From pixels to semantics: a realistic predictive coding model of evoked responses during reading

Marijn van Vliet[1,2], Laura Rautiainen[1], Gayane Ghazaryan[1], Samer Nour Eddine[3], Aino Saranpää[1], Gina Kuperberg[4,5], Riitta Salmelin[1,2]

[1] Department of Neuroscience and Biomedical Engineering, Aalto University, Finland
[2] Aalto Neuroimaging, Aalto University, Finland
[3] Maryland Language Science Center, University of Maryland, USA
[4] Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, USA
[5] Department of Psychology, Tufts University, USA

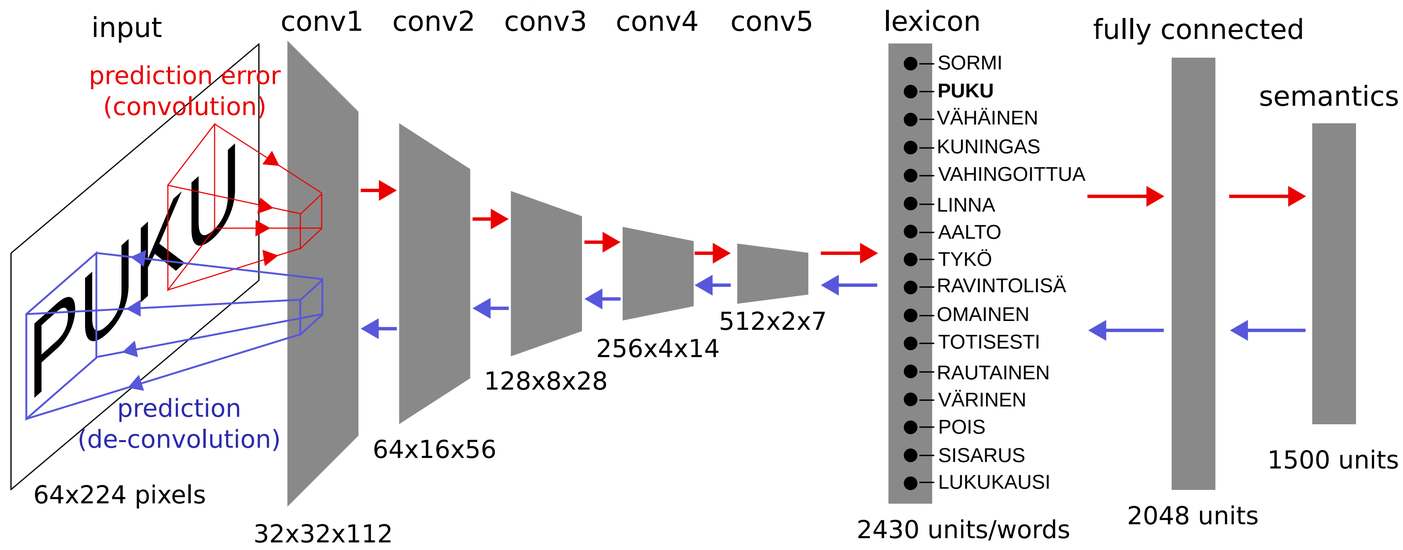
INTRODUCTION

- Reading can be modeled with a Convolutional Neural Network (CNN) [1]
- The N400 can be modeled with predictive coding (PC) [2]
- Can we model the entire evoked response using a CNN with PC?
- Can we achieve enough realism to obtain a signal that resembles MEG data?

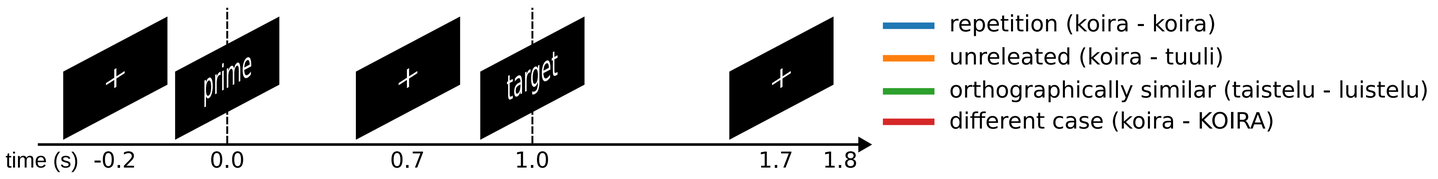
METHODS

- We built a deep CNN with predictive coding mechanics, starting from pixels, passing through a lexicon (2430 Finnish words) and ending in abstract semantics (sparsified word2vec vectors) (**A**)
- Training data: 1M bitmap images of words (varying typeface and size)
- MEG data was collected from 20 participants in a repetition-priming paradigm (**B**)
- The stimuli were presented to the model using the same exact sequence as during the MEG experiment, including stimulus on/off timing and the fixation cross.
- The total prediction error within each model layer simulates the components of the evoked response (**C**)
- PC cycle duration was set to 3.7 ms, retina-V1 delay was set to 90 ms.
- Linear regression was used to map model layers to cortical locations (**D**)
- For each cortical location, the mixture of simulated responses is compared to the actual MEG timecourse (**E**).

