# Prior Scene Context Modulates the Dynamic Interplay Between Bottom-Up and Top-Down Neural Processes in Face Detection

**Sule Tasliyurt-Celebi (sule.tasliyurt-celebi@psychol.uni-giessen.de)**
Department of Psychology
Justus Liebig University Giessen, Giessen, 35394, Germany

**Daniel Kaiser (danielkaiser.net@gmail.com)**
Department of Mathematics and Computer Science, Physics, Geography
Justus Liebig University Giessen, Giessen, 35392, Germany
Center for Mind, Brain and Behavior
Universities of Marburg, Giessen, and Darmstadt, Marburg, 35032, Germany

**Katharina Dobs (katharina.dobs@uni-giessen.de)**
Department of Mathematics and Computer Science, Physics, Geography,
Justus Liebig University Giessen, 35392, Germany
Center for Mind, Brain and Behavior
Universities of Marburg, Giessen, and Darmstadt, Marburg, 35032, Germany

## Abstract

**Within a fraction of a second, we detect faces in our environment. How is this remarkably fast process implemented in the brain, and is it modulated by top-down mechanisms? Here, we used electroencephalography (EEG) to probe how prior scene context shapes temporal dynamics of neural face representations in natural settings. Participants viewed images of natural scenes containing a single face (on the left or right) that followed either a faceless preview (preview condition) or a gray screen (no-preview condition), while performing a face detection task (~10% foils). Using MVPA decoding, we were able to decode the face location (left vs. right) shortly after target onset. Critically, decoding accuracy of face location was initially higher in the preview condition, while the no-preview condition showed increased accuracy at later processing stages. Moreover, time-frequency analyses showed an enhanced decodability of face location in the preview condition in the alpha band (8–13 Hz), consistent with enhanced spatial orienting. Our findings suggest that prior scene context modulates face detection via distinct neural mechanisms that affect both bottom-up sensory integration and top-down spatial attention, thereby highlighting the dynamic interplay between contextual cues and neural processing.**



**Keywords:** face perception; time-frequency analysis; top-down processing; MVPA; EEG


## Introduction

Humans rapidly and accurately detect faces across diverse scenes (Bindemann & Lewis, 2013). Saccades to faces can occur as early as 100 ms after stimulus onset, faster than to any other object category (Crouzet et al., 2010; Martin et al., 2018). Such rapid performance reflects the efficiency of our visual system, which relies on both bottom-up sensory input and top-down influences. While bottom-up processing is mainly driven by immediate sensory information, top-down processing incorporates expectations, task demands, and prior knowledge. Yet, does the rapid and automatic nature of face detection preclude the influence of top-down processes? Indeed, it has been argued that rapid face detection might be minimally influenced by top-down factors, such as scene context (Crouzet & Thorpe, 2011). In contrast, recent evidence suggests that prior information impacts face perception (Garlichs & Blank, 2024; Mares et al., 2024; Tasliyurt-Celebi et al., 2024). Thus, face detection is not only a fundamental perceptual skill but also an ideal paradigm for examining the dynamic interplay between bottom-up and top-down processes.

Here, we combine MVPA decoding with high-temporal-resolution EEG to probe the neural mechanisms underlying rapid face detection. By modulating the presence of prior scene context, we address two key questions: (i) Does prior scene context modulate neural representations of faces during detection? and (ii) Which neural mechanisms underlie the dynamic interplay between bottom-up and top-down processing in face detection?

## Methods

**Task procedure.** To examine how prior scene context affects the neural mechanisms of face detection, we measured EEG responses during a face detection experiment (N=44). We curated a large set of 784 natural scene images, each featuring a single target face (left or right) and created faceless counterparts (preview) by manually editing out both the face and body. To control for face location, all images were mirrored, and the use of which version as preview versus no-preview was counterbalanced. In each trial (Fig. 1), participants viewed either a scene preview (preview condition) or a gray screen (no-preview condition) for 250 ms before the target scene was presented. The experiment also included 200 foil trials (no face), during which participants pressed a button; these trials were excluded from further analyses.

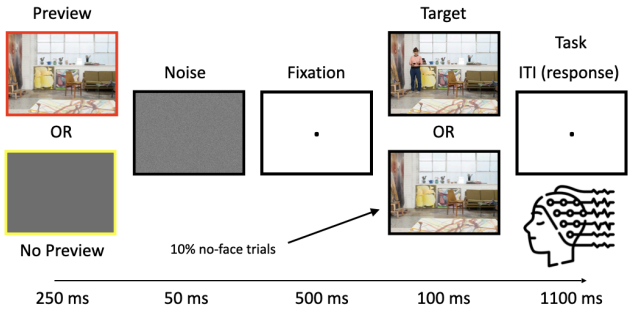


Figure 1: EEG face detection task (N=44).

