Electrophysiological signatures of speech recognition

Kathryn Schelonka, Chris Graulty, Enriqueta Canseco-Gonzales, and Michael Pitts

Reed College, Portland, Oregon

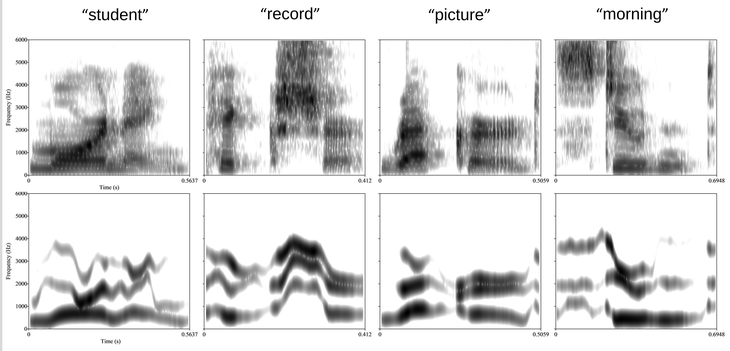
SCALPLAB
Sensation Cognition Attention Language Perception

Introduction

- Previous research supports the existence of specialized neural mechanisms for processing speech vs. non-linguistic sounds.
- The time course for when speech processing diverges from sound processing is unknown.
- Sine-wave-speech (SWS) is specially filtered speech that naïve listeners hear as "electronic whistles," but informed listeners can hear as speech.
- SWS is an ideal tool for this research because a physically identical stimulus can be perceived as either speech or sound.
- Event-related potentials (ERPs) allow noninvasive and temporally precise observations of neural activity.
- **The purpose of this study was to combine SWS and ERPs to determine when the brain first differentiates speech from sound.**

Methods

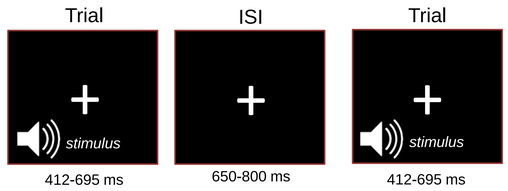
- *Participants* (n=16): 5 male, age range 18-30 years old.
- *Stimuli.* Four two-syllable nouns with similar frequency and concreteness ratings (unfiltered words top row, SWS bottom row).



- *Procedure*.

1. Participants were initially told that the stimulus sounds were randomly generated by a computer.
2. Participants then completed a target detection task on the SWS stimuli in "sound mode."
3. Next, participants were queried to determine if they perceived any of the stimuli as speech.
4. Participants were then informed about the linguistic nature of the stimuli and were trained to hear them as speech.
5. Finally, participants completed the same target-detection task in "speech mode."

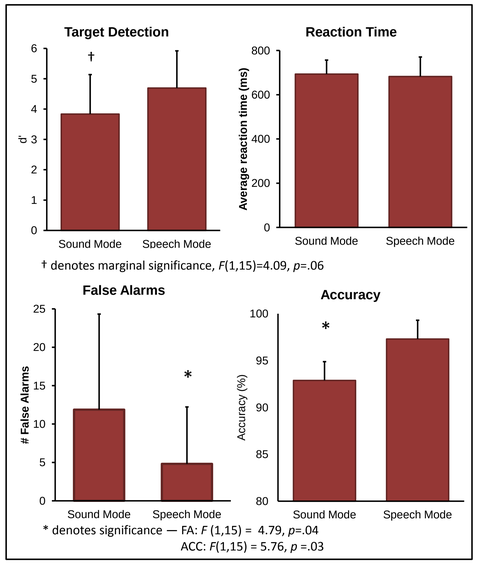
Target detection task



- The four SWS stimuli were presented in random order (each 25% probable)
- Four blocks of trials in "sound mode" and four blocks in "speech mode."
- Target switched each block (order counterbalanced across subjects)

Results

1. Behavioral measures indicate improved performance and significant perceptual change after switching to speech mode.



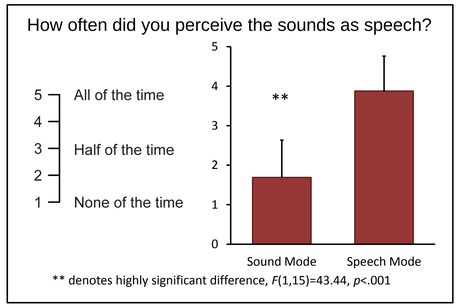
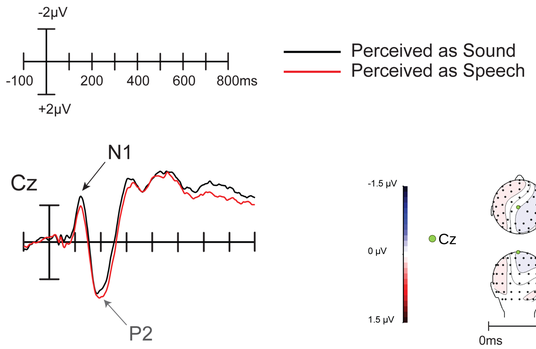


