Using a vertical three-mass computational model of the vocal folds to match human phonation of three adult males

Brittany L. Perrine[1,a)] and Ronald C. Scherer[2]

[1]*Department of Communication Sciences and Disorders, Baylor University, One Bear Place #97332, Waco, Texas 76798, USA*

[2]*Department of Communication Sciences and Disorders, Bowling Green State University, Ridge Street, Bowling Green, Ohio 43403, USA*

ABSTRACT:

Computer models of phonation are used to study various parameters that are difficult to control, measure, and observe in human subjects. Imitating human phonation by varying the prephonatory conditions of computer models offers insight into the variations that occur across human phonatory production. In the present study, a vertical three-mass computer model of phonation [Perrine, Scherer, Fulcher, and Zhai (2020). J. Acoust. Soc. Am. **147**, 1727–1737], driven by empirical pressures from a physical model of the vocal folds (model M5), with a vocal tract following the design of Ishizaka and Flanagan [(1972). Bell Sys. Tech. J. **51**, 1233–1268] was used to match prolonged vowels produced by three male subjects using various pitch and loudness levels. The prephonatory conditions of tissue mass and tension, subglottal pressure, glottal diameter and angle, posterior glottal gap, false vocal fold gap, and vocal tract cross-sectional areas were varied in the model to match the model output with the fundamental frequency, alternating current airflow, direct current airflow, skewing quotient, open quotient, maximum flow negative derivative, and the first three formant frequencies from the human production. Parameters were matched between the model and human subjects with an average overall percent mismatch of 4.40% (standard deviation = 6.75%), suggesting a reasonable ability of the simple low dimensional model to mimic these variables.
© 2023 Acoustical Society of America. https://doi.org/10.1121/10.0020847

I. INTRODUCTION

A primary goal that laryngeal modeling attempts to meet is the accurate simulation of human phonation. Such simulations would suggest that the modeling itself reflects accurate (1) properties of the vocal folds and their dynamic movement, (2) aerodynamics of air pressures and airflows, and (3) output phonatory acoustics. If neuromuscular inputs are also included, then accurate simulation of human phonation also suggests that such neuromuscular inputs are reasonable approximations.

Computer laryngeal models have been compared to human data in several ways as computer laryngeal modeling is related to such factors as age (Döllinger *et al.*, 2017) and sex (Döllinger *et al.*, 2017; Lucero and Koenig, 2005a), and various voice disorders (Cataldo and Soize, 2018; Gunter, 2004; Kanduri *et al.*, 2021; Stepp *et al.*, 2010) have provided information about the effects of mass and stiffness on the movement of the vocal folds. Asymmetry of movement of the vocal folds has been studied (e.g., Ishizaka and Isshiki, 1976), as well as the mucosal wave (most models, e.g., Titze, 1988; Story and Titze, 1995), phonation threshold pressure (e.g., Titze, 1988; Lucero, 1999; Perrine *et al.*, 2020), and vocal registers (e.g., Tokuda *et al.*, 2007). Other areas of study include muscle tensions (e.g., Story and Titze, 1995), intraglottal pressures (Pelorson *et al.*, 1994),

the transglottal pressure coefficient (Guo and Scherer, 1993), supraglottal velocities and airflows (e.g., Alipour *et al.*, 2011), and numerous other motion, fluid-structure, aerodynamic, and aeroacoustic phenomena (for a review of many studies and further applications, see Hanson *et al.*, 2001, and Alipour *et al.*, 2011).

Computer modeling, simple to complex, therefore, has attempted to imitate many of the important characteristics of phonation and has done so successfully. However, the computational power required for many of these models limits their practical application (e.g., Sadeghi *et al.*, 2019). Simple models with fast run times also allow for descriptions of the control of parameters like open quotient, glottal airflow, fundamental frequency, and output acoustics and could be used to determine outcomes for humans in clinical settings. To accomplish both, there is a need to determine how closely a computer model can match phonatory behaviors of specific individuals. The approach taken here is to explore a low dimensional computer model of phonation and the vocal tract to see how well it can be used to match human aerodynamic and acoustic signals. More specifically, the study explores a comparison of measures of the glottal airflow waveforms, acoustic output waveforms, and formant frequencies between three male humans and a low dimensional computer model.

The phonatory model used in this study is based on the model published by Ishizaka and Flanagan (1972), henceforth I&F72, with significant modifications. In their model,

a)Email: brittany_perrine@baylor.edu

the motivation was to offer a self-oscillating synthesis model that would "make the synthetic voice as natural sounding as possible" (p. 1234) and applicable to "successful medical diagnosis" (p. 1234). Those have been the goals for nearly all phonatory models since that time. In the study here, it is thought that successful matching of human phonatory characteristics using a low dimensional computer model would extend the aforementioned motivations.

II. METHODS

A. The vertical three-mass (V3M) computational model

The V3M model used in the present study is described in Perrine *et al.* (2020) (cf. Tokuda *et al.*, 2007). Briefly, the model was based on the two-mass model by I&F72 of the vocal folds with the addition of a third vocal fold mass below the glottis that was the same size as the mass just superior to it (see Fig. 1 taken from Perrine *et al.*, 2020). The uppermost mass was the smallest with a mass and vertical length ratio of 1–5 relative to each of the two lower masses. The vocal tract algorithms were the same as those used in I&F72 (see below).

As in the I&F72 model, the two vocal folds and glottis were modeled symmetrically with symmetric motion. The vocal tract sections were cylindrical. Masses $m1$ and $m2$ and masses $m2$ and $m3$ were coupled with linear springs. All three masses were coupled to the sidewall by nonlinear springs (as in I&F72). Damping between the masses and sidewall represented equivalent viscous resistances (as in I&F72). Motion was only lateral in I&F72, but in our model, the lowest mass had an additional spring and damper attached to a lower "surface," allowing vertical movement because in model $M5$, surface air pressures used in the simulations were obtained along the inferior vocal fold surfaces (pressure taps 1–5 in Fig. 1) as well as on the medial surface of the vocal folds. It is acknowledged that this model does not represent the layered structure of the vocal folds but represents more of an indented motion of the mucosal wave (represented by stress-strain studies of lateral displacements, e.g., Oren *et al.*, 2014). The vocal fold collision mechanism of I&F72 was adopted, which also is a nonlinear restoring force. Viscous losses increased stepwise at vocal fold contact.

In the I&F72 model, the pressure distribution along the glottis was modeled with the Bernoulli one-dimensional equation with losses obtained by the limited glottal configuration study of van den Berg *et al.* (1957). For our model, the empirical values of the surface pressures on the $M5$ vocal fold surfaces corresponded to specific values of transglottal pressure and glottal configuration (angle and diameter; Scherer *et al.*, 2001, 2002; Scherer *et al.*, 2004; Scherer 2010). Model $M5$ is a two-dimensional physical model of the larynx made of Plexiglas® and is 7.5 times the size of the human larynx. The vocal folds of model $M5$ contain multiple pressure taps (16 taps at the surface of the model and vocal folds on 1 side from the trachea to downstream of the glottis). To obtain the intraglottal air pressures and
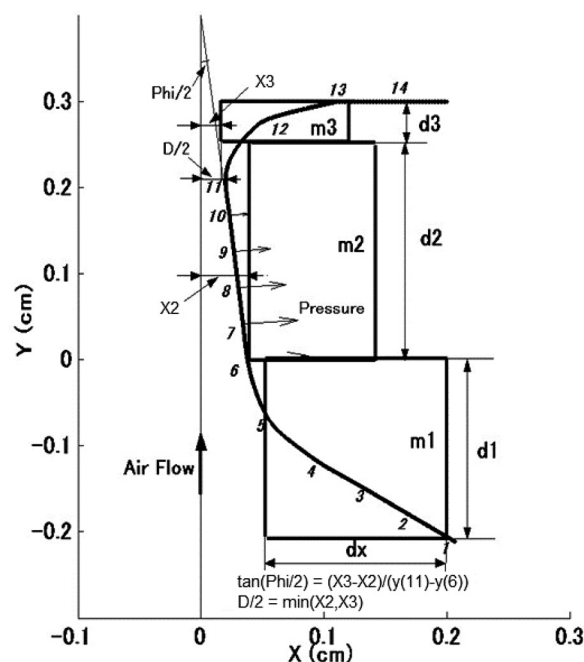

FIG. 1. Graphical representation of the V3M computer model used in the present study. The M5 pressure taps, which provided the empirical driving pressures, are labeled as numbers 1–14 along the medial edge of the vocal folds. Mass $m1$ is inferior and mass $m3$ is superior. Not shown are the springs and dashpots, which are described in the text. Reprinted with permission from Perrine *et al.* (2020). J. Acoust. Soc. Am. **147**, 1727–1737. Copyright 2020 Acoustical Society of America.

corresponding glottal airflows, 63 configurations (7 glottal diameters and 9 glottal angles) and a range of transglottal pressures (1, 3, 5, 10, 15, and 25 cm H_2O) were used to obtain pressures at each tap and were configured in a "lookup" table such that any configuration and transglottal pressure within the empirical ranges could be used to interpolate the surface pressures and airflow. At each time step during a run of the V3M model, the geometry was given, and new surface air pressures and airflows and a new configuration were obtained. The use of such models is supported by the quasi-steady approximation except for the case of a very narrow glottis during the phonatory cycle or high fundamental frequencies (Ishizaka and Flanagan, 1972; Flanagan, 1958; Mongeau *et al.*, 1992, 1997; Pelorson *et al.*, 1994; Zhang *et al.*, 2002; Vilain *et al.*, 2004; Honda *et al.*, 2022; Wang *et al.*, 2021a).

The vocal tract was modeled as an equivalent circuit (transmission line of six cylindrical sections) as in I&F72 [their Fig. 5 and Eqs. (12), (17), and (18); they used only four sections in their simulations] with the subglottal pressures taken as given values (the model used here lacked subglottal resonances) and inductances and capacitances calculated for each of the six sections. Vocal tract viscous losses were taken as serial resistances; our study used the same expressions as in I&F72. A radiation load at the lips terminated the transmission line. I&F72 provides the differential equations pertinent to obtaining the airflows within the system, as well as the equations of motion for the vocal folds.

Brittany L. Perrine and Ronald C. Scherer

The glottal (laryngeal) airflow in the V3M model is given by

$$U_g = U_{gg} + U_{dpx} + U_{dpy} + U_{cg},$$

where U_{gg} is the glottal airflow associated with the $M5$ empirical data for instantaneous geometry and transglottal pressure conditions, U_{dpx} is the horizontal displacement airflow, U_{dpy} is the vertical displacement airflow, and U_{cg} is the airflow through the posterior glottis {using a simple Bernoulli formula approximation of $U_{cg} = A_{cg}[(2P_t)/(\rho k)]0.5$, where A_{cg} is the posterior glottal area, P_t is the transglottal pressure, ρ is air density, and $k = 1.7$; Perrine et al., 2020; Scherer et al., 2013}. As a first approximation, the posterior glottis was given a triangular shape (called the posterior glottal gap) with the base of the triangle at the posterior wall. The triangle height (i.e., the anterior-posterior dimension) was set to 1 cm (Scherer et al., 2013). The area was the area of the triangle-like posterior gap with the base at the posterior wall. The laryngeal airflow resistance for the false vocal fold gap was calculated based on Eq. (1) in Agarwal (2004) and Agarwal et al. (2004).

The V3M model had a vocal tract made up of six sections for this study. Each cylindrical section had a set *a priori* vertical length of 2.475 cm for a total vocal tract length of 14.85 cm. This length is relatively short for adult male vocal tracts (typically given as 17 cm but is vowel dependent; Story et al., 1996). Fitch and Giedd (1999) report an average length for 19–25 year old males of 15.54 cm with a range of 14.6–16.0 cm, and Groll et al. (2020) estimated a range of 14.15–17.58 cm. The first section of the vocal tract corresponded to the epilaryngeal or larynx tube region. It is clear that the vocal tract model used here (a la I&F72) did not include the configurations and inertance of more recent models and, thus, it may have created a greater challenge to match the human data (due basically to the probable less inertive effects to skew the airflow). The cross-sectional area of each section of the vocal tract was varied in this study from a default value of 5 cm^2 to match the human formant values. Therefore, the vocal tract is a gross simplification of the complex cross-sectional area function of the human vocal tract during vowel production.

It is noted that the airflow resistance effect of the false vocal folds was included (Agarwal, 2004; Agarwal et al., 2004). For false vocal fold gaps greater than approximately 0.5 cm, the airflow resistance effect was negligible according to the results by Agarwal. The average false vocal fold gap determined by the operator in this study was 0.459 cm, and the correlation was essentially zero between the false vocal fold gap size and the variables f_o, alternating current (AC) airflow, direct current (DC) airflow, open quotient, skewing quotient, and maximum flow negative derivative (MFND). Thus, in this study, the false vocal fold gap appears to have had a limited if not negligible effect.

The differential equations were approximated by difference equations as in I&F72. The update interval for recalculations was 0.1 ms. It is noted that the transglottal pressure was the difference between the subglottal pressure and supraglottal pressure at each update in time and, hence, includes the effects of inertance of the supraglottal vocal tract. The fundamental frequency is primarily dependent on the natural frequency of the masses and their coupling, where the frequency for one mass alone could be stated approximately as $f_o = (1/2 \ \pi)(k/m)^{0.5}$, where k is the spring constant and m is the mass.

B. Recordings from human participants

Three adult males ($S1$–$S3$) without laryngeal or respiratory problems at the time of the recording produced /pi:/ syllable repetitions smoothly, evenly, and all in one breath. Four different production conditions were elicited: (1) normal pitch, normal loudness; (2) higher pitch, normal loudness; (3) normal pitch, louder; and (4) normal pitch, softer. The recordings were made by Dr. Robert Hillman and Dr. Daryush Mehta at the Center for Laryngeal Surgery and Voice Rehabilitation at the Massachusetts General Hospital and sent to the Bowling Green State University (BGSU) laboratory for analysis. All analog signals were simultaneously recorded and digitized at a sampling rate of 120 000 samples per second into a digital acquisition board with 16-bit quantization and a ±10 V dynamic range (6259M series; National Instruments, Austin, TX). Prior to digitization, the signals were conditioned with an anti-aliasing low pass filter with a 3-dB cutoff frequency of 30 000 Hz (CyberAmp model 380; Danaher Corp., Washington, DC). The audio signal was recorded using a head mounted condenser microphone (model MKE104; Sennheiser electronic GmbH and Co. KG., Wedemark, Germany) that was placed 4 cm from the mouth at a 45 deg angle from the lips and connected to a pre-conditioner (model 302 dual microphone preamplifier; Symetrix Inc., Mountlake Terrace, WA). Oral air pressure and oral airflow were collected using a modified version of the Glottal Enterprises aerodynamic system (mask, model MA-1L; transducers, model PT-series; electronics unit, model MS-100A2; Glottal Enterprises, Syracuse, NY). The modified version of the mask is described in Zañartu et al. (2011). Electroglottography, accelerometer, and high-speed imaging recordings were also simultaneously made but were not included in the study reported here.

