Computational models of vision do not explain the effect of expertise on neural processing of visual Braille

Filippo Cerpelloni (filippo.cerpelloni@kuleuven.be)
KU Leuven, Belgium; UCLouvain, Belgium

Olivier Collignon
UCLouvain, Belgium

Hans Op de Beeck
KU Leuven, Belgium

Abstract

The human visual stream adapts to process letters and words at different processing stages (Vinckier et al., 2007), even when the stimuli do not share canonical script features, like Braille (Cerpelloni et al., 2024). This supports an interactive account of the Visual Word Form Area (VWFA). Here we expand these findings to test the organization of peculiar visual features in computational models. By training a benchmark convolutional neural network (AlexNet) to classify words in the Latin script (literacy) and then in the Braille script (expertise), we model the processing of reading visual Braille and explore the network's representations at different stages. We observe a similar degree of clustering between models before and after training on Braille. The lack of alignment between the visual processing of the computational models and the effect of expertise highlighted by neural data suggests that the fundamental processing of reading cannot be fully explained by the visual characteristics of the script, but necessarily relies on other mechanisms, among which language connections.

Keywords: convolutional neural networks; object recognition; reading; words

Introduction

In humans, the Visual Word Form Area (VWFA) supports the processing of written scripts (Cohen et al., 2002). Models of how VWFA acquire its selectivity for orthography either emphasise the progressive integration of line-junctions (Bola et al., 2017; Szwed et al., 2009, 2011) happening in the visual stream and culminating in these region (Dehaene et al., 2005; Vinckier et al., 2007) or instead the connectivity of this area with the language network (Price & Devlin, 2003; Saygin et al., 2016; Wang et al., 2022). Recently, high-resolution 7T fMRI has found that VWFA represents different scripts with multiple sub-patches (Zhan et al., 2023). Moreover, the multivariate pattern analysis of different scripts (Latin-based and Braille) shows similar coding principles based on the statistical regularities of the stimuli, but relying on a segregated organization of scripts in VWFA and across the visual stream (Figure. 1; Cerpelloni et al., 2024).

We adapted recent computational studies, which explored the visual representations of letters (Janini et al., 2022) and created synthetic models of the visual stream to read words (Agrawal & Dehaene, 2024; Hannagan et al., 2021), to test the impact of expertise with visual features in processing Braille. We trained AlexNet (Krizhevsky et al., 2017) models to process Latin alphabet and then Braille words, paralleling expert visual Braille reading. We observe that the statistical regularities of the stimuli, but not the training, explain the increased clustering of representations of Braille script.

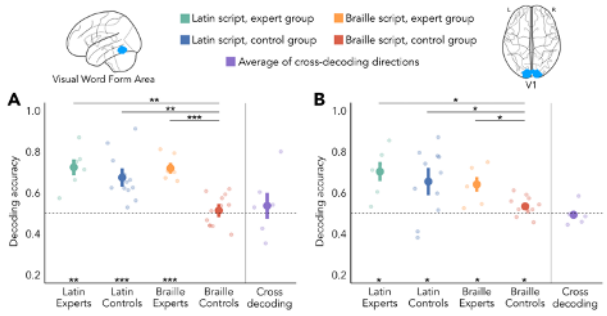

Figure 1: Neural organization of visual Braille. *Pairwise decoding accuracies for the neural activation of expert visual Braille readers and naïve controls in VWFA and V1. When a participant was able to read the script, the linguistic information can be decoded. In V1, decoding of Braille stimuli is possible for naïve participants too, but much weaker than in experts.*

Methods

Stimuli

For training, we replicated the stimulus set used by Agrawal and Dehaene (2024) with CORnet models. We included one thousand Dutch words in four font variations of the Latin alphabet (Arial, Times New Roman, American Typewriter, Futura) and then in the Braille alphabet; five variations in size; eleven and five variations on the x and y axes respectively. Similarly to Cerpelloni and colleagues (2024), we developed a test set of ***real words*** (from the training set), ***pseudo words***, ***non words***, and a ***fake script*** condition in both the Latin (Arial) and the Braille scripts (Figure 2), to test the effect

of the statistical regularities that differentiate these conditions at different stages of the network. All stimuli underwent the same size and position variations of the training set.

