MEG responses track lexical processing of continuous narrative speech

Computational Sensorimotor Systems Lab

Christian Brodbeck[1], L. Elliot Hong[2] & Jonathan Z. Simon[1]

[1]University of Maryland, College Park; [2]University of Maryland School of Medicine, Baltimore

Introduction

Aim: study the neural basis of lexical processing of continuous speech

- Generalize and extend results gained from controlled studies to a more natural setting
- Develop framework to assess speech processing at multiple levels simultaneously
- Study influence of factors that affect speech comprehension in realistic contexts

Background: known brain responses to continuous speech

- Acoustic features (Ding & Simon, 2012)
- Phoneme identity (Di Liberto et al., 2015)
- Semantic processing (Broderick et al., 2018)

Gap: lexical processing – transformation from representations based on acoustic features to lexical representations

Approach: natural listening to audiobook segments

1. Clean speech

- Assess predictor variables that index lexical processing
- Identify variables that significantly modulate brain responses

2. "Cocktail party" – two speakers mixed at equal loudness, listeners attend to one while ignoring the other

- Is unattended speech processed lexically?
- Does lexical processing change in the context of a second (distractor) speaker?

Methods

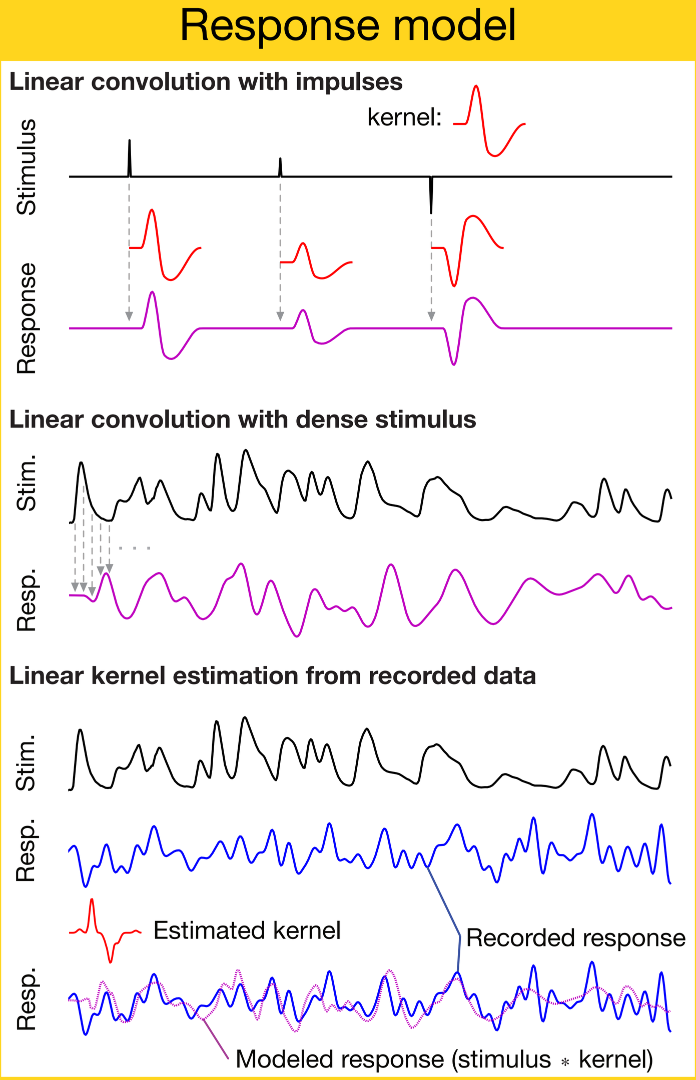
Acoustic predictor variables:

- **Acoustic envelope**: Auditory spectrogram based on model of the auditory periphery (Yang et al., 1992) averaged in 8 frequency bands
- **Acoustic onsets**: derivative of acoustic envelope, with negative values set to 0

