

# Intelligibility Modulates Early Attentional Filtering of Competing Speech

Benjamin N. Richardson<sup>1</sup>, Maanasa Guru Adimurthy<sup>2</sup>, Christopher A. Brown<sup>3</sup>, Merri J. Rosen<sup>4</sup>, Barbara G. Shinn-Cunningham<sup>1,5</sup>

<sup>1</sup>Neuroscience Institute, Carnegie Mellon University; <sup>2</sup>Department of Psychiatry, University of Pittsburgh; <sup>3</sup>Department of Communication Sciences and Disorders, University of South Florida;

<sup>4</sup>Department of Neurobiology, Northeast Ohio Medical University; <sup>5</sup>Mellon College of Science, Carnegie Mellon University

## Overview

- Intelligible speech masks target speech more than unintelligible speech [1,2,3], but manipulations of intelligibility vary in the literature: time reversal, N-talker babble, vocoding, filtering, etc. [4,5,6,7].
- Here, we controlled for acoustic differences between intelligible and unintelligible maskers, and measured neural responses.
- Intelligible speech, which more readily masked target speech, modulated the P1-N1 response but not the P300 response, suggesting an early attentional filtering mechanism.

## Stimuli & Color Word Detection Task

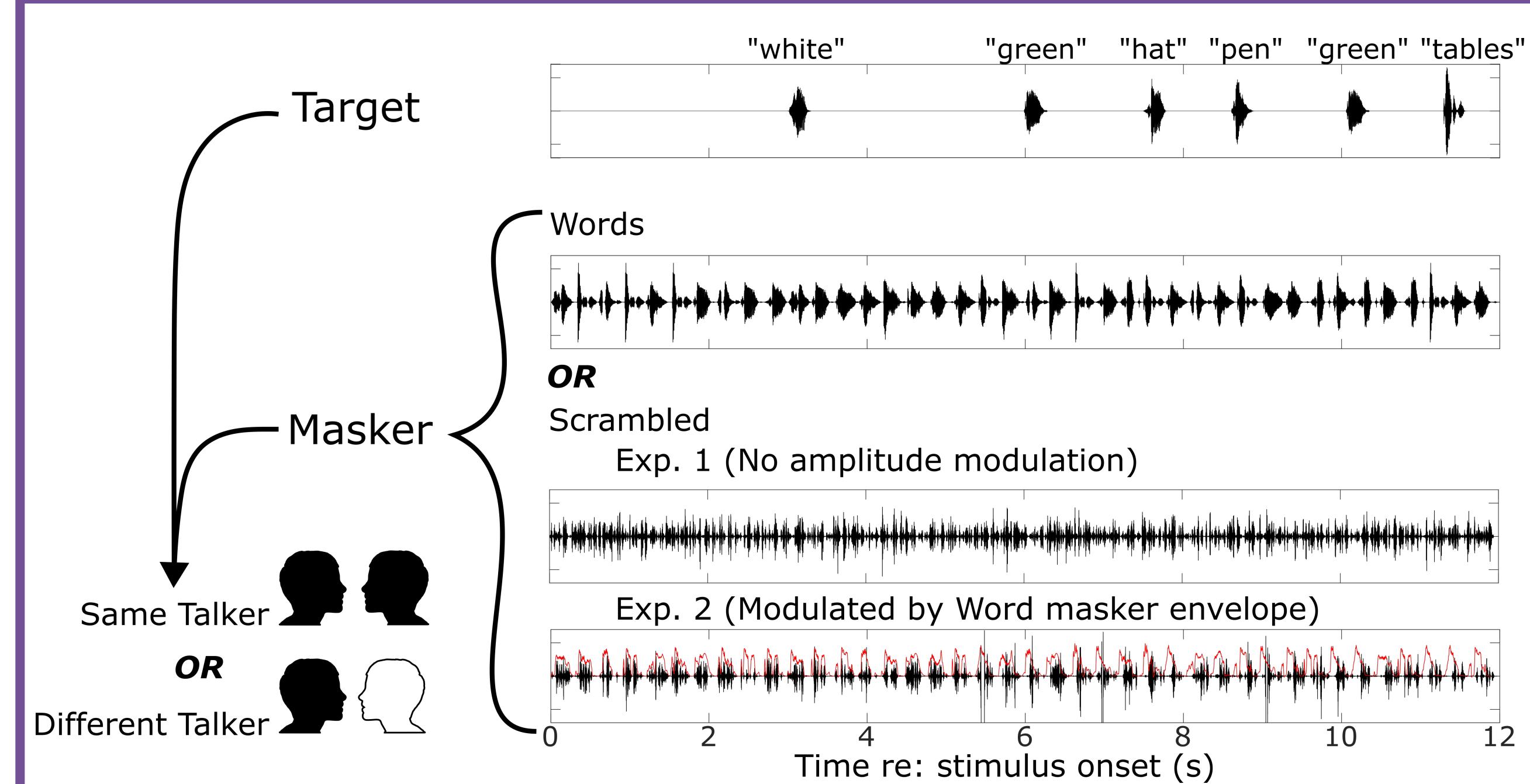


Fig. 1. Schematic of a single task block. Listeners attended to the sporadic target stream (top row), and ignored a simultaneous, isochronous Masker that depended on condition (bottom rows).

- N = 20 normal hearing listeners in each of two experiments.
- Listeners responded to color words in the target stream (Fig. 1).
- Listeners ignored an isochronous masker: either words from the same set ("Words"), or scrambled speech ("Scrambled").
- Target and masker were spoken by the same or different talkers.
- In Exp. 2, masker was amplitude modulated with a Word envelope
- Target and masker were filtered into non-overlapping bands (Fig. 2A).
- Scrambling (Fig. 2B) preserved local spectrotemporal density.

