Acoustic correlates of talker height in children's voices

Abbey L. Thomas[1], Peter Assmann[1], Santiago Barreda[2]

[1]*University of Texas at Dallas* [2]*University of California, Davis*

Introduction

In a recent study[1] we investigated the ability of listeners to judge the height of the talker from /hVd/ syllables spoken by children between 5 and 18 years. The present study extends that research by examining acoustic variables that predict veridical talker height. We analyzed recordings of the sustained vowel /ɑ/ from each child.

Research Questions

Knowledge of the sex of the talker plays an important role in the perception of age[1] and height[2] for children's voices. Models incorporating fundamental frequency (F0) and the geometric mean of the lowest 3 formants (GMFF) provide an explanation for perceptual judgments. The present study asks :

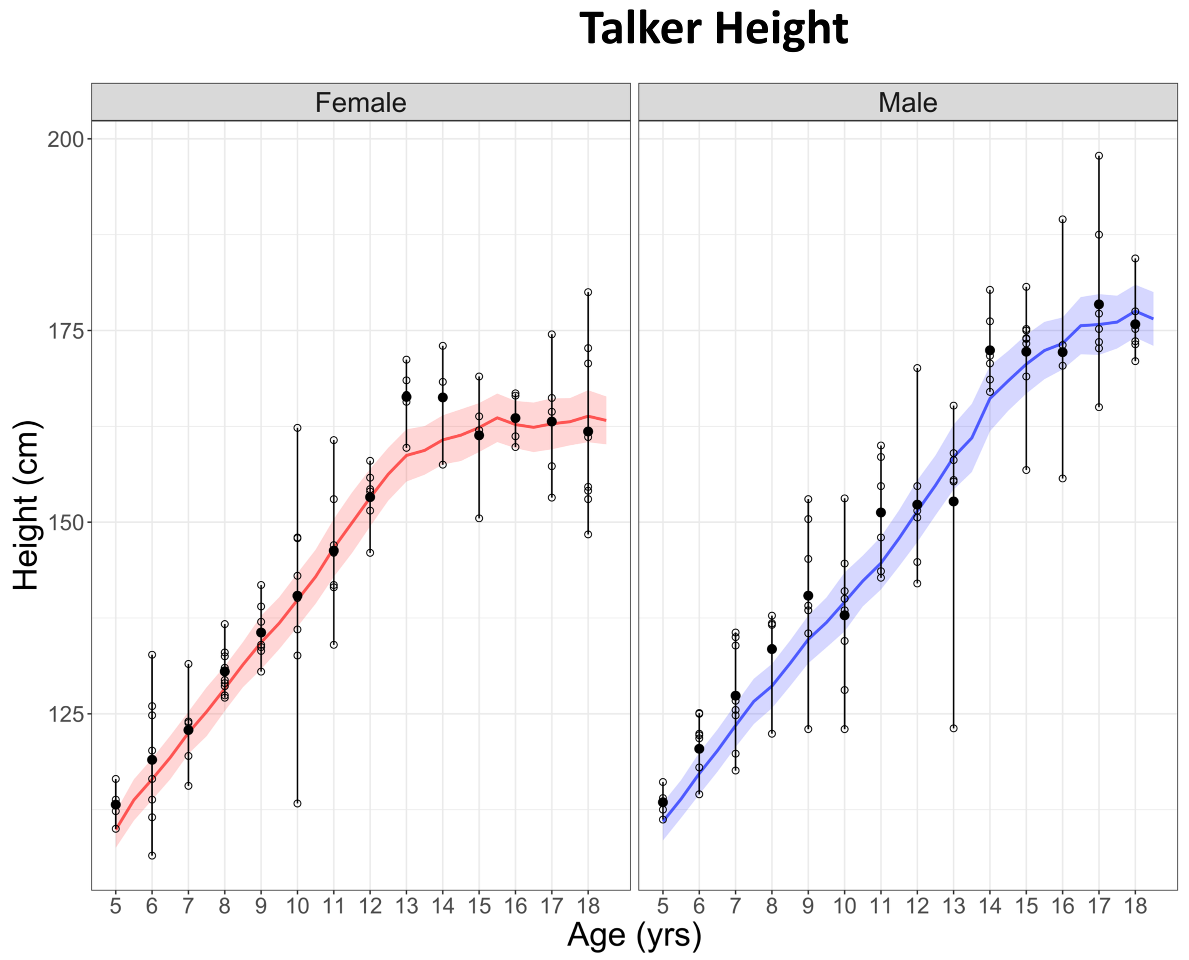
(1) whether models specifying talker sex predict veridical talker height more accurately than models that exclude sex; and