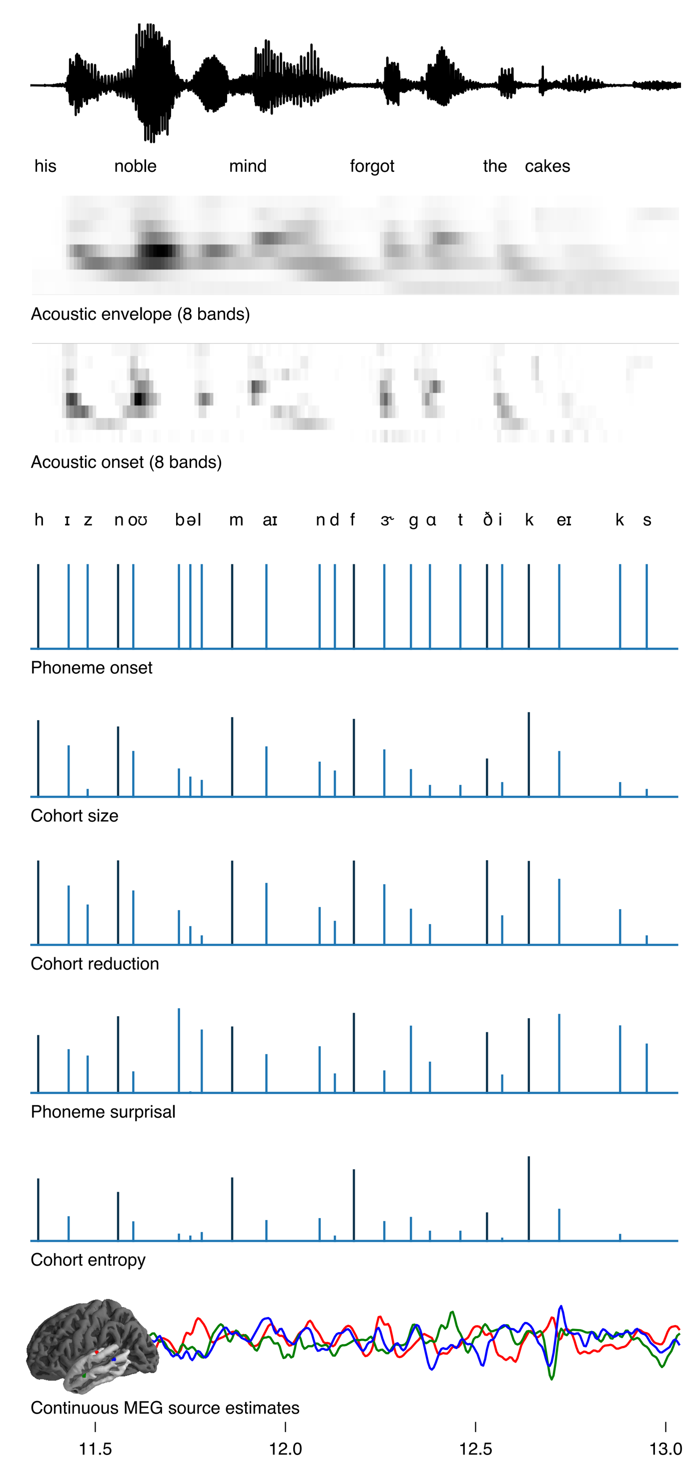
Cohort-based predictor variables: Assume form-based lexicon (word frequencies from SUBTLEX; Brysbaert et al., 2009) and ideal listener; instantiation of the cohort at the first phoneme of each word and update at each subsequent phoneme

- **Cohort size**: Log of the number of words in the cohort after considering this phoneme
- **Cohort reduction**: Log of the number of words removed from the cohort by this phoneme
- **Phoneme surprisal**: inverse of the conditional probability of this phoneme; associated with predictive coding
- **Cohort entropy**: entropy of the cohort; associated with lexical competition
- First phoneme of each word modeled separately