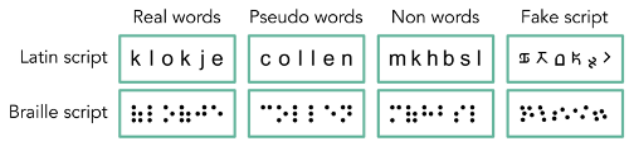

***Figure 2: Example of stimuli used in the test set, with 4 conditions: Real words** possess a lexical entry (output node), high-frequency phonological units, orthography. **Pseudo words** do not possess a lexical entry; **non words** are not made of frequent phonological units; **fake script** is composed out of the same lines of non words but arranged in a novel structure.*

Networks, training, analyses

We used five instances of AlexNet previously trained on ImageNet (Krizhevsky et al., 2017). After resetting the last layer's weights, we trained the networks to classify Dutch words in the Latin alphabet (literacy acquisition; network naïve to Braille) and, in a second step, added the same words presented in the Braille alphabet and mapped to the same output units (expert network).

We extracted the activations from the networks' ReLU stages at the last epoch of training, presenting the test set of stimuli. We then computed the Euclidean distance between stimuli (across their variations) following the methods of Janini and colleagues (2022) for the identity of Latin alphabet letters. We used the resulting dissimilarity matrices (RDMs) to compute a measure of the clustering of representations (average dissimilarity between conditions minus the dissimilarity within conditions, divided by the average dissimilarity) in the four different conditions (words, pseudowords, nonwords, and fake script) at different processing stages.

Results

We extracted the network's representations for Braille in AlexNet models expert or naïve to visual Braille (Figure 3A). We observe a main effect of the layer ($p < 0.001$) and no general difference between expertise level of the networks ($p = 0.8$). We do note a significant interaction between the factors ($p < 0.001$), to be attributed to the differences in the last convolutional layer and in the last fully connected layer (Figure 3B).

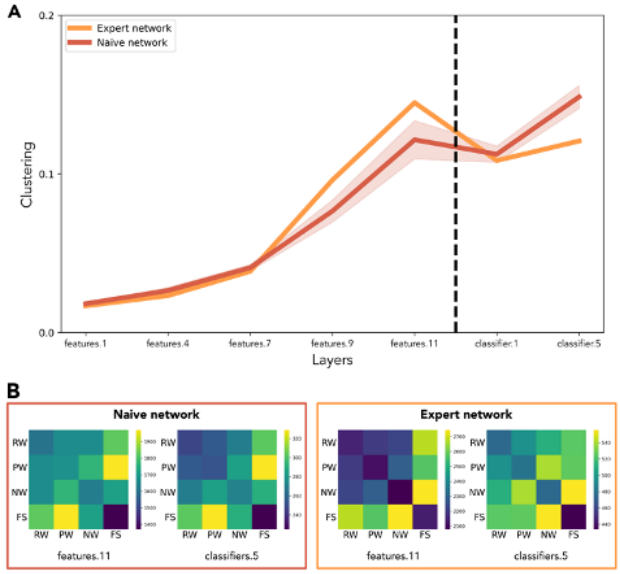

***Figure 3: Clustering of dissimilarity between layers across AlexNet expertise.** A. Comparison of the degree of clustering between networks trained on Braille or naïve to it. Shaded area indicates 95% confidence interval. B. Representations at crucial nodes of the networks. Colorbars indicate category dissimilarity within a network / layer*

Conclusions

Overall, this preliminary result indicates that visual strategies to process Braille in artificial networks lead to a similar clustering of stimuli with different statistical regularities (linguistic properties), independently of whether or not the visual network was trained with Braille. Such similarity is at odds with the neural data that showed a strong difference in neural encoding as a result of expertise (Cerpelloni et al., 2024). Here we assume that the clustering index should show a similar pattern as the decoding approach used with fMRI, as a strong decoding requires clustering. This discrepancy hints at a possible role of linguistic processing in the visual brain, given that such processing is not included in visual artificial networks. Further simulation studies will pinpoint the necessary and sufficient conditions to reproduce the effect of expertise in the human brain.

