Temporal dynamics of face recognition: Insights from combining MEG and Artificial Neural Networks

Hamza Abdelhedi (hamza.abdelhedi@umontreal.ca)
Dept. Computer Science and Operational Research, Université de Montréal, Montreal, Quebec, Canada
Mila (Quebec AI institute), Montreal, Quebec, Canada

Shahab Bakhtiari (shahab.bakhtiari@umontreal.ca)
Dept. Psychology, Université de Montréal, Montreal, Quebec, Canada
UNIQUE (Unifying Neuroscience and AI), Montreal, Quebec, Canada
Mila (Quebec AI institute), Montreal, Quebec, Canada

Karim Jerbi (karim.jerbi@umontreal.ca)
Dept. Psychology, Université de Montréal, Montreal, Quebec, Canada
UNIQUE (Unifying Neuroscience and AI), Montreal, Quebec, Canada
Mila (Quebec AI institute), Montreal, Quebec, Canada

Abstract

Humans excel at face recognition, relying on a specialized system whose representation of familiar and unfamiliar faces have been debated. Some would argue that unfamiliar and familiar faces are processed in the same way, while others claim otherwise. Similarly, artificial neural networks (ANNs) have shown remarkable abilities, prompting discussions on their similarities to human face processing and their use as models of the brain. This study employs Convolutional Neural Networks (CNNs) and Magnetoencephalography (MEG) to explore the signatures of face recognition and familiarity, and investigate whether the face selective areas in the brain are specialized only for faces (domain-specific) or are developed for more general purposes (domain-general). Our findings reveal distinct brain responses: occipital areas distinguish faces from non-faces, while fusiform and inferior temporal areas engage in familiar face recognition. When training our ANNs on face recognition, we observe significant results supporting the idea of domain-specific brain regions for faces. Our findings shed new light on the temporal dynamics of familiar versus unfamiliar face processing in the visual cortex and more globally, highlight the potentials of combining ANNs and MEG to uncover the neural mechanisms that mediate face processing in humans.

Introduction

Humans are remarkably adept at face recognition. We effortlessly recognize and categorize familiar faces. This ability relies on group of brain regions with different degrees of specialization (Haxby et al., 2001; Apurva et al., 2004; Yovel & Kanwisher, 2005). The concept of face familiarity has been a topic of extensive research. Some studies have suggested that the an electrophysiology face perception component (M170) only appears in response to familiar faces, while others claim that it appears for all faces, with a later component (M400) linked more to familiarity (Megreya & Burton, 2006; Johnston & Edmonds, 2009; Dobs, Isik, Pantazis, & Kanwisher, 2019). Another key debate centers on how the brain represents faces. Some studies have advocated for domain-specific brain specialization dedicated to face processing, while others have proposed domain-general face representation (Tsao & Livingstone, 2008; Dobs, Martinez, Kell, & Kanwisher, 2022; Vinken, Prince, K., & Livingstone, 2023; Shi et al., 2023).

However, humans are not the only face recognition experts. Convolutional neural networks (CNNs), inspired by the brain's structure and function, have recently surpassed human performance on many tasks, including face recognition (Dyck & Gruber, 2023), and have also been proposed as models of the visual cortex, including the face system (Yamins et al., 2014; Richards et al., 2019; Storrs, Kietzmann, Walther, Mehrer, &

Kriegeskorte, 2020). While some studies have shown that training on face recognition is not necessary for achieving a similar Brain-CNN face representation (Grossman et al., 2019), others have shown that it is necessary in order to capture the behavioral properties of human face perception (Dobs, Yuan, Martinez, & Kanwisher, 2023). However, most of these studies used unfamiliar face stimuli which neglected the question of familiarity impact on how the brain processes and represents faces. Additionally, most findings rely on fMRI or behaviroal data (Tsantani et al., 2021), which limits the understanding of the temporal dynamics of face processing. Previous work suggests that magnetoencephalography's (MEG) high spatiotemporal resolution is crucial to investigate the links between CNNs and brain dynamics (Kietzmann et al., 2019).

