

Speech features are predicted by the brain even when attention is focused elsewhere



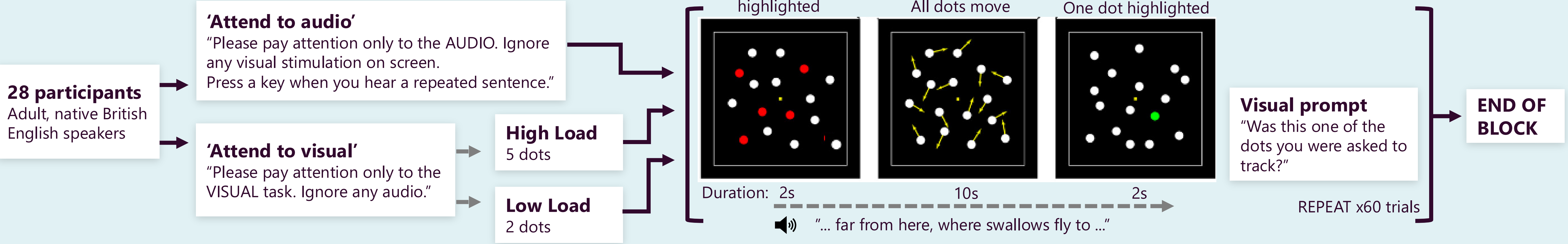
Dominic West¹, Ed Sohoglu¹, Sophie Forster¹, Jeff Mitchell²

1. School of Psychology, University of Sussex, Brighton, UK
2. School of Engineering and Informatics, University of Sussex, Brighton, UK

Background

- **Sensory prediction:** an automatic, pre-attentive process?¹²³
- The majority of recent work investigating speech processing has focused on **within-modality manipulations of attention**⁴⁵⁶. These studies suggest that **predictive speech processing is strongly dependent on attention**.
- However, it may be that the visual and auditory sensory domains share **common attentional mechanisms** and demonstrate similar object-based selection effects⁷.
- Here we explore whether predictive speech processing is similarly dependent on attention when **attention is manipulated cross-modally** (across audition and vision).

