Perception of voice gender in children's voices by cochlear implant users

Daniel Guest, Vahid Montazeri, Michelle Kapolowicz, Peter Assmann

The University of Texas at Dallas, School of Behavioral and Brain Sciences

Questions

When listening to the speech of children

- What acoustic cues do listeners use to identify the gender of the talker?
- How does age perception connect to voice gender perception (VGP)?
- What happens when these cues are distorted or eliminated, as occurs in cochlear implant (CI) processing?

Background

- **Acoustic cues of voice gender**
  - Acoustic cue manipulation paradigms can reveal which cues listeners use [2, 5, 9]
  - Fundamental frequency (F0) and formant frequencies (FFs) are key [5]
  - However, F0 and FFs are not the only cues listeners use [2, 9]
- **Voice gender in children's speech and the role of age perception**
  - In children, average F0 and FFs vary systematically with age as well as gender (see Figure 1)
  - Previous research showed accurate age perception plays an important role [3]
  - Relatively accurate age perception has been demonstrated in children's speech [1]
- **VGP by CI users**
  - CI users have limited VGP abilities [4, 7]
  - Poor VGP possibly due to limited access to F0 and FFs [7, 4]
  - Good temporal resolution could help CI users use F0 in VGP [4]
  - Good spectral resolution could help CI users use FFs in VGP [4]
- **Significance**
  - VGP tasks may be useful to measure availability of spectrotemporal cues in CI users [8]
  - Talker identity cues play a role in speech perception

Methods

- **Stimuli:** /hVd/ syllables spoken by children from the North Texas area (age range: 5-18 years, medial vowels: /i/, /a/, /u/)

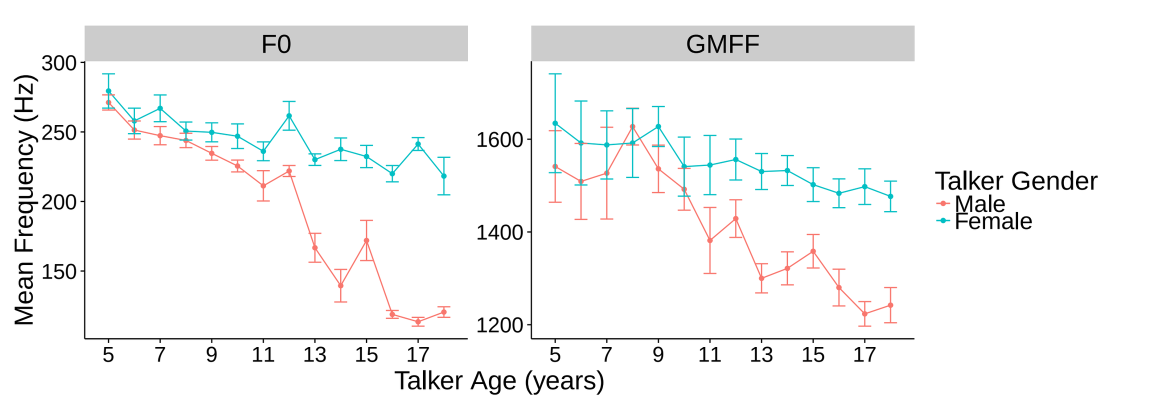


Figure 1: Mean F0, geometric mean of first three formants (GMFF) Values averaged across talkers and tokens within each gender. Error bars show ±1 standard error of the mean.

- **Task:** Participants listened to isolated syllables presented in randomized order and made two responses: (1) **Voice Gender** (two-alternative forced choice) and (2) **Age** (continuous scale, graphical slider, range: 5 – 18 years)

- **Acoustic Cue Conditions:** STRAIGHT vocoder [6] used to scale F0 contours and/or FF contours to opposite-sex averages at each talker age level

Unswapped	Original F0 and FFs
FF Swapped	Original F0, scaled FFs
F0 Swapped	Scaled F0, original FFs
F0FF Swapped	Scaled F0 and FFs

- **Experiments:**
  - **NH** 81 normal hearing (NH) adults
  - **VO** 37 NH adults attending to tone vocoder CI simulation
  - **CI** 5 CI users

Age	60	70	55	60	74
Age of implantation (L/R)	53/55	65/64	57/56	57/∅	52/∅
Age of hearing loss (L/R)	30/30	52/52	0/0	2/2	1/1
Device type (L/R)	N6/N6	N5/N5	N5/N5	Naida/∅	N6/∅

