AUTOMATIC HEIGHT ESTIMATION USING THE SECOND SUBGLOTTAL RESONANCE

Harish Arsikere[1] *Gary K. F. Leung*[1] *Steven M. Lulich*[2] *Abeer Alwan*[1]

[1] Department of Electrical Engineering, University of California, Los Angeles, USA.
[2] Department of Psychology, Washington University, Saint Louis, USA.
harishan@ucla.edu, garyleung@ucla.edu, slulich@wustl.edu, alwan@ee.ucla.edu

ABSTRACT

This paper presents an algorithm for automatically estimating speaker height. It is based on: (1) a recently-proposed model of the subglottal system that explains the inverse relation observed between subglottal resonances and height, and (2) an improved version of our previous algorithm for automatically estimating the second subglottal resonance (Sg2). The improved Sg2 estimation algorithm was trained and evaluated on recently-collected data from 30 and 20 adult speakers, respectively. Sg2 estimation error was found to reduce by 29%, on average, as compared to the previous algorithm. The height estimation algorithm, employing the inverse relation between Sg2 and height, was trained on data from the above-mentioned 50 adults. It was evaluated on 563 adult speakers in the TIMIT corpus, and the mean absolute height estimation error was found to be less than 5.6*cm*.

***Index Terms*—** speaker height, second subglottal resonance, uniform tube model, automatic estimation

1. INTRODUCTION

Automatic estimation of an unknown speaker's height from the speech signal can benefit forensics, and can also provide useful supplementary information about a speaker in tasks such as analysis of 911 telephone calls, and automatic speaker identification. Over the last few decades, several researchers have attempted to identify acoustic features of speech that can convey information about speaker height. Most of these efforts rely on the assumption that an anatomical correlation exists between height and vocal tract length (VTL). In fact, a study based on magnetic resonance imaging has shown strong evidence in favor of this assumption [1]. Several studies have analyzed the correlation between speaker height and formant frequencies [2, 3, 4]. Although speech production theory assumes a correlation between formant frequencies and VTL, these studies report only a weak correlation between formant frequencies and speaker height. A few studies have also investigated the relation between height and average fundamental frequency (F0), but have found no significant correlation between the two [3, 5]. Correlations between height and other commonly used acoustic features such as Mel-frequency cepstral coefficients (MFCCs) and linear prediction coefficients (LPCs) have been reported in a more recent study [6], which shows that 57% of the variance in height can be explained by combining the first 10 MFCCs, 16 LPCs and the first 5 formant frequencies. A few studies have proposed algorithms for automatically estimating speaker height. In [7], speech was parameterized using the first 19 MFCCs, and height-dependent Gaussian mixture models (GMMs) were trained using data from all speakers in the TIMIT corpus [8]. The height of a given test speaker was then estimated using the maximum *a posteriori* classification rule. With this approach, the estimation error was found to be 5*cm* or less for 72% of the speakers. However, it should be noted that the *same* set of speakers was used for both training and evaluation. In [9] and [10], regression models were proposed for height estimation. The models were trained and evaluated using data from 462 and 168 speakers, respectively, in the TIMIT corpus. Using a 50-dimensional feature vector, a mean absolute error of 5.3*cm* was achieved. The features consisted mostly of means, standard deviations, percentiles and quartiles of MFCCs, F0 and voicing probability. Although the algorithm yields good results, the relation between these features and speaker height is not clear.

This paper presents a novel algorithm that does *not* rely on the correlation between VTL and height. It is based on a recently proposed uniform tube model of the subglottal system that explains the inverse relation observed between speaker height and subglottal resonances (SGRs) [11]. The model assumes a correlation between height and the 'acoustic length' of the subglottal system, which can be defined as the length of an equivalent uniform tube (closed at one end) whose resonant frequencies closely match the actual SGR frequencies. This assumption is more meaningful than assuming a correlation between height and formant frequencies since the 'acoustic length' of the subglottal system does not vary considerably during speech production. Another important feature of this work is that the amount of training data and the number of features used for height estimation are very small in comparison with previous studies. Section 2 describes the data used. Section 3 explains the uniform tube model, and an improved algorithm for automatically estimating Sg2. Experiments and results of automatic height estimation are discussed in Section 4. Section 5 concludes the paper.