To complement this research, we use CNNs as tools to explore the neuromagnetic signatures of face recognition in the brain, as well as domain generality versus specificity of face representation.

Methods

MEG Data   We used publicly available MEG data acquired from 16 subjects performing arecognition task with familiar (famous), unfamiliar, and scrambled face stimuli (Wakeman & Henson, 2015). We followed the preprocessing steps from Jas et al. (2018). This involved projecting the data to source space to estimate the underlying neural activity for voxels. For each subject, we segmented into 1s epochs per voxel. Voxels were then assigned to 450 regions of interest (ROIs) based on the Destrieux atlas (Destrieux, Fischl, Dale, & Halgren, 2010).

Network training and activations' extraction   We used the backbone of FaceNet (Schroff, Kalenichenko, & Philbin, 2015) for our analysis due to it's success in modeling the face system (Abdelhedi & Jerbi, 2022; Jiahui et al., 2023). We trained the same architecture on three distinct tasks. A Face Recognition task (FR-task) trained on VGGFace dataset (Cao, Shen, Xie, Parkhi, & Zisserman, 2017) and fine-tuned on a distribution of celebA dataset (gray scale, cropped to include only the face,similar to the stimuli used above) (Liu, Luo, Wang, & Tang, 2015). An Object Recognition task (OR-task) trained on ImageNet (Deng et al., 2009) with excluding human-face stimuli. Finally, a Dual Task (Du-Task) by augmenting the ImageNet with a human-face class selected from celebA. To extract the layer activations necessary for subsequent similarity analyses, we fed the three types of face stimuli used in the MEG experiment through the networks. The responses from all layers were collected and stored for further analysis.

Representational Similarity Analysis (RSA)   Inline with previous research, we employed RSA (Kriegeskorte, Mur, & Bandettini, 2008) to evaluate the similarity between the activation patterns of the artificial (FaceNet) and biological (MEG) systems when presented with the same stimuli. We quantified pairwise dissimilarities between the brain and the CNN activations for each stimulus using representational dissimilarity matrices (RDMs). We computed three distinct RDMs: One

per each type of face stimuli (Familiar: Fam-RDM, Unfamiliar: Unfam-RDM, and Scrambled: Scram-RDM) for both MEG and FaceNet. Layer-wise RDMs were computed for the three training objectives (explained above) and for a randomly initialized FaceNet. For the MEG data, we computed the three RDMs for each time point for every ROI by considering the activity of all voxels assigned to it. All correlations within the RDM cells and across RDMs were calculated using Pearson correlation. **Noise Celing (NC)** To quantify the potential influence of noise on our data, we calculated an upper bound NC using the method described in (Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016). We present the results as the proportion of variance explained by dividing the similarity score by the NC.

Results

To compare neural representations of familiar versus unfamiliar faces, we plot the CNN-brain representational similarities across time for FaceNet trained on Face Recognition, when presented with three distinct types of stimuli. The results depicted in Figure 1A, demonstrate that the FR-task FaceNet exhibits peak similarity for both familiar (0.59, p<0.05) and unfamiliar (0.6, p<0.05) stimuli in comparison with the Occipital ROI. Notably, the peak for unfamiliar stimuli occurs earlier at 114ms, roughly 30ms before the peak elicited by familiar stimuli (154ms). However, in Figure 1B and 1C, which show comparisons of Fusiform and Inferior Temporal (IT), peak similarities are only observed for familiar stimuli. These peaks occur at 170ms (0.5, p<0.05) for the fusiform area and 365ms (0.48, p<0.05) for the IT, with a second smaller peak for the fusiform at 490ms (0.4, p < 0.05). Examining the brain maps in Figure 2, demonstrating similarity scores obtained from Figure 1 at the peak time points but for all brain ROIs, a clear difference between unfamiliar and familiar stimuli becomes evident. When using unfamiliar stimuli (Figure 2A), only the Occipital ROI shows a strong correlation with FaceNet activations. In contrast, familiar stimuli (Figure 2B) elicit a peak of similarity in the same region along with an additional cluster of regions also showing high similarity with FaceNet including The fusiform (Figure 2C) and IT (Figure 2D) regions.

