INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Accelerometer-Based Measurements of Voice Quality in Children during Semi-Occluded Vocal Tract Exercise with a Narrow Straw in Air

Steven M. Lulich, Rita R. Patel

Department of Speech, Language and Hearing Sciences, Indiana University, U.S.A.
slulich@iu.edu, patelrir@iu.edu

Abstract

Non-invasive measures of voice quality, such as H1-H2, rely on oral flow signals, inverse filtered speech signals, or corrections for the effects of formants. Voice quality measures play especially important roles in the assessment of voice disorders and the evaluation of treatment efficacy. One type of treatment that is increasingly common in voice therapy, as well as in voice training for singers and actors, is semi-occluded vocal tract exercises (SOVTEs). The goal of SOVTEs is to change patterns of vocal fold vibration and thereby improve voice quality and vocal efficiency. Accelerometers applied to the skin of the neck have been used to investigate subglottal acoustics, to inverse-filter speech signals, and to obtain voice quality metrics. This paper explores the application of neck-skin accelerometers to measure voice quality without oral flow, inverse filtering, or formant correction. Accelerometer-based measures (uncorrected K1-K2 and corrected K1*-K2*, analogous to microphone-based H1-H2 and H1*-H2*) were obtained from typically developing children with healthy voice, before and during SOVTEs. Traditional microphone-based H1-H2 measures (corrected and uncorrected) were also obtained. Results showed that K1-K2 and K1*-K2* were not substantially affected by vocal tract acoustic changes in formant frequencies.

Index Terms: voice production, semi-occluded vocal tract, voice quality, voice disorders, child voice

1. Introduction

The prevalence of dysphonia in school-age children ranges from 0.12% [1] to 15.80% [2]. Clinical evaluation of voice quality involves a comprehensive battery of tests including invasive methods of laryngeal endoscopy which can challenging in the pediatric population [3]. Objective assessments of non-invasive methods of evaluating vibratory kinematics in children are extremely limited and important for timely assessment of voice disorders and evaluation of treatment efficacy. SOVTEs are widely used in voice therapy for rehabilitation and habilitation of dysphonia and for training vocal performers in the pediatric population [4].

The goal of SOVTEs with small diameter tubes is to change the vocal fold vibrations where the top edges of the vocal folds are slightly abducted thereby reducing vocal fold collision and vibratory amplitude, increasing the open quotient, and thus improving voice quality and vocal efficiency [5, 6]. Reasonable evidence exists in adults that the open quotient of glottal vibration is correlated with the amplitude difference between the first and second harmonics of microphone-based speech signals, after correcting for the effects of vocal tract resonances (H1*-H2*) [7, 8]. Mehta et al (2012) [9] first reported the measurement of H1-H2 from an accelerometer placed on the skin of the neck, in one vocally healthy male participant, based on the impedance-based inverse-filtering method of [10]. In 79 vocally healthy adult females Mehta et al (2019) [11] reported a significant high correlation (r = 0.72) between the H1-H2 derived from the accelerometer signal and that derived from the glottal airflow signal, thus providing empirical evidence that H1-H2 derived from the accelerometer signal reflects properties of vocal fold vibrations such as the glottal open quotient. Lulich and Patel (2021) [12] reported microphone- and accelerometer-based H1-H2 measures for adults during SOVTEs, and found that the accelerometer-based measures were not impacted by the semi-occlusion at threshold phonation, and that observed changes in microphone-based measures were strongly associated with estimated changes in F1. Together, previous studies suggest that reliable accelerometer-based measures of voice quality can be obtained from adults, perhaps without the need for inverse filtering or formant correction.

Since the anatomy and physiology of the vocal folds [13, 14] and the vocal tract [15] are different between children and adults, it is important to examine the characteristics of accelerometer-based measures in children. Anatomical differences between children and adults may affect the nonlinear source-filter interaction and impact the outcomes of SOVTEs in children [16].

