Context shapes the linguistic structuring and neural alignment of self-supervised speech representations

Marianne de Heer Kloots*[,1], Atlas Kazemian[3], Will Turner[3], Josef Parvizi[2], Laura Gwilliams[3]

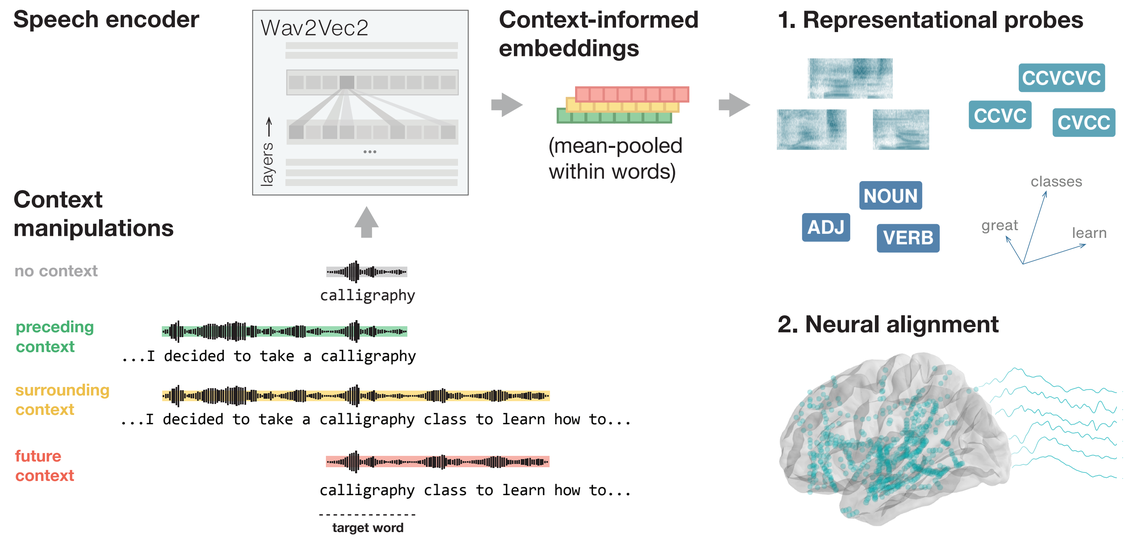
[1]Institute for Logic, Language and Computation; University of Amsterdam, The Netherlands
[2]Department of Neurology and Neurological Sciences, Stanford University
[3]Department of Psychology; Stanford Data Science; Wu Tsai Neurosciences Institute, Stanford University

*m.l.s.deheerkloots@uva.nl

What kinds of linguistic context shape the representation of words in human and artificial speech processing?

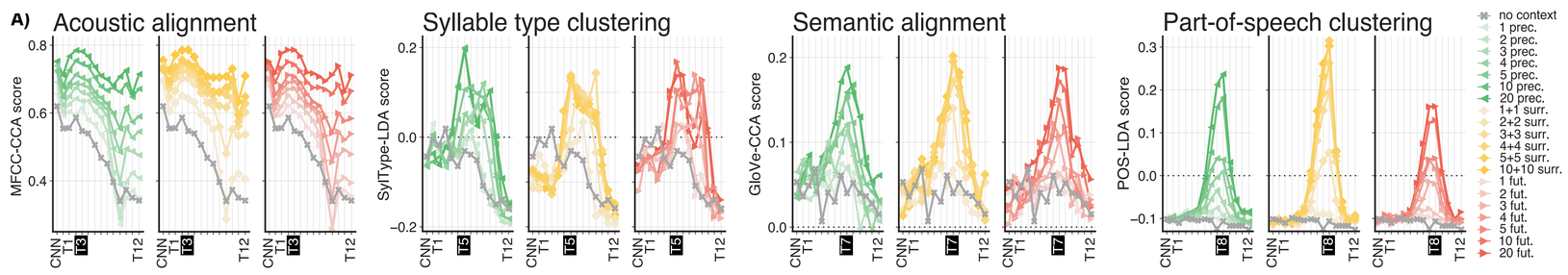
- Speech comprehension involves continuously integrating incoming input with available contextual information. Past *and* future context helps in accurately decoding linguistic units from the speech signal, for both machine learning models[1] and human brains[2].
- Do human and artificial neural networks align in their integration of contextual cues during speech processing?
- We explore how context *size* and *type* affect Wav2Vec2[3]'s representation of speech content from a spoken narrative, as well as its alignment to intracranially-recorded neural activity of humans listening to the same narrative.