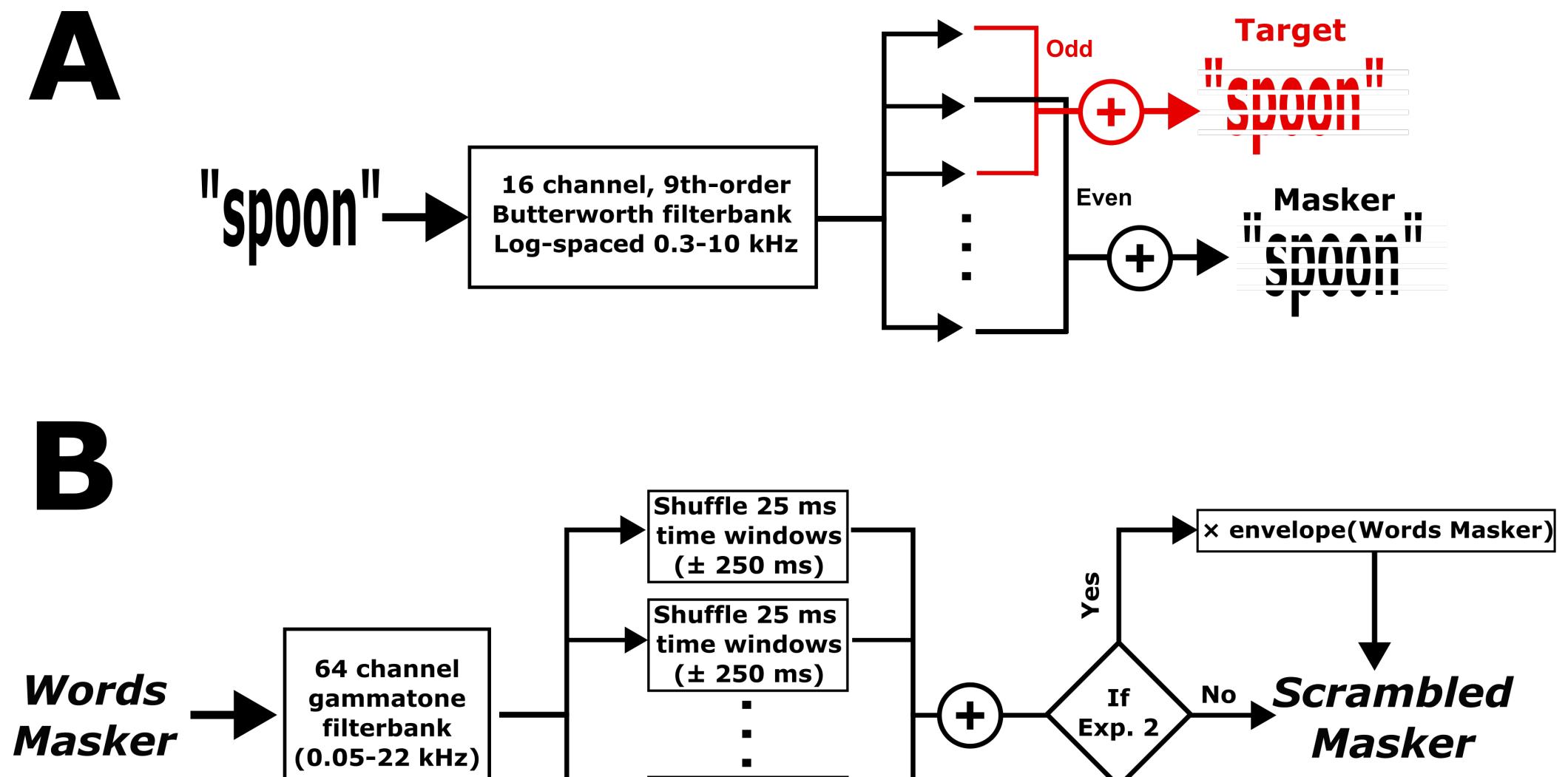


Fig. 2. Controls for energetic masking and spectrotemporal properties. (A) Target and masker were filtered into non-overlapping frequency bands. (B) The scrambling process resulted in similar spectrotemporal density between Word and Scrambled maskers.

## EEG Methods and ERP Analysis

- We measured 32-channel EEG throughout the task.
- Preprocessing included manual artifact rejection, eye blink/saccade rejection using ICA, and rejection of ERPs with magnitude  $\pm 100$  V.
- We measured event-related potentials (ERPs), and calculated...
  - P1-N1 sensory onset responses in frontocentral electrodes
  - P300 color word recognition responses in parietooccipital electrodes

