ANALYSIS OF A VOWEL DATABASE

Peter F. Assmann[1], Terrance M. Nearey[2] and Sneha Bharadwaj[1]
[1]School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson TX 75083-0688 USA
[2]Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada T6G 2E7

1. INTRODUCTION

To study developmental changes in the acoustic properties of vowels, we have collected a database of recordings from adults and children from the Dallas, Texas region. Across age and gender classes, we find a systematic relationship between the geometric mean of the formant frequencies (F1-F3) across all of the vowels for a given speaker (a measure related to vocal tract length) and the geometric mean fundamental frequency (f_0, reflecting the rate of vocal fold vibration). The aim of this study is to provide a preliminary analysis of the developmental changes in these acoustic properties.

2. VOWEL DATABASE

2.1 Speakers

To date we have recorded 163 children and 39 adults. The children ranged in age from 5 to 18 years of age, with at least 5 boys and 5 girls at each age. The adults were undergraduate students at the University of Texas at Dallas ranging in age from 19-45 years. All spoke English as their first language and were long-term residents of the North Texas region.

2.2 Vowels

Vowels were recorded in hVd words, both in isolation and in a carrier sentence, "Please say the word __ again." The recordings included the 12 monophthongal vowels of American English, /ɪ/, *heed*; /I/, *hid*; /ɛ/, *hayed*; /E/, *head*; /Θ/, *had*; /ɷ/, *hud*; /™/, *herd*; /A/, *hod*; /□/, *hawed*; /o/, *hoed*; /Y/, *hood*; /ʊ/, *who'd;* and three diphthongs, /Aɪ/, *hide*; /Aʊ/, *how'd*; and /□ɪ/, *hoyed*.

2.3 Recording procedure

Recordings were carried out over a 2.5-year period at the University of Texas at Dallas. Each recorded token was produced following a screen prompt that displayed the orthographic representation of the hVd word or sentence, along with an audio example spoken by an adult female from the Dallas area. Five repetitions of each of the 12 vowels were elicited from each of the children, and at least 10 repetitions were collected from each adult. Recordings were made in a sound-treated room using a Shure SM-94 microphone, Symetrix SX202 dual-microphone pre-amplifier and Tucker-Davis Technologies data acquisition hardware (MA1, RP2.1). The digital waveforms were stored on computer disk at a rate of 48,828 Hz and 16-bit resolution. Following data collection, each recording was screened by listening and tokens judged to be misarticulated or of low recording quality were omitted from subsequent analysis.

2.4 Acoustical analysis

A semi-automated procedure was used to mark the onset and offset of the vowel in the isolated hVd words. Vowel onset was defined as the beginning of the first pitch period of the voiced segment of the syllable. Vowel offset was defined as the end of the last pitch period before the silent interval created by the final /d/ or by a substantial drop (>15 dB) in the levels of the higher formants (F2-F5) in cases where pre-voicing filled the silent interval. Formant frequencies were measured every 2 ms using an automatic formant tracking program[1]. Fundamental frequency was estimated with 1-ms resolution using a procedure developed by Kawahara.[2]

3. RESULTS

Figure 1 plots the geometric mean f_0 across all of the frames, tokens, vowels, and talkers as a function of age and sex. Consistent with earlier findings[3], there is a progressive decline in f_0 as a function of age. For males there is a sharp break around age 13 and an increase in variability at ages 13 and 14, while females show a more gradual decline up to age 18.

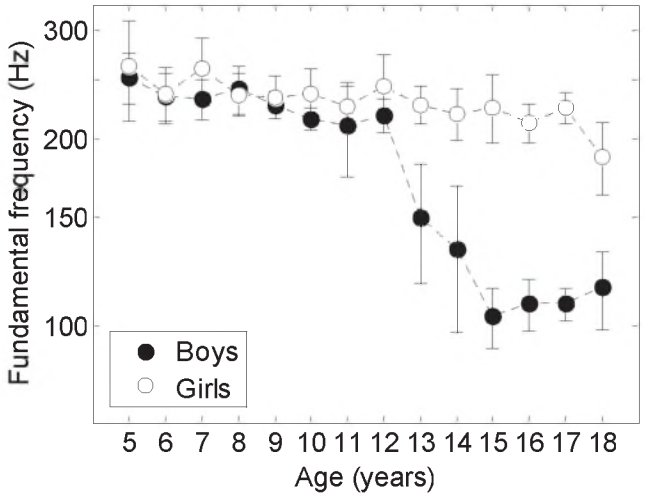

Fig. 1. Geometric mean f_0 averaged across frames, tokens, vowels and talkers. Error bars show standard errors across talkers.

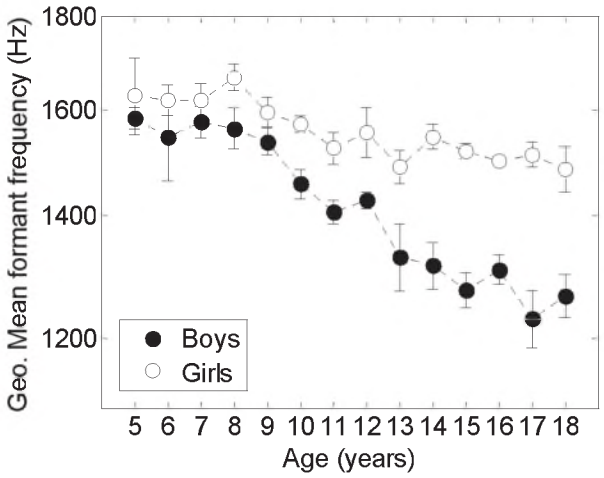

Fig. 2. Geometric mean of the formant frequencies F1, F2, and F3 averaged across frames, tokens, vowels and talkers. Error bars show standard errors across talkers.

