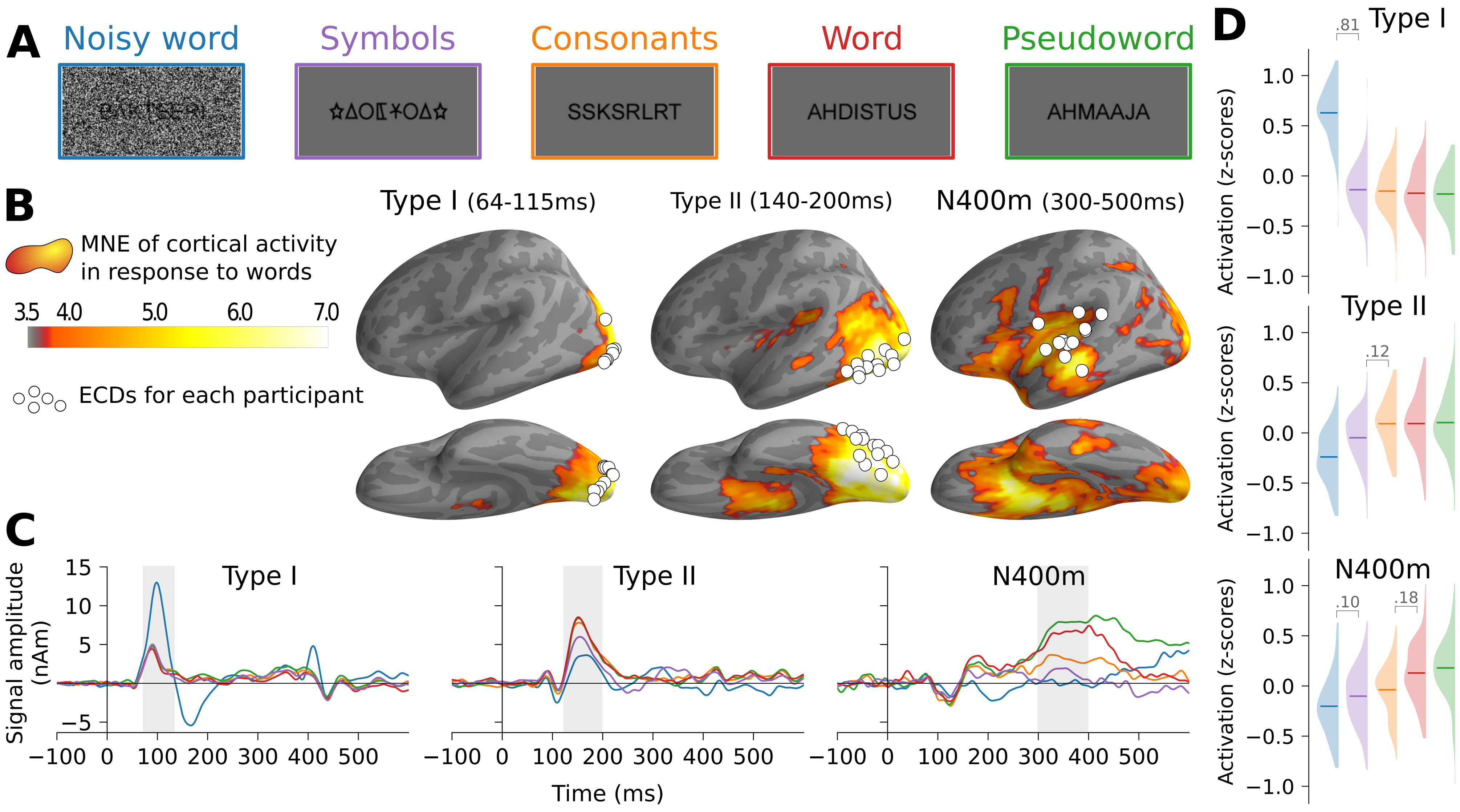
# Convolutional networks can model the functional modulation of MEG responses during reading

Aalto University
School of Science

Marijn van Vliet[1], Oona Rinkinen[1], Takao Shimizu[1], Anni-Mari Niskanen[1], Barry Devereux[2], Riitta Salmelin[1]

