D71 - Cochlea-scaled Entropy Predicts Robust Neural Coding of Speech Envelopes

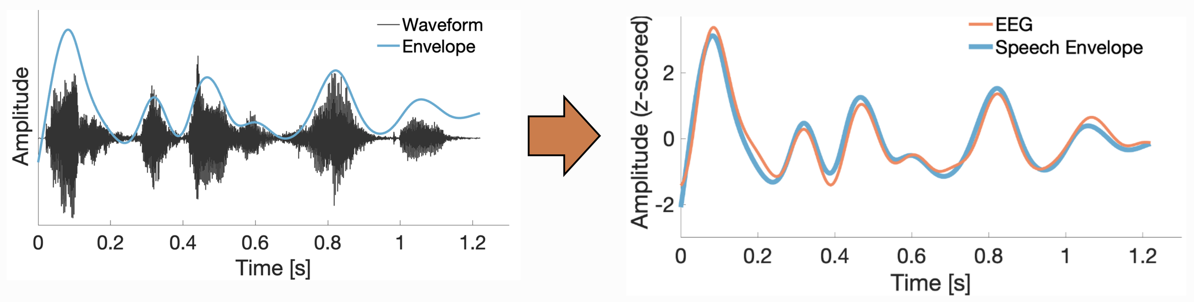
Baorian Nuchged[1], Christian Stilp[2], Keith Kluender[3], Fernando Llanos[1*];

[1]The University of Texas at Austin, [2]Marquette University, [3]Purdue University, *Corresponding author: fllanos@utexas.edu

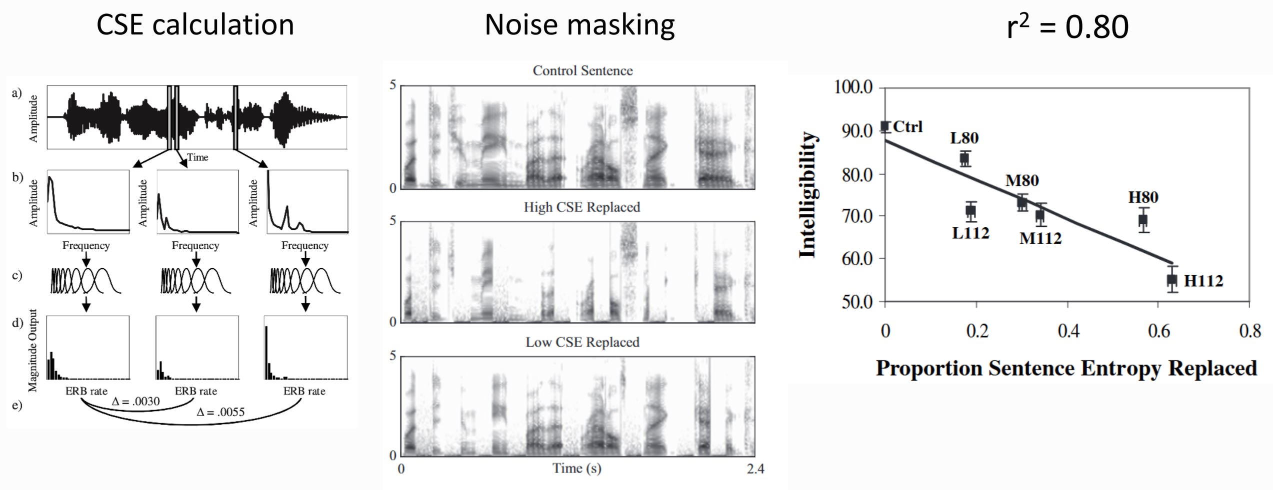
INTRODUCTION

- During processing of connected speech, such as sentences or spoken narratives, low-frequency cortical oscillations synchronize in phase with amplitude fluctuations in the broadband speech envelope.
- The degree of synchronization between brain and envelope oscillations is modulated by higher-order cognitive factors, such as attention and memory, independently of the acoustic signal. This suggests that real-time encoding of speech envelopes reflects more than mere sensory transduction of fluctuations in signal amplitude [1-3].
- Previous research has proposed that neural tracking of the speech envelope reflects parsing of speech primitives, such as phonemes, syllables, words or peaks in the envelope derivative [2,4].
- Here, we propose that neural coding of temporal speech patterns is predicted by spectrally-local changes in relative energy – operationalized by cochlear-scaled entropy (CSE) [5].
  - This metric reflects the fact that sensorineural systems respond predominantly to change.
  - High-CSE segments have been shown contribute more to intelligibility than vowels and consonants.

