Saccade onset, not fixation onset, best explains early sensory responses across the human visual cortex during naturalistic vision

Carmen Amme [1,a,✉], **Philip Sulewski** [1,2,a], **Eelke Spaak** [3], **Martin N. Hebart** [2,4,5], **Peter König** [1,6,a], **and Tim C. Kietzmann** [1,a]

[a]P.S. and C.A., as well as P.K., and T.C.K., contributed equally to this work.; [1]Institute of Cognitive Science, Osnabrück University, 49090 Osnabrück, Germany; [2]Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany; [3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 HR Nijmegen, The Netherlands; [4]Department of Medicine, Justus Liebig University Giessen, 35392 Giessen, Germany; [5]Center for Mind, Brain and Behavior, Universities of Marburg, Giessen, and Darmstadt, 35032 Marburg, Germany; [6]Department of Neurophysiology and Pathophysiology, Center of Experimental Medicine, University Medical Center Hamburg-Eppendorf, 20251 Hamburg, Germany

Visual processing has traditionally been investigated using static viewing paradigms, where participants are presented with streams of randomized stimuli. Observations from such experiments have been generalized to naturalistic vision, which is based on active sampling via eye movements. In studies of naturalistic vision, visual processing stages are thought to be initiated at the onset of fixations, equivalent to a stimulus onset. Here we test whether findings from static visual paradigms translate to active, naturalistic vision. Utilizing head-stabilized magnetoencephalography (MEG) and eye tracking data of 5 participants who freely explored thousands of natural images, we show that saccade onset, not fixation onset, explains most variance in latency and amplitude of the early sensory component M100. Source-projected MEG topographies of image and saccade onset were anticorrelated, demonstrating neural dynamics that share similar topographies but produce oppositely oriented fields. Our findings challenge the prevailing approach for studying natural vision and highlight the role of internally generated signals in the dynamics of sensory processing.

active vision | eye movements | natural scenes | M100 / P100 | MEG
Correspondence: ***carmenamme@outlook.de***

Introduction

To explore our environment, we sample visual information by moving our eyes across the visual field multiple times each second. Visual processing is, however, often examined using static paradigms where participants maintain central fixation while unpredictable stimuli are presented, and brain activity is investigated in relation to stimulus onset. In contrast, naturalistic paradigms allow free eye movements, leading to sequences of saccades and fixations. Fixation onset is thereby used to analyze this data, similar to stimulus onset [1–3]. However, due to the self-initiated nature of eye movements, it is conceivable that the brain utilizes knowledge about the generated saccade and, therefore, processing is effectively started at saccade onset [4].

To investigate the dynamics of naturalistic visual processing, we collected magnetoencephalography (MEG) and eye tracking data from 5 participants who visually explored 4,080 natural scenes, yielding data for around 200,000 saccade and fixation events. Our analysis shows that the M100, closely related to the P100, one of the earliest and most studied visual components in neuroscience, is better explained as being initiated already at the onset of the saccade rather than the fixation onset. Moreover, analyses of source-projected data indicate significant differences between neural sources involved in saccade and fixation onsets compared to scene onsets, indicating qualitative differences between these events. This evidence suggests that classic paradigms may not align with natural viewing conditions, highlighting new research avenues related to natural vision.

Results

M100 latency shifts with saccade duration

For each participant, we selected the sensor with the highest amplitude in the traditional fixation-locked response within 60 and 110 ms after event onset and computed the median-based event-related fields (ERFs) across 160 equally-sized bins based on the preceding saccade duration (Fig. 1A). The M100 deflections were better aligned with the preceding saccade onset than the fixation onset (Fig. 1A). The same observation held when selecting sensors based on the saccade onset-locked ERF (Fig. 1B). This observation was further supported by a negative correlation between the M100 latency and the saccade duration (Fig. 1C). A direct comparison between fixation and saccade onset ERFs for different saccade durations yielded much-improved alignment of saccade-locked ERFs (Fig. 1D, E).

