An EEG-fMRI Investigation of the Spatiotemporal Hierarchy of Social Actions

Emalie McMahon (emaliemcmahon@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N Charles St
Baltimore, MD, USA 21218

Elizabeth Jiwon Im (imelizabeth@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N Charles St
Baltimore, MD, USA 21218

Michael F. Bonner (mfbonner@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N Charles St
Baltimore, MD, USA 21218

Leyla Isik (lisik@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N Charles St
Baltimore, MD, USA 21218

Abstract

Recent work has argued that in addition to the dorsal and ventral visual streams, there is a third visual stream projecting laterally from primary visual cortex to the superior temporal sulcus that is specialized for dynamic social content. A key characteristic of the dorsal and ventral streams is hierarchical computations. To investigate whether the lateral visual stream also has hierarchical computations, we combine the spatial precision of fMRI with the temporal precision of EEG to investigate the direction of information flow through lateral regions of the brain. We find evidence of a temporal and spatial dissociation between features computed early and late in both ventral and lateral regions of the brain providing evidence of hierarchical computations in the lateral visual stream and insight into visual processing of dynamic, social scenes.

Keywords: EEG, fMRI, social actions, encoding models

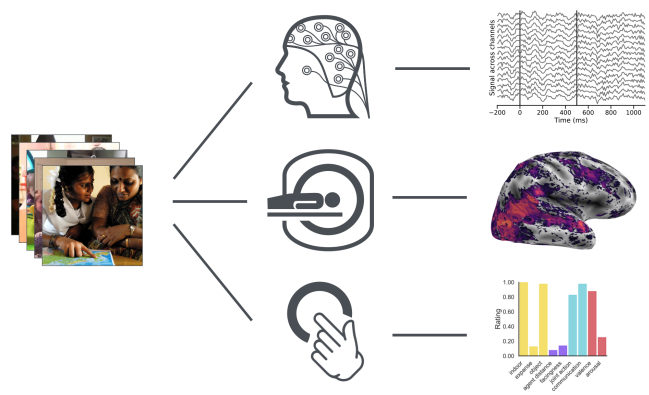

Figure 1: Participants in EEG and fMRI viewed videos of dyadic social actions (500 ms and 3 s, respectively). Participants online rated features of the social visual scene.

Introduction

The ability to understand the actions and interactions of others is an important part of daily life. The social actions of others, such as talking, hugging, or waving goodbye, are priortized in the brain and behavior (Dima, Tomita, Honey, & Isik, 2022; Wurm, Caramazza, & Lingnau, 2017; Tarhan & Konkle, 2020; McMahon & Isik, 2023). The extent to which an action is directed at someone else is an important dimension of action understanding in the lateral occipital temporal cortex (LOTC), a region implicated in action understanding (Dima et al., 2022; Wurm et al., 2017; Tarhan & Konkle, 2020). Further, the superior temporal sulcus (STS), slightly anterior to LOTC, there are regions that selectively respond to two or more people engaged in a social interaction (Isik, Koldewyn, Beeler, & Kanwisher, 2017; Walbrin, Downing, & Koldewyn, 2018). This and related work have led to recent theoretical proposal of a lateral visual pathway that is specialized for recognizing agentic action (Wurm & Caramazza, 2022) or dynamic social perception more generally (Pitcher & Ungerleider, 2021).

Pitcher and Ungerleider (2021) proposed that lateral pathway representations are organized hierarchically from low-level features in early visual cortex (EVC) to high-level features in the STS, analogous to what has been found in ventral stream representations (DiCarlo, Zoccolan, & Rust, 2012). A prior study utilized functional magnetic resonance imaging (fMRI) to reveal that features of social actions are organized hierarchically in an increasingly abstract manner along the anterior-to-posterior axis of the lateral visual stream (McMahon, Bonner, & Isik, 2023). However, due to the poor temporal resolution of fMRI, the direction of information flow through these regions is unknown. To investigate this question, in the current study, we combine the high temporal resolution of EEG and the high spatial resolution of fMRI using a cross-validated encoding method for EEG-fMRI fusion. In particular, we decode visual and social features from the EEG signal and predict voxel-wise fMRI responses using the EEG signal at each time point.

