

Subglottal resonances of American English speaking children

Gary Yeung, Steven M. Lulich, Jinxi Guo, Mitchell S. Sommers, and Abeer Alwan

Citation: *The Journal of the Acoustical Society of America* **144**, 3437 (2018); doi: 10.1121/1.5082289

View online: <https://doi.org/10.1121/1.5082289>

View Table of Contents: <https://asa.scitation.org/toc/jas/144/6>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[A comparison of electroglottographic and glottal area waveforms for phonation type differentiation in male professional singers](#)

The Journal of the Acoustical Society of America **144**, 3275 (2018); <https://doi.org/10.1121/1.5066456>

[An acoustic camera approach to studying nasality in speech: The case of Polish nasalized vowels](#)

The Journal of the Acoustical Society of America **144**, 3603 (2018); <https://doi.org/10.1121/1.5084038>

[Effects of listener age and native language on perception of accented and unaccented sentences](#)

The Journal of the Acoustical Society of America **144**, 3191 (2018); <https://doi.org/10.1121/1.5081711>

[Manner and place differences in Kannada coronal consonants: Articulatory and acoustic results](#)

The Journal of the Acoustical Society of America **144**, 3221 (2018); <https://doi.org/10.1121/1.5081686>

[Speech: A Dynamic Process](#)

The Journal of the Acoustical Society of America **144**, 3210 (2018); <https://doi.org/10.1121/1.5082302>

[Putting Laurel and Yanny in context](#)

The Journal of the Acoustical Society of America **144**, EL503 (2018); <https://doi.org/10.1121/1.5070144>



CAPTURE WHAT'S POSSIBLE
WITH OUR NEW PUBLISHING ACADEMY RESOURCES

Learn more 



Subglottal resonances of American English speaking children

Gary Yeung^{a)}

Department of Electrical and Computer Engineering, University of California, Los Angeles, California 90095, USA

Steven M. Lulich

Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405, USA

Jinxi Guo

Department of Electrical and Computer Engineering, University of California, Los Angeles, California 90095, USA

Mitchell S. Sommers

Department of Psychology, Washington University, St. Louis, Missouri 63130, USA

Abeer Alwan

Department of Electrical and Computer Engineering, University of California, Los Angeles, California 90095, USA

(Received 3 October 2017; revised 5 October 2018; accepted 16 November 2018; published online 21 December 2018)

This paper presents an investigation of children's subglottal resonances (SGRs), the natural frequencies of the tracheo-bronchial acoustic system. A total of 43 children (31 male, 12 female) aged between 6 and 18 yr were recorded. Both microphone signals of various consonant-vowel-consonant words and subglottal accelerometer signals of the sustained vowel /a/ were recorded for each of the children, along with age and standing height. The first three SGRs of each child were measured from the sustained vowel subglottal accelerometer signals. A model relating SGRs to standing height was developed based on the quarter-wavelength resonator model, previously developed for adult SGRs and heights. Based on difficulties in predicting the higher SGR values for the younger children, the model of the third SGR was refined to account for frequency-dependent acoustic lengths of the tracheo-bronchial system. This updated model more accurately estimates both adult and child SGRs based on their heights. These results indicate the importance of considering frequency-dependent acoustic lengths of the subglottal system.

© 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5082289>

[LK]

Pages: 3437–3449

I. INTRODUCTION

The subglottal airways, consisting of the trachea, bronchi, and lungs, serve as the main source of airflow that activates the larynx and vocal tract during speech. Previous research has shown that the adult subglottal system and its acoustic properties play an important role in speech acoustics (Arsikere *et al.*, 2013; Chi and Sonderegger, 2007; Cranen and Boves, 1987; Ishizaka *et al.*, 1976; Klatt and Klatt, 1990; Lulich, 2010; Lulich *et al.*, 2011a; Lulich *et al.*, 2011b; Lulich *et al.*, 2012; Zhang *et al.*, 2006). However, only a few of these studies have focused on the measurement and analysis of subglottal acoustics in children (Lulich, 2010; Lulich *et al.*, 2011b).

A more detailed understanding of children's subglottal acoustics may assist in a number of applications. In medical fields, further research into children's subglottal acoustics can complement existing studies on children's tracheo-bronchial anatomy, which may prove useful to applications such as pediatric anaesthesiology and pulmonology (Henry and Royston, 2017, 2018a,b; Kim *et al.*, 2013; Mahajan

et al., 2007). Additionally, studies on adult subjects have shown that the subglottal system physically influences the speech acoustic system through coupling with the vocal tract (Fant, 1960; Lulich, 2013; Lulich *et al.*, 2009; Stevens, 1998; Titze, 2006, 2008; Titze *et al.*, 2008; Zañartu *et al.*, 2011), resulting in acoustically unstable frequency regions that speakers may avoid to produce perceptually salient phonological contrasts (Chi and Sonderegger, 2007; Csapó *et al.*, 2009a; Csapó *et al.*, 2009b; Dogil *et al.*, 2011; Grácsi *et al.*, 2011; Jung, 2009; Lulich, 2010; Lulich *et al.*, 2007; Madsack *et al.*, 2008; Stevens, 1998). A better understanding of children's subglottal acoustics can provide insight into the development of children's speech, which can benefit the field of speech-language pathology. Furthermore, previous studies revealed that subglottal acoustics provide valuable information when applied to both automatic speech and speaker recognition systems for adults (Guo *et al.*, 2016; Wang *et al.*, 2009a; Wang *et al.*, 2009b). For children's speech recognition, the use of subglottal acoustics has been shown to help combat the mismatch of systems trained on adult speech data and applied to children (Arsikere *et al.*, 2012a; Guo *et al.*, 2015). A more complete knowledge of children's subglottal acoustics is expected to benefit the development of such technologies.

^{a)}Electronic mail: garyyeung@g.ucla.edu

Subglottal resonances (SGRs) are the natural frequencies of the subglottal system. When observed in the subglottal acoustic input impedance, SGRs manifest themselves as pairs of complex-conjugate poles. As the measurement of SGRs requires considerable effort through the recording of accelerometer signals or other more invasive techniques, several studies have attempted to model SGRs using speech parameters that are easily measured from microphone signals. Wang *et al.* (2009b) used the fact that the second SGR contributed a zero in the microphone signal within the range of the second formant. This was used to estimate the location of the second SGR for adults by identifying where a discontinuity occurred in the second formant's trajectory. Arsikere *et al.* (2013) used the differences between the first three formants (in bark scale) to estimate the first three SGRs for adults. Similarly, both Lulich *et al.* (2011b) and Guo *et al.* (2015) used the differences between the first three formants to estimate the first three SGRs for children. However, the difficulty of estimating formants for speakers with high fundamental frequencies (Makhoul, 1975; Monsen and Engbretonson, 1983) is a major limitation for these models.

An alternative approach to modeling SGRs was taken by Lulich *et al.* (2011a) and Lulich *et al.* (2012) where speaker height was used as the main variable instead of speech parameters. In those studies, speaker height was shown to be the single most contributing factor in the estimation of SGRs for adults. The first three SGRs were predicted using a single quarter-wavelength resonator tube model with the tube length being dependent on only height. The tube was intended to approximate the acoustic length of the tracheo-bronchial tree (as defined by Lulich *et al.*, 2011a), roughly half of which is comprised of the trachea. The result was consistent with Griscom and Wohl (1985, 1986), which showed a strong correlation between trachea length and height and no correlation between trachea length and gender when the effect of height was eliminated. In a recent study by Hanna *et al.* (2018), SGRs estimated using this tube model were found to be in agreement with SGRs obtained from input impedance measurements performed at the lips. However, a consequence of using this tube model is that the ratios between the first three SGRs are assumed to be constant across all speakers, which was a major implication of Lulich *et al.* (2012). This implication was never formally investigated by that study.

Lulich *et al.* (2011a) and Lulich *et al.* (2012) made two main observations when constructing the SGR tube model for adults. These observations were embedded into the model as parameters to be estimated using the measured SGR and height data. The first observation was that the acoustic length of the tracheo-bronchial tree of a speaker was directly proportional to the height of that speaker and could be modeled by a single scaling factor, denoted as k_a . The second observation was that the wave propagation velocity of the first SGR differed from the speed of sound due to the inertive properties of the subglottal system tissues. This velocity was denoted as c_w . In both studies, the parameters were estimated using the minimum root-mean-square (RMS) error criterion on adult SGR data measured from subglottal accelerometer recordings. The scope

of these tube model studies has not been extended to include children's data.

The range of children's SGRs is a major cause for concern when applying the tube model to children's data. Notably, children, especially younger and shorter children, are known to have higher SGRs due to having a shorter tracheo-bronchial system compared to adults. In Fredberg and Moore (1978), high frequencies were found to penetrate deeper into the tracheo-bronchial system than low frequencies. Similar frequency-dependent penetrations are well documented in the acoustics of horns (Benade, 1990; Pyle, 1975), particularly relevant due to the similarities between horns and the subglottal system (Jackson *et al.*, 1978; Van den Berg, 1960). This frequency-dependent penetration depth is analogous to the effective acoustic length of the subglottal system or tube length of the quarter-wavelength resonator. As the model proposed by Lulich *et al.* (2011a) has a fixed tube length for each speaker, the model is incapable of addressing the frequency-dependent effects that may be necessary in the modeling of higher frequencies.

This study presents a model of the SGRs in children and adolescents, extending the previous studies by investigating frequency-dependent effects when the tube model is applied to children's SGRs. In this paper, we present SGR data from 43 children between the ages of 6 and 18 yr. Speech acoustics were recorded using a microphone, and subglottal acoustics were recorded using a non-invasive technique with an accelerometer placed on the skin of the subject's neck under the glottis. Relationships between the first three SGRs and height were investigated, and the simple tube model of Lulich *et al.* (2011a) and Lulich *et al.* (2012) was refined to better reflect the physics of sound propagation in the subglottal airways and account for the children's SGR data without compromising the estimation of adult SGRs.