## Behavior Results

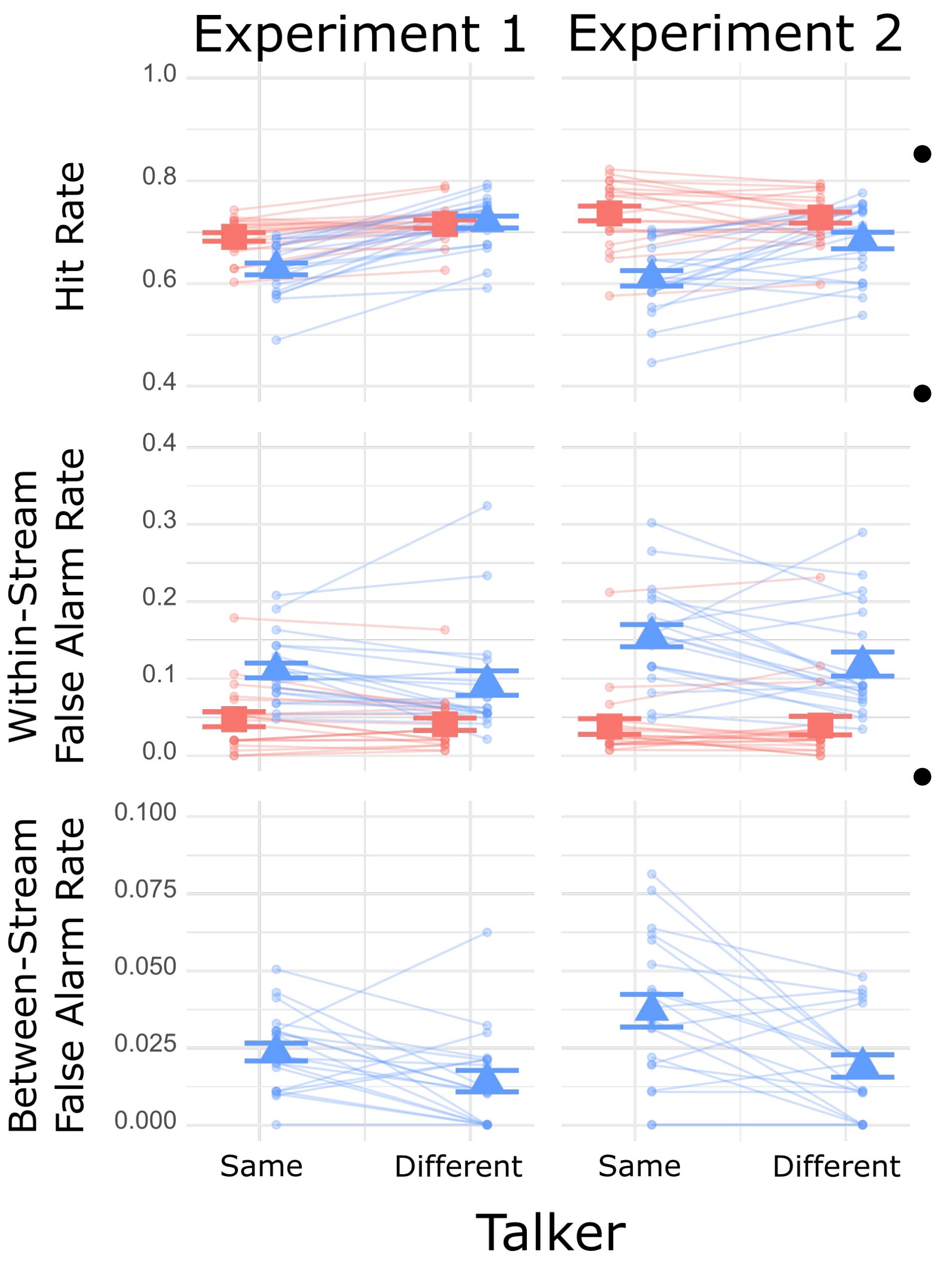


Fig. 3. Individual behavioral data (small symbols connected by thin lines) and across-subject average and standard error (large symbols with error bars).