2. DATA USED

The WashU-UCLA corpus [12, 13] comprises simultaneous recordings of microphone and subglottal accelerometer signals from 50 adult speakers (25 males, 25 females) of American English. Every speaker, aged between 18 and 24 years, was recorded in two sessions: one with 14 *hVd* words (10 monophthongs - in which we include the approximant [ɹ] - and 4 diphthongs) and the other with 21 *CVb* words (4 monophthongs and 3 diphthongs, in three different consonant contexts). Every word, embedded in the carrier phrase, *"I said a ____ again"*, was recorded 10 times. Only the *hVd* words and the corresponding carrier phrases were used in this study. Speaker heights in the corpus range from 165*cm* to 188*cm* for males, and from 152*cm* to 175*cm* for females. The WashU-UCLA corpus was used for two purposes. (1) Data from all 50 speakers in the corpus were used to model the inverse relation between SGRs and

Work supported in part by NSF Grant No. 0905381.

height. Since the modeling involved steady-state measurements of SGRs, only accelerometer recordings of the 10 monophthongs were required. (2) Data from 30 speakers (15 males, 15 females) were used for training the improved Sg2 estimation algorithm, while data from the remaining 20 speakers were used for evaluation. Since algorithm training involved steady-state measurements of Sg2 as well as F0 and formant frequencies, it required both microphone and accelerometer recordings of monophthongs. However, since Sg2 (and hence height) had to be estimated from continuous speech, evaluation was performed on microphone recordings of the carrier phrases.

To evaluate the height estimation algorithm, data from 563 adult speakers (390 males, 173 females) in the TIMIT corpus were used. The heights of all the test speakers were within the height range spanned by the 50 training speakers mentioned above. To assess the algorithm's ability to estimate height from telephone speech, a narrowband evaluation set was also generated by filtering the TIMIT data with the ITU-T G.712 filter [14], which has a flat frequency response between 300 and 3400*Hz*.

3. METHODS

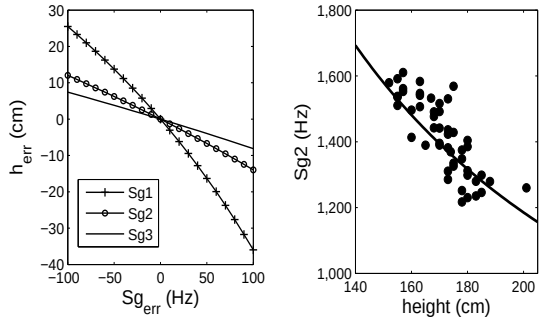
3.1. Uniform tube model of the subglottal system

Subglottal resonances are the poles of the input impedance of the subglottal system measured by looking down from the top of the trachea. Although the subglottal system consists of the trachea and a complex bronchial tree, it was shown in [11] that the first 3 SGRs can be predicted well by an equivalent uniform tube that is closed at the glottis and open at the inferior end. The following equation was used to model the relation between SGRs and speaker height h,

$$\text{SgN} = \frac{(2N-1)c}{4L_a} = \frac{(2N-1)c}{4h/k_a},\ \text{N} = 1, 2, 3 \qquad (1)$$

where SgN denotes the N^{th} SGR, c denotes the propagation velocity of sound waves in the subglottal airways, L_a denotes the length of the equivalent uniform tube (or the 'acoustic length'), and k_a is an empirically determined scale factor that relates h (height) and L_a.

Inspired by [11], the parameters of the uniform tube model represented by Eq. (1), were derived as follows: (1) SGRs of all 50 speakers in the WashU-UCLA corpus were measured manually in the accelerometer signals of 10 different monophthongs. The measurement procedure involved visual inspection of discrete Fourier transforms (DFTs), LPC spectra and smoothed spectral envelopes (see [15] for details). Sg1 and Sg2 were measured in anywhere between 10 and 30 tokens (14 on average) per speaker. Since the low-pass nature of the accelerometer signal made the measurement of Sg3 difficult, it was measured in fewer tokens (more than 6 on average) per speaker; for one speaker, no Sg3 measurements could be obtained. (2) The *actual* Sg1, Sg2 and Sg3 of a given speaker were taken to be the averages of the speaker's SGR measurements. (3) The height scaling factor was determined by minimizing the root mean squared (RMS) error incurred in fitting Eq. (1) to the actual values of Sg2, Sg3 (but not Sg1) and h, with N set to 2 (for Sg2) or 3 (for Sg3), and c set to c_0 = 35900*cm/s*, which is the free-field value of the speed of sound in humid air at body temperature. The scale factor was found to be equal to 8.803; this value is henceforth denoted as $\widetilde{k}_a$. Sg1 was not used in deriving $\widetilde{k}_a$ because, as shown in [11], the relation between Sg1 and h cannot be modeled by setting c to c_0, but by increasing it to a higher value in order to account for the non-rigid nature of the subglottal airway walls at low frequencies. (4) The increased wave propagation velocity, $\widetilde{c}$, required to model the relation between Sg1 and h was determined by minimizing the RMS error incurred in fitting Eq. (1) to the actual values of Sg1 and h, with k_a set to $\widetilde{k}_a$. The value of $\widetilde{c}$ was found to be 46572*cm/s*. In [11], the values corresponding to $\widetilde{k}_a$ and $\widetilde{c}$ were reported to be 8.508 and 46900*cm/s*, respectively. Since the values in [11] were derived using SGR measurements from just two accelerometer recordings (per speaker) of the sustained [ɑ] vowel, the parameters derived in the present study were used for automatic height estimation.