A custom program for MATLAB software (Natick, MA), BGSigplot, was used for data analysis. The middle three syllables of a string of 8–12 syllables were used for subglottal pressure analysis. For the estimate of subglottal pressure from oral air pressure, the calibrated pressure signal was smoothed using a moving average of ten samples to the left and right of the target sample. The subglottal pressure was estimated from the average of the pressure value of the right corner of the /p/ occlusion before the vowel and the left corner of the /p/ occlusion after the vowel (Rothenberg, 1973; Löfqvist et al., 1982; Perrine et al., 2019). The airflow signal was used to visually determine the right and left corners of the /p/ occlusion. The right corner was the last moment in the pressure signal before the increase in airflow

Brittany L. Perrine and Ronald C. Scherer   1507

in the oral airflow signal. The left corner was the first moment in the pressure signal where there was no airflow in the oral airflow signal. A relatively flat pressure plateau and no airflow during the lip occlusion were requisites for inclusion in the analysis.

The middle vowel of the string of syllables that was analyzed for subglottal pressure was used for the airflow analysis. The smoothed wideband airflow signal was inverse filtered using TF32 (Paul Milenkovic, Madison, WI). Ten consecutive glottal airflow cycles were chosen from a stable (consistent airflow amplitude) part of the middle of the vowel /i/. The closed phase of the glottal airflow was defined as a period of consistent and minimum airflow between consecutive glottal cycles. Peak airflow and airflow onset of each glottal cycle (first instance of glottal airflow rise prior to the glottal airflow peak after a period of minimum airflow) and airflow offset for the glottal cycle (last moment of airflow decrease before a period of minimum airflow) were manually selected and used to measure AC airflow and DC airflow and calculate the open quotient, skewing quotient, and MFND. Open quotient was calculated as the time of the airflow offset minus the time of the airflow onset divided by the total time for one cycle. Skewing quotient was calculated as the glottal airflow rise time divided by the glottal airflow fall time (i.e., the time of peak airflow minus the time of airflow baseline onset, divided by the time of airflow offset minus the time of the peak airflow).

Fundamental frequency was estimated from the microphone signal using Praat (Boersma, 2001) by dividing the frequency of the tenth harmonic by ten. Fundamental frequency calculations of the vowel /i/ were made for the same three syllables from which air pressure and airflow measurements were made. The formant frequency values were determined from visual examination of the spectrum in Praat of the /i/ vowel for the human recordings and model estimations. Extractions of aerodynamic and acoustic data could not be made for two recordings (normal pitch, louder; and normal pitch, softer) for one participant ($S2$) due to unknown reasons (that is, when considering the data for publication, we could not determine with certainty the reasons those data were unusable).

C. Using the model to match signals

Six features of the acoustic and aerodynamic signals were targeted for matching as well as the first three formants. One vowel in the center of a syllable string produced by the human participant was matched for each condition. The subglottal pressure was set to the average subglottal pressure used by the participant across the syllable string. Thus, the modeled subglottal pressure was always specified as the same average value as empirically determined for the participant and not used to calculate percent match as was performed for the other variables. The fundamental frequency of the selected vowel was matched from the microphone signal. From the inverse filtered airflow, the AC airflow, DC airflow, open quotient, skewing quotient, and

MFND were also targeted. The airflow characteristics (e.g., onset and offset) were extracted and calculated as described above. The values of vocal fold features not varied in the model are presented in Table I. The vocal fold length of 1.2 cm was the same as in model $M5$ and is in the range for male vocal fold length during phonation (Hollien, 2014; Nishizawa, 1988). The posterior glottal airflow constant ($k = 1.7$) was determined by an earlier study (Scherer *et al.*, 2013). The other values in Table I are from I&F72. When matching the synthesized and human glottal airflow waveforms, emphasis was given to matching the descent of the airflow on the right side of the waveform due to the acoustic importance of that portion of the glottal airflow.

To determine the vocal fold and vocal tract settings that best matched the human data, iterative runs of the model were completed by the operator using the same subglottal pressure but different combinations of each variable in Table II. Based on the modeling and synthesis studies referenced in this project, the ranges for the variables in Table II were established as bounds when running the V3M model. The cross-sectional area of each of the six vocal tract segments was varied to match the vowel and acoustic pressure signal to the human microphone signal. Based on Mermelstein (1967), the downstream region was set to be narrower and the upstream region was generally set to be wider as all participants attempted to produce the vowel /i/ while the rigid scope was in the mouth. The cross-sectional area of the vocal tract segments was varied until the overall shape of the output spectra of the model visually matched the spectra from the corresponding vowel produced by the human. Formant frequency measures of the human and model spectra were made in Praat.

Matching the fundamental frequency and features of the airflow signal (AC airflow, open quotient, DC airflow, skewing quotient, and MFND) involved varying tissue properties and prephonatory glottal configurations. The tension of the springs between the vocal fold masses, the springs between the lateral "wall" and masses, and the spring between the "floor" and lowest mass were varied equally by a certain percent increase or decrease to each spring setting. Masses

TABLE I. The V3M model variables and values used in the present study for the variables not modified to create matches. Viscous resistances of the masses are specified as in Ishizaka and Flanagan (1972) as $r = 2\zeta(mk)^{0.5}$, where the values of ζ are given herein. The vocal fold length of 1.2 cm matches that for model $M5$. The posterior glottal airflow correction factor $k = 1.7$ was based on hydraulic equations for flow through triangular tubing (Scherer *et al.*, 2013).

Parameter	Value used in present study
Vocal fold length	1.2 cm
Cubic term for springs	Included
Displacement airflow	Included
Damping ratio for viscous loss, $\zeta 0$	0.05
Damping ratio for viscous loss, $\zeta 1$	0.1
Damping ratio for viscous loss, $\zeta 2$	0.1
Damping ratio for viscous loss, $\zeta 3$	0.6
Added damping ratio during closure	1.0
Posterior glottal airflow (U_{cg}), k	1.7

TABLE II. The V3M model variables, ranges, and default values for the variables modified to create matches. Each vocal tract segment was 2.475 cm in vertical length. The ranges are based on studies referenced in this paper.

Parameter	Ranges	Default
Subglottal pressure	0-50 cm H_2O	8 cm H_2O
Vocal tract segment 1 (upstream near glottis)	0–15 cm^2	5 cm^2
Vocal tract segment 2	0–15 cm^2	5 cm^2
Vocal tract segment 3	0–15 cm^2	5 cm^2
Vocal tract segment 4	0–15 cm^2	5 cm^2
Vocal tract segment 5	0–15 cm^2	5 cm^2
Vocal tract segment 6 (downstream at lips)	0 – 15 cm^2	5 cm^2
Spring between the "floor" and $m1$ ($K0$)	1000–400 000 dyn/cm	160 000 dyn/cm
Spring between "lateral wall" and $m1$ ($K1$)	1000–400 000 dyn/cm	80 000 dyn/cm
Spring between lateral wall and $m2$ ($K2$)	1000–100 000 dyn/cm	40 000 dyn/cm
Spring between lateral wall and $m3$ ($K3$)	0–100 000 dyn/cm	4000 dyn/cm
Spring between $m1$ and $m2$ (Kcp)	0–100 000 dyn/cm	12 500 dyn/cm
Spring between $m2$ and $m3$ (Kc)	1000–100,000 dyn/cm	12 500 dyn/cm
Prephonatory glottal angle	20° divergent to 20° convergent	0°
Mass of $m1$	0.01–0.2 g	0.0625 g
Mass of $m2$	0.01–0.2 g	0.0625 g
Mass of $m3$	0.001–0.2 g	0.0125 g
Prephonatory glottal diameter	0–0.08 cm	0.04 cm
False vocal fold gap	0.1–2.0 cm	0.6 cm
Prephonatory abduction of cartilaginous glottis	0–0.4 cm	0.04 cm

$m1$ and $m2$ were varied such that they always had the same mass value, and the mass of $m3$, the smallest uppermost mass, was varied proportionately. Prephonatory mass and spring settings were the same for both vocal folds in the present study (to provide symmetric vocal fold motion). Other prephonatory settings were varied independently (although these settings are the same for the left and right vocal folds), including the vertical angle created by the medial surfaces of $m2$ and $m3$ (that is, half of the prephonatory glottal angle), the minimal gap between the two vocal folds (called the prephonatory minimal glottal diameter, given by the vocal process gap), the space between the false vocal folds (false vocal fold gap), and the posterior glottal gap (the prephonatory abduction of the cartilaginous glottis).

The process of making a match involved the operator entering the prephonatory conditions (Table II) for the model based on the aerodynamic and acoustic measures obtained from the human. The starting prephonatory conditions were determined based on tests that were performed using the V3M model, which determined the impact of individual prephonatory settings on airflow measures. For example, increasing the diameter in the range of 0.005–0.08 cm while other prephonatory features were held constant (subglottal pressure $= 8$ cm H_2O; prephonatory glottal angle $= 5$ deg convergent) resulted in increases in the airflow peak, open quotient, DC airflow, and a more circular trajectory of motion of the surfaces of the three masses. Increasing the angle of convergence from 0 deg uniform to 20 deg convergent while other prephonatory features were held constant (subglottal pressure $= 8$ cm H_2O; prephonatory glottal diameter $= 0.04$ cm) resulted in an increase in the open quotient and a broadening of the left side of the glottal airflow pulse. When the convergent angle became large enough (i.e., greater than 10 deg convergent), mass $m2$ did not reach midline in its trajectory of motion. As $m2$ represents the lower portion of the glottis, this would be equivalent to a glottis that did not close in the inferior aspect. It also represents a reduction in the effective thickness of the vocal fold as well as an effective decrease in the vertical glottal distance (because the upper portion of the glottis is what is closest or contacting during the cycle).

Each run of the model took approximately 5 min to complete. The duration of the simulated waveforms was 1000 ms. The operator extracted the six airflow measurements from the glottal airflow signal and used Praat to visually determine the approximate values of the first, second, and third formant frequencies from the acoustic pressure signal. Based on the results, the prephonatory settings of the model were adjusted and the model was run again. A match was considered appropriate when the fundamental frequency was matched within 2% and the overall percent difference between the model and human airflow measurements was within 10%. Further runs were then completed by varying the prephonatory settings in smaller increments to improve the overall percent difference. After a minimum of five runs in which the overall percent difference did not decrease after varying the prephonatory model settings, the matching process for that production was terminated by the operator and the "match" declared. There was one case, the higher pitch, normal loudness condition for participant $S2$, in which the model was not able to create a match with the human data that had an overall percent difference lower than 10%. In this case, the matching process was terminated after 117 runs, the last 5 of which resulted in no improvement in the overall percent difference.

TABLE III. Comparison of human and model phonatory data for participant $S1$ for all conditions. "% Diff" represents the percent difference between the human and model values for that particular variable. Key: Psub, subglottal pressure; SD, standard deviation; % Diff, percent difference.

		Psub (cm H$_2$O) Value	Fundamental frequency (Hz) Value	Fundamental frequency (Hz) % Diff	AC airflow (ml/s) Value	AC airflow (ml/s) % Diff	Open quotient Value	Open quotient % Diff	DC airflow (ml/s) Value	DC airflow (ml/s) % Diff	Skewing quotient Value	Skewing quotient % Diff	MFND (L/s^2) Value	MFND (L/s^2) % Diff	Overall % diff (SD)
Normal pitch,	Human	6.50	106.10	0.28%	417.43	0.10%	0.46	10.87%	95.55	0.07%	1.72	2.91%	−565.2	34.94%	8.20% (13.75)
normal loudness	Model		106.40		417.85		0.51		95.62		1.67		−367.7		
Higher pitch,	Human	5.54	146.76	1.55%	404.50	0.39%	0.92	22.13%	142.70	2.47%	1.97	22.52%	−524.7	1.39%	8.41% (10.80)
normal loudness	Model		149.03		406.07		0.72		146.23		1.53		−517.4		
Normal pitch, loud	Human	8.61	92.72	1.68%	559.77	20.83%	0.42	13.43%	31.27	4.19%	2.47	0%	−1138.0	11.42%	8.59% (8.03)
	Model		91.16		676.35		0.47		32.58		2.47		−1008.0		
Normal pitch, soft	Human	2.82	94.36	0.74%	261.12	0.57%	0.67	8.78%	111.00	2.10%	1.59	5.66%	−187.0	14.17%	5.34% (5.37)
	Model		95.06		259.63		0.61		113.33		1.50		−160.5		
Overall % Diff (SD)				1.06% (0.67)		5.47% (10.24)		13.80% (5.87)		2.21% (1.69)		7.77% (10.10)		15.48% (14.09)	

III. RESULTS

The average number of runs to create an acceptable match was 85.5 runs (range, 19–148), each with different tissue properties and prephonatory glottal configurations. Each run was performed by the operator (not automated). Matches for four conditions were obtained for two participants, $S1$ and $S3$: (1) normal pitch, normal loudness; (2) higher pitch, normal loudness; (3) normal pitch, louder; and (4) normal pitch, softer. Matches for (1) normal pitch, normal loudness and (2) higher pitch, normal loudness were obtained for participant $S2$. Matches between the human and model were evaluated for overall percent mismatch. The model run with the lowest overall percent mismatch is presented below for each condition. The normal pitch, normal loudness condition was used as a comparison point for understanding the relationship between the model settings and the aerodynamic and acoustic variables.

A. Participant $S1$

The matches between participant $S1$ and the model are presented in Table III and the model settings are presented in Table IV. The model was able to match the normal pitch, normal loudness within 8.20% [standard deviation (SD) = 13.75%] for this participant (Fig. 2). Fundamental frequency, AC airflow, and DC airflow were able to be matched within 2.00%. MFND was the most poorly matched variable for $S1$ in the normal pitch, normal loudness condition (34.94% difference).