Experiment Procedure



Temporal Response Function (TRF) Procedure

1. Extract speech features

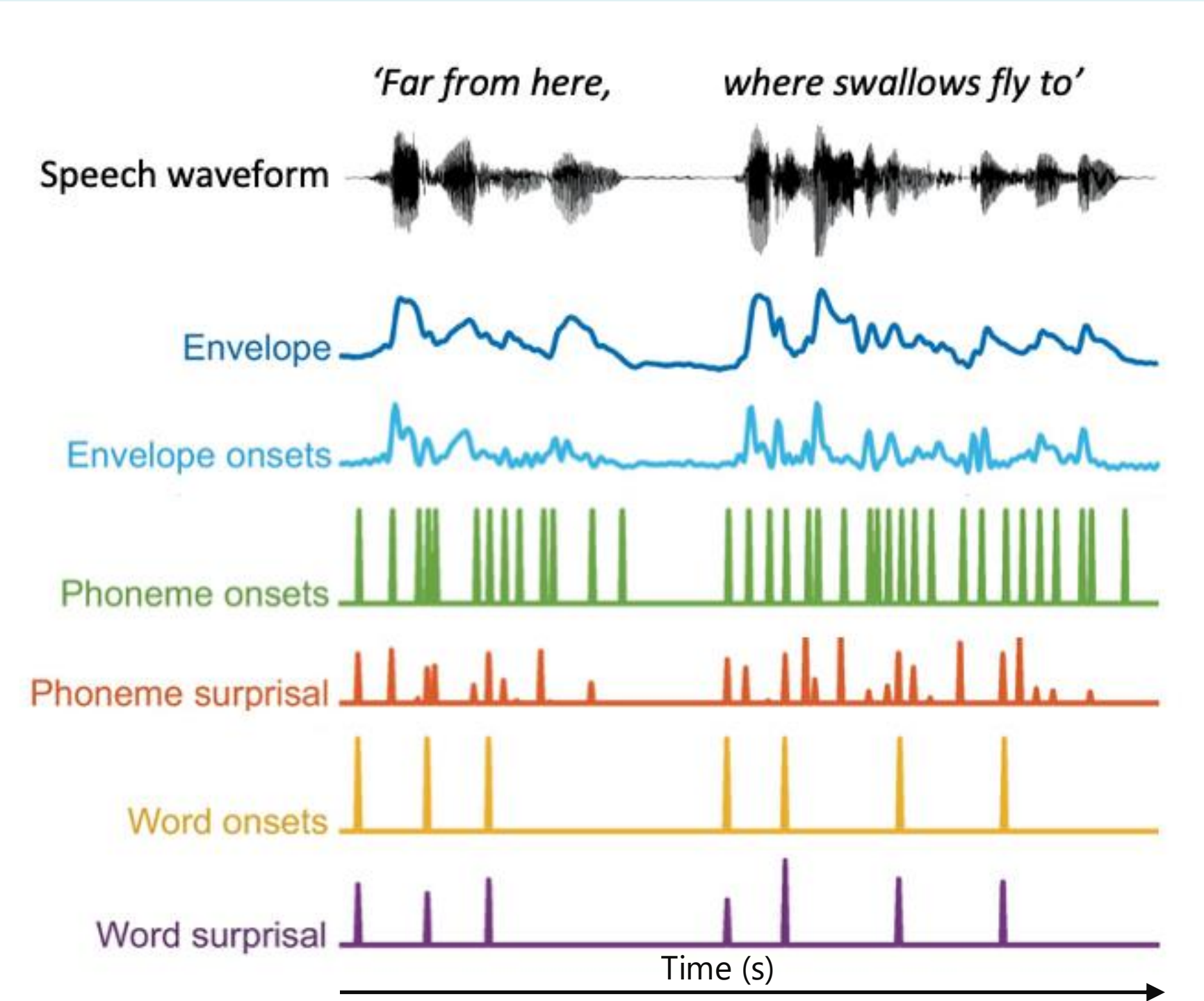
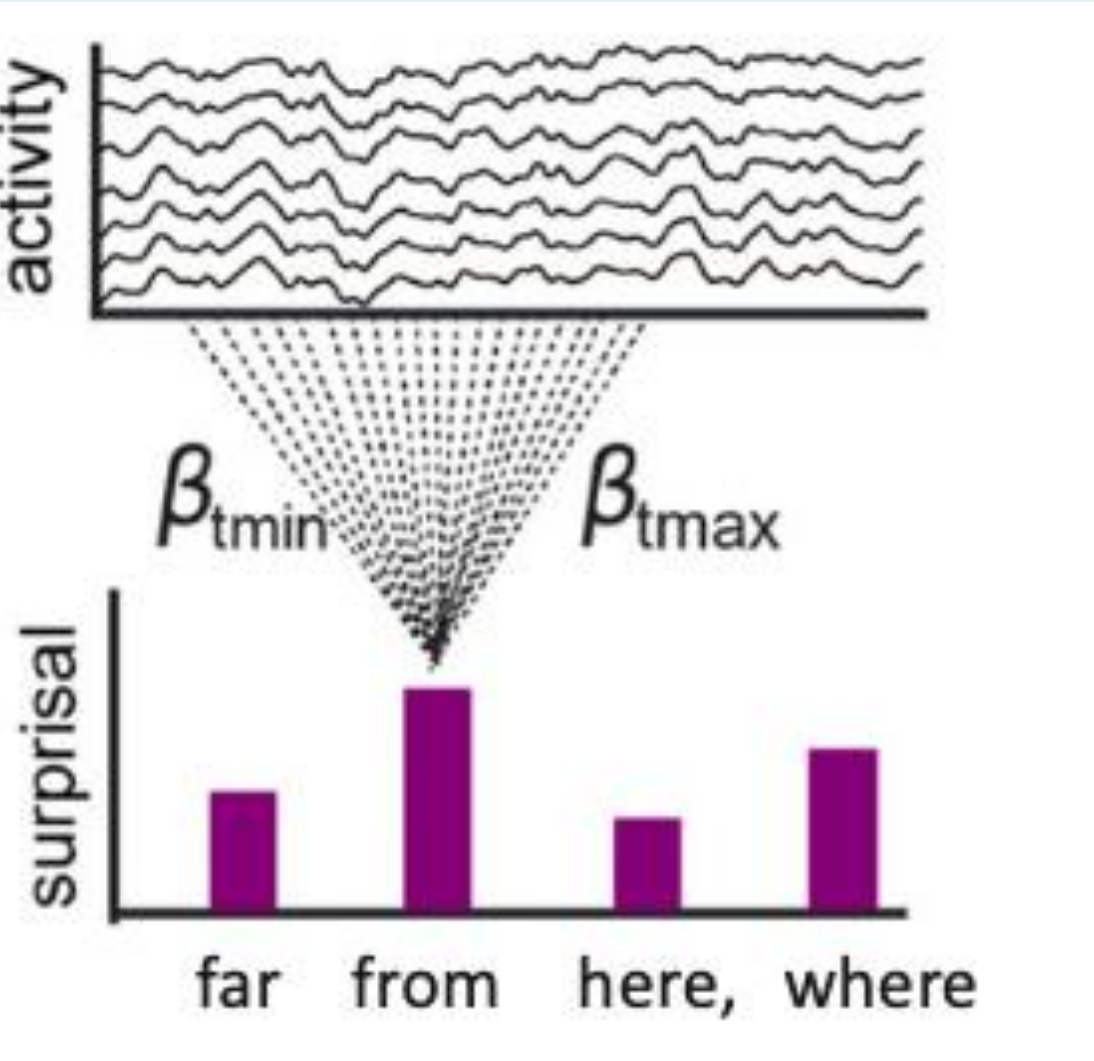