(2) whether predictions can be improved by including additional measures related to the voicing source.

Data

- ***Speech***: 1 sec. excerpt from sustained vowel /ɑ/. 1 token per talker
- ***Talkers***: 175 children, 5-18 years of age, similar numbers of males and females at each age level:

	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Sum
M	5	8	8	5	9	8	6	6	6	6	8	4	7	6	92
F	4	8	5	9	7	8	7	6	5	3	4	5	5	7	83
Sum	9	16	13	14	16	16	13	12	11	9	12	9	12	13	*175*



Despite some obvious outliers, the mean height of our talkers at each age level did not differ significantly from the mean height data provided by the CDC [4].

Males: $t(26) = -.10$, $p = .92$; Females: $t(26) = -.08$, $p = .94$.

Figure 1. U.S. children's height data by age from the CDC (shown in red and blue), overlaid with our talkers' height data (mean = filled circles, individuals = open circles).

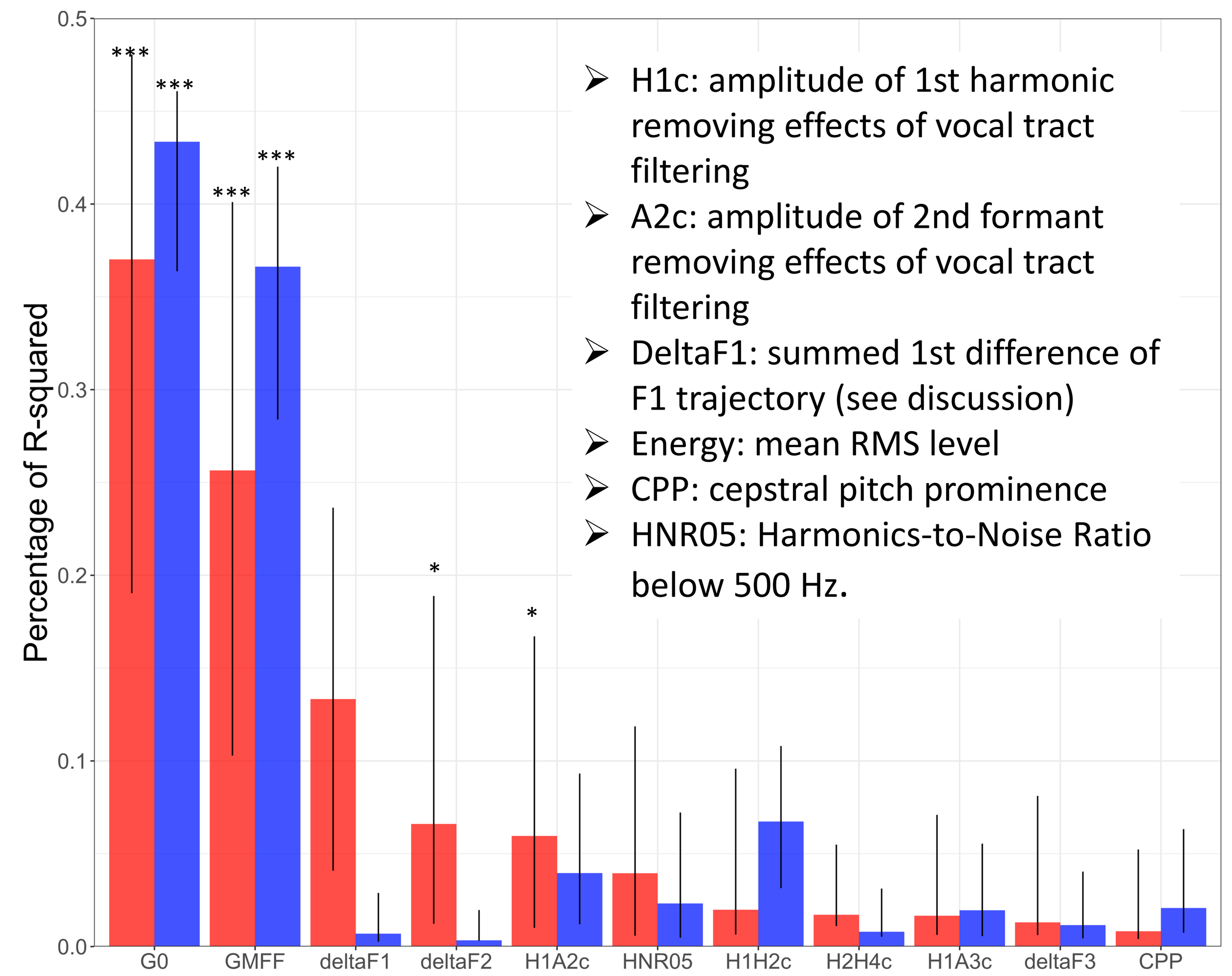
Analysis

Acoustic Predictors

Voicing properties were measured with the Voice Sauce package[3]. Measures were chosen because of their relationship to properties of the voicing source: e.g., breathiness (H1c - H2c) and spectral tilt (H1c – A3c)[5]. GMFF and changes in formants 1-3 across the signal were calculated using the Nearey formant tracker. G0 was estimated using the STRAIGHT algorithm.

Modeling

Collinearity between variables was measured with Variance Inflation Factor (VIF)[6]. All variables in the final model were below the suggested threshold of VIF < 10.