In the current study, we differentiate between microphone- and accelerometer-based measures of harmonic amplitudes by referring to the accelerometer-based measures as K1 and K2, while reserving the traditional notation H1 and H2 for the microphone-based measures. Asterisks are used to indicate correction for the effects of formants (e.g. H1*) and subglottal resonances (e.g. K1*). Figure 1 illustrates the measurement of uncorrected H1, H2, K1, K2, the first formant frequency (F1), and the first subglottal resonance frequency (Sg1). This study aims to quantify H1-H2, H1*-H2*, K1-K2, and K1*-K2* before and during SOVTEs in children.

2. Method

2.1. Participants

Participant data from our prior study [4] were used for this current study. Our prior study [4] focused on evaluating the most salient acoustic (fundamental frequency, first & second format frequency, first subglottal resonance, peak-to-peak amplitude ratio, and various tongue and pharyngeal ultrasound measurements during SOVTEs in children (6-9 years). A total of 21 vocally normal, typically developing children were recruited in the Department of Speech, Language, and Hearing Sciences at Indiana University after signing IRB approved informed

http://dx.doi.org/10.21437/Interspeech.2021-1918

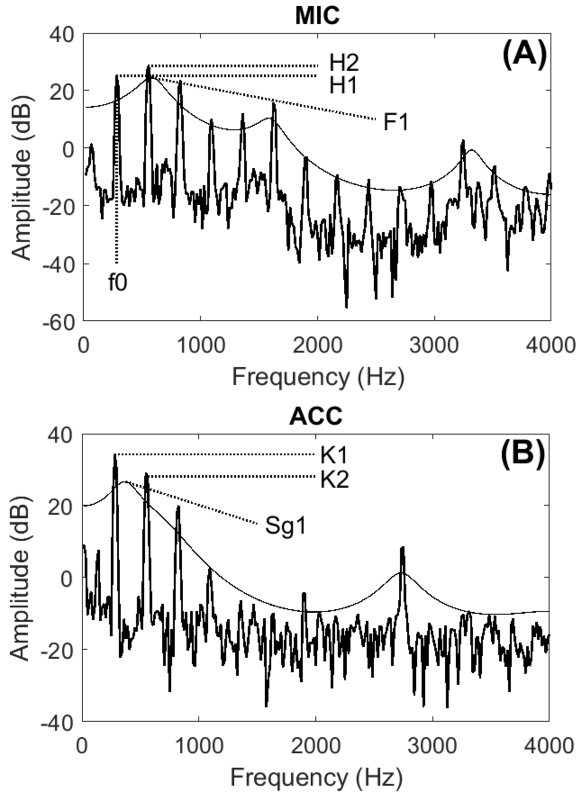

Figure 1: *Illustration of measurements H1, H2, K1, K2, F1, Sg1 (first subglottal resonance), and f0 (fundamental frequency) from microphone (A) and accelerometer (B) signals.*

consent/assent forms. Data from three participants were excluded from analysis because of medical issues (n = 1) and errors in the recording protocol (n = 2). The participants in this study included 18 children, 8 boys (8.5 years±11 months) and 10 girls (8.4 years±12 months).

2.2. Experimental protocol

Simultaneous recordings of 3D/4D ultrasound, acoustic, and accelerometer were conducted with a Philips EPIQ 7G ultrasound system and xMatrix x6-1 digital 3D/4D transducer, a Shure KSM32 microphone, and a K&K Sound HotSpot accelerometer, respectively in a double-walled sound treated booth. The cardioid condenser microphone was placed at a fixed distance of 60cms in front of the participant on a stand at an angle of 45 degrees to avoid distortions due to high oral airflow. The accelerometer was secured against the skin of the next at the level of the cricoid cartilage below the thyroid notch [4, 17]. The sampling rate was 48kHz and the quantization was 16 bits/sample for the microphone and the accelerometer. The participants were asked to produce the following tasks at self-selected vocal frequency and sound levels, in order to minimize the effects of increased effort during targeted productions of vocal frequency and sound level [18]: (1) three sustained productions of the vowel /o:/ and (2) three sustained productions of the vowel /o:/ phonated into a straw of 2.53mm inner diameter and 124mm length, while wearing a nose clip to ensure no leakage of air through the nose [19]. Each vowel production was sustained for 5 seconds. Prior to recording with the small diameter straw, practice trials with a large diameter straw (5.18mm inner diameter and 196.85 mm length) followed by the narrow diameter straw were conducted by a fellowship-trained speech language pathologist in the area of voice.

