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

Daniel Guest
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?
- 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 [1, 5, 9]
  - Fundamental frequency (F0) and formant frequencies (FFs) are key [5]
  - However, F0 and FFs are not the only cues listeners use [1, 9]
  - In children, average F0 and FFs depend on age as well as gender (see Figure 1)
  - It is not entirely clear how listeners utilize these cues when listening to children's speech
- **Voice gender perception (VGP) by CI users**
  - CI users have poor VGP [4, 7]
  - This is 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]
  - VGP tasks may be useful to measure availability of spectrotemporal cues in CI users [8]
- **Previous research**
  - Listeners employ flexible strategies in the presence of acoustic cue shifts [3]
  - Accurate age perception plays an important role [2]

Methods

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

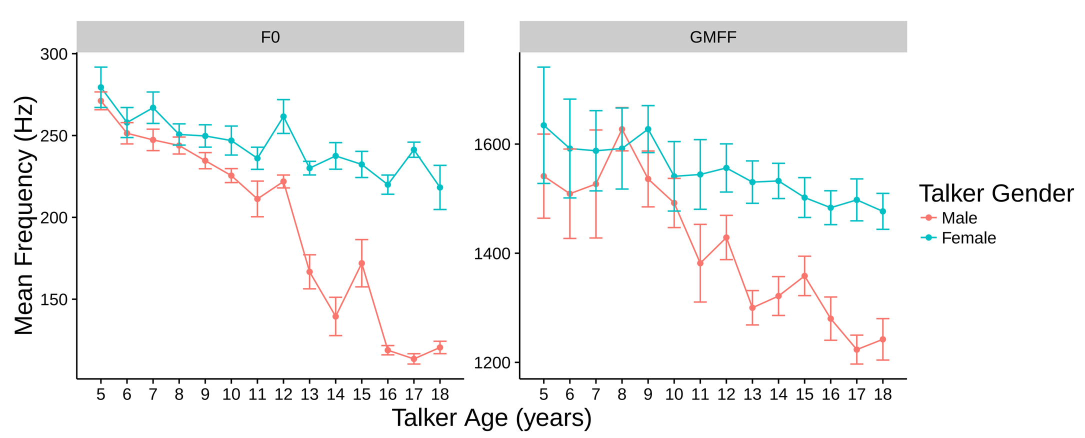


Figure 1: Mean F0 and geometric mean of the first three formants (GMFF) at each age level, averaged across talkers and tokens within each gender. Error bars show standard error of the mean.

- **Shifting Conditions:** STRAIGHT vocoder [6] used to 'swap' F0 and/or FFs to opposite-sex averages at each age level

Name	*Acronym*	*Description*
Unswapped	UN	Original F0 and FFs
F0 Swapped	F0	Swapped F0, original FFs
FF Swapped	FF	Original F0, swapped FFs
F0FF Swapped	FX	Swapped F0 and FFs

- **Task:** Participants listened to isolated syllables presented in randomized order and then made two responses:

Response	*Description*
Voice Gender	Two-alternative forced choice (2AFC)
Age	Continuous scale (range: 5 – 18 years)

- **Experiments:** 3 experiments; normal-hearing (NH) adults, NH adults attending to an 8 channel tone vocoder, and adult CI users

Name	NH	VO	CI
Participants	NH UTD students	NH UTD students	CI users
Processing	None	Tone vocoder	None
Shifting	Between-subjects	Between-subjects	Within-subjects
# of stimuli	420	420	672
n	81	37	5

- **Other Details**:
  - Tone vocoder implemented according to specifications in [4]
  - Stimuli were presented monaurally over headphones to normal-hearing listeners and in free field to CI users via a Tucker-Davis System 3 and RP2.1 hardware system
  - Listeners made their responses via a computer interface generated in MATLAB

