# Modeling the perception of speaker age and sex in children's voices

Peter Assmann[1] Santiago Barreda[2] Terrance Nearey[3]

*[1]University of Texas at Dallas [2]University of Arizona [3]University of Alberta*

## Background

At previous meetings we presented data on the perception of speaker sex and age in children's voices. The stimuli common to these experiments were /hVd/ syllables in isolation and in sentence context. Here we present the results of a modeling study in which acoustic measurements of the /hVd/ syllables were used to predict listener judgments of age and sex. Logistic regression models were constructed to predict listeners' judgments of speaker sex, and linear regression models for speaker age.

## Experiments

**Syllable stimuli**

- **Age** 5-18 years (14 age levels)
- **Sex** Equal numbers of male & female speakers
- **Vowel** /hid/, /hɑd/, and /hud/
- **Talker** 5 speakers per age group, drawn from a vowel database of 208 speakers[1]
- **Experiment 1: Perception of speaker sex**
- **Experiment 2: Perception of speaker age**
- All conditions randomly interspersed; stimuli presented monaurally using headphones with Tucker-Davis System 3 and RP2.1 hardware.
- Listeners used a 2-alternative button box to indicate speaker sex, and a graphical slider for estimating the speaker's age.
- **Speaker sex info** Listeners in Experiment 1 were either informed/not informed of the speaker's **age**.
- **Speaker age info** Listeners in Experiment 2 were either informed/not informed of the speaker's **sex**.

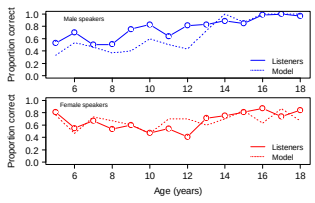
## Acoustic measures

| Label | Description | Reference |
|---|---|---|
| dur | Duration (ms) | |
| F0 | Log average fundamental frequency (Hz) * | Kawahara et al., 2008 |
| GMFF | Geometric mean of F1 F2 F3 (Hz) | Assmann et al. 2008 |
| H1H2c | Corrected magnitude difference between harmonic 1 and harmonic 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 | Harmonics-to-noise ratio below 500 Hz (dB) | de Krom, 1993 |

* In all models F0 is represented by a linear spline with a single knot corresponding to 175 Hz.

## Modeling sex judgments

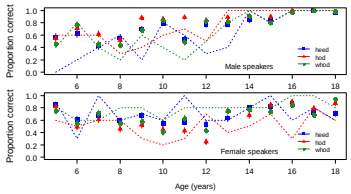
- Sex recognition accuracy: **age x sex interaction**




Mean sex recognition accuracy as a function of age and sex.

- *Consistent with listeners, the model predicts more accurate sex recognition for older males compared to females, and an overall improvement with age for males and older females. Both listeners and model show a tendency to label the youngest speakers as female and older females as male. Some but not all age-related irregularities in listeners' judgments are predicted, suggesting that these variations are partially linked to within-group acoustic differences.*

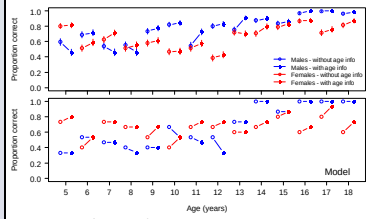
- Sex recognition accuracy: **age x sex x vowel**




Predicted accuracy (dotted lines) and observed accuracy (symbols) as a function of age, sex, and vowel.

- *Vowel category has a complex influence on sex recognition accuracy. For example listeners tend to judge speaker sex more accurately when the syllable is "hod" for some groups of male speakers. The model tends to overestimate differences between the syllable types.*

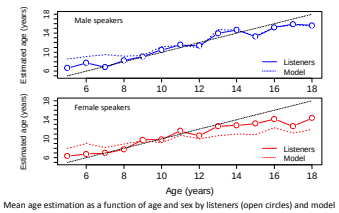
- Sex recognition accuracy: **age information**




- *Knowledge of the speaker's age provides a modest improvement across the age range for most age/sex groups. The model predicts more accurate sex recognition with age information, but only for female speakers.*

## Modeling age judgments

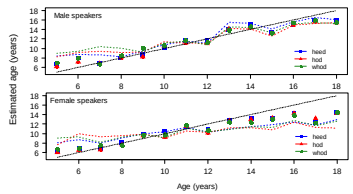
- Age estimation: **age x sex interaction**




Mean age estimation as a function of age and sex by listeners (open circles) and model predictions (dashed lines) for male (upper panel) and female speakers (lower panel). The black dotted line indicates where perceived age = chronological age.