- **Other Details**:
  - **NH** and **VO** listeners completed one condition each
  - **CI** users completed a reduced set including all 4 conditions using only best implanted ear
  - 8 channel tone vocoder implemented according to specifications in [4], 160 Hz envelope cutoff
  - Stimuli presented monaurally over headphones (NH listeners) or in free field (CI listeners)

Results

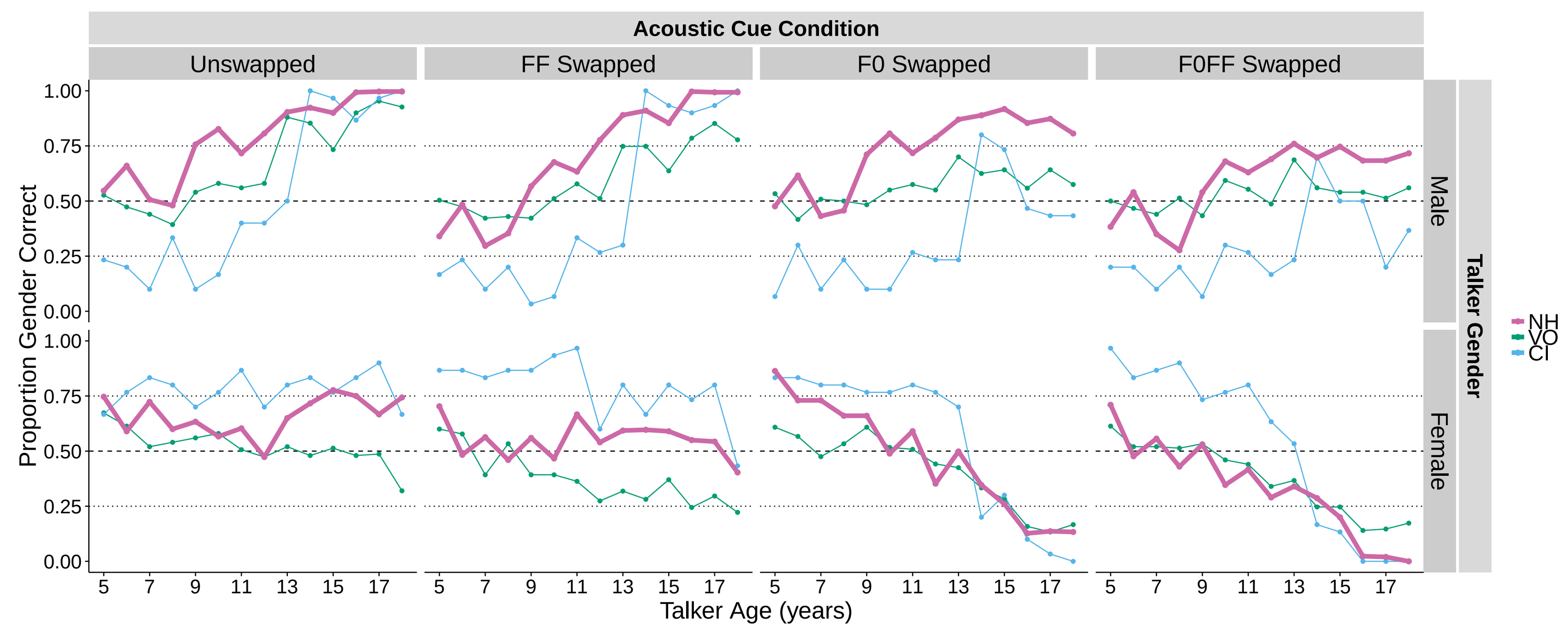


Figure 2: Proportion gender correct Each panel shows a different combination of acoustic cue condition and talker gender, while line color indicates listener group.

- Majority of male talkers correctly identified...
  - by NH listeners after 9 years of age
  - by CI listeners after 13 years of age
- CI listeners answered female for most younger talkers
  - Explains why CI listeners had higher proportion gender correct than NH listeners for female talkers
- Listeners answered male for many older females in Unswapped
- Across talker age, talker gender, and experiment, scaling F0 had larger effect than scaling FFs
- Effect of acoustic cue condition interacted with talker age, talker gender, and experiment