(a) (b)

Fig. 1. (a) The dependence of height estimation error (h_{err}) on the errors in estimating Sg1, Sg2 and Sg3; h_{err} is most sensitive to Sg1 errors and least sensitive to Sg3 errors. (b) Scatter plot of Sg2 versus height, and the corresponding empirical relation between them.

3.1.1. Sensitivity analysis of the uniform tube model

The SGR frequencies used to derive the uniform tube model were obtained from accelerometer signals. However, automatic estimation of speaker height requires SGRs to be estimated automatically from speech signals. Therefore, it is important to analyze how sensitive the uniform tube model is to SGR estimation errors. It is clear from Eq. (1) that height estimation using SGRs involves a relation of the form $h = K/x$, where K is a constant and x is the frequency of the SGR used. Hence, the sensitivity of the model to SGR estimation errors is proportional to a quantity of the form, $-K/x^2$ (the derivative of K/x). This means that height estimation is more sensitive to SGR errors when the SGR frequencies are small. Typically, SGR frequencies of adults are observed to be in the following ranges: 500-700*Hz* for Sg1, 1200-1600*Hz* for Sg2, and 2000-2400*Hz* for Sg3 [12]. Figure 1(a) shows a plot of height estimation error, h_{err}, versus SGR estimation error, Sg_{err}, assuming that the actual values of the first three SGRs are 600, 1400 and 2200*Hz*. Naturally, for a given $|Sg_{err}|$, the smallest $|h_{err}|$ can be achieved by using Sg3. However, existing automatic algorithms can estimate Sg1 [16] and Sg2 [13] only. Since h_{err} is much more sensitive to Sg1 errors than to Sg2 errors (see Fig. 1(a)), only the inverse relation between Sg2 and height was used in this study. Figure 1(b) shows the scatter plot of Sg2 versus height and the corresponding empirical relation based on Eq. (1): $Sg2 = \frac{3\times 35900\times 8.803}{4\times h}$. This empirical relation, which allows speaker height to be predicted from just one feature - Sg2 - accounts for 59% of the variance in the data. In comparison, [6] showed that 57% of the variance in height can be explained by using a 31-dimensional feature vector consisting of MFCCs, LPCs and formants. This suggests that the 'acoustic length' of the subglottal system is more reliable than VTL for estimating speaker height.

3.2. Improved algorithm for the automatic estimation of Sg2

An automatic algorithm for estimating Sg2 in continuous speech was proposed in [13]. The algorithm was based on the following central

idea: Sg2 acts as a boundary between *front* and *back* vowels [17], so that two acoustic features characterizing vowel backness - the Bark difference between the third and second formants (F3 and F2) and the Bark difference between F2 and Sg2 - are correlated. These two acoustic features were denoted in [13] as ${}_{f3}D_{f2}$ and ${}_{f2}D_{s2}$, respectively. For ease of representation, the two features will henceforth be denoted as B_{32} and $B_{2,S2}$, respectively. In [13], an empirical equation was derived to predict $B_{2,S2}$ from a linear combination of the first three powers of B_{32}, and a constant term. The empirical relation allowed Sg2 to be estimated from a speech signal once the formants F2 and F3 were tracked automatically.