TABLE IV. Model settings for matches for participant $S1$ across all conditions. *Diff* represents the difference in the model setting for a given condition from the normal pitch, normal loudness condition. Positive difference values indicate an increase in that parameter and negative difference values indicate a decrease in that parameter. Conv., convergent.

|		Normal pitch, normal loudness Model settings	Higher pitch, normal loudness Model settings	Higher pitch, normal loudness Diff	Normal pitch, loud Model settings	Normal pitch, loud Diff	Normal pitch, soft Model settings	Normal pitch, soft Diff
Model prephonatory	Vocal tract segment 1 (cm^2)	9	5	−4	5	−4	5	−4
settings	Vocal tract segment 2 (cm^2)	7	7	0	7	0	7	0
	Vocal tract segment 3 (cm^2)	3	7	+4	7	+4	4	+1
	Vocal tract segment 4 (cm^2)	1.7	1.2	−0.5	1.2	−0.5	1	−0.7
	Vocal tract segment 5 (cm^2)	1.7	1.2	−0.5	1.2	−0.5	2	+0.3
	Vocal tract segment 6 (cm^2)	5	2	−3	2	−3	3.5	−1.5
	$K0$ (dyne/cm)	100 000	100 000	0	60 000	−40 000	60 000	−40 000
	$K1$ (dyne/cm)	50 000	50 000	0	30 000	−20 000	30 000	−20 000
	$K2$ (dyne/cm)	25 000	25 000	0	15 000	−10 000	15 000	−10 000
	$K3$ (dyne/cm)	2500	2500	0	1500	−1000	1500	−1000
	Kcp (dyne/cm)	7812.50	7812.50	0	4687.5	−3125	4687.5	−3125
	Kc (dyne/cm)	7812.50	7812.50	0	4687.5	−3125	4687.5	−3125
	Angle (deg) (pos = divergent; neg = convergent)	0	−13 (conv.)	−13	0	0	−5 (conv.)	−5
	$m1$ and $m2$ (g)	0.092188	0.04375	−0.04813	0.09375	+0.001562	0.060625	−0.031563
	$m3$ (g)	0.018438	0.00875	−0.009688	0.01875	+0.00031	0.012125	−0.006313
	Glottal diameter (cm)	0.0051	0.045	+0.0399	0	−0.0051	0.026	+0.0209
	False vocal fold gap (cm)	0.4	0.4	0	0.45	+0.05	0.3	−0.1
	Posterior glottal gap (cm)	0.0699	0.12	+0.0501	0.02	−0.0499	0.131	+0.0611
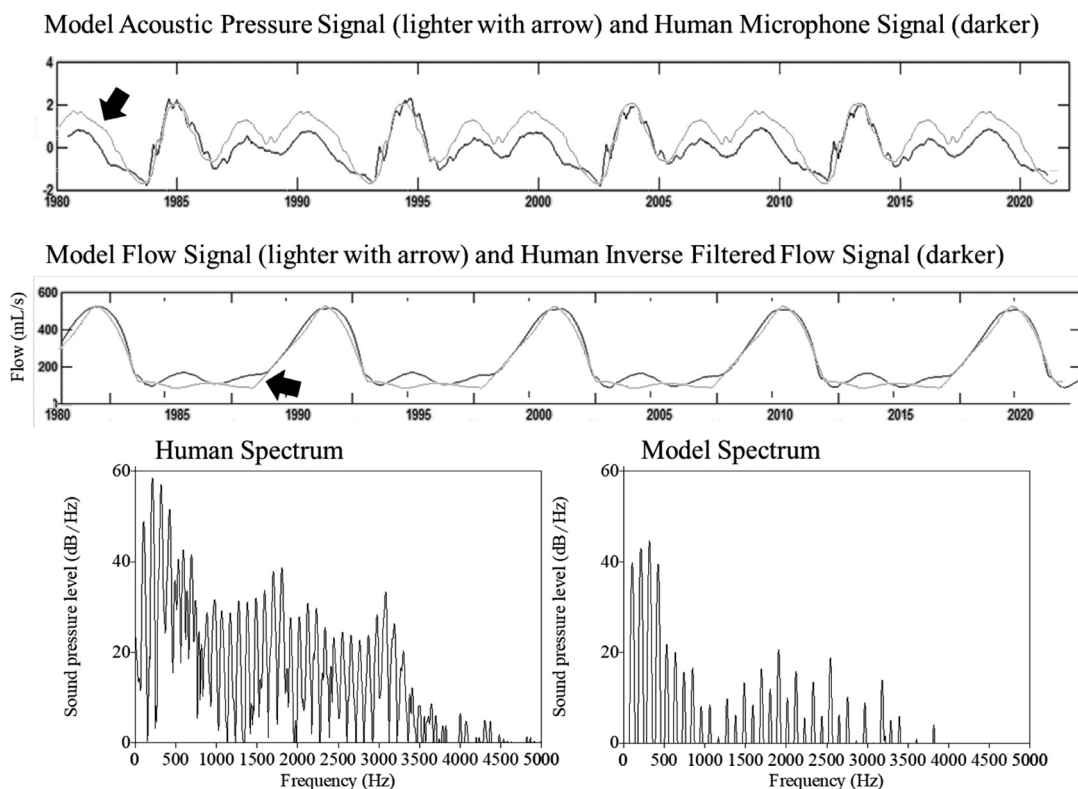
FIG. 2. Normal pitch and normal loudness match for $S1$. The bold arrows point to the modeled signals.

When increasing pitch, participant $S1$ decreased the subglottal pressure by $0.96\,cm\ H_2O$ and increased fundamental frequency by 5.32 semitones (ST) compared to the normal pitch, normal loudness condition. To model this change, the tension settings of the model were not changed, but the mass of all three masses was decreased by 53.00% from the normal pitch, normal loudness condition. The glottal angle also became more convergent, and the glottal diameter increased. This resulted in an overall percent mismatch between the human production and model output of 8.41% (SD $= 10.80\%$). The match between the human and model's acoustic pressure, airflow, and spectrum can be visualized in Fig. 3. From Fig. 3, it is within 3% and depicted in Table III). The poorest matched features were the open quotient and skewing quotient. The human had an open quotient of 0.92 while the model's open quotient was 0.72. The skewing quotient of the human was 1.97, whereas glottal opening in the model takes slightly less time compared to glottal closing, resulting in a skewing quotient of 1.53.

When producing the normal pitch, loud condition, $S1$ increased subglottal pressure by $2.11\,cm\ H_2O$ and fundamental frequency was decreased by 2.30 ST compared to the normal pitch, normal loudness condition. To model this decrease in fundamental frequency, likely secondary to an increase in the amount of tissue in vibration, tension was reduced by 40.00% and mass was increased slightly (1.70% increase) in the loud condition. The fundamental frequency and skewing quotient were matched within 2%. The AC

airflow and open quotient in the model output were higher compared to the human production despite setting the glottal diameter to $0\,cm$. AC airflow was the worst matched feature in this condition (20.83% difference). The open quotient had a 13.43% mismatch. The skewing quotient increased from the normal loudness condition to the loud condition. When modeling this condition, the first vocal tract segment was reduced from the normal loudness condition to the louder condition, suggesting that the skewing may be due to the increase in vocal tract inertance just above the glottis (Titze, 2004a). The overall percent mismatch was 8.60% (SD $= 8.00\%$). The match between the human and model acoustic pressure, airflow, and spectra are displayed in Fig. 4.

The soft loudness condition for $S1$ was produced with a $3.68\,cm\ H_2O$ reduction in subglottal pressure and a 2.01 ST ($11.65\,Hz$) decrease in fundamental frequency compared to the normal pitch, normal loudness condition. In this condition, $S1$ used a low subglottal pressure of $2.82\,cm\ H_2O$, which is near the phonation threshold pressure of males using a lower pitch (Chang and Karnell, 2004). This suggests that the small convergent angle that was used (5 de convergent) may provide a highly efficient glottal contour for voice production near phonation threshold pressure. The decrease in fundamental frequency was achieved by a 40.0% lower tension setting for all springs despite a 34.2% lower mass compared to the normal pitch, normal loudness condition. For $S1$, the prephonatory glottal diameter in the softer condition was around 50.00% less than that in the higher pitch condition. The match with the lowest percent difference between the
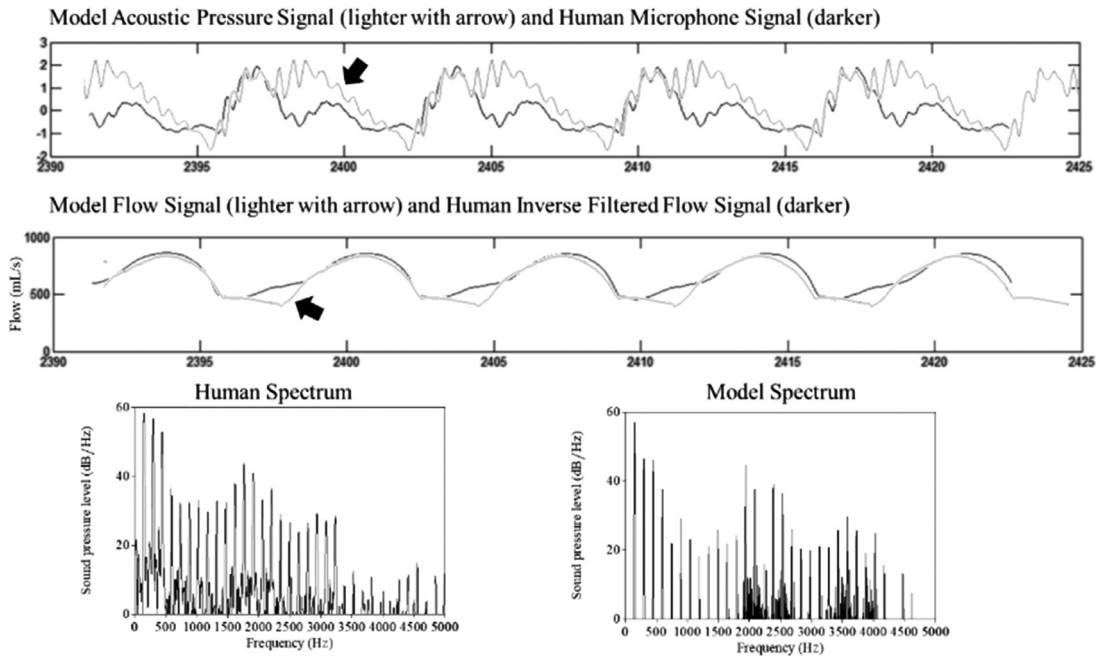
JASA

FIG. 3. Higher pitch and normal loudness match for *S*1. The bold arrows point to the modeled signals.

human and model had an overall percent difference of 5.34% (SD = 5.37%) and is shown in Fig. 5. The fundamental frequency and AC airflow were matched within 1.00% between the human and the model. The feature that had the poorest match was the MFND with a 14.17% mismatch.

B. Participant *S*2

Participant *S*2 was studied for two conditions, the normal pitch, normal loudness condition and the higher pitch, normal loudness condition. Two matches are presented for the higher pitch, normal loudness condition in
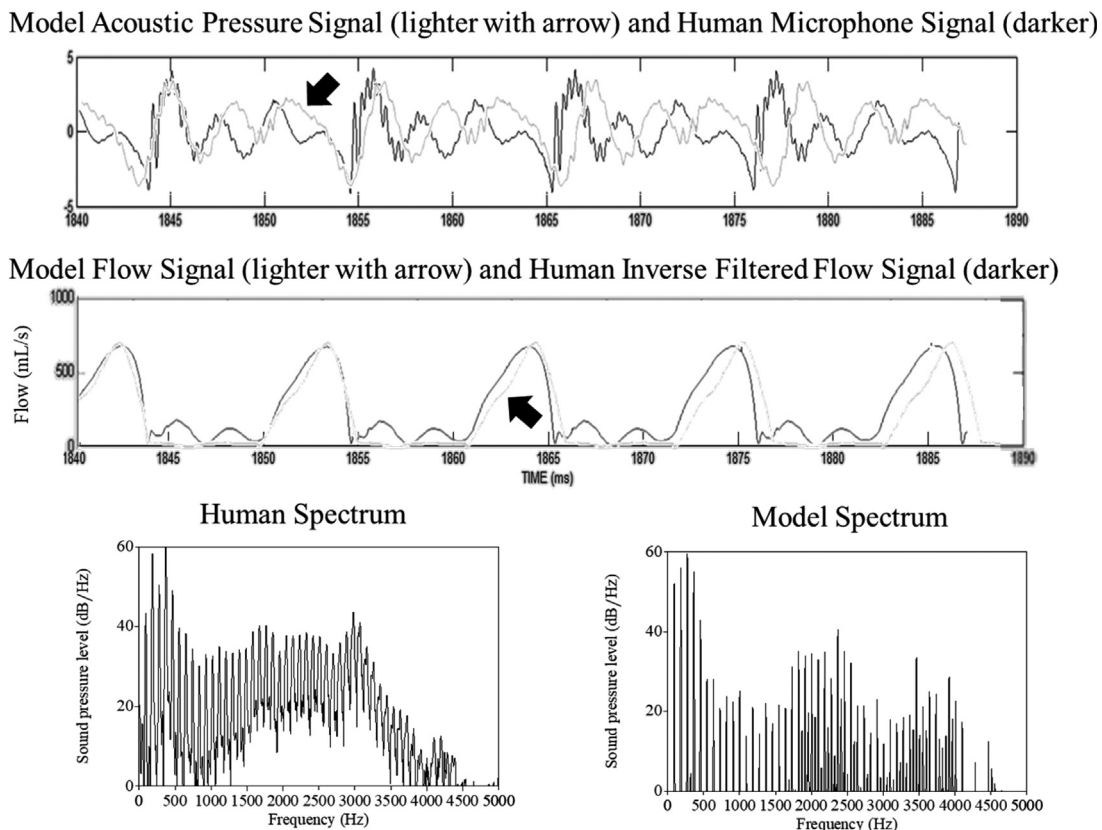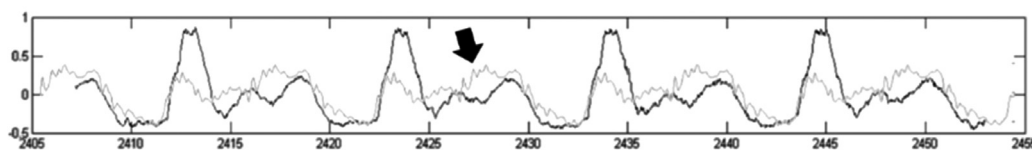

FIG. 4. Normal pitch and loud match for *S*1. The bold arrows point to the modeled signals.

