The visual processing streams for objects and words may interact in the human brain

Aalto University
School of Science

Shristi Baral Riitta Salmelin Marijn van Vliet

Department of Neuroscience and Biomedical Engineering
Aalto University, Finland

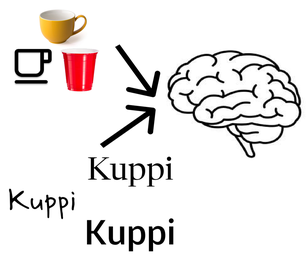
INTRODUCTION

The brain's ability to process written words and objects has been well-studied, but often in isolation. A key question remains: how does orthographic processing interact with visual object recognition, given that both share the same pathways?

Since object recognition is already developed when we learn to read as a child, the visual system must reorganize to accommodate written words, enabling both modalities to coexist and integrate efficiently.

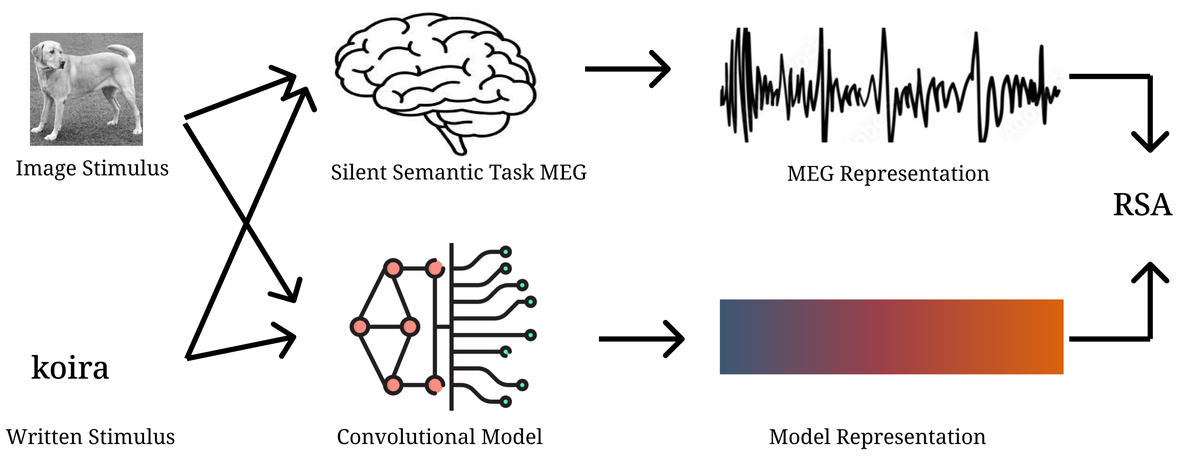
Written words derive meaning from real-world objects and concepts, requiring the brain's ventral stream to reorganise and process both modalities.

The question is how written words (e.g., "CUP") and visual objects (e.g., a picture of a cup) lead to the **same amodal neural representation.**



We investigated this using **Representational Similarity Analysis (RSA) between a CNN-based fusion model and MEG data.**

METHOD



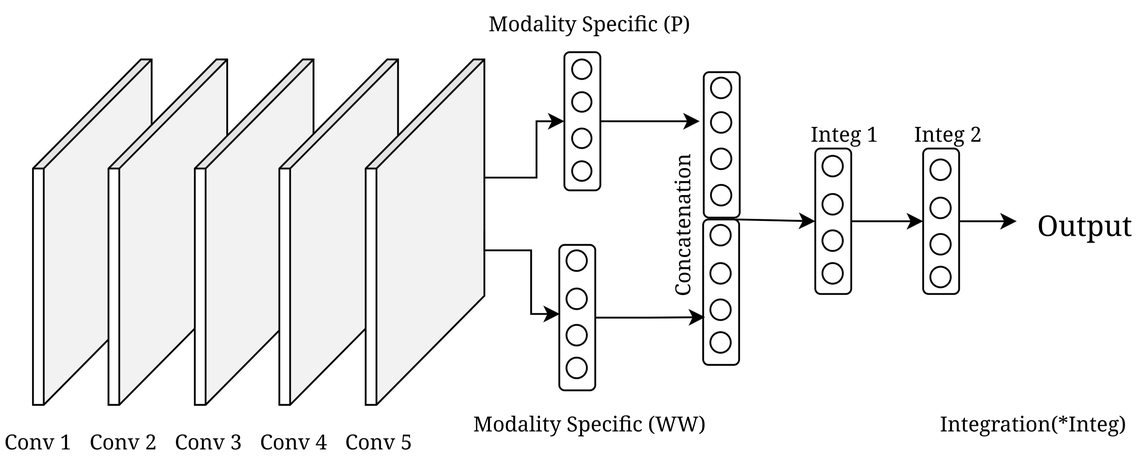
MEG

- 20 participants recruited for picture task and 20 for written word task
- Subjects were asked to silently think about the semantic properties of the stimulus while MEG recorded their response.
- Data cleaning and pre-processing steps are explained in Ghazaryan et.al (2023a, 2023b)