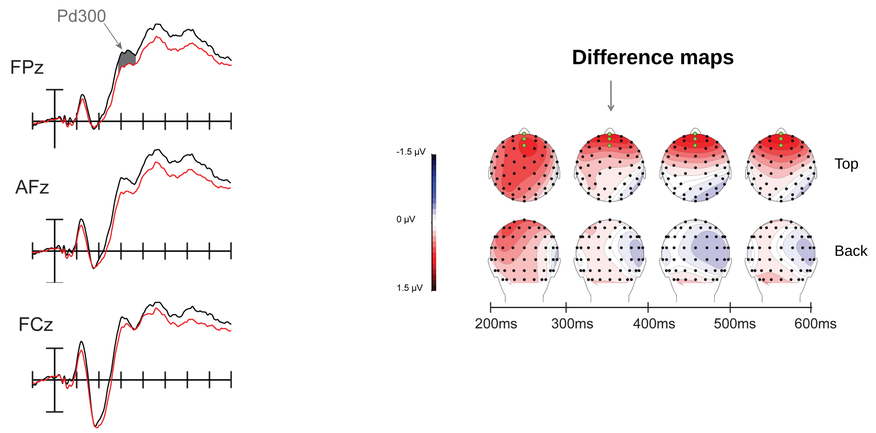
Table of means

	Non-Speech Mode	Speech Mode
Accuracy	92.86 ± 6.16	97.30 ± 3.56
Reaction Time	693.44 ± 63.22	682.49 ± 88.17
False Alarms	11.86 ± 12.45	4.81 ± 7.41
d'	3.84 ± 1.30	4.69 ± 1.23
Perception Rating	1.69 ± 0.95	3.86 ± 0.89

2. In the speech mode, the stimuli elicited a significantly smaller N1 (p<.05) and a marginally enhanced P2 (p<0.1) compared to sound mode.



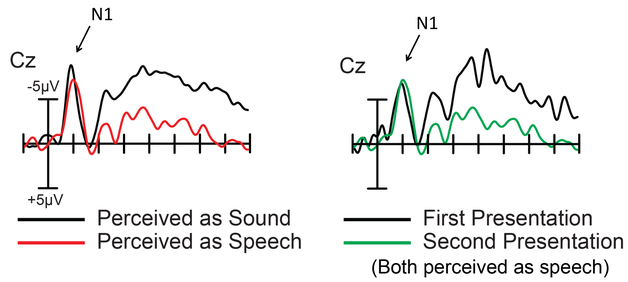
3. The next significant difference observed was a positive shift that first peaked at 300 ms over the frontal scalp.



Follow-Up

- A second study to control for percept order, repetition, and task difficulty was devised and piloted.
- *Participants* (n = 4): 2 male, age range 19-21 years old.
- *Stimuli*. 185 two-syllable nouns matched for frequency and concreteness + 20 reversed two-syllable nouns.
- *Procedure*. Participants were instructed to rate the intelligibility (on a scale of 1 to 4) of the sine-wave-speech before and after the unfiltered word was revealed.

3. The N1 difference was observed only after perceptual change, while the Pd300 was present even when participants reported spontaneously perceiving the SWS stimuli as speech.



Discussion

- The N1 reduction observed in speech mode (Experiment 1) could be due to decreased task difficulty, but the effect's persistence in the pilot study makes it a good candidate for an early speech-related difference.
- The N1 reduction is likely not habituation due to repetition, as the component did not decrease over time within perceptual modes in Experiment 1.
- The marginal P2 effect in Experiment 1 is consistent with previous literature describing P2 enhancement after sound discrimination training.
- The lasting positive difference beginning at 300ms is likely caused by stimulus repetition because it is observed even without perceptual change.
- However, future research is necessary to verify the underlying processes indexed by these ERP differences.

Selected References

Barker, J. and M. Cooke (1999). "Is the sine-wave speech cocktail party worth attending?" *Speech Commun.* 27(3-4): 159-174.

Boersma P & Weenink D (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.16, retrived from http://www.praat.org/

Celsis, P., Doyon, B., Boulanouar, K., Pastor, J., Demonet, J.F., & Nespoulous J.L. (1999). ERP correlates of phoneme perception in speech and sound contexts. *NeuroReport*, 10(7) pp. 1523–1527.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage,* 24, 21–33.

Remez, R. E., P. E. Rubin, et al. (1981). "Speech-Perception without Traditional Speech Cues." *Science* 212(4497): 947-950.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J Cog Neurosci*, 13(7), 994–1005.

Acknowledgments

This project was funded by a Reed College Science Research Fellowship for Faculty-Student Collaborative Research.