The remainder of the paper is organized as follows. Section II explains the data collection, labeling, and measurement procedures. Section III presents the data analysis and modeling results. Finally, Sec. IV concludes the work with a brief summary and describes directions for future work.

II. METHODS

A. Participants

A total of 43 native speakers of American English (31 males, 12 females) between the ages of 6 and 18 yr were recorded for this study. The children were recruited through the Washington University in St. Louis (WashU) psychology department subject pool, as well as through advertisements and flyers posted in public spaces around the St. Louis, MO, area. Parents of the children were asked if their children had any history of speech or hearing disorders, and none were reported. Each speaker's standing height, age, and gender were recorded. All participants were recorded on the WashU campus.

While we aimed for a balanced representation, the recruitment effort returned fewer responses than we had hoped for. Notably, while the 31 males had reasonable representation across the 6–18 yr age range, the 12 females were all under the age of 14 yr. Since male and female subglottal

airways do not begin to display sexual dimorphisms until approximately 14 yr old (Griscom and Wohl, 1986), it is reasonable to conclude that the data presented here are representative of all children up to approximately the age of 14 yr, but only for males above that age.

B. Recordings

To capture the speech and subglottal acoustics of the participants, recordings were made using a free-standing SHURE PG27 microphone (Shure, Niles, IL) and a K&K Sound HotSpot accelerometer (K&K Sound Systems, Coos Bay, OR) while participants sat in a double-walled sound attenuating booth. The accelerometer is advertised to have a flat magnitude response for up to 15 kHz, and Wade *et al.* (2017) verified that the magnitude response was at least flat in the range of 350–2000 Hz. Any deviation from the flat response of the accelerometer would only affect the magnitude of a resonance but not the frequency. Both microphone and accelerometer signals were sampled at a sampling rate of 48 kHz and quantized at 16 bits per sample. Recordings were made using a two-channel M-Audio MobilePre USB pre-amplifier (M-Audio, Cumberland, RI) connected to a computer running Windows Vista (Microsoft, Redmond, WA) and were recorded via MATLAB (MathWorks, 2016). The microphone was placed approximately 20 cm in front of the speaker and slightly to the side to avoid distortion due to high airflow sounds (e.g., the fricative /s/ or the plosive /p/). Each speaker was instructed to hold the accelerometer firmly against the skin of his or her neck at the cricoid cartilage just below the glottis. This placement helped prevent formants from interfering significantly with the accelerometer signal, which is common when formants and SGRs are located near each other (Chi and Sonderegger, 2007). The recording setup was identical to the adult recording procedure in Lulich *et al.* (2012).

A number of consonant-vowel-consonant (CVC) words were embedded in the carrier phrase “I said a CVC again” and were either displayed to the speaker to be read aloud or read to the speaker to be repeated. In most cases, a computer monitor was placed directly in front of the speaker, and the sentences to be read were displayed on the screen. However, some of the younger speakers had not yet learned to read. For these speakers, one of the researchers sat in the sound booth with the child, and the “repeat after me” approach was taken. The complete list of all CVC words recorded (hVd, bVb, dVb, gVb) is identical to the words recorded for adults in Lulich *et al.* (2012). For each of the CVC words, both microphone and accelerometer signals were recorded, and each child had at least six repetitions of each word. An example wideband spectrogram of the accelerometer signal of a 13 yr old male saying “hod again” is shown in Fig. 1. Notably, the subglottal acoustics remain relatively stationary during phonation, which is characteristic of the subglottal system (Arsikere *et al.*, 2013; Lulich *et al.*, 2011a; Lulich *et al.*, 2012).

The accelerometer signal quality for the CVC words was variable. To account for this, additional subglottal accelerometer recordings of each child sustaining the vowel /a/ were recorded with an emphasis on high-quality signals. The quality of the signal was optimized by allowing the subject and

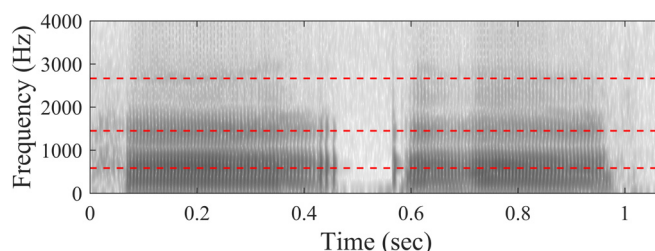


FIG. 1. (Color online) A wideband spectrogram of the accelerometer signal of a 13 yr old male saying “hod again.” A 2048 length discrete Fourier transform (DFT), window length of 6 ms, and frame shift of 1 ms were used. The dashed lines show the means of the speaker’s first three SGRs. Notably, the subglottal acoustics remain relatively stationary across different phonemes.

experimenter to interact while visually inspecting the spectrogram of the produced signal. Before each recording, a live spectrogram of the accelerometer signal was shown to the experimenter. This display was used as an indicator of how to further improve the quality of the accelerometer recording in real time. The position of the accelerometer, loudness of the speaker, and fundamental frequency of the speaker were all adjusted until the highest quality spectrogram was achieved. The recording of the accelerometer signal followed immediately afterward. This accelerometer recording procedure of the sustained vowel /a/ was repeated twice for each participant. Since the accelerometer recordings of the CVC words were of variable quality, this paper will only use the subglottal accelerometer recordings of the sustained vowel /a/.

C. Microphone analysis

The fundamental frequency (f_0) and first three formants (F1, F2, F3) from the vowels /a/, /æ/, /ʌ/, /ɛ/, /ɪ/, /i/, /u/, and /ʊ/ embedded in their respective hVd words were measured using the recorded microphone signals. The measurements and measurement details are included in Appendix A for comparability to past studies. In general, these measurements were similar to previous measurements of f_0 , F1, F2, and F3 for children reported in Lee *et al.* (1999). Additionally, since the relationship between formants and SGRs from the perspective of quantal theory (Stevens, 1972, 1989, 1998) has attracted some attention in the past (Chi and Sonderegger, 2007; Csapó *et al.*, 2009a; Csapó *et al.*, 2009b; Dogil *et al.*, 2011; Grácsi *et al.*, 2011; Jung, 2009; Lulich, 2010; Lulich *et al.*, 2007; Madsack *et al.*, 2008), we present a brief analysis of the quantal behavior of children in Appendix C.

D. Accelerometer analysis

The high-quality subglottal accelerometer signals from the sustained vowel /a/ were used to measure the first three SGRs (Sg1, Sg2, Sg3). Each token was downsampled to 8 kHz before measurement as the first three SGRs were all less than 4 kHz. Downsampling to 8 kHz also allowed for easier and more focused visual inspection of the accelerometer signals. Each accelerometer token was inspected manually with a visual analysis of the discrete Fourier transform (DFT) and linear predictive coding (LPC) spectral envelope, with order varied from 6 to 24, at the steady-state portions of the sustained vowel using WAVESURFER

(Sjölander and Beskow, 2000). If the SGRs could be observed visually as a peak in the LPC envelope after varying the order of the LPC polynomial, the SGRs were measured manually. Solutions to additional measurement challenges are presented and discussed in Appendix B.

III. RESULTS AND DISCUSSION

A. SGR measurement analysis

Table I shows the mean and range of the SGR measurements, height rounded to the nearest inch (converted to cm), and age rounded to the nearest month for each participant. The ranges of Sg1, Sg2, and Sg3 across all participants were approximately 500–900 Hz, 1250–2200 Hz, and 1950–3300 Hz, respectively. For most cases with two reliable SGR measurements from different files for a single participant, the difference between the measurements did not exceed 100 Hz. The results are also consistent with the claim that shorter individuals have higher SGRs (Lulich *et al.*, 2011a; Lulich *et al.*, 2012), and the participants in Table I have been sorted by height to display this result.

The Pearson’s correlation coefficients between the various SGRs were computed using all data from Table I. Similar to the results for adults in Lulich *et al.* (2012), the correlations of Sg1 with Sg2 ($r=0.773$) and Sg1 with Sg3 ($r=0.758$) were strong, but not as strong as the correlation of Sg2 with Sg3 ($r=0.929$). While these correlations are stronger than those of adults reported in Lulich *et al.* (2012), this may be due to the fact that the adult data were measured from the noisier hVd signals while the children’s measurements in this paper are based on the cleaner sustained /a/ signals. The weaker correlations involving Sg1 are likely due to how tissue inertance interacts with the lower frequency values of Sg1 (Lulich and Arsikere, 2015; Lulich *et al.*, 2012).

B. Quarter-wavelength resonance model of SGRs and height

In large-scale investigations of adult SGRs, Lulich *et al.* (2011a) and Lulich *et al.* (2012) verified that adult SGRs were well-modeled by a single quarter-wavelength resonator (tube with one open end) as follows:

TABLE I. Mean Sg1, Sg2, and Sg3 (Hz) values from two accelerometer signals of the sustained vowel /a/, organized by gender and sorted by height (cm). The difference (in Hz) between the two values is in parentheses if more than one value was successfully measured. Also listed are the speaker identification (ID) number and age (years;months).