MODEL

Model Architecture

- Pre-trained Model based on VGG-11 is used as a model of cortical processing
- Model is trained in three different way by initializing pre-trained weights and continue with
  - Only Pictures as input (PN)
  - Only written words as input (WWN)
  - Both picture and written words input resulting fusion network (FN)



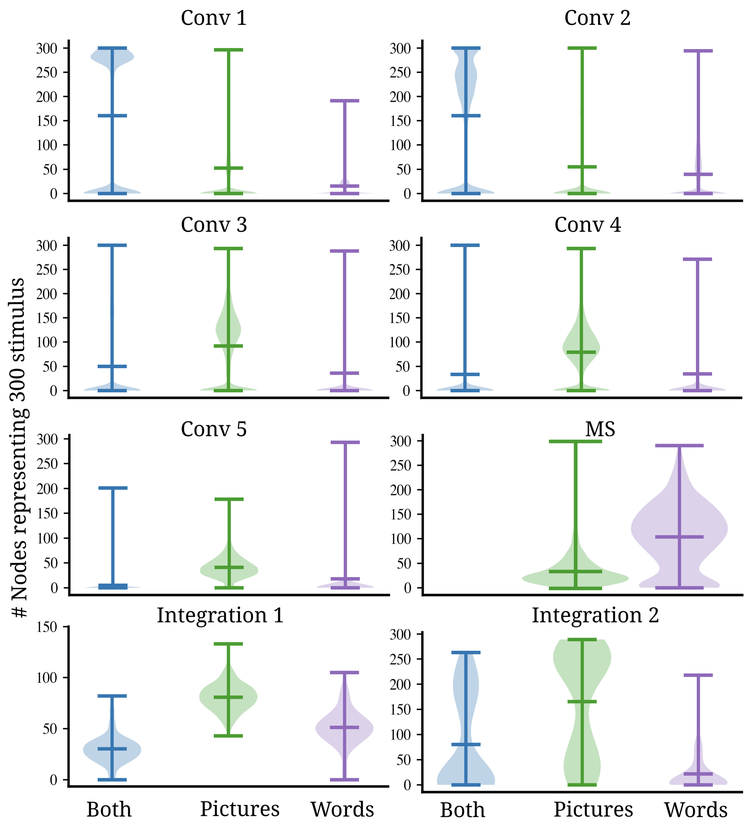
Training Parameter

- Training data was generated by modifying Image-net, consolidating similar categories e.g., all breeds of dogs into one category 'dog'.
- Class names were translated to Finnish language to generate written word dataset.
- Written word data was generated by rendering words in a canvas.
- Pre-trained weights were unfrozen gradually as training approached epoch 5, 20, 25.
- Differential learning rate was applied (picture: 0.005, written words: 0.001)
- Each model was trained for 30 epochs.

