J Neurophysiol 136: 414–422, 2026.
First published June 27, 2026; doi:10.1152/jn.00043.2026

american physiological society

JNP JOURNAL OF NEUROPHYSIOLOGY

SHORT REPORT

Abrupt scene onsets and gradually emerging scene information produce distinct EEG decoding dynamics

Ilker Duymaz,[1,2] Micha Engeser,[1,2] and Daniel Kaiser[1,2,3,4]

[1]*Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Giessen, Germany;* [2]*Center for Mind, Brain and Behavior, Universities of Giessen, Marburg and Darmstadt, Germany;* [3]*Center for Applied Computer Science and Data Science, Justus Liebig University Giessen, Germany; and* [4]*Cluster of Excellence "The Adaptive Mind", Universities of Giessen, Marburg and Darmstadt, Germany*

Abstract

Multivariate analyses of magneto-/electroencephalography (M/EEG) data are typically performed on neural responses time-locked to discrete stimulus onsets. Such designs usually reveal high decoding performance during the initial transient response (0–500 ms), which subsequently drops to a lower, sustained level. Here, we examined time-resolved EEG decoding of natural scene processing when scenes gradually enter the visual field without a clear onset. We created video sequences in which one scene category (e.g., a beach) smoothly transitioned into another category (e.g., a forest) by blending two scenes into a single composite panorama and moving a square aperture across it. We compared EEG decoding for the first scenes within the transitions, which appeared with a sudden, artificial onset, to the second scenes, which emerged naturalistically as the videos progressed. For the first scenes, we observed robust category decoding from 60 ms after onset with a clear peak structure. For the second scene, category decoding was markedly weaker and showed no discernible peak structure. Realigning the appearance of category-diagnostic content for the second scene using deep neural networks did not enhance decoding or recover a peak structure. Furthermore, classifiers trained on the first scene generalized to the second, but with a broad, temporally diffuse pattern, instead of a diagonal pattern more consistent with a shared hierarchical processing timeline. Together, these findings demonstrate that time-resolved EEG decoding is sensitive to stimulus-presentation context. Accordingly, temporal decoding patterns obtained in conventional trial-based paradigms may not generalize unchanged to conditions involving gradual scene transitions and continuous visual input.

NEW & NOTEWORTHY Using multivariate EEG decoding, we show that scene-category information follows different temporal profiles across two presentation regimes: abrupt scene appearance after a grayscale screen and gradual scene emergence during continuous visual input. Our findings demonstrate that time-resolved decoding is sensitive to stimulus-presentation context and should not be interpreted as an invariant signature of a fixed visual processing hierarchy.

EEG decoding; multivariate pattern analysis; scene perception; temporal processing

INTRODUCTION

During the past decade, multivariate decoding of magneto-/electroencephalographic (M/EEG) signals has become an essential tool for probing the temporal dynamics of visual representations in the human brain (1, 2). M/EEG decoding studies, by and large, employ trial-based designs that treat stimuli as discrete, independent events, separated by blank interstimulus intervals. Stimuli are thus presented with abrupt, artificial onsets, where a blank screen is suddenly replaced by the stimulus. When time-locked to these sharp onsets, M/EEG signals typically yield robust decoding shortly after the stimulus onset, with a consistent discernible peak structure within the first 500 ms, followed by a sustained but gradually diminishing trend toward chance-level accuracy even when the stimulus remains visible and consciously perceived (1, 3–6).

Many decoding studies draw their primary conclusions from the relative timing of decoding within the first 500 ms of processing, interpreting it as a window into perceptual encoding and category-specific representation (7, 8). For instance, Cichy et al. (4) contrasted superordinate and subordinate categories, interpreting differences in the timing of onset-locked early peaks to chart hierarchical processing of objects. Likewise,

Correspondence: I. Duymaz (ilker.duymaz@uni-giessen.de).
Submitted 27 January 2026 / Revised 5 May 2026 / Accepted 21 June 2026

Kaiser et al. (9) varied object position and observed a robust, onset-locked advantage for typical locations at ~140 ms, demonstrating that object processing is sensitive to the statistical structure of the environment even at early stages of visual processing. These and many other studies rely on abrupt stimulus onsets and report strictly onset-locked processing latencies. It remains unclear to what extent the nature of these conclusions depends on this specific design choice.

The visual system is highly sensitive to onsets, exhibiting greater detection accuracy (10) and enhanced event-related potential (ERP) components (11) compared with other visual events such as visual offsets or spontaneous changes in the luminance of already visible stimuli. This advantage has been attributed to the high salience of onsets for attentional capture (12, 13) and to neural mechanisms tuned to transient, flicker-like cues (14, 15). