UNIVERSITY of WASHINGTON
Listen Lab

Acoustic cues underlying the adjustment to talker sex in perception of fricative consonants

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

SPEECH AND HEARING SCIENCES

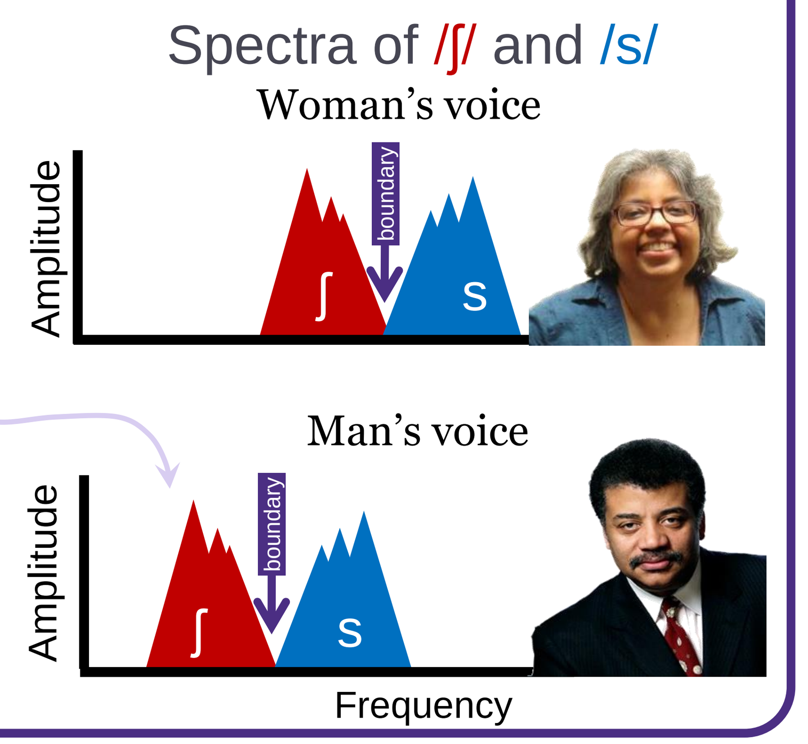
Acoustical Society of America
2016

INTRODUCTION

We hear the same speech sound using different acoustic parameters depending upon if it is spoken by a woman or a man

This adaptation is called…

PHONETIC ACCOMODATION OF TALKER GENDER



A well known example of this is seen in fricatives: /s/ and /ʃ/ 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 differences in speech production

The problem:

There are a lot of acoustic differences between women's and men's voices
We do NOT know which acoustic cues drive this behavior

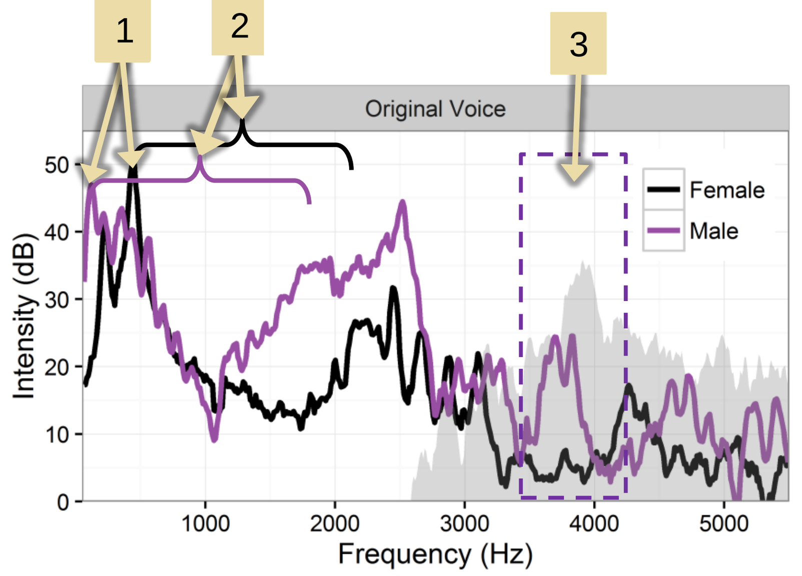
Do listeners adjust based on basic cues to gender such as voice pitch?
Do we normalize to the vocal tract dimensions?
Are we simply sensitive to basic peripheral auditory contrast?