2.3. Data Analysis

Synchronized measurements from 108 microphone and 108 accelerometer recordings were analyzed from the steady-state portion of the vowel. Corresponding fundamental frequency, formant frequency, and subglottal resonance frequency measurements were previously reported [4]. Data analysis was conducted by a trained undergraduate student majoring in Speech Language Pathology using WAVESURFER (version 1.8.8p4) [20]. The H1 and H2 were measured directly from the discrete Fourier transform (DFT) line spectrum using a hamming window and 512 points after down-sampling to 8000Hz to obtain the H1-H2 index as a correlate of the glottal open quotient (Figure 1A). A second index of glottal open quotient K1-K2 was obtained from analogous measurements of the first and second harmonic amplitudes in the accelerometer signal (Figure 1B).

Both H1-H2 and K1-K2 were corrected based on formant and subglottal resonance frequencies, yielding measures of H1*-H2* and K1*-K2*. The bandwidth (B) corresponding to both the first formant (F1) and the first subglottal resonance (Sg1) was assumed to be 100 Hz, and the second harmonic was assumed to lie at a frequency twice that of the fundamental frequency (f0). Three measurements of fundamental frequency, F1, and Sg1 were not available due to a clerical error, and there were therefore 105 measurements of H1*-H2* and K1*-K2*, rather than 108. The correction for H1* was accomplished using the following formula [21]:

$$H1^* = H1 - 20 \cdot log_{10}\left(\frac{F1^2+(B/2)^2}{\sqrt{(f_0-F1)^2+(B/2)^2}\cdot\sqrt{(f_0+F1)^2+(B/2)^2}}\right)$$

For H2*, the value 2f0 was substituted for f0. Correction for K1* and K2* used the same formula, but with the frequency of the first subglottal resonance, Sg1, substituted for F1.

2.3.1. *Reliability of Measurements*

A total of 18 trials (16.7%) were pseudo-randomly selected for intra-rater reliability for the microphone and accelerometer measurement. The Pearson correlation coefficient was greater than .99 for these measures, and the mean absolute difference (MAD) was .0125 dB. Similar intra-rater reliability was found in a companion study of SOVTEs in adults [21], in which inter-rater reliability was also high (for H1 and K1: r>=.94, MAD<=.86 dB; for H2 and K2: r > .8, MAD<=3.08 dB). As reported in [4], intra-rater and interrater reliability were high for f0 (r >=.9, MAD<=6.64 Hz) and F1 (r>=.9, MAD<=43.18 Hz), and moderate for Sg1 (.5>=r>=.38, MAD<=275.36).

2.3.2. *Statistical Analyses*

Shapiro-Wilks tests were used to determine the normality of the outcome variables H1-H2, H1*-H2*, K1-K2, and K1*-K2*. H1-H2 was non-normally distributed (*p*<.005), while the remaining three variables were normally distributed (*p*>.1). Because one of the four outcome variables was non-normal, further inferential statistical analyses were carried out using non-parametric Kruskal-Wallis tests, in addition to descriptive statistics and the Pearson correlation between the normally-distributed variables H1*-H2* and K1-K2. Main effects of gender (male vs. female) and condition (without straw vs. with straw) were examined for each of the 4 outcome variables. Statistical significance was determined after Bonferroni correction, with $\alpha=.05/8 = .0125$.