RESULT

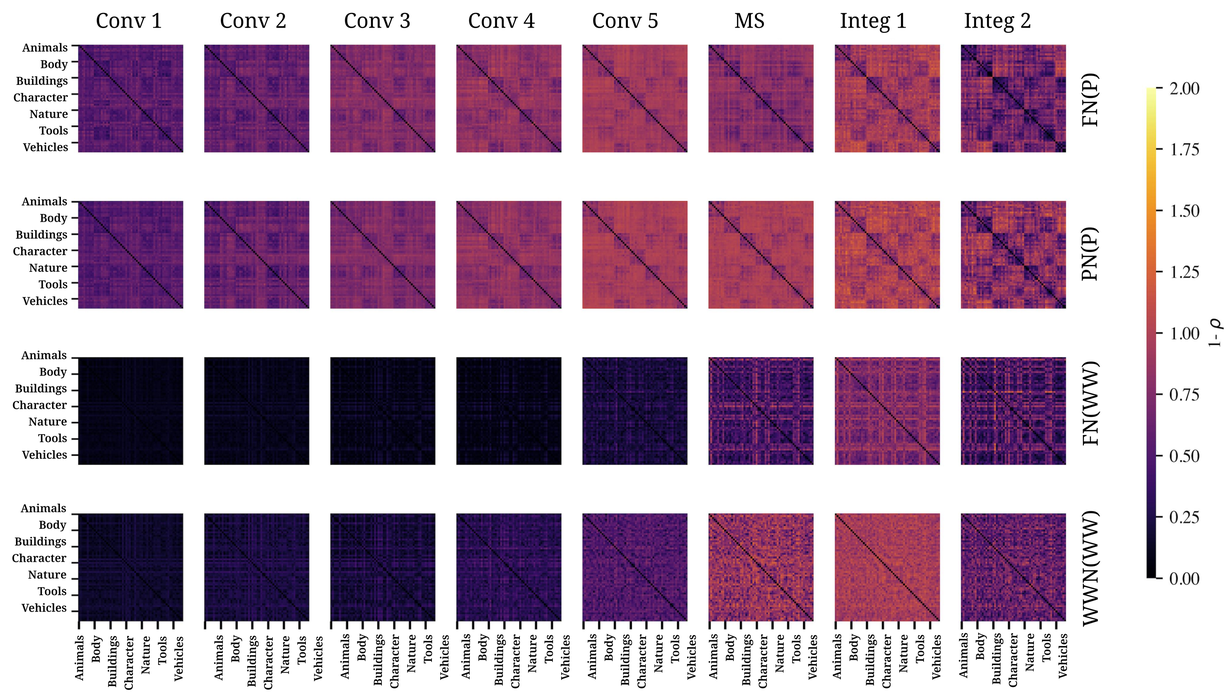
Node Allocation In The Network

- In the initial layers, units respond to basic visual features present in both written words and visual objects.
- In later layers, specialization emerged as more units respond to either written words or visual objects.
- In the integration layers, these specialized streams merged and more units once again respond to both written words and visual objects, some nodes still responding to pictures and written words only.



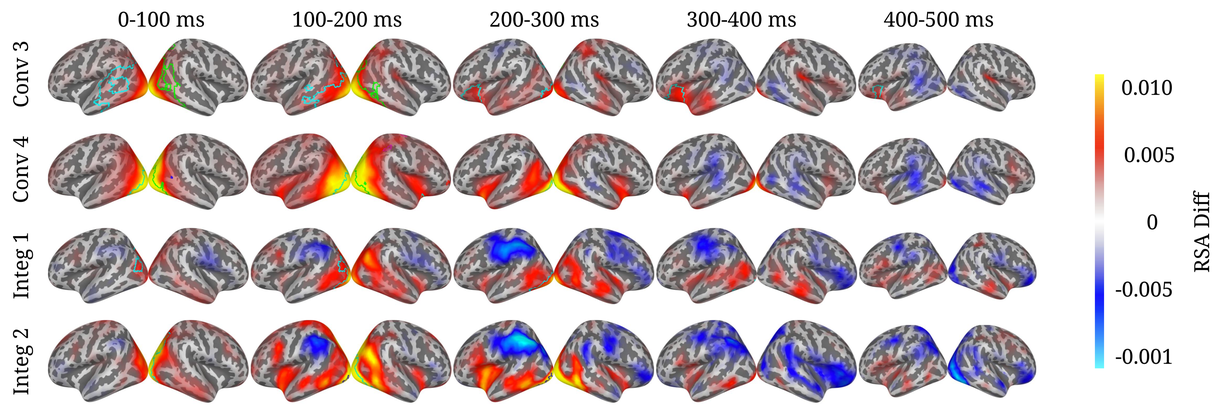
Model RDMs

More distinct different semantic categories in FN(P) compared to PN(P)



RSA Difference FN(P)- PN(P)

The boundary highlights the areas with significant differences in RSA P(N) and FN(P) using MEG-picture data which suggests the fusion model, trained on both pictures and written words, aligns better with brain data compared to the model trained solely on pictures. This indicates representation of both modalities may persist along the processing pathway, some areas with fewer shared nodes.



p<0.05; (Only RSA difference of 4 layers shown due to lack of space)

CONCLUSION

In the fusion model, the internal representations of the picture became more distinct between semantic categories, implying that the existence of units tuned to written language that also sharpened the semantic representations of pictures. The RSA shows the fusion model that incorporated written words achieved more brain-like representations than only pictures.

Just like the early layers of a model, some neurons in the visual system may initially respond to basic features of written words and images. Further along the processing pathway specialized neurons may exist which primarily respond to one modality. However, later processing may form amodal representation where some neurons respond to both.

a) Gayane Ghazaryan, Marijn van Vliet, Aino Saranpää, Lotta Lammi, Tiina Lindh-Knuutila, Annika Hultén, Sasa Kivisaari & Riitta Salmelin (14 Jun 2023): Trials and tribulations when attempting to decode semantic representations from MEG responses to written text, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2023.2219353
b) Ghazaryan, G., van Vliet, M., Lammi, L. et al. Cortical time-course of evidence accumulation during semantic processing. Commun Biol 6, 1242 (2023). https://doi.org/10.1038/s42003-023-05611-6