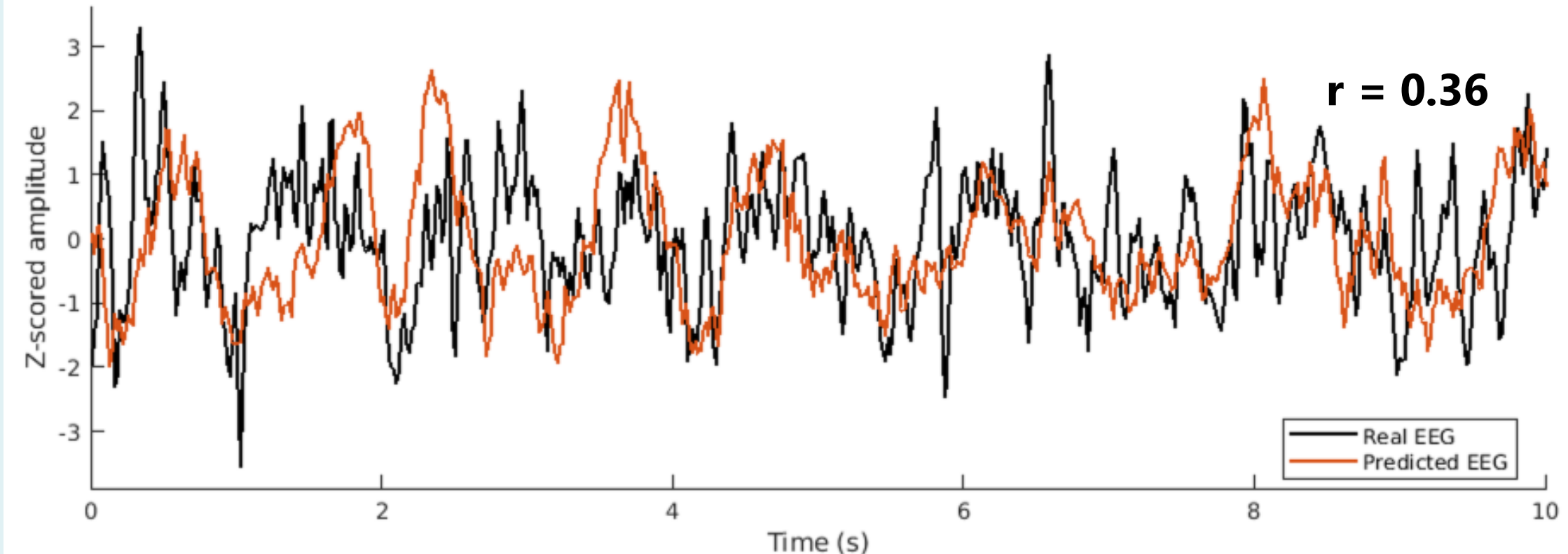
Figure adapted from Kries et al. (2024)⁸.