In this study, the empirical relation involving B_{32} and $B_{2,S2}$ was derived again using data from 30 speakers in the WashU-UCLA corpus (as opposed to just 11 speakers in [13]). For each speaker, F2 and F3 were measured in the steady-state region of 5 tokens from every monophthong except [ɪ]. Snack [18] was used for measuring formants. The measured formants, along with the measured Sg2 values (see Sec. 3.1), were used to derive the following equation.

$$B_{2,S2} = -0.004(B_{32})^3 + 0.134(B_{32})^2 - 1.958(B_{32}) + 6.182 \quad (2)$$

Let the Sg2 estimation algorithm that makes use of the basic regression model represented by Eq. (2), be denoted as A1. Algorithm A1 was improved in this study by adding two more variables - F3 and F0 (measured using Snack), in Hertz - to the above multi-linear regression. The motivation behind using F3 and F0 is that they carry some speaker-related information. Inclusion of speaker-related information was deemed necessary because, when the empirical equation involving $B_{2,S2}$ and B_{32} was derived on a speaker-by-speaker basis, the coefficients of the equation were found to vary considerably from one speaker to another. The basic regression model (Eq. (2)) resulted in an r-squared (r^2) value of 0.891. When F3 and F0 (in that order) were added incrementally to the model, the value of r^2 increased to 0.943 ($p < 0.001$) and 0.971 ($p < 0.001$). Thus, the improved Sg2 estimation algorithm, A2, employs a more complete regression model that is represented by Eq. (3).

$$\begin{aligned} B_{2,S2} = {} & 0.001(B_{32})^3 + 0.009(B_{32})^2 - 1.083(B_{32}) \\ & + 0.002(F3) - 0.007(F0) - 0.019 \end{aligned} \quad (3)$$

Given a speech signal, the steps involved in estimating Sg2 using algorithm A2 can now be summarized as follows: (1) Track F0, F2 and F3 automatically using Snack by setting the frame length and spacing to 30 and 5*ms*, respectively. (2) Select voiced frames using the binary parameter 'probability of voicing' returned by Snack. (3) Estimate Sg2 for every voiced frame using Eq. (3). (4) Compute the average of all the frame-level Sg2 estimates obtained in Step 3 to arrive at an estimate of Sg2 for the given speech signal. Figure 2 shows the spectrogram of a speech signal along with the Sg2 measured in the corresponding accelerometer signal, and the estimated Sg2.

3.2.1. Performance analysis of algorithms A1 and A2

The Sg2 estimation algorithms, A1 and A2, were evaluated on 20 speakers of the WashU-UCLA corpus using two performance metrics: (1) Mean absolute error (MAE) and (2) Average standard deviation (ASD). MAE = $\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|Sg2_a^i - Sg2_e^{ij}\right|$, where $Sg2_a$ and $Sg2_e$ denote the actual and estimated Sg2 values, and M and N denote the number of speakers and the number of Sg2 estimates per speaker, respectively. ASD = $\frac{1}{M}\sum_{i=1}^{M}\sigma_i$, where σ_i is the standard deviation of Sg2 estimates of the i^{th} speaker. To compare A1 and A2, every Sg2 estimate was obtained using one sentence of data ($<$ 2 seconds). Table 1 shows the results of Sg2 estimation for both A1 and A2. Clearly, the inclusion of F3 and F0 in A2 reduces the overall MAE and ASD significantly. Algorithm A2 was also evaluated by obtaining Sg2 estimates from more than one sentence of data. When 2 sentences were used per Sg2 estimate, ASD reduced from 13*Hz* to 9*Hz*, and when 5 sentences were used per estimate, it reduced further to 5*Hz*. MAE did not decrease significantly; nevertheless, the reduction in ASD suggests that the performance of A2 improves with the amount of data provided.

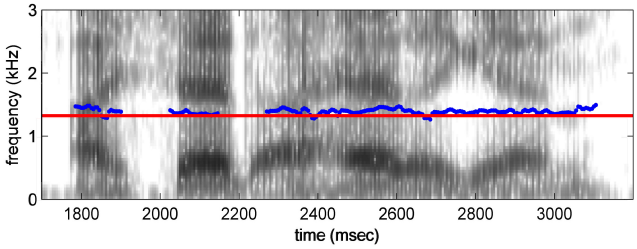


Fig. 2. Spectrogram of a microphone recording of *"I said a head again"* (speaker 12), superimposed with the Sg2 track from the accelerometer signal (blue), and the Sg2 estimated using A2 (red). The actual and estimated Sg2 values are 1385 and 1325*Hz*, respectively.