- *Listeners provided fairly accurate estimates of speaker age across the age range, apart from a tendency to underestimate the ages of older girls. The model exaggerates this pattern and also overestimates the ages of younger children.*

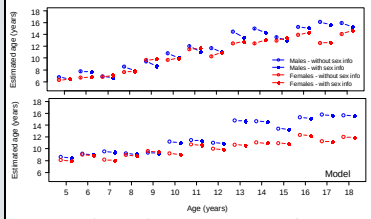
- Age estimation: **age x sex interaction x vowel**




Mean age estimation as a function of age and sex by listeners (open circles) and model predictions (dashed lines) for male (upper panel) and female speakers (lower panel). The black dotted line indicates where perceived age = chronological age.

- *The sex x age x vowel interaction was significant, but vowel category has a relatively small influence on age estimation for listeners and model.*

- Age recognition accuracy: **sex information**




- *Knowledge of the speaker's sex leads to lower age estimates for older males and higher estimates for older females for some age/sex groups. The model predicts little or no effect of sex information.*

## Model summary

- Pooled, fixed-effects models were fit to speaker sex (a logistic model) and age (a linear model) listener responses.
- The importance of factors in these judgments is compared by observing the change in the variance (or deviance) explained when the factor is dropped out of the model.

**Sex**

| Factor | df | Dev. | Δ% Deviance |
|---|---|---|---|
| dur | 1 | 40.1 | 2.1 |
| F0 | 3 | 981.3 | 51.2 |
| GMFF | 1 | 67.6 | 1.1 |
| H1H2c | 1 | 105.3 | 5.5 |
| H1A3C | 1 | 4.5 | 0.2 |
| CPP | 1 | 6.2 | 0.3 |
| HNR05 | 1 | 20.1 | 1.1 |
| Age Info | 1 | 1.2 | 1.7 |
| Vowel | 2 | 31.7 | 6.3 |

**Age**

| Factor | df | SS | Δ% Variance |
|---|---|---|---|
| dur | 1 | 612.8 | 11.4 |
| F0 | 3 | 1069.9 | 20.0 |
| GMFF | 1 | 152.1 | 2.8 |
| H1H2c | 1 | 87.9 | 1.6 |
| H1A3C | 1 | 357.7 | 6.7 |
| CPP | 1 | 136.3 | 2.5 |
| HNR05 | 1 | 43.8 | 0.8 |
| Sex Info | 1 | 7.7 | 0.1 |
| Vowel | 2 | 297.8 | 5.6 |

**Relative contributions of factors for predicting sex and age judgments**

**Preliminary assessment of statistical significance with random listener effects**

- Least squares regression of perceived age on predictor variables with random intercepts for listener confirmed highly significant effects for F0 (including non-linear F0 effects represented by a linear spline with a single knot at 175 Hz), formant frequencies and all measures related to the voicing source for both age and sex.
- Logistic regression (Laplace approximation) of perceived sex on predictor variables with random intercepts for listeners showed a similar pattern of significant effects.
- Note: Including random effects for talkers was not technically feasible.

## Summary and conclusions

- Consistent with earlier findings[2], both listeners and model showed more accurate recognition of speaker sex for males compared to females.
- Both listeners and model provided fairly accurate estimates of speaker age across the age range, apart from a tendency to underestimate the ages of older girls.
- Modeling results confirmed the importance of F0, formant frequencies and measures related to the voicing source for both age and sex.
- Overall, relatively simple regression models incorporating acoustic measures predict overall trends in the data rather well. However some systematic discrepancies remain and will be the focus of future research.

**References**

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

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

**Acknowledgments**

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