Normalization to talker gender and F0: Phonetic category adjustment by cochlear implant users

UMCP NACS PROGRAM NEUROSCIENCE AND COGNITIVE SCIENCE

Matthew B. Winn Ariane E. Rhone William J. Idsardi Monita Chatterjee

CEBH CENTER FOR COMPARATIVE AND EVOLUTIONARY BIOLOGY OF HEARING

University of Maryland College Park

Background

Focus: the perception of the /s-ʃ/ contrast (“see-she” “sue-shoe”)

Perception of this contrast is affected by context and talker information such as vowel coarticulation and talker gender :

Since lip-rounding and large vocal tracts have a frequency-lowering effect, the perceptual boundary between /s-ʃ/ is lowered when the consonant is spoken before /u/ or spoken by a male [1]

vowel-context effect *gender-context effect*

This adjustment also occurs (to a lesser extent) in response to visual stimuli [2]

Phonetic boundary adjustment in these contexts requires sensitivity to spectral contrasts (vowel formants, gender-related spectral shape differences, etc).

Listeners who use cochlear implants have poor spectral resolution, but can still recognize vowels and talker gender adequately [3,4] .

The Questions

Can CI listeners use auditory gender (voice) cues to make phonetic boundary adjustments?

If not, do they use visual gender (face) cues to compensate?

Do CI listeners use lip-reading cues more than NH listeners?

Do CI simulations predict listeners’ sensitivity to phonetic context cues?

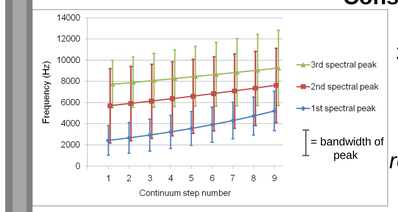
If some listeners cannot make adjustments to vowels and gender, how do we reconcile that with good performance on vowel and gender identification tasks?

Methods

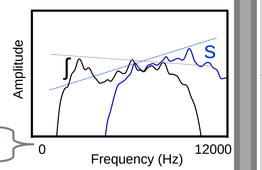
Participants: 10 normal-hearing (NH) listeners, 7 cochlear implant (CI) listeners

Stimuli: words varying between “see” “she” “sue” “shoe”

Consonants: continuum of s/ʃ sounds

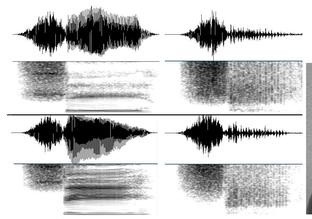


Each fricative had 3 spectral peaks, varying by: *center frequency, bandwidth,* and *relative amplitude (spectral tilt)*

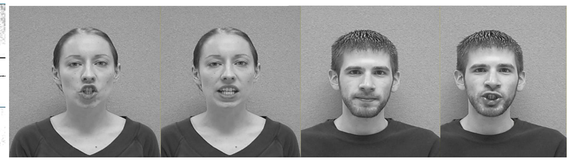


Fricatives were appended to naturally-spoken **vowels** comprising various **contexts**:

Auditory task:
Four **talkers** (2F, 2M)
Two **vowels** (/i/, /u/)
Two **formant transitions**
(for s, ʃ in each vowel context)



Audio-visual task:
Two **talkers** (1F, 1M)
(half aud/vid gender mis-matches)
Two **vowels** (/i/, /u/)
(always matched the video)
Two levels of **lip rounding**
(rounded, unrounded)



Task

4- alternative forced-choice identification
“See” “She” “Sue” “Shoe”
Each of 144 stimuli heard 5x each
(NH listeners also heard 5x in CI simulation)

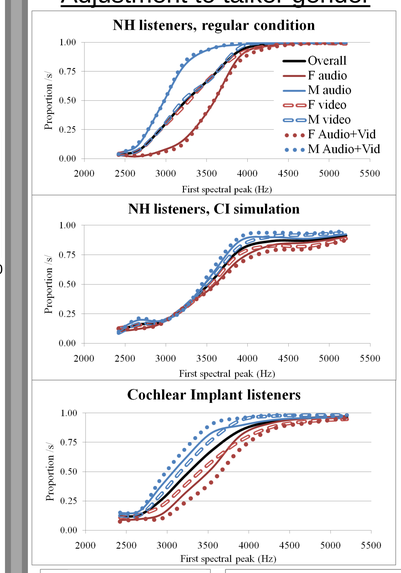
Extra task: Talker gender identification;
F0 was kept natural or neutralized

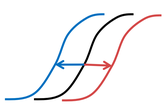
Signal degradation
(NH listeners / CI simulation)