The strategy to solve the problem:

We are independently manipulating parameters of voice acoustics to see which are the strongest contributors to this effect

Parameters to Explore

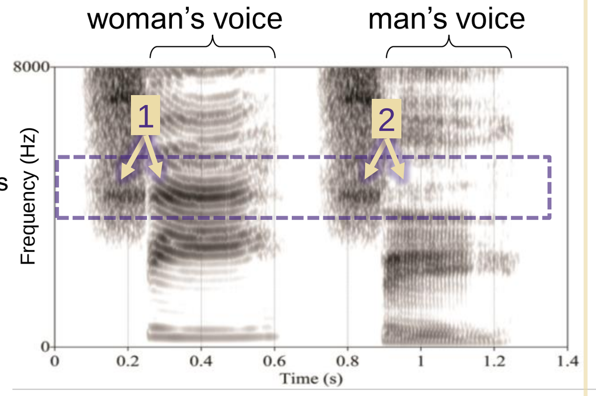
1. Fundamental Frequency (F0)
2. Vocal Tract Length (Formant Spacing)
3. Vowel energy near the fricative spectral peak (local spectral contrast)



This is what we mean by "vocal tract length", and how it relates to the fricative

1. vowel formant for the woman's voice (left) aligns with the fricative spectral peak, reducing spectral contrast (leading to /ʃ/ perception).