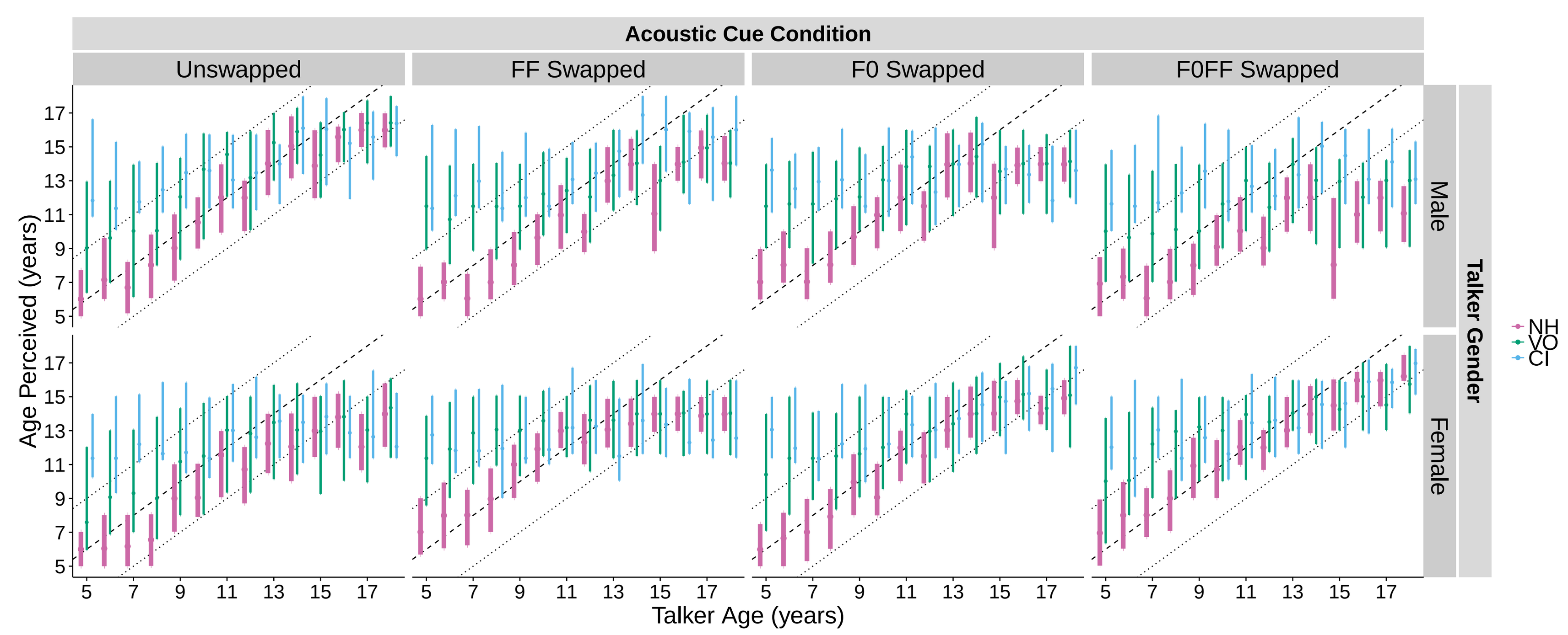


Figure 3: Median age perceived Each panel shows a different combination of acoustic cue condition and talker gender, while line color indicates listener group. Error bars indicate upper and lower quartiles.

- VO and CI listeners overestimated age of younger talkers
- Most listeners underestimated age of older female talkers in Unswapped
- Across talker age, talker gender, and experiment, scaling F0 had larger effect than scaling FFs
- Results in Unswapped mirror results for opposite talker gender in F0FF Swapped
- Correspondence between errors in VGP and errors in age perception suggests link

Model of VGP Data

- **Type:** Multilevel logistic (GLRM) model
- **Predictors:** Talker age, talker gender, age estimation error, acoustic cue condition, experiment (and interactions)
- **Implementation:** Implemented via lme4 package in R
  - Correct voice gender responses coded with a 1 and incorrect responses with a 0
  - Outcome variable was denoted $y_{i(jk)}$, for the i-th measurement from listener j to a stimulus spoken by talker k
  - Intercepts were allowed to vary between listeners and talkers

$$\text{logit}(y_{i(jk)}) = X_{i(jk)}\beta + u_j + v_k \mid u_j \sim N(0, \sigma_u^2), v_k \sim N(0, \sigma_k^2)$$

Model Analysis

Questions: To what extent did particular listeners make correct voice gender responses? To what extent was the voice gender of particular talkers correctly identified?