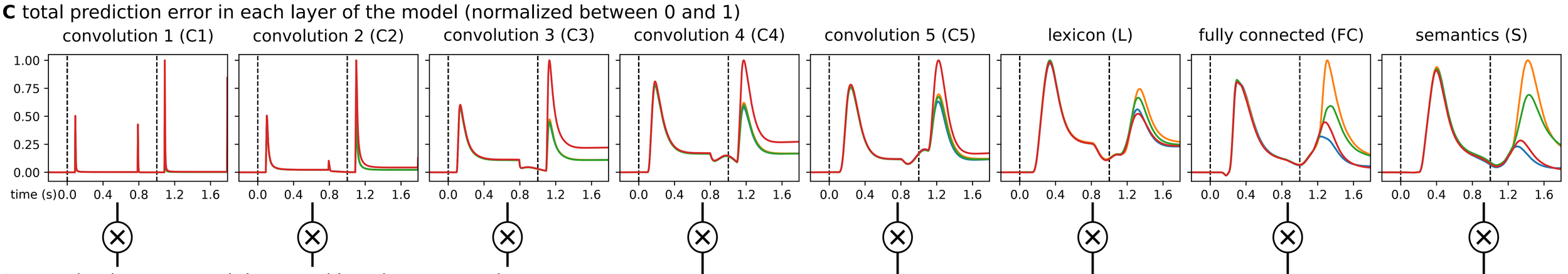
A architecture of the predictive coding model



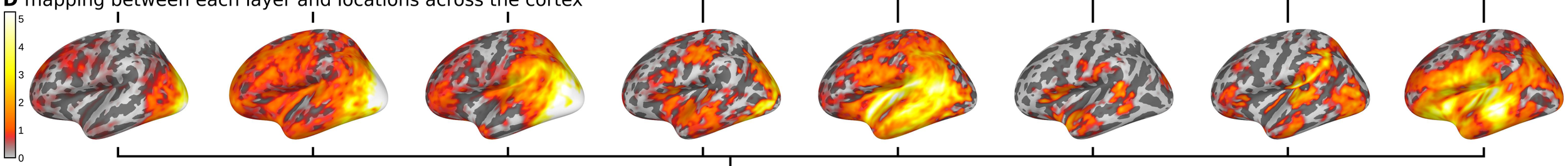
B experimental paradigm and experimental conditions