Females						Males					
ID	Age	Height	Sg1	Sg2	Sg3	ID	Age	Height	Sg1	Sg2	Sg3
137	7;6	127	803(34)	2031(27)	3106(43)	106	7;1	107	719(19)	2202(36)	3308(17)
105	10;8	130	703(44)	2036(12)	3156(—)	128	6;1	114	802(—)	1888(—)	3063(—)
124	8;5	130	696(—)	1793(—)	2745(—)	110	6;10	117	771(59)	1980(23)	3097(129)
146	6;11	130	756(11)	1991(56)	2916(—)	122	7;7	119	886(3)	2053(112)	3207(3)
119	9;10	135	797(19)	1802(4)	2825(—)	136	7;6	123	737(37)	1968(42)	2914(33)
107	10;10	137	651(6)	1960(54)	2894(174)	134	11;3	127	944(75)	1997(40)	3103(29)
126	13;3	147	697(9)	1646(47)	2599(94)	145	8;7	127	653(40)	1642(9)	2670(56)
118	10;11	152	662(38)	1669(31)	2533(35)	115	8;11	130	778(36)	1830(24)	2865(135)
113	12;4	154	610(36)	1570(81)	2280(—)	108	8;2	130	573(14)	1787(101)	2778(58)
121	11;11	155	728(12)	1750(47)	2712(77)	114	8;11	132	654(—)	1607(—)	2704(—)
140	12;4	158	686(—)	1594(48)	2655(39)	139	8;3	133	808(66)	1786(47)	2725(38)
130	13;6	169	638(1)	1506(93)	2225(47)	116	8;7	135	750(74)	1877(47)	2873(32)
						138	7;7	135	776(—)	1761(53)	2939(24)
						143	9;8	138	842(7)	2050(57)	2866(136)
						112	10;7	140	736(2)	1677(11)	2761(22)
						102	11;5	143	550(42)	1547(87)	2345(—)
						132	9;9	146	815(—)	1810(—)	2718(—)
						123	10;3	147	587(31)	1712(29)	2611(103)
						135	11;1	155	709(13)	1743(30)	2702(19)
						144	12;0	156	798(78)	1759(—)	2598(—)
						120	14;4	160	695(68)	1620(1)	2549(73)
						127	13;3	163	652(19)	1481(40)	2642(62)
						109	13;1	165	583(28)	1441(2)	2474(79)
						141	16;3	166	526(50)	1383(4)	2177(47)
						117	13;4	168	555(34)	1454(38)	2494(94)
						133	13;7	169	710(11)	1489(25)	2374(53)
						111	15;3	170	589(6)	1393(6)	2351(101)
						125	13;7	173	598(52)	1551(55)	2504(165)
						129	16;0	173	467(23)	1270(27)	2206(34)
						103	16;3	174	533(58)	1462(4)	2387(71)
						104	17;8	182	502(5)	1270(8)	1988(46)

$$\text{Sg}N = \frac{(2N - 1)c_N}{4h/k_a}, \quad (1)$$

where $N \in \{1,2,3\}$, c_N is the propagation velocity of the wave for the N th SGR, h is the height of the speaker, and k_a is a scaling factor relating the acoustic length of the subglottal system (l_a) to speaker height ($l_a = h/k_a$). It was noted that in adults, while the wave propagation velocity for Sg2 and Sg3 was approximately the speed of sound at body temperature, $c_0 = 35900$ cm/s, the wave propagation velocity for Sg1 was larger than c_0 due to the inertive properties of the subglottal system tissue walls in the Sg1 frequency range (Lulich *et al.*, 2011a; Lulich and Arsikere, 2015). The value of c_1 was denoted as c_w for the “walls” of the subglottal system.

In Lulich *et al.* (2011a) and Lulich *et al.* (2012), the two main modeling parameters to be estimated were k_a and c_w . These parameters were estimated using Eq. (1), adult SGR and height data, and the minimum RMS error criterion. The value of k_a was first estimated from adult Sg2, Sg3, and height data, along with the assumption that $c_2 = c_3 = c_0$. Then, c_w was estimated using Sg1 data and the resulting value of k_a . For the remainder of this paper, we will also use the minimum RMS criterion to estimate model parameters.

To analyze the effectiveness of Eq. (1) for modeling children’s SGRs with height, we estimated c_w and k_a using children’s data. Sg2 and Sg3 were assumed to have a wave propagation velocity of c_0 . As with the adults, the values of k_a for all SGRs and c_w for Sg1 were chosen to minimize the RMS error between the actual and estimated values of the SGRs. An initial estimate was derived using all the data from Table I. A second estimate was derived using the data from 12 children (6–18 yr old; 5 male, 7 female) reported in Lulich *et al.* (2011b) (not including B1 and G1 from that paper as those speakers were noted to have lower quality accelerometer recordings), henceforth referred to as dataset “C,” in addition to the data from Table I. The children’s SGR values in dataset C were measured using a combination of LPC, autocorrelation-based smooth spectra, and visual inspections of the DFT spectra of the accelerometer signals of the vowels of various hVd words. To examine the model’s ability to simultaneously model the SGRs of both adults and children, a third estimate was derived using the adult SGR data from 50 adults (18–25 yr old;

25 male, 25 female) reported in Lulich *et al.* (2012), henceforth referred to as dataset “A,” as well as the children’s SGR data from the 12 children in dataset C and data from Table I. The adult SGR values in dataset A were measured with a visual inspection of the DFT, LPC, and wideband power spectral density (WPSD) of the accelerometer signals of the vowels of various hVd words.

The top three rows of Table II display these parameter estimates, along with the RMS estimation errors for each of the three datasets used. The modeling curves for all three estimates, along with all SGR vs height data, are shown in Fig. 2. The resulting models are similar to the model reported in Lulich *et al.* (2012). The quarter-wavelength resonator approximation seems to fit the relationship between Sg1 and height well. However, based on a qualitative inspection of the data in Fig. 2, it is clear that the model systematically underestimates Sg2 based on height, especially for the children’s data. Additionally, the model seems to overestimate Sg3 for some of the shorter children.

When studying SGR estimation algorithms based on formants for children, Guo *et al.* (2015) noted that better results were produced when different algorithms were used for children under and above the age of 11 yr. To investigate the possibility of age-dependent interactions between SGRs and height, the minimum RMS values of c_w and k_a were re-estimated six times using the data from Table I with children under 10, 11, and 12 yr old and children above 10, 11, and 12 yr old. The bottom six rows of Table II display these parameter estimates, along with the RMS estimation errors for the three datasets mentioned earlier. In all cases, computed values of c_w differed from the value computed using the data in Table I by less than 1.7%. Similarly, computed values of k_a differed from the value computed using the data in Table I by less than 3.1%. There was no clear dependence on age for c_w . This could suggest that the inertive properties of the airway wall tissues described by c_w are independent of age (and height). Notably, the use of the older children’s data raised the value of k_a closer to the adult values reported in both Lulich *et al.* (2012) and Arsikere *et al.* (2012b) while the use of the younger children’s data lowered the value of k_a . This suggests that the relationship between a child’s height and the acoustic length of the subglottal airways is

TABLE II. Minimum RMS estimates of k_a and c_w for Eq. (1) using a combination of various SGR datasets, along with RMS errors of the used datasets for each of the SGRs. The first two rows are computed from combinations of the data from Table I, dataset A, and dataset C. The remaining datasets are subsets of Table I, restricted by age.

Data	c_w (cm/s)	k_a	TABLE I ϵ_{RMS} (Hz)			A ϵ_{RMS} (Hz)			C ϵ_{RMS} (Hz)		
			Sg1	Sg2	Sg3	Sg1	Sg2	Sg3	Sg1	Sg2	Sg3
TABLE I + A + C	45 400	8.760	88	136	190	49	75	108	73	138	160
TABLE I + C	44 648	8.735	87	138	187	55	77	107	69	141	162
TABLE I	45 325	8.656	87	146	179	53	84	109	70	154	172
Under 10 years	45 581	8.411	89	175	175	60	112	137	67	195	220
Under 11 years	44 872	8.511	89	162	172	62	100	123	67	177	198
Under 12 years	45 235	8.556	88	157	173	57	95	117	68	170	189
Above 10 years	45 060	8.920	89	126	216	48	68	118	76	115	153
Above 11 years	46 072	8.910	93	126	215	46	68	117	84	116	153
Above 12 years	45 546	8.922	91	126	217	46	68	118	80	115	153

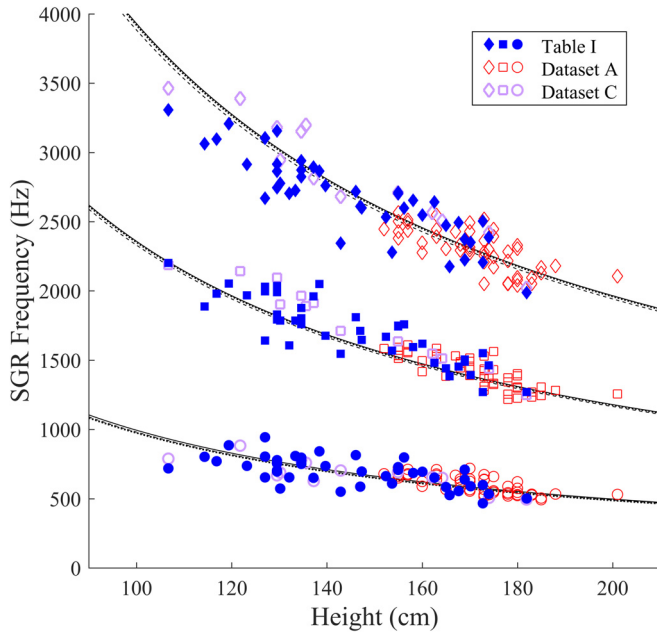


FIG. 2. (Color online) Sg1 (○), Sg2 (□), and Sg3 (◇) vs height for all data from Table I, dataset A, and dataset C. The modeling curves shown are estimated using Eq. (1) with all data shown in this figure (solid), only data from Table I and dataset C (dotted), and only data from Table I (dashed).

nonlinear, with a disproportionately greater acoustic length in early childhood and a decelerating rate of growth of the acoustic length relative to height.

