Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 2aSC: Linking Perception and Production (Poster Session)

2aSC15. Perception of speaker age in children's voices

Peter Assmann*, Santiago Barreda and Terrance Nearey

***Corresponding author's address: Cognition and Neuroscience, University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080-3021, assmann@utdallas.edu**

To study the perception of speaker age in children's voices, adult listeners were presented with vowels in /hVd/ syllables, either in isolation or in a carrier sentence. Listeners used a graphical slider to register their estimate of the speaker's age. The data showed a moderate correlation of perceived age and chronological age. For isolated syllables, age estimation accuracy was fairly constant across age up to about age 11, but there was a systematic tendency for listeners to underestimate the ages of older girls. This error pattern was actually exaggerated when listeners were informed of the speaker's sex. Age estimation accuracy was higher for syllables embedded in a carrier sentence, and knowledge of the speaker's sex had little effect. Linear regression analyses were conducted using acoustic measurements of the stimuli to predict perceived age. These analyses indicated significant contributions of fundamental frequency, duration, vowel category, formant frequencies as well as certain measures related to the voicing source. The persistent underestimation of age for older girls, and the effect knowledge of speaker sex has on this underestimation suggest that acoustic information is combined with expectations regarding speakers of a given sex in arriving at an estimate of speaker age.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800918]
Received 4 Feb 2013; published 2 Jun 2013

INTRODUCTION

When attending to a recording of an unfamiliar voice, listeners form an immediate impression of the speaker's sex, age, and physical size, along with other personal attributes often collectively referred to as indexical properties (Abercrombie, 1967). Information about the speaker is extracted in tandem with processing of the linguistic message; the overall aim of the present research is to study how these processes interact. The perception of age in children's voices is particularly interesting because age-related changes in the voice are correlated with substantial changes in physical size. Children's vocal tracts are shorter, leading to higher formant frequencies, and their larynges are smaller, resulting in higher average fundamental frequency (f0). These properties determine the phonetic properties of speech but also affect the perceived age, sex, and size of the speaker. As part of a larger study to investigate the interaction between indexical and phonetic aspects we presented a sample of speech sounds spoken by children ranging in age from 5 through 18 years to adult listeners and asked them to judge the age of the speaker.

The literature on the perception of vocal age is relatively sparse. Several studies have found that listeners can estimate the age of the speaker with varying degrees of success depending on the speech material, the characteristics of the speech sample and the task (for reviews see Linville, 2001; Schötz, 2007). Some studies have asked listeners to assign voices to discrete categories or age ranges (e.g., Ptacek and Sanders, 1966) while others have used direct magnitude estimation (e.g., Harnsberger et al., 2008). In these studies, perceived age generally shows a moderate correlation with chronological age (with correlation coefficients of r=0.7 or higher in several studies). However, few studies have examined children's voices. One exception is a recent study by Amir et al. (2012) who found better than chance accuracy for age identification in a sample of speech (vowels and sentences) recorded from 120 children, including boys and girls from six age groups (ages 8, 10, 12, 14, 16, and 18). Age recognition accuracy (defined in terms of age categories spanning 2 years) was fairly low (40% for sentences, 35% for vowels) with the lowest performance for the oldest group where there was a tendency to systematically underestimate the perceived age in female voices. However, the authors noted that a large proportion of the errors involved assigning the speaker to an adjacent age category. Gender recognition was fairly accurate (85% for sentences, 78% for vowels) with overall higher accuracy for the older children, but the results showed lower accuracy for the older girls compared to the older boys. Age estimation was more accurate for sentences compared to vowels (about 5 percentage points improvement, on average).

The present study was designed to replicate and extend the findings of Amir et al. (2012) using 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, under some circumstances, for judging whether the speaker is male or female. Here we ask if information about the speaker's sex provides a corresponding benefit for judgments of perceived age. (2) *To what extent is perceived age dependent on phonetic 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. (2012) found higher accuracy for the perception of age in children's voices based on complete sentences rather than isolated vowels, sustained tokens of /a/ and /i/. In the present study we compared hVd syllables spoken in isolation with those same syllables embedded in a carrier sentence.

METHOD

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: /hid/ ("heed"), /hɑd/ ("hod"), and /hud/ ("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 ("Please say the word _____ again"). The number of speakers was reduced in the sentence condition to keep the experiment to a reasonable length.

Participants: Two separate groups of 24 listeners completed the syllable and sentence conditions; of these, 12 were provided with gender information on each trial prior to responding, and 12 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. They subsequently 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.