Model Acoustic Pressure Signal (lighter with arrow) and Human Microphone Signal (darker)

Model Flow Signal (lighter with arrow) and Human Inverse Filtered Flow Signal (darker)
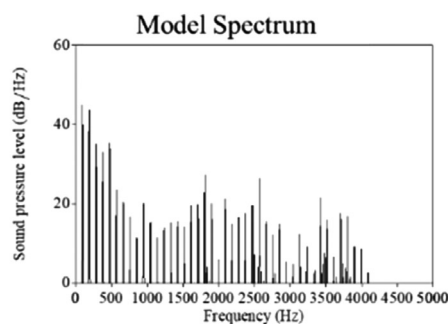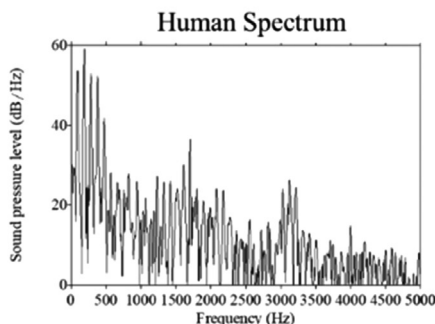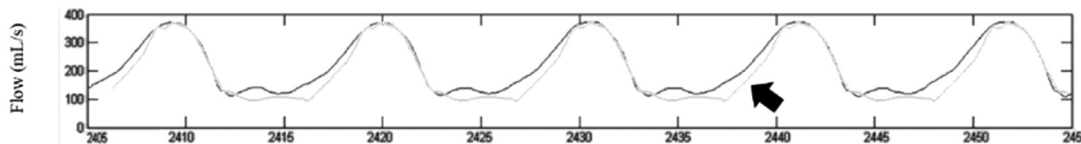
Human Spectrum

Model Spectrum

FIG. 5. Normal pitch and soft match for $S1$. The bold arrows point to the modeled signals.

Table V. Table VI presents the model settings for these runs.

For the normal pitch, normal loudness condition (Fig. 6), the model was able to match the production of $S2$ with a 4.90% (SD = 6.71%) mismatch, which is a lower percent mismatch than any condition for $S1$ and $S2$. Three features were able to be matched with less than 1% mismatch: fundamental frequency, AC airflow, DC airflow, and skewing quotient. The model had an open quotient that was less than that for the human (12.70% mismatch) and a lower MFND during the airflow shut off compared to the human (14.36% mismatch).

Participant $S2$ increased fundamental frequency by 5.33 ST and increased subglottal pressure by 0.98 cm H_2O

when increasing the pitch. Two matches for this condition are presented to demonstrate the difficulty that the model had in matching certain features (Tables V and VI and Figs. 7 and 8). In both runs, the tension of all the springs was increased and the mass of all the masses was decreased from the normal pitch, normal loudness condition. The alternate run for the higher pitch, normal loudness condition had tension settings between the normal pitch, normal loudness condition and the other higher pitch, normal loudness run, and mass settings lower than both (Table VI).

The model came close to matching the DC airflow used by the human with a difference less than 2 ml/s in the higher pitch, normal loudness condition. For both matches presented, the posterior glottal gap was reduced to 0 cm, but

TABLE V. Comparison of human and model phonatory data for participant $S2$ for all conditions. "% Diff" represents the percent difference between the human and model values for that particular variable. Psub, subglottal pressure; SD, standard deviation; % Diff, percent difference; *not included in overall % difference calculation.

		Psub (cm H_2O) Value	Fundamental frequency (Hz) Value	% Diff	AC airflow (ml/s) Value	% Diff	Open quotient Value	% Diff	DC airflow (ml/s) Value	% Diff	Skewing quotient Value	% Diff	MFND (L/s^2) Value	% Diff	Overall % Diff (SD)
Normal Pitch,	Human	7.76	133.05	0.62%	507.36	0.62%	0.63	12.70%	22.74	0.44%	1.55	0.65%	–544.0	14.36%	4.90% (6.71)
Normal Loudness	Model		133.87		504.22		0.55		22.64		1.56		–465.9		
Higher Pitch,	Human	8.74	180.87	0.20%	482.25	5.02%	0.82	15.12%	0.00	—*	2.23	35.54%	–525.9	9.36%	10.87% (13.38)
Normal Loudness	Model		180.51		506.47		0.70		1.53		1.44		–476.7		
Higher Pitch,	Human	8.74	180.87	3.73%	482.25	2.46%	0.82	3.55%	0.00	—	2.23	43.95%	–525.9	14.57%	11.38% (16.73)
normal loudness alternate run*	Model		187.62		494.11		0.85		1.62		1.25		–449.3		
Overall % Diff (SD)			0.41% (0.30)		2.82% (3.11)		13.91% (1.71)		0.44%		18.09% (24.67%)		11.86% (24.67)		

J. Acoust. Soc. Am. **154** (3), September 2023

Brittany L. Perrine and Ronald C. Scherer   1513

TABLE VI. Model settings for matches for participant *S*2 across all conditions. *Diff* represents the difference in the model setting for a given condition from the normal pitch, normal loudness condition. Positive difference values indicate an increase in that parameter and negative difference values indicate a decrease in that parameter.

		Normal pitch, normal loudness Model settings	Higher pitch, normal loudness		Higher pitch, normal loudness alternate run	
			Model settings	Diff	Model settings	Diff
Model Prephonatory Settings	Vocal tract segment 1 (cm^2)	5	4	−1	6	+1
	Vocal tract segment 2 (cm^2)	7	8	+1	8	+1
	Vocal tract segment 3 (cm^2)	3	8	+5	8	+5
	Vocal tract segment 4 (cm^2)	1.5	2	+0.5	1	−0.5
	Vocal tract segment 5 (cm^2)	1.5	2	+0.5	1	−0.5
	Vocal tract segment 6 (cm^2)	4	6	+2	4	0
	$K0$ (dyne/cm)	160 000	280 000	+120 000	256 000	+96 000
	$K1$ (dyne/cm)	80 000	140 000	+60 000	128 000	+48 000
	$K2$ (dyne/cm)	40 000	70 000	+30 000	64 000	+24 000
	$K3$ (dyne/cm)	4000	7000	+3000	6400	+2400
	Kcp (dyne/cm)	12 500	21 875	+9375	20 000	+7500
	Kc (dyne/cm)	12 500	21 875	+9375	20 000	+7500
	Angle (deg) (pos = divergent; neg = convergent)	−5	0	−5	0	−5
	$m1$ and $m2$ (g)	0.085	0.071875	−0.013125	0.053125	−0.031875
	$m3$ (g)	0.017	0.014375	−0.002625	0.010625	−0.006375
	Glottal diameter (cm)	0.01	0.05	+0.04	0.065	+0.055
	False vocal fold gap (cm)	0.5	0.6	+0.06	0.6	+0.06
	Posterior glottal gap (cm)	0.013	0	−0.013	0	−0.013

there was still a small amount of DC airflow produced by the model. When the human produced the higher pitch sound, the airflow signal suggests that the vocal folds only touched for a moment to bring the airflow to 0 ml/s. The

model was able to match the fundamental frequency within 1% for the higher pitch run presented in the top panel of Fig. 8, but that run was unable to capture the large open quotient (15.12% difference) and the skewing quotient (35.54%
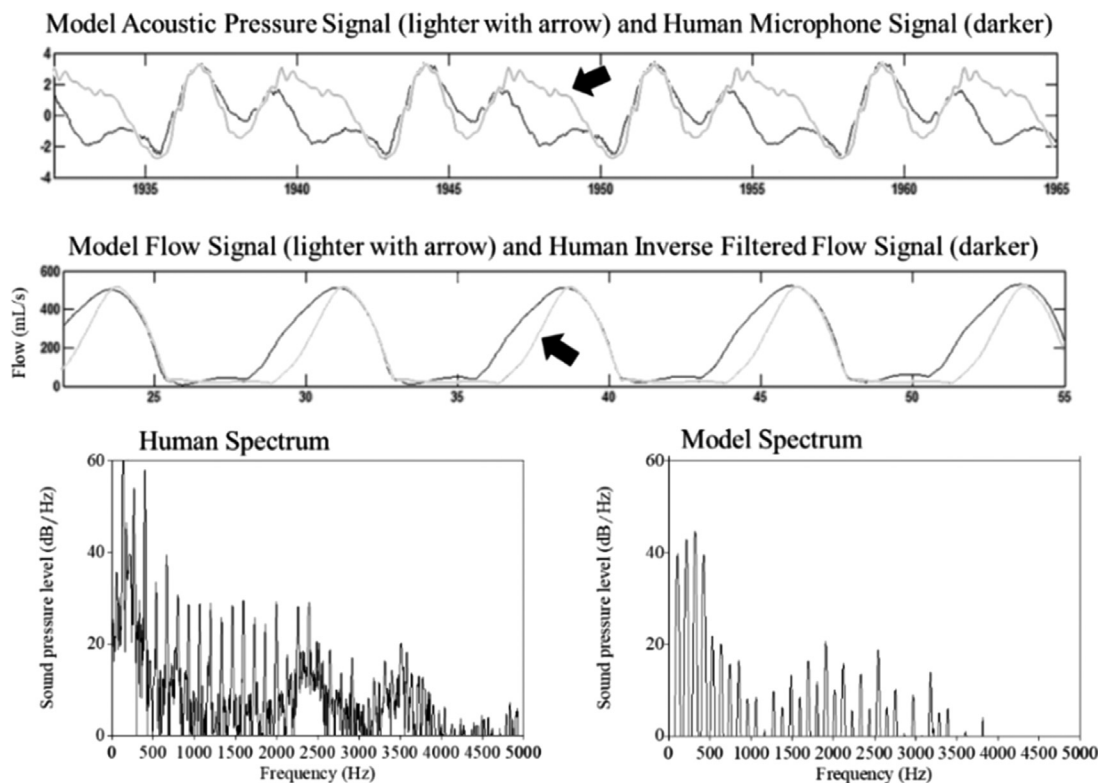

FIG. 6. Normal pitch and normal loudness match for *S*2. The bold arrows point to the modeled signals.
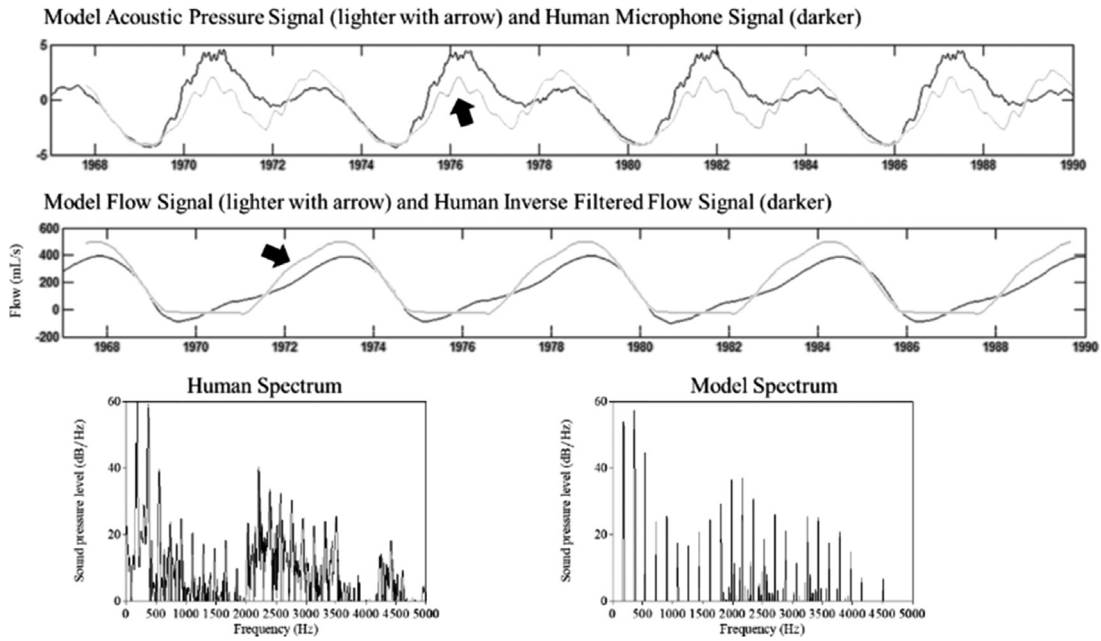
FIG. 7. Higher pitch and normal loudness match for *S*2. The bold arrows point to the modeled signals.

difference; Table V). Stated otherwise, this run of the model did not accurately capture the behavior of the human because the vocal folds were in contact too long in the model run.

In the alternate run (bottom panel of Fig. 8), the model was better able to capture the reduced glottal closed time that occurs at higher pitches. In addition to lowering the tension and mass, this run had a wider glottal diameter than the presented first match (Table VI). In this run, the open quotient for the model was higher than the human produced (3.56% mismatch, Table V), but the match for the skewing quotient (43.95% mismatch) remained poor.

In both runs, the skewing quotient could not be matched well for the higher pitch, normal loudness condition of *S*2. *S*2 had a skewing quotient of 2.23 in this condition and the model could only produce a skewing quotient of 1.44 and
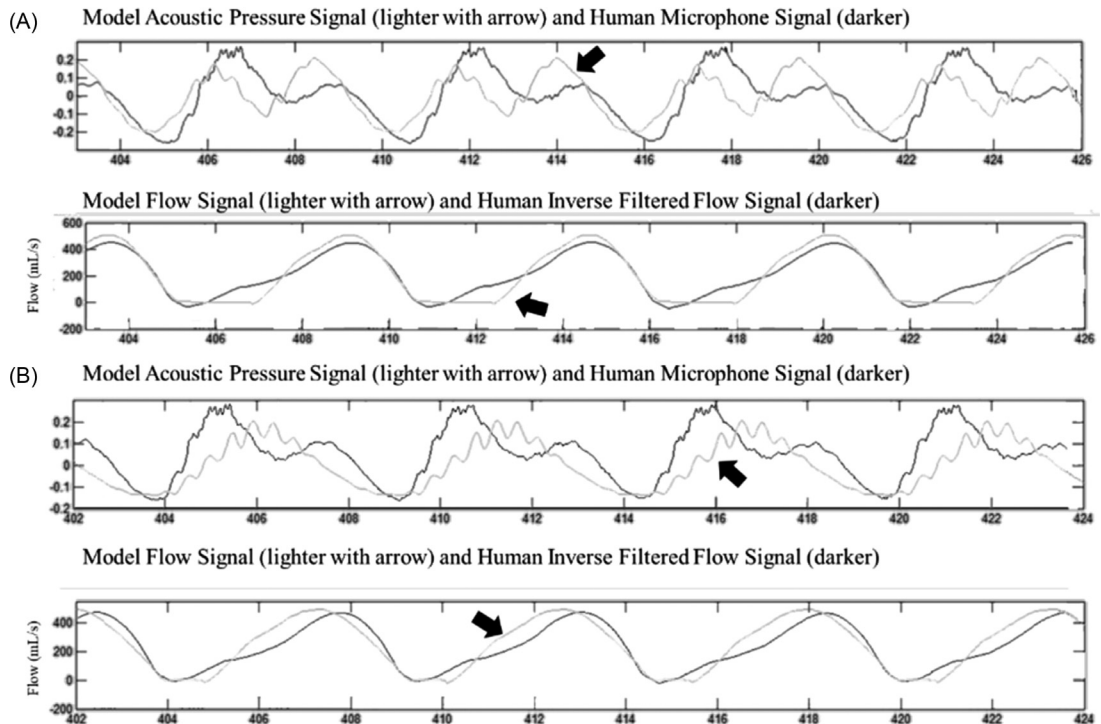

FIG. 8. Higher pitch, normal loudness matches for *S*2. (A) shows the match that is presented in Table IV and (B) shows a match in which mass, tension, and prephonatory glottal diameter are changed to improve the matches between AC airflow and open quotient. The bold arrows point to the modeled signals.