Methods

Participants

Participants (n = 21, 5 Males, M = 21.4 years, SD = 2.9 years) gave informed consent prior to partition in accordance with the Johns Hopkins University Institutional Review Board and were either given course credit or monetary compensation for their time. One participant was excluded from subsequent analyses due to excessive movement.

EEG Experiment

EEG signals were recording using a 64-channel cap. During the experiment, participants viewed 250 short video clips of social actions. The videos are the same as in McMahon et al. shortened from 3 s to the central 500 ms to facilitate time locking to the EEG signal (2023). The training-test split (200 and 50 videos in train and test sets, respectively) from the original fMRI experiment was maintained. Videos in the training set were repeated four times, and videos in the test set were repeated sixteen times.

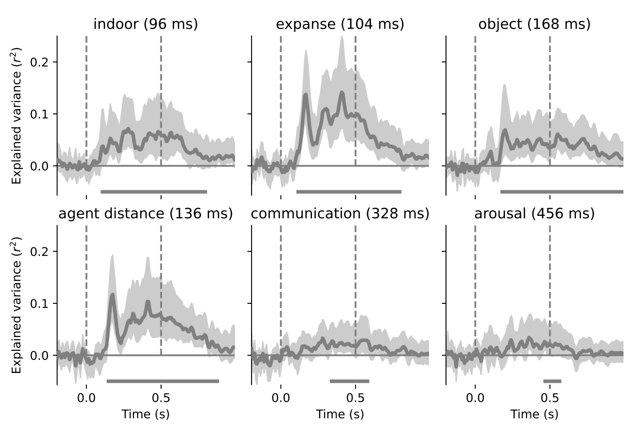

Figure 2: Decodability of select features from the EEG signal across time. Shaded regions are the 95% confidence intervals from bootstrapped variance distributions. Bold horizontal lines indicate significant decoding at the level of p ¡ 0.05. The time in parentheses is the first time point of significant decoding.

Analysis

EEG preprocessing EEG data preprocessing was performed using MATLAB R2023b and FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). The EEG data were aligned to stimulus onset and cut to 1.2 s (0.2 s pre-stimulus to 1 s post-stimulus onset), baseline-corrected using the 0.2 s prior to stimulus onset, and high-pass filtered at 0.1 Hz. Trials with muscle artifacts or high variability were removed. Eye movement artifacts were removed using independent component analysis (ICA). Finally, data were median centered, 30 Hz low-pass filtered, resampled to 250 Hz, and temporally smoothed over five consecutive samples with a step size of 2.

Decoding & Encoding For EEG decoding, at each time point in the EEG signal, we used the 64 channel EEG activity to predict visual and social features using ridge regression. The encoded fMRI data were previously published in McMahon et al. (2023) and are publicly available. Using ridge regression, we modeled the activity within all voxels with reliable signal with the the EEG signal as the predictor for each time point. Across both analyses, the ridge penalty was fit using optimized leave-one-out prediction in sklearn (Pedregosa et al., 2011) in the training set, and evaluation was performed in the test set. Thus, we used a cross-validated metric to compare neural responses across modality.

Performance was determined as the sign-squared correlation of the predicted signal and true signal. We used permutation testing and bootstrap resampling across the predicted values to determine significance and estimate variance of the model performance within each EEG subject. We performed cluster correction to control for multiple comparisons across time.

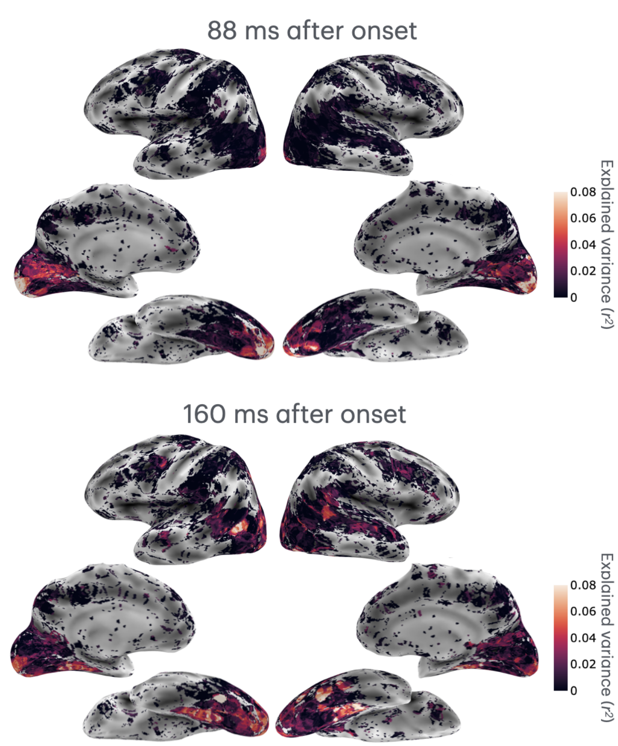

Figure 3: EEG-fMRI fusion results in one representative fMRI subject at two select time points after stimulus onset (88 ms: top and 160 ms: bottom).