3. Results

Kruskal-Wallis tests revealed no significant main effect of gender ($p>.48$), which is not surprising given the prepubescent ages of the participants. As previously reported, the first formant frequency (F1) was significantly dependent on condition [4], and consequently there was a significant main effect of condition on H1-H2 ($p \ll .0001$), with a nearly 21 dB mean increase during the SOVTE. (There was no effect of condition on f0 [4]). There was also a significant main effect of condition on the corrected H1*-H2* ($p<.0002$), with a 4 dB mean decrease during the SOVTE, indicating either a failure of the correction to adequately account for the change in formant frequency (e.g. due to inaccurate formant measurements or poor choice of assumed formant bandwidth) a change in the voice quality of the participants, or a Type I error.

As previously reported, the first subglottal resonance frequency was significantly dependent on condition [4], though with a much smaller effect size than for F1. There was a significant main effect of condition on the uncorrected K1-K2 ($p<.0004$), with a nearly 5 dB mean increase during the SOVTE. There was no main effect of condition on the corrected K1*-K2* ($p>.02$), with no more than a 2.5 dB mean increase during the SOVTE.

Descriptive statistics characterizing the variables in both conditions (with straw, without straw) are presented in Table 1, while results of inferential statistical tests are presented in Table 2. Box plots illustrate the variables' distributions in Figure 2. Additionally, the Spearman correlation coefficient was calculated to determine the strength of the relationship between H1*-H2* and K1-K2, analogous to the relation probed by Mehta et al (2019) [11]. Unlike in Mehta et al (2019) [11], who found a strong correlation (r=.72) among 79 adult women phonating at three loudness levels, the correlation between H1*-H2* and K1-K2 among the children phonating at one loudness level in the present study was poor (r=.20).

4. Discussion

As previously reported for the same data set analyzed in the present study, the first formant and the first subglottal resonance frequencies were statistically significantly different in the without- and with-straw conditions, while f0 was not [4]. The present study expands on Patel et al (2019) [4] by examining uncorrected and corrected voice quality measures related to open quotient. Two of these measures (H1-H2, H1*-H2*) were obtained from the microphone signal, and two (K1-K2, K1*-K2*) were obtained from the accelerometer signal.

Although accelerometer-based measures of open quotient have been reported before [9-11], this is the first study to examine such measures in children, or in semi-occluded vocal tract exercises. Mehta et al (2019) [11] reported a strong correlation between inverse-filtered glottal airflow-based measures of open quotient (analogous, though not equivalent, to H1*-H2*) and raw accelerometer signal-based measures (equivalent to K1-K2) in adults. In contrast, the correlation between H1*-H2* and K1-K2 observed in this study in children was poor. The difference in observed correlation strength could be due to different analysis methods (e.g. the method of obtaining corrected or inverse-filtered H1*-H2* values), different ranges of loudness (3 levels in Mehta et al (2019) [11], 1 level in the present study), different populations (adults vs. children), or different tasks (spoken syllables vs. sustained vowels and SOVTEs). Further research on this topic is necessary to determine the effects and important of these factors.

K1*-K2* revealed no statistically significant dependence on condition (without straw vs. with straw), and K1-K2 revealed a small, though statistically significant dependence on condition, suggesting a possible increase in open quotient during the SOVTE exercises. In contrast, H1-H2 was strongly dependent on condition, while H1*-H2*, like K1-K2, was weakly dependent on condition. Together, these findings suggest 1) an increase in the open quotient during SOVTE, and 2) K1-K2 and K1*-K2* are both adequate measures of open quotient. Both measures have an advantage over traditional H1-H2 and H1*-H2* measures, in that they are a function of the acoustic properties of the subglottal airways, which do not vary substantially over time [4, 22-24]. The uncorrected measure K1-K2 has an additional advantage over K1*-K2* in that it does not require inverse filtering of any kind, and it therefore does not depend upon the accuracy of formant or subglottal resonance frequency estimation. On the other hand, use of uncorrected K1-K2 may result in a degradation of accuracy relative to K1*-K2*, as evidenced by the fact that the dependence of K1-K2 on condition was still statistically significant, although small. Thus, although the present results contrast with those of Mehta et al (2019) [11] with regard to the strength of the correlation between H1*-H2* and K1-K2, the present study supports the conclusions of Mehta et al (2019) [11] with regard to the appropriateness and utility of raw accelerometer-based measures of open quotient.