Results

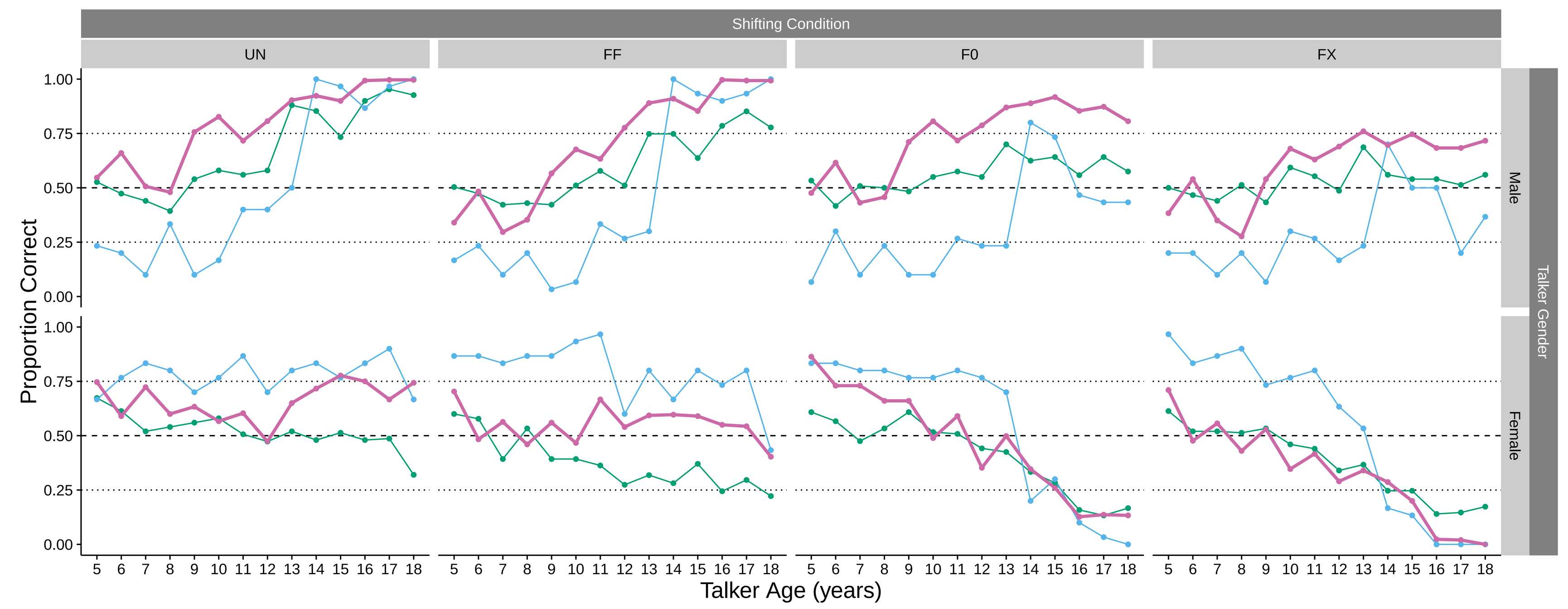


Figure 2: Proportion gender correct at each age level. Each panel shows a different combination of talker gender and shifting condition, while line color indicates experiment.

- Overall proportion correct:

	NH	VO	CI
UN	0.72	0.59	0.65
FF	0.62	0.49	0.62
F0	0.6	0.49	0.44
FX	0.46	0.45	0.4

- VO listeners at chance level for most female stimuli in UN Swapped
- CI listeners responded female for most younger children
- Effect of shifting condition interacts with talker gender, talker age, and experiment

