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

I.M Dushyanthi Karunathilake[1], Joshua P. Kulasingham[2], Jonathan Z. Simon[1,3]

[1]Department of Electrical and Computer Engineering, [3]Department of Biology, University of Maryland, College Park, [2]Department of Electrical Engineering, Linköping University, SE

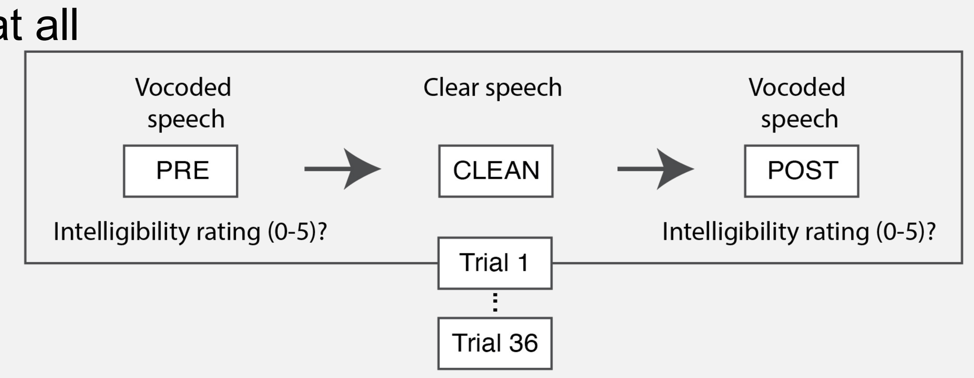
INTRODUCTION

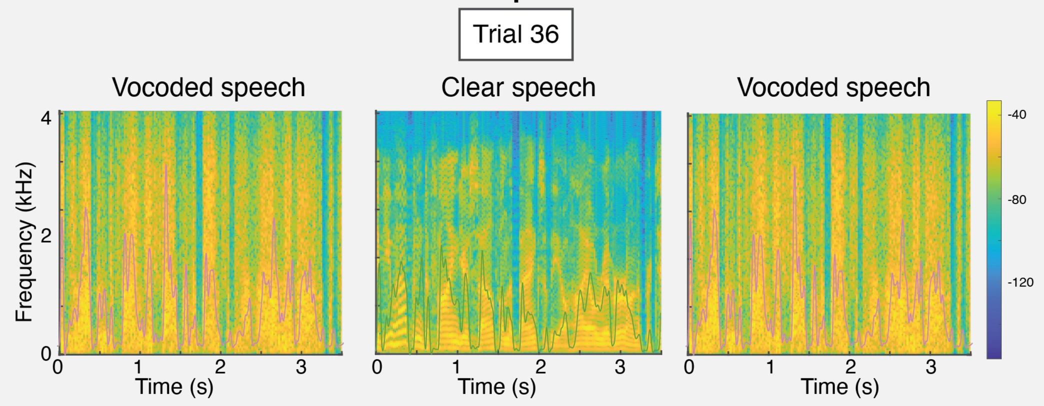
Neural speech tracking has advanced our understanding of how our brains rapidly map an acoustic speech signal onto linguistic representations and ultimately meaning [1]. However, it remains unclear how *speech intelligibility* is related to the corresponding neural responses. Many studies addressing 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. Here we,

- manipulate intelligibility *while keeping the acoustical structure unchanged*.
- investigate both *acoustic and linguistic* based neural responses that might be related to speech intelligibility.