2. More local spectral contrast for the male vowel formants, leading to /s/ perception



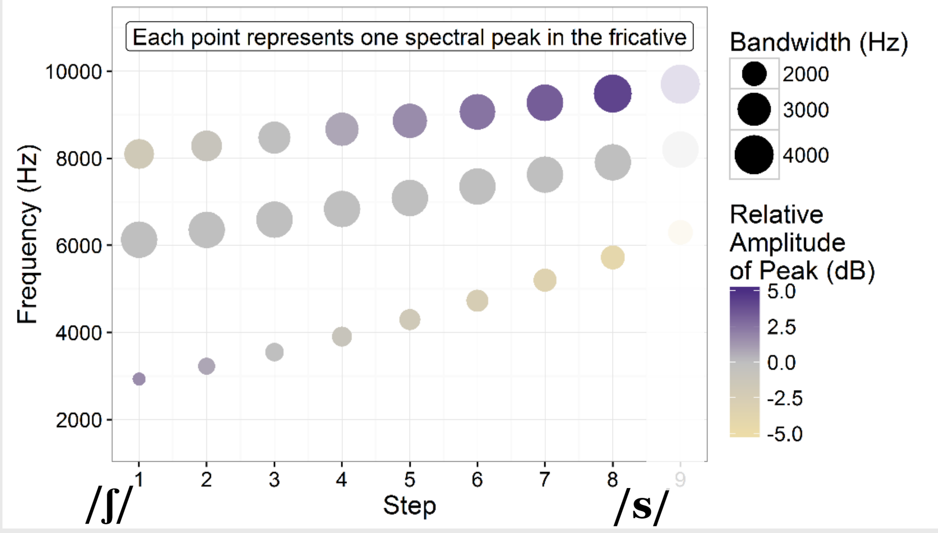
METHODS

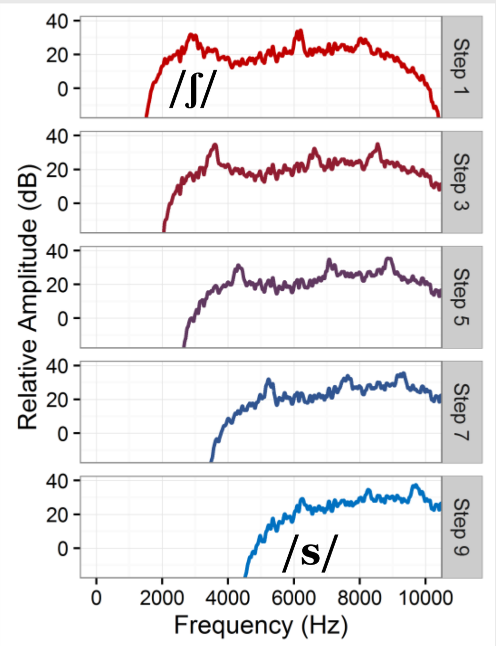
PARTICIPANTS: 10 listeners with normal hearing (ages 18 – 30 y)

PROCEDURE: Click on the word that is spoken →

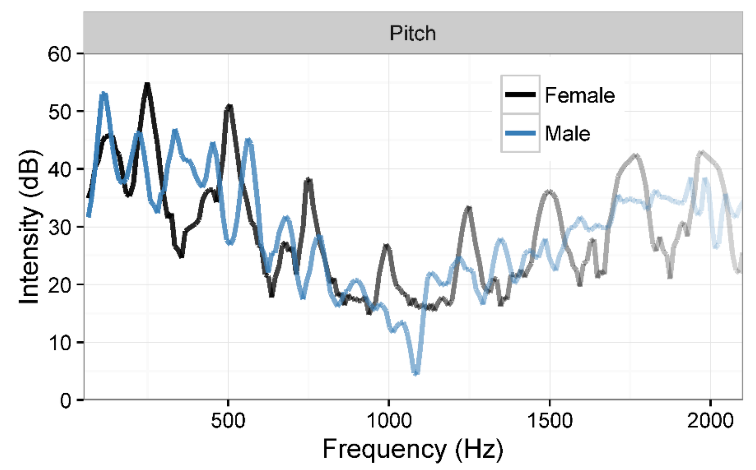
STIMULI: **8-step continuum of fricative sounds ranging from /ʃ/ ("sh") to /s/**
appended to /i/ and /u/ vowels
spoken by a female or a male talker