[1] Aalto University
[2] Queen's University Belfast




## INTRODUCTION: MEG responses during reading

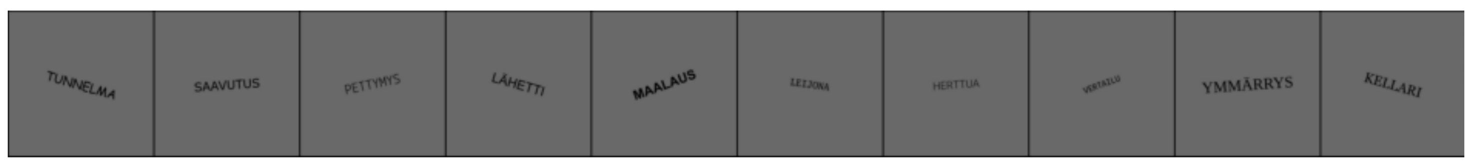
Reading elicits a series of evoked responses along the left ventral stream. In MEG, notable ones are the Type I, Type II and N400m. The location (Fig. 1B), timing (Fig. 1C) and functional behaviour (Fig. 1D) of these responses to different stimuli (Fig. 1A) tells the story of a processing pipeline starting with basic visual analysis (Type I) to letter detection (Type II) to lexical analysis (N400m). In this study, we sought to understand this pipeline better by implementing it as a computational model. In contrast to previous models, ours starts with raw pixels, which is required if one wants to reproduce all three afore-mentioned evoked responses. By presenting the same stimuli to both human and model, we evaluated the model's accuracy both qualitatively (response patterns to experimental contrasts) and quantitatively (correlation with MEG evoked response amplitudes).

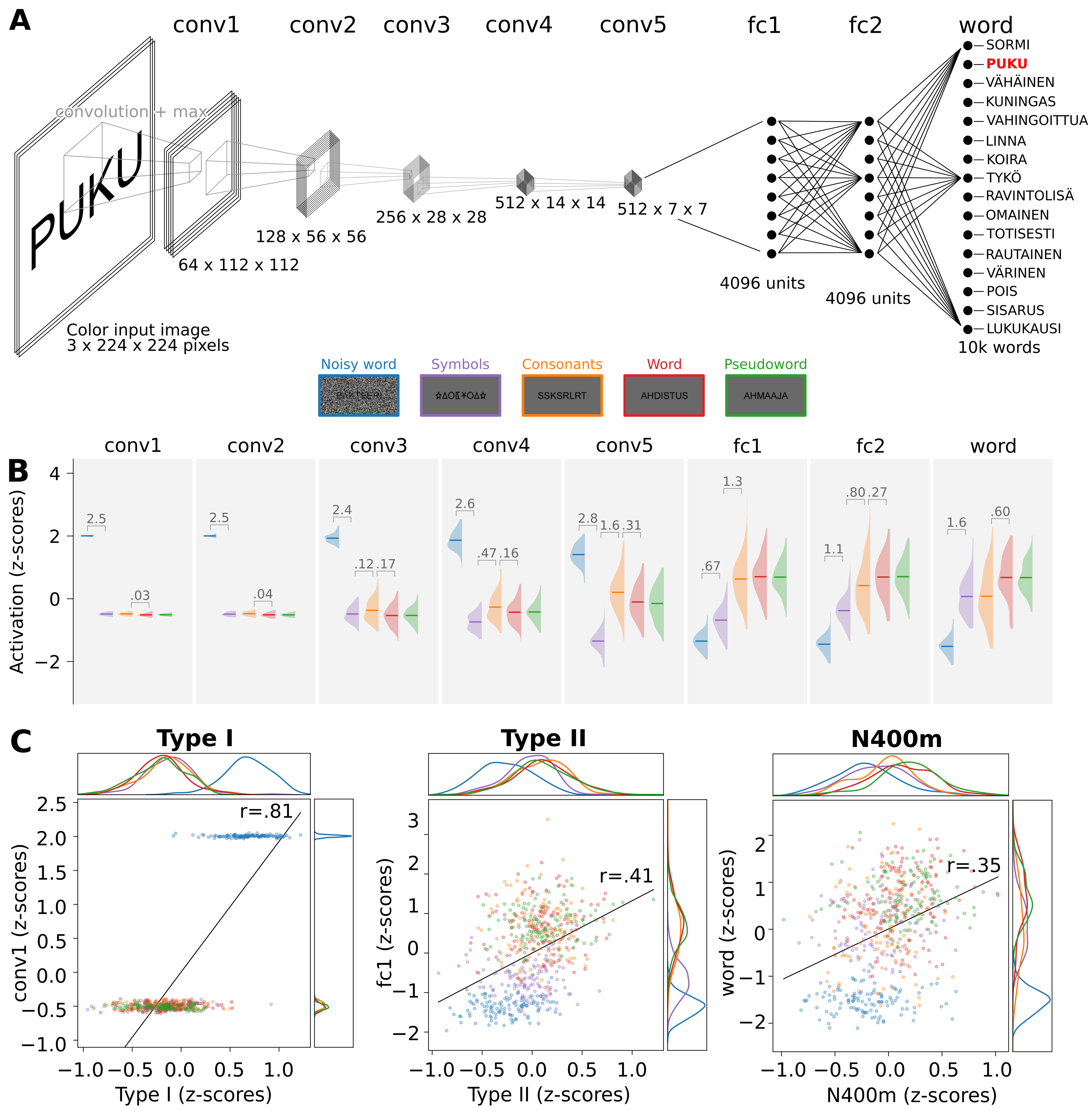
◀ **Fig. 1: Summary of the MEG results obtained by Vartiainen et al. (2011).**
**A:** Examples of stimuli used in the MEG experiment. Each stimulus contained 7-8 letters or symbols.
**B:** Source estimate of the evoked MEG activity, using MNE-dSPM. For each time interval, white circles indicate the location of the most representative left-hemisphere equivalent current dipole (ECD) for each participant.
**C:** Grand-average time course of signal strength for each group of ECDs in response to the different stimulus types. Shaded regions indicate time periods over which statistical analysis was performed.
**D:** For each group of ECDs shown in B, the distribution (and mean) of the grand-average response amplitude to each of the stimulus types, obtained by integrating the ECD signal strength over the time intervals highlighted in C. Significant differences between means are annotated.