A parametric search for the optimal event latency aligns with saccade onset rather than fixation onset

To identify the time point that best aligns the M100s for different saccade durations, we aligned the data based on artificial events t_α, with $t_\alpha = t_{fix} + \alpha * (t_{fix} - t_{sac})$. To obtain these artificial events, we interpolated

event onsets between saccade and fixation onset, corresponding to α = -1 and α = 0, respectively. Additionally, we extrapolated between fixation onset (α = 0) and an artificial event after fixation onset at twice the saccade duration (α = 2; same for saccade onset, where we extrapolated between an artificial event at twice the saccade duration before saccade onset (α = -2), and saccade onset (α = -1)). The M100 explained most variance when closely locked to saccade onset (Fig. 1F; participant 4: $\overline{\alpha}$ = -.87, 95% CI = [-.92, -.83]; across participants: $\overline{\alpha}$ = -.92, 95% CI = [-.95, -.87]), a finding consistent across most gradiometers (Fig. 1G). The saccade onset-locked ERF was observable across most sensors and participants (Fig. 1H, I).

Small, separate contribution of fixation onset to the visual response

While saccade onset better explained the M100 and subsequent components, we further investigated whether fixation onset contributed to the visual response. By subtracting saccade-locked ERFs, we obtained fixation-triggered ERFs from the residuals. All participants exhibited a non-zero residual fixation onset-locked component (Fig. 1J). Therefore, although the majority of cortical processing observed in MEG recordings was time-locked to saccade onset, a minor fraction associated with fixation onset remained.

Similar latency and amplitude, but opposite polarities, for saccade ERF and stimulus onset ERF

We compared the traditional stimulus onset ERF to the saccade ERFs to investigate their relationship. To account for different saccade durations, we extrapolated the data to a hypothetical saccade ERF with a duration of zero ms (Fig. 2A). The amplitude and peak latency of the zero-duration saccade ERF were similar to those of the scene onset ERF, with only slight differences in peak latencies (Fig. 2B). However, a notable difference was the opposite polarities between scene onset and zero-duration saccade ERF, indicating qualitative differences between these events and their underlying physiological process. A similar observation could be made when comparing the residual fixation ERF's first (positive) deflection to the scene onset ERF.

Negative relationship between source projections of saccade ERF and stimulus onset ERF

To move beyond analyses of individual sensors, we applied Dynamic Statistical Parametric Mapping (dSPM) to source reconstruct the saccade, fixation and scene onset ERFs. In source space, we projected the activations onto the cortical surface normal and averaged the data between 90-100 ms after event onset. Visual inspection (Fig. 2C, D, E) and quantitative analyses (Fig. 2F, G) revealed a positive relationship across cortical vertices between a hypothetical saccade with a duration of zero and residual fixation events in early, ventral, lateral, parietal, and posterior cingulate areas (grand average: all $m_{saccade,fixation}$ > .24, p < .001, R^2 > .07). In contrast, both residual fixation and saccade-based source activations were negatively related to scene onset source activations (grand average: all $m_{fixation,scene}$ < -.50, p < .001, R^2 > .24, and all $m_{saccade,scene}$ < -.07, p < .05, R^2 > .01, respectively). Thus, this systematic sign flip indicates qualitatively differences between saccade/fixation onset responses and scene onset responses.

Discussion

Contrary to the focus on fixation onset as an equivalent to stimulus onset of many studies, our findings indicate that saccade onsets are the critical events for understanding MEG signatures of natural visual processing. We found that the majority of the M100 amplitude was locked to saccade onset rather than fixation onset, highlighting the importance of internally-generated processes during natural vision, initiated before fixation onset. These results align with previous studies showing latency modulation of fixation-based neural responses based on saccade duration from EEG and iEEG recordings [5–7]. Similar conclusions were drawn from local field potentials in area V1 of macaques during free eye movements [8]. Jointly, these data argue for a paradigm shift towards an action-based, i.e., saccade-initiated, perspective on sensory processing.

