# PERCEPTION OF SPEAKER AGE IN CHILDREN'S VOICES

## Peter F. Assmann[1], Terrance M. Nearey[2]
[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

Children's speech differs from adult speech in several respects. Their vocal tracts are shorter, leading to higher formant frequencies, and their larynges are smaller, resulting in higher average fundamental frequency (f0). These properties are linked to phonetic properties of speech but also affect indexical properties (such as the age, sex, and size of the speaker). As part of a larger study to investigate the interaction between indexical and phonetic properties we presented a sample of speech sounds spoken by children ranging in age from 5 to 18 years to adult listeners and asked them to judge the age of the speaker.

### 1.1. Background

Previous studies have shown that listeners can judge the age of a speaker fairly accurately in adult voices (Ptacek and Sander, 1966), but few studies have examined children's voices. One exception is a recent study by Amir et al. (2012) who investigated age and gender recognition of a sample of Hebrew-speaking children between 8 and 18 years. Age recognition accuracy (±2 years) was fairly low (40% for sentences, 35% for isolated vowels) with the lowest performance for the oldest group where there was a tendency to systematically underestimate the perceived age in female voices. Gender recognition was fairly accurate (85% for sentences, 78% for vowels) but again showed lower accuracy for older girls compared to older boys.

### 1.2. Research questions

The aim of the present study was to replicate and extend the research reported by Amir et al. (2011) with a sample of American English speaking children to answer the following questions: (1) *Does knowing the sex of the speaker help determine their age?* We previously found (Assmann and Nearey, 2011) that providing information about the age of the speaker provides a small benefit, in some conditions, for determining whether the speaker is male or female. Here we ask if knowledge of the speaker's sex also provides a benefit when listeners are asked to judge the speaker's age. (2) *To what extent is perceived age influenced by context?* Previous studies (Hillenbrand and Clark, 2009; Assmann and Nearey, 2011) have shown that listeners can identify the sex of the speaker more accurately from sentences than from single syllables. Amir et al. (2011) found higher accuracy for the perception of age in children's voices based on complete sentences rather than isolated vowels. In the present study we compared isolated /hVd/ syllables with the same syllables embedded in a fixed carrier sentence ("Please say the word _____ again").

## 2. METHODS

Stimuli

The stimuli were recorded syllables and sentences drawn from a vowel database (Assmann et al., 2008) of 208 children ranging in age from 5 to 18 years. In the syllable condition, 140 speakers (5 boys and 5 girls at each age level) contributed 3 syllables (heed, hod, and who'd) for a total of 420 stimuli. In the sentence condition (tested with a separate group of listeners) a subset of 84 of speakers was included, for a total of 252 stimuli (3 boys and 3 girls at each age level, each speaking the same 3 syllables in a carrier sentence) to keep the experiment to a reasonable length.

Participants

Separate sets of 20 listeners completed the syllable and sentence conditions; of these, 10 were provided with age information prior to responding, and 10 were not. The listeners were undergraduate students at the University of Texas at Dallas, native speakers of American English with normal hearing who received experimental credits for their participation. Prior to the experiment they completed a hearing screen and a questionnaire to provide information along their age, sex, younger siblings and exposure to children's voices on a daily basis.

Procedure

Stimuli were presented monaurally using earphones with Tucker-Davis System 3 and RP2.1 hardware. All stimulus conditions were randomly interspersed. Listeners used a graphical slider to register their estimate of the speaker's age, after which they checked one of five buttons indicating their confidence level. The experiment was self-paced, with an optional break in the middle, and lasted about 50 minutes.

## 3. RESULTS

Age judgments showed a fairly close match to chronological age across the age range for boys, but progressively underestimated chronological age for older girls, with the discrepancy reaching nearly four years on average by age 18. This is consistent with findings reported by Amir et al. (2012). Informing listeners about the sex of the speaker reduced the discrepancy slightly, although this

information did not lead to substantially improved age estimation. Figure 1 shows that perceived age judgments (pooled across talkers and listeners) were well fit by a quadratic function ($R^2$=0.93) although age judgments for boys' voices were close to a linear function of age. When /hVd/ syllables were presented in a carrier sentence, perceived age was closer to chronological age and the degree of underestimation in older girls' voices was reduced, though not eliminated entirely.
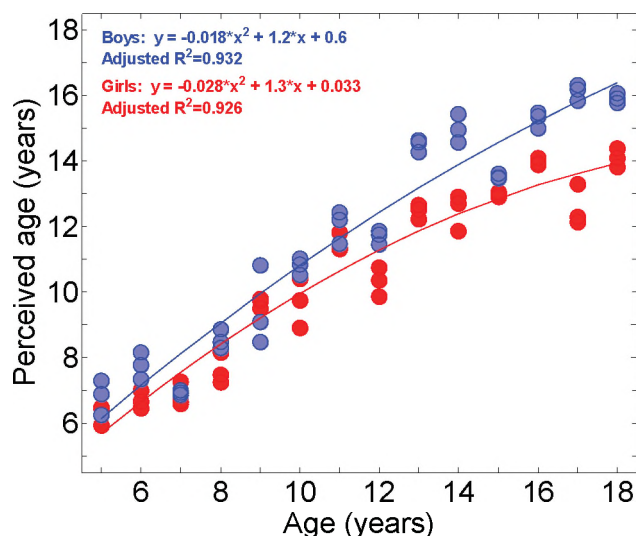