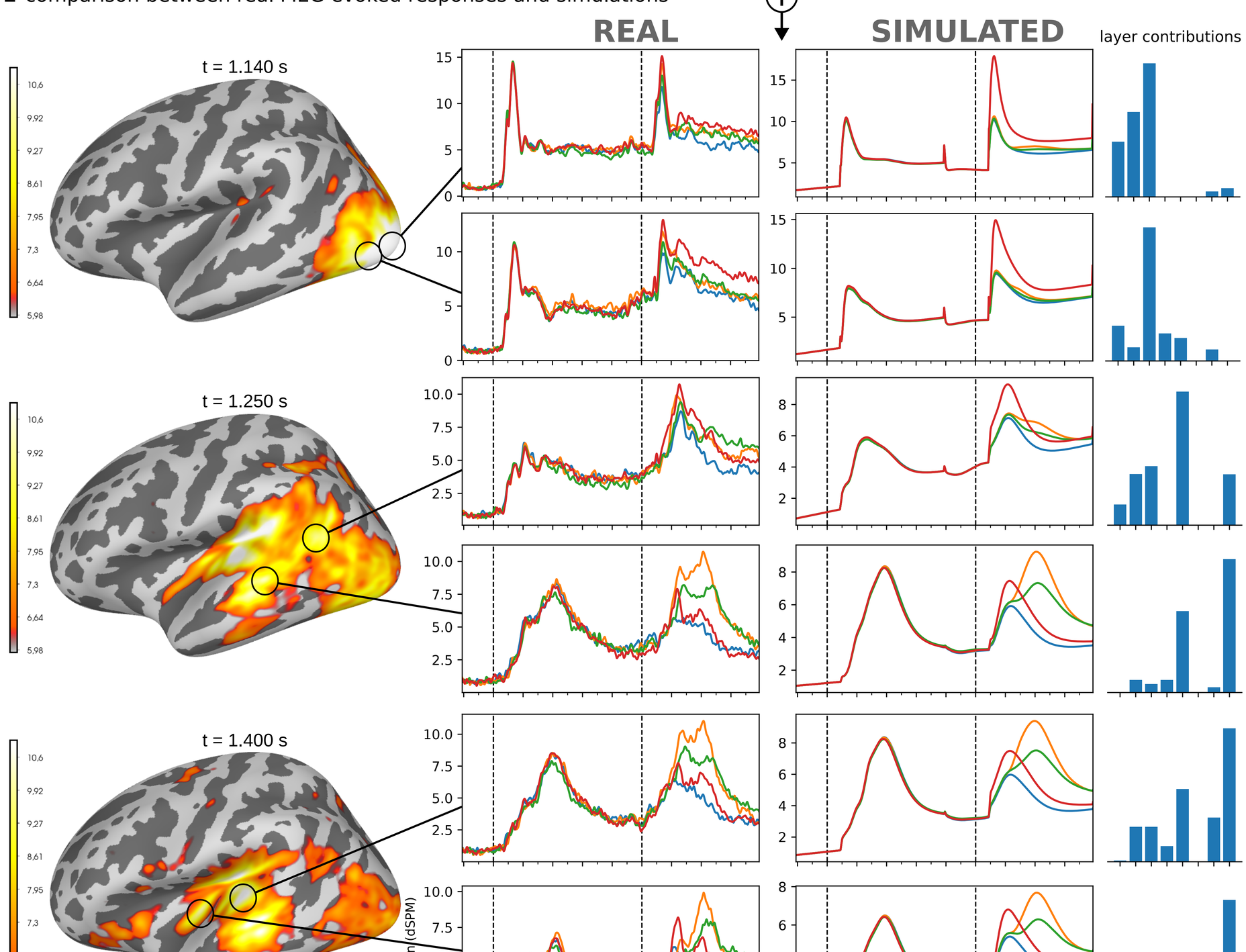
C total prediction error in each layer of the model (normalized between 0 and 1)



D mapping between each layer and locations across the cortex



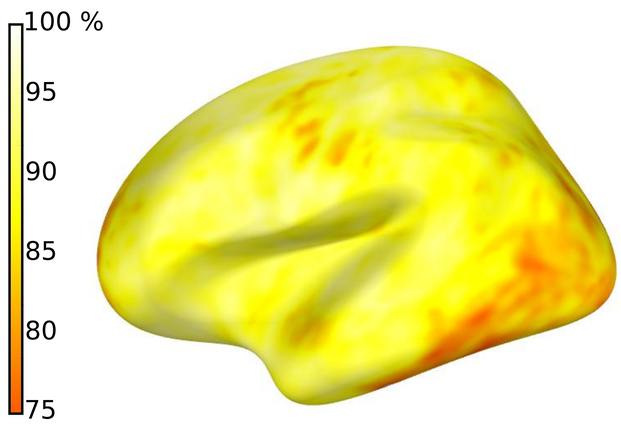
E comparison between real MEG evoked responses and simulations



RESULTS

- The model produces very evoked-like responses, with early responses being fast and later responses being slow.
- The repetition priming effect is only seen in later layers because of the 300-ms gap, which is enough to reset the early layer, but not the late ones.
- Layers are mapped to cortical locations in a manner that follows the ventral stream.
- The mixture of evoked response resembled the actual MEG timecourses (**F**), making a direct model-brain comparison possible.
- Experimental effects were mostly replicated.
- Differences between model and brain remain, e.g. the model responds too much to uppercase letters.

F explained variance (r^2)



REFERENCES

[1] Nour Eddine, S., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (2024). A predictive coding model of the N400. Cognition, 246, 105755. https://doi.org/10.1016/j.cognition.2024.105755

[2] van Vliet, M., Rinkinen, O., Shimizu, T., Niskanen, A.-M., Devereux, B., & Salmelin, R. (2025). Convolutional networks can model the functional modulation of the MEG responses associated with feed-forward processes during visual word recognition. eLife, 13, RP96217. https://doi.org/10.7554/eLife.96217