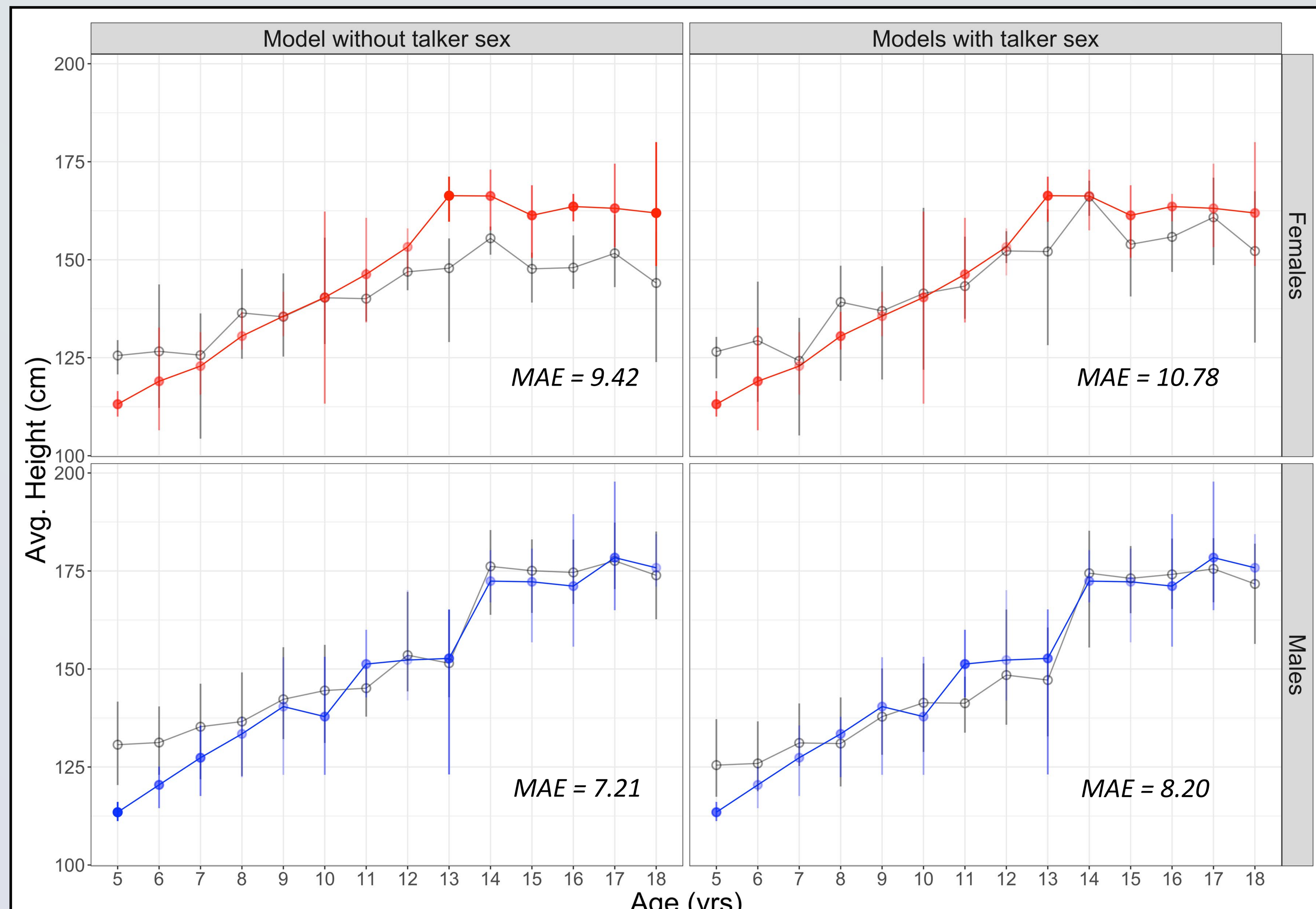


- H1c: amplitude of 1st harmonic removing effects of vocal tract filtering
- A2c: amplitude of 2nd formant removing effects of vocal tract filtering
- DeltaF1: summed 1st difference of F1 trajectory (see discussion)
- Energy: mean RMS level
- CPP: cepstral pitch prominence
- HNR05: Harmonics-to-Noise Ratio below 500 Hz.

Figure 2 . % of variance explained by each predictor. Error bars show 95% confidence intervals. *** $p < .005$; * $p < .05$. Model for female talkers = red; male talkers = blue.

Three linear regression models were fitted: one ignoring talker sex, one for males, and one for females. A bootstrapping method was used for each model to determine the stability of the model and the relative importance of predictors by averaging their contributions to the experimental R-squared over 1000 iterations of the model selection procedure[7].

Results



Formula: Talker height (cm) ~ G0 + GMFF + DeltaF1 + DeltaF2 + Delta F3 + (H1c-A2c) + (H1c-A3c) + (H1c-H2c) + (H2c-H4c) + HNR05 + CPP

Figure 3. Predictions (open circles, black lines) of linear regression analyses with all acoustic parameters as predictors of talker height (actual heights and variability shown with colored lines and circles). Differences between models with and without sex shown in columns.