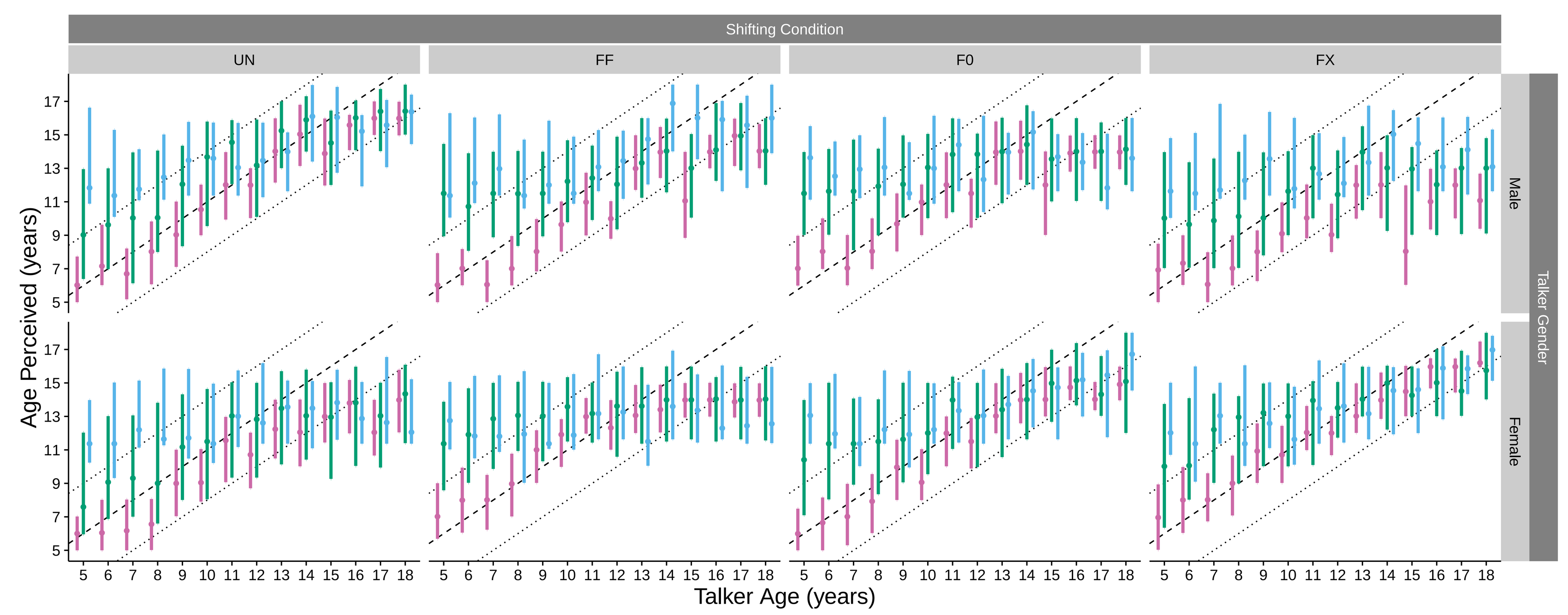


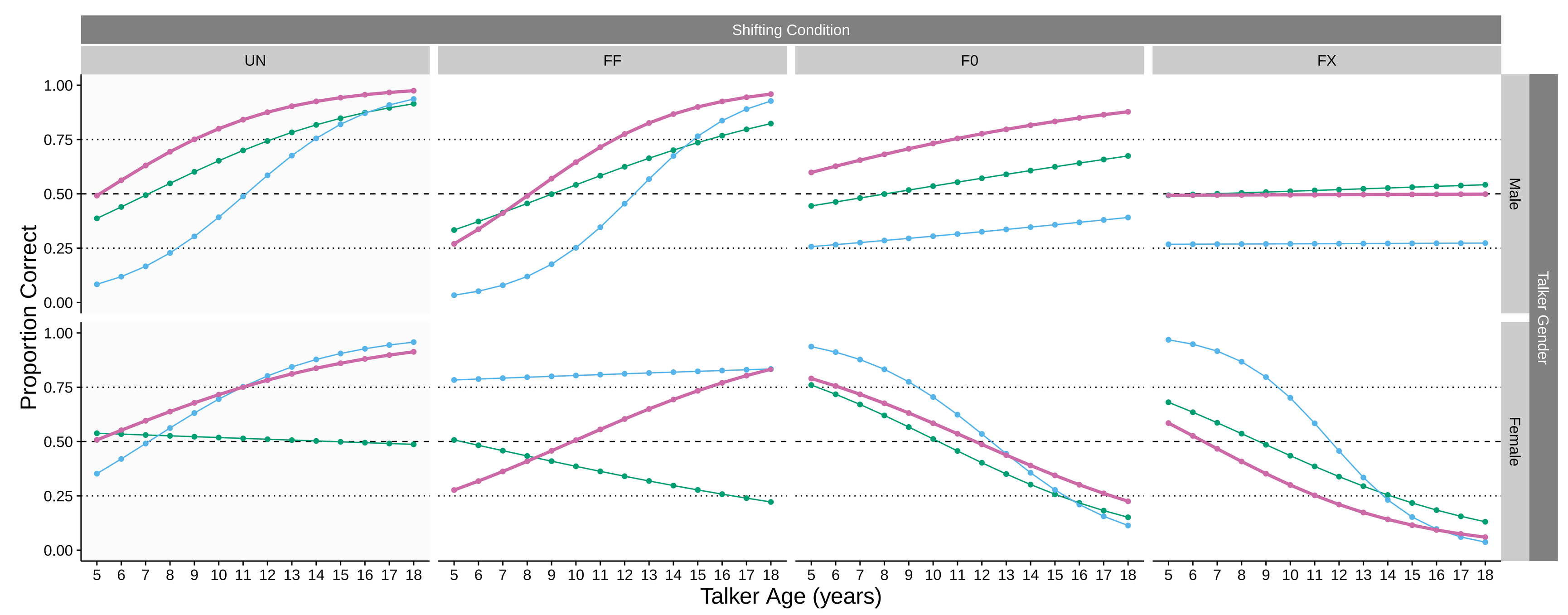
Figure 3: Median age perceived at each age level, with error bars indicating upper and lower quartiles. Each panel shows a different combination of talker gender and shifting condition, while line color indicates experiment.