## ERPs to the masker stream

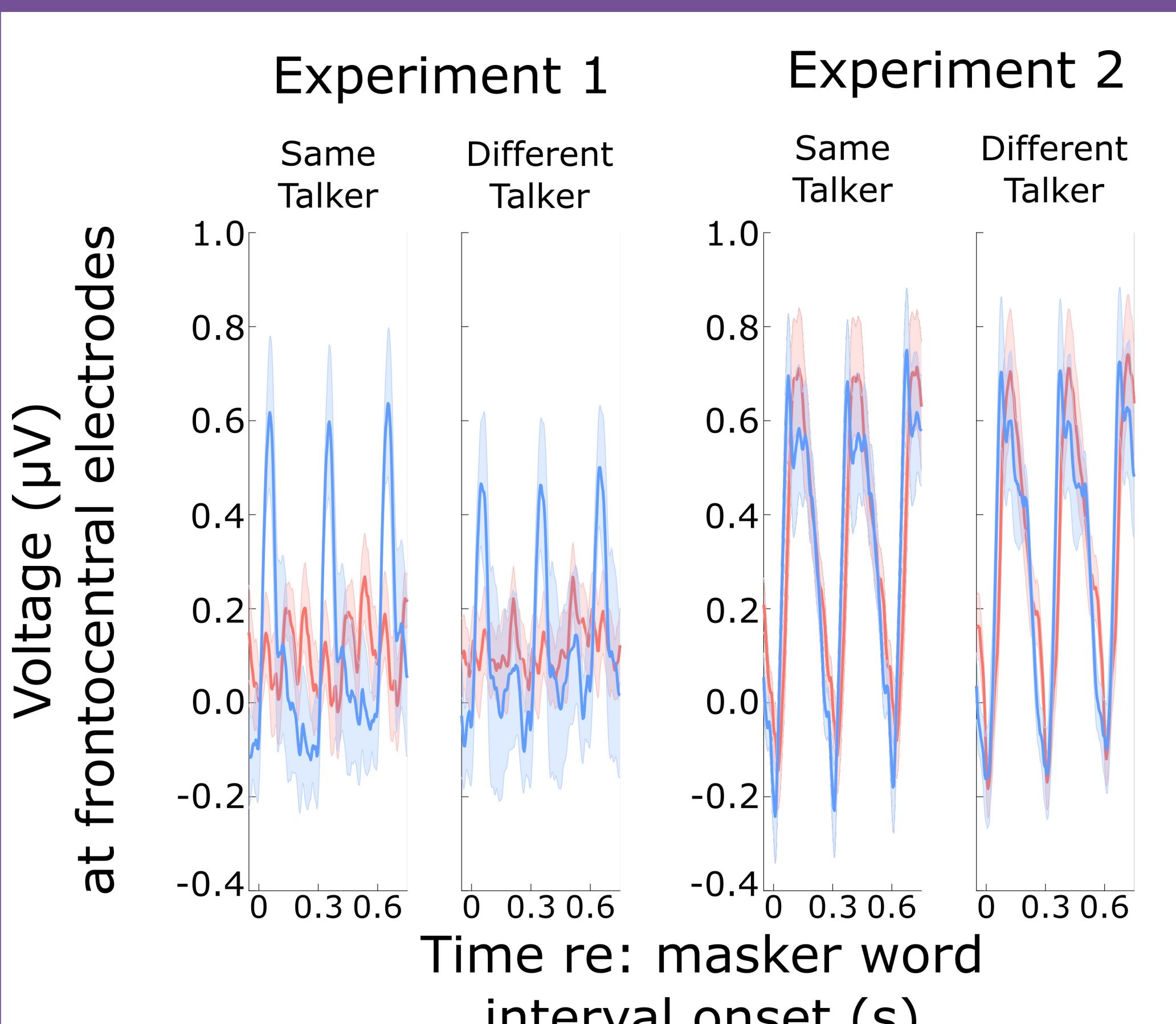