The tube model of Eq. (1) suggests a fixed ratio relationship between each of the SGRs, with the inertive properties of the subglottal system in the Sg1 frequency range maintained across the range of Sg1 values. Ideally, the SGR ratios Sg2/Sg1, Sg3/Sg1, and Sg3/Sg2 should be $3c_0/c_w$, $5c_0/c_w$, and $5/3$, respectively. Figure 3 shows the SGR data along with solid lines through the origin with slopes $3c_0/c_w$, $5c_0/c_w$, and $5/3$ for Sg2 vs Sg1, Sg3 vs Sg1, and Sg3 vs Sg2, respectively, using the value of $c_w = 45\,400$ cm/s estimated using all collected data. The lines using the other estimates of c_w

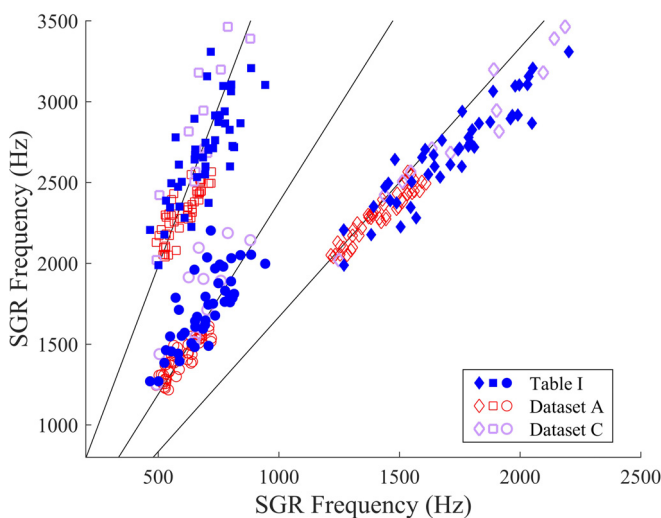


FIG. 3. (Color online) Sg2 vs Sg1 (○), Sg3 vs Sg1 (□), and Sg3 vs Sg2 (◇) for all data from Table I, dataset A, and dataset C. The lines shown are the linear ratio approximations according to Eq. (1) and the value of c_w estimated using all datasets.

are not plotted as they are almost identical to the ones shown. While the ratio approximates Sg2 vs Sg1 and Sg3 vs Sg1 reasonably well, the data for Sg3 vs Sg2 are below the ratio approximation of $5/3$ for higher SGR values. The mean value of Sg3/Sg2 for the children's data from Table I is 1.57. Lulich *et al.* (2012) also reported that the mean value of Sg3/Sg2 for adults was 1.62, which is slightly below the expected value of $5/3$ or 1.67. This result was interpreted as evidence of a constant $5/3$ ratio for Sg3 vs Sg2, suggesting that the subglottal system behaves like an equivalent uniform tube with yielding walls and frequency-independent length. In light of the children's data presented here, the low value of Sg3/Sg2 is better interpreted as evidence that the acoustic length of the uniform tube has a more complex interaction with shorter speakers and higher SGRs.

Instead, these results suggest that the rate of increase in Sg3 with respect to the decrease in height should become smaller relative to the rate of increase in Sg2. A possible solution is found in Fredberg and Moore (1978), who found that the penetration depth of sound into the tracheo-bronchial tree increased as frequency increased. That is, the acoustic resonances of the subglottal system penetrate further into the subglottal system (bronchi and lungs) in children than in adults for the higher frequency SGRs. This would imply that the effective tube length of the subglottal system increases for each successive SGR.

C. Height-adjusted resonance model

Section III B and Figs. 2 and 3 revealed that Eq. (1) does not properly model the highest Sg3 values. We propose a refined tube-resonance model for Sg3, in which the effective acoustic length of the subglottal system is modified for the modeling of Sg3, thus incorporating the penetration depth findings of Fredberg and Moore (1978). To this extent, l_a is introduced as the acoustic length of the subglottal system for Sg1 and Sg2, previously defined as $l_a = h/k_a$. To maintain consistency with Eq. (1), the refined tube model replaces l_a with $l_{a,Sg3}$, which represents the acoustic length for the Sg3 model. Specifically, as height decreases, and Sg3 correspondingly increases, the effective acoustic length of the tube model, $l_{a,Sg3}$, should decrease slower than l_a as higher frequencies penetrate further into the subglottal system than lower frequencies. This is implemented by defining the acoustic length of Sg3 as a function of l_a . The model is implemented as follows:

$$Sg3 = \frac{5c_0}{4l_{a,Sg3}(l_a)}, \quad (2)$$

$$l_{a,Sg3}(l_a) = l_a + f(l_a), \quad (3)$$

where the effective acoustic length for Sg3 can be derived from the acoustic length used to model Sg1 and Sg2 and some additive function representing the additional penetration depth of the higher frequencies.

The function $f(l_a)$ must follow certain assumptions within reasonable values of l_a . First, $f(l_a)$ should always be positive. Second, $f(l_a)$ should always be much smaller than l_a . Third, $f(l_a)$ should approach 0 as l_a becomes large.

TABLE III. Minimum RMS estimates of k_a , c_w , α , and β using Eq. (1) for Sg1 and Sg2 and Eqs. (2)–(4) for Sg3, computed from combinations of various SGR datasets, along with RMS errors of the used datasets for each of the SGRs. The datasets used are the data from Table I, dataset A, and dataset C.

Data	c_w (cm/s)	k_a	α (1/cm)	β	TABLE I ϵ_{RMS} (Hz)			A ϵ_{RMS} (Hz)			C ϵ_{RMS} (Hz)		
					Sg1	Sg2	Sg3	Sg1	Sg2	Sg3	Sg1	Sg2	Sg3
TABLE I + A + C	43 849	9.070	0.235	0.805	88	122	147	49	68	107	73	98	186
TABLE I + C	42 735	9.126	0.298	1.704	87	123	147	55	71	116	69	93	179
Table I	43 251	9.071	0.326	2.245	87	122	143	53	68	111	70	98	205

Finally, the derivative of $f(l_a)$ should always be greater than -1 to ensure that $l_{a,\text{Sg3}}(l_a)$ is monotonically increasing. Our proposed candidate function is the following:

$$f(l_a) = \frac{l_a}{1 + e^{(\alpha l_a - \beta)}}, \quad (4)$$

where the additional parameters α and β can be computed by finding the minimum RMS values over the Sg3 data. Equation (4) can be thought of as l_a multiplied by a reversed logistic function. As such, $f(l_a)$ can be controlled to be a small percentage of l_a . For reasonable values of α and β , Eq. (4) can satisfy all four requirements.

To solve for a complete SGR model using the new model for Sg3, the values of k_a , c_w , α , and β were derived incrementally in order to obtain values properly adjusted for each of the three SGR ranges. First, the minimum RMS value of k_a was found using Eq. (1) and c_0 over the Sg2 data. Next, the minimum RMS value of c_w was found using Eq. (1) and k_a over the Sg1 data. Finally, the minimum RMS values of α and β were found using Eqs. (2)–(4), k_a , and c_0 over the Sg3 data.

Three sets of parameters were derived with the incremental derivation stated above using the same combinations of SGR data as in the first three rows of Table II used in

Sec. III B. The resulting parameters and RMS errors for all three estimates are shown in Table III. Comparing the RMS errors of Sg2 in Tables II and III shows that the new modeling procedure provides a better estimate of Sg2 for all three datasets. The RMS error of Sg3 for dataset A decreases only slightly, suggesting that the original tube model was suitable for estimating Sg3 for adults. However, the RMS error of Sg3 for the data in Table I decreases substantially from 190 Hz to 147 Hz when using the new model (and all datasets for estimation), indicating that the modified model is more suitable for estimating Sg3 for children (and shorter individuals). The RMS error of Sg3 for dataset C increases slightly when using the new model. However, this may be due to the fact that these SGR measurements used accelerometer signals with hVd words, which are likely noisier than the sustained vowel signals used in this study.

Figure 4 displays the new modeling curves along with all data used. The modeling curves for Sg2 and Sg3 fit the data better than the curves in Fig. 2. Figure 5 displays the new ratio curves along with all Sg2 vs Sg1, Sg3 vs Sg1, and Sg3 vs Sg2 data. The estimated Sg3 vs Sg2 curves from the new model fit the Sg3 vs Sg2 data better than the 5/3 ratio shown in Fig. 3, especially at lower heights. In general, it seems that the model parameters estimated using both adult and children’s SGR data had the best fit. This model appears suitable to model human SGRs for subjects as short as 100 cm to subjects as tall as 210 cm.

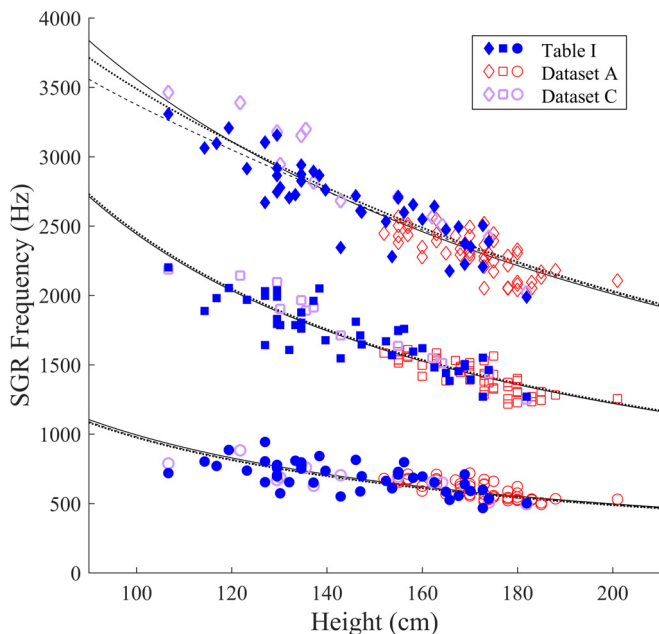


FIG. 4. (Color online) Sg1 (○), Sg2 (□), and Sg3 (◇) vs height for all data from Table I, dataset A, and dataset C. The modeling curves shown are estimated using Eq. (1) for Sg1 and Sg2 and Eqs. (2)–(4) for Sg3 with all data shown in this figure (solid), only data from Table I and dataset C (dotted), and only data from Table I (dashed).