RESULTS AND DISCUSSION

Figure 1 shows that listeners' age judgments were fairly close matched to chronological age for boys, but underestimated chronological age for older girls. Informing listeners about the sex of the speaker did not lead to substantially improved age estimation, and in fact increased the discrepancy between perceived and chronological age for the isolated syllables, though not noticeably for the sentence condition.

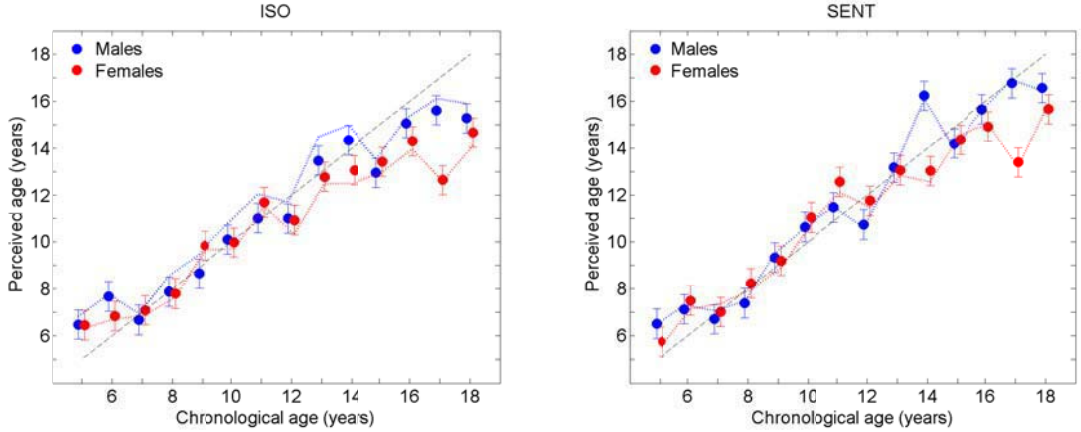

FIGURE 1. Perceived age as a function of chronological age. Circles and error bars indicate means and standard errors across listeners who were not provided with information about the speaker's sex. Dotted lines indicate mean age estimates in the condition where speaker sex information was provided (blue for boys, red for girls). The diagonal indicates perfect performance (perceived age = chronological age). The left panel shows the results for syllables in isolation; results for sentence context are shown on the right.

When /hVd/ syllables were presented in a carrier sentence, the discrepancy between perceived age and chronological age was smaller and the degree of underestimation in older girls' voices was reduced, though not eliminated entirely. The contribution of gender information was also greatly reduced in sentence context; the discrepancy between conditions with and without gender information all but disappeared.

An analysis of variance was carried out using age estimation accuracy (defined as the absolute difference between perceived and chronological age) as the dependent variable and including three within-subjects factors: age (with 14 levels, 5-18 years), sex (male, female), vowel (/i/, /ɑ/, /u/), and two between-subjects factors: gender information (provided or not provided) and context (syllable, sentence). The data were pooled across talkers. There was a significant main effect of context, with more accurate age estimates from sentences compared to syllables, $F(1, 44) = 4.20$; $p<.05$. Age estimation accuracy was relatively constant with chronological age up to 16 years but increased abruptly for 17 and 18-year old female speakers, giving rise to a significant age by sex interaction, $F(13, 572) = 67.44$; $p<.01$. Gender information did not lead to an overall improvement, but there was a significant four-way interaction of age by sex by gender information by context, $F(13, 572) = 2.63$; $p<.01$. Figure 1 shows that the provision of gender information exaggerates the tendency for listeners to underestimate the speaker's age in older girls, but only for the isolated syllables (as indicated by the deviation of dotted lines from solid circles). For the sentence condition, the dotted lines and solid circles essentially coincide, indicating that gender information had little impact on age judgments for these stimuli.

Relating Chronological and Perceived Age to Acoustic Properties

Previous studies 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; Harnsberger et al., 2006). Lee et al. (1999) analyzed a large sample of children's vowels and recorded systematic changes in f0 and formant frequencies as a function of age and sex as well as increased spectral and temporal variability in younger children. Iseli et al. (2007) extended these findings to examine developmental changes in source properties, including measures of the open quotient, defined by the difference between the first and second harmonics, H1*–H2*, where the asterisk denotes spectral magnitudes "corrected" for the effects of the vocal tract transfer function (formants), and H1*–A3*, the corrected magnitude difference between the first harmonic and the third formant peak, related to source spectral tilt. Male speakers showed a drop of approximately 5 dB in H1*–H2* around age 15. Males also show a systematic decline in H1*–A3* of about 10 dB between ages 8 and 39 years while females show a smaller decline of about 4 dB. Shue and Iseli (2008) have shown that these source measures, when used in combination with formant measures and f0, can lead to improved automatic gender classification in vowels spoken by children of different ages. They noted that the contribution of these source measures declined as F0 differences between boys and girls became more prominent.