Figure 2 plots the geometric mean of the formant frequencies (F1, F2, and F3) across all of the frames, tokens, vowels, and talkers as a function of age and sex. This measure is related to vocal tract length and also shows the predicted decline as a function of age. It is noteworthy that differences between boys and girls in mean formant frequencies appear as early as age 8, consistent with findings of other recent studies[4].

The scatterplot in Figure 3 shows that the geometric mean f_0 increases linearly with the geometric mean of the formant frequencies when the data for males and females are combined. The relationship is weaker for females alone (r=.46) compared to males alone (r=.87). It also appears to be weaker in younger children.

4. DISCUSSION

In a recent study[5] we have investigated the perceptual effects of upward and downward scaling of the formants in combination with changes in f_0. The results indicate that coordinated shifts (formant frequencies and f_0 scaled in the same direction) resulted in higher identification accuracy than uncoordinated shifts (formants and f_0 scaled in opposite directions). One explanation is that listeners are sensitive to the statistical covariation illustrated in Figure 3.

REFERENCES

[1] Nearey, T.M., Assmann, P.F. and Hillenbrand, J.M. (2002). Evaluation of a strategy for automatic formant tracking. *J. Acoust. Soc. Am.* 112: 2323.

[2] Kawahara H., de Cheveigné A., Banno H., Takahashi T. and Irino T. (2005). Nearly Defect-free F0 Trajectory Extraction for Expressive Speech Modifications based on STRAIGHT, Proc. Interspeech2005, Lisboa, pp. 537-540, Sept. 2005.

[3] Lee S., Potamianos A., Narayanan S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. J. Acoust. Soc. Am. 105: 1455-1468.

[4] Perry T.L., Ohde R.N. and Ashmead, D.H. (2001). The acoustic bases for gender identification from children's voices. J. Acoust. Soc. Am. 109: 2988- 2998.

[5] Assmann, P.F. and Nearey, T.M. (2008). Identification of frequency-shifted vowels. Submitted to *J. Acoust. Soc. Am.*

ACKNOWLEDGMENTS

This research was funded by NSF and SSHRC.

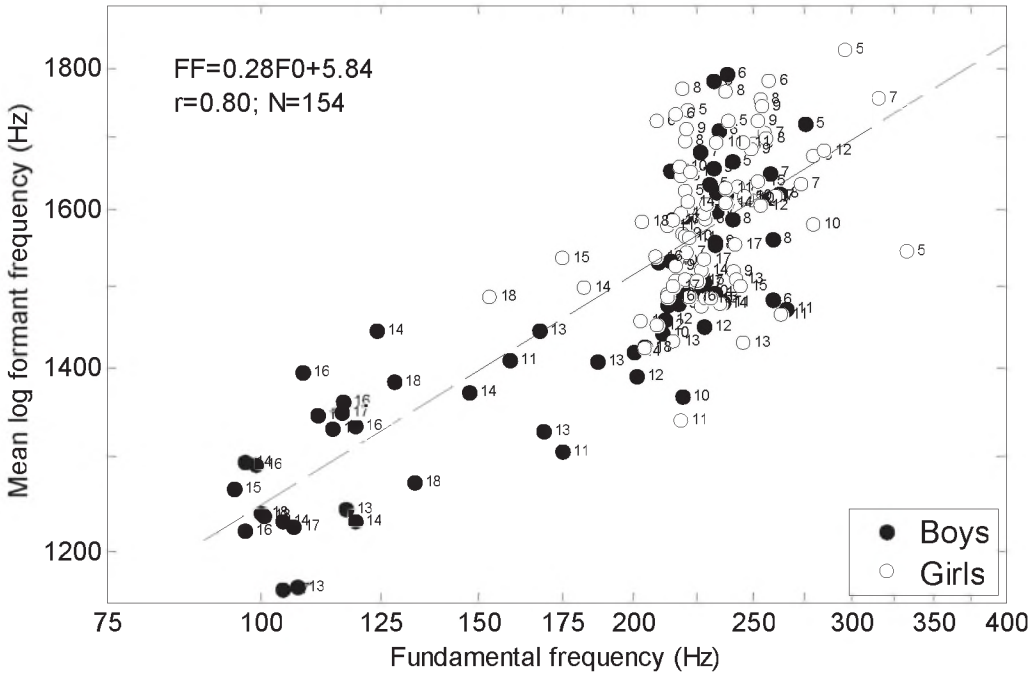

Fig. 3. Geometric mean of the formant frequencies F1, F2, F3 against the geometric mean f_0, both averaged across frames, tokens, and vowels separately for each of the children in the database. The numbers next to each circle indicate the talker's age.