AGE- AND GENDER-DEPENDENT ANALYSIS OF VOICE SOURCE CHARACTERISTICS

Markus Iseli, Yen-Liang Shue, and Abeer Alwan

University of California Los Angeles
Dept. of Electrical Engineering
405 Hilgard Ave., Los Angeles, CA 90095
iseli@icsl.ucla.edu, yshue@ee.ucla.edu, alwan@ee.ucla.edu

ABSTRACT

The effects of age, gender, and vocal tract configurations on the glottal excitation signal are still only partially understood. In this paper we examine some of these effects, and show that the voice source parameters, such as fundamental frequency (F_0), open quotient (related to $H_1^* - H_2^*$), and spectral tilt (related to $H_1^* - A_3^*$), are not only affected by age and gender but are also intercorrelated (the asterisk superscript denotes correction for the influence of various formants). Recordings of 92 male and female speakers from three age groups (8, 15, 20-39) are analyzed. The main observations are: for low-pitched talkers $H_1^* - H_2^*$ (hence, the open quotient) is proportional to F_0, while for high-pitched talkers $H_1^* - H_2^*$ is proportional to F_1 (high to low vowels) for $F1 < 700$ Hz. The parameter $H_1^* - A_3^*$ showed a strong dependence on F_2 and F_3 for all talkers and age groups: increasing F_2 or F_3 yielded an increase in $H_1^* - A_3^*$. Spectral tilt was seen to be vowel dependent and for male talkers, spectral tilt changed dramatically with age. A better understanding of the dependencies of voice source parameters on age and gender will help improve voice source parameter estimation and analysis for a variety of speech processing and medical applications.

1. INTRODUCTION

The acoustic characteristics of the glottal excitation signal have been shown to be gender dependent [1] and are believed to change with age. A better understanding of these age- and gender-dependencies will help improve voice source estimation and analysis for a variety of speech processing and medical applications.

A study of the effects of age on speech acoustics was presented in [2]. Amongst other things, it analyzed the fundamental frequency (F_0) and formant frequencies for 10 monothong vowels from a relatively large database [3] with about 490 subjects in the age range of 5 - 50 years old. The study shows that children have higher F_0 and formant frequencies, and greater temporal and spectral variability than adults. These findings are attributed to vocal-tract anatomical differences and possible differences in the ability to control speech articulators.

Another study compares the effects of gender on voice source parameters for about 21 adult male and female talkers for the three vowels /eh/, /ae/, and /ah/ [1]. The main voice source parameters studied were open quotient (OQ) and spectral tilt (SL). The study shows that OQ and SL are generally higher for female than for male talkers.

The focus of this paper is on the analysis of voice source characteristics (F_0, OQ and SL). Age-, gender-, and vowel dependencies are evaluated on the CID database [3] with the 5 vowels /iy/, /ih/, /eh/, /ae/, and /uw/. The paper is organized as follows: Section 2 presents the statistics of the data in terms of the age and gender of the talkers. The voice source parameters and their estimation methods are described in Section 3. Results are then presented and discussed in Section 4. Section 5 concludes the paper.

Work supported in part by the NSF.

2. SPEECH DATA

Speech signals recorded from 92 people (54 males, 38 females) in three age groups, ages 8, 15, and 20–39, from the CID database [3] were analyzed. The carrier sentence was "I say uh, bVt again", where the vowel was /ih/ (bit), /eh/ (bet), /ae/ (bat), and /uw/ (boot). 'uh' is used before the target word to maximize vocal tract neutrality. The corner vowel /iy/ in 'bead' was also analyzed. Most utterances were repeated twice by each speaker. No pronunciation instructions were given to the speakers beforehand. In total, 879 utterances were analyzed. The sampling frequency was 16 kHz. The distribution of analyzed talkers (males/females) was: 25/11 (age 8), 11/11 (age 15), and 18/16 (ages 20–39).