Results

Feature Decoding

We find that we were able to decode many features of the visual-social scene from the EEG signal. Following cluster correction: indoor, spatial expanse, object directedness, agent distance, communication, and arousal have significant decoding across participants (Figure 2). We see a progression in the onset timing from early decoding of lower-level features of the scene (indoor, $t_{onset} = 96$ ms and spatial expanse, $t_{onset} = 104$ ms) to late decoding of higher level features (object directedness, $t_{onset} = 168$ ms, agent distance, $t_{onset} = 138$ ms, communication, $t_{onset} = 328$ ms, and arousal, $t_{onset} = 456$ ms) indicating a temporal hierarchy in the representation of social action features consistent with prior fMRI results (McMahon et al., 2023).

fMRI Encoding

At each time point for each EEG participant, we predicted the voxel-wise response in fMRI data from each subject in McMahon et al. (2023). We then averaged the prediction across EEG participants for every reliable voxel in the brain. We found a dissociation between early activation in posterior regions and late activation in more anterior visual regions both in the ventral and lateral regions of the brain (Figure 3). These results are consistent with the dissociation between lower-level and higher-level visual features.

Discussion

Here we investigated the spatiotemporal hierarchy of social actions in the lateral stream of the human brain. Using EEG decoding, we find that features of social actions are computed hierarchically. Using a novel EEG-fMRI fusion method, we also find a temporal dissociation between between posterior regions and anterior visual regions of the brain. Together, these results provide support for a spatiotemporal hierarchy in the lateral stream that mirrors similar findings in ventral temporal cortical responses. In future analyses, we will investigate the shared variance that the EEG signal predicts in the visual-social features and the fMRI responses.

Acknowledgments

This work was funded in part by NSF GRFP DGE-1746891 awarded to E.M. and NIMH R01MH132826 awarded to L.I.

References

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012, February). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. doi: 10.1016/j.neuron.2012.01.010

Dima, D. C., Tomita, T. M., Honey, C. J., & Isik, L. (2022, May). Social-affective features drive human representations of observed actions. *eLife*, *11*, e75027. (Publisher: eLife Sciences Publications, Ltd) doi: 10.7554/eLife.75027

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017, October). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), E9145–E9152. doi: 10.1073/pnas.1714471114

McMahon, E., Bonner, M. F., & Isik, L. (2023, December). Hierarchical organization of social action features along the lateral visual pathway. *Current Biology*, *33*(23), 5035–5047.e8. (Publisher: Elsevier) doi: 10.1016/j.cub.2023.10.015

McMahon, E., & Isik, L. (2023, December). Seeing social interactions. *Trends in Cognitive Sciences*, *27*(12), 1165–1179. doi: 10.1016/j.tics.2023.09.001

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. doi: 10.1155/2011/156869

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pitcher, D., & Ungerleider, L. G. (2021, February). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, *25*(2), 100–110. doi: 10.1016/j.tics.2020.11.006

Tarhan, L., & Konkle, T. (2020). Sociality and interaction envelope organize visual action representations. *Nature Communications*, *11*(1). doi: 10.1038/s41467-020-16846-w

Walbrin, J., Downing, P., & Koldewyn, K. (2018, April). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*, 31–39. doi: 10.1016/j.neuropsychologia.2018.02.023

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends in Cognitive Sciences*, *26*(2), 103–116. doi: https://doi.org/10.1016/j.tics.2021.10.003

Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action Categories in Lateral Occipitotemporal Cortex Are Organized Along Sociality and Transitivity. *Journal of Neuroscience*, *37*(3), 562–575. (Publisher: Society for Neuroscience _eprint: https://www.jneurosci.org/content/37/3/562.full.pdf) doi: 10.1523/JNEUROSCI.1717-16.2016