Table 1: *Descriptive statistics of [min, mean, max] values for each of the harmonic amplitude differences (in decibels).*

Variable	w/o straw	w/ straw
H1-H2	[-15.8, -1.7, 10.4]	[5.9, 18.9, 30.8]
H1*-H2*	[-3.6, 5.5, 15.7]	[-10.8, 1.5, 13.4]
K1-K2	[-9.0, 3.8, 22.5]	[-12.8, 8.5, 28.1]
K1*-K2*	[-5.3, 10.1, 28.7]	[-9.0, 12.5, 29.9]

Table 2: *Inferential statistics. P-values for significant results are highlighted.*

Variable	Factor	N	df (error)	χ^2	*p*
H1-H2	Gender	108	106	.33	.57
H1*-H2*	Gender	105	103	.46	.50
K1-K2	Gender	108	106	.48	.49
K1*-K2*	Gender	105	103	.37	.54
H1-H2	Condition	108	106	79.0	≪.001
H1*-H2*	Condition	105	103	14.6	≪.001
K1-K2	Condition	108	106	12.6	≪.001
K1*-K2*	Condition	105	103	5.32	.021

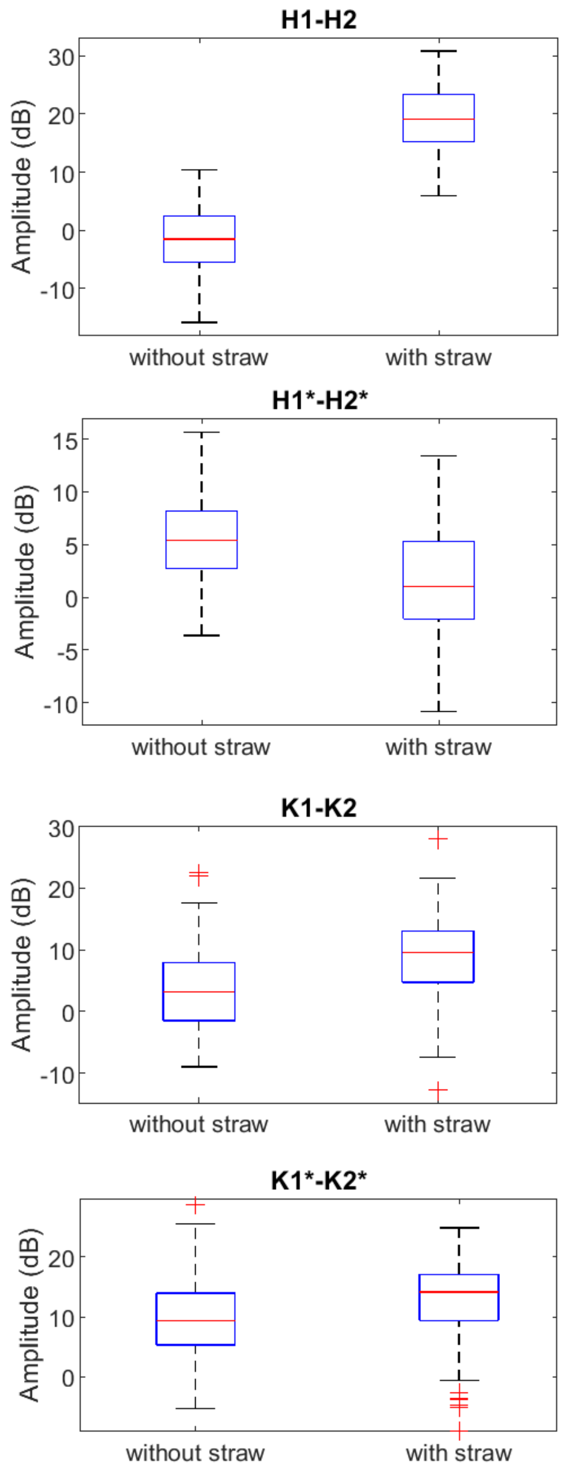

Figure 2: *Box plots of uncorrected (H1-H2, K1-K2) and corrected (H1*-H2*, K1*-K2*) measurements, separated by condition.*

Thus, the accelerometer-based measures of K1-K2 and K1*-K2* could be potentially used for objective, non-invasive clinical voice assessment of vibratory kinematics in children after further investigations.