Examining the impact of different training tasks, we observe that using only face recognition as the training objective yields the highest similarity peaks in the ROIs (Figure 1A, 1B, and 1C). In contrast, smaller similarity peaks are seen at 153ms for the dual-task and object recognition training objective. However, these peaks are comparable to the results obtained with an untrained architecture (Figure 1A). When analyzing the distribution of scores obtained using the dual-task trained FaceNet (Figure 1B), we can observe a cluster of similarities in occipital areas, although less significant than the cluster obtained with the face recognition-trained FaceNet. This cluster becomes even less prominent in the object recognition-trained FaceNet.
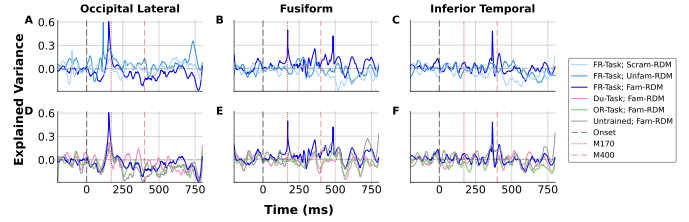

Figure 1: Explained variance (similarity score normalized by noise ceiling) across time for three ROIs and FaceNet (FR-task). Results depicted for the layer that gave highest scores but the pattern is sustained for other layers. Panels **(A-C)** compare brain activity with FaceNet (FR-Task) across stimuli types. Panels **(D-F)** quantify MEG-FaceNet similarity change for different training tasks using Fam-RDM. Columns denote brain region. Panels **(A, D)** are for Lateral Occipital, **(B, E)** for Fusiform, **(C, F)** for Inferior Temporal (IT).
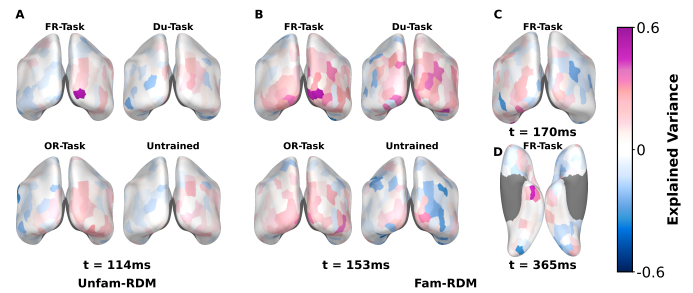

Figure 2: Brain maps of the time points with peak similarity. **(A)**: 114ms, (unfamiliar stimuli). **(B)**:153ms (familiar stimuli). **(C)**: 170ms (familiar stimuli). **(D)**: 365ms (familiar stimuli).

Discussion

Our study investigated how brain-ANNs representation similarity was affected by training conditions and stimulus types. First, the reported peak similarities exceeded those obtained for controls (scrambled faces and untrained network). Second, comparing familiar versus unfamiliar stimuli revealed how face perception temporally evolved in the brain. Our results supported the hypothesis that both M170 and M400 components are modulated by face familiarity. Regarding domain-general versus domain-specific, our results favored the latter. The domain-general hypothesis would predict that the OR-task ANN would do as well in terms of similarity with the visual system during face recognition, while our results showed the FR-task as necessary to capture the signatures of face processing.

We plan to extend our analysis in several directions. First, we will examine the frequency bands of the MEG to explore the spectral signatures of familiar versus unfamiliar face perception. Second, we will investigate if the results hold using various CNNs. Finally, we believe this work contributes to a much-needed effort in the field: building in silico models of the brain to understand its function and test related hypotheses.