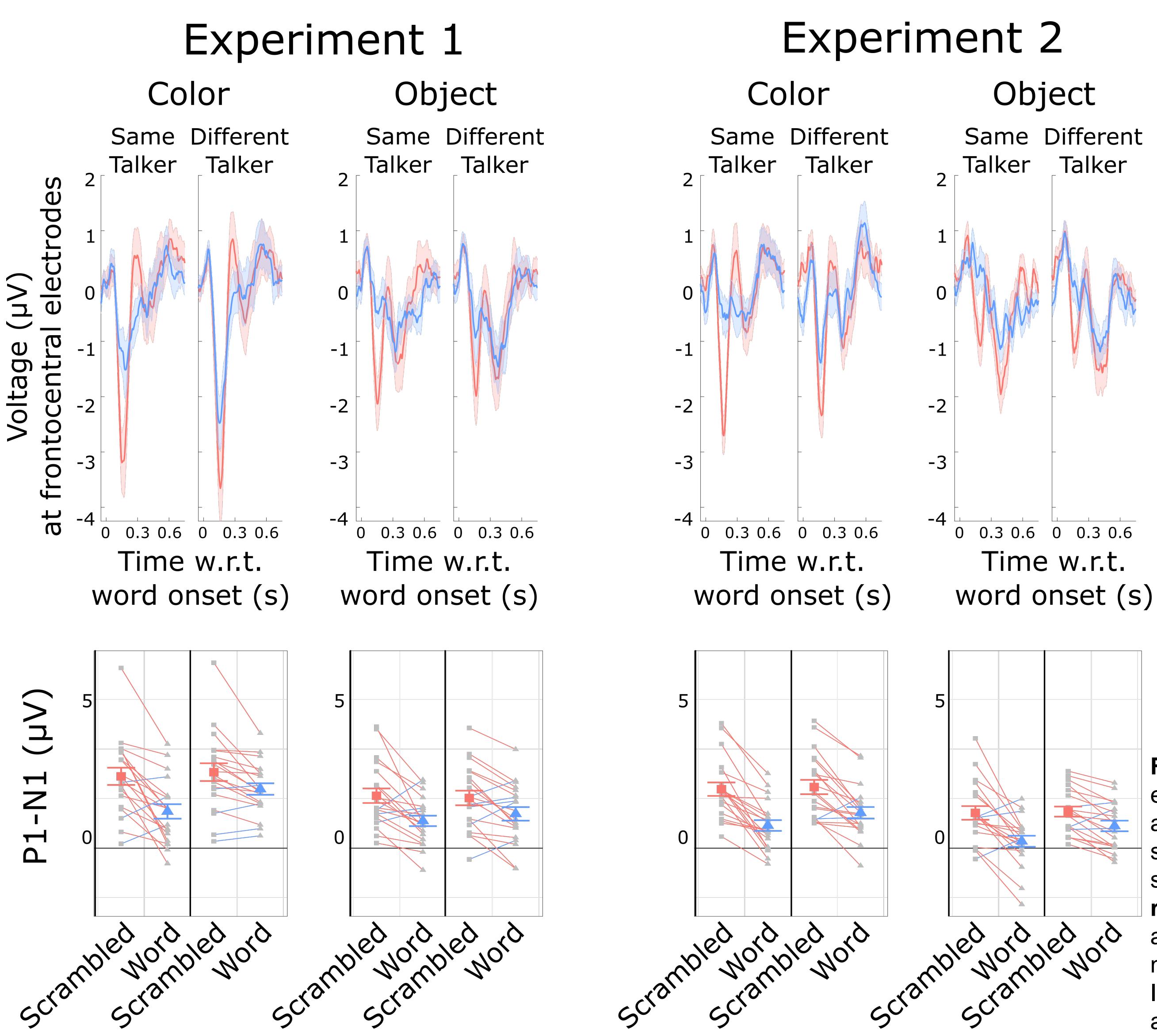