**EEG data processing.** EEG data were recorded using an Easycap system with 64 channels and a Brain Products amplifier at a sampling rate of 1000 Hz. AFz served as the ground electrode, while Fz was used as the reference. Data were bandpass filtered between 0.1 and 40 Hz and then segmented into epochs ranging from –100 ms to 1100 ms relative to target onset, with baseline correction applied using the between –100 to 0 ms pre-target interval. Independent component analysis (ICA) was conducted to isolate and remove eye movement artifacts through visual inspection.

**Decoding analyses.** We applied MVPA to extract temporal information about face location from the EEG data at the subject level. To increase reliability, we constructed pseudo-trials at each time point, averaging

the data within the same location (left vs. right) and condition (preview vs. no-preview) into 10 folds. For evoked responses, each pattern consisted of the sensor activations for one pseudo-trial and one condition computed using a sliding window of 50 ms width with a 5 ms resolution to enhance the signal-to-noise ratio. In addition, time-frequency analysis was performed using a Morlet wavelet transformation (6-cycle length) to extract activity in the theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz) and gamma (31–100 Hz) bands over the interval –100 to 1100 ms relative to target onset. Frequency-resolved data were then averaged across all frequencies and electrodes. For both analyses, linear support vector machine (SVM) classifiers were trained within each condition using 10-fold cross-validation to decode face location (left vs. right), and decoding accuracy was evaluated against chance level using sign permutation tests (10,000 iterations). All analyses were conducted using the MNE Toolbox in Python.

## Results

We successfully decoded face location from evoked responses shortly after target onset and throughout the trial duration (preview: max. decoding accuracy of 58.3% at 146 ms; no-preview: max. decoding accuracy of 59.5% at 442 ms; Fig. 2A). Crucially, decoding accuracy was higher in the preview condition at early stages (171–292 ms, $d$ = .480), whereas the no-preview condition showed greater accuracy at later stages (392–1040 ms, $d$ = .580). This pattern suggests early facilitation by scene previews, followed by a compensatory processing in the absence of prior context.

Time-frequency decoding revealed higher decoding accuracy in the preview condition in the alpha band (8–13 Hz; 322–553 ms; $d$ = .422; Fig. 2B), with a similar effect in the beta band (14–30 Hz; 397–497 ms), consistent with top-down spatial orienting. In contrast, face location could not be reliably decoded from theta or gamma band activity.

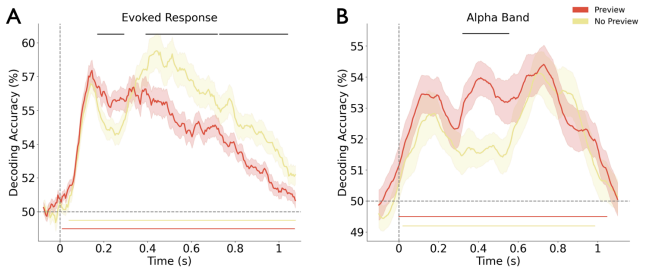


Figure 2: Time-resolved face location decoding accuracy based on evoked responses (**A**) and alpha frequency band (**B**) relative to target onset in the preview (red) and no-preview (yellow) condition. Red yellow horizontal lines: significant clusters, black horizontal lines: significant difference clusters ($p$ < 0.05, sign permutation test). Shaded areas: SEM. Gray dotted horizontal lines: chance level.

## Discussion

Our findings show that prior scene context enhances neural representations of faces at early processing stages—within the first 300 ms after target stimulus onset. This early enhancement in the preview condition suggests that top-down mechanisms, likely mediated by prior expectations and contextual cues (Garlichs & Blank, 2024; Manes et al., 2024), rapidly facilitate face processing. Interestingly, this top-down modulation is also reflected in the alpha and beta band, with enhanced decodability from around 300 to 500 ms. This finding is consistent with studies linking alpha activity to top-down processes (Stecher et al., 2025) and an enhanced spatial orienting effect which follows the earlier enhancement of location decoding for the target face (Battistoni et al., 2020). In contrast, the no-preview condition elicited a stronger face representation at later processing stages, which might reflect an increased reliance on bottom-up sensory integration in the absence of contextual cues.

These findings inform broader questions about the dynamic interplay between bottom-up and top-down processing in visual perception (Peters et al., 2024). The observation that top-down influences are integrated into the rapid feedforward sweep of information processing is consistent with recent computational models where top-down predictions are continuously compared with incoming sensory evidence (e.g., Spratling, 2017).