Acknowledgments

References

Abdelhedi, H., & Jerbi, K. (2022). *Linking human and artificial neural representations underlying face recognition: Insights from meg and cnns.* doi: 10.32470/ccn.2022.1320-0

Apurva, N., Murty, R., Teng, S., Beeler, D., Mynick, A., Oliva, A., ... Beeck, H. O. D. (2004). Visual experience is not necessary for the development of face-selectivity in the lateral fusiform gyrus. *PNAS*, *117*, 23011-23020. doi: 10.1073/pnas.2004607117/-/DCSupplemental

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2017). *Vggface2: A dataset for recognising faces across pose and age.* doi: 10.48550/ARXIV.1710.08092

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016, 6). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*. doi: 10.1038/srep27755

Deng, J., Dong, W., Socher, R., Li, L., Kai, L., & Li, F. (2009). Imagenet: A large-scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition.* doi: 10.1109/CVPR.2009.5206848

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010, 10). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*. doi: 10.1016/j.neuroimage.2010.06.010

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019, 12). How face perception unfolds over time. *Nature Communications*. doi: 10.1038/s41467-019-09239-1

Dobs, K., Martinez, J., Kell, A. J. E., & Kanwisher, N. (2022). *Brain-like functional specialization emerges spontaneously in deep neural networks.*

Dobs, K., Yuan, J., Martinez, J., & Kanwisher, N. (2023, 8). Behavioral signatures of face perception emerge in deep neural networks optimized for face recognition. *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.2220642120

Dyck, L. E. V., & Gruber, W. R. (2023, 10). Modeling biological face recognition with deep convolutional neural networks. *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn$_{a0}$2040

Grossman, S., Gaziv, G., Yeagle, E. M., Harel, M., Mégevand, P., Groppe, D. M., ... Malach, R. (2019, 12). Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nature Communications*. doi: 10.1038/s41467-019-12623-6

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001, 9). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. doi: 10.1126/science.1063736

Jas, M., Larson, E., Engemann, D. A., Leppäkangas, J., Taulu, S., Hämäläinen, M., & Gramfort, A. (2018). A reproducible meg/eeg group study with the mne software: Recommendations, quality assessments, and good practices. *Frontiers in Neuroscience*. doi: 10.3389/fnins.2018.00530

Jiahui, G., Feilong, M., Matteo, V., Nastase, S. A., Haxby, J. V., & Gobbini, M. I. (2023). Modeling naturalistic face processing in humans with deep convolutional neural networks. *Proceedings of the National Academy of Science*. doi: 10.1073/pnas.2304085120

Johnston, R. J., & Edmonds, A. J. (2009). *Familiar and unfamiliar face recognition: A review.* doi: 10.1080/09658210902976969

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1905544116

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. doi: 10.3389/neuro.06.004.2008

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of international conference on computer vision (iccv).*

Megreya, A., & Burton, A. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory Cognition*. doi: doi.org/10.3758/BF03193433

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A. J., ... Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*. doi: https://doi.org/10.1038/s41593-019-0520-2

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering.. doi: 10.1109/cvpr.2015.7298682

Shi, Y., Bi, D., Hesse, J. K., Lanfranchi, F. F., Chen, S., & Tsao, D. Y. (2023). Rapid, concerted switching of the neural code in inferotemporal cortex 1. Retrieved from https://doi.org/10.1101/2023.12.06.570341 doi: 10.1101/2023.12.06.570341

Storrs, K. R., Kietzmann, T. C., Walther, A., Mehrer, J., & Kriegeskorte, N. (2020). Diverse deep neural networks all predict human it well, after training and fitting. *bioRxiv*. doi: 10.1101/2020.05.07.082743

Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A. L., McGettigan, C., & Garrido, L. (2021, 3). Ffa and ofa encode distinct types of face identity information. *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.1449-20.2020

Tsao, D. Y., & Livingstone, M. S. (2008, 7). *Mechanisms of face perception.* Annual Reviews Inc. doi: 10.1146/annurev.neuro.30.051606.094238

Vinken, K., Prince, J. S., K., T., & Livingstone, M. S. (2023). *The neural code for "face cells" is not face-specific.*

Wakeman, D., & Henson, R. (2015, 01). A multi-subject, multi-modal human neuroimaging dataset. *Scientific Data*. doi: 10.1038/sdata.2015.1

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1403112111

Yovel, G., & Kanwisher, N. (2005, 12). The neural basis of the behavioral face-inversion effect. *Current Biology*. doi: 10.1016/j.cub.2005.10.072