Our results demonstrate that neural responses to typical stimulus onsets differ systematically from gaze events in naturalistic vision. Both saccade- and fixation-based components were negatively correlated with stimulus onset responses, suggesting that common analyses in response to stimulus onsets may only partially transfer to natural vision. The key distinction is that eye movement-triggered components are planned and self-initiated, based on predictions and prior exploration of peripheral contextual cues. These components may stem from internally generated signals, relating to predictions of incoming visual information in the fashion of the predictive coding framework [9–11], contextual 'priming' [12, 13] and memory [14] guided by top-down information flow, or a readiness potential initiated at saccade onset through corollary discharge to prepare the visual areas for the upcoming visual information [8, 15]. Future work is needed to investigate these or other possibilities.

Taken together, our results imply that neural processing in natural vision is guided by an internally generated signal, whereby self-initiated actions are a key to sensory processing. Studying active visual paradigms, both in terms of experimental data and computational modeling, is hence needed for understanding visual information processing as it occurs naturally in everyday settings.

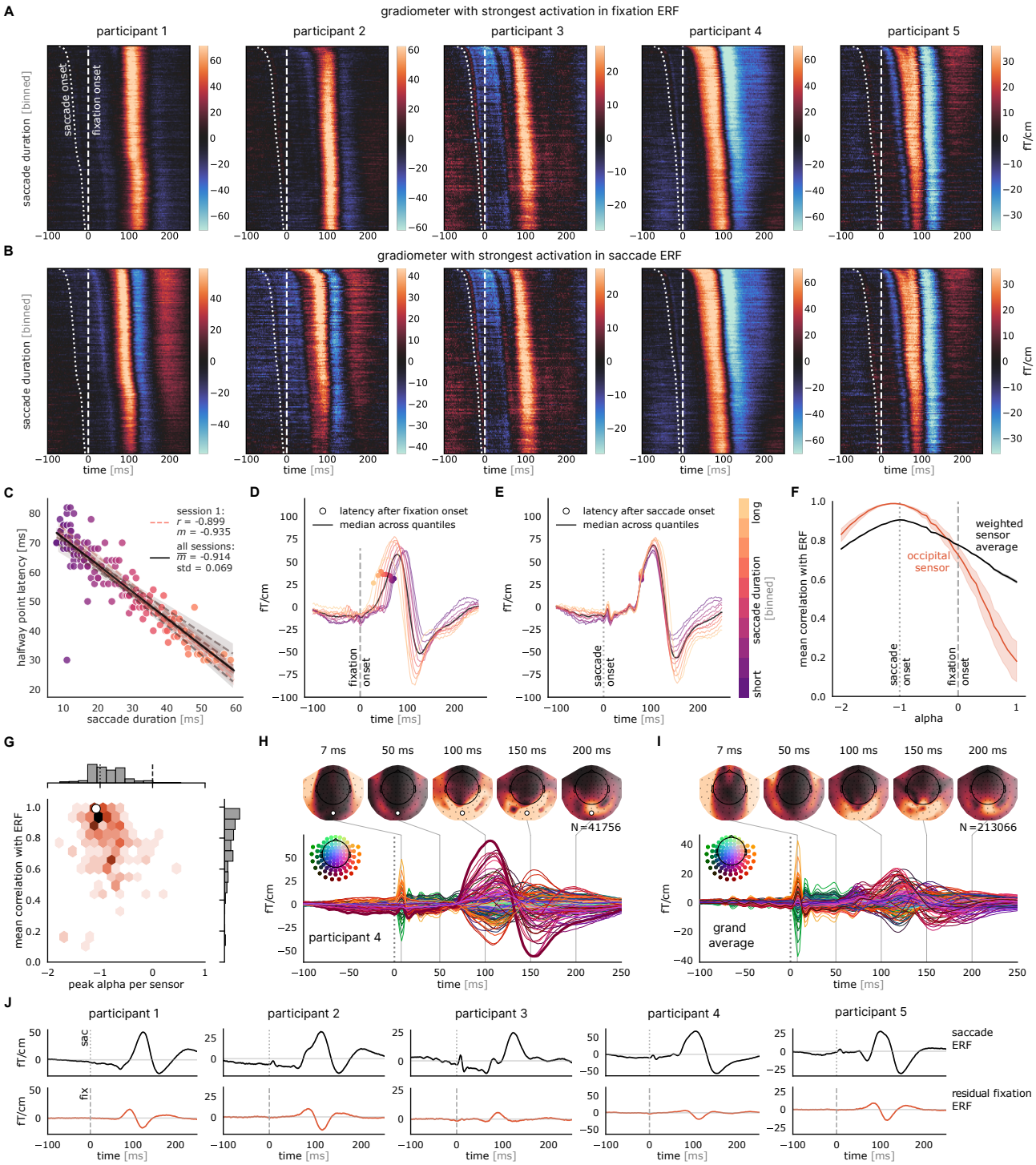

Figure 1. Overview of saccade and fixation onset-locked ERFs, highlighting the saccade-locking of the M100. *(A, B)* The saccade duration binned median-based ERF deflections followed the onset of saccades (dotted line) rather than the onset of fixations (dashed line). Each