Acoustic cues underlying the adjustment to talker sex in perception of fricative consonants in listeners with cochlear implants

Ashley N. Moore, Au.D. CCC-A Matthew B. Winn, Au.D., Ph.D.

University of Washington Listen Lab

Conference on Implantable Auditory Prosthesis 2017

INTRODUCTION

Different talkers have different voices. This introduces variability that we need to accommodate in order to identify speech. People with normal hearing can do this, and CI users also accommodate, despite a degraded signal.

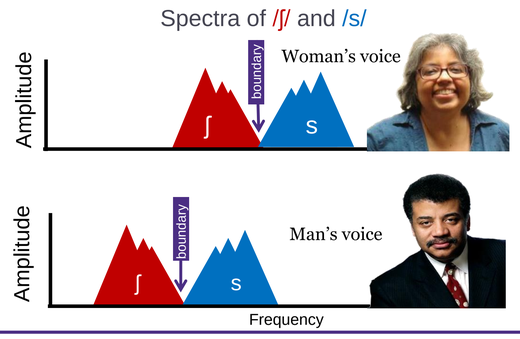
Do CI listeners adapt using the same strategy as individuals with normal hearing?

This adaptation is called...

PHONETIC ACCOMODATION OF TALKER GENDER

A well known example of this is seen in fricatives: /s/ and /ʃ/ ("sh") have different acoustic properties when spoken by a man compared to a woman; Frequency peaks are lower for a man's voice

A shift in the ***perceptual boundary*** between /ʃ/ and /s/ will reflect perception of subtle acoustic differences between talkers [1,2,3]



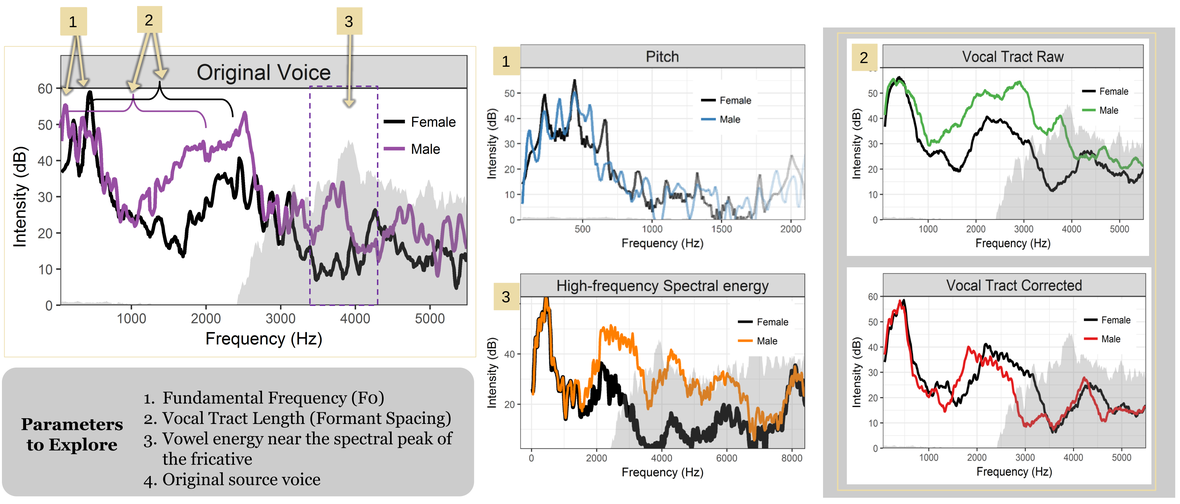
The problem:

- Perception of talker difference should be based on the perception of vocal tract size (formant spacing)
- CI listeners do not have the spectral resolution to be able to reliably perceive formant spacing [4]
- NH and CI listeners might accommodate using **different** acoustic features of the talker's voice.

The strategy to solve the problem:

We are independently manipulating parameters of voice acoustics to see which are the strongest contributors to accommodation of talker sex

Isolating potential acoustic cues for talker gender



Parameters to Explore

1. Fundamental Frequency (Fo)
2. Vocal Tract Length (Formant Spacing)
3. Vowel energy near the spectral peak of the fricative
4. Original source voice

