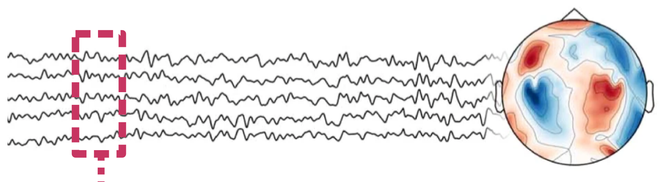
A predictive learning model can simulate temporal dynamics and context effects found in neural representations of continuous speech

Oli Danyi Liu[1], Hao Tang[1], Naomi Feldman[2], Sharon Goldwater[1] [1]University of Edinburgh [2]University of Maryland College Park oli.liu@ed.ac.uk

How is continuous speech represented ...

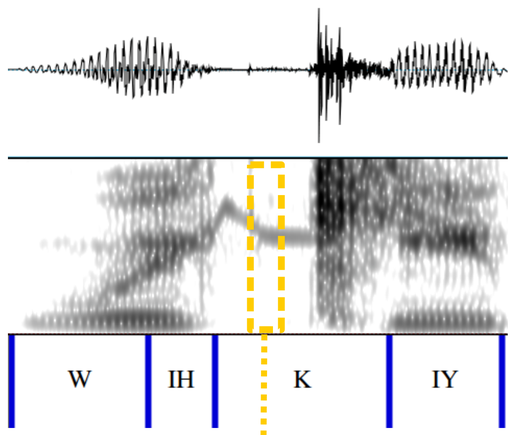
... *in human brains?* (Gwilliams et al., 2022)



MEG signals

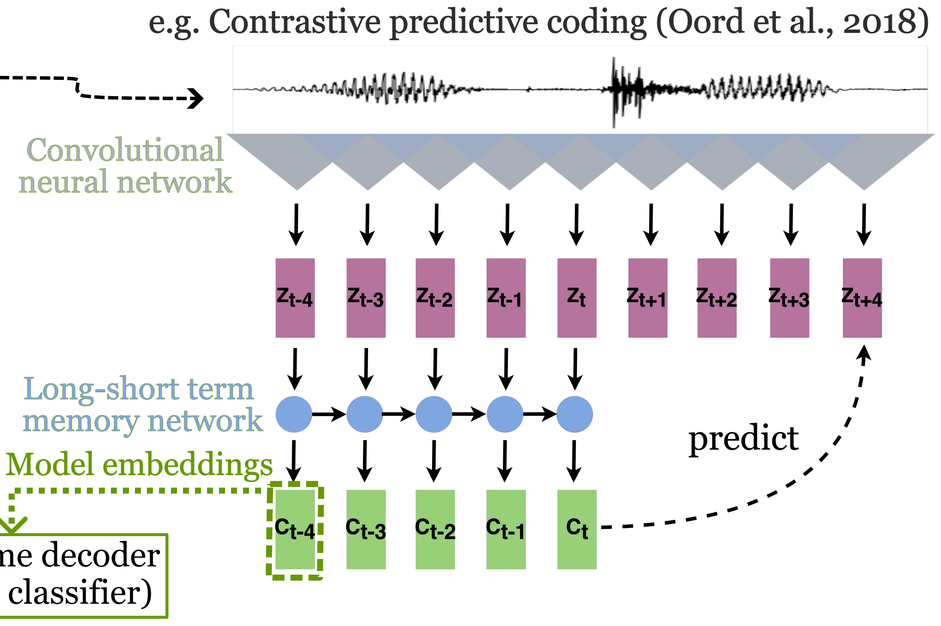
Gwilliams et al. (2022)
Neural dynamics of phoneme sequences reveal position-invariant code for content and order

They identified certain **temporal dynamics** and **context effects** in the neural encoding of human listening to short narratives.



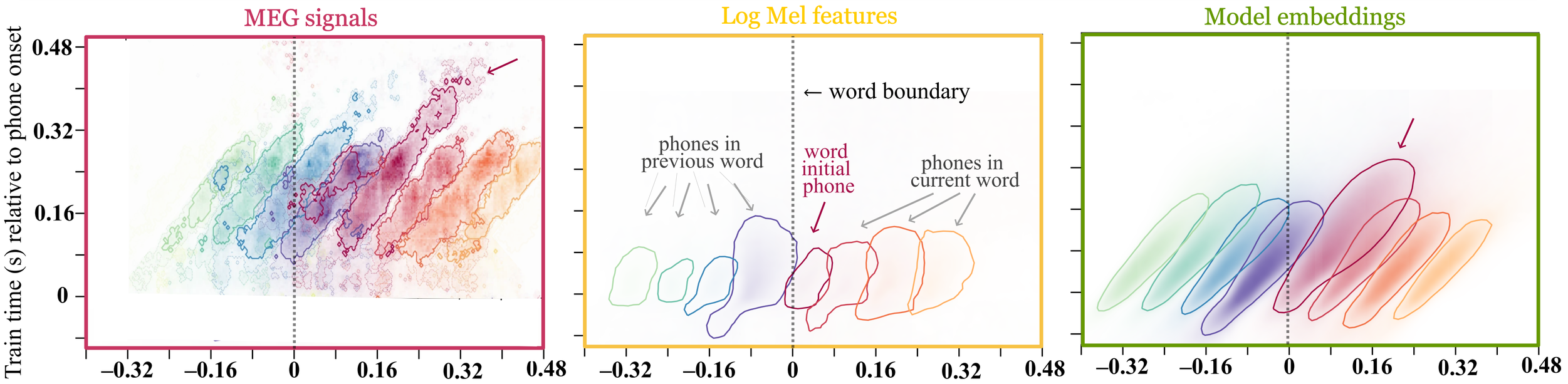
... *in a neural network?* (This work)

e.g. Contrastive predictive coding (Oord et al., 2018)