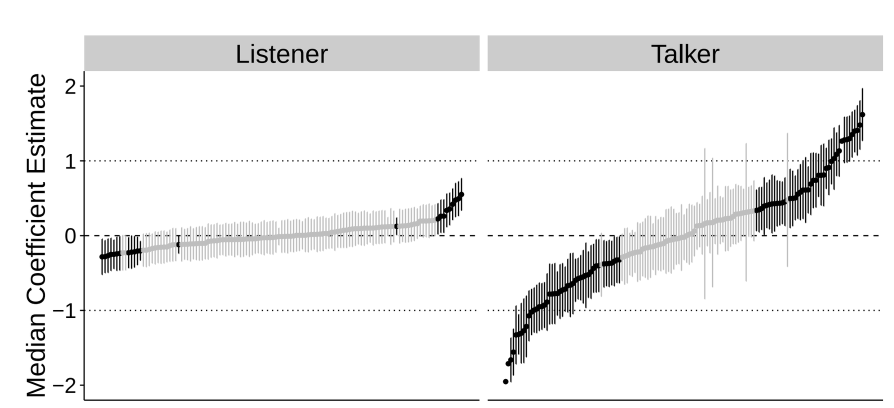


Figure 4: Median estimates of random intercepts for talkers and listeners Error bars indicate ±2 standard deviations. Grey intervals indicate intersection with 0.

- More variation explained by talker than listener intercepts

Question: How did listeners perform in the Unswapped condition when they identified talker age accurately?

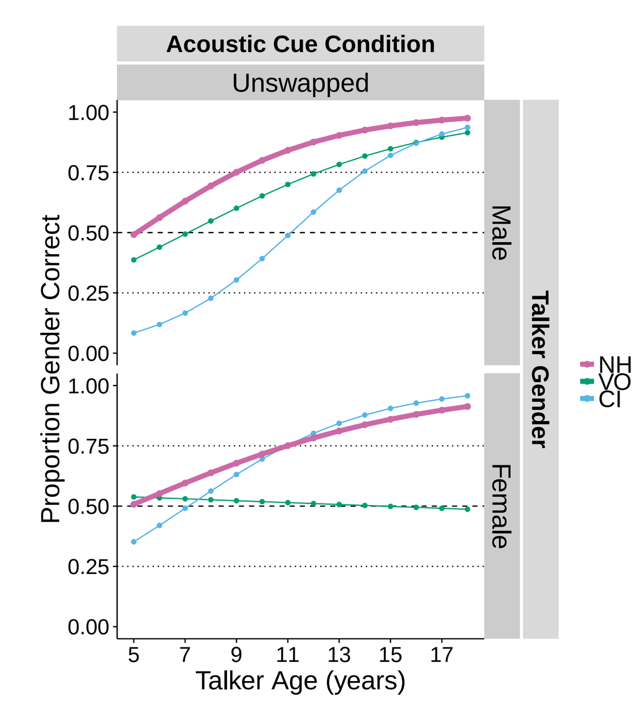
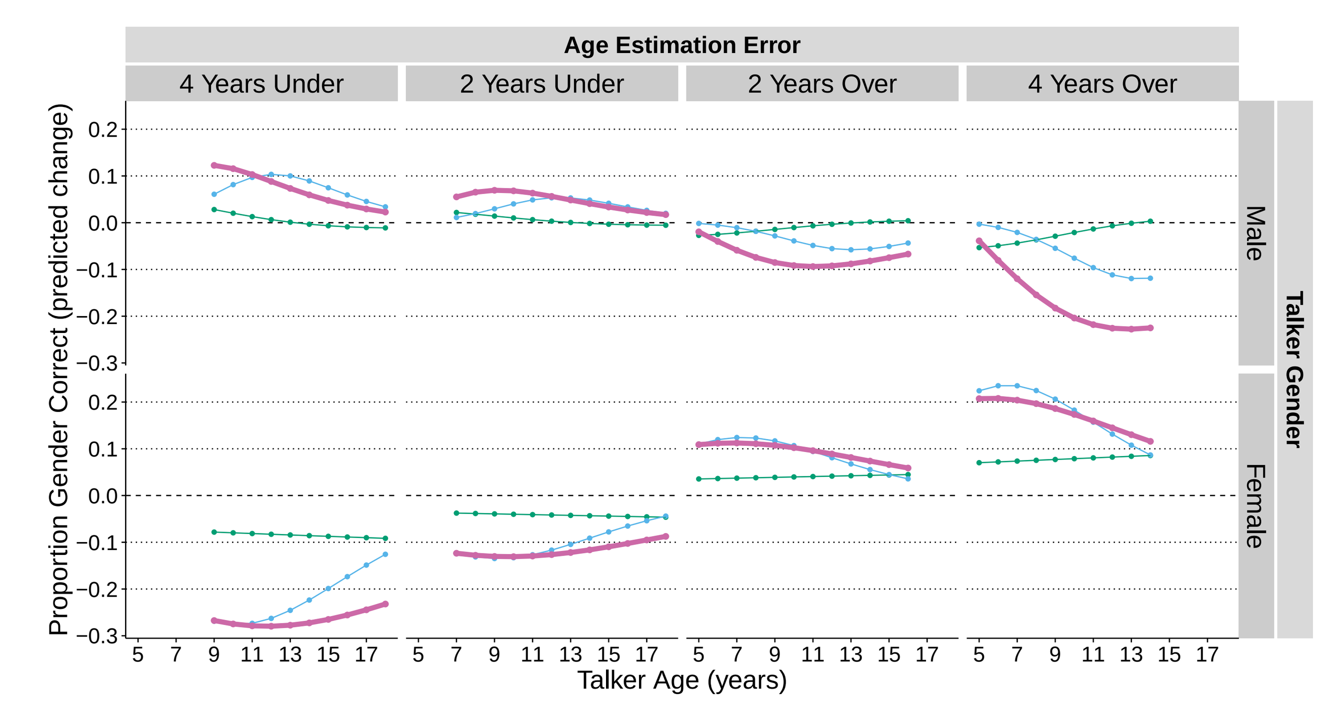