Response model



Model and predictor variables



Modeling response functions

- 157 axial gradiometer whole head MEG (KIT, Kanazawa, Japan)
- 26 adults, one-minute-long audiobook segments (8 solo, 16 mix)
- Distributed minimum norm source estimates
- Each source element modeled as a sum of linear responses to different predictor variables with coordinate descent (cf. Brodbeck et al., 2018; David et al., 2007)

Statistical analysis:

- Significance of each predictor variable assessed by comparing its predictive power against models in which the variable was shuffled
  - Acoustic variables and phoneme onset were temporally misaligned
  - Word onsets were randomly reassigned to different phonemes
  - For cohort-based predictors, temporal locations were kept but values were shuffled
- Localization tests with threshold-free cluster enhancement (Smith & Nichols, 2009) and null distribution based on 10.000 permutations
- Variable selection in clean speech: step-wise removal of non-significant variable with largest *p*-value until only significant variables remained

Results

Responses to clean speech reflect lexical processing

Process localization

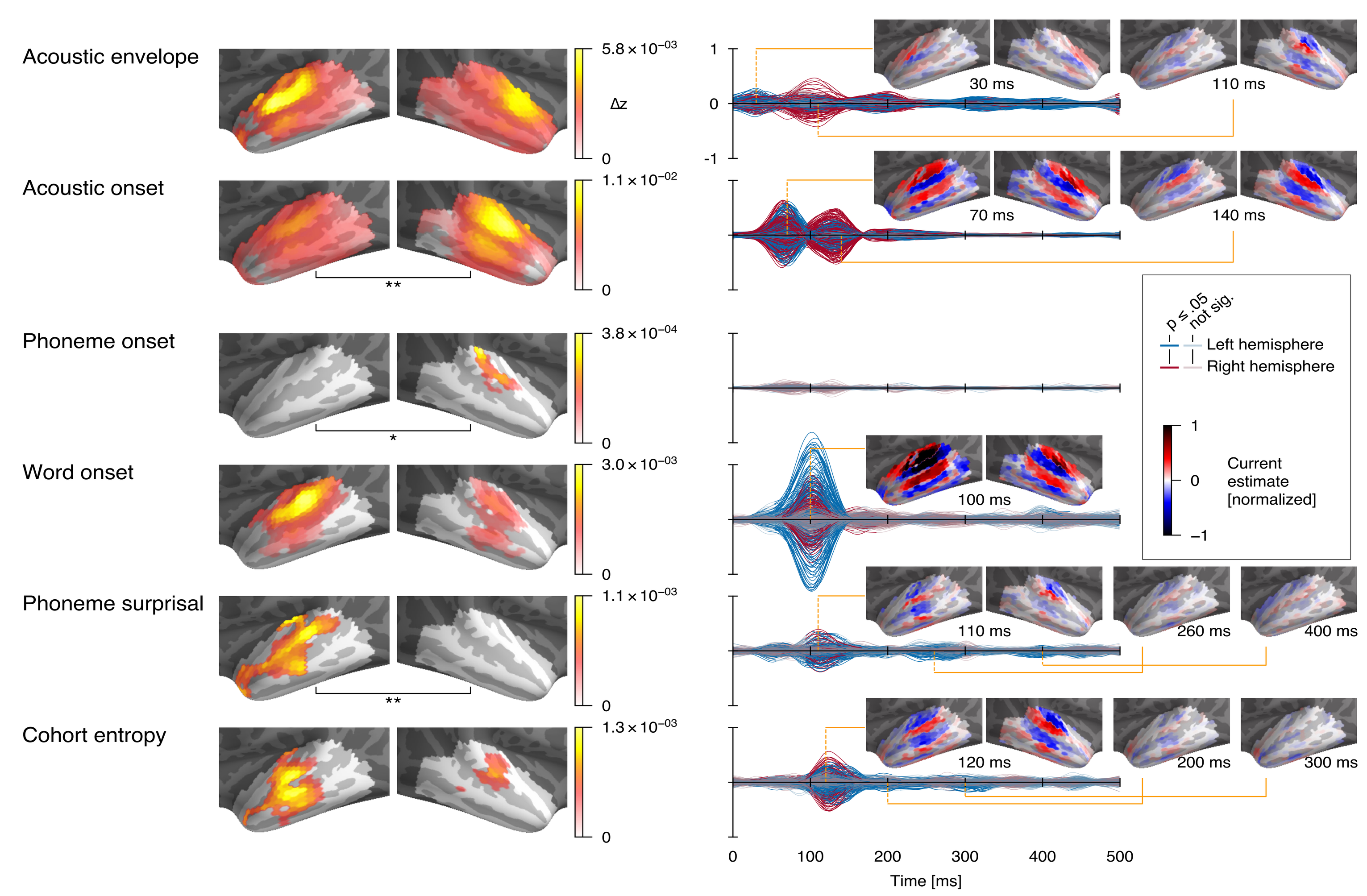
The region in which each predictor variable had significant predictive power (analysis restricted to the temporal lobes).