Temporal dynamics

(1) Multiple successive phones are encoded simultaneously (2) The encoding pattern evolves over time

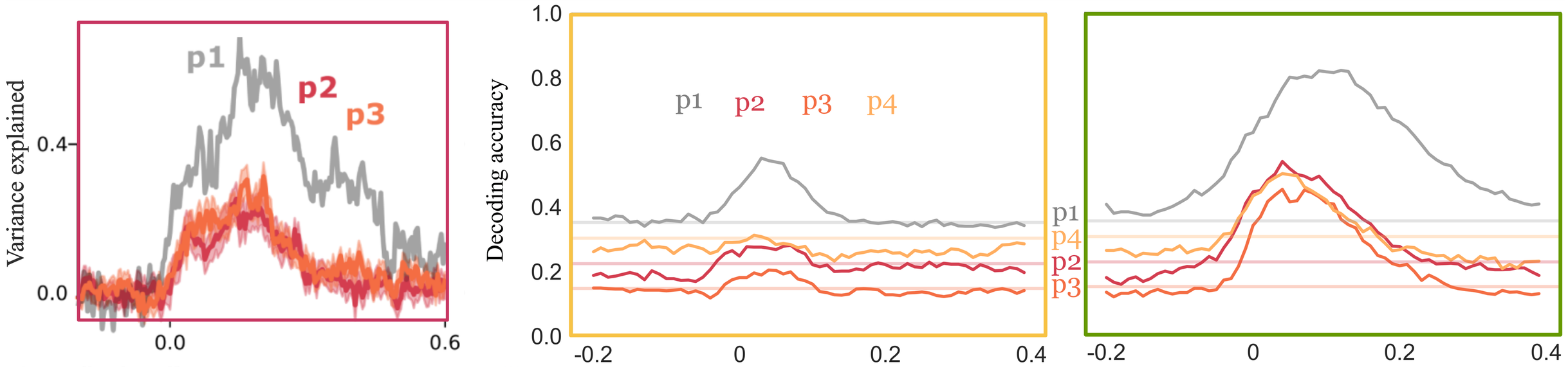


Similar characterstics were found in the model but not in the acoustic features.

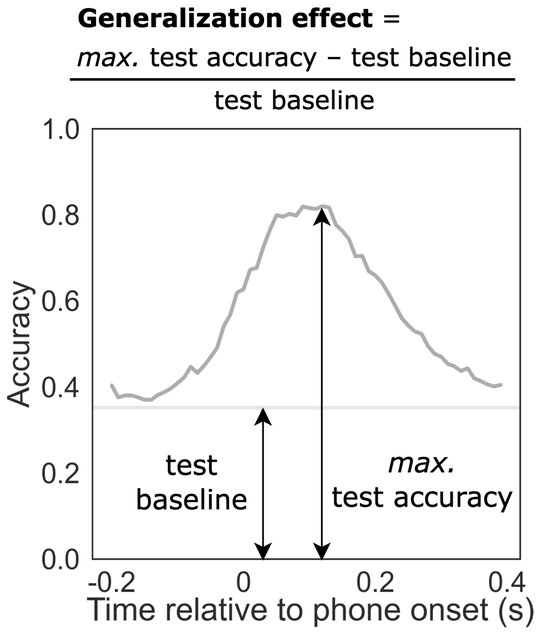
These temporal dynamics can arise through predictive learning without top-down linguistic knowledge.

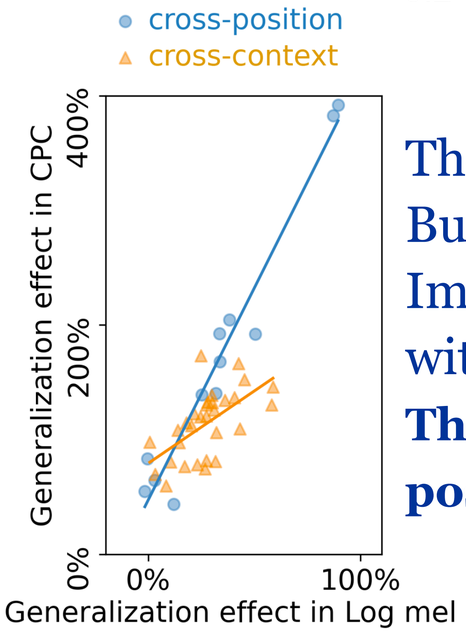
Context effects

The encoding pattern support generalization across phone position to some extent.



Generalization effect =
max. test accuracy − test baseline / test baseline





The model showed cross-position (and cross-context) generalizability.
But the acoustic features also showed a small degree of generalization.
Importantly, the generalization effect in the model correlates strongly with the acoustic similarity between train & test position/context.

These results alone are insufficient for concluding position/context-invariance.