## METHODS: Building the model

The model has a VGG11 architecture (Fig 3A). It was trained using a series of bitmap images containing Finnish words in different fonts, sizes and rotations (Fig. 2). Different models were tried (see extended results) and the final model has a vocabulary of 10k words, has noisy activations and during training, words were repeated proportional to their respective frequency of occurrence in Finnish texts. After training, we presented the model with the stimuli used in the MEG study and recorded the total ReLu activity ($\ell 2$ norm) in each layer. Note that the model has not seen any of the stimulus types in Fig. 1A other than proper words during training.

▼ **Fig 2. Examples of training images for the model.**






## RESULTS: Comparing model and brain

The three different types of layers in the model (Fig. 3A) produce response patterns (Fig. 3B) that closely match those of the Type I, Type II and N400m:

- The first 3 conv. layers match the Type I response. Only noisy stimuli evoke a large response.
- The two fully connected layers match the Type II response. Noisy stimuli now evoke a small response, and symbols evoke smaller re-sponses than stimuli containing letters.
- The output layer matches the N400m response. Noisy stimuli and symbols evoke smaller responses than letters, and additionally letter strings that don't follow proper consonant-vowel patterns (consonant strings) evoke smaller responses than those that do (words and pseudowords).

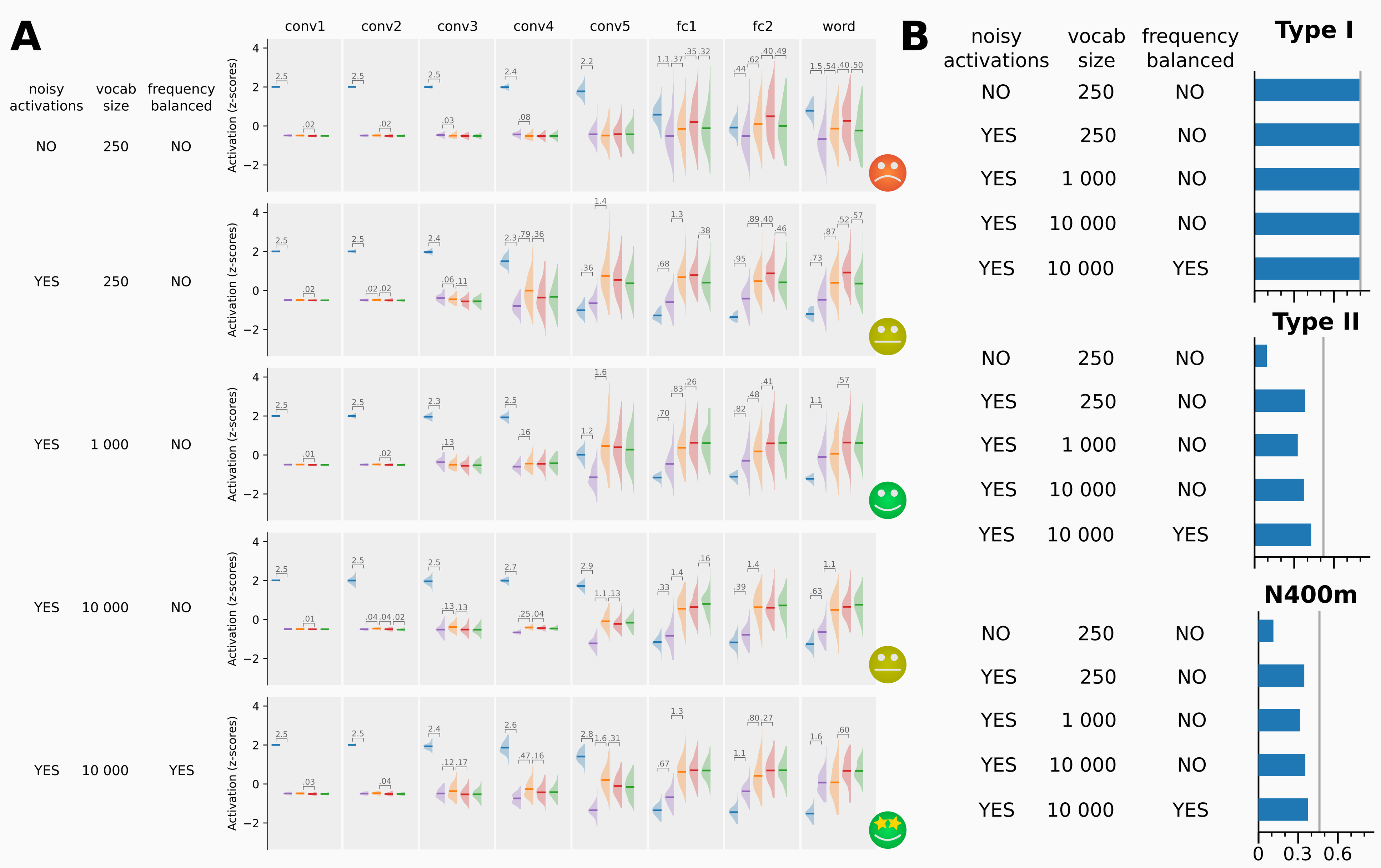
Because the exact stimuli of the MEG study can be presented to the model, model responses could be correlated directly to MEG response amplitudes (Fig. 3C). For the Type I response, correlation is at the noise ceiling, for the Type II and N400m, the model captures a good chunk of the variance, but not all (see extended results). Brain-wide correlations between the model and MNE source estimates (Fig. 3D) show the expected correlations with the three aforementioned evoked components, and not much spurious correlations elsewhere.