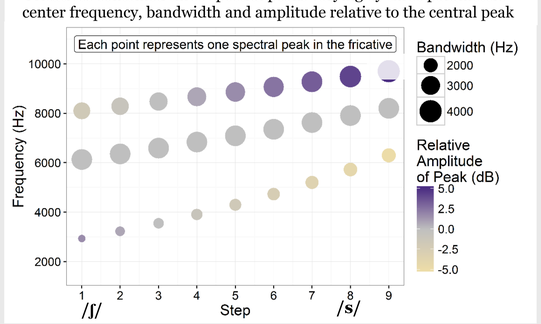
METHODS

PARTICIPANTS: 8 listeners with cochlear implants (ages 44 – 87 y)
20 listeners with normal hearing (ages (18 – 50)

PROCEDURE: Click on the word that is spoken

STIMULI: **8-step continuum of fricative sounds ranging from /ʃ/ ("sh") to /s/** appended to /i/ and /u/ vowels
vocal cues to gender (in the vowel) were acoustically manipulated

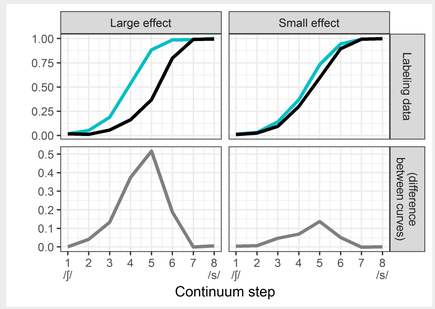
Fricatives contained three spectral peaks varying by three parameters: center frequency, bandwidth and amplitude relative to the central peak



Listener	Sex	Age	Device Type	Implanted Ear(s)	Etiology of Deafness	CI Experience
C101	F	54	MedEl	Bilateral	Sudden SNHL	5 yr
C102	F	64	Cochlear	Right	Idiopathic	2 yr
C103	F	53	AB Hi Res 90k	Bilateral	Genetic	22 yr
C104	M	64	Advanced Bionics C1	Bilateral	Ototoxicity	15 yr
C105	F	47	Cochlear N-6	Bilateral	Progressive SNHL	5 yr
C106	M	87	AB	Bilateral	Noise Induced SNHL	30 yr
C107	M	67	Cochlear N-6	Bilateral	Progressive SNHL	4 yr
C108	F	71	Cochlear N-6	Left	Genetic	26 yr
C109	M	44	AB Naida C90	Right	Genetic	1 yr
C110	M	78	Cochlear N-6	Bilateral	Progressive SNHL	14 yr

It's not all about /s/ and /sh/!
These fricative continuums allow us to probe the effects of gender *cues within the vowel*.