2. Regress a speech feature against the EEG response at each electrode, for a range of time-lags



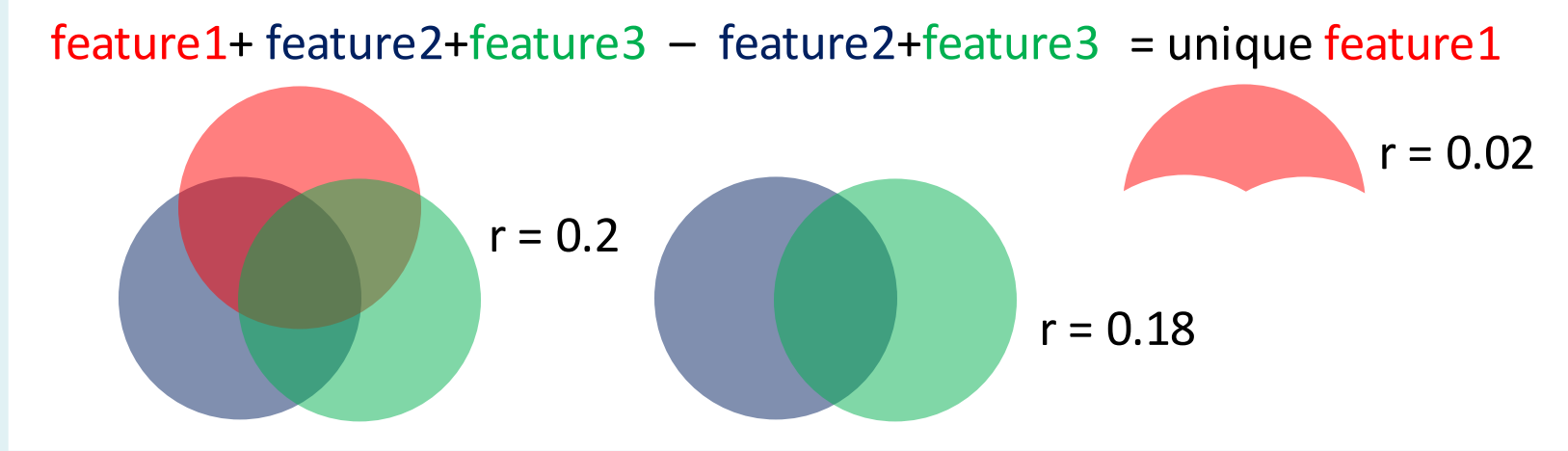
Time-lag range = -100ms to 1000ms.
Figure adapted from Heilbron et al. (2022)⁹.

3. Use leave-one-out cross-validation to compute a single model accuracy value per trial and electrode, integrated across time-lags