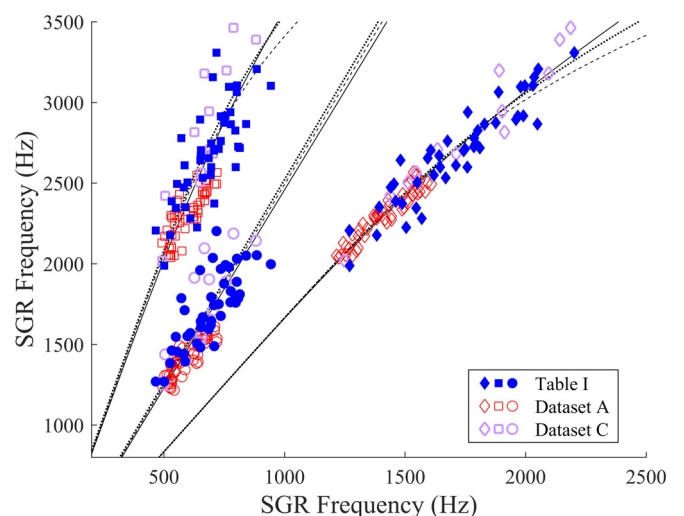


FIG. 5. (Color online) Sg2 vs Sg1 (○), Sg3 vs Sg1 (□), and Sg3 vs Sg2 (◇) for all data from Table I, dataset A, and dataset C. The curves shown are the ratio approximations according to Eq. (1) for Sg1 and Sg2 and Eqs. (2)–(4) for Sg3, estimated using all data shown in this figure (solid), only data from Table I and dataset C (dotted), and only data from Table I (dashed).

IV. CONCLUSION

This paper makes two main contributions to the science of children’s speech and subglottal acoustics. First, it reports a comprehensive description of the measurement and analysis of children’s SGRs across a wide range of ages with a larger and more representative dataset than previously investigated. SGRs were measured non-invasively with an accelerometer placed on the neck under the glottis. We investigated how the SGRs relate to each other as well as to other pertinent characteristics of children (e.g. height). These relationships were largely similar to those found for adults (Lulich *et al.*, 2012) with the notable exception that the frequency ratio $Sg3/Sg2$ decreased as height decreased.

This finding prompted the second contribution, which was a refined empirical model of subglottal acoustics. Lulich *et al.* (2011a) found that SGR frequencies as a function of height could be modeled using a simple uniform tube as in Eq. (1), with a correction applied to $Sg1$ to account for the inertive effects of the subglottal tissues on wave propagation velocity. In this paper, the findings from Fredberg and Moore (1978)—namely, that wave penetration depth in the subglottal airways increases with frequency—were used to refine the $Sg3$ model by including a height-related correction to the tube’s effective acoustic length, modeled as Eqs. (2)–(4). This correction was made so that the model can accurately estimate SGRs of both children and adults from height measurements. The refined SGR model demonstrates the importance of considering factors such as wave penetration depth in vocal and respiratory tract acoustic modeling.

The height-related correction was empirically derived and modeled to account for the observed decrease in $Sg3/Sg2$ ratio. However, derivation of a correction from the physical principles of horn acoustics (Benade, 1990; Pyle, 1975) may provide additional insight into the subglottal acoustics of both children and adults. This should be addressed in future research.

Several additional directions can be considered for future work and applications. Although the number of children involved in this study was large compared with previous studies (Arsikere *et al.*, 2012a; Lulich, 2010; Lulich *et al.*, 2011b), sample sizes were still only moderate (especially for females) in comparison with large microphone-only studies such as Lee *et al.* (1999). Because children have high SGRs and fundamental frequencies, measurement of SGRs was more difficult than in adults. Furthermore, the accelerometer signals were not of as uniformly high quality as similar data from adults (Lulich *et al.*, 2012), leading to the decision to limit measurements to sustained /a/ productions for the present study. In the future, measurements from additional children in a range of phonetic contexts will be important for determining how children’s SGRs change as they grow, how SGRs interact with speech acoustics, and how children use these interactions to produce intelligible speech.

Previous studies of SGRs among both adults and children have shown that automatic speaker normalization based on SGRs is an effective and computationally efficient way to improve automatic speech recognition systems (Arsikere

et al., 2012a; Guo *et al.*, 2015; Wang *et al.*, 2009a; Wang *et al.*, 2009b). The data presented in this paper should be useful for developing more accurate and empirically grounded models for child automatic speech recognition applications. Similar directions can be considered for automatic speaker recognition systems.

The data presented here also provide a new perspective on the relationship between children’s growth and the geometric properties of the tracheo-bronchial tree, measured indirectly through acoustic analysis. Future studies should compare the modeled “effective acoustic length” of the tracheo-bronchial tree with anatomical lengths of the trachea and bronchi. It may be possible with future research to use accelerometer signals to determine airway lumen dimensions non-invasively (unlike methods using acoustic reflectometry), which could impact the practice of pediatric anaesthesiology (Henry and Royston, 2017, 2018a,b).

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation Grant Nos. 0905381 and 1551113. The authors extend their thanks to John R. Morton for the extensive task of recording and labeling the children’s data and Qi Wang, Evan Xue, and Luchuan Zhang for assisting with initial measurements. Additionally, the authors thank the reviewers and editor for their content and organizational suggestions.

APPENDIX A: MICROPHONE MEASUREMENTS AND ANALYSIS

This appendix follows up on Sec. II C and discusses the general measurement procedure of speech parameters and resulting measurements from the hVd microphone recordings.

For the vowels /a/, /æ/, /ʌ/, /ɛ/, /ɪ/, /i/, /ʊ/, and /u/ from their respective hVd words, time stamps for the start and end of the target vowel were labeled by trained students from the University of California, Los Angeles (UCLA). The start of the vowel was labeled where the formants became visible and the waveform displayed a clear deviation from the aspirated qualities of the /h/ phoneme. Similarly, the end of the vowel was labeled just before where the energy vanished due to the plosive /d/. The center of the steady-state portion of the vowel was also labeled, chosen as the point where formants were clearly visible and constant. This was usually near the midpoint between the starting and ending time stamps. Labels were verified by the first author to ensure consistency. All labeling was done using PRAAT (Boersma and Weenink, 2017), and each label file was saved in the TextGrid format. As such, each microphone recording had an associated label file indicating the start of the vowel, center of the steady-state region of the vowel, and end of the vowel. In general, labeling was accomplished through careful inspection of the spectrogram, but labelers also listened to the recordings to verify accuracy.

Before measuring f_o , F1, F2, and F3, all microphone signals were downsampled to 10 kHz as the first three formants for all recorded vowels were less than 5 kHz. Downsampling to 10 kHz also allowed for easier and more focused visual

inspection of the microphone signals. Then, f_o , F1, F2, and F3 were extracted automatically from the microphone signals using PRAAT's cross-correlation-based "To Pitch (cc)" function and LPC-based "To Formant (burg)" function. A window size of 49 ms was used to capture the vowels at steady state. A window shift of 5 ms was used. The minimum and maximum f_o parameters were set to 75 Hz and 600 Hz, respectively. The values of f_o , F1, F2, and F3 were chosen as the average values over five frames around the point labeled as the steady-state of the vowel. As automatic formant estimation is often difficult when applied to young children with high f_o values (Makhoul, 1975; Monsen and Engebretson, 1983), all values were verified by the first author to ensure accuracy and consistency.

Several metrics were used to identify incorrect formant measurements. As each speaker pronounced each hVd word a minimum of six times, at least six different formants measurements were available for the evaluation of a single speaker's vowel formants. If one of the measurements was a clear outlier compared to the other measurements, that file would be corrected manually. Formant measurements were also compared to Lee *et al.* (1999) based on vowel and age. Files with measurements that differed by over 2 standard deviations from the values reported in Lee *et al.* (1999) would be evaluated manually. Additionally, previous studies on children's formants (e.g., Vorperian and Kent, 2007) were used as references to evaluate the quality of formant measurements. In general, f_o measurements were more accurate than formant measurements and required few corrections.

Manual corrections were aided by a visual analysis of the DFT and LPC spectral envelope at the steady-state portions of the vowel using WAVESURFER. Previous studies have recommended varying the LPC polynomial order for formant estimation of both adults (Vallabha and Tuller, 2002) and children (Buder, 1996). As such, the LPC order was varied from 6 to 24 for the visual inspection. Automatic formant values that were judged to be questionable were either corrected manually if the formants could be observed through such a visual examination of the spectra and LPC or discarded if the files were too noisy.

Approximately 25% of the formant measurements were corrected across all speakers with less than half of these being discarded. However, the number of formant corrections tended to differ depending on the age of the speaker. In the most extreme case, 50% of the formant measurements from the youngest speaker (speaker 128; 6 yr old) either needed to be corrected or were discarded due to low recording quality. In general, older speakers needed fewer corrections than younger speakers. The most common mistakes made by the automatic formant estimator when estimating formants of younger children were misidentifying a harmonic as a formant and identifying some lower (non-existent) formant, likely caused by noise. As such, the formant estimator tended to underestimate the formants of younger children.