Furthermore, our results have implications for computational models of face perception. Many state-of-the-art deep neural networks for face perception operate primarily on a feedforward basis (van Dyck and Gruber, 2023; O'Toole and Castillo, 2021), often lacking mechanisms for incorporating context. Our results highlight the potential benefits of integrating recurrent or feedback connections into these computational models (Kar et al., 2019; Kietzmann et al., 2019, Tugsbayar, et al., 2024).

In sum, our findings not only advance our understanding of the neural mechanisms underlying face detection but also emphasize the importance of prior contextual information in shaping perceptual processes, thereby offering powerful constraints on computational models of human visual perception.

## Acknowledgments

This work was supported by the ERC Starting Grant DEEPFUNC (ERC-2023-STG-101117441), the

Hessian Ministry of Higher Education, Research, Science and the Arts (LOEWE Start Professorship and Excellence Program “The Adaptive Mind”), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-project number 222641018-SFB/TRR 135 TP C9.

## References

Battistoni E, Kaiser D, Hickey C, Peelen MV. (2020). The time course of spatial attention during naturalistic visual search. *Cortex, 122,* 225–234. https://doi.org/10.1016/j.cortex.2018.11.018

Bindemann, M., & Lewis, M. B. (2013). *Face detection differs from categorization: Evidence from visual search in natural scenes. Psychonomic Bulletin & Review, 20*(6), 1140–1145. https://doi.org/10.3758/s13423-013-0445-9

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. https://doi.org/10.1167/10.4.16

Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology*, *2*(11). https://doi.org/10.3389/fpsyg.2011.00342

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature neuroscience*, *22*(6), 974–983. https://doi.org/10.1038/s41593-019-0392-5

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K., 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, 116(43), 21854-21863. https://doi.org/10.1073/pnas.1905544116

Garlichs, A., & Blank, H. (2024). Prediction error processing and sharpening of expected information across the face-processing hierarchy. *Nature Communications, 15*(1), 3407. https://doi.org/10.1038/s41467-024-47749-9

Mares, I., Smith, F.W., Goddard, E.J., Keighery, L., Pappasava, M., Ewing, F.W., & Smith, M.L. (2024). Effects of expectation on face perception and its association with expertise. *Scientific Reports, 14*, 9402. https://doi.org/10.1038/s41598-024-59284-0

Martin, J. G., Davis, C. E., Riesenhuber, M., & Thorpe, S. J. (2018). Zapping 500 faces in less than 100 seconds: Evidence for extremely fast and sustained continuous visual search. *Scientific Reports*, *8*(1), 12482. https://doi.org/10.1038/s41598-018-30245-8

O'Toole, A. J., & Castillo, C. D. (2021). Face recognition by humans and machines: three fundamental advances from deep learning. *Annual Review of Vision Science, 7(*1), 543-570. https://doi.org/10.1146/annurev-vision-093019-111701

Peters, B., DiCarlo, J. J., Gureckis, T., Haefner, R., Isik, L., Tenenbaum, J., Konkle, T., Naselaris, T., Stachenfeld, K., Tavares, Z., Tsao, D., Yildirim, I., & Kriegeskorte, N. (2024). How does the primate brain combine generative and discriminative computations in vision?. *ArXiv*. https://doi.org/10.48550/arXiv.2401.06005

Spratling, M. W. (2017). A hierarchical predictive coding model of object recognition in natural images. *Cognitive Computation*, *9*(2), 151-167. https://doi.org/10.1007/s12559-016-9445-1

Stecher R, Cichy RM, Kaiser D. (2025). Decoding the rhythmic representation and communication of visual contents. *Trends in Neuroscience, 48*(3), 178-188. https://doi.org/10.1016/j.tins.2024.12.005

Tasliyurt-Celebi, S., de Haas, B., Võ, M. L.-H., & Dobs,K. (2024). Using CNNs to understand how bottom-up and top-down processes shape human face detection. *Cognitive Computational Neuroscience (CCN).* Boston, USA.

Tugsbayar, M., Li, M., Muller, E. B., & Richards, B. (2024). Top-down feedback matters: Functional impact of brainlike connectivity motifs on audiovisual integration. bioRxiv, 2024-10. https://doi.org/10.1101/2024.10.01.615270

van Dyck, L. E., & Gruber, W. R. (2023). Modeling biological face recognition with deep convolutional neural networks. *Journal of Cognitive Neuroscience*, *35*(10), 1521-1537. https://doi.org/10.1162/jocn_a_02040