- UN Swapped mean absolute deviation (years):

NH	VO	CI
2.23	3.42	3.88

- CI listeners (and to a lesser extent VO listeners) overestimated the age of younger talkers
- Most listeners underestimated the age of older female talkers in UN Swapped

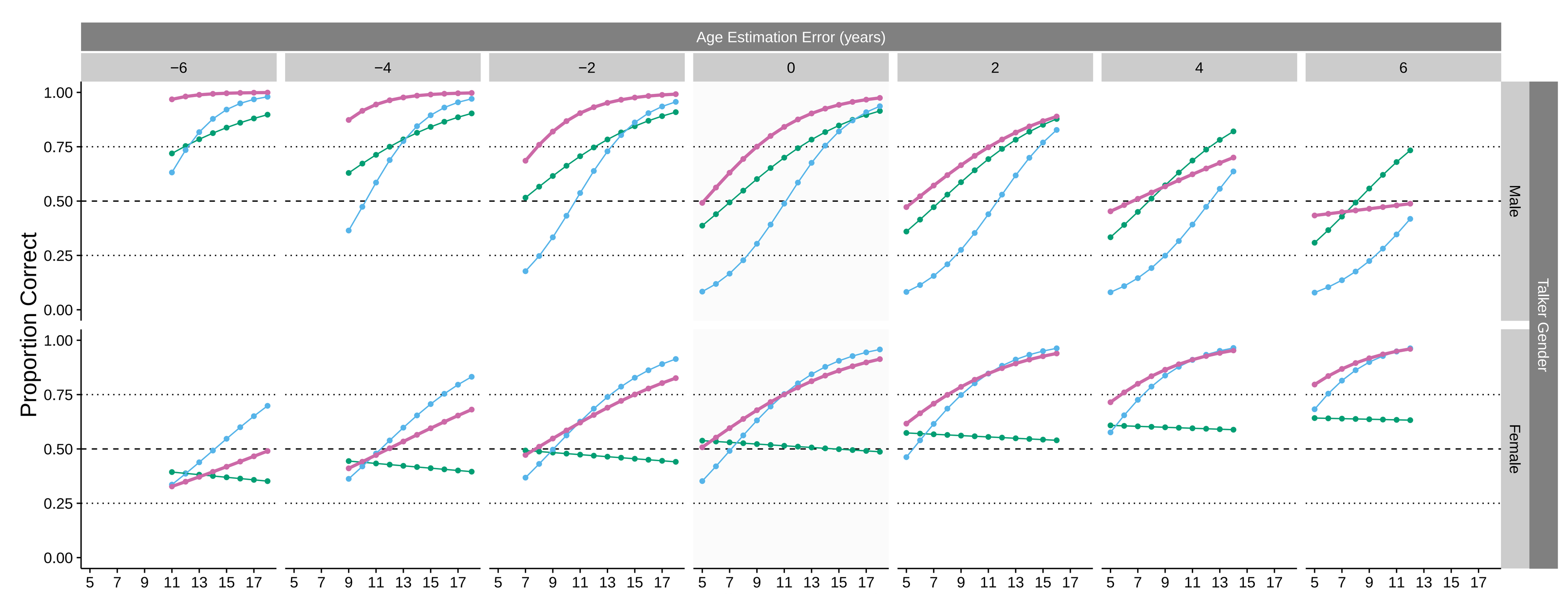
Model



Question: How do the shifting conditions affect VGP?

Figure 4: Model predictions for VGP at each age level assuming perfect age estimations. Each panel shows a different combination of talker gender and shifting condition, while line color indicates experiment.

- Generally:
  - Swapping FFs had a larger impact on younger male talkers
  - Swapping F0 had a larger impact on older talkers
  - Swapping both cues had a larger effect than swapping individual cues
- For VO listeners, but not CI or NH listeners, F0 and FF shifts had a similar impact on older female talkers
- Disproportionately large effect of F0 swap for CI listeners
- Swapping both cues 'flipped' VGP for VO and CI listeners
- Swapping both cues did not 'flip' VGP for NH listeners



Question: How does age perception connect to VGP?

Figure 5: Model predictions for VGP at each age level for varying amounts of age estimation error in the UN shifting condition. Each panel shows a different combination of talker gender and shifting condition, while line color indicates experiment. Predictions for impossible combinations of age estimation error and talker age are not shown.

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

Model (continued)

- **Type:** Multilevel generalized linear model with a logistic link function
- **Predictors:** Talker age, talker gender, age estimation error, shifting condition, and experiment
  - Coefficients for each predictor variable alone and all possible interactions were estimated
- **Mathematical description:**
  - The outcome variable was denoted $y_{i(jk)}$, for the i-th measurement from listener j to a stimulus spoken by talker k, and coded as...

$$y_{i(jk)} = \begin{cases} 1, \text{correct gender identification} \\ 0, \text{incorrect gender identification} \end{cases} \tag{1}$$

  - Intercepts were allowed to vary by listener and talker, so the model was written...