A significant limitation of the present study is the lack of oral airflow data and direct imaging of vocal fold vibration. Based on the accelerometer-based measures of open quotient (K1-K2 and K1*-K2*), the findings suggest a possible main effect of condition on open quotient. We suggest that high-speed endoscopic video should be collected simultaneously with microphone- and accelerometer-based measures of voice quality to validate or qualify our conclusions. Another limitation is the restriction to one particular set of voice quality measures, namely those traditionally related to open quotient. Future studies should examine additional measures, such as spectral tilt [25] or Maximum Flow Declination Rate (MFDR) [9], obtained from accelerometer signals applied to children and to SOVTEs.

5. Conclusions

This paper explored the application of neck-skin accelerometers to measures of voice quality without oral flow, inverse filtering, or formant correction. It was found that K1-K2 is adequate as a measure of open quotient, although K1*-K2* corrected for the first subglottal resonance is probably even more accurate. Findings suggests a possible increase in the open quotient during SOVTE in children and the potential usefulness of K1-K2 and K1*-K2* for objective, non-invasive evaluation of treatment efficacy of the SOVTE exercises in children; however, the accuracy of this needs to be confirmed with direct measurement of open quotient with simultaneous high-speed videoendoscopy and accelerometer-based measurements.

6. Acknowledgements

The authors would like to thank Alessandra Verdi, who helped with the data collection and measurement of f0, F1, and Sg1 reported by Patel et al (2019) [4], and Alyssa Arick, who helped make the measurements of H1, H2, K1, and K2 for this study. We also thank the three anonymous reviewers for their feedback. This work was supported by an Undergraduate Research Grant from the Indiana University, Department of Speech, Language and Hearing Sciences.

7. References

[1] D. H. McKinnon, S. McLeod, and S. Reilly, "The prevalence of stuttering, voice, and speech-sound disorders in primary school students in Australia," *Lang Speech Hear Serv Sch,* vol. 38, pp. 5-15, Jan 2007.

[2] E. Kallvik, E. Lindström, S. Holmqvist, J. Lindman, and S. Simberg, "Prevalence of Hoarseness in School-aged Children," *J Voice,* vol. 29, pp. 260.e1-260.e19, Jul 10 2014.

[3] C. J. Hartnick and S. M. Zeitels, "Pediatric video laryngo-stroboscopy," *Int J Pediatr Otorhinolaryngol,* vol. 69, pp. 215-9, Feb 2005.

[4] R. Patel, S. M. Lulich, and A. Verdi, "Vocal tract shape and acoustic adjustments of children during phonation into narrow flow-resistant tubes," *J Acoust Soc Am,* vol. 146, pp. 352-68, Jul 2019.

[5] V. Radolf, A. M. Laukkanen, J. Horáček, and D. Liu, "Air-pressure, vocal fold vibration and acoustic characteristics of phonation during vocal exercising. Part 1: Measurement in vivo," *Eng. Mech,* vol. 21, pp. 53-59, 2014.

[6] I. Titze, "Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings," *J Speech Lang Hear Res,* vol. 49, pp. 448-59, Apr 2006.

[7] H. M. Hanson, "Glottal characteristics of female speakers: acoustic correlates," *J Acoust Soc Am,* vol. 101, pp. 466-81, Jan 1997.

[8] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. C. Guiod, and S. L. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *J Speech Hear Res,* vol. 38, pp. 1212-23, Dec 1995.