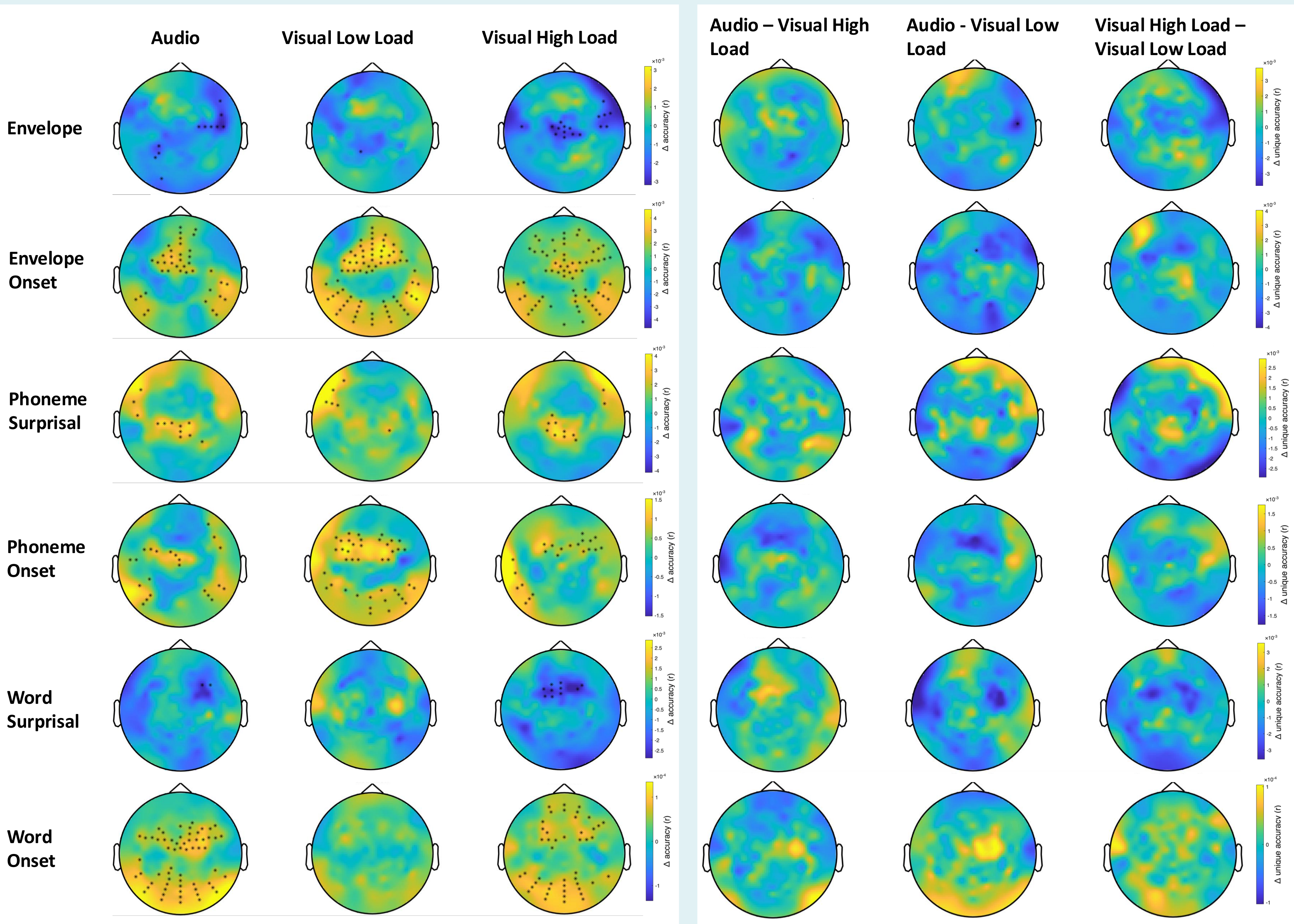
Subj01, trial 1, audio condition, phoneme surprisal, electrode D12

4. Derive the unique contribution of the feature



Attending to a visual task does not reduce the neural tracking of the phoneme predictability

No detectable effect of either attentional focus or attentional load on the neural tracking of speech features...

