Perception of speaker age in children's voices

Peter Assmann[1] Santiago Barreda[2] Terrance Nearey[2]

[1]University of Texas at Dallas [2]University of Alberta

Background

When attending to an unfamiliar voice, listeners form an immediate impression of the speaker's sex, age, and size. Such indexical properties are extracted in tandem with the linguistic message. The overall aim of our 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. In a recent study Amir et al. (2012) [1] obtained age judgments for vowels and sentences spoken by 120 children between 8 and 18 years. Listeners' responses were fairly accurate, apart from a tendency to underestimate the ages of older girls.

Research questions

1. Does knowledge of speaker sex lead to more accurate age estimates?
2. Is speaker age judged more accurately from sentences compared to isolated /hVd/ syllables?
3. How accurately can age judgments be predicted from acoustic measures?

Method

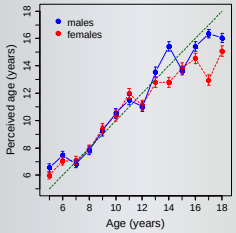
Stimuli Children's vowel database (208 speakers)[2]

- **Age** 5-18 years (14 age levels)
- **Sex** Equal numbers of male & female speakers
- **Vowel** /hid/, /had/, and /hud/
- **Speaker sex info** One group of listeners was told whether the speaker was male or female; the other group was not.
- **Context** Syllables embedded in sentences ("Please say the word ___ again") or in isolation
  - **ISO**: 5 males + 5 females x 14 age levels = 140
  - **SENT**: 3 males + 3 females x 14 = 84 speakers†
- Stimuli presented monaurally using headphones with Tucker-Davis System 3 and RP2.1 hardware
- Listeners used a graphical slider to register their estimate of the speaker's age
- All conditions randomly interspersed

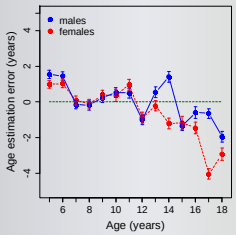
† reduced number of speakers to keep listening sessions ≤1 hour

Results

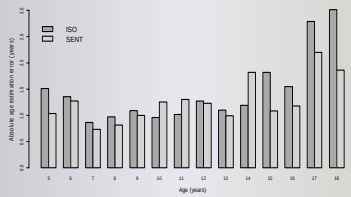
- Perceived age matched chronological age fairly well across most of the age range



- Age estimation error (actual age – perceived age) mostly ± 2 years; however listeners underestimated the ages of the older girls.

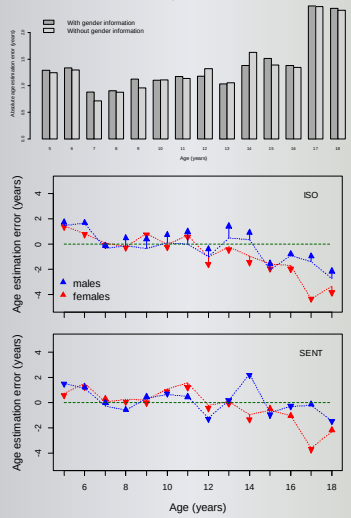


- Listeners estimated speaker age more accurately from sentences (light gray bars) than syllables (dark gray bars), with a mean improvement in age estimation of 0.2 years.



Results

- No overall benefit of speaker sex info



- The 4-way interaction of *age* x *sex* x *context* x *speaker sex info* was significant. Arrows show the change in age estimation error when speaker sex info is provided. Knowledge of speaker sex did not lead to more accurate estimates.

Regression Model

Relating chronological and perceived age to acoustic properties

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

Regression Model

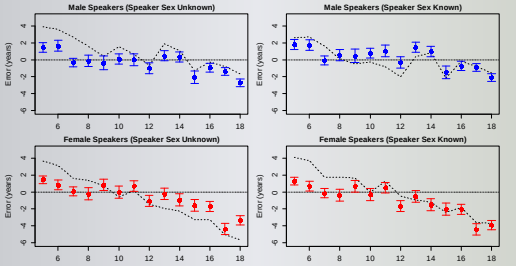
- Least squares regression model (fixed effects)
- Actual (chronological) age predicted from acoustic measures
- Results for syllable data only; separate analyses for males and females

	Males		Females	
Factor	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

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

Predictions for conditions where speaker sex is unknown

Predictions where speaker sex is known



Summary and conclusions

- Listeners are reasonably accurate in gauging the ages of children from their speech, but there are systematic discrepancies, notably underestimation of the ages of older girls. Informing listeners of the sex of the speaker does not lead to improved estimates (and in some conditions leads to lower accuracy). Age is more accurately perceived in sentence context compared to isolated syllables.
- For syllables, chronological age is relatively well predicted by acoustical measures Regression models of chronological age on acoustic measures result in error patterns broadly similar to those of human listeners.

References

[1] Amir, O., Engel, M., Shabtai, E., & Amir, N. (2012). "Identification of children's gender and age by listeners," J. Voice 26, 313-321.

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

Acknowledgments

Work supported by the National Science Foundation Grant #1124479. Thanks to Daniel Hubbard and Shaikat Hossain for assistance in data collection and analysis.