8-channel noise-band vocoder [3,5]
Freq. range 150 – 10,000 Hz (used in Nucleus implants)
300 Hz temporal envelope filter
(sufficient to encode F0 for all talkers)

Results

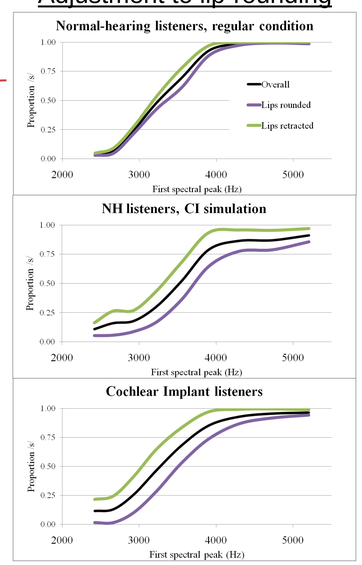
Adjustment to talker gender

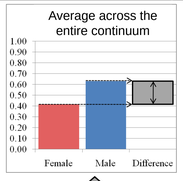




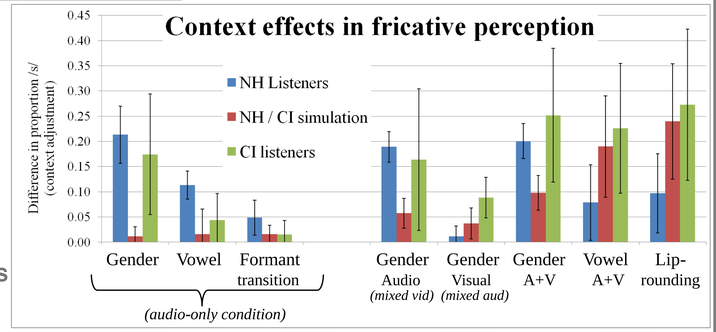
Divergence from overall average response (*black line*) indicates adjustment to the effect shown on the graph.

Adjustment to lip-rounding





Simplified Context effect difference scores shown here →



Gender identification task:

	NH listeners	NH/CI simulation	CI Listeners
Natural F0	**99.1%**	**83.3%**	**90.4%**
F0-neutralized	**97.7%**	**65.3%**	**68.9%**

GLMM results [6]

NH listeners made significantly more use of the auditory gender (voice) cues.*
CI listeners made significantly more use of visual gender (face) cues.*
CI listeners made significantly more use of visual lip-rounding cues.*
Lip-rounding cue was stronger for /i/-context, where it is incompatible with the vowel.*
Effects of the vowel context and gender cues were strongest for ambiguous fricatives (interaction between fricative level and context cues). *
Visual lip-rounding cue was stronger for female face than for the male face. *
= all significant at the p < 0.001 level or better)

Conclusions

Normal-hearing (NH) listeners adjust the perceptual /s/-/ʃ/ frequency boundary in response to talker gender, vowel context, and other cues.

Cochlear Implant (CI) listeners do not show the same amounts of context-sensitive phonetic boundary adjustment based on auditory cues.

CI listeners appear to compensate for this by making relatively *more* use of visual cues like lip-rounding and visual cues to gender.

When gender is cued by both audio and visual cues, CI listeners show phonetic context shifts comparable to those by NH listeners.

NH listeners in CI simulations show virtually *no* phonetic context adjustment; CI listeners possess an ability not captured in the simulation.

Diminished auditory gender context effect by CI listeners and NHs in simulations may arise from reliance on F0 for gender identification; This cue is not sufficient to support phonetic context effects.

CI listeners can recognize vowels and talker gender, but not in a way that facilitates normal context-sensitive phonetic boundary shifting.

1. Mann, V.A., Repp, B.H. (1980) Influence of vocalic context on perception of the [ʃ]-[s] distinction. Perc & Psychophys 28(3), 213-228.
2. Strand, E.A., Johnson, K. (1996) Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), Natural language processing and speech technology: Results of the 3rd KONVENS conference, Bielefeld, Oct. 1996 (pp. 14- 26). Berlin: Mouton.
3. Friesen et al. (2001). “Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,” JASA, 110, 1150-1163.
4. Fu et al. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. JASA 118, 1711–1718
5. Shannon et al. (1995). Speech recognition with primarily temporal cues. Science, 270, 303–304.
6. Bates, D. and Maechler, M. (2010). “lme4: Linear mixed-effects models using S4 classes.” R package version 0.999375-37. http://CRAN.R-project.org/package=lme4

This research is supported by National Institutes of Health R01 DC 004786 (Monita Chatterjee) and T32 DC000046-17 (Arthur N. Popper).