Figure 5: Predicted proportion gender correct given perfect age estimation, Unswapped Panels and line color indicate the same as in Figures 2 and 3. Intercepts for listener and talker were selected to be zero (i.e., predictions here correspond to 'average' talkers and listeners).

- Listeners made more correct responses to older talkers and male talkers
- Constitutes a baseline for comparisons in Figures 6 and 7

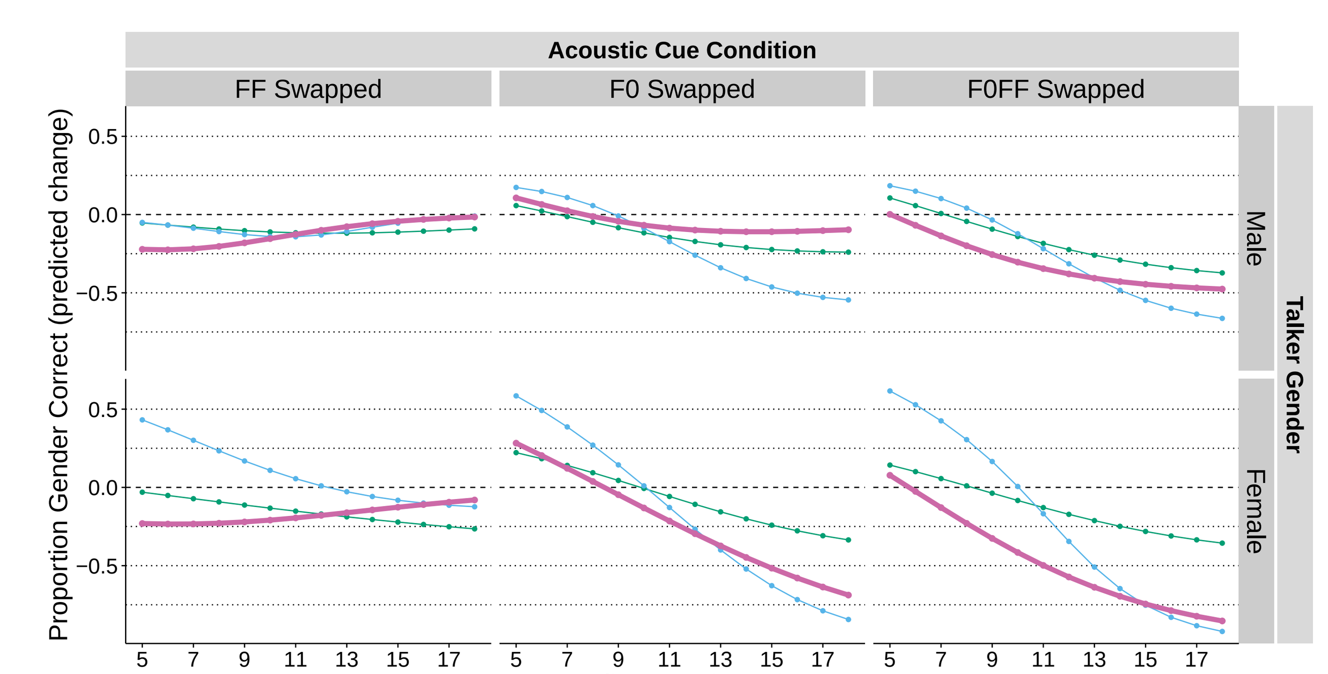
Model Analysis (continued)



