel finite element model of vocal fold vibration to a transversely isotropic stiffness model. *J. Acoust. Soc. Am.* **2017**, *142*, 1376–1383. [CrossRef]

43. Titze, I.R. *The Myoelastic Aerodynamic Theory of Phonation*; National Center for Voice and Speech: Denver, CO, USA, 2006; pp. 1–337.

44. Story, B.H.; Titze, I.R.; Hoffman, E.A. Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Am.* **1996**, *100*, 537–554. [CrossRef]

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

46. Sato, K.; Hirano, M. Histological investigation of the macula flava of the human vocal fold. *Ann. Otol. Rhinol. Laryngol.* **1995**, *104*, 138–143. [CrossRef] [PubMed]

47. Gray, S.D.; Alipour, F.; Titze, I.R. Biomechanical and histological observations of vocal fold fibrous proteins. *Ann. Otol. Rhinol. Laryngol.* **2000**, *109*, 77–85. [CrossRef] [PubMed]

48. Alipour, F.; Berry, D.A.; Titze, I.R. A finite element model of vocal fold vibration. *J. Acoust. Soc. Am.* **2000**, *108*, 3003–3012. [CrossRef] [PubMed]

49. Liljencrants, J. Speech Synthesis with a Reflection-Type Line Analog. Ph.D. Thesis, Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, Sweden, 1985.

50. Story, B.H. Physiologically Based Speech Simulation Using an Enhanced Wave Reflection Model of the Vocal Tract. Ph.D. Thesis, University of Iowa, Iowa City, IA, USA, 1995.

51. Titze, I.R.; Palaparthi, A.; Smith, S. Benchmarks for time-domain simulation of sound propagation in soft-walled airways: Steady configurations. *J. Acoust. Soc. Am.* **2014**, *136*, 3249–3261. [CrossRef] [PubMed]

52. Titze, I.R.; Palaparthi, A. Radiation efficiency for long-range vocal communication in mammals and birds. *J. Acoust. Soc. Am.* **2018**, *143*, 2813–2824. [CrossRef]

53. Palaparthi, A.; Riede, T.; Titze, I.R. Combining multiobjective optimization and cluster analysis to study vocal fold functional morphology. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2199–2208. [CrossRef]

54. Palaparthi, A.; Smith, S.; Mau, T.; Titze, I.R. A computational study of depth of vibration into vocal fold tissues. *J. Acoust. Soc. Am.* **2019**, *145*, 881–891. [CrossRef]

55. Flanagan, J.L. *Speech Analysis, Synthesis, and Perception*; Springer: New York, NY, USA, 1972.

56. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [CrossRef]

57. Titze, I.R. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.* **1988**, *83*, 1536–1552. [CrossRef]

58. Titze, I.R.; Sundberg, J. Vocal intensity in speakers and singers. *J. Acoust. Soc. Am.* **1992**, *91*, 2936–2946. [CrossRef] [PubMed]

59. Bjorklund, S.; Sundberg, J. Relationship between subglottal pressure and sound pressure level in untrained singers. *J. Voice* **2016**, *30*, 15–20. [CrossRef] [PubMed]

60. Titze, I.R.; Riede, T.; Mau, T. Predicting fundamental frequency ranges in vocalization across species. *PLoS Comput. Biol.* **2016**, *12*, e1004907. [CrossRef] [PubMed]