Tables IV and V show the means and standard deviations of the f_o , F1, F2, and F3 measurements for the vowels of the various hVd words of the male and female participants, respectively. Both tables were sorted by age for comparability to previous studies that also grouped by age

(e.g., Huber *et al.*, 1999; Lee *et al.*, 1999; Vorperian and Kent, 2007). Most of the values in Tables IV and V are within 2 standard deviations of the values reported in Lee *et al.* (1999). As for the measurements that were different from the findings in Lee *et al.* (1999), there are several possible explanations. First, sample sizes for each age group were small, so measurements were susceptible to bias due to accents or outliers. For instance, the 6 yr old girl (speaker 146) was incapable of pronouncing "had," instead pronouncing the word as /h ε d/. As this speaker was the only speaker with obvious mispronunciations and was very young, we have included her utterances for completeness. Additionally, the subjects were representative of a period of rapid growth in children (10–16 yr old for males, 8–14 yr old for females) (Barbier *et al.*, 2015; Vorperian *et al.*, 2009), and the rate of growth could vary widely between children. This would account for the large changes in formants seen between age groups. Finally, we note that Lee *et al.* (1999) performed measurements using bVC and pVC words while we performed measurements using hVd words. This may have resulted in different coarticulatory effects.

APPENDIX B: ACCELEROMETER MEASUREMENTS AND ANALYSIS

Section IID outlines the general measurement procedure of the accelerometer recordings of the sustained vowel /a/. This appendix follows up by discussing possible measurement difficulties and additional details that may interest some readers.

The use of the LPC spectral envelope with order varied from 6 to 24 was based on the same procedure used for the microphone signals. The procedure was found to be mostly successful in measuring the SGRs from the sustained vowel accelerometer signals. In a small number of cases, accelerometer files either contained too much noise or some of the SGRs were not visually obvious after applying the LPC envelope. In these cases, only the SGRs that resulted in clear resonance peaks in the LPC envelope were recorded. Each of the first three SGRs was measured from at least one recording for each participant across the two repetitions of the accelerometer recordings of the vowel /a/. The SGRs were measured several times across the sustained vowel /a/ and averaged within each vowel utterance to ensure that measurements were accurate. In general, varying the time location of the SGR measurements in a single file did not change the measurements.

One of the main difficulties found in measuring children's SGRs from the accelerometer signals was due to high fundamental frequencies, similar to the difficulties found in measuring children's formants. Specifically, the major difficulty in measuring Sg1 came from interference with the glottal signal structure, as strong low-frequency harmonics were often close to Sg1. This difficulty was compounded by the fact that higher-pitched children had larger spacing between harmonics, decreasing the sampling of the subglottal spectral envelope. The main difficulty in measuring Sg3 came from the low-pass properties of the neck tissue, which often attenuated Sg3 and caused difficulties in visualizing the spectrum

TABLE IV. Mean f_o , F1, F2, and F3 (Hz) values for the hVd vowel pronunciations of the male participants, sorted by age (years;months). Standard deviations of the measurements are in the parentheses. Also listed are the number of participants in each age group.

Age	Number		ɑ	æ	ʌ	ɛ	ɪ	i	ʊ	u
6;0–6;11	2	f_o	222(32)	230(38)	244(26)	242(35)	244(44)	254(47)	250(40)	252(25)
		F1	1048(120)	934(174)	743(121)	739(93)	618(55)	432(49)	583(40)	469(60)
		F2	1716(176)	2545(101)	1577(73)	2395(34)	2722(88)	3159(201)	1374(109)	1328(182)
		F3	3262(139)	3563(112)	3361(216)	3515(182)	3667(181)	3753(275)	3336(402)	3266(318)
7;0–7;11	4	f_o	215(32)	221(13)	223(34)	228(22)	226(14)	243(18)	234(14)	241(12)
		F1	1006(119)	967(93)	811(85)	802(95)	633(62)	456(103)	654(90)	445(49)
		F2	1516(184)	2318(212)	1595(186)	2387(210)	2628(202)	3189(279)	1502(270)	1547(172)
		F3	3400(245)	3430(244)	3644(298)	3621(369)	3612(318)	3867(524)	3619(238)	3251(189)
8;0–8;11	6	f_o	226(25)	220(26)	230(20)	226(24)	230(28)	238(20)	237(34)	244(44)
		F1	939(103)	955(132)	741(119)	710(113)	504(45)	444(67)	547(59)	416(50)
		F2	1412(161)	2153(243)	1623(133)	2394(150)	2651(133)	3203(152)	1587(216)	1326(364)
		F3	3064(175)	3170(252)	3351(186)	3339(209)	3443(135)	3595(225)	3287(219)	3227(181)
9;0–9;11	2	f_o	230(9)	218(19)	229(23)	236(8)	235(11)	242(24)	231(13)	245(12)
		F1	933(51)	833(119)	668(54)	671(65)	478(34)	395(49)	514(57)	368(95)
		F2	1543(140)	2254(99)	1623(81)	2331(35)	2746(88)	3219(201)	1656(85)	1114(199)
		F3	3265(186)	3076(78)	3561(187)	3543(200)	3749(199)	3691(303)	3388(227)	3128(124)
10;0–10;11	2	f_o	206(14)	197(21)	206(16)	207(17)	213(20)	215(33)	210(18)	212(25)
		F1	899(72)	821(108)	587(77)	578(74)	453(26)	340(36)	467(63)	368(32)
		F2	1517(57)	2346(289)	1512(77)	2398(136)	2601(140)	2876(269)	1559(57)	1379(328)
		F3	3027(366)	3155(281)	3252(280)	3235(147)	3257(288)	3317(136)	3195(169)	3210(226)
11;0–11;11	3	f_o	227(22)	231(28)	233(29)	238(34)	235(24)	248(39)	243(27)	247(20)
		F1	1004(145)	871(108)	701(67)	720(88)	542(71)	344(55)	586(71)	382(86)
		F2	1462(129)	2216(139)	1526(114)	2230(116)	2417(162)	2975(125)	1630(117)	1363(303)
		F3	2947(304)	3128(172)	3236(263)	3106(175)	3372(237)	3705(375)	3100(191)	3021(198)
12;0–12;11	1	f_o	209(8)	216(12)	220(5)	224(7)	232(8)	220(6)	220(2)	229(7)
		F1	806(20)	896(33)	677(56)	672(29)	490(54)	420(30)	596(31)	416(22)
		F2	1432(41)	1905(71)	1811(39)	2071(41)	2207(45)	2907(104)	1779(117)	1342(52)
		F3	2413(110)	2706(82)	3124(178)	3186(150)	3286(45)	3375(110)	3264(151)	3170(94)
13;0–13;11	5	f_o	157(30)	158(32)	161(30)	159(29)	163(31)	168(30)	164(34)	166(34)
		F1	790(80)	698(109)	607(106)	533(102)	430(74)	315(45)	467(91)	336(52)
		F2	1276(131)	1876(159)	1349(152)	2030(168)	2222(191)	2639(117)	1438(195)	1141(297)
		F3	2585(240)	2704(170)	2900(112)	2839(156)	2916(172)	3145(229)	2914(153)	2941(169)
14;0–14;11	1	f_o	183(12)	174(10)	195(10)	191(17)	206(16)	204(19)	202(23)	208(21)
		F1	769(47)	832(38)	667(50)	698(50)	519(58)	337(42)	523(57)	401(35)
		F2	1236(44)	1912(106)	1357(34)	2084(31)	2427(51)	2965(83)	1433(58)	1410(94)
		F3	3150(117)	2938(133)	3020(118)	3091(196)	3099(82)	3672(110)	3034(51)	2936(95)
15;0–15;11	1	f_o	113(6)	118(5)	122(4)	125(9)	128(9)	141(7)	136(7)	133(12)
		F1	843(50)	642(34)	478(18)	496(55)	332(20)	287(9)	358(18)	297(26)
		F2	1334(71)	2101(38)	1404(70)	2242(43)	2477(43)	2709(81)	1431(51)	893(59)
		F3	2355(206)	2792(64)	3208(35)	3072(72)	3042(96)	3453(100)	3188(80)	3442(137)
16;0–16;11	3	f_o	136(18)	166(81)	137(16)	137(16)	140(16)	150(20)	140(16)	144(15)
		F1	773(55)	709(51)	622(32)	600(43)	442(23)	294(36)	448(43)	342(57)
		F2	1307(84)	1964(161)	1326(105)	1948(168)	2268(114)	2764(180)	1316(144)	1316(112)
		F3	2761(248)	2875(191)	2928(161)	2951(187)	2956(156)	3407(212)	2838(165)	2838(204)
17;0–17;11	1	f_o	107(3)	109(5)	114(3)	113(4)	117(4)	115(4)	114(2)	123(4)
		F1	782(24)	619(23)	585(25)	547(24)	457(7)	312(16)	460(5)	358(13)
		F2	1284(19)	2025(47)	1253(24)	2001(30)	2161(16)	2625(34)	1276(26)	1207(114)
		F3	2555(25)	2702(11)	2700(41)	2739(35)	2747(24)	3053(98)	2637(19)	2487(20)

around Sg3. Sg2 tended to be much easier to measure than Sg1 or Sg3. These difficulties in measurement were almost identical to those reported in the measurement of adult SGRs (Lulich *et al.*, 2012). In most of the accelerometer recordings, these difficulties in measuring SGRs could be resolved. Doing so required both the varying of the LPC polynomial order to assist with visualization and an estimate of the approximate locations of the SGRs from previous studies (Lulich *et al.*, 2011a; Lulich *et al.*, 2011b; Lulich *et al.*, 2012).

APPENDIX C: QUANTAL THEORY IMPLICATIONS

The results of this paper may contain implications for the quantal theory of vowel production (Stevens, 1989). These are briefly documented in this appendix.

In adults, SGRs have been shown to be used as boundaries dividing the (F1,F2) vowel space into four quadrants (Arsikere *et al.*, 2013; Chi and Sonderegger, 2007). Sg1 is used as a threshold for F1 dividing low and high vowels, and

TABLE V. Mean f_o , F1, F2, and F3 (Hz) values for the hVd vowel pronunciations of the female participants, sorted by age (years;months). Standard deviations of the measurements are in the parentheses. Also listed are the number of participants in each age group.