TABLE VII. Comparison of human and model phonatory data for participant S3 for all conditions. "% Diff" represents the percent difference between the human and model values for that particular variable. Psub, subglottal pressure; SD, standard deviation; % Diff, percent difference.

		Psub (cm H$_2$O) Value	Fundamental frequency (Hz)		AC airflow (ml/s)		Open quotient		DC airflow (ml/s)		Skewing quotient		MFND (L/s^2)		Overall % Diff (SD)
			Value	% Diff	Value	% Diff	Value	% Diff	Value	% Diff	Value	% Diff	Value	% Diff	
Normal pitch, normal loudness	Human	11.7	109.13	0.48%	403.84	1.27%	0.62	8.06%	184.8	0.45%	1.40	14.29%	−339.3	2.30%	4.47% (5.59)
	Model		109.65		408.97		0.57		185.63		1.60		−331.5		
Higher pitch, normal loudness	Human	14.82	201.71	0.76%	456.80	3.87%	0.82	2.44%	22.49	0.93%	1.87	0.53%	−556.8	1.10%	1.61% (1.30)
	Model		203.25		439.10		0.8		22.70		1.86		−562.9		
Normal pitch, loud	Human	12.83	117.96	0.08%	476.22	6.81%	0.55	4.34%	44.16	0.23%	2.25	1.78%	−607.1	3.46%	2.78% (2.61)
	Model		118.06		508.67		0.53		44.26		2.21		−586.1		
Normal pitch, soft	Human	7.29	95.19	0.23%	263.19	0.24%	0.75	6.21%	68.89	0.55%	1.67	22.75%	−129.8	1.54%	5.75% (10.06)
	Model		94.97		262.56		0.70		69.27		1.24		−127.8		
Overall % Diff (SD)			0.39% (0.30)		3.05% (2.94)		5.26% (2.42)		0.54% (0.30)		10.59% (11.86)		2.10% (1.03)		

1.25 with the model settings used. The model was able to replicate a skewing quotient above two for the normal pitch, loud conditions of S1 and S3. Both runs with skewing quotients above two (the normal pitch, normal loudness matches for S1 and S2) had model settings with less tension in all of the springs compared to the higher pitch, normal loudness model run for S2. The skewing of the glottal airflow is dependent on glottal area changes and vocal tract inertance (Titze, 2015). Based on the higher skewing quotient, lower MFND, and slightly lower AC airflow compared to the model, it seems likely that the human had greater vocal tract inertance in this condition than the model provided. This could have been mediated by reducing the prephonatory glottal diameter, which also may have resulted in a better match for the AC airflow.

C. Participant S3

Table VII presents the matches between the model and the human for each of the four conditions for participant S3. The model settings for those matches are presented in Table VIII. For the normal pitch, normal loudness condition, the model was able to match the human production with 4.5% (SD = 5.6%) mismatch (Fig. 9). Fundamental frequency and DC airflow were matched within 1%, AC airflow was matched within 2%, and MFND was matched within 3.5%. The model had a higher skewing quotient compared to the human, which resulted in a 14.29% mismatch for that aerodynamic feature.

When changing from the normal pitch, normal loudness to the higher pitch, normal loudness condition, participant S3 increased fundamental frequency by 10.63 ST and

TABLE VIII. Model settings for matches for participant S3 across all conditions. Diff represents the difference in the model setting for a given condition from the normal pitch, normal loudness condition. Positive difference values indicate an increase in that parameter and negative difference values indicate a decrease in that parameter. Conv., convergent.

|		Normal pitch, normal loudness Model settings	Higher pitch, normal loudness		Normal pitch, loud		Normal pitch, soft	
			Model settings	Diff	Model settings	Diff	Model settings	Diff
Model prephonatory settings	Vocal tract segment 1 (cm^2)	3	3	0	3	0	3	0
	Vocal tract segment 2 (cm^2)	5	5	0	7	+2	2	−3
	Vocal tract segment 3 (cm^2)	5	5	0	7	+2	1	−4
	Vocal tract segment 4 (cm^2)	2	2	0	2	0	0.5	−1.5
	Vocal tract segment 5 (cm^2)	1	1	0	1	0	0.6	−0.4
	Vocal tract segment 6 (cm^2)	0.5	0.5	0	0.5	0	1.2	+0.7
	$K0$ (dyne/cm)	160 000	260 800	+100 800	80 000	−80 000	208 000	+48 000
	$K1$ (dyne/cm)	80 000	130 400	+50 400	40 000	−40 000	104 000	+24 000
	$K2$ (dyne/cm)	40 000	65 200	+25 200	20 000	−20 000	52 000	+12 000
	$K3$ (dyne/cm)	4000	6520	+2520	2000	−2000	5200	+1200
	Kcp (dyne/cm)	12 500	20 375	+7875	6250	−6250	16 250	+3750
	Kc (dyne/cm)	12 500	20 375	+7875	6250	−6250	16 250	+3750
	Angle (deg) (pos = divergent; neg = convergent)	0	−2.5 (conv.)	−2.5	0	0	0	0
	$m1$ and $m2$ (g)	0.115625	0.046875	−0.06875	0.0625	−0.05313	0.157188	+0.041563
	$m3$ (g)	0.023125	0.00925	−0.013875	0.0125	−0.01063	0.031438	+0.008313
	Glottal diameter (cm)	0.004	0.0225	+0.0185	0	−0.004	0.0235	+0.0195
	False vocal fold gap (cm)	0.4	0.4	0	0.4	0	0.6	+0.2
	Posterior glottal gap (cm)	0.111	0.0093	−0.1017	0.02	−0.091	0.051	−0.06
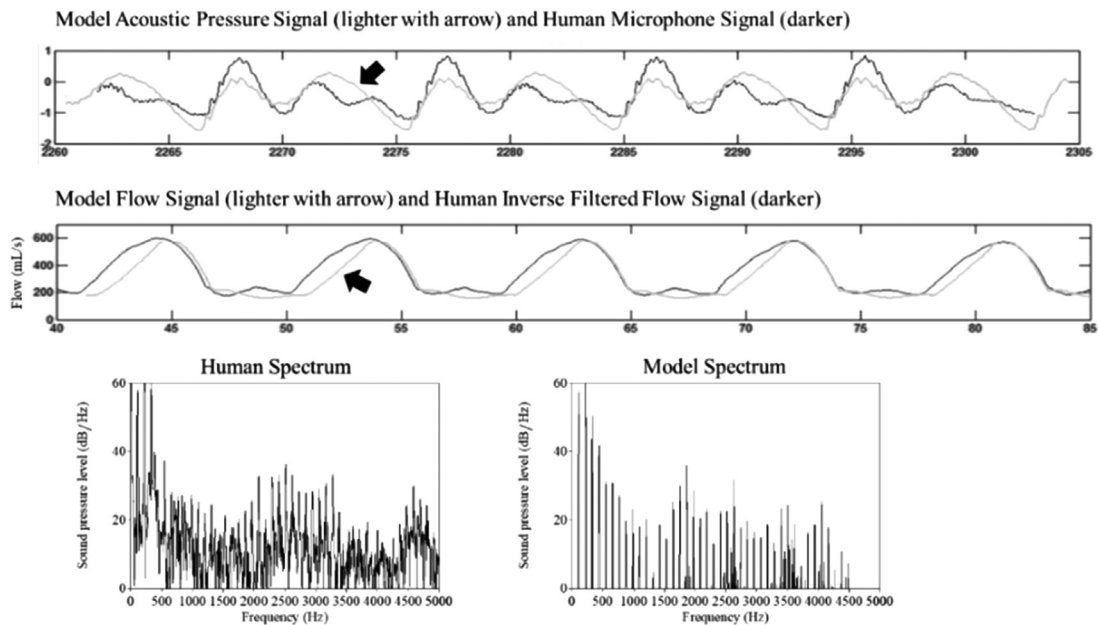
FIG. 9. Normal pitch and normal loudness match for $S3$. The bold arrows point to the modeled signals.

increased subglottal pressure by 3.12 cm H_2O. The model was able to produce a match with an overall percent difference of 1.61% (SD $= 1.30$%), as depicted in Fig. 10, which is the match with the lowest percent difference in the study. The fundamental frequency, DC airflow, and skewing quotient were matched within 1%. The remaining features were matched within 4%. To create this match, the tension was increased, and the mass was decreased compared to the normal pitch, normal loudness condition. The glottal diameter was increased slightly from the normal pitch, normal loudness condition to model the slight increase in airflow used, and the posterior glottal gap was reduced to 0.0093 cm to model the decrease in DC airflow used in this condition. The diameters of the vocal tract segments were unchanged.
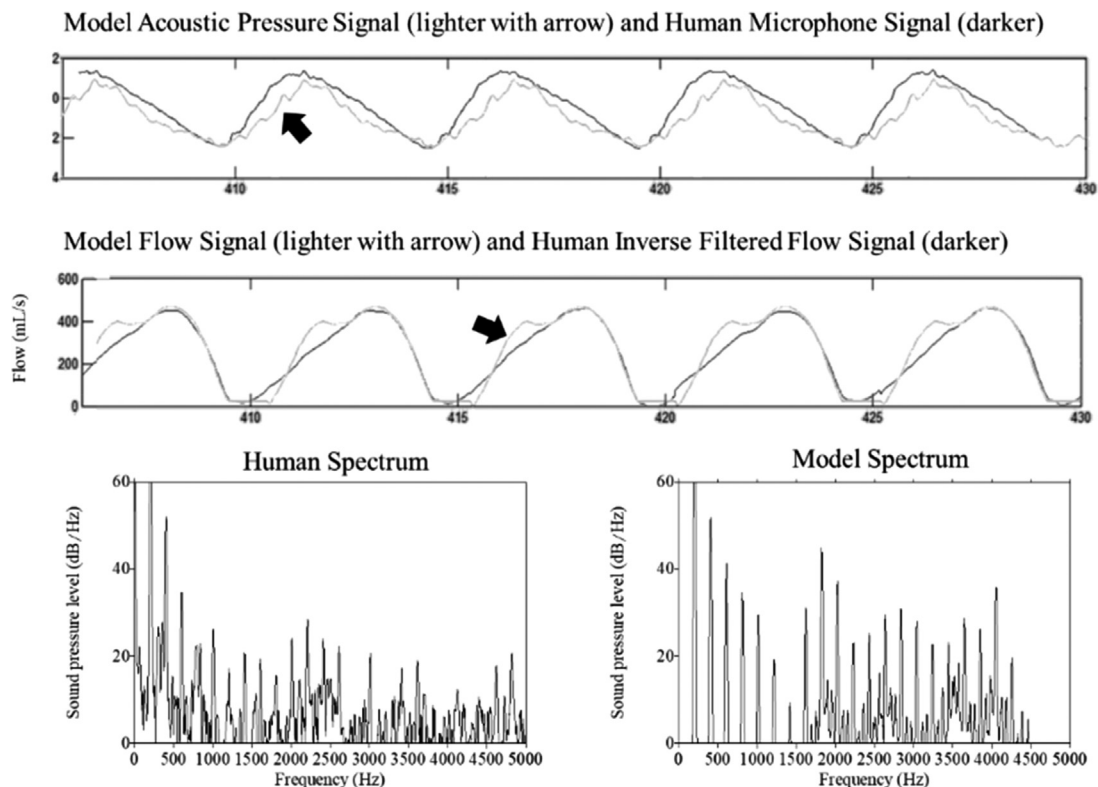

FIG. 10. Higher pitch and normal loudness match for $S3$. The bold arrows point to the modeled signals.

$S3$ only increased subglottal pressure by 1.13 cm H_2O from the normal pitch, normal loudness to normal pitch, loud condition. The fundamental frequency for $S3$ was 1.35 ST higher in the louder condition compared to the normal condition. This condition was able to be matched with a 2.78% (SD = 2.61%) mismatch by the model (Fig. 11). All features except for AC airflow were matched within 5%. For $S3$, despite the small increase in fundamental frequency, the best match for the loud condition involved a 43.95% reduction in mass and 50.00% reduction in tension from the normal loudness match. The fundamental frequency rise may have been due to the increase in subglottal pressure, which would increase the functional length of the vocal folds during maximum lateral excursion. Again, $S3$ used a lower DC airflow compared to the normal pitch, normal loudness condition such that the posterior glottal gap diameter was decreased. The glottal diameter was set to 0 cm for this condition, but the AC airflow in the model remained higher than the human, resulting in a 6.81% mismatch. For this condition, the skewing quotient increased from the normal pitch, normal loudness condition. Given that there was no change in vocal tract inertance across these two conditions in this participant, it is possible that the increase in skewing quotient in participant $S3$'s production was due to the phase closure of the upper and lower edges of the vocal fold tissue, which was not captured in the model output data.

$S3$ decreased subglottal pressure by 4.41 cm H_2O and fundamental frequency by 2.37 ST (13.94 Hz) when changing from the normal pitch, normal loudness condition to the normal pitch, soft condition. The model was able to match

the acoustic and aerodynamic features of this production with a 5.75% (SD = 10.06%) mismatch (Fig. 12). The best match involved a 30.00% increase in tension and a 35.95% increase in mass compared to the normal pitch, normal loudness condition. The glottal diameter and false vocal fold gap were also increased. While fundamental frequency, AC airflow, and DC airflow were able to be matched within 1.00% by the model, the skewing quotient was lower (22.75% mismatch) in the model compared to the human.

D. Matches between the formant frequencies of the humans and model

In addition to varying the properties of the vocal folds, the cross-sectional areas of the six vocal tract segments were modified to create formant frequency matches between the model and the human. The percent mismatch between the human and the model ranged from 1.4% to 20.9% for the first formant frequency ($F1$), 0% to 22.5% for the second formant frequency ($F2$), and 2.1% to 25.1% for the third formant frequency ($F3$). Results are presented in Table IX.