Question: How was VGP affected when listeners under- or overestimated the age of the talker?

Figure 6: Predicted change in proportion gender correct at different levels of age estimation error, Unswapped Plotted relative to baseline in Figure 5. Each panel shows a different combination of age estimation error and talker gender. Line color indicates listener group. Impossible combos of age estimation error and talker age are not shown.

- Underestimating age of older female talkers linked to VGP errors
- Overestimating age of younger male talkers linked to VGP errors



Question: How was VGP affected by different acoustic cue conditions?

Figure 7: Predicted change in proportion gender correct in different acoustic cue conditions Plotted relative to baseline in Figure 5. Panels and line color indicate the same as in Figures 2 and 3. Here, age estimation error is fixed at 0 years.

- Swapping both cues usually had larger effect than swapping one cue for NH listeners
- Swapping both cues often had similar effect to swapping only F0 for VO and CI listeners

Summary and Conclusions

- For NH listeners...
  - Accurate age perception tied to accurate VGP
  - Older female talkers frequently mis-identified as younger male talkers
  - Utilized both acoustic cues, but FFs more important for VGP of younger talkers and F0 for VGP of older talkers
- For CI listeners...
  - Depended more on F0 than on FFs in VGP task
  - Younger male talkers frequently mis-identified as older female talkers
  - Errors likely due to combination of poor age perception and heavier reliance on F0
- Future directions
  - For NH listeners, scaling both cues did not flip VGP of older male talkers — may need to investigate modifying other acoustic cues (e.g., properties of the voicing source)
  - Utilize more advanced acoustic simulations of CIs to address VO - CI discrepancies
  - Statistical modeling of link between acoustic properties and VGP/age perception

References

[1] Ofer Amir et al. "Identification of Children's Gender and Age by Listeners". In: *Journal of Voice* 26.3 (2011), pp. 313–321.

[2] Peter F. Assmann, Terrance M. Nearey, and Sophia Dembling. "Effects of frequency shifts on perceived naturalness and gender information in speech". In: *Proceedings of the Ninth International Conference of Spoken Language Processing* (2006), pp. 889–892.

[3] Peter F. Assmann et al. "Links between the perception of speaker age and sex in children's voices". In: *The Journal of the Acoustical Society of America* 138 (2015). Poster presented at the ASA Jacksonville 2015 Meeting, p. 1811. DOI: http://dx.doi.org/10.1121/1.4933751.

[4] Qian-Jie Fu, Sherol Chinchilla, and John J. Galvin. "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users". In: *The Journal of the Association for Research in Otolaryngology* 5.3 (2004), pp. 253–260. DOI: 10.1007/s10162-004-4046-1.

[5] James M. Hillenbrand and Michael J. Clark. "The role of F0 and formant frequencies in distinguishing the voices of men and women". In: *Attention, Perception, & Psychophysics* 71 (2009), pp. 1150–1166. DOI: 10.3758/APP.71.5.1150.

[6] Hideki Kawahara. "STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds". In: *Acoustical Science and Technology* 27.6 (2006), pp. 349–353.

[7] Damir Kovacic and Evan Balaban. "Voice gender perception by cochlear implantees". In: *The Journal of the Acoustical Society of America* 126 (2009), pp. 762–775. DOI: 10.1121/1.3158855.

[8] Tianhao Li and Qian-Jie Fu. "Voice gender discrimination provides a measure of more than pitch-related perception in cochlear implant users". In: *International Journal of Audiology* 50 (2011), pp. 498–502.

[9] Verena G. Skuk and Stefan R. Schweinberger. "Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender". In: *Journal of Speech, Language, and Hearing Research* 57 (2014), pp. 285–296. DOI: 10.1044/1092-4388(2013/12-0314).

Work supported by NSF grant 1124479