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

Answers

When listening to the speech of children...

- NH listeners primarily used FFs to identify voice gender in younger talkers and a mixture of F0 and FFs to identify voice gender in older talkers
- CI listeners identified many younger male talkers as females and relied mostly on F0 to identify voice gender

Discussion and Future Directions

- **Role of age perception**
  - Common errors in age perception linked to common errors in VGP
  - Statistical modeling of age perception may clarify its role
- **Talker gender asymmetry**
  - Acoustic cue swaps affected VGP of female talkers more than VGP of male talkers
  - Due to bias to respond 'male', acoustic cues which mark male voices, or flexible listening strategies?
- **Other acoustic cues in VGP**
  - Swapping both cues did not completely 'flip' VGP for NH listeners
  - There are likely additional acoustic cues of voice gender (e.g., spectral slope [9])
  - May not be readily available to VO or CI listeners
  - Might identify candidates by modeling link between acoustics and VGP
- **VO vs. CI results**
  - VO listeners more affected by shifted FFs than CI listeners
  - CI listeners and VO listeners showed different response pattern for younger talkers
  - May want to use more advanced acoustic simulations of CIs in future studies

Acknowledgements

- Supported by National Science Foundation Grant #1124479, Undergraduate Research Scholar Award, and Buhrmester Summer Research Award
- Thanks to fellow lab members and everyone who participated in this research
- R: knitr, ggplot2, dplyr, grid, gridExtra, lme4, reshape2, merTools
- LaTeX: beamer, beamerposter, graphicx, tabularx, paralist

References

[1] 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.

[2] 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.

[3] Peter F. Assmann et al. "Perception of speaker sex in resynthesized children's voices". In: *The Journal of the Acoustical Society of America* (2014). Poster presented at the ASA 2014 Meeting.

[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).