Age	Number		a	æ	Λ	ε	ɪ	i	u	u
6;0–6;11	1	f_o	282(17)	271(25)	285(14)	257(11)	268(21)	272(10)	289(11)	287(3)
		F1	999(69)	677(57)	642(45)	686(30)	552(28)	450(38)	591(16)	534(48)
		F2	1698(77)	2594(141)	1696(47)	2637(125)	2788(60)	3342(80)	1771(51)	1298(173)
		F3	3617(162)	3489(148)	4053(60)	3547(93)	3615(108)	3997(65)	3697(44)	2854(91)
7;0–7;11	1	f_o	233(13)	238(14)	246(17)	254(17)	251(12)	269(8)	257(19)	215(83)
		F1	925(43)	1008(77)	724(99)	759(61)	522(30)	565(41)	511(39)	508(28)
		F2	1238(251)	2199(51)	1543(114)	2455(75)	2680(151)	3290(147)	1652(110)	1216(135)
		F3	3070(393)	3556(135)	3471(194)	3515(123)	3218(233)	3693(112)	3363(42)	3234(171)
8;0–8;11	1	f_o	290(94)	244(5)	248(8)	246(9)	251(4)	259(3)	253(3)	251(13)
		F1	1174(65)	999(50)	846(78)	892(77)	550(100)	498(7)	564(99)	484(7)
		F2	1632(103)	2252(128)	1616(80)	2476(43)	2585(93)	3109(21)	1504(114)	1644(121)
		F3	3110(129)	3199(117)	3354(122)	3192(127)	3390(139)	3592(42)	3197(76)	3145(101)
9;0–9;11	1	f_o	202(10)	202(7)	208(17)	213(21)	228(7)	221(8)	226(4)	227(10)
		F1	964(59)	1017(22)	804(57)	702(12)	622(45)	557(147)	661(18)	435(13)
		F2	1471(58)	2315(52)	1566(36)	2503(108)	2691(34)	3360(24)	1555(102)	1485(38)
		F3	3010(149)	3283(54)	3260(95)	3577(87)	3692(45)	3717(99)	3141(80)	3255(199)
10;0–10;11	3	f_o	224(21)	228(21)	235(23)	233(19)	237(18)	240(11)	241(17)	244(15)
		F1	1039(120)	1024(123)	776(75)	746(77)	568(57)	421(127)	661(47)	426(53)
		F2	1456(162)	2093(154)	1648(102)	2363(148)	2610(106)	3198(141)	1609(98)	1609(166)
		F3	3402(120)	3157(181)	3360(74)	3337(86)	3467(88)	3797(171)	3302(107)	3155(149)
11;0–11;11	1	f_o	211(12)	212(8)	222(9)	224(17)	222(4)	249(4)	228(8)	236(12)
		F1	934(73)	803(42)	671(27)	669(47)	470(17)	310(14)	510(38)	385(41)
		F2	1427(69)	2315(46)	1593(43)	2270(61)	2529(76)	3022(43)	1567(66)	1145(68)
		F3	3229(217)	3156(168)	3374(55)	3250(95)	3323(69)	3600(95)	3290(57)	3094(107)
12;0–12;11	2	f_o	188(35)	186(36)	210(20)	206(23)	199(27)	216(27)	214(24)	215(22)
		F1	928(132)	751(82)	638(106)	596(62)	472(36)	420(43)	562(74)	429(38)
		F2	1495(163)	2161(185)	1630(102)	2228(151)	2352(155)	2831(211)	1618(96)	1350(126)
		F3	2813(201)	3055(272)	3235(210)	3082(128)	3194(136)	3439(95)	2975(362)	2911(345)
13;0–13;11	2	f_o	206(26)	206(46)	214(27)	212(27)	212(24)	214(19)	218(22)	216(23)
		F1	730(322)	777(47)	541(79)	610(69)	459(48)	412(33)	474(99)	384(22)
		F2	1482(368)	2196(151)	1790(151)	2285(247)	2524(224)	3020(50)	1787(253)	1483(282)
		F3	3094(121)	3154(90)	3330(146)	3262(80)	3335(94)	3442(109)	3258(108)	3280(76)

Sg2 is used as a threshold for F2 dividing back and front vowels. To analyze the vowel spaces of children, the (F1,F2) placement of the four corner vowels /a/, /æ/, /i/, and /u/ were examined in a formant-normalized space to account for age-related differences.

Figure 6 shows the (F1,F2) coordinates of all measured corner vowels produced by the participants of this study with F1 normalized by Sg1 and F2 normalized by Sg2. A majority of each vowel's (F1,F2) coordinate locations is in its predicted quadrant, and Sg1 and Sg2 act as natural thresholds for the vowel divisions in most cases.

Table VI shows the percentage of vowel formant measurements within their expected quadrants. For the vowel /i/, both F1 and F2 were in the predicted quadrant in 100% of cases. For the vowel /u/, F1 was in the predicted quadrant in 100% of cases, but F2 was in the predicted quadrant only 84.4% of cases. For /a/ and /æ/, F1 was in the predicted quadrant in 93.9% and 90.2% of cases, respectively, and F2 was in the predicted quadrant in 87.8% and 96.6% of cases, respectively. These findings for the four corner vowels are similar to those reported for adult speakers of German (Madsack et al., 2008) and Hungarian (Csapó et al., 2009b). The other vowels in Table VI (more mid and central) tended to follow the quantal patterns much less than the corner vowels.

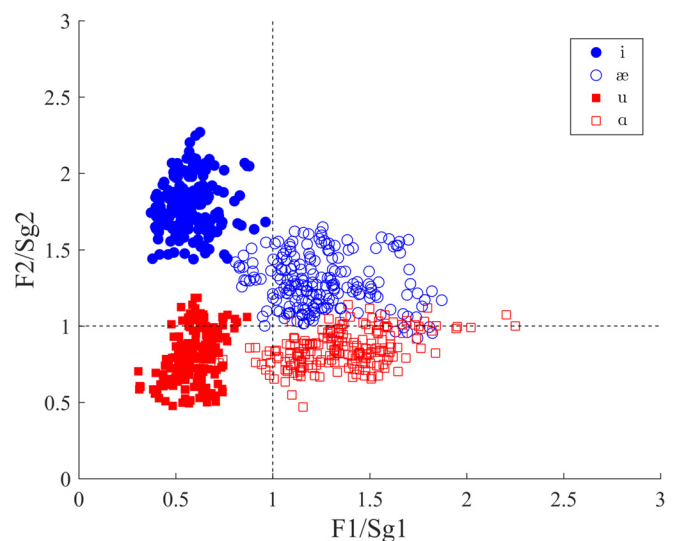


FIG. 6. (Color online) F2/Sg2 vs F1/Sg1 plot of the four corner vowels /a/, /æ/, /i/, and /u/ produced by participants of this study. The SGR boundaries (normalized to 1) are plotted as dashed lines. The SGRs appear to serve as boundaries for the corner vowels in the formant space.

TABLE VI. Percentage of vowel utterances across all participants with typical F1 and F2 placements compared to the SGR boundaries.

Formant	ɑ	ʌ	æ	ɛ	i	ɪ	u	ʊ
F1	93.9%	48.4%	90.2%	43.0%	100%	97.1%	100%	90.0%
F2	87.8%	74.4%	96.6%	100%	100%	100%	84.4%	73.2%
Both	81.7%	33.3%	86.8%	43.0%	100%	97.1%	84.4%	66.4%

Jung (2009) found that children’s low vowels are situated within their quantal-theoretic quadrants by the age of 4 yr, which is younger than any of the children in the present study. The children’s corner vowel formants used in our study generally followed this quantal pattern as well. However, for a few of the corner vowels, F1 and F2 came very close to or slightly crossed over the SGRs, observed in Fig. 6, and the remaining four vowels did not follow the quantal patterns as closely. This may suggest a weaker version of quantal theory, which takes into consideration additional competing factors such as coarticulation (Grácz et al., 2011) or other possible interactions between formants and SGRs. This may also be due to the fact that children are still developing speech and articulatory skills and cannot fully adhere to the quantal pattern yet.

Arsikere, H., Leung, G. K. F., Lulich, S. M., and Alwan, A. (2012a). “Automatic estimation of the first two subglottal resonances in children’s speech with application to speaker normalization in limited-data conditions,” in *Proc. of INTERSPEECH*, pp. 4616–4619.

Arsikere, H., Leung, G. K. F., Lulich, S. M., and Alwan, A. (2012b). “Automatic height estimation using the second subglottal resonance,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3989–3992.

Arsikere, H., Leung, G. K. F., Lulich, S. M., and Alwan, A. (2013). “Automatic estimation of the first three subglottal resonances from adults’ speech signals with application to speaker height estimation,” *Speech Commun.* **55**(1), 51–70.

Barbier, G., Boë, L.-J., Captier, G., and Laboissière, R. (2015). “Human vocal tract growth: A longitudinal study of the development of various anatomical structures,” in *Proc. of INTERSPEECH*, pp. 364–368.

Benade, A. H. (1990). *Fundamentals of Musical Acoustics*, 2nd ed. (Dover, Mineola, NY), pp. 405–409.

Boersma, P., and Weenink, D. (2017). “PRAAT: Doing phonetics by computer (version 6) [computer program],” <http://www.praat.org> (Last viewed July 2017).

Buder, E. H. (1996). “Experimental phonology with acoustic phonetic methods: Formant measures from child speech,” in *Proc. of the UBC International Conference on Phonological Acquisition*, pp. 254–265.

Chi, X., and Sonderegger, M. (2007). “Subglottal coupling and its influence on vowel formants,” *J. Acoust. Soc. Am.* **122**(3), 1735–1745.

Cranen, B., and Boves, L. (1987). “On subglottal formant analysis,” *J. Acoust. Soc. Am.* **81**(3), 734–746.