▶ **Fig 3. Comparison between model and MEG responses.**
**A:** Architecture of a VGG11 convolutional network.
**B:** Response patterns obtained from the best model. For each layer, the magnitude of ReLu activations in response to the same stimuli as used in the MEG experiment. Compare with Fig 1, panel D. Significant differences in means are shown.
**C:** Correlation between the response patterns obtained from evoked MEG activity and the response patterns obtained from three layers of the model.

## EXTENDED RESULTS: Hunting for the correct model

A vanilla VGG11 network architecture is not enough to capture all three evoked responses (Fig 4A, top row). Especially the response to noisy stimuli is incorrect. Adding noise helps (second row). To properly capture the response to pseudowords, a large enough vocabulary is necessary (third row). However, increasing the vocabulary too much produces incorrect responses to consonant strings (fourth row), unless repetition of training stimuli follows word frequencies (bottom row).

Such qualitative analysis was a more important tool for guiding our model design than raw model-brain correlations (Fig 4B). A better fit to one stimulus type can mask the total lack of fit to another stimulus type.




▶ Fig 4. Evaluating different models.
**A:** Response patterns obtained from various models. Emoticons indicate qualitative comparison with Fig 1D. Significant differences in means are shown.
**B:** For each model, the maximum correlation between all layers of the model and the MEG evoked responses. Estimated noise ceilings are indicated with vertical lines

## CONCLUSIONS

We present a computational model that shows how a series of convolution-and-pooling operations, followed by a sequence of linear transformations, is a viable method of visual word recognition that produces activity that mimics three well-studied MEG evoked responses. This is an important platform for refining our theories of word processing in the brain and a jumping off point for further explorations into semantic processing.

Full paper:

NBE
QUEEN'S UNIVERSITY BELFAST
ACADEMY OF FINLAND
ASC
CMHC