Neural Tracking Measures of Speech Intelligibility: Manipulating Intelligibility while Keeping Acoustics Unchanged

I.M Dushyanthi Karunathilake[1], Joshua P. Kulasingham[2], Jonathan Z. Simon[1]

[1]Department of Electrical and Computer Engineering, University of Maryland, College Park, [2]Department of Electrical Engineering, Linköping University, SE

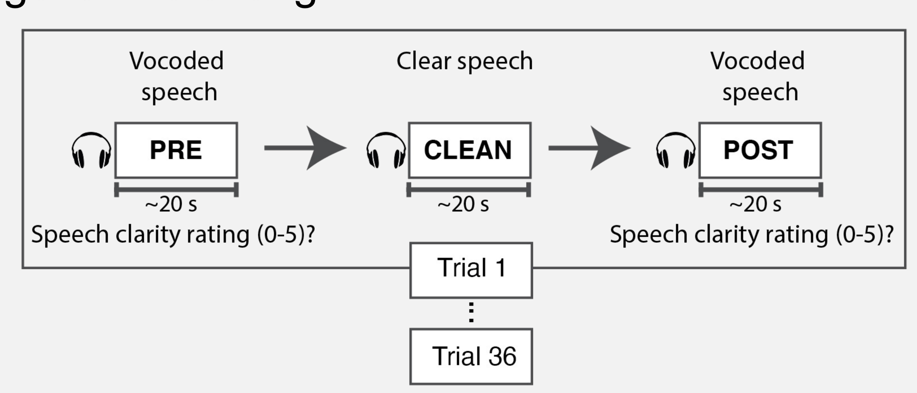
INTRODUCTION

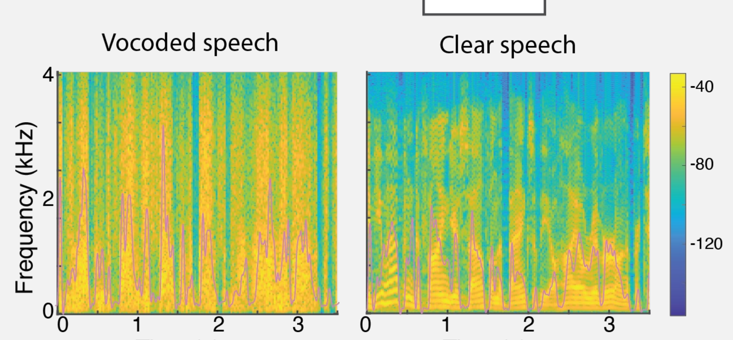
Neural speech tracking has advanced our understanding of how our brains rapidly map an acoustic speech signal onto linguistic representations and ultimately meaning[1]. However, it remains unclear how speech intelligibility is related to the corresponding neural responses. Many studies addressing this question have varied the level of intelligibility by manipulating the acoustic waveform, (i.e., by changing the linguistic content, speech rate, background noise) making it difficult to cleanly distinguish effects of intelligibility from the underlying acoustical confounds.

- Study neural measures of speech intelligibility by manipulating intelligibility while keeping the acoustic structure unchanged.
- Investigate both acoustic and linguistics based neural responses that might be related to speech intelligibility.