Neural tracking of amplitude modulation



Cochlea-scaled entropy predicts intelligibility

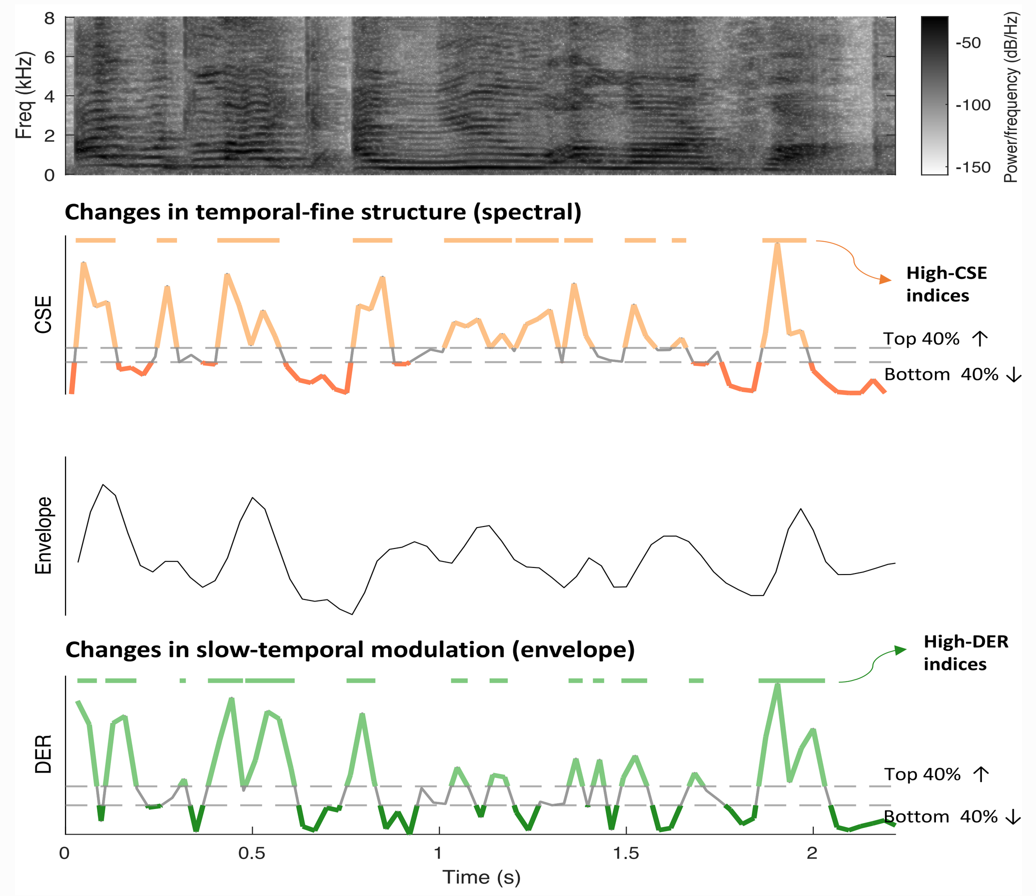


METHODS

STUDY 1: Does CSE predicts robust neural tracking?

First, we tested whether amplitude modulation is more robustly tracked during the presentation of (1) high vs. low CSE frames, or (2) high vs. low peak-derivative (DER) frames