3. ANALYSIS AND PARAMETER DESCRIPTION

The voice source parameters F_0, $H_1^* - H_2^*$, and $H_1^* - A_3^*$ were estimated. These parameters are of significant importance in the areas of voice perception and voice synthesis ([4] and [5]). $H_1^* - H_2^*$, the difference between the spectral magnitudes of the first 2 source harmonics, is related to the open quotient (OQ) [5]. $H_1^* - A_3^*$, the difference between the spectral magnitudes of the first harmonic and the third formant peak, is related to the spectral tilt [5]. The asterisk denotes that the corresponding spectral magnitudes (H_1, H_2, A_3) have been corrected for the effect of the first and second formants (F_1 and F_2) with the formula described in [6]. This formula has no restrictions on formant locations. For comparison, A_3^* additionally was normalized to a neutral vowel, as was done in [1]. The calculation of the three parameters requires the estimation of the first 3 formant frequencies (F_1, F_2, F_3), the bandwidths B_1 and B_2, and F_0. Formant frequencies F_1, F_2, and F_3, as well as F_0 were estimated using the "Snack Sound Toolkit" software [7] with these settings: the pre-emphasis coefficient was 0.9, the length of the analysis window was 25 ms, and the window shift was 10 ms. The amplitudes H_1, H_2 and A_3 were extracted from the signal spectrum using values of F_0 and F_3 as reported by Snack. Since the Snack bandwidth estimates were sometimes too high, bandwidths were calculated from their corresponding formant frequency applying the formula in [8]. This reduced the bandwidth variance and therefore the variance of results depending on bandwidth. Analysis segments were chosen at the steady-state part of the vowel, where the context-influence was small.

The estimates of F_0, F_1, F_2, and F_3 were manually checked for every utterance. The number of formant estimate corrections in percent, for 8 year old children, was: 86% for /iy/, 44% for /eh/, 32% for /ih/, and 2% for /uw/. Most formant estimation errors occurred with children speech. With /iy/, Snack typically allocated 2 formants to the first spectral peak resulting in a much lower 2nd formant frequency. In addition to the mis-indentification of F2, there were 3 utterances of /uw/ spoken by 8 year old females which needed an adjustment of F_0. The formant values for the vowels are not listed in this paper as the results are similar to what was reported in [2].

4. RESULTS

In this section, we will refer to high-pitched talkers (age 8 both genders, females 15 and older, with usually $F_0 > 175$ Hz) as "group 1" and to low-pitched talkers (males 15 and older, with usually $F_0 < 175$ Hz) as "group 2". The 3 source parameters, F_0, $H_1^* - H_2^*$ and $H_1^* - A_3^*$ are evaluated as a function of age, gender, formant frequencies and vowel type.

4.1. $\mathbf{F_0}$

Table 1 shows the range of F_0 values. Note that F_0 changes significantly with age (by about 130 Hz) for male talkers, while the change is less dramatic for female talkers (about 50 Hz). This agrees with the results in [2]. We noticed that the very high F_0 values (above 300 Hz) are due to high lexical stress on the target word. In those cases F_0 was around 300 Hz for the rest of the sentence, but increased for the target word.

Table 1. *Min/Mean/Max of F_0 (in Hz) per age group.*

Age	F_0 males	F_0 females
8	182/255/422 Hz	181/281/419 Hz
15	95/124/248 Hz	179/228/303 Hz
20–39	87/127/189 Hz	158/233/335 Hz

Average F_0 values are highest for /uw/, and higher for /iy/ than for /eh/ and /ae/. The trend of increasing F_0 as the tongue moves from a front to a back position and from open to closed vowels, has been described for German talkers in [9]. This trend can be seen for all ages and genders for the vowels in this study and may partly be explained by vowel-dependent intrinsic pitch [10]. Note that F_0 was not normalized for lexical stress.

Fig. 1 shows F_1 versus F_0 for group 1. Interestingly, F_1 is often close to an integer multiple of F_0 (while it is not for group 2). This effect could be due to a combination of source-vocal tract interaction and LPC formant estimation errors. The former effect has been shown in [11] to be more pronounced for high pitched voices and the latter effect is described in [12].