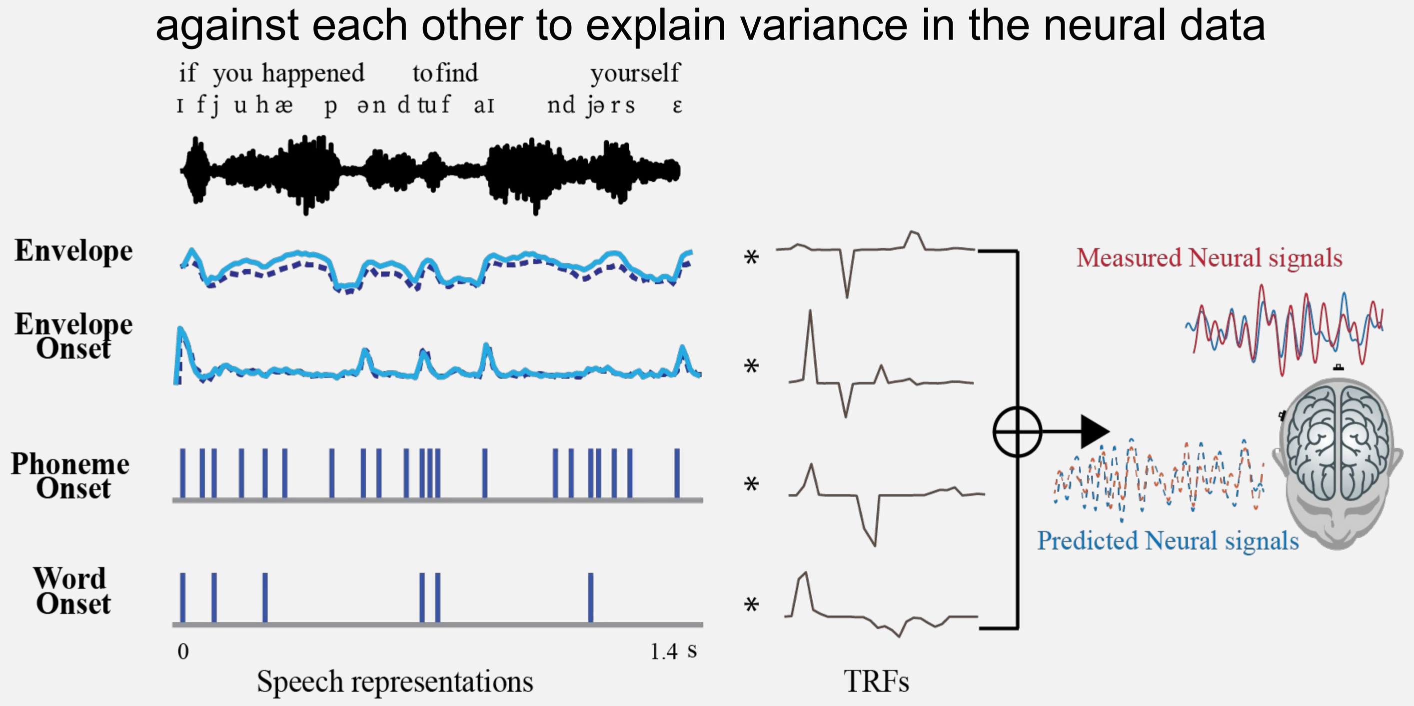
METHODS

- Magnetoencephalography (MEG) data were recorded from 24 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.
- Speech intelligibility is rated on a scale 0 - 5, where 0 means not intelligible at all





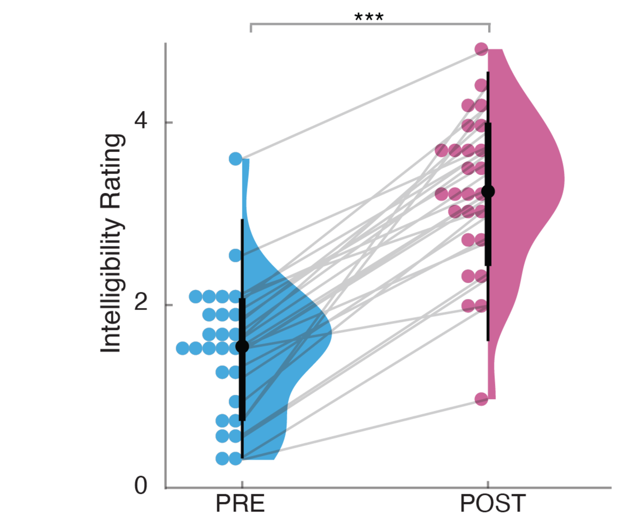
- Data were analyzed using multivariate Temporal Response Functions (mTRFs) using Eelbrain [2]
  - TRFs relate how the brain responds to different speech features ("representations") [3]. Both acoustic and linguistic speech representations are included. All features simultaneously compete against each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