heatmap displays the data for the sensor with the highest amplitude based on the fixation (A) and saccade onset-locked ERF (B). *(C)* Session-wise linear fit for the previously selected sensor for participant 4 between the duration of the preceding saccade and the latency of the fixation onset-locked ERF per saccade duration bin. The ERF latency is defined as the halfway point of the slope leading to the M100 (see dots in D and E). *(D, E)* The fixation (D) and saccade (E) onset-locked ERF per saccade duration bin (n_{bins} = 10 for visualization). *(F, G)* The mixing factor analysis inter- and extrapolates event onsets around fixation (α = 0) and saccade onset (α = -1). For each event onset, the 160 binned saccade duration ERFs were correlated with the grand average ERF of that event onset. The sensor average is weighted based on maximum correlation strength (F; with 95% CI). A distribution of peak α values and mean ERF correlation for each sensor (G). *(H)* The saccade onset-locked ERFs for all sensors for participant 4, with the in (A) selected sensor in bold. *(I)* The grand average across all participants of the saccade onset-locked ERF. *(J)* Saccade onset-locked ERF for each participant and selected sensor (upper row). This ERF was subtracted from each epoch, and the epochs were realigned to fixation onset. The resulting median-based residual fixation ERF was obtained (lower row; with 95% CI).

Methods

Five participants (right-handed; m_{age} = 27.8) visually explored 4,080 natural scenes from the Natural Scenes Dataset [16] in 10 measurement sessions each while recording MEG (Megin VectorView MEG system; 102 magnetometers, 204 planar gradiometers), and eye-