4.2. $\mathbf{H_1^* - H_2^*}$

4.2.1. $\mathbf{H_1^* - H_2^*}$ *vs.* $\mathbf{F_0}$

The effects of age and gender on $H_1^* - H_2^*$ are shown in Figs. 2–4. Comparing the histogram distributions, it is interesting to observe that the $H_1^* - H_2^*$ separation between genders is the clearest at age 15. The mean $H_1^* - H_2^*$ value drops by about 3 dB for male talkers between ages 8 and 20–39, whereas for female talkers it remains relatively unchanged.

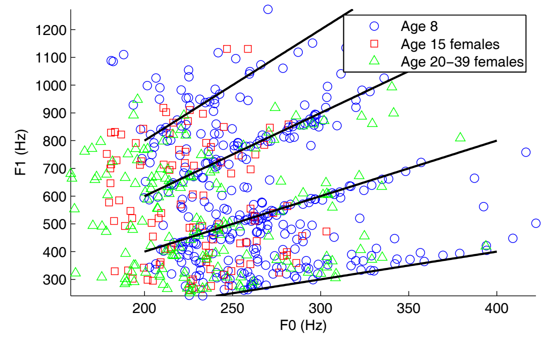

Fig. 1. Relation between F_1 and F_0 for 8 year old children, and females ages 15 and 20–39. The lines correspond to $F_1 = kF_0$ with $k = 1, \ldots 4$.

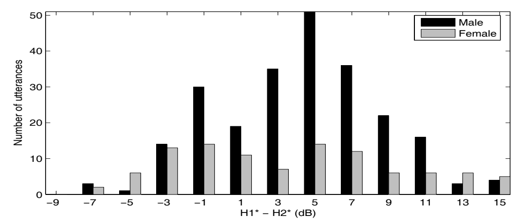

Fig. 2. Histogram of $H_1^* - H_2^*$ values at age 8. The mean for males is at 4.3 dB and for females, it is at 3.7 dB.

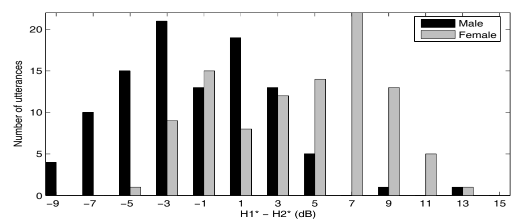

Fig. 3. Histogram of $H_1^* - H_2^*$ values at age 15. The mean for males is at -1.5 dB and for females, it is at 4.1 dB.

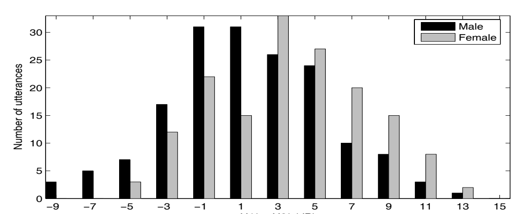

Fig. 4. Histogram of $H_1^* - H_2^*$ values at age 20-39. The mean value for males is at 1.5 dB, and for females it is at 3.5 dB. Compared with Fig. 3, there is a greater overlap between male and female talkers.

The difference between genders may be attributed to the fact that F_0 drops significantly between age 8 and 15 for males while it does not change as much for females [2].

Fig. 5 shows the relationship between $H_1^* - H_2^*$ and F_0 for both groups. In general $H_1^* - H_2^*$ seems to be higher for high F_0. It

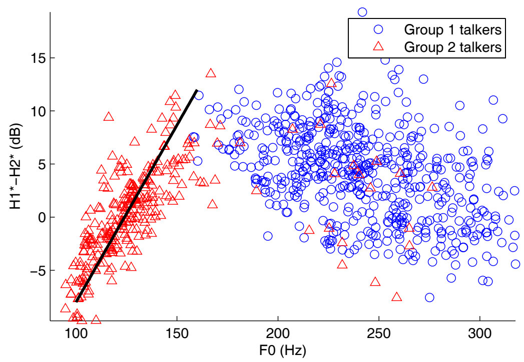

Fig. 5. Relation between $H_1^* - H_2^*$ and F_0 for high-pitched (group 1) talkers and low-pitched (group 2 talkers). A linear relationship for F_0 between 80 and 175 Hz is observed: see Eq. 1.