Csapó, T. G., Bárkányi, Z., Grácz, T. E., Böhm, T., and Lulich, S. M. (2009a). “Relation of formants and subglottal resonances in Hungarian vowels,” in *Proc. of INTERSPEECH*, pp. 484–487.

Csapó, T. G., Grácz, T. E., Bárkányi, Z., Beke, A., and Lulich, S. M. (2009b). “Patterns of Hungarian vowel production and perception with regard to subglottal resonances,” *The Phonetician* **99–100**, 7–28.

Dogil, G., Lulich, S. M., Madsack, A., and Wokurek, W. (2011). “Crossing the quantal boundaries of features: Subglottal resonances and Swabian diphthongs,” in *Tones and Features: Phonetic and Phonological Perspectives* (De Gruyter Mouton, The Hague, Netherlands), pp. 137–148.

Fant, G. (1960). *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations* (Mouton, The Hague, Netherlands), pp. 1–328.

Fredberg, J. J., and Moore, J. A. (1978). “The distributed response of complex branching duct networks,” *J. Acoust. Soc. Am.* **63**(3), 954–961.

Grácz, T. E., Lulich, S. M., Csapó, T. G., and Beke, A. (2011). “Context and speaker dependency in the relation of vowel formants and subglottal resonances—Evidence from Hungarian,” in *Proc. of INTERSPEECH*, pp. 1901–1904.

Griscom, N. T., and Wohl, M. E. B. (1985). “Dimensions of the growing trachea related to body height: Length, anteroposterior, and transverse diameters, cross-sectional area, and volume in subjects younger than 20 years of age,” *Am. Rev. Respir. Dis.* **131**(6), 840–844.

Griscom, N. T., and Wohl, M. E. B. (1986). “Dimensions of the growing trachea related to age and gender,” *Am. J. Roentgenol.* **146**(2), 233–237.

Guo, J., Paturi, R., Yeung, G., Lulich, S. M., Arsikere, H., and Alwan, A. (2015). “Age-dependent height estimation and speaker normalization for children’s speech using the first three subglottal resonances,” in *Proc. of INTERSPEECH*, pp. 1665–1669.

Guo, J., Yeung, G., Muralidharan, D., Arsikere, H., Afshan, A., and Alwan, A. (2016). “Speaker verification using short utterances with DNN-based estimation of subglottal acoustic features,” in *Proc. of INTERSPEECH*, pp. 2219–2222.

Hanna, N., Smith, J., and Wolfe, J. (2018). “How the acoustic resonances of the subglottal tract affect the impedance spectrum measured through the lips,” *J. Acoust. Soc. Am.* **143**(5), 2639–2650.

Henry, B., and Royston, T. J. (2017). “A multiscale analytical model of bronchial airway acoustics,” *J. Acoust. Soc. Am.* **142**(4), 1774–1783.

Henry, B., and Royston, T. J. (2018a). “Localization of adventitious respiratory sounds,” *J. Acoust. Soc. Am.* **143**(3), 1297–1307.

Henry, B., and Royston, T. J. (2018b). “Erratum: A multiscale analytical model of bronchial airway acoustics [J. Acoust. Soc. Am. **142**, 1774–1783 (2017)],” *J. Acoust. Soc. Am.* **143**(3), 1427.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (1999). “Formants of children, women, and men: The effects of vocal intensity variation,” *J. Acoust. Soc. Am.* **106**(3), 1532–1542.

Ishizaka, K., Matsudaira, M., and Kaneko, T. (1976). “Input acoustic-impedance measurement of the subglottal system,” *J. Acoust. Soc. Am.* **60**(1), 190–197.

Jackson, A. C., Butler, J. P., and Pyle, R. W. (1978). “Acoustic input impedance of excised dog lungs,” *J. Acoust. Soc. Am.* **64**(4), 1020–1026.

Jung, Y. (2009). “Acoustic articulatory evidence for quantal vowel categories: The features [low] and [back],” Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology.

Kim, E. J., Kim, S. Y., Kim, W. O., Kim, H., and Kil, H. K. (2013). “Ultrasound measurement of subglottic diameter and an empirical formula for proper endotracheal tube fitting in children,” *Acta Anaesthesiol. Scand.* **57**(9), 1124–1130.

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**(2), 820–857.

Lee, S., Potamianos, A., and Narayanan, S. (1999). “Acoustics of children’s speech: Developmental changes of temporal and spectral parameters,” *J. Acoust. Soc. Am.* **105**(3), 1455–1468.

Lulich, S. M. (2010). “Subglottal resonances and distinctive features,” *J. Phonetics* **38**(1), 20–32.

Lulich, S. M. (2013). “Estimation of lumped vocal fold mechanical properties from non-invasive microphone recordings,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 8041–8045.

Lulich, S. M., Alwan, A., Arsikere, H., Morton, J. R., and Sommers, M. S. (2011a). “Resonances and wave propagation velocity in the subglottal airways,” *J. Acoust. Soc. Am.* **130**(4), 2108–2115.

Lulich, S. M., and Arsikere, H. (2015). “Tracheo-bronchial soft tissue and cartilage resonances in the subglottal acoustic input impedance,” *J. Acoust. Soc. Am.* **137**(6), 3436–3446.

Lulich, S. M., Arsikere, H., Morton, J. R., Leung, G. K. F., Alwan, A., and Sommers, M. S. (2011b). “Analysis and automatic estimation of children’s subglottal resonances,” *Proc. Mtgs. Acoust.* **6**, 060007.

Lulich, S. M., Bachrach, A., and Malyska, N. (2007). “A role for the second subglottal resonance in lexical access,” *J. Acoust. Soc. Am.* **122**(4), 2320–2327.

Lulich, S. M., Morton, J. R., Arsikere, H., Sommers, M. S., Leung, G. K. F., and Alwan, A. (2012). “Subglottal resonances of adult male and female native speakers of American English,” *J. Acoust. Soc. Am.* **132**(4), 2592–2602.

Lulich, S. M., Zañartu, M., Mehta, D. D., and Hillman, R. E. (2009). “Source-filter interaction in the opposite direction: Subglottal coupling

- and the influence of vocal fold mechanics on vowel spectra during the closed phase," *Proc. Mtgs. Acoust.* **6**, 060007.
- Madsack, A., Lulich, S. M., Wokurek, W., and Dogil, G. (2008). "Subglottal resonances and vowel formant variability: A case study of high German monophthongs and Swabian diphthongs," in *Proc. of Laboratory Phonology 11*, pp. 91–92.
- Mahajan, A., Hoftman, N., Hsu, A., Schroeder, R., and Wald, S. (2007). "Continuous monitoring of dynamic pulmonary compliance enables detection of endobronchial intubation in infants and children," *Anesth. Analg.* **105**(1), 51–56.
- Makhoul, J. (1975). "Linear prediction: A tutorial review," in *Proc. of the IEEE*, pp. 561–580.
- MathWorks (2016). "MATLAB (version 9.1) [computer program]," Natick, Massachusetts, <https://www.mathworks.com> (Last viewed September 2016).
- Monsen, R. B., and Engebretson, A. M. (1983). "The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction," *J. Speech Hear. Res.* **26**(1), 89–97.
- Pyle, R. W. (1975). "Effective length of horns," *J. Acoust. Soc. Am.* **57**(6), 1309–1317.
- Sjölander, K., and Beskow, J. (2000). "WAVESURFER—An open source speech tool," in *Proc. of the International Conference on Spoken Language Processing*, Vol. 4, pp. 464–467, <http://www.speech.kth.se/wavesurfer/> (Last viewed July 2017).
- Stevens, K. N. (1972). "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication: A Unified View*, edited by P. B. Denes and E. E. David, Jr. (McGraw-Hill, New York), pp. 51–66.
- Stevens, K. N. (1989). "On the quantal nature of speech," *J. Phonetics* **17**(1), 3–45.
- Stevens, K. N. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA), pp. 1–607.
- Titze, I. R. (2006). *The Myoelastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Denver, CO), pp. 1–424.
- Titze, I. R. (2008). "Nonlinear source-filter coupling in phonation: Theory," *J. Acoust. Soc. Am.* **123**(5), 2733–2749.
- Titze, I. R., Riede, T., and Popolo, P. (2008). "Nonlinear source-filter coupling in phonation: Vocal exercises," *J. Acoust. Soc. Am.* **123**(4), 1902–1915.
- Vallabha, G. K., and Tuller, B. (2002). "Systematic errors in the formant analysis of steady-state vowels," *Speech Commun.* **38**(1–2), 141–160.
- Van den Berg, J. W. (1960). "An electrical analogue of the trachea, lungs and tissues," *Acta Physiol. Pharmacol. Neerl.* **9**, 361–385.
- Vorperian, H. K., and Kent, R. D. (2007). "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," *J. Speech, Lang. Hea. Res.* **50**(6), 1510–1545.
- Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., and Gentry, L. R. (2009). "Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study," *J. Acoust. Soc. Am.* **125**(3), 1666–1678.
- Wade, L., Hanna, N., Smith, J., and Wolfe, J. (2017). "The role of vocal tract and subglottal resonances in producing vocal instabilities," *J. Acoust. Soc. Am.* **141**(3), 1546–1559.
- Wang, S., Lee, Y.-H., and Alwan, A. (2009a). "Bark-shift based nonlinear speaker normalization using the second subglottal resonance," in *Proc. of INTERSPEECH*, pp. 1619–1622.
- Wang, S., Lulich, S. M., and Alwan, A. (2009b). "Automatic detection of the second subglottal resonance and its application to speaker normalization," *J. Acoust. Soc. Am.* **126**(6), 3268–3277.
- 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**(1), 326–339.
- Zhang, Z., Neubauer, J., and Berry, D. A. (2006). "The influence of subglottal acoustics on laboratory models of phonation," *J. Acoust. Soc. Am.* **120**(3), 1558–1569.