References

Agrawal, A., & Dehaene, S. (2024). Cracking the neural code for word recognition in convolutional neural networks. *PLOS Computational Biology*, *20*(9), e1012430. doi: 10.1371/journal.pcbi.1012430

Bola, Ł., Radziun, D., Siuda-Krzywicka, K., Sowa, J. E., Paplińska, M., Sumera, E., & Szwed, M. (2017). Universal Visual Features Might Be Necessary for Fluent Reading. A Longitudinal Study of Visual Reading in Braille and Cyrillic Alphabets. *Frontiers in Psychology*, *8*. doi: 10.3389/fpsyg.2017.00514

Cerpelloni, F., Van Audenhaege, A., Matuszewski, J., Gau, R., Battal, C., Falagiarda, F., Op de Beeck, H., & Collignon, O. (2024). Expertise in reading visual Braille co-opts the reading network. *bioRxiv*. doi: 10.1101/2024.05.08.593104

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*(5), 1054–1069. doi: 10.1093/brain/awf094

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, *9*(7), 335–341. doi: 10.1016/j.tics.2005.05.004

Hannagan, T., Agrawal, A., Cohen, L., & Dehaene, S. (2021). Emergence of a compositional neural code for written words: Recycling of a convolutional neural network for reading. *Proceedings of the National Academy of Sciences*, *118*(46), e2104779118. doi: 10.1073/pnas.2104779118

Janini, D., Hamblin, C., Deza, A., & Konkle, T. (2022). General object-based features account for letter perception. *PLOS Computational Biology*, *18*(9), e1010522. doi: 10.1371/journal.pcbi.1010522

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. doi: 10.1145/3065386

Kubilius, J., Schrimpf, M., Nayebi, A., Bear, D., Yamins, D. L. K., & DiCarlo, J. J. (2018). *CORnet: Modeling the Neural Mechanisms of Core Object Recognition*. Neuroscience. doi: 10.1101/408385

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, *19*(3), 473–481. doi: 10.1016/S1053-8119(03)00084-3

Saygin, Z. M., Osher, D. E., Norton, E. S., Youssoufian, D. A., Beach, S. D., Feather, J., Gaab, N., Gabrieli, J. D. E., & Kanwisher, N. (2016). Connectivity precedes function in the development of the visual word form area. *Nature Neuroscience*, *19*(9), 1250–1255. doi: 10.1038/nn.4354

Szwed, M., Cohen, L., Qiao, E., & Dehaene, S. (2009). The role of invariant line junctions in object and visual word recognition. *Vision Research*, *49*(7), 718–725. doi: 10.1016/j.visres.2009.01.003

Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabrègue, R., Amadon, A., & Cohen, L. (2011). Specialization for written words over objects in the visual cortex. *NeuroImage*, *56*(1), 330–344. doi: 10.1016/j.neuroimage.2011.01.073

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. *Neuron*, *55*(1), 143–156. doi: 10.1016/j.neuron.2007.05.031

Wang, S., Planton, S., Chanoine, V., Sein, J., Anton, J.-L., Nazarian, B., Dubarry, A.-S., Pallier, C., & Pattamadilok, C. (2022). Graph theoretical analysis reveals the functional role of the left ventral occipito-temporal cortex in speech processing. *Scientific Reports*, *12*(1), 20028. doi: 10.1038/s41598-022-24056-1

Zhan, M., Pallier, C., Agrawal, A., Dehaene, S., & Cohen, L. (2023). Does the visual word form area split in bilingual readers? A millimeter-scale 7-T fMRI study. *Science Advances*.