has been observed in [13] that increased tension of the cricothyroid muscle in the larynx induces a simultaneous increase of F_0 and OQ, and therefore also of $H_1^* - H_2^*$. However, we observed a *linear* relationship only for low F_0 values. The Pearson product correlation (PPC) between $H_1^* - H_2^*$ and F_0 yields a value of 0.56 for group 2 (low-pitched talkers). An approximate mapping is:

$$H_1^* - H_2^* \approx \frac{1}{4}F_0 - 32 \qquad \text{for } F_0 \text{ between 80–175 Hz} \qquad (1)$$

4.2.2. $\mathbf{H_1^* - H_2^*}$ *vs.* $\mathbf{F_1}$

Fig. 6 shows the relationship between $H_1^* - H_2^*$ and F_1 for group 1 (high-pitched talkers). $H_1^* - H_2^*$ increases simultaneously with F_1 when F_1 is less than 700 Hz; the PPC is 0.61. As F_1 increases, related to shifting the tongue from a high to a low position, $H_1^* - H_2^*$ increases by about 10 to 15 dB. No such relationship can be seen for F_1 values above 700 Hz nor was it observed for group 2 speakers.

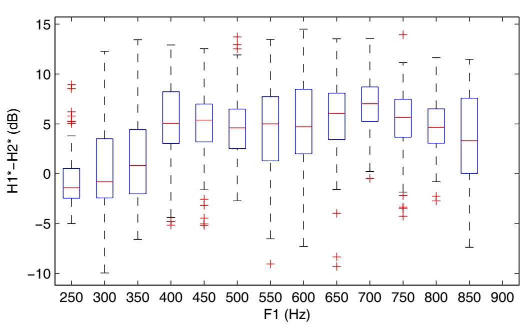

Fig. 6. Relation between $H_1^* - H_2^*$ and F_1 for high-pitched talkers (group 1). $H_1^* - H_2^*$ monotonically increases, on average, by about 10-15 dB when F1 increases between 250–700 Hz. The boxes start at the first quartile of the data and end at the third quartile. The line in the box denotes the median value and the whiskers represent the data range.

Fig. 7 depicts $H_1^* - H_2^*$ as a function of vowel for high-pitched talkers. The values for /iy/ and /uw/ are the lowest compared to the other vowels, which would confirm that high vowels have low OQ. As F_1 increases for vowels /uw/, /ih/, /eh/, and /ae/, it can be seen that the trend in $H_1^* - H_2^*$ is consistent with Fig. 6. For low-pitched talkers, on the other hand, no significant correlations between the $H_1^* - H_2^*$ values and vowel height could be observed.

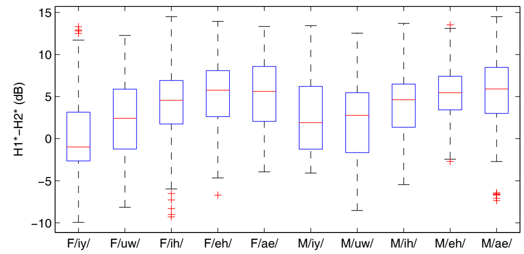

Fig. 7. $H_1^* - H_2^*$ as a function of vowel for high-pitched talkers. 'F' and 'M' denotes the values for female and male talkers respectively. Note the low values for the high front vowel /iy/ (compare to Fig. 6).

The lack of significant trends of $H_1^* - H_2^*$ values with low-pitched talkers may be due to the physiology associated with voice production in different genders. Another study [14] utilizing electroglottography (EGG) of Zapotec speakers showed that females produce phonation differences by altering OQ while males don't. This is a possible explanation for the trends in Fig. 6 and 7 only appearing for high-pitched talkers.