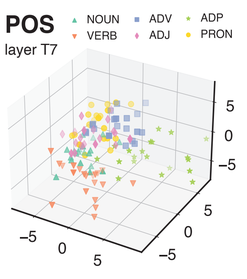




➔ 1. Beyond-word context is crucial for the representation of word-level information in speech model embeddings, and especially for the encoding of more *abstract* linguistic features.

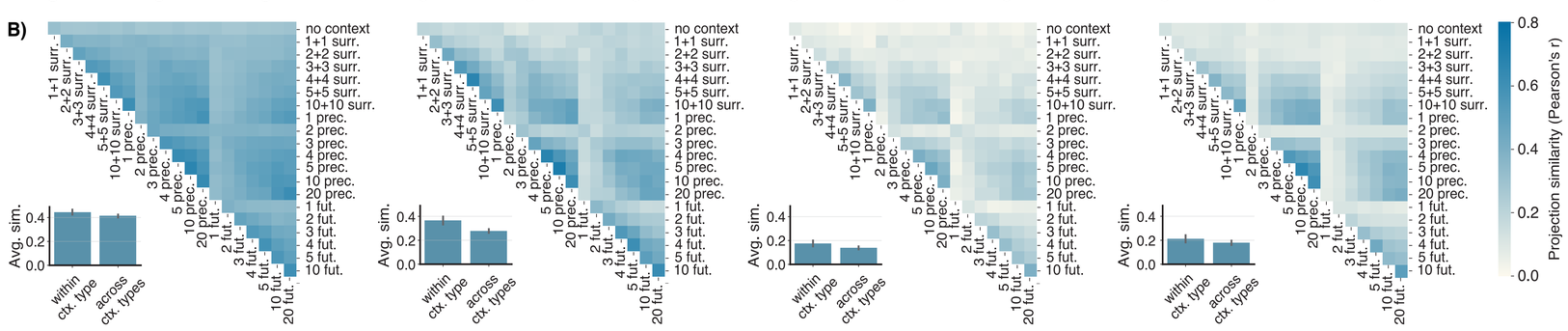
We use *Canonical Correlation Analysis* (CCA) and *Linear Discriminant Analysis* (LDA) to analyze to what extent interpretable word-level feature structures can be extracted from model-internal representations, and to what extent information related to these features is similarly represented across context conditions.



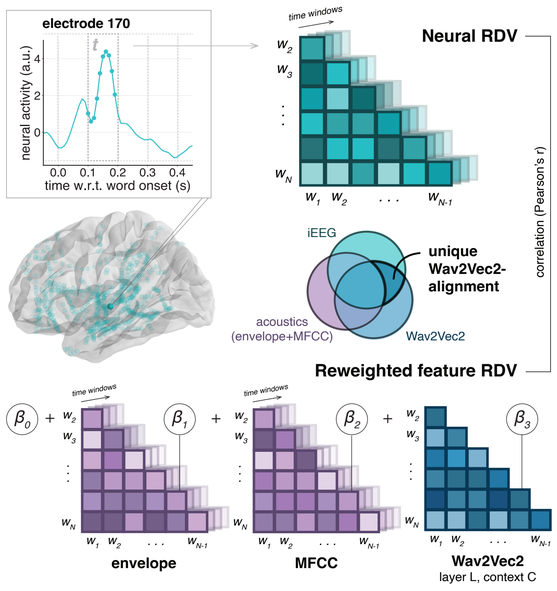


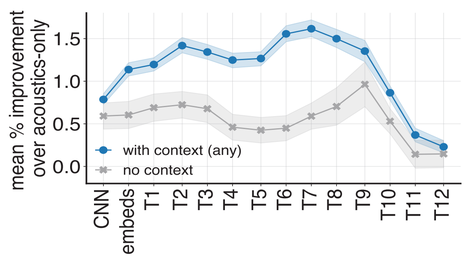
More abstract linguistic features are encoded in *higher model layers*, but only when the model input contains enough context.

Projection similarities are generally higher *within* than *across* context types.



➔ 2. Context-informed but not context-uninformed speech model embeddings explain unique variance in human neural activity, *beyond acoustics*.