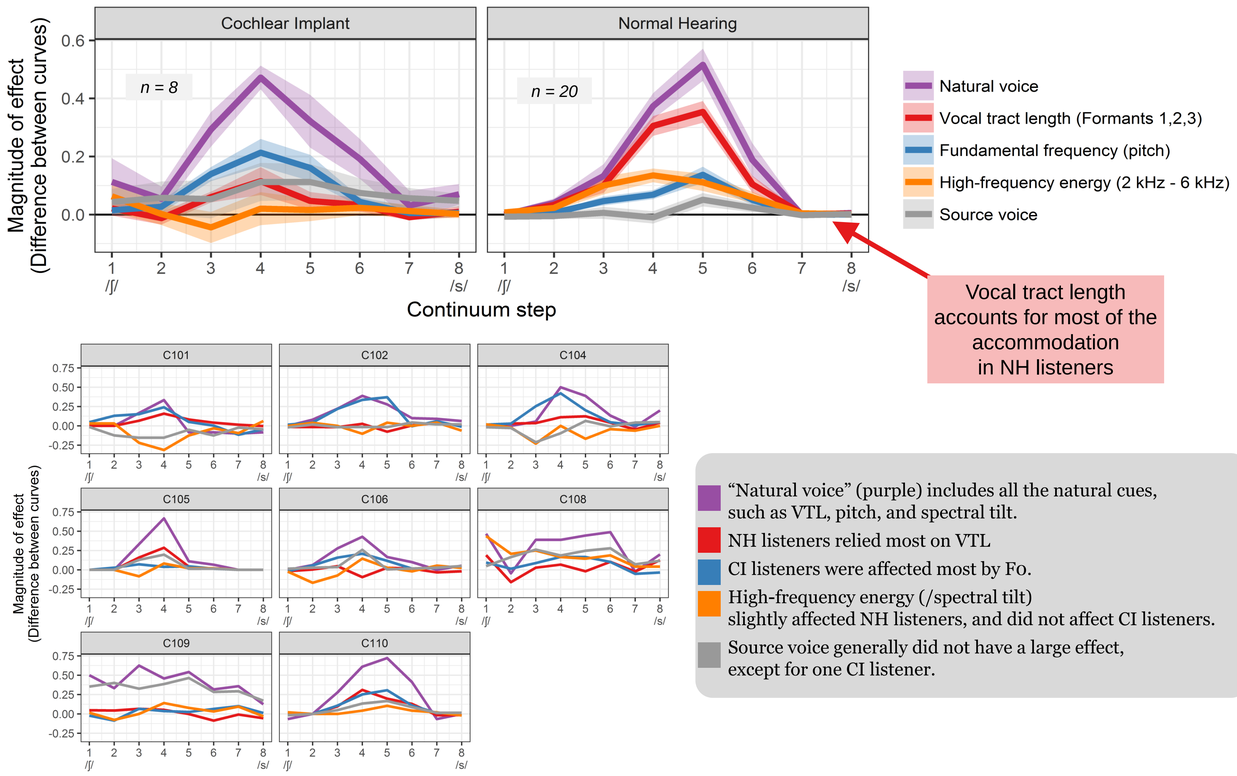
ANALYSIS



- Fricatives are labeled in the context of the vowel, which contains all acoustic cues for talker gender.
- Vowel context influences how the listener labels fricatives (the psychometric function shifts to the left or right).
- The effect should be greater for more ambiguous stimuli in the middle of the continuum.

More space between the two functions means a greater effect of that acoustic cue

RESULTS



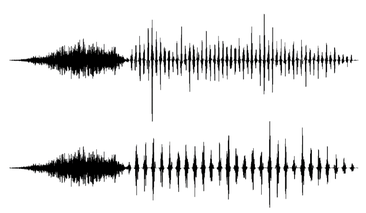
Vocal tract length accounts for most of the accommodation in NH listeners

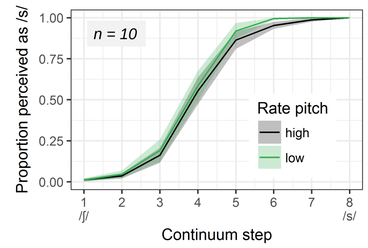
- "Natural voice" (purple) includes all the natural cues, such as VTL, pitch, and spectral tilt.
- NH listeners relied most on VTL
- CI listeners were affected most by Fo.
- High-frequency energy (/spectral tilt) slightly affected NH listeners, and did not affect CI listeners.
- Source voice generally did not have a large effect, except for one CI listener.

Rate Pitch? In MY phonetic categorization?

- If CI listeners were using pitch, and they lack harmonic pitch, they must have been using rate (temporal) pitch.
- Is rate pitch a "proxy" cue for vocal tract size?
- Can NH listeners use rate pitch to accommodate phonetic perception?

- Here, inharmonic noise was fully amplitude modulated (80 or 160 Hz) and filtered to sound like /i/ or /u/





- Rate pitch was not effectively used by listeners with normal hearing as a cue to accommodate phonetic perception, indicating that the use of rate pitch for accommodation might be *learned* rather than automatic.
- Fos used (80 Hz and 160 Hz) are within the limits of rate pitch perception

CONCLUSIONS

- CI listeners use a different strategy to accommodate differences in voice acoustics.
  - NH listeners rely primarily on vocal tract length
  - CI listeners rely primarily on Fo (pitch)
    - Fo may be used as a proxy for VTL when the signal is too degraded to extract formant information
- The strategy used by CI listeners might explain their difficulties in everyday environments
  - Fo is not the most direct index of vocal tract differences, and is not easy to perceive with a CI
    This might explain some difficulty of CI listeners in perceiving multiple talkers .
- Although NH listeners use pitch as a strong cue for *identification* of gender, they did not utilize pitch to *accommodate* to different talkers' voices when it was isolated from VTL information.
  - CI users appear to learn to rely on Fo when access to VTL information is insufficient.

[1] Mann,V., & Repp, B. (1980). Influence of vocalic context on perception of the /ʃ/ -/s/ distinction. Perception & Psychophysics
[2] Munson, B. Jefferson, S., & McDonald, E. (2006). The influence of perceived sexual orientation on fricative identification. JASA
[3] Winn, Rhone, Chatterjee, & Idsardi (2013). The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. *Frontiers in Psychology*
[4] Gaudrain, E. & Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant simulations. JASA

Steven Gianakas, Tiffany Mitchell and D.J. Audet assisted with data collection. Financial support for this project is provided by NIH-NIDCD R03 DC014309 (Winn)