4.3. $\mathbf{H_1^* - A_3^*}$

The age and gender effects on $H_1^* - A_3^*$ (spectral tilt) are shown in Table 2 as mean/standard deviation pairs. Values for adults (20–39) presented in [1] are in parentheses.

The mean $H_1^* - A_3^*$ value drops for male talkers by about 20 dB between ages 8 and 20–39, whereas for female talkers it drops by about 8 dB.

Table 2. *Mean/standard deviation in dB for $H_1^* - A_3^*$ at ages 8, 15, and 20–39. For comparison, parameters from [1] are given in parentheses. Large differences from [1] are in bold.*

Gender	8	15	20–39
$H_1^* - A_3^*$			
M	33.8/8.5	15.5/7.5	13.0(13.8)/**8.5**(4.8)
F	31.0/9.9	20.4/7.9	23.5(23.4)/**9.1**(6.6)

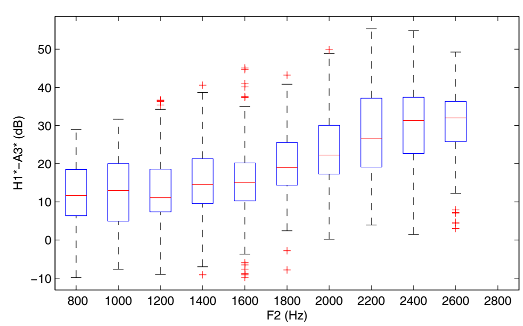

Fig. 8. Relation between $H_1^* - A_3^*$ and F_2 for all talkers.

Compared to [1], for adults, our standard deviation for $H_1^* - A_3^*$ is significantly higher. This may be explained by the inclusion of five vowels in our analysis, while [1] studied three vowels.

The PCC between $H_1^* - A_3^*$ and F_0 was very small for low-pitched talkers (0.22) and zero for high-pitched talkers. A similar relationship was observed with F_1.

A much stronger correlation was observed with F_2: PPC is 0.61. This is shown in Fig. 8 which shows $H_1^* - A_3^*$ gradually increasing for $F_2 > 1600\ Hz$.

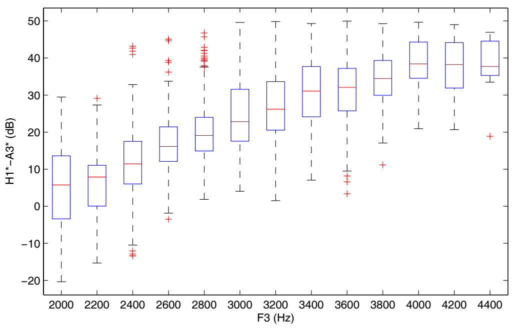

Fig. 9. Relation between $H_1^* - A_3^*$ and F_3 for all talkers. A steady rise of the parameter can be seen for increasing F_3

In Fig. 9 $H_1^* - A_3^*$ is plotted as a function of F_3 for all talkers (PCC was 0.6). It shows that, with increasing F_3, $H_1^* - A_3$ increases monotonically until $F_3 > 3800\ Hz$ where the plot flattens out. An explanation for this effect around 4 kHz could be the presence of higher formants, such as F_4, for which the parameter was not corrected, and which would boost the value of A_3 when evaluated close to F_4. However, a physiological explanation is still missing.

$H_1^* - A_3^*$ is depicted in Fig. 10 as a function of vowel. It can be seen that values for $H_1^* - A_3^*$ decrease from /iy/, to /uw/. This result is consistent with the plots shown in Figs. 8 and 9. As F_2 and F_3 are lowest for /uw/, the corresponding $H_1^* - A_3^*$ values are also at the minimum of the five vowels. This trend was observed for all three age groups.