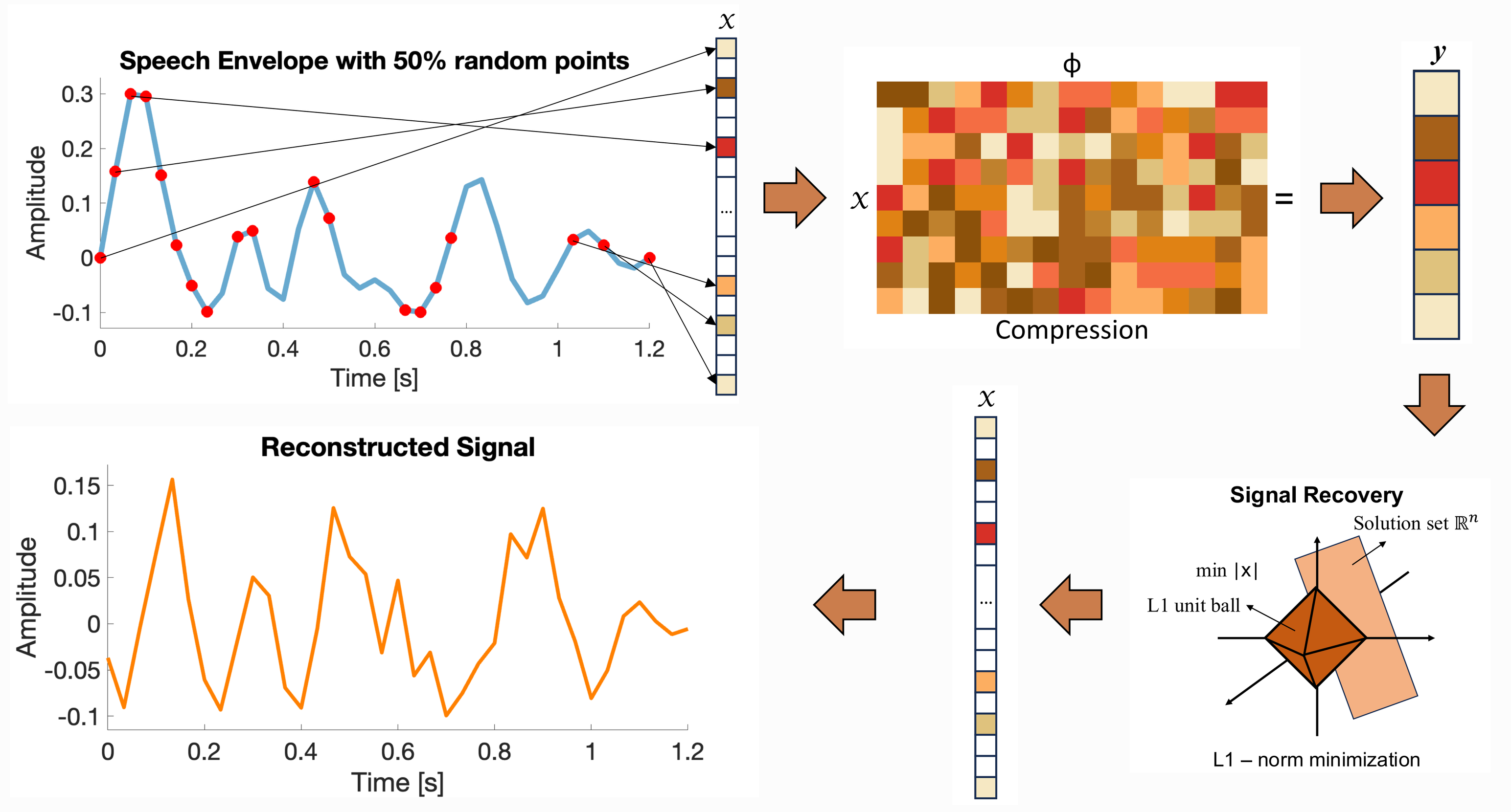
- EEG Dataset: 15 subjects, 30 sentences x 36 repetitions, 32 channels
- Neural tracking was assessed by correlating the speech envelopes of high-CSE, low-CSE, or peak-derivative frames with the low-frequency EEG oscillations evoked during their presentation



STUDY 2: Does high-CSE provide efficient neural coding of envelopes?

Additionally, we tested whether brain oscillations are more robustly predicted from from sparse representations of original broadband envelopes derived from high-CSE or high-DER frames

- Original envelopes were reconstructed via compressed sensing (l1-magic) from 50%, 35%, and 25% of envelope datapoints showing higher-CSE or higher-DER
- Two control reconstructions: (1) random selection of datapoints, (2) lower-CSE percentiles