Fig. 4. ERPs evoked by masker streams over frontocentral electrodes, time locked to word onsets in the Word masker.

- Hit rates were higher with a Scrambled than a Word masker, but only with same talker.
- Within-stream false alarm rates (target object words) were higher with a word than scrambled masker, and greater with the same talker.
- Between stream false alarm rates (masker color words) were low, but higher when the talker was the same.

**Masker**  
Word  
Scrambled

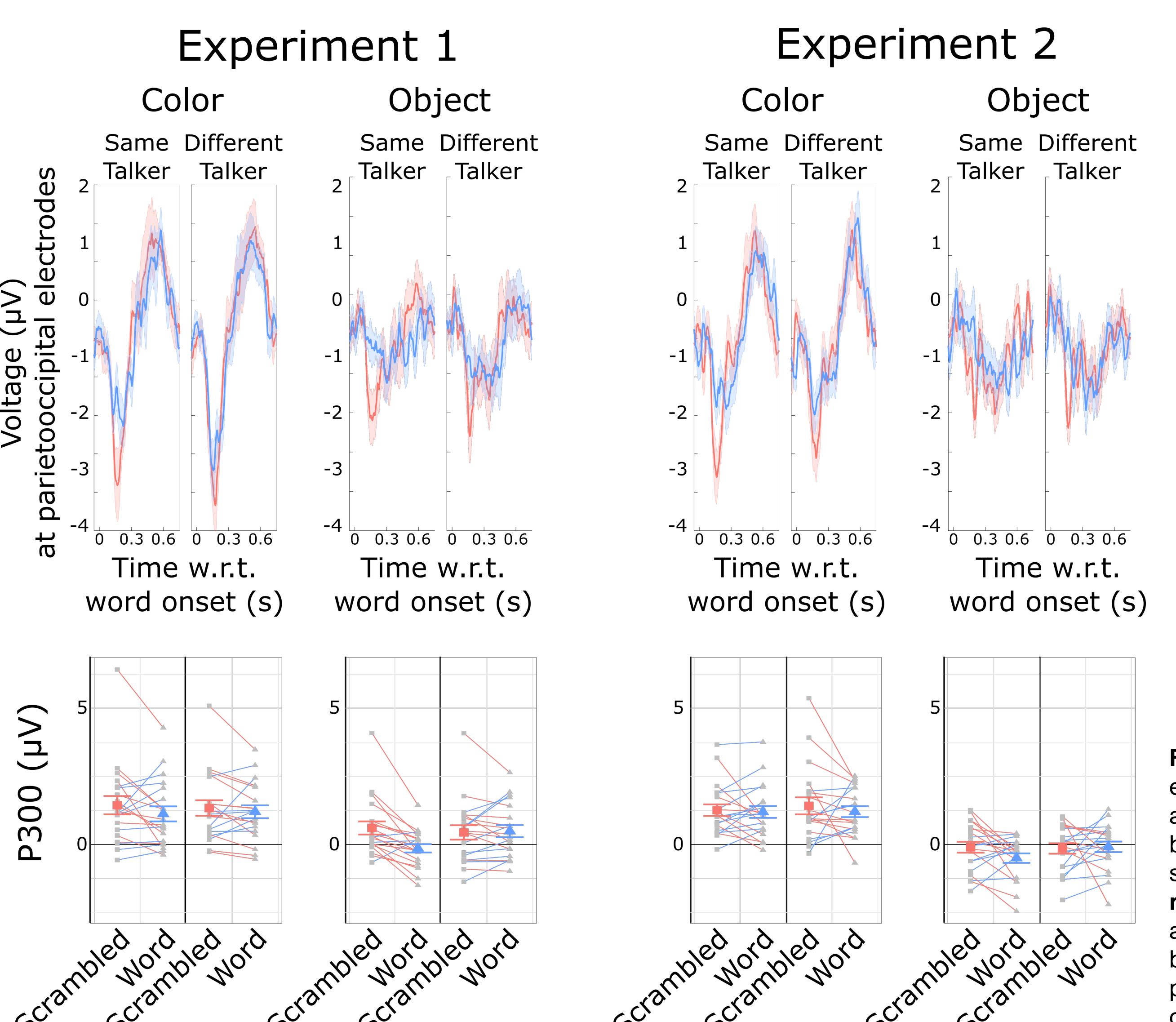
## Sensory (P1-N1) responses to target words