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

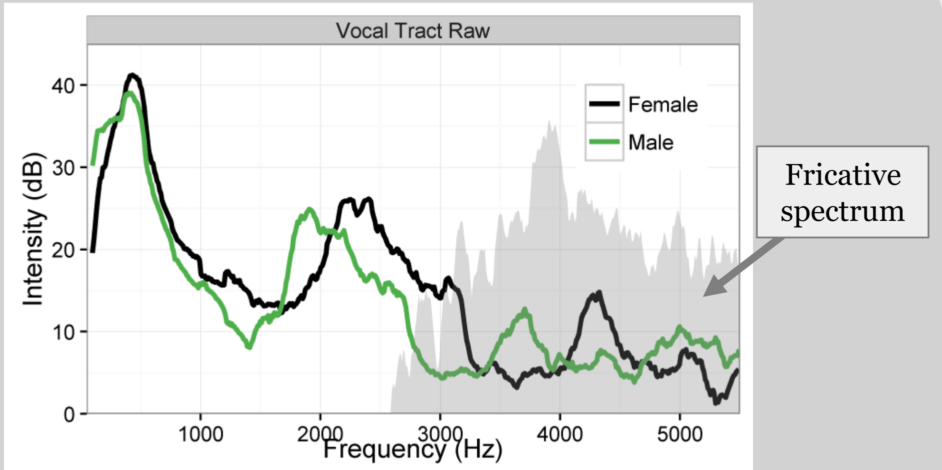




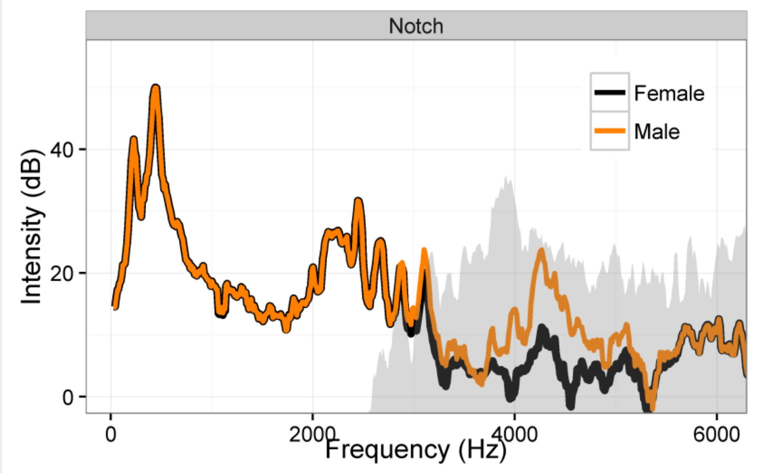
Isolating potential acoustic cues for talker gender



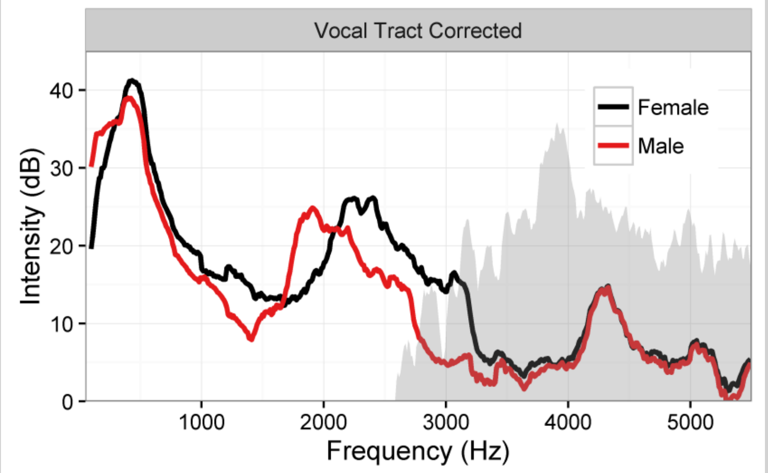
Pitch contour was controlled using PSOLA in Praat. Fundamental frequencies of each voice were multiplied by the ratio of fundamental across female and make voices



To transition the vocal tract length toward being more like the opposite sex, all frequencies were multiplied by the ratio of the formants of the female and male talkers. Note how all frequencies are affected.

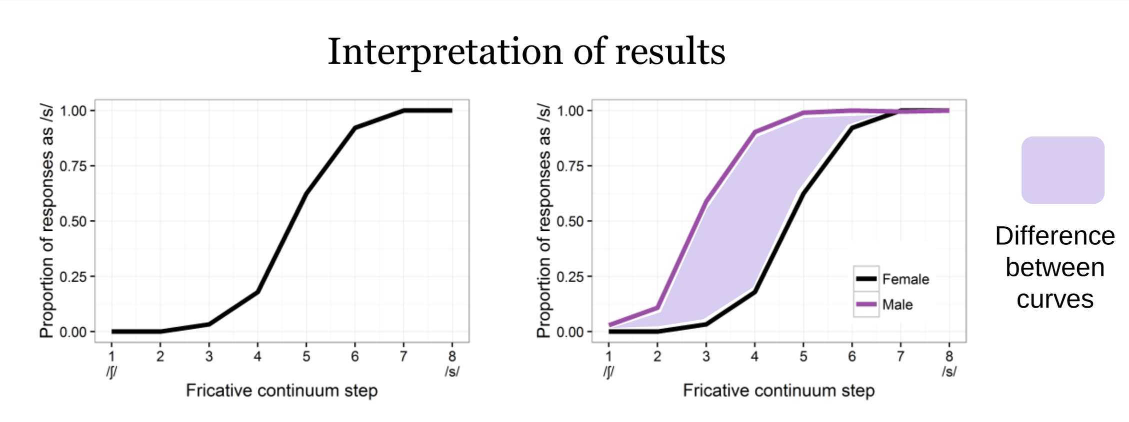


Notch filtering +/– 8 dB intensity of 4000 Hz (+/- 3rd octave band) was used to create varying degrees of local spectral contrast near the first frequency peak of the most ambiguous fricative