**Boys: y = -0.018*x$^2$ + 1.2*x + 0.6**
**Adjusted R$^2$=0.932**

**Girls: y = -0.028*x$^2$ + 1.3*x + 0.033**
**Adjusted R$^2$=0.926**

**Figure 1. Perceived versus chronological age (/hVd/ syllables).**

Accuracy of perceived age was quantified in terms of the absolute deviation of perceived age from chronological age. Analysis of variance revealed a significant effect of context, with more accurate age estimates from sentences compared to syllables, $F(1,40)$=4.99; $p<.05$. There was a significant interaction of context, sex and age, $F(13,520)$=8.97; $p<.01$, reflecting an increased benefit of context for older girls whose ages were often underestimated. Gender information did not lead to an overall improvement, but the interaction with sex, age and context was significant, $F(13,520)$=2.92; $p<.01$, with slightly improved estimates for some age/sex combinations when gender information and context were both available.

## 4. DISCUSSION AND CONCLUSIONS

The results provide further support for several findings reported by Amir et al. (2012). First, age estimates are more accurate for boys' voices than for those of girls. Second, there is a consistent trend for listeners to underestimate the ages of older girls. It is unlikely that the discrepancy is caused by a tendency for listeners to avoid the extreme responses on the slider, since the pattern is present only at the high end of the scale and only for girl's voices. Moreover, Amir et al. found a similar pattern when listeners were instructed to respond by selecting one of six response buttons representing the age ranges of the speakers.

The pattern in Figure 1 suggests that listeners can gauge the age of children's voices with relatively high accuracy. However, it should be noted that each data point is averaged across 5 talkers and 24 listeners. The inclusion of talkers and listeners introduces additional sources of variability in the data. Fitting a quadratic function to the entire dataset provides similar coefficients but reduced $R^2$ (0.66 for boys and 0.56 for girls). We are currently investigating acoustic factors that might account for variability across talkers and demographic factors – for example, familiarity and exposure to children's voices at different ages – to account for listener variability.

Previous findings have indicated that mean f0 and formant frequencies as well as durational properties provide important cues for the perception of speaker age (Linville, & Fisher, 1985; Harnberger et al., 2006). Studies with vocoded speech have indicated that upward scaling of the frequencies of the formants in a male voice, when coupled with an appropriate increase in f0, can raise the probability that the voice will be perceived as female. Similarly, downward scaling of the formants and f0 in a female voice raises the probability that the voice will be perceived as male. However, a change in F0 or formants alone is generally not sufficient to produce a compelling conversion of speaker sex (Hillenbrand & Clark, 2009). We are currently investigating whether a similar pattern holds for the perception of speaker age, and we are investigating properties of children's voices that might help to explain the discrepancy between perceived and chronological age in female voices and the effects of context.

## REFERENCES

Amir, O., Engel, M., Shabtai, E., & Amir, N. (2012). *Identification of children's gender and age by listeners.* Journal of Voice 26(3), 313-21. Epub 2011 Aug 12.

Assmann, P.F., Nearey T.M. & Bharadwaj, S. (2008). *Analysis and classification of a vowel database.* Canadian Acoustics 36(3), 148-149.

Assmann P.F. & Nearey T.M. (2011). *Perception of speaker sex in children's voices.* Journal of the Acoustical Society of America 130(4), 2446(A).

Harnsberger, J. D., Shrivastav, R., Brown, Jr., W. S., Rothman, H. & Hollien, H. (2006). *Speaking rate and fundamental frequency as speech cues to perceived age.* Journal of Voice 22, 58-69.

Hillenbrand, J. M. & Clark, M. J. (2009). *The role of f0 and formant frequencies in distinguishing the voices of men and women.* Attention, Perception, & Psychophysics 71, 1150-1166.

Linville, S. E., and Fisher, H. B. (1985). *Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females* Journal of the Acoustical Society of America 78, 40–8.

Ptacek, P. H., & Sander, E. K. (1966). *Age recognition from voice.* Journal of Speech and Hearing Research 9(2), 273-277.