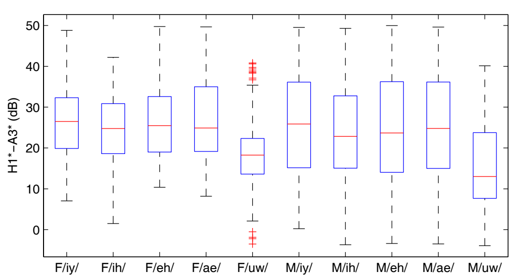

Fig. 10. $H_1^* - A_3^*$ as a function of vowel for all talkers. /iy/ has the highest value, while /uw/ has the lowest value. This could be related to the dependence of the parameter on F_2 and F_3.

5. SUMMARY AND CONCLUSIONS

In this paper, we examined three voice source parameters: fundamental frequency (F_0), $H_1^* - H_2^*$, and $H_1^* - A_3^*$ for their dependence on age, gender, and vocal tract configuration. For low-pitched talkers $H_1^* - H_2^*$ (hence, the open quotient) is proportional to F_0 while for high-pitched talkers, it is proportional to F_1 for $F1 < 700$ Hz. The parameter $H_1^* - A_3^*$ (hence spectral tilt) showed a strong dependence on F_2 and F_3 for all talkers and age groups: increasing F_2 or F_3 yielded an increase in $H_1^* - A_3^*$. Hence spectral tilt is vowel dependent. Future work will examine the effect of context and prosody on voice source parameter estimates.

6. REFERENCES

[1] H. M. Hanson and E. S. Chuang, "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," *J. Acoust. Soc. Am.*, vol. 106, pp. 1064–1077, 1999.

[2] S. Lee, A. Potamianos, and S. Narayanan, "Acoustics of childrens speech: Developmental changes of temporal and spectral parameters," *J. Acoust. Soc. Am.*, vol. 105, no. 3, pp. 1455–1468, March 1999.

[3] J. Miller, S. Lee, R. Uchanski, A. Heidbreder, B. Richman, and J. Tadlock, "Creation of two children's speech databases," in *Proceedings of ICASSP*, vol. 2, May 1996, pp. 849–852.

[4] G. Fant and A. Kruckenberg, "Voice source properties of speech code," *TMH-QPSR. Report 4*, pp. 45–56, 1996.

[5] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. Guiod, and S. L. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *J. Speech Hear. Res.*, vol. 38, pp. 1212–1223, 1995.

[6] M. Iseli and A. Alwan, "An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation," in *Proceedings of ICASSP*, vol. 1, Montreal, Canada, May 2004, pp. 669–672.

[7] K. Sjölander, "Snack sound toolkit," KTH Stockholm, Sweden, 2004, http://www.speech.kth.se/snack/.

[8] R. H. Mannell, "Formant diphone parameter extraction utilising a labelled single speaker database," in *Proceedings of the ICSLP*, vol. 5. Sydney, Australia: ASSTA, 1998, pp. 2003–2006.

[9] K. Marasek, "Glottal correlates of the word stress and the tense-lax opposition in German," in *Proceedings ICSLP*, Philadelphia, PA, Oct. 1996, pp. 1573–1576.

[10] I. Lehiste and G. E. Peterson, "Some basic considerations in the analysis of intonation," *J. Acoust. Soc. Am.*, vol. 33, no. 4, pp. 419–425, April 1961.

[11] H. Hatzikirou, W. T. Fitch, and H. Herzel, "Voice instabilities due to source-tract interactions," in *International Conference on Voice Physiology and Biomechanics - Modeling Complexity*, Marseille, 2004, pp. 63–70.

[12] G. Vallabha and B. Tuller, "Systematic errors in the formant analysis of steady-state vowels," *Speech communication*, vol. 38, pp. 141–160, 2002.

[13] J. Koreman, "Decoding linguistic information in the glottal airflow," Ph.D Thesis, University of Nijmegen, 1996.

[14] C. Esposito, "An acoustic and electroglottographic study of phonation in Santa Ana del Valle Zapotec," Poster at the 79th meeting of the Linguistic Society of America, 2005, 2005.