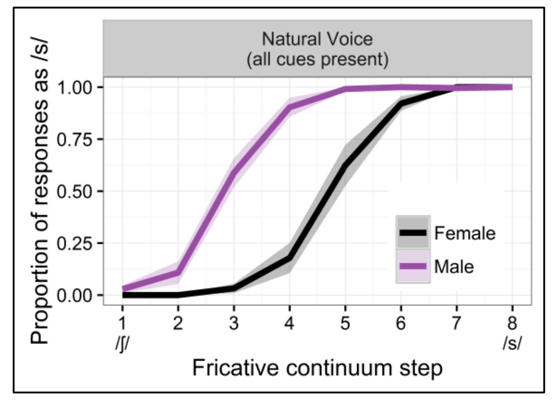
Across VTL manipulations, high-frequency energy above 3200 Hz was equalized to control for differences in energy that could confound formant spacing vs. high-frequency spectral contrast

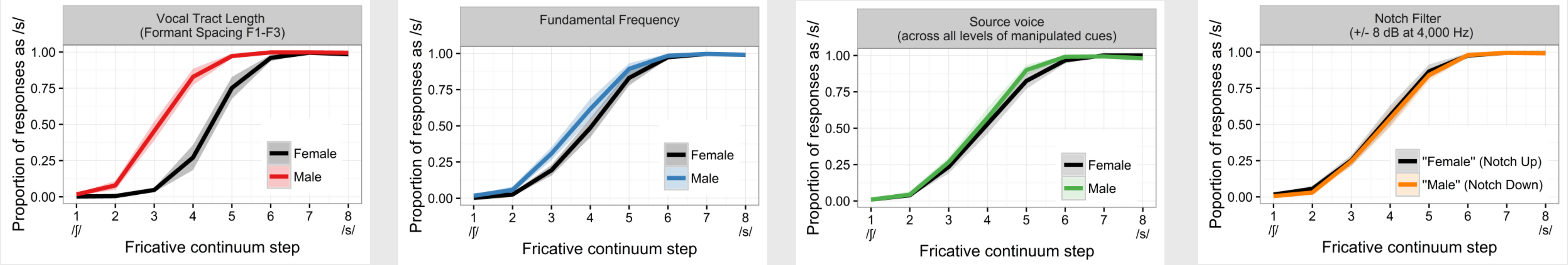
RESULTS



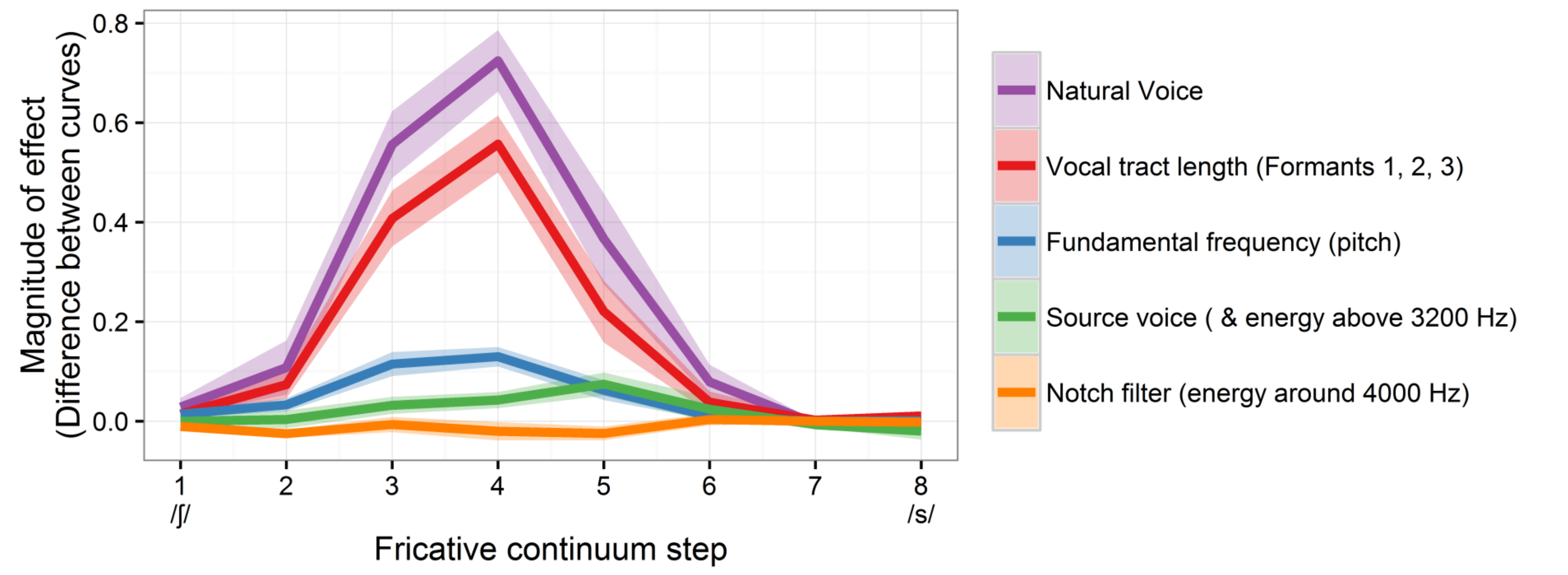
A full change of talker sex produces the plot on the right. This includes all natural cues, such as vocal tract length, pitch, and spectral shape.

This is the FULL accommodation effect





- Formant shifts due to changes in vocal tract length had the greatest effect on the perception of /s/ vs /ʃ/.
- A modest effect was seen for fundamental frequency.
- The effects of spectral contrasts introduced by notch filtering or source voice (energy above 3200 Hz) was minimal.



Differences in formant spacing imposed by differences in vocal tract length had the greatest effect on the perception of /s/ vs /ʃ/, in fact, this cue accounted for most of the accommodation.

CONCLUSIONS

- There are multiple acoustic cues that can inform accommodation of talker gender when perceiving fricative sounds
- Vocal tract length (namely F1 and F2 spacing) is the cue that accounts for the majority of the talker accommodation effect.
- Further investigation will be conducted to evaluate the effect of vocal quality and spectral contrasts (energy above 3200 Hz) between the spectral peak of the fricative and tilt of the vowel spectrum for each talker.

[1] Mann,V., & Repp,B.(1980). Influence of vocalic context on perception of the /ʃ/ -/s/ distinction. Perception & Psychophysics
[2] Munson, B. Jefferson, S.V., & McDonald, E.C.(2006). The influence of perceived sexual orientation on fricative identification. Journal of the Acoustical Society of America
[3] Strand, A., Johnson, K (1996). Gradient and visual speaker normalization in the perception of fricatives. *Results of the 3rd KOVENS Conference, Bielefeld, October, 1996*. Berlin: Mouton de Gruyter, pp 14-26.
[4] 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*

ACKNOWLEDGEMENTS
A special measure of thanks is extended to lab members who assisted with this project: Steven Gianakas assisted with data collection. Tiffany Mitchell, who assisted with audio equipment set up and support, D.J. Audet who assisted with data management, and Gabrielle O'Brien for input and collaboration.
Financial support for this project is provided by NIH-NIDCD R03 DC014309 (Winn)