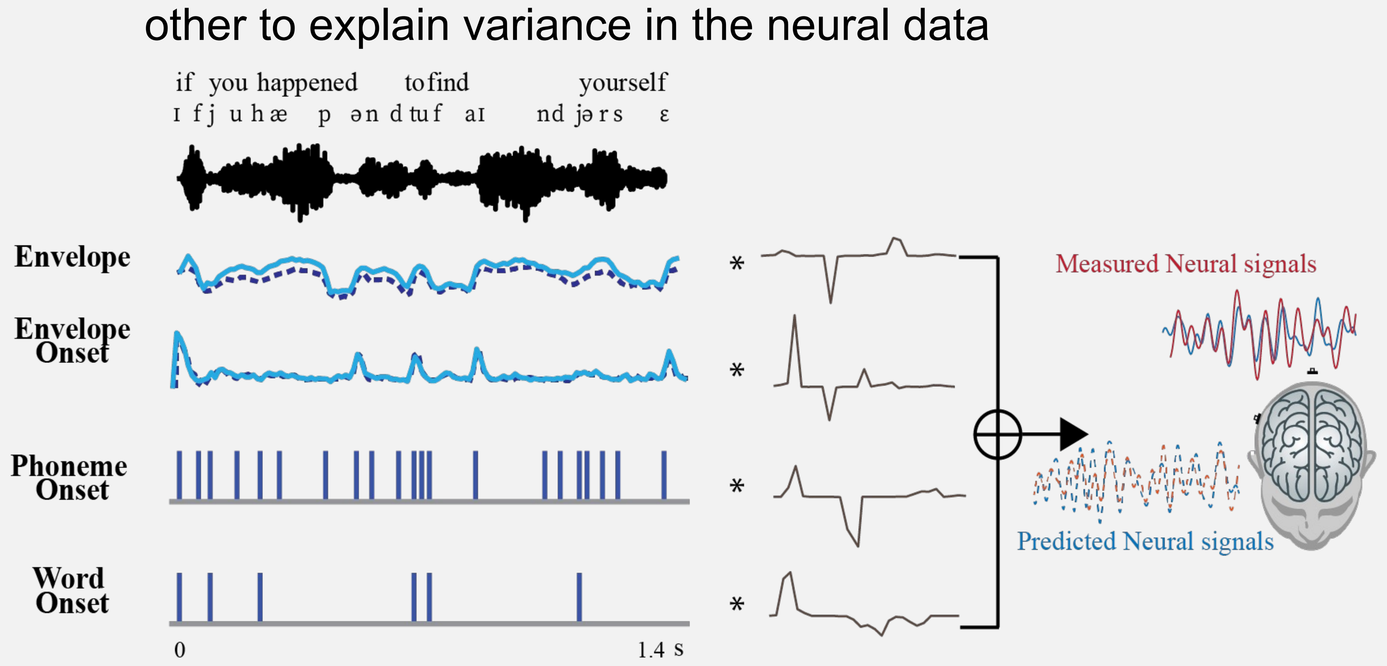
METHODS

- MEG (Magnetoencephalography) data were recorded from 25 native English speaking younger adults (age: 18-26 y)
- Acoustically identical degraded speech (3-band noise vocoded) is presented twice, but the second presentation is preceded by the original (clear speech) recording of the speech (priming).
- Each passage is ~20 s long.





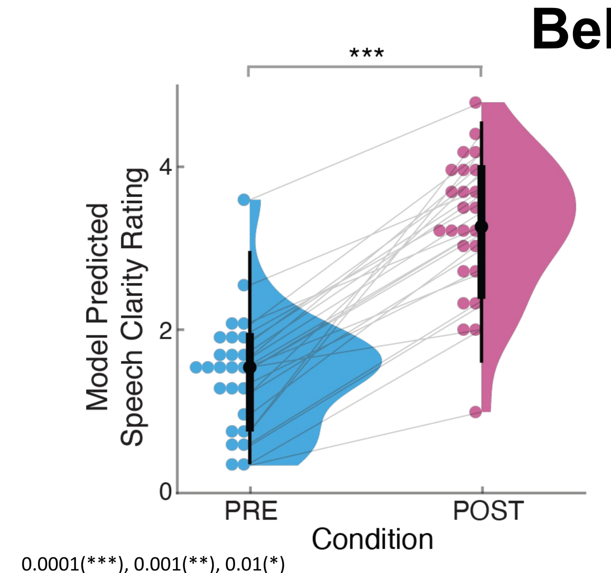
- Data were analyzed using multivariate Temporal Response Functions (mTRFs) using Eelbrain[2]
  - TRFs relate how the brain responds to different speech features[3]. Both acoustic and linguistic speech features are included. All predictors simultaneously compete against each other to explain variance in the neural data



  - Significance of each predictor is assessed by comparing the predictive power to a reduced model without that predictor
  - Statistical tests in source space were performed using TFCE[4]