- P1-N1 to target color words was larger than to object words in the target stream
- P1-N1 to the target stream was significantly larger with a Scrambled masker than with a Word masker, especially when the talker was the same.

Fig. 5. ERPs over frontocentral electrodes. Top row: Time traces averaged over participants, with bands showing the across-participant standard error of the mean. Bottom row: P1-N1 magnitude. Group average and standard error of the mean shown by large symbols. Individual participant data (thin lines) are colored according to which P1-N1 is larger.

## Target recognition (P300) responses to target words



- P300 was elicited strongly by color words in the target stream, but not by object words.
- P300 was not modulated by masker intelligibility: No differences between P300 to the target between Word and Scrambled masker conditions, except for Exp 1. (Object, Same Talker)

Fig. 6. ERPs over parietooccipital electrodes. Top row: Time traces averaged over participants, with bands showing the across-participant standard error of the mean. Bottom row: P300 magnitude. Group average and standard error of the mean shown by large symbols. Individual participant data (thin lines) are colored according to which P300 is larger.

## References

- Summers, Robert J., and Brian Roberts. 2020. "Informational Masking of Speech by Acoustically Similar Intelligible and Unintelligible Interferers." *The Journal of the Acoustical Society of America* 147 (2): 1113.
- Van Engen, Kristin J., Jasmina E. B. Phelps, Rajika Smiljanic, and Bharath Chandrasekaran. 2014. "Enhancing Speech Intelligibility: Interactions among Context, Modality, Speech Style, and Masker." *Journal of Speech, Language, and Hearing Research*. JSLR 57 (6): 1908–18.
- Rennies, Jan, Virginia Best, Elin Rovnerud, and Gerald Kidd Jr. 2019. "Enhancing and Informational Components of Speech-on-Speech Masking in Binaural Speech Intelligibility and Perceived Listening Effort." *Trends in Hearing* 23 (January). <https://doi.org/10.1177/232432181985459>.
- Seidman, Steven, Steve Rossen, Harrison Lang, and Richard L. Kraus. 2005. "Neural Correlates of Intelligibility in Speech Investigated with Noise Vocoded Speech—A Positron Emission Tomography Study." *The Journal of the Acoustical Society of America* 120 (2): 1075–83.
- Pollack, Irvin. 2005. "Effects of High Pass and Low Pass Filtering on the Intelligibility of Speech in Noise." *The Journal of the Acoustical Society of America* 120 (3): 2595–66.
- Cooke, Martin, and Youyi Lu. 2010. "Spectral and Temporal Changes to Speech Produced in the Presence of Energetic and Informational Maskers." *The Journal of the Acoustical Society of America* 128 (4): 2059–69.
- Van Engen, Kristin J., and Ann R. Bradlow. 2007. "Sentence Recognition in Native- and Foreign-Language Multi-Talker Background Noise." *The Journal of the Acoustical Society of America* 121 (1): 519–26.

## Acknowledgements



- F31-DC022481 to BNR
- R21-DC018408 to CAB
- R01-DC015988 to BGSC
- R01-DC022699 to CAB & BGSC
- ONR-N00014-23-1-2065 to BGSC
- R01 DC019126 to MJR & BGSC

## Contact

Your future postdoc might be just a scan away...