Metric	**Males**		**Females**		**Overall**		
(in *Hz*)	A1	A2	A1	A2	A1	A2	Reduction
MAE	79	61	84	56	82	58	29%
ASD	56	14	40	11	48	13	73%

Table 1. Mean absolute error (MAE) and average standard deviation (ASD) for the Sg2 estimation algorithms, A1 and A2. Every estimate was obtained using only one sentence of data, per speaker.

4. EXPERIMENTS, RESULTS AND DISCUSSION

Given a speech signal, speaker height was estimated as follows. (1) Sg2 was estimated using the improved algorithm A2. (2) An estimate of the height, h_e, was obtained using the height scaling factor $\widetilde{k}_a$ = 8.803, and the estimated Sg2 ($Sg2_e$):

$$h_e = \frac{3 \cdot c_0 \cdot \widetilde{k}_a}{4 \cdot Sg2_e} = \frac{3 \times 35900 \times 8.803}{4 \times Sg2_e}. \quad (4)$$

The algorithm incurs two kinds of errors, namely, the error in estimating Sg2 from speech, and the error in estimating height using the uniform tube model. Taking this into account, the algorithm was assessed using two metrics: (1) Speaker-level MAE (MAE_{sp}) and (2) Sentence-level MAE (MAE_{st}). Both the metrics are essentially the same, equal to $\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|h_a^i - h_e^{ij}\right|$, where h_a is the actual height, and M and N denote the number of speakers and the number of height estimates per speaker, respectively. The only difference is that in the calculation of MAE_{sp}, the effect of Sg2 estimation errors was mitigated by estimating Sg2 using all the data available (about 20 seconds per speaker), while to calculate MAE_{st}, Sg2 and height were estimated on a sentence-by-sentence basis in order to incorporate the effect of both the errors. Table 2 shows the results of automatic height estimation. The overall MAE_{sp} is less than 5.6*cm*. The overall MAE_{st} is slightly worse as expected, but is still acceptable.

Since $\widetilde{k}_a$ was obtained using very little data (50 samples), it was surmised that a different scale factor might provide better estimation accuracy. A search was performed for the height scaling factor that

Metric (in *cm*)	**Males**	**Females**	**Overall**
k_a **= 8.803**			
$\mathbf{MAE}_{sp}$	5.88	4.90	5.58 (3.23%)
$\mathbf{MAE}_{st}$	6.25	5.18	5.93 (3.43%)
k_a **= 8.699 (optimized for TIMIT)**			
$\mathbf{MAE}_{sp}$	5.33	5.45	5.37 (3.10%)
$\mathbf{MAE}_{st}$	5.73	5.69	5.72 (3.29%)
k_a **= 8.803; G.712-filtered speech**			
$\mathbf{MAE}_{sp}$	5.70	4.93	5.46 (3.16%)
$\mathbf{MAE}_{st}$	6.11	5.21	5.84 (3.37%)

Table 2. Speaker-level (MAE_{sp}) and sentence-level (MAE_{st}) mean absolute errors for the proposed height estimation algorithm. The numbers in parentheses denote average percentage errors.

resulted in the smallest MAE_{sp} for the TIMIT evaluation set. The 'optimal' scale factor was $k_a{}' = 8.699$, corresponding to an MAE_{sp} of 5.37*cm*; this is only 0.07*cm* worse than the best known result (reported in [9] and [10]). Since $\widetilde{k}_a$ differs from $k_a{}'$ by less than 1.5%, a scale factor of 8.803 is expected to yield reasonably good results in general. More importantly, it must be noted the proposed height estimation algorithm was trained on just 50 speakers and required only 3 features, F0, F2 and F3 (effectively, only 1 feature – Sg2), while the algorithms in [9] and [10] were trained on 462 speakers and required 50 features. Also, it is clear from Table 2 that there is little degradation in the algorithm's performance after G.712 filtering, which means that the algorithm can potentially be used in the analysis of telephone-based speech.

MAE as a metric of height estimation accuracy is useful, but it is also limited in some respects. For instance, even a naive estimate equal to the mean height of all 563 TIMIT speakers (174.4*cm*) yields an MAE of 6.99*cm*. As an additional metric, the correlation between the actual and estimated heights can be particularly informative: across all 563 TIMIT speakers, the proposed algorithm resulted in a correlation coefficient of 0.72 ($r^2 = 0.52$), indicating that roughly 52% of the variance in height was successfully accounted for. This is only slightly less than the variance accounted for in the training set of 50 speakers (59%), indicating that the algorithm is fairly robust when generalizing to new speakers.