[9] D. D. Mehta, M. Zanartu, S. W. Feng, H. A. Cheyne, 2nd, and R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," *IEEE Trans Biomed Eng,* vol. 59, pp. 3090-6, Nov 2012.

[10] M. Zañartu, J. C. Ho, D. D. Mehta, R. E. Hillman, and G. R. Wodicka, "Subglottal Impedance-Based Inverse Filtering of Voiced Sounds Using Neck Surface Acceleration," *IEEE Trans Audio Speech Lang Process,* vol. 21, pp. 1929-1939, Sep 2013.

[11] D. D. Mehta, V. M. Espinoza, J. H. Van Stan, M. Zanartu, and R. E. Hillman, "The difference between first and second harmonic amplitudes correlates between glottal airflow and neck-surface accelerometer signals during phonation," *J Acoust Soc Am,* vol. 145, pp. EL386, May 2019.

[12] S. M. Lulich and R. Patel, "Semi-Occluded Vocal Tract Exercises in Healthy Young Adults: Articulatory, Acoustic, and Aerodynamic Measurements During Phonation at Threshold," *J Acoust Soc Am, vol. 149, pp. 3213-27,May, 2021*.

[13] M. E. Boseley and C. J. Hartnick, "Development of the human true vocal fold: depth of cell layers and quantifying cell types within the lamina propria," *Ann Otol Rhinol Laryngol,* vol. 115, pp. 784-8, Oct 2006.

[14] M. Hirano, S. Kurita, and T. Nakashima, *Growth, Development, and Aging of Human Vocal Fold*: College-Hill Press, San Diego, California, 1983.

[15] H. Vorperian, S. Wang, M. K. Chung, R. B. Durtschi, E. M. Schimek, L. R. Gentry*, et al.*, "Relational growth of oral and pharyngeal structures with vocal-tract length during the first two decades of life as visualized from MRI and CT studies," ed: University of Wisconsin Madison, 2008.

[16] B. H. Story and K. Bunton, "Formant measurement in children's speech based on spectral filtering," *Speech Commun,* vol. 76, pp. 93-111, 2015.

[17] L. Wade, N. Hanna, J. Smith, and J. Wolfe, "The role of vocal tract and subglottal resonances in producing vocal instabilities," *J Acoust Soc Am,* vol. 141, pp. 1546, Mar 2017.

[18] D. G. Hanson, B. R. Gerratt, and G. S. Berke, "Frequency, intensity, and target matching effects on photoglottographic measures of open quotient and speed quotient," *J Speech Hear Res,* vol. 33, pp. 45-50, Mar 1990.

[19] I. Titze, "Phonation threshold pressure measurement with a semi-occluded vocal tract," *J Speech Lang Hear Res,* vol. 52, pp. 1062-72, Aug 2009.

[20] K. Sjölander and J. Beskow, "Wavesurfer-an open source speech tool," in *Proceedings of International Conference on Spoken Language Processing*, 2000, pp. 464-467.

[21] G. Fant, Ed., *Acoustic theory of speech production*. The Hague, Netherlands: Mouton, 1960, pp. 1-328.

[22] S. M. Lulich and H. Arsikere, "Tracheo-bronchial soft tissue and cartilage resonances in the subglottal acoustic input impedance," *J Acoust Soc Am,* vol. 137, pp. 3436, Jun 2015.

[23] S. M. Lulich, J. R. Morton, H. Arsikere, M. S. Sommers, G. K. Leung, and A. Alwan, "Subglottal resonances of adult male and female native speakers of American English," *J Acoust Soc Am,* vol. 132, pp. 2592-602, Oct 2012.

[24] G. Yeung, S. M. Lulich, J. Guo, M. S. Sommers, and A. Alwan, "Subglottal resonances of American English speaking children," *Journal of the Acoustical Society of America,* vol. 144, pp. 3437-3449, in press.

[25] M. Iseli, Y. L. Shue, and A. Alwan, "Age, sex, and vowel dependencies of acoustic measures related to the voice source," *J Acoust Soc Am,* vol. 121, pp. 2283-95, Apr 2007.