The vocal tract cross-sectional areas are presented in Fig. 13. The setting for the most upstream segment ranged from 3 to 9 cm^2. For the /i/ vowel, the first section of the vocal tract (from the glottis to c. 3 cm above the glottis) has been measured in humans at c. 4 cm^2 (Story et $al.$, 1996) and as wide as 8–10 cm^2 (Fant, 1960; Baer et $al.$, 1991; Mainka et $al.$, 2015). In Fant (1960), the initial section of the vocal tract above the glottis was just over 2 cm^2 with widening to c. 3 cm^2 above the glottis. This large widening
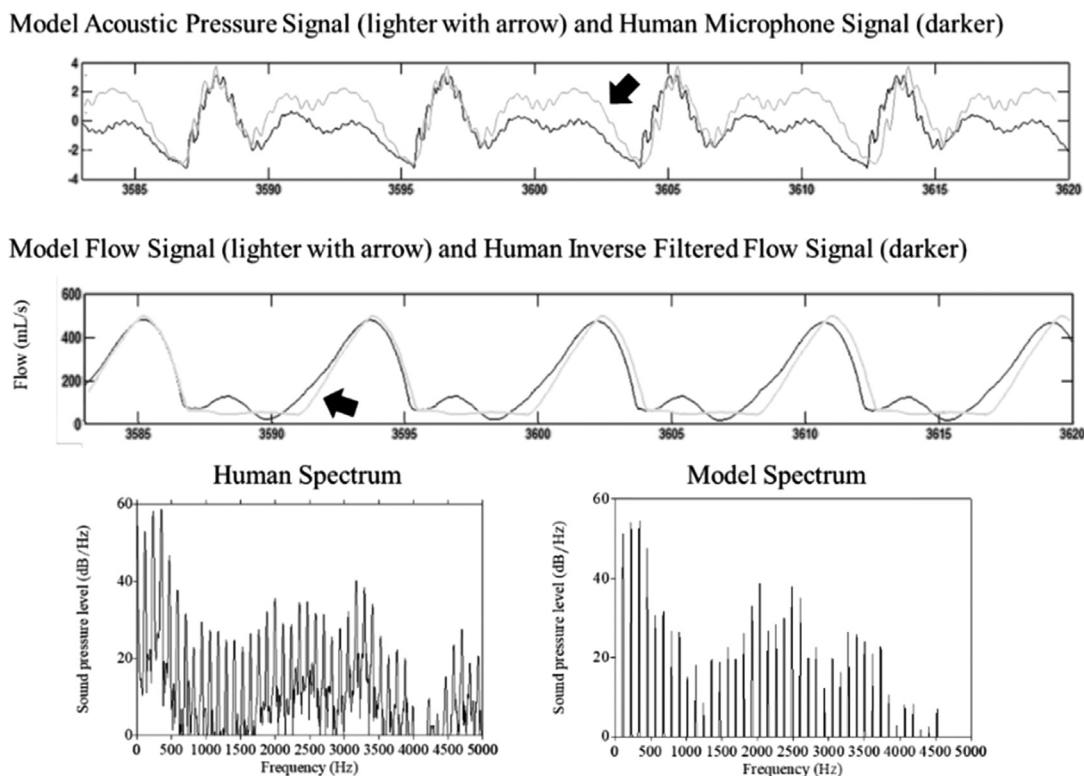

FIG. 11. Normal pitch and loud match for $S3$. The bold arrows point to the modeled signals.

1518   J. Acoust. Soc. Am. **154** (3), September 2023

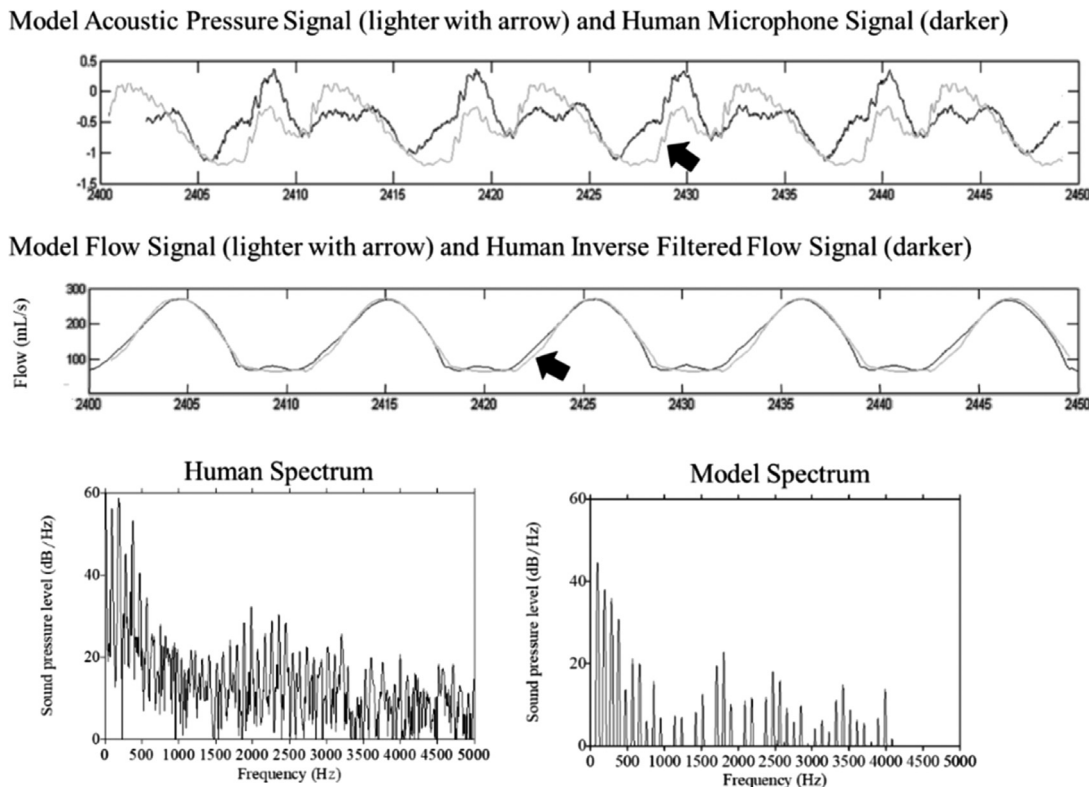Brittany L. Perrine and Ronald C. Scherer

FIG. 12. Normal pitch and soft match for S3. The bold arrows point to the modeled signals.

within 3 cm above the glottis also happened for the /u/ vowel in Fant (1960), where the vocal tract widened to just under 10 cm². The next two segments were set to a range 2–8 cm² and 1–8 cm², respectively. This section has ranged from around 4 cm² (Story *et al.*, 1996) to 10 cm² (Fant, 1960). The next two segments downstream were generally set to narrower values to represent the higher tongue position for the vowel /i/. The cross-sectional areas of these segments ranged from 0.5 to 2 cm². This segment has been measured at less than 1 cm² in many different human studies

(Baer *et al.*, 1991; Story *et al.*, 1996; Yang and Kasuya, 1994). The most downstream segment, segment 6, was set to a cross-sectional area that ranged from 0.5 to 5 cm², with the aforementioned studies measuring this segment at around 3 cm².

Checking the formant frequency values from the model compared to the human served is a verification that the vocal tract shape is plausible for the vowel produced. A six-section vocal tract is a relatively gross approximation to the actual cross-sectional area function of the participants' vocal

TABLE IX. Formant frequency value comparison between the human and model data for each condition.

|	F1 (Hz)			F2 (Hz)			F3 (Hz)		
	Human	Model	Diff (% diff)	Human	Model	Diff (% diff)	Human	Model	Diff (% diff)
Normal pitch, normal loudness									
S1	359	339	20 (5.6%)	2457	1914	543 (22.1%)	2828	2570	258 (9.1%)
S2	355	343	12 (3.4%)	2192	1935	257 (11.7%)	3305	2474	831 (25.1%)
S3	299	247	52 (17.4%)	2325	1816	509 (21.9%)	3105	2578	527 (17.0%)
Higher pitch, normal loudness									
S1	357	325	32 (9.0%)	1732	1926	−194 (11.2%)	2120	2424	−304 (14.3%)
S2	325	385	−60 (18.5%)	2054	2135	−81 (3.9%)	2705	2761	−56 (2.1%)
S3	430	443	−13 (3.0%)	2373	1840	533 (22.5%)	2738	2501	237 (8.7%)
Normal pitch, loud									
S1	384	309	75 (19.5%)	1727	1866	−139 (8.0%)	2479	2373	106 (4.3%)
S3	345	340	5 (1.4%)	2010	2010	0 (0.0%)	2620	2550	70 (2.7%)
Normal pitch, soft									
S1	303	347	−44 (14.5%)	1730	1743	−13 (0.8%)	3174	2468	706 (22.2%)
S3	325	257	68 (20.9%)	1919	1726	193 (10.1%)	2331	2484	−153 (6.6%)
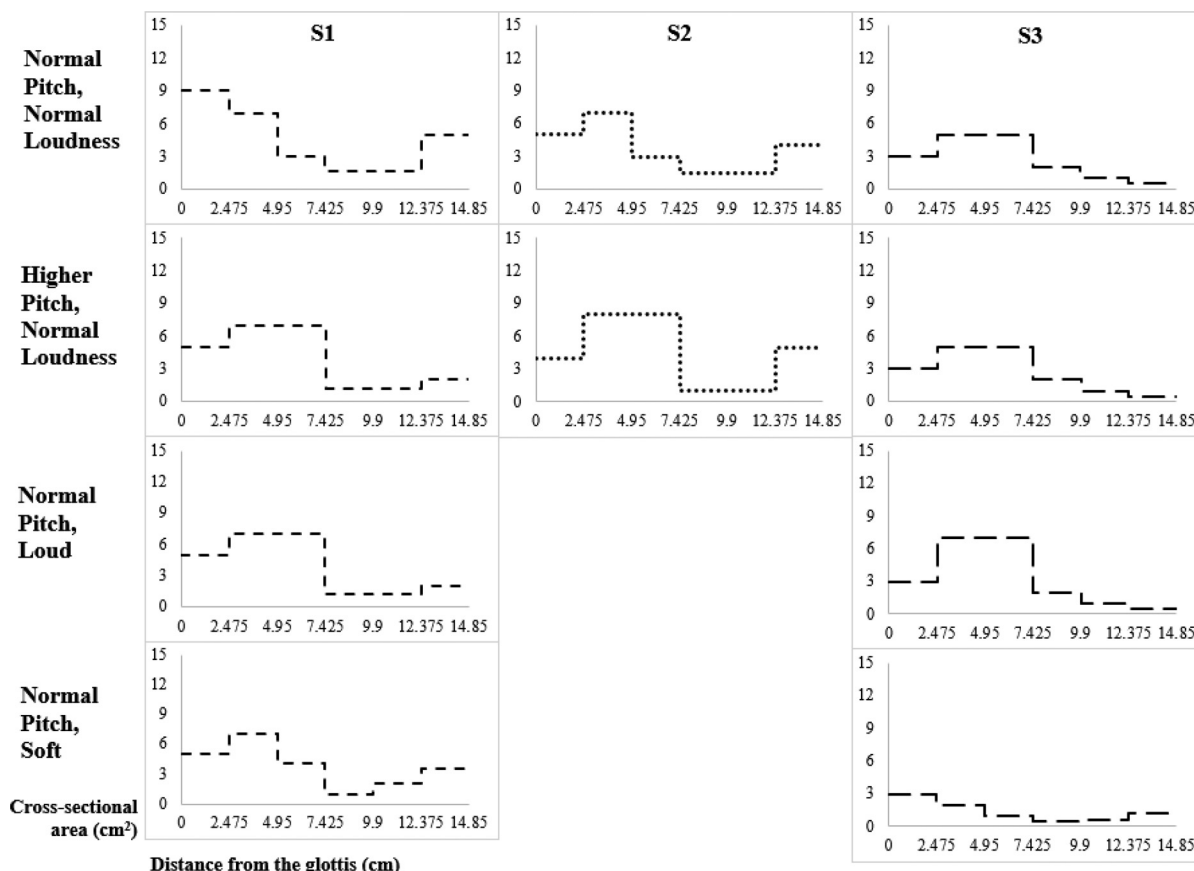
FIG. 13. Cross-sectional area of vocal tract segments for all matches and all conditions.

tracts, however. It is noted that the four-section vocal tract of I&F72 also could provide realistic formant values for a variety of vowels.

IV. GENERAL DISCUSSION

In this research, a low dimensional V3M vocal fold model with a six-segment vocal tract was used to simulate the approximate fundamental frequency, AC airflow, DC airflow, open quotient, skewing quotient, MFND, and acoustic output from human productions of a prolonged /i/-like vowel under different phonatory conditions (varying pitch and loudness). By varying the mass, tension, prephonatory glottal angle, glottal diameter, posterior glottal gap, false vocal fold gap, and the six segments of the vocal tract of the computer model, ten human production conditions were approximated with an average overall mismatch (average of the six matched parameters) of 1.61%–10.87% for the three male participants. These results suggest that despite the simplicity of the V3M model, the model using model $M5$ empirical intraglottal pressure values and a six-section vocal tract can approximate multiple human aerodynamic and acoustic parameters simultaneously. Looking across the three subjects, the average percent mismatch was the lowest for fundamental frequency (mean = 0.39%; SD = 0.26%) followed by DC airflow (mean = 0.48%; SD = 0.29%), AC airflow (mean = 3.00%; SD = 2.61%),

MFND (mean = 4.05%; SD = 4.36%), open quotient (mean = 6.99%; SD = 4.19%), and skewing quotient (mean = 11.49%; SD = 12.46%).