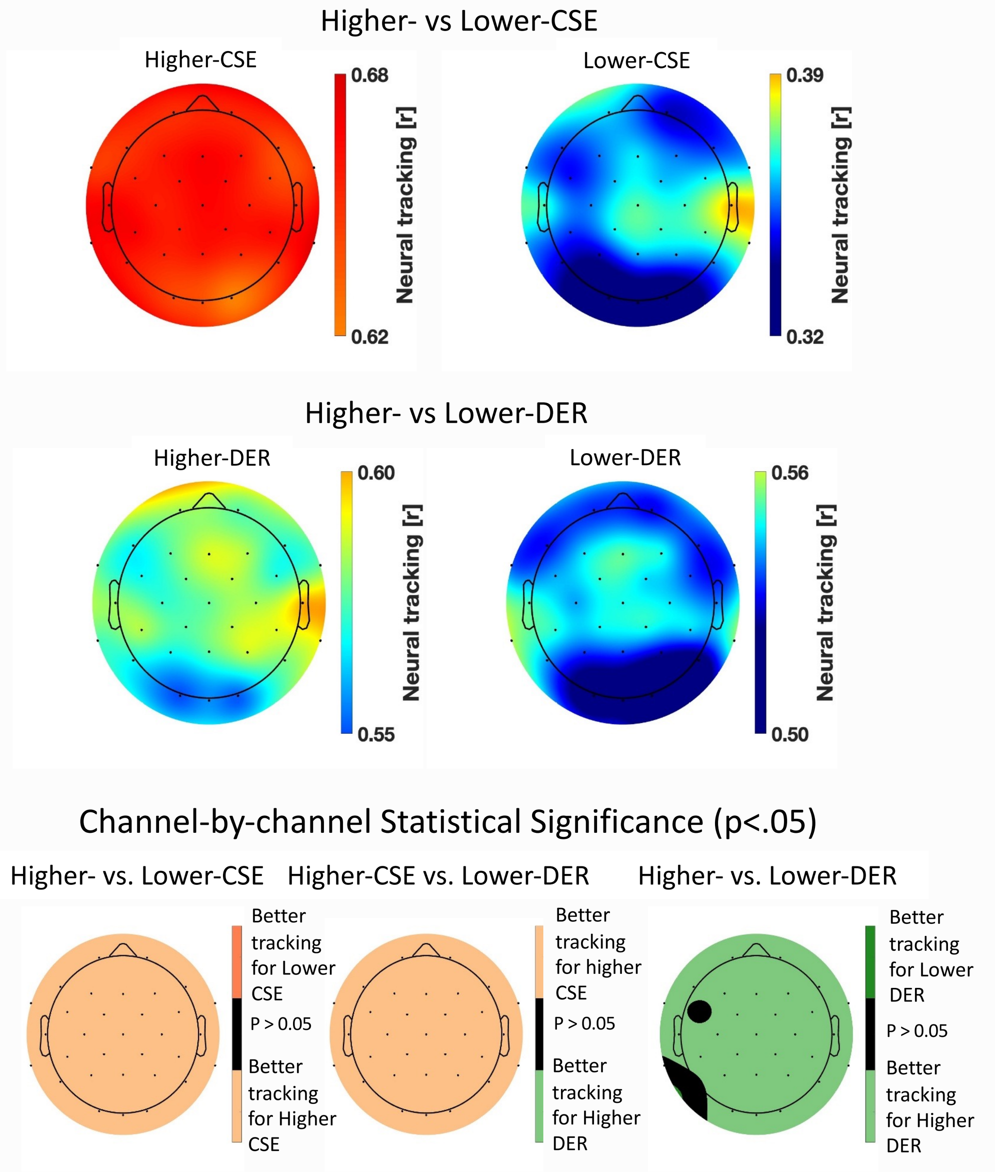


RESULTS

STUDY 1

High-CSE spots the segments with more robust tracking

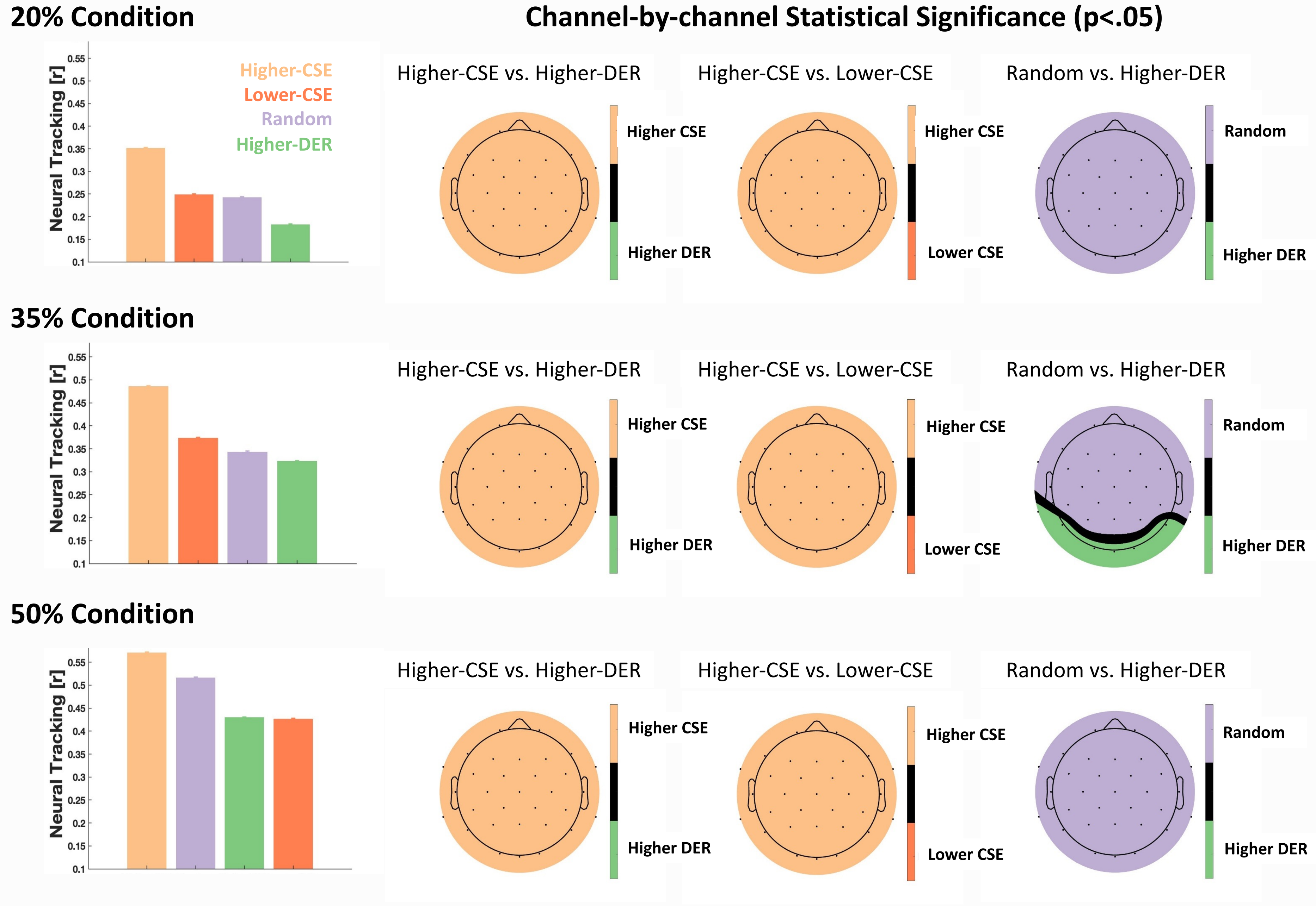
- Higher-CSE (r = 0.66) > Lower-CSE (r = 0.36, significant in all electrodes)
- Higher-DER (r = 0.57) > Lower-DER (r = 0.54, significant in 29/32 electrodes)
- *Higher-CSE (r = 0.66) > Higher-DER* (r = 0.57, significant in all electrodes)



STUDY 2

Neural oscillations are best predicted from sparse envelope representations derived from high-CSE

- Higher CSE (20%: r = 0.35; 35%: r = 0.49; 50%: r = 0.57) > Random (20%: r = 0.24; 35%: r = 0.34; 50%: r = 0.52) > Higher DER (20%: r = 0.18; 35%: r = 0.32; 50%: r = 0.43) ≈ Lower CSE (20%: r = 0.25; 35%: r = 0.37; 50%: r = 0.43)
- Effect consistent at 20%, 35%, and 50% sampling
- Strongest correlations at 50% Higher CSE



DISCUSSION

- We aimed to investigate the relationship between discrete and spectrally-local changes in relative energy, operationalized by CSE, and the neural coding of amplitude modulation patterns
- Study 1: Results suggest that neural tracking is strongest for higher CSE segments. Since these segments are also the ones most critical for intelligibility [5], this finding suggests that neural tracking may reflect the brain's efficient encoding of speech patterns.
- Study 2: Results show that high-CSE provides better neural reconstruction derived from sparse coding principles. This suggests that CSE provides an efficient way to compress speech representations, increasing neural encoding bandwidth
- These results support a model of temporal sparse coding in which the brain selectively enhances the processing of speech segments that convey greater informational value. By prioritizing higher-CSE segments—those most critical for intelligibility—this encoding strategy enables listeners to maximize neural encoding bandwidth while minimizing intelligibility loss

REFERENCES

[1] Ding, N., & Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. Journal of Neuroscience,

[2] 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)

[3] Teng, X., Tian, X., Doelling, K., & Poeppel, D. (2019). Speech fine structure contains critical temporal cues to support speech segmentation. NeuroImage, 201, 116037.

[4] Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. Science advances, 5(11)

[5] Stilp, C. E., & Kluender, K. R. (2010). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. Proceedings of the National Academy of Sciences, 107(27)

Poster presented at the SNL conference, Washington DC, Fall 2025