5. CONCLUSION

The Sg2 estimation algorithm proposed previously in [13] improves when F3 and F0 are incorporated; the average error reduces by 29%. Speaker height can be automatically estimated using: (1) the improved algorithm for estimating Sg2, and (2) the inverse relation between Sg2 and height. The proposed height estimation algorithm performs equally well for wideband and narrowband speech. With sufficient data (about 20 seconds), speaker height can be estimated to within 5.58*cm* using the empirically determined scale factor (8.803), and to within 5.37*cm* using the 'optimal' scale factor (8.699), on average. The novelty of the algorithm is its dependence on the correlation between height and the 'acoustic length' of the subglottal system (and not VTL). The proposed algorithm is much simpler than the best existing algorithms (yet achieves comparable performance) because it requires very little training data and just three features – F0, F2 and F3. The results are likely to improve with more training data and better Sg2 estimation algorithms. The methods presented in this paper can be extended to children's speech as well, if a sizable corpus of simultaneous speech and subglottal acoustics is available.

6. REFERENCES

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

[2] W. A. van Dommelen and B. H. Moxness, "Acoustic parameters in speaker height and weight identification: sex-specific behaviour," *Language and Speech*, vol. 38, pp. 267–287, 1995.

[3] J. González, "Formant frequencies and body size of speaker: a weak relationship in adult humans," *Journal of Phonetics*, vol. 32, pp. 277–287, 2004.

[4] D. Rendall, S. Kollias, C. Ney, and P. Lloyd, "Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," *J. Acoust. Soc. Am.*, vol. 117, pp. 944–955, 2005.

[5] H. J. Künzel, "How well does average fundamental frequency correlate with speaker height and weight?," *Phonetica*, vol. 46, pp. 117–125, 1989.

[6] S. Dusan, "Estimation of speaker's height and vocal tract length from speech signal," in *Proc. of Interspeech*, 2005, pp. 1989–1992.

[7] B. L. Pellom and J. H. L. Hansen, "Voice analysis in adverse conditions: the centennial Olympic park bombing 911 call," in *40^{th} Midwest Symp. on Circuits and Sys.*, 1997, pp. 873–876.

[8] J. S. Garofolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database," *National Institute of Standards and Technology (NIST)*, 1988.

[9] T. Ganchev, I. Mporas, and N. Fakotakis, "Audio features selection for automatic height estimation from speech," *Artificial Int.: Theories, Models and Applications*, pp. 81–90, 2010.

[10] T. Ganchev, I. Mporas, and N. Fakotakis, "Automatic height estimation from speech in real-world setup," in *Proc. of the 18^{th} European Sig. Proc. Conf.*, 2010, pp. 800–804.

[11] S. M. Lulich, A. Alwan, H. Arsikere, J. R. Morton, and M. S. Sommers, "Resonances and wave propagation velocity in the subglottal airways," *J. Acoust. Soc. Am. (in press; Online: http://nautilus.icsl.ucla.edu/sgr/papers/sos.pdf)*, 2011.

[12] S. M. Lulich, J. R. Morton, M. S. Sommers, H. Arsikere, Y.-H. Lee, and A. Alwan, "A new speech corpus for studying subglottal acoustics in speech production, perception, and technology (A)," *J. Acoust. Soc. Am.*, vol. 128, pp. 2288(A), 2010.

[13] H. Arsikere, S. M. Lulich, and A. Alwan, "Automatic estimation of the second subglottal resonance from natural speech," in *Proc. of ICASSP*, 2011, pp. 4616–4619.

[14] ITU-T recommendation G.712, "Transmission performance characteristics of pulse code modulation channels," 2001.

[15] S. M. Lulich, H. Arsikere, J. R. Morton, G. Leung, M. S. Sommers, and A. Alwan, "Analysis and automatic estimation of children's subglottal resonances," in *Proc. of Interspeech*, 2011, pp. 2817–2820.

[16] H. Arsikere, S. M. Lulich, and A. Alwan, "Automatic estimation of the first subglottal resonance," *J. Acoust. Soc. Am. (Express Letters)*, vol. 129, pp. 197–203, 2011.

[17] S. M. Lulich, "Subglottal resonances and distinctive features," *Journal of Phonetics*, vol. 38, pp. 20–32, 2010.

[18] K. Sjölander, "The Snack sound toolkit," *KTH, Stockholm, Sweden (Online: http://www.speech.kth.se/snack/)*, 1997.