A. Variable relationships for the data sets across subjects

Each of the ten conditions (three normal pitch, normal loudness; three higher pitch, normal loudness; two normal pitch, loud; and two normal pitch, soft) across the three subjects described above were independent runs to match phonatory and acoustic values to the human subjects. It is acknowledged that there may be numerous combinations of prephonatory conditions that match any particular condition for a subject. This raises the question of whether or not the independently obtained matching values (using the V3M computer model) are consistent with reasonable notions of phonation. For each of the ten conditions, there are ten data points for each measure and, thus, graphs can be made to show the general relationships among the variables. Figure 14 shows these relationships for open quotient vs mass of $m2$, tension of $K2$ and glottal diameter, and glottal angle [Figs. 14(a)–14(d)], skewing quotient vs tension of $K2$ and glottal diameter [Figs. 14(e) and 14(f)], AC airflow vs posterior glottal gap [Fig. 14(g)], and DC airflow vs posterior glottal gap [Fig. 14(h)]. Figure 14(i) shows the relationship between the fundamental frequency estimated based on
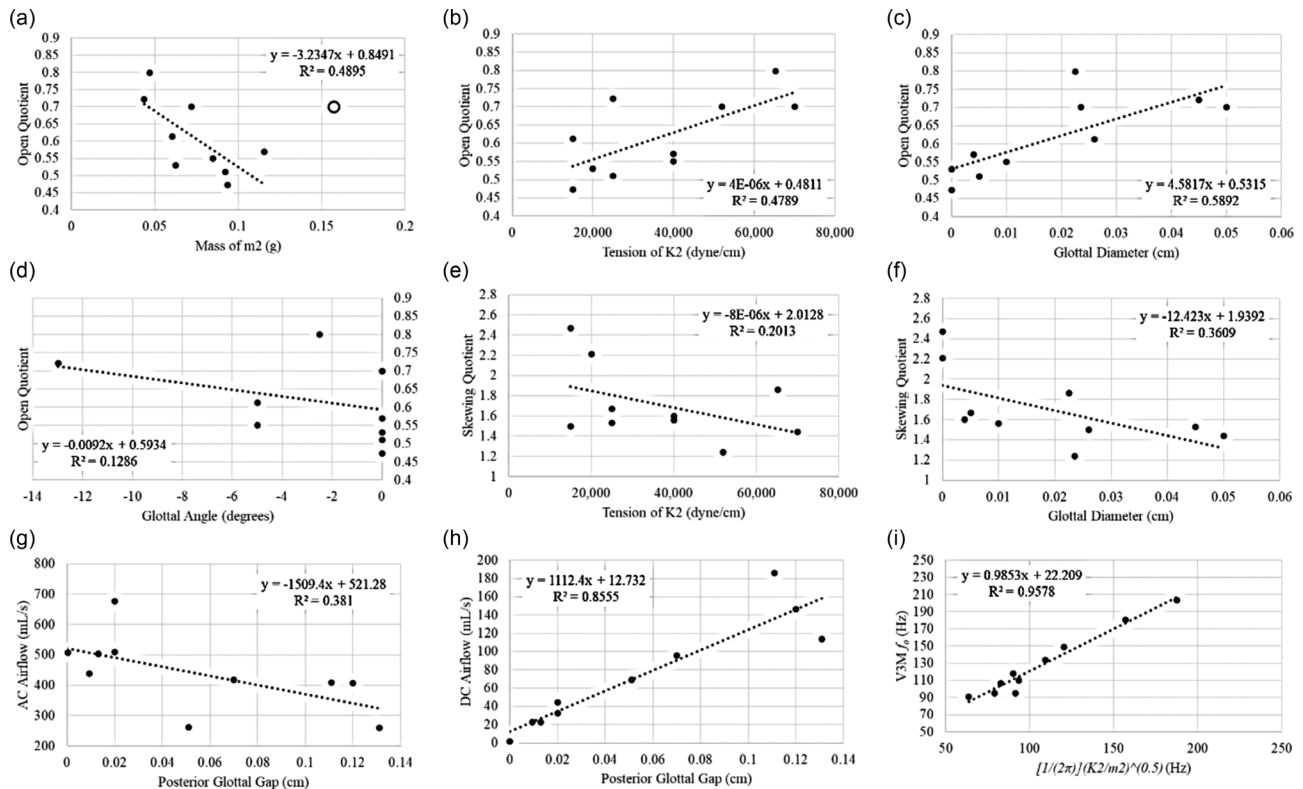
1520   J. Acoust. Soc. Am. **154** (3), September 2023

Brittany L. Perrine and Ronald C. Scherer

FIG. 14. Comparison of the prephonatory model settings (*x* axis) and measured variables (*y* axis) for the three-mass model, showing (a) open quotient vs mass, (b) open quotient vs tension of *K*2, (c). open quotient vs diameter, (d) open quotient vs glottal angle, (e) skewing quotient vs tension of *K*2, (f) skewing quotient vs diameter, (g) AC airflow vs posterior glottal gap, and (h) DC airflow vs posterior glottal gap.

the tension (*K*2) and mass (*m*2) of the model and the fundamental frequency measured from model output.

The linear relationships in Fig. 14(a) (excluding the outlier value indicated by an open circle; $r = 0.70$) suggests that the open quotient decreased (providing a longer glottal closed time) as the mass of the tissue in motion increased. As Fig. 14(b) indicates, the open quotient rose with an increase in the tension of the masses ($r = 0.69$). This may appear to be logical due to a presumed reduced lateral excursion and reduced contact time between the two vocal folds with, then, tension or stiffness increases. However, as pointed out by Wang *et al.* (2021b) and Zhang (2017), it depends on which layers of the vocal folds are given an increase in stiffness. The open quotient also increased as the glottal diameter increased [Fig. 14(c), $r = 0.77$], which is logical due to the greater prephonatory separation of the vocal folds (e.g., Klatt and Klatt, 1990). The relationship between the prephonatory glottal angle and the open quotient[Fig. 14(d), $r = 0.36$] has too few points within the angle range to show a significant relationship but it shows a decreasing trend, which is consistent with findings by Titze (2006, p. 214). The skewing quotient tended to decrease with an increase in the tension of the mass[Fig. 14(e), $r = 0.45$], consistent with the open quotient increasing with tension. The skewing quotient also decreased with an increase in the glottal diameter[Fig. 14(f), $r = 0.60$], which is again consistent with the open quotient increasing with glottal diameter. Furthermore, the AC airflow tended to

decrease with an increase in the posterior glottal gap[Fig. 14(g), $r = 0.62$], and the DC airflow tended to increase with the posterior glottal gap[Fig. 14(h), $r = 0.92$], consistent with modeling work performed by Zañartu *et al.* (2014). These results suggest that the operator process used in this study was relatively consistent regarding general trends for the relation among variables related to phonation. Figure 14(i) indicates further that the model is consistent ($r = 0.98$) relative to the fundamental frequency and its dependence on a wide range of mass (*m*2) and tension (*K*2) values [i.e., f_o is proportional to $(1/2 \pi)(k/m)^{0.5}$, the term for the natural frequency of a mass *m*2 with spring *K*2; Story and Titze, 1995; Titze and Story, 2002].

B. Comparisons of model settings across people

When modeling the higher pitch condition, the average percent mismatch ranged from 1.61% for *S*3 to 10.87% for participant *S*2. The change in fundamental frequency in the participants from "normal" to higher pitch, which averaged 60.38 Hz (7.19 ST), did not appear to be primarily driven by the change in the participants' subglottal pressure, which increased by only 1.04 cm H_2O (range, –0.96–3.12 cm H_2O) on average (cf. Titze, 1989). Thus, prephonatory mass and tension were varied to adjust the fundamental frequency. Specifically, the mass setting of all masses decreased in all participants when they increased pitch. The tension settings increased by an average of 69% in two of the three

participants in the higher pitch condition, the exception being $S1$. Tension settings were not changed but the mass of $S1$'s vocal folds was decreased by 52.54% from the normal to higher pitch condition.

In addition, the prephonatory glottal angle was more convergent in $S1$ and $S3$ but not in $S2$ compared to the normal pitch, normal loudness condition. This more convergent glottal angle for $S1$ and $S3$ makes sense because greater cricothyroid (CT) muscle activity would result in a more convergent glottal angle as the vocal folds are slightly stretched with the increased fundamental frequency (it is noted that vocal fold length was not altered in the model, however). In addition, when the glottal angle is more convergent, the time for effective glottal closure is reduced, which would aid in increasing the open quotient (Zhang, 2009).

For the normal pitch, loud condition, $S1$ had an overall percent mismatch of 8.59% (SD = 8.03%) and $S3$ had an overall percent mismatch of 2.78% (SD = 2.61%). The prephonatory glottal diameter for both participants in this condition was set to 0 cm, which was the only time the glottis was set to be completely closed in the present study. This prephonatory glottal diameter (of 0 cm) would be consistent with the higher subglottal pressure, lower open quotient, higher skewing quotient, and lower MFND.

A major reason that the glottal diameter being set to 0 cm for this condition was to control the AC airflow produced by the model at the higher subglottal pressures. Despite this, AC airflow was the poorest matched parameter between the human and the model for both participants in this condition with the model output for both having higher AC airflows than the humans (by 20.83% for $S1$ and 6.81% for $S3$). To further reduce the AC airflow of the model and reduce the DC airflow from the normal pitch, normal loudness condition, the posterior glottal gap was narrowed for both participants. Modeling work by Zañartu *et al.* (2014) found a similar reduction in airflow based on reducing the posterior glottal gap but with less magnitude. It is possible that increasing the stiffness of the springs in the model would have further reduced the AC airflow when the subglottal pressure increased (Zhang, 2015).

For the loud condition, both participants decreased the open quotient from the normal loudness condition. This relationship was not unexpected as it was noted in Sulter and Wit (1996). However, if all the data (i.e., the best match cases for all participants and all conditions) for the present study were taken into consideration, as other conditions had higher subglottal pressure than the loud condition, the relationship between open quotient and loudness was not observed. Unlike Sulter and Wit (1996), the participants in the present study had a positive relationship between open quotient and fundamental frequency, which influenced efforts to match the former.

The normal pitch, soft condition was matched with an overall percent mismatch of 5.34% (SD = 5.37%) for $S1$ and 5.75% (SD = 10.06%) for $S3$. The prephonatory glottal diameter setting for both participants ($S1$ and $S3$) increased in the softer condition from the two normal pitch conditions

(normal and louder loudness) to better match the wider open quotients used in this condition compared to the other normal pitch conditions. Open quotient was also the greatest for the soft condition in Holmberg *et al.* (1988), thus, this increase in open quotient was not unexpected. For both participants, the model had a smaller open quotient than was observed in the human. However, AC airflow was matched within 1% for both, which explains why the glottal diameter was not increased further to increase the open quotient values produced by the model.

The reduction in fundamental frequency from the normal pitch, normal loudness condition in $S1$ (11.74 Hz, 2.01 ST) and $S3$ (13.94 Hz, 2.37 ST) is likely due to the pitch-countering reduction in subglottal pressure in both (3.68 cm H_2O in $S1$; 4.41 cm H_2O in $S3$). This seems to be especially the case in $S3$ for whom the best match involved a 30.00% increase in tension and a 35.95% increase in mass compared to the normal pitch, normal loudness condition. In $S1$, the decrease in fundamental frequency is supported by a 40.00% lower tension setting despite a 34.24% lower mass in the softer condition.

C. Matching of specific timing and airflow features

The model was able to match the AC airflow and DC airflow with acceptable accuracy. Airflow measures that varied more consistently with the prephonatory variables (e.g., open quotient generally increased as diameter increased) were more challenging to match. The airflow features with the poorest matches (highest average percent mismatch) between the model and the human were the open quotient, skewing quotient, and MFND.

The two worst matches for open quotient were in the higher pitch conditions for $S1$ and $S2$. In both cases, the model produced open quotients lower than the human open quotient. Across all conditions in the study, open quotient was positively related to the tension settings of the springs; as tension went up, the open quotient also increased. This pattern has also been observed in the canine larynx (Slavit *et al.*, 1990). Although the fundamental frequency was matched within 2% for $S1$ and the first match for $S2$ in this condition, increasing the tension (and reducing the mass) may have improved the matches for open quotient. Alternatively, the prephonatory glottal diameter could have been widened further to help increase the open quotient (Slavit *et al.*, 1990). In the alternative run for the higher pitch condition of $S2$, the diameter was increased (among other variables) relative to the normal pitch, normal loudness, and other higher pitch condition for this participant. The result was an open quotient produced by the model, which was higher than the human open quotient.

The skewing quotient has been related to changes in vocal tract inertance, the glottal area function, and potentially in subglottal pressure (Titze, 1992, 2004a, 2015). In the V3M model, the skewing quotient tended to decrease as the glottal diameter increased [Fig. 14(f)]. In the present study, the skewing quotient was well matched by the model

Brittany L. Perrine and Ronald C. Scherer

for $S1$ loud condition (0% difference) but presented with a 35.54% mismatch for $S2$'s higher pitch condition.

In all but one match, the higher pitch condition for $S3$, the MFND output of the model was less negative compared to the MFND of the human. MFND, as expected, decreased from the normal condition to the loud condition and increased from the normal to the soft condition (Holmberg *et al.*, 1988; Sapienza and Stathopoulos, 1994; Sulter and Wit, 1996). The higher MFNDs produced by the model likely represent a limitation in the model. It is possible that the epilaryngeal vocal tract inertance was not capitalized on to result in larger MFND negative values as observed in the human outputs (Titze, 2004b).

D. Physiological settings of the model

The masses of the vocal fold segments were varied to create the matches in the present study. Variations in the total vocal fold mass in motion occur as fundamental frequency changes (Titze, 2011). In this modeling work, the mass represents the amount of tissue in vibration with an average total mass of 0.18245 g (95% CI, 0.135508–0.229392). In their modeling work, Lucero and Koenig (2005a) used 0.125 and 0.025 g for the two (lower and upper, respectively) masses of their male vocal folds, based on I&F72. In the V3M model, the values also were loosely based on I&F72 and ranged from 0.11 to 0.05 g for the lower glottal mass and from 0.023 to 0.012 g for the upper glottal mass, depending on production conditions. In contrast, in their cover-body model, Story and Titze (1995) include two cover masses, each approximately 0.01 g, and one body mass of around 0.05 g. It is evident, then, that the amount of mass used in different models has varied, and based on the results of the current study, the amount of mass in motion may vary over a wide range depending on pitch, loudness, adduction, and quality.

The fundamental frequency in case A in Story and Titze (1995) is the most similar to that of the males used in the present study. In their case A, which is considered soft phonation produced with a low pitch, the stiffness settings for the cover and body are set to low levels. $S1$ in the soft phonation condition uses lower stiffness and mass settings compared to most of his other productions, like case A but with more stiffness and heavier mass settings. In contrast, the match for $S3$ in the soft phonation case involved higher tension and greater mass than all other conditions. Despite the differences in mass and tension, $S3$ and $S1$ have similar fundamental frequencies and AC airflow values in this soft condition. As has been found in other modeling work, there are numerous combinations of mass and stiffness settings that give similar and realistic fundamental frequencies (Lucero and Koenig, 2005a,b). Our modeling work also indicates that there are many potential prephonatory variable combinations that result in similar glottal airflow waveforms. This suggests that there may be numerous natural ways for an individual to produce a desired glottal airflow, some of which may be healthier than others.

Lucero and Koenig (2005a) also used a slightly modified I&F72 modeling approach (with only two vocal tract sections) with mass settings that were lower than the present study for their male and female configurations. Their work included mass and tension settings that are most similar to the higher pitch condition of the present study. In this condition, $S1$ and $S3$ have less mass than the male in Lucero and Koenig (2005a) while $S2$ has a very similar mass. Participants $S2$ and $S3$ have similar tension settings to Lucero and Koenig (2005a) for this condition. Specifically, the tension setting of the spring that connects the lower mass $m2$ to the "wall" is 70 000 dynes/cm for $S2$ and was 80 000 dynes/cm in the study by Lucero and Koenig. The coupling spring (between the lower two masses, $m1$ and $m2$, in this study) was also similar, with a setting of 25 000 dynes/cm in their study and 21 875 dynes/cm in the present work. The resulting fundamental frequency of around 180 Hz for $S2$ is higher than the fundamental frequency measured from the model by Lucero and Koenig, which was in the range of 100–150 Hz. A major difference between these two studies is that Lucero and Koenig (2005a) were studying the ability of a model to match human productions of /aha/. To see the wide range of tissue properties used in various computer models, refer to Alipour *et al.* (2011).