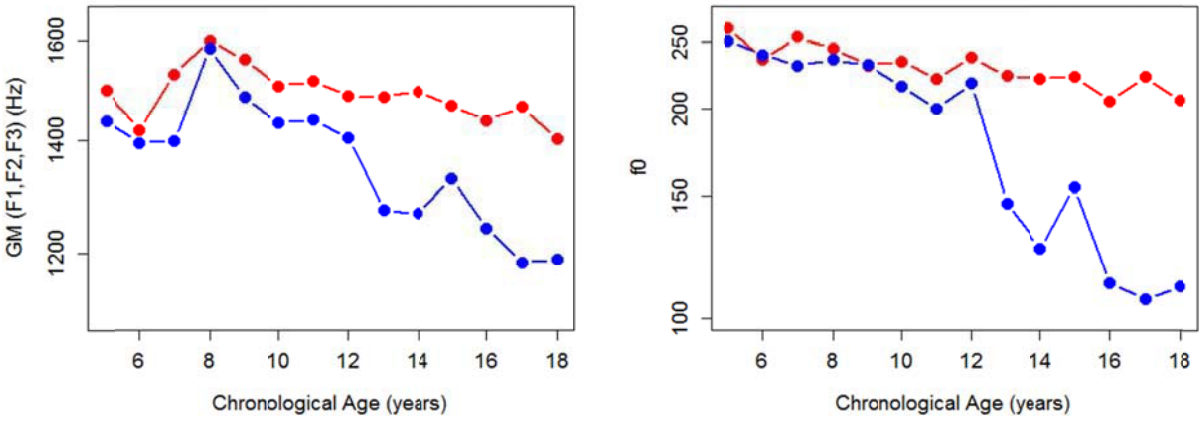

FIGURE 2. Average geometric mean (GM) of the first three formant frequencies, and average f0 for male (blue) and female (red) speakers as a function of age.

To model the present data, a large number of acoustical properties were measured, including duration, average f0, F1, F2, F3, and the geometric mean of F1, F2 and F3 (frequency measures all log transformed) of each syllable. Figure 2 shows the average geometric mean of F1, F2 and F3 and the average f0 for all age groups. It is clear that both age and gender differences are correlated with trends in these plots. Unsurprisingly, older children show lower frequency values than younger children and males show lower frequency values than females.

In addition, we incorporated measures related to glottal source properties beyond pitch derived from the VoiceSauce package for Matlab©, developed by Shue and colleagues (Shue, 2012). A preliminary relative importance analysis (Grömping, 2006) was run to screen candidate acoustic measures for predicting perceived age from the measures. Table 1 provides a list of measures that were selected by this screening for a prediction model.

TABLE 1. Acoustic measurements used in prediction of chronological age.

Label	Description	Reference
dur	duration (ms)	
F0	average fundamental frequency (Hz)	Kawahara et al., 1999
GMFF	geometric mean of F1 F2 F3 (Hz)	Assmann et al., 2008
H1H2c	Corrected magnitude difference between harmonics 1 and 2 (dB)	Iseli et al., 2007
H1A3c	Corrected magnitude difference between harmonic 1 and F3 peak (dB)	Iseli et al., 2007
CPP	Cepstral pitch prominence (dB)	Hillenbrand et al., 1994
HNR05	Harmonic to noise ratio (dB)	de Krom, 1993

Comparing a regression model of chronological age to listeners' judgments when speaker sex is known

The acoustic measures in Table 1 were fitted together with the categorical factor for vowel quality to an ordinary (fixed effects) least squares regression model that predicts true chronological age from those measures. Separate analyses were conducted on female and male talkers. Statistical summaries in terms of variance accounted for are provided in Table 2.

TABLE 2. Relative contribution of factors in predicting chronological age given speaker gender. SS: sum of squares; %Variance: percentage of total variance accounted for.

	Males		Females	
	SS	**% Variance**	**SS**	**% Variance**
CPP	612.4	1.50	140	0.21
dur	348.6	0.86	6515	9.69
F0	17242.5	42.30	2809	4.18
GMFF	420.3	1.03	1458	2.17
H1A3c	194.2	0.48	4685	6.97
H1H2c	161.4	0.40	1354	2.01
HNR05	455.8	1.12	150	0.22
Vowel	318.7	0.78	2878	4.28
Residuals	21008.2	51.54	47242	70.27
Total	40762.1	100.00	67231	100.00

This statistical model can be naively interpreted as a hypothesis about how listeners might make judgments of age. If we pool predictions across male and female speakers within each age group, we can compare errors of prediction with errors of judgments. If listeners are using information used by the statistical models in similar ways to estimate age, then we might expect a good correspondence in the error patterns. A graphic comparison of just such errors is shown in Figure 3.

Figure 3 plots the data in the left panel of Figure 1 in terms of age estimation accuracy, defined as the signed difference between perceived and chronological age. From this figure it can be seen that the underestimation of age starts around 14 years.

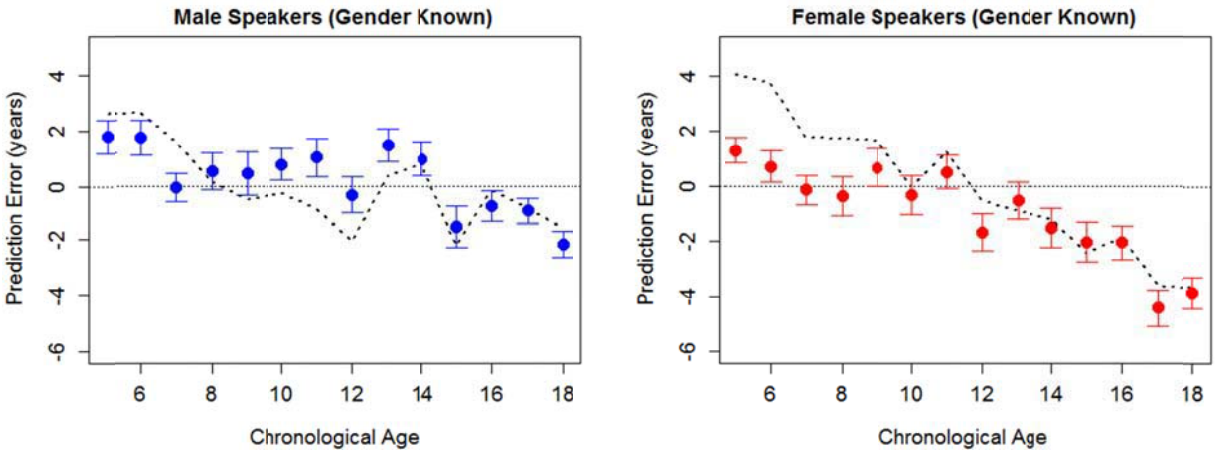

FIGURE 3. Circles and error bars indicate average judgment errors, and standard errors across listeners who were provided with information about the speaker's sex. Dotted lines indicate mean errors in prediction made by the sex-known model.

Comparing a regression model of chronological age to listeners' judgments when speaker sex is unknown

As a naive model of listener's judgments of age when sex is unknown, we fit a regression model similar to that in Table 2 to all of the data with no information about speaker sex. A statistical summary in terms of percentage of variance accounted for is shown in Table 3.

TABLE 3. Relative contribution of factors in predicting chronological age given no information about speaker gender.
SS: sum of squares; %Variance: percentage of total variance accounted for.

Factor	SS	% Variance
CPP	903	0.94
dur	6433	6.70
GMFF	1473	1.53
H1A3c	4746	4.94
H1H2c	2135	2.22
HNR05	0	0.00
strF0	10919	11.37
Vowel	2554	2.66
Residuals	85141	88.63
Total	96060	100.00

If we again pool predictions across male and female speakers within each age group, we can compare errors of prediction with errors of judgments. If listeners are using information employed by the statistical models in similar ways to estimate age, then we might expect a good correspondence in the error patterns. A graphic comparison of just such errors is shown in Figure 4.