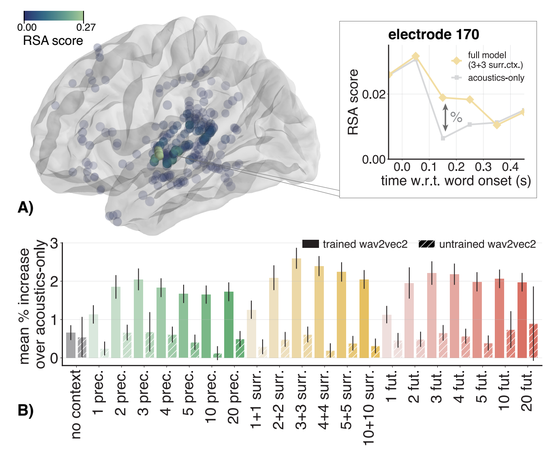
We measure alignment to stereo-EEG (sEEG) data recorded while participants listened to audiobooks. We aggregate data across 10 participants and analyze a subset of 355 electrodes, mostly concentrated in early auditory and language processing regions.

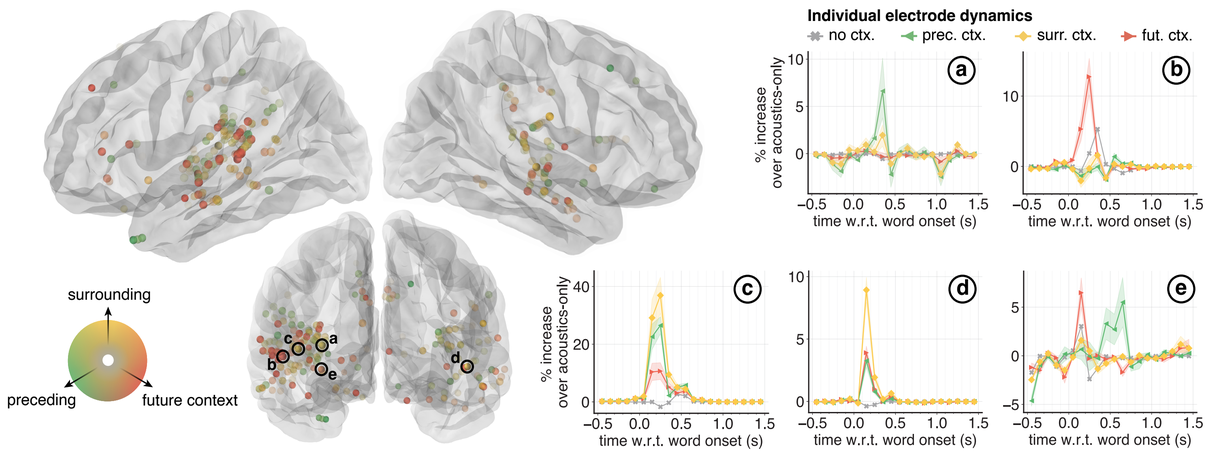
We model distances in *neural representation space* as a weighted combination of distances in *feature representation space*, and quantify **unique Wav2Vec2-alignment** as the relative increase in correlation between these spaces as compared to an acoustics-only model.

When controlling for variance explained by acoustics, the unique contribution of Wav2Vec2 peaks in a *higher model layer* (T7).

Within auditory cortex, unique Wav2Vec2-alignment is substantially higher in the left hemisphere, and for context-informed embeddings over context-uninformed embeddings.

Electrodes best modelled by different context types (future, preceding, surrounding) are *spatially intermixed*.





References

[1] de Heer Kloots, M., & Zuidema, W. (2024). Human-like linguistic biases in neural speech models: Phonetic categorization and phonotactic constraints in Wav2Vec2.0. In *Proc. INTERSPEECH*. https://doi.org/10.21437/Interspeech.2024-2490

[2] Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience, 38*(35), 7585-7599. https://doi.org/10.1523/JNEUROSCI.0065-18.2018

[3] Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems, 33*, 12449-12460. https://proceedings.neurips.cc/paper/2020/hash/92d1e1eb1cd6f9fba3227870bb6d7f07-Abstract.html

Acknowledgements

MdHK is funded by the Netherlands Organization for Scientific Research (NWO), through a Gravitation Grant 024.001.006 to the Language in Interaction Consortium. LG is funded by Esther A. & Joseph Klingenstein Fund; Whitehall Foundation 2024-08-043; BRAIN Foundation A-0741551370.