Predictive power quantified as model fit difference when including the true predictor or a shuffled version; hemispheric lateralization indicated when significant.

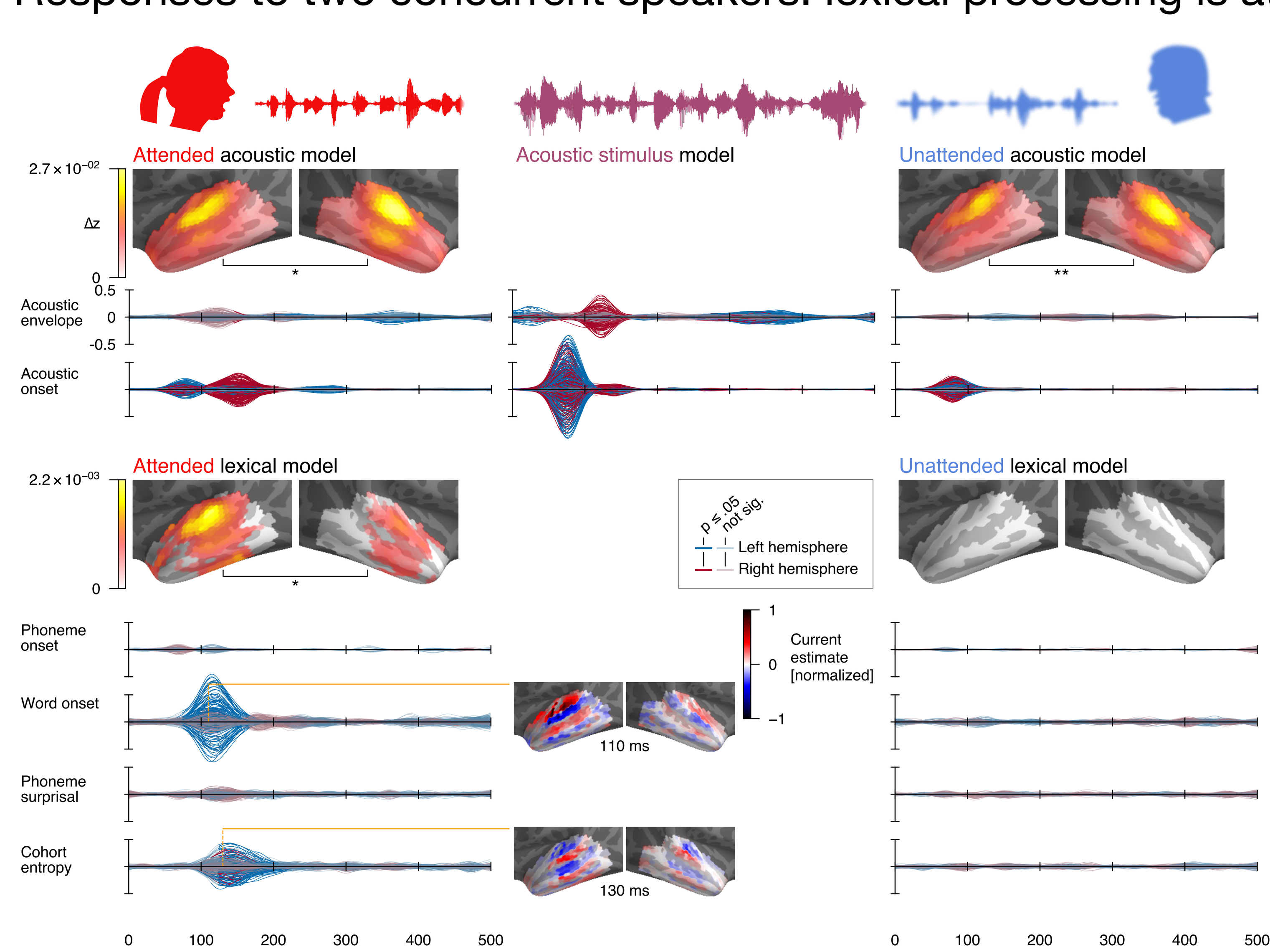
Response time course

Estimated neural response to a unit change in the predictor variable.

Each line represents one neural current source element. Current direction is shown relative to the cortical surface, leading to alternating direction due to cortical folding.

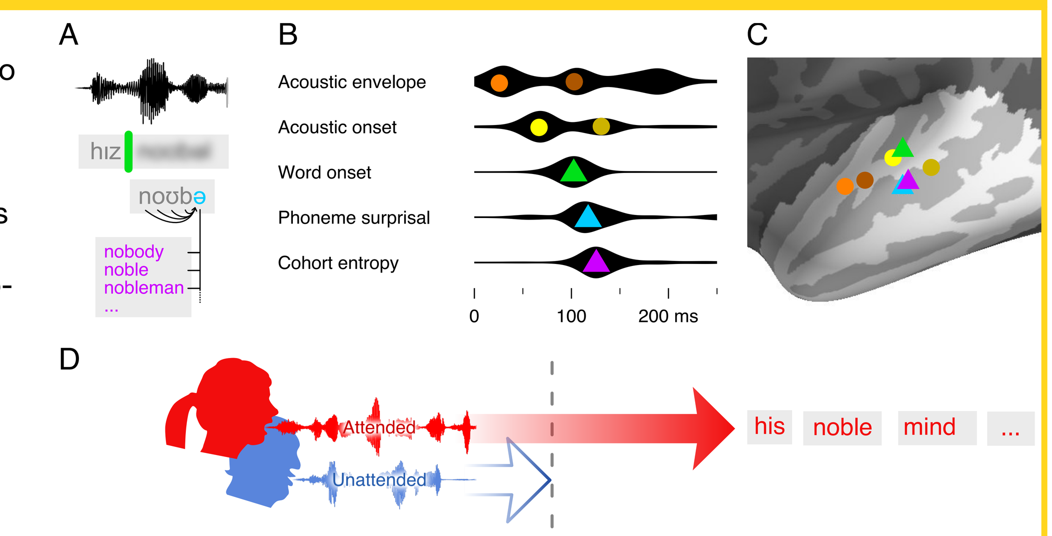


Responses to two concurrent speakers: lexical processing is attention-dependent



Summary of results

- MEG responses to continuous speech can be decomposed into responses to acoustic features and responses associated with lexical processing (A)
- Lexical processing of phonemes occurs in the superior temporal lobe within ~100-150 ms of phoneme onset (B, C)
- In the two-talker condition, acoustic responses reflect both talkers; lexical processing variables reflect attended speech only (D)