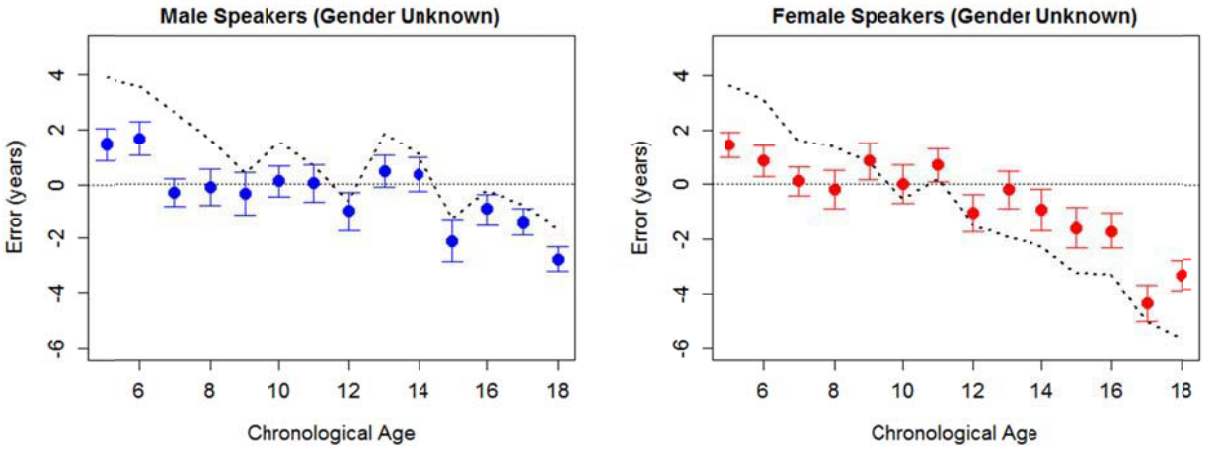

FIGURE 4. Circles and error bars indicate average judgment errors, and standard errors across listeners who were not provided with information about the speaker's sex. Dotted lines indicate mean errors in prediction made by the sex-unknown model.

SUMMARY AND CONCLUSIONS

- Listeners are reasonably accurate in gauging the ages of children from their speech.
- There are some systematic discrepancies, notably underestimation of the ages of older girls.
- Age is more accurately perceived in sentence context compared to isolated syllables.

For isolated syllables:

- Chronological age is relatively well predicted by acoustical measures.
- Simple regression models of chronological age on acoustic members result in error patterns broadly similar to those of human listeners. We plan analogous modeling for sentence context when acoustic measures are completed.

ACKNOWLEDGMENTS

Work supported by the National Science Foundation, Grant No. 1124479. Thanks to Daniel Hubbard and Shaikat Hossain for assistance in data collection and analysis.

REFERENCES

Amir, O., Engel, M., Shabtai, E., & Amir, N. **(2012)**. "Identification of children's gender and age by listeners," J. Voice **26**, 313-321. Epub 2011 Aug 12.

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

Assmann P.F. & Nearey T.M. **(2011)**. "Perception of speaker sex in children's voices," J. Acoust. Soc. Am. **130**, 2446(A).

de Krom, G. **(1993)**. "A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals," J Speech Hear Res 1993 36: 254-266.

Grömping, U. **(2006)**. "Relative Importance for Linear Regression in R: The Package relaimpo," Journal of Statistical Software 17: 1-27.

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," J. Voice **22**, 58-69.

Hillenbrand, J. M., Cleveland, R. A., Erickson, R. L. **(1994)**. "Acoustic Correlates of Breathy Vocal Quality," J. Speech Hear. Res. 37: 769-778.

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.

Iseli, M., Shue, T.-L. and Alwan A., **(2007)**. "Age, sex, and vowel dependencies of acoustic measures related to the voice source," J. Acoust. Soc. Am. **121**, 2283-2295.

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.

Linville, S. E., and Fisher, H. B. **(1985)**. "Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females," J. Acoust. Soc. Am. **78**, 40–8.

Linville, S.E. **(2001)**. *Vocal Aging*. Singular Thomson Learning, San Diego, CA.

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

Schötz, S. **(2007)**. "Acoustic analysis of adult speaker age," In C. Müller (Ed.): *Speaker Classification I*, LNAI 4343, pp. 88–107, Springer-Verlag, Berlin Heidelberg.

Shue, Y.L. **(2012).** "VoiceSauce - A program for voice analysis," Retrieved January 22, 2013, from http://www.ee.ucla.edu/~spapl/voicesauce/index.html

Shue, Y.-L. and Iseli, M. **(2008)**. "The role of voice source measures on automatic gender classification," in *Proceedings of ICASSP*, 2008, pp. 4493-4496.