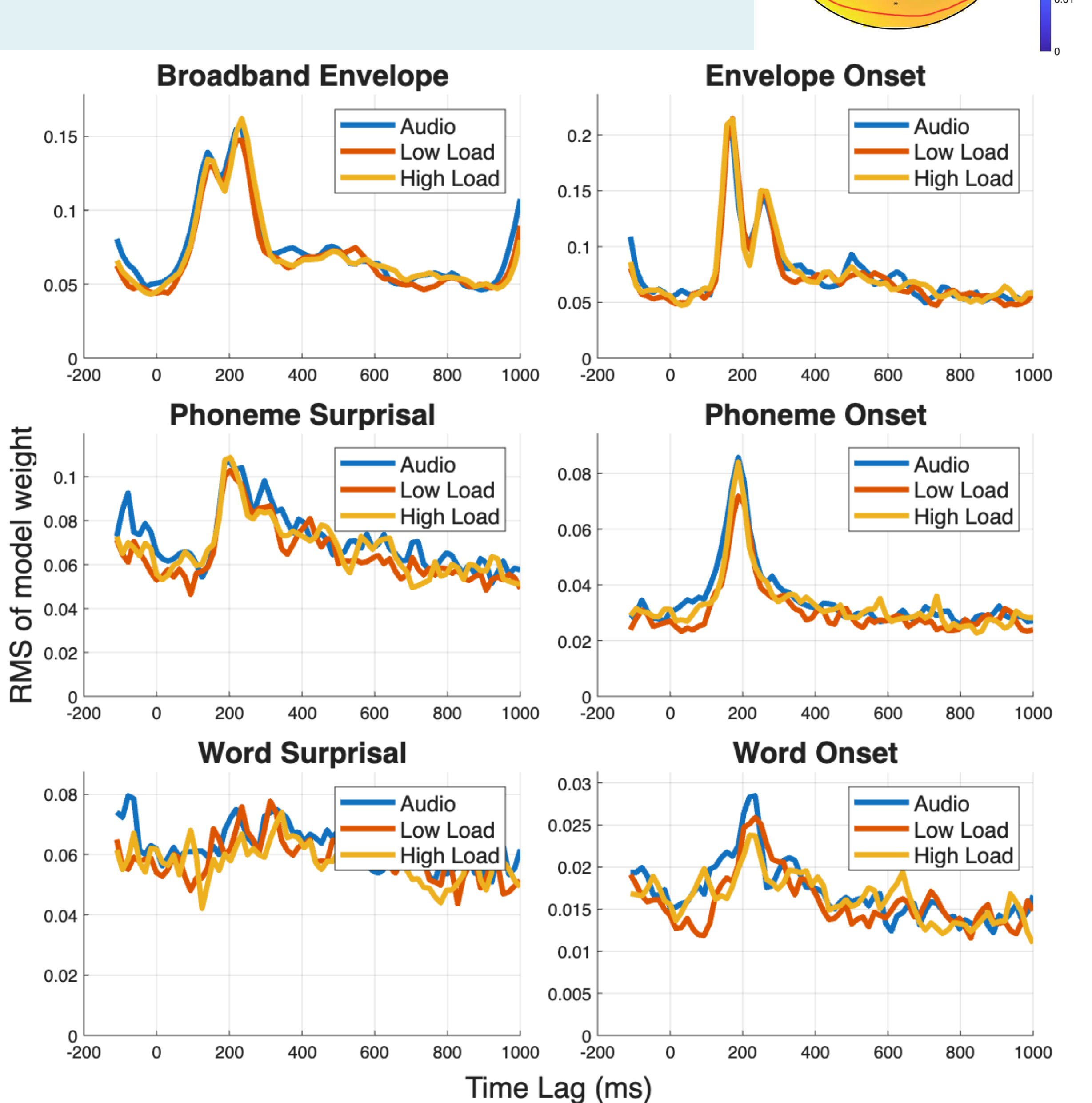


* = $p < 0.05$, corrected for multiple comparisons using Threshold-Free Cluster Enhancement (5000 permutations), $N = 28$ participants.

References

1. Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, 46, 219–227.
2. Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, 113(5).
3. Tivadar, R. I., Knight, R. T., & Tzovara, A. (2021). Automatic Sensory Predictions: A Review of Predictive Mechanisms in the Brain and Their Link to Conscious Processing. *Frontiers in Human Neuroscience*, 15, 702520.
4. Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. *Current Biology*, 28(24), 3976–3983. e5.
5. Brodbeck, C., Jiao, A., Hong, L. E., & Simon, J. Z. (2020). Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. *PLOS Biology*, 18(10), e3000883.
6. Broderick, M. P., Anderson, A. J., Di Liberto, G. M., 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.
7. Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5), 182–186.
8. Kries, J., De Clercq, P., Gillis, M., Vanthornhout, J., Lemmens, R., Francart, T., & Vandermosten, M. (2024). Exploring neural tracking of acoustic and linguistic speech representations in individuals with post-stroke aphasia. *Human Brain Mapping*, 45(8), e26676.
9. Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, 119(32), e2201968119.

...or the temporal processing of speech features



Root-mean-square (RMS) of model weights across speech-sensitive electrodes. Weights extracted from all-features model. Tests for differences between condition-pairs using Threshold-Free Cluster Enhancement (10,000 permutations) failed to reveal any significant time ranges. $N = 28$.

Funding

This project is funded by the Leverhulme Trust, as part of the Sussex 'be.AI' programme.

LEVERHULME TRUST