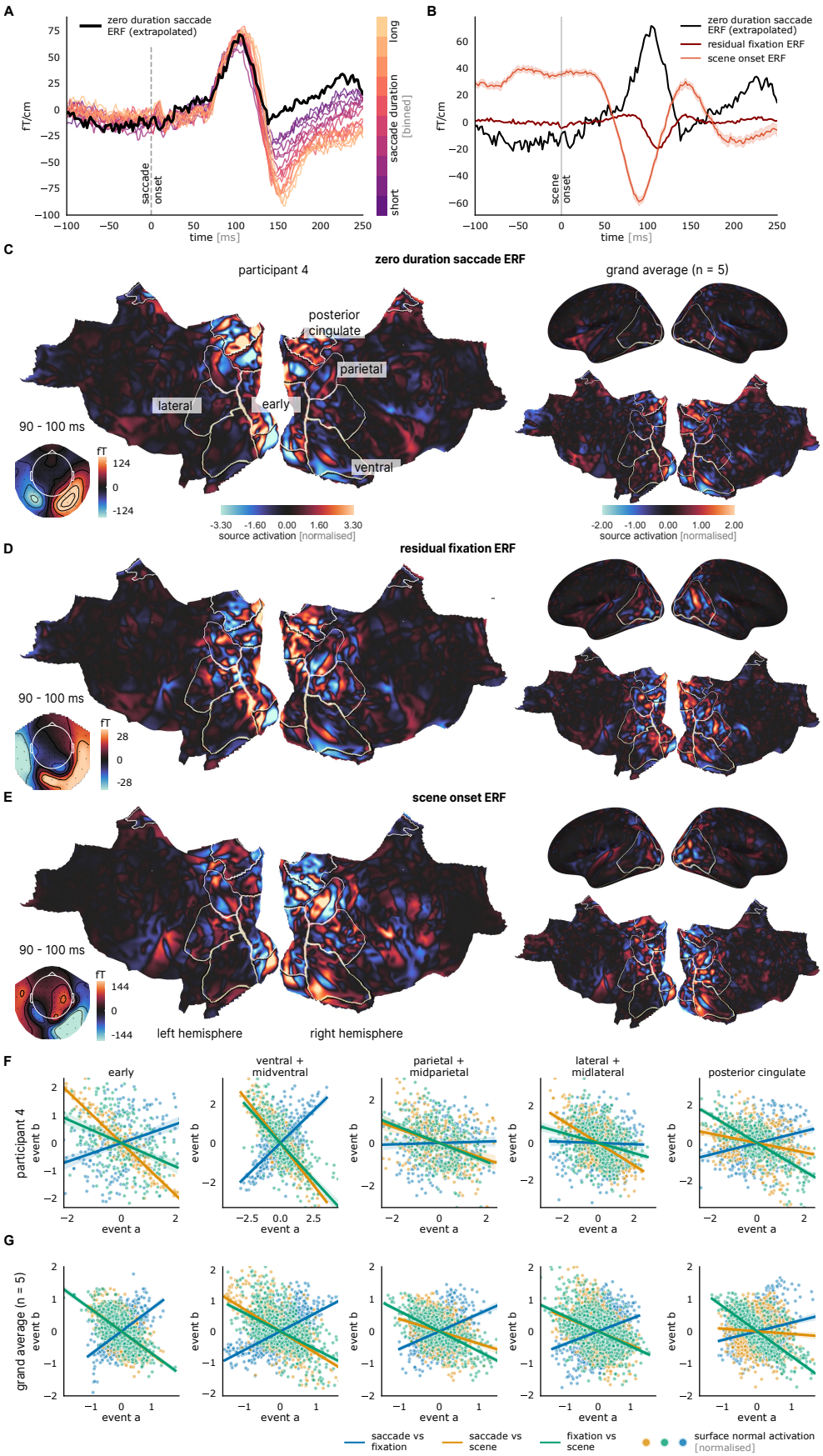

Figure 2. Topographical differences between stimulus onset and fixation/saccade ERFs. *(A)* Extrapolation of an ERF of a saccade with a duration of zero ms based on 160 binned saccade duration ERFs. Here: participant 4 and the earlier selected sensor (see Fig. 1A, B) *(B)* The zero-duration saccade ERF (see A) is contrasted to the stimulus onset ERF and the residual fixation onset ERF (see Fig. 1J; participant 4). *(C, D, E)* Flattened cortical sheets with the surface normal projections of the source reconstructed topographies for the three different ERFs. Lower left: Magnetometer topographies for each ERF. Right: The grand average (n = 5) on the inflated and flattened cortical sheet. *(F)* Comparison of the surface normal projections for each vertex between three events for early visual, ventral, lateral, parietal, and posterior cingulate areas with regression fits (with 95% CI). *(G)* Same as F but based on the averaged (n = 5) surface normal activation.

tracking (SR Research EyeLink 1000). The measurements were supported by individually fitted head stabilizers. Standard preprocessing steps were applied to the MEG data, including tSSS (MaxFilter™; Elekta Oy, Finland), interpolation of bad channels, resampling, filtering (0.2 to 200 Hz), and applying independent component analysis (see Supporting Information). Saccades and fixations events were identified using [17]. The MEG data was then epoched based on those events to obtain ERFs.

Acknowledgements

C.A., P.S., and T.C.K. are supported by the ERC Starting Grant TIME (101039524). P.S. is supported by the Federal Ministry of Education and Research (BMBF) and the Max Planck Society (MPG), and Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) - GRK 2340. M.N.H. is supported by a research group grant by the Max Planck Society, the ERC Starting Grant COREDIM (101039712), the Hessian Ministry of Higher Education, Science, Research and Art (LOEWE Start Professorship to M.N.H. and Excellence Program 'The Adaptive Mind'). We thank Burkhard Maess, Yvonne Wolff-Rosier and Olaf Hauck for support and advice on all the MEG-related aspects of this projects.