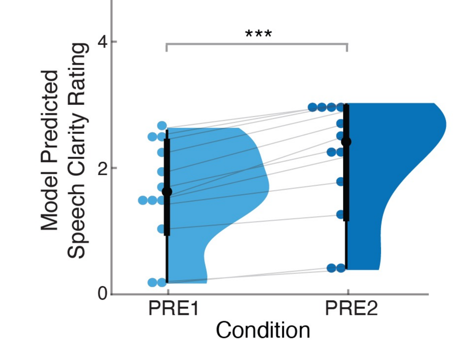
RESULTS

Behavioral Responses



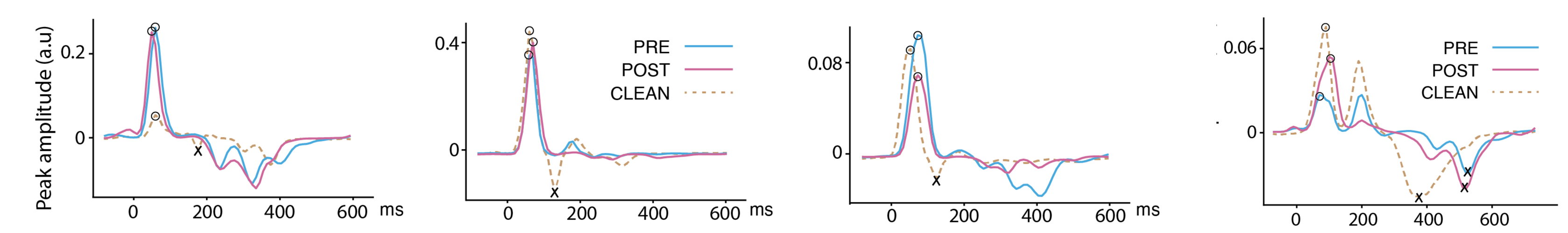
Speech clarity ratings improves for the POST vocoded speech compared to PRE vocoded speech

Improvements in speech clarity with repetition is significantly smaller than the priming effect



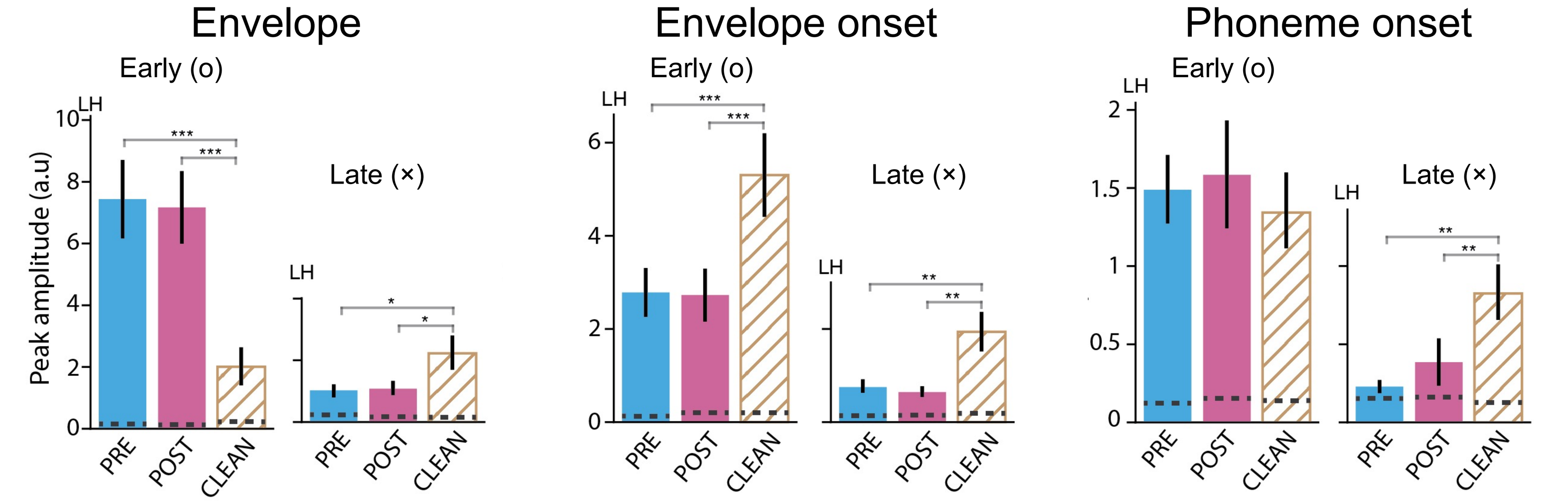
TRFs for a representative subject

Envelope | Envelope onset | Phoneme Onset | Word onset



- TRFs from a representative subject in source space, visualized as a single time series using principal component analysis (PCA)
- TRF peaks (Early (o) and Late (×)) were compared between PRE, POST, and CLEAN speech.

Neural Responses to Acoustic features and Phoneme Onsets

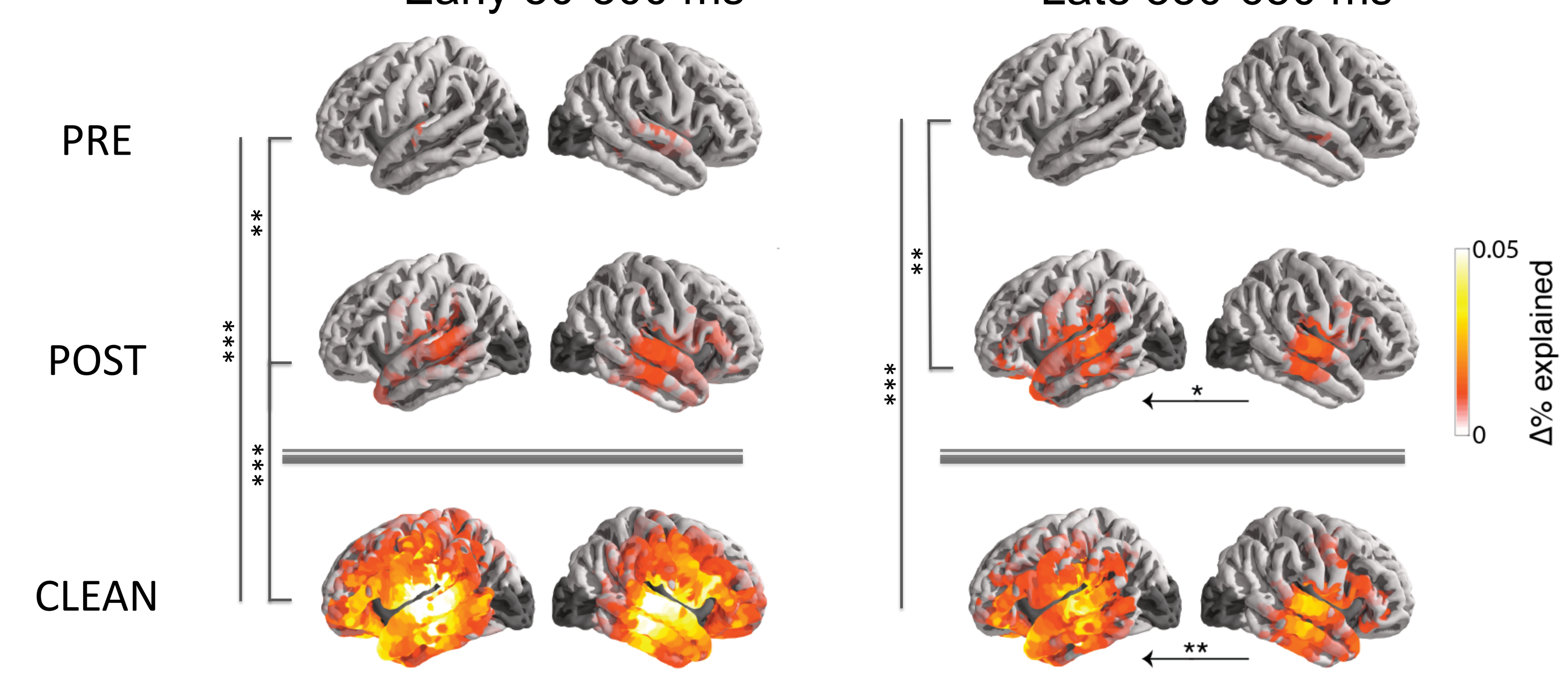


- TRF peak amplitudes were extracted as the maximum peak of the sum of absolute current dipole strengths across sources with a specific polarity, where the polarity was determined from the current directions from the source TRFs.
- The dashed lines within the bars represent the noise floor

- No significant differences between PRE and POST vocoded speech responses
- Significant difference between vocoded and clean speech responses
  - Envelope and envelope onset responses are influenced by the acoustics of stimuli
- Opposite patterns between vocoded vs clean passages for envelope and envelope onset early responses
  - Envelope onset – Loss of salient acoustic onsets
  - Envelope – Low spectral variability

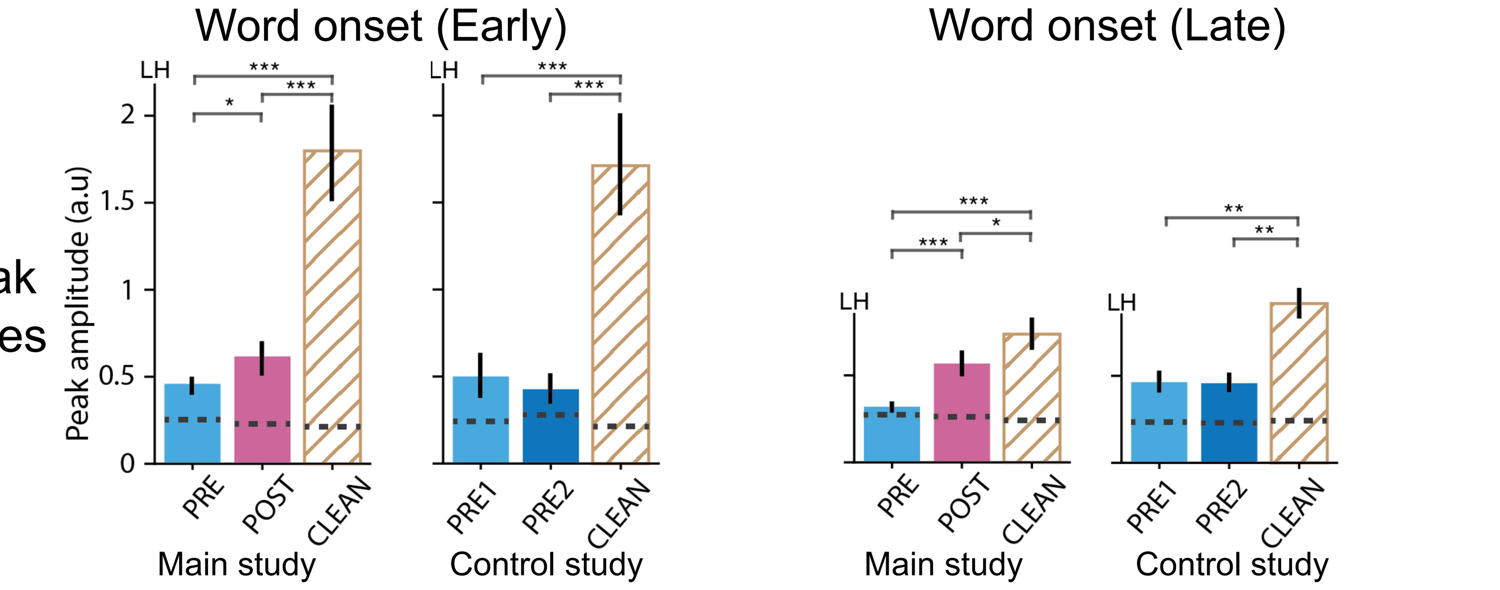
Neural Responses to Word Onsets

Significant contribution of word onset predictor to the model fit, at early and late processing stages relative to the word onset



- Significant difference between PRE and POST vocoded speech word onset processing, for early (superior temporal gyrus (STG)) and more specially at later stage processing (STG and pre frontal cortex (PFC)). Late processing stage also shows left hemispheric lateralization only for POST and CLEAN speech.

Word onset TRF peak amplitudes comparisons for early and late processing stage



- Results are further validated by the TRF peak amplitudes comparisons
- In order to determine that observed differences between PRE vs POST are not due to a side effect of passage repetition, a control study was performed, where passages were repeated back to back (PRE1, PRE2), before clean speech.
  - No differences between PRE1 vs PRE2
  - Improvements in intelligibility generate increased word onset responses over and beyond repetition effects
- Similar effects of intelligibility for contextual word surprisal late responses
  - Evidence for comprehension linked processing in addition to and beyond lexical segmentation

References

[1]. Brodbeck, C., Hong, L. E. & Simon, J. Z. Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. *Current Biology* 28, 3976-3983.e5 (2018).
[2]. Brodbeck, C., Eelbrain 0.30. Zenodo, 2019
[3]. David, S. V., Mesgarani, N. & Shamma, S. A. Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network: Computation in Neural Systems* **18**, 191–212 (2007).
[4]. Smith, S. & Nichols, T. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* **44**, 83–98 (2009).
[5]. Kösem A, Dai B, McQueen JM, Hagoort P. Neural tracking of speech does not unequivocally reflect intelligibility. doi:10.1101/2022.07.25.501422.
[6]. Hakonen, M. *et al.* Predictive processing increases intelligibility of acoustically distorted speech: Behavioral and neural correlates. *Brain Behav* 7, e00789 (2017).
[7]. Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K. & McGettigan, C. Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General* 134, 222–241 (2005).

CONCLUSION

- The experimental paradigm allows to change the level of intelligibility while keeping the acoustics unchanged
- Late neural responses of word segmentation better reflects the speech intelligibility
- Acoustic feature responses are mostly modulated by the acoustics of stimuli and not necessarily on intelligibility
- Lexical representations may provide objective measures of speech comprehension.

This work was supported by the grants R01-DC019394 and SMA 1734892

Contact – dushk@umd.edu

Find preprint at : https://www.biorxiv.org/content/10.1101/2023.05.18.541269