Discussion

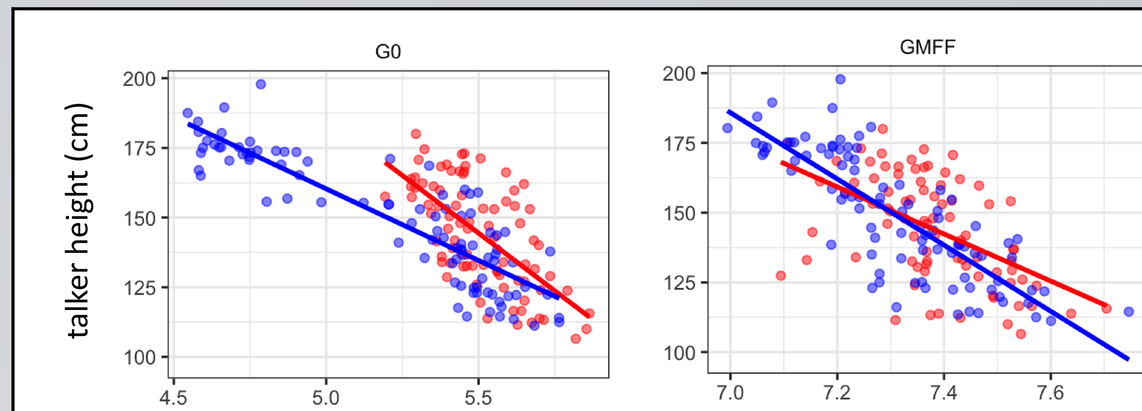
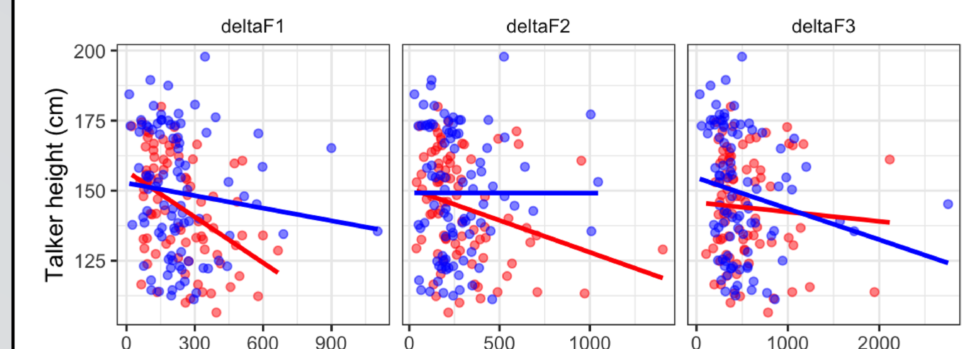


Figure 4. Relationships between G0 and GMFF and talker height/sex.

G0 and GMFF far outperform the other acoustic variables as predictors of talker height. Together they account for 80% of the variance explained by all 14 predictors included in the model for male talkers and 63% of the variance for female talkers.

Formant change: $\Delta Fn = \sum_{n=1}^{11} (Fn_i - Fn_{i-1})$ where Fn_i is the frequency of the n^{th} formant sampled at 10% intervals over the vowel



A potentially important predictor of height unique to sustained vowels.

Figure 5. Relationships between formant change and talker height/sex.

Summary and Conclusions

- Models trained on acoustic measures related to talker height produce significantly better predictions of veridical height when information about talker sex is included.
- We confirm previous results indicating that G0 and GMFF make a reliable contribution to predictions of talker height in children's voices.
- Spectral slope (H1A2c) and formant change (ΔF2) make small but reliable contributions to height predictions for female voices.

References

[1] S. Barreda & P.F. Assmann (2018). Modeling the perception of children's age from speech acoustics. The Journal of the Acoustical Society of America 143, EL361 (2018); https://doi.org/10.1121/1.5037614

[2] P.F. Assmann, M.R. Kapolowicz and S. Barreda (2018). Perception of talker height and sex from children's voices. The Journal of the Acoustical Society of America 144, 1964 (2018); https://doi.org/10.1121/1.5068584.

[3] Shue, Y.-L., P. Keating , C. Vicenik, K. Yu (2011). VoiceSauce: A program for voice analysis, Proceedings of the ICPhS XVII, 1846-1849.

[4] Kuczmarski RJ et al. (2002). 2000 CDC Growth Charts for the United States: methods and development. Vital Health Stat 11(246), 1-190.

[5] Iseli, M., Shue, Y.L., Alwan, A. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. J. Acoust. Soc. Am. 121, 2283-2295.

[6] Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.

[7] Lindeman, R.H., Merenda, P.F. and Gold, R.Z. (1980). *Introduction to Bivariate and Multivariate Analysis*, Glenview IL: Scott, Foresman.

Acknowledgments

Work supported by National Science Foundation Grant #1124479. We thank Danni Yang, Christina Mai, Katelyn Foxworth, Riya Mahajani, Shravya Kichena for help with data collection and acoustic analyses.