Author Contributions

Conceptualization: C.A., P.S., E.S., P.K., and T.C.K.; Formal Analysis: C.A. and P.S.; Investigation: C.A. and P.S.; Resources: M.N.H. and T.C.K.; Writing - Original Draft: C.A., P.S., P.K., and T.C.K.; Writing - Review & Editing: E.S. and M.N.H.; Funding Acquisition: M.N.H., P.K., and T.C.K.

The authors declare no competing interest.

Bibliography

1. Jack L. Gallant, Charles E. Connor, and David C. Van Essen. Neural activity in areas v1, v2 and v4 during free viewing of natural scenes compared to controlled viewing. *NeuroReport*, 9(7):1673, 1998.
2. Daniel Kaiser, Nikolaas N. Oosterhof, and Marius V. Peelen. The neural dynamics of attentional selection in natural scenes. *Journal of Neuroscience*, 36(41):10522–10528, 2016.
3. William E. Vinje and Jack L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276, 2000.
4. E. Spaak. *On the role of alpha oscillations in structuring neural information processing.* 2015. Dissertation (Radboud Universiteit Nijmegen, NL).
5. Chaim N Katz, Kramay Patel, Omid Talakoub, David Groppe, Kari Hoffman, and Taufik A Valiante. Differential generation of saccade, fixation, and image-onset event-related potentials in the human mesial temporal lobe. *Cerebral Cortex*, 30(10):5502–5516, 2020.
6. Koji Kazai and Akihiro Yagi. Comparison between the lambda response of eye-fixation-related potentials and the p100 component of pattern-reversal visual evoked potentials. *Cognitive, Affective, & Behavioral Neuroscience*, 3(1):46–56, 2003.
7. G. W. Thickbroom, W. Knezevič, W. M. Carroll, and F. L. Mastaglia. Saccade onset and offset lambda waves: relation to pattern movement visually evoked potentials. *Brain Research*, 551(1):150–156, 1991.
8. J. Ito, P. Maldonado, W. Singer, and S. Grun. Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. *Cerebral Cortex*, 21(11):2482–2497, 2011.
9. Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013.
10. Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.
11. M. W. Spratling. A review of predictive coding algorithms. *Brain and Cognition*, 112:92–97, 2017.
12. Seppo P. Ahlfors, Stephanie R. Jones, Jyrki Ahveninen, Matti S. Hämäläinen, John W. Belliveau, and Moshe Bar. Direction of magnetoencephalography sources associated with feedback and feedforward contributions in a visual object recognition task. *Neuroscience Letters*, 585:149–154, 2015.
13. M. Bar, K. S. Kassam, A. S. Ghuman, J. Boshyan, A. M. Schmid, A. M. Dale, M. S. Hämäläinen, K. Marinkovic, D. L. Schacter, B. R. Rosen, and E. Halgren. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2):449–454, 2006.
14. Miriam L. R. Meister and Elizabeth A. Buffalo. Getting directions from the hippocampus: The neural connection between looking and memory. *Neurobiology of Learning and Memory*, 134:135–144, 2016.
15. Anna Mayer, Caspar M. Schwiedrzik, Michael Wibral, Wolf Singer, and Lucia Melloni. Expecting to see a letter: Alpha oscillations as carriers of top-down sensory predictions. *Cerebral Cortex*, 26(7):3146–3160, 2016.
16. Emily J. Allen, Ghislain St-Yves, Yihan Wu, Jesse L. Breedlove, Jacob S. Prince, Logan T. Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, J. Benjamin Hutchinson, Thomas Naselaris, and Kendrick Kay. A massive 7t fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, 25(1):116–126, 2022.
17. Ralf Engbert and Konstantin Mergenthaler. Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences*, 103(18):7192–7197, 2006.