Responses track acoustic features

- Significant responses to acoustic envelope and acoustic onsets
- Response to acoustic onsets localized posterior to response to envelope, suggesting different underlying neural populations (cf. Hamilton et al., 2018)

Responses track lexical segmentation

- Significant response of word onset
- ~103 ms peak latency suggests immediate awareness of lexical boundaries

Responses track lexical activation

- Significant effects of cohort-based phoneme surprisal and lexical entropy
- Surprisal (~114 ms) precedes entropy (~125 ms)
- Effect of phoneme surprisal might reflect predictive coding based on lexical statistics
- Effect of entropy might reflect lexical competition

Responses to the two-speaker condition were assessed by grouping variables reflecting acoustic and lexical processing levels

Responses track acoustic features of both speakers

- Responses to acoustic envelope track mix of the two speakers (i.e., the sound as presented to the listeners)
- ~68 ms response to acoustic onsets predominantly tracks mix
- ~131 ms response to acoustic onsets predominantly tracks attended speaker, indicating selective processing of acoustic features in attended stream (cf. Ding & Simon, 2012)

Responses track lexical features of attended speech only

- Significant effect of lexical processing model for attended speech
- Lexical model of unattended speech does not improve model fit (the attended lexical model is significantly more predictive than the unattended lexical model)
- Lexical responses delayed relative to single speaker
  - Word onset: 118 vs 103 ms
  - Entropy: 140 vs 125 ms
  - Response to surprisal not significant

Discussion

- Brain responses to acoustic and lexical features of continuous speech can be separated
- Evidence for rapid transformation from acoustic to lexical representations
  - Phoneme surprisal ~114 ms
  - Lexical cohort entropy ~125 ms
- Low latencies could be partly due to
  - Coarticulation: the acoustic signal often contains information about upcoming phonemes before proper phoneme onset (not modeled here)
  - Prediction: context in natural speech allows predicting and anticipating upcoming words
- Two speaker mix condition
  - Only attended speech is lexically processed
  - Lexical processing of attended speech is slowed down by ~15 ms
- Future potential
  - Modelling different levels of speech processing simultaneously
  - The present study used audiobook segments, but more controlled stimuli could be generated to study specific questions
  - Method might improve modeling of electrophysiological responses to more controlled stimuli as well (e.g. account for acoustic differences)

References

Brodbeck, C., Presacco, A., & Simon, J. Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. NeuroImage, 172, 162–174. https://doi.org/10.1016/j.neuroimage.2018.01.042

Broderick, M. P., Anderson, A. J., Liberto, G. M. D., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Current Biology, 28(5), 803-809.e3. https://doi.org/10.1016/j.cub.2018.01.080

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behav Res Methods, 41, 977–990. https://doi.org/10.3758/BRM.41.4.977

David, S. V., Mesgarani, N., & Shamma, S. A. (2007). Estimating sparse spectro-temporal receptive fields with natural stimuli. Network: Computation in Neural Systems, 18(3), 191–212. https://doi.org/10.1080/09548980701609235

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Current Biology, 25(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A, 109, 11854–11859. https://doi.org/10.1073/pnas.1205381109

Hamilton, L. S., Edwards, E., & Chang, E. F. (2018). A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. Current Biology, 28(12), 1860-1871.e4. https://doi.org/10.1016/j.cub.2018.04.033

Yang, X., Wang, K., & Shamma, S. A. (1992). Auditory representations of acoustic signals. IEEE Transactions on Information Theory, 38(2), 824–839. https://doi.org/10.1109/18.119739

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage, 44(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Acknowledgements

This work was supported by a National Institutes of Health grant R01-DC-014085 (to JZS) and by a University of Maryland Seed Grant (to LEH and JZS). We would like to thank Krishna Puvvada for his assistance in designing and preparing the stimuli, and Natalia Malley for her help in collecting data and for excellent technical support.

A preprint of this work is available from https://doi.org/10.1101/326785