An interesting qualitative observation of the spectra for the human and modeled vowels is that of the ten spectra depicted in this study, the overall intensity (observed primarily as the height of the components in the first formant region) is similar between the human and modeled spectra for four (Figs. 7 and 9–11), the human spectra have greater intensity for five (Figs. 2, 3, 5, 6, and 12), and the human spectra has less intensity for one (Fig. 4). The MFND values for these ten conditions are relatively consistent with this observation. That is, when the difference between MFND values is more than 10%, the spectral intensities are observed to be different with the human spectra having higher intensities. For the four similar intensities, the MFND differences are 1%–11% different. When the human spectra have the observed higher intensity, the MFND is greater (negatively) for four out of the five spectra.

E. Limitations

The V3M model used in the present study involves three vertically stacked masses with intraglottal pressures from empirical studies using a physical model of the larynx (model $M5$) instead of Bernoulli-based or Navier-Stokes equations. This is a simplification of reality and a limitation of the present study, particularly because there was no three-dimensional geometry variation, such as a change in the length and shape of the vocal folds or a complex vocal tract. The cross-sectional area for the first vocal tract section for 2 of the 11 conditions was relatively large (9 cm^2) but within the range of 8–10 cm^2 reported in the literature. A simplification of reality is that all the spring tension settings were changed to the same degree. It may have been more realistic

J. Acoust. Soc. Am. **154** (3), September 2023

Brittany L. Perrine and Ronald C. Scherer   1523

to vary tissue tensions, as well as the amount of mass in motion, independently as conceived in vocal fold layered models. Also, because of the human interaction to determine the parameter values to best match the human glottal airflows, frequencies, and formants, optimal decisions may not have been achieved, whereas a widely generative process via computer iteration might result in more satisfactory matches. The present study is also limited because of the small sample size of humans used to create the matches and because only male participants were matched.

V. CONCLUSIONS

The goal of this study was to determine if a low dimensional phonatory model could mimic human laryngeal airflow and acoustics by manipulation of model variables. Recordings for three adult male participants were used. A modified Ishizaka and Flanagan two-mass model with electronic component vocal tract sections (six) was adapted for this purpose. A third (lower) vocal fold mass was added. This mass represents vocal fold tissue below the glottis proper for more complete vocal fold involvement and vertical movement. The intraglottal and transglottal pressures were obtained from empirical studies using the physical model $M5$. The manipulated variables were tissue mass and tension (damping was held constant), subglottal pressure, prephonatory glottal diameter, posterior glottal gap, and the false vocal fold gap. Dependent variables included fundamental frequency, AC and DC airflows, skewing quotient, MFND, open quotient, and the first three formant frequencies.

The strategies used in this study created matches to human phonation characteristics within 4.40% ($SD = 6.75\%$), indicating that the model generated aerodynamic and acoustic values similar to that of human phonation. In addition to presenting generally physiologically reasonable approximations of human phonation, the results of this study indicate that airflow measures can be approximated using a combination of prephonatory parameters.

ACKNOWLEDGMENTS

Agarwal, M. (**2004**). "The false vocal folds and their effect on translaryngeal airflow resistance," Ph.D. thesis, Bowling Green State University, Bowling Green, OH.

Agarwal, M., Scherer, R. C., and Witt, K. J. (**2004**). "The effects of the false vocal folds on translaryngeal airflow resistance," in *Proceedings of the International Conference on Voice Physiology and Biomechanics*, Marseille, France.

Alipour, F., Brücker, C., Cook, D. D., Gömmel, A., Kaltenbacher, M., Mattheus, W., Mongeau, L., Nauman, E., Schwarze, R., Tokuda, I., and Zörner, S. (**2011**). "Mathematical models and numerical schemes for the simulation of human phonation," Curr. Bioinform. **6**, 323–343.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (**1991**). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," J. Acoust. Soc. Am. **90**, 799–828.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot. Int. **5**, 341–345.

Cataldo, E., and Soize, C. (**2018**). "Stochastic mechanical model of vocal folds for producing jitter and for identifying pathologies through real voices," J. Biomech. **74**, 126–133.

Chang, A., and Karnell, M. P. (**2004**). "Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: A study of vocal fatigue," J. Voice **18**, 454–466.

Döllinger, M., Gómez, P., Patel, R. R., Alexiou, C., Bohr, C., and Schützenberger, A. (**2017**). "Biomechanical simulation of vocal fold dynamics in adults based on laryngeal high-speed videoendoscopy," PLoS One **12**, e0187486.

Fant, G. (**1960**). *The Acoustic Theory of Speech Production* (Mouton Co., The Hague, The Netherlands).

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Flanagan, J. L. (**1958**). "Some properties of the glottal sound source," J. Speech Hear. Res. **1**, 99–116.

Groll, M. D., McKenna, V. S., Hablani, S., and Stepp, C. E. (**2020**). "Formant-estimated vocal tract length and extrinsic laryngeal muscle activation during modulation of vocal effort in healthy speakers," J. Speech. Lang. Hear. Res. **63**, 1395–1403.

Gunter, H. E. (**2004**). "Modeling mechanical stresses as a factor in the etiology of benign vocal fold lesions," J. Biomech. **37**, 1119–1124.

Guo, C. G., and Scherer, R. C. (**1993**). "Finite element simulation of glottal flow and pressure," J. Acoust. Soc. Am. **94**, 688–700.

Hanson, H. M., Stevens, K. N., Kuo, H. K. J., Chen, M. Y., and Slifka, J. (**2001**). "Towards models of phonation," J. Phon. **29**, 451–480.

Hollien, H. (**2014**). "Vocal fold dynamics for frequency change," J. Voice **28**, 395–405.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am. **84**, 511–529.

Honda, T., Kanaya, M., Tokuda, I. T., Bouvet, A., Van Hirtum, A., and Pelorson, X. (**2022**). "Experimental study on the quasi-steady approximation of glottal flows," J. Acoust. Soc. Am. **151**, 3129–3139.

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

Ishizaka, K., and Isshiki, N. (**1976**). "Computer simulation of pathological vocal cord vibration," J. Acoust. Soc. Am. **60**, 1193–1198.

Kanduri, V. S., Emilian, J., and Jagadish, V. (**2021**). "Fatigue analysis of vocal-folds using discretized aeroelastic model," in *Trends in Mechanical and Biomedical Design*, Lecture Notes in Mechanical Engineering, edited by E. Akinlabi, P. Ramkumar, and M. Selvaraj (Springer, Singapore).

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

Löfqvist, A., Carlborg, B., and Kitzing, P. (**1982**). "Initial validation of an indirect measure of subglottal pressure during vowels," J. Acoust. Soc. Am. **72**, 633–635.

Lucero, J. C. (**1999**). "A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset," J. Acoust. Soc. Am. **105**, 423–431.

Lucero, J. C., and Koenig, L. L. (**2005a**). "Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control," J. Acoust. Soc. Am. **117**, 1362–1372.

Lucero, J. C., and Koenig, L. L. (**2005b**). "Phonation thresholds as a function of laryngeal size in a two-mass model of the vocal folds," J. Acoust. Soc. Am. **118**, 2798–2801.

Mainka, A., Poznyakovskiy, A., Platzek, I., Fleischerer, M., Sundberg, J., and Murbe, D. (**2015**). "Lower vocal tract morphologic adjustments are relevant for voice timbre in singing," PLoS One **10**, e0132241.

Mermelstein, P. (**1967**). "Determination of the vocal-tract shape from measured formant frequencies," J. Acoust. Soc. Am. **41**, 1283–1294.

Mongeau, L., Coker, C. H., and Kubli, R. A. (**1992**). "Experimental study of the aerodynamics of a larynx model," J. Acoust. Soc. Am. **92**, 2391.

Mongeau, L., Franchek, N., Coker, C. H., and Kubli, R. A. (**1997**). "Characteristics of a pulsating jet through a small modulated

1524   J. Acoust. Soc. Am. **154** (3), September 2023

Brittany L. Perrine and Ronald C. Scherer

orifice, with application to voice production," J. Acoust. Soc. Am. **102**, 1121–1134.

Nishizawa, N. (**1988**). "Vocal fold length in vocal pitch change," in *Physiology: Voice Production, Mechanisms, and Function*, edited by O. Fujimura (Raven, New York), pp. 75–82.

Oren, L., Dembinski, D., Gutmark, E., and Khosla, S. (**2014**). "Characterization of the vocal fold vertical stiffness in a canine model," J. Voice **28**, 297–304.

Pelorson, X., Hirschberg, A., van Hassel, R., Wijnands, A., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Perrine, B., Scherer, R. C., and Whitfield, J. A. (**2019**). "Signal interpretation considerations when estimating subglottal pressure from oral air pressure," J. Speech Lang. Hear. Res. **62**, 1326–1337.

Perrine, B. L., Scherer, R. C., Fulcher, L. P., and Zhai, G. (**2020**). "Phonation threshold pressure using a 3-mass model of phonation with empirical pressure values," J. Acoust. Soc. Am. **147**, 1727–1737.

Rothenberg, M. (**1973**). "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," J. Acoust. Soc. Am. **53**, 1632–1645.

Sadeghi, H., Kniesburges, S., Kaltenbacher, M., Schützenberger, A., and Döllinger, M. (**2019**). "Computational models of laryngeal aerodynamics: Potentials and numerical costs," J. Voice **33**, 385–400.

Sapienza, C. M., and Stathopoulos, E. T. (**1994**). "Comparison of maximum flow declination rate: Children versus adults," J. Voice **8**, 240–247.

Scherer, R. C. (**2010**). "Static physical modeling of laryngeal aerodynamics," in *Ninth International Conference: Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*, September 10–11, Erlangen, Germany.

Scherer, R. C., Frazer, B., and Zhai, G. (**2013**). "Modeling flow through the posterior glottal gap," Proc. Mtgs. Acoust. **19**, 060240.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2002**). "Intraglottal pressure distributions for a symmetric and oblique glottis with a uniform duct (L)," J. Acoust. Soc. Am. **112**, 1253–1256.

Scherer, R. C., Zhai, G., Fulcher, L., and Agarwal, M. (**2004**). "A vertical three-mass model of phonation based on empirical intraglottal pressures," in *Proceedings of the International Conference on Voice Physiology and Biomechanics*, August 18–20, Marseille, France.

Slavit, D. H., Lipton, R. J., and McCaffrey, T. V. (**1990**). "Phonatory vocal fold function in the excised canine larynx," Otolaryngol.-Head. Neck Surg.**103**, 947–956.

Stepp, C. E., Hillman, R. E., and Heaton, J. T. (**2010**). "A virtual trajectory model predicts differences in vocal fold kinematics in individuals with vocal hyperfunction," J. Acoust. Soc. Am. **127**, 3166–3176.

Story, B., and Titze, I. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.

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

Sulter, A. M., and Wit, H. P. (**1996**). "Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age," J. Acoust. Soc. Am. **100**, 3360–3373.

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

Titze, I. R. (**1989**). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906.

Titze, I. R. (**1992**). "Phonation threshold pressure: A missing link in glottal aerodynamics," J. Acoust. Soc. Am. **91**, 2926–2935.

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

Titze, I. R. (**2004b**). "A theoretical study of $F0$-$F1$ interaction with application to resonant speaking and singing voice," J. Voice **18**, 292–298.

Titze, I. R. (**2006**). *The Myoelastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Iowa City, IA).

Titze, I. R. (**2011**). "Vocal fold mass is not a useful quantity for describing $F0$ in vocalization," J. Speech Hear. Res. **54**, 520–522.

Titze, I. R. (**2015**). "Sensitivity of odd-harmonic amplitudes to open quotient and skewing quotient in glottal airflow," J. Acoust. Soc. Am. **137**, 502–504.

Titze, I. R., and Story, B. (**2002**). "Rules for controlling low-dimensional vocal fold models with muscle activation," J. Acoust. Soc. Am. **112**, 1064–1076.

Tokuda, I., Horáček, J., Švec, J., and Herzel, H. (**2007**). "Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments," J. Acoust. Soc. Am. **122**, 519–531.

van den Berg, J. W., Zantema, J. T., and Doornenbal, P., Jr. (**1957**). "On the air resistance and the Bernoulli effect of the human larynx," J. Acoust. Soc. Am. **29**, 626–631.

Vilain, C. E., Pelorson, X., Fraysse, C., Deverge, M., Hirschberg, A., and Willems, J. (**2004**). "Experimental validation of a quasi-steady theory for the flow through the glottis," J. Sound Vib. **276**, 475–490.

Wang, W., Ziang, W., Zheng, X., and Xue, Q. (**2021b**). "A computational study of the effects of vocal fold stiffness parameters on voice production," J. Voice **35**, 327.e1–327.e11.

Wang, X., Palaparthi, A., Titze, I. R., Zheng, X., and Xue, Q. (**2021a**). "Testing the validity of the quasi-steady assumption in vocal fold vibration," in *Proceedings of the 14th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL)*, June 7–10, Bogota, Colombia, p. 61.

Yang, C.-S., and Kasuya, H. (**1994**). "Accurate measurement of vocal tract shapes from magnetic resonance images of child, female, and male subjects," Proc. ICSLP **94**, 623–626.

Zañartu, M., Galindo, G. E., Erath, B. D., Peterson, S. D., Wodicka, G. R., and Hillman, R. E. (**2014**). "Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction," J. Acoust. Soc. Am. **136**, 3262–3271.

Zañartu, M., Mehta, D. D., Ho, J. C., Wodicka, G. R., and Hillman, R. E. (**2011**). "Observation and analysis of *in vivo* vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study," J. Acoust. Soc. Am. **129**, 326–339.

Zhang, Z. (**2009**). "Characteristics of phonation onset in a two-layer vocal fold model," J. Acoust. Soc. Am. **125**, 1091–1102.

Zhang, Z. (**2015**). "Regulation of glottal closure and airflow in a three-dimensional phonation model: Implications for vocal intensity control," J. Acoust. Soc. Am. **137**, 898–910.

Zhang, Z. (**2017**). "Effect of vocal fold stiffness on voice production in a three-dimensional body-cover phonation model," J. Acoust. Soc. Am. **142**, 2311–2321.

Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002**). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," J. Acoust. Soc. Am. **112**, 1652–1663.

J. Acoust. Soc. Am. **154** (3), September 2023

Brittany L. Perrine and Ronald C. Scherer   1525