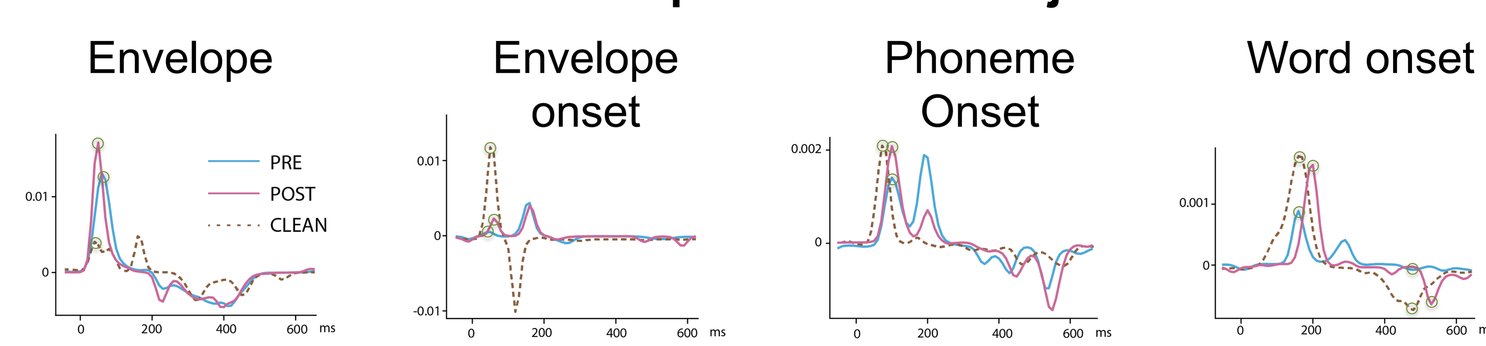
Behavioral Responses



0.0001(***), 0.001(**), 0.01(*), 0.05(.)

- Intelligibility ratings improve for the POST vocoded speech compared to PRE vocoded speech
- Intelligibility ratings also improve *over trials* for both PRE and POST vocoded speech

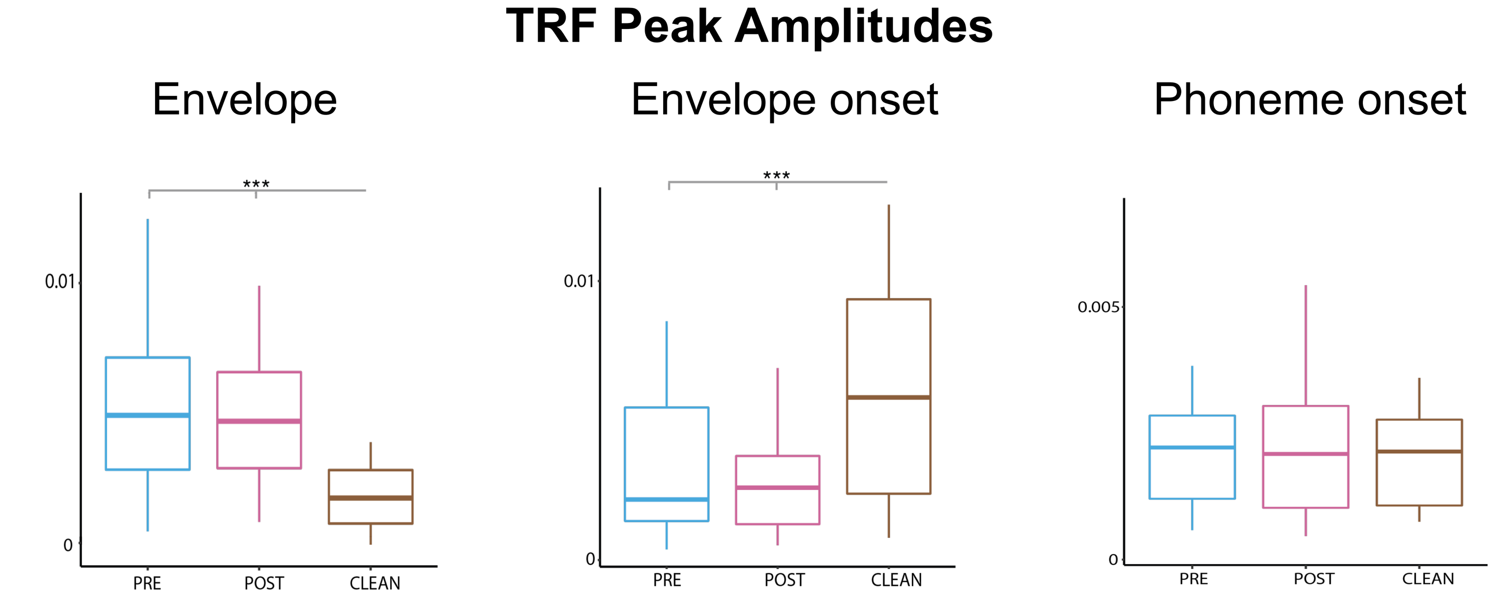
TRFs for a representative subject



TRF peak amplitudes (marked in circles) were extracted and compared for PRE, POST and clean speech

- Envelope, envelope onset and phoneme onset : maximum peak in 30-150 ms range
- Word onset : early peak (80-200 ms) and late peak (350-600 ms)

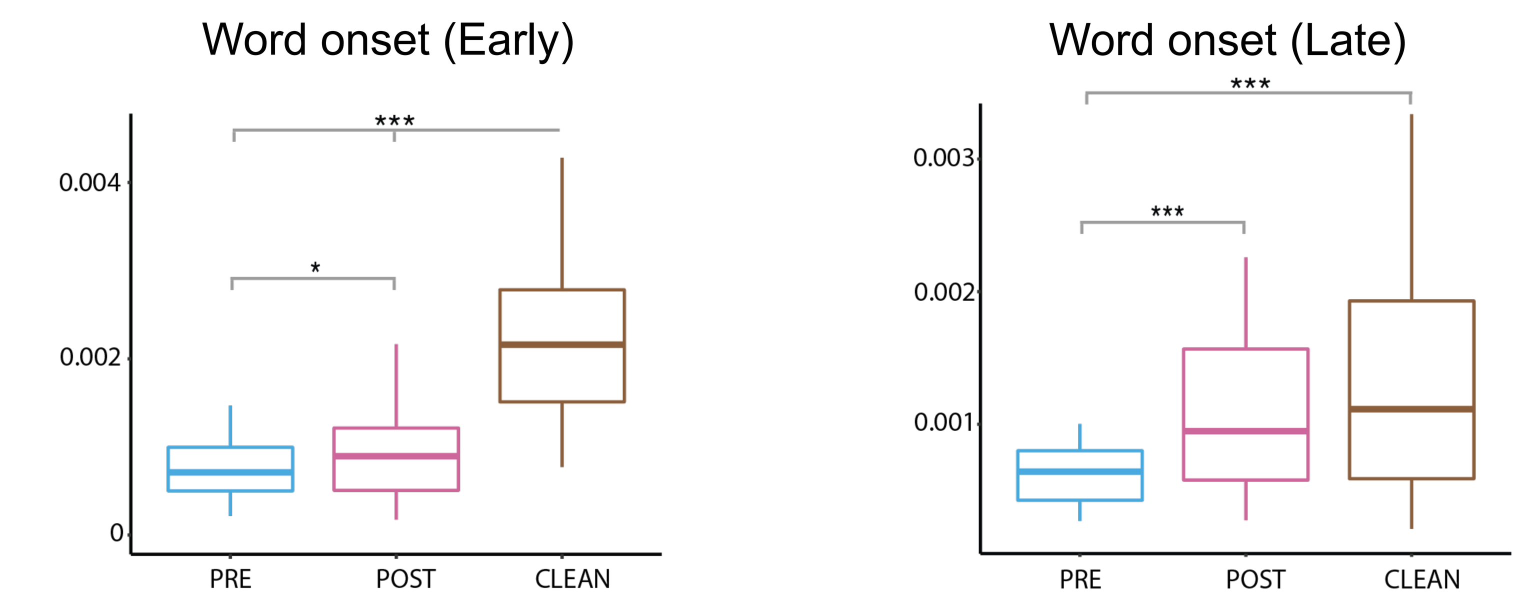
TRF Peak Amplitudes



- No significant differences between PRE and POST vocoded speech responses for these three speech features.
- Significant difference between vocoded and clean speech responses for envelope and envelope onset
  - ➢ Envelope and envelope onset responses are influenced by the acoustics of stimuli

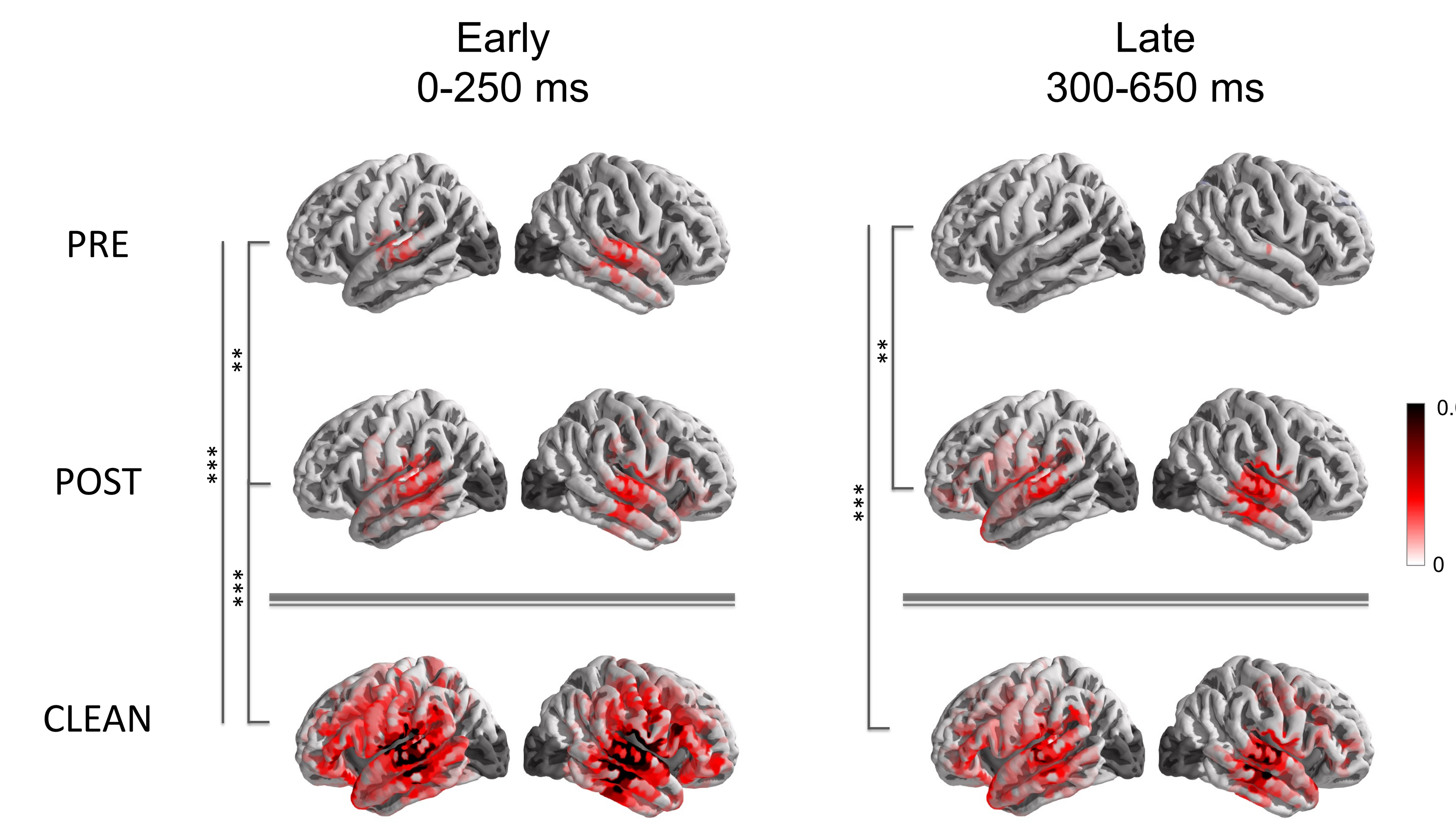
Word Onset Responses

Word onset TRF peak amplitudes for early and late peaks



- Significant difference between PRE and POST vocoded speech word onset processing, for early and later stage, validating the prediction accuracy comparisons
  - ➢ Word onset responses are influenced by the intelligibility of speech

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 even more prominently at later stage processing (STG and pre frontal cortex (PFC))

CONCLUSION

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

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. *et al. Eelbrain: A Python toolkit for time-continuous analysis with temporal response functions*. (2021) doi:10.1101/2021.08.01.454687.
[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., et al. 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).

This work was supported by the National Institutes of Health grant R01-DC019394 and National Science Foundation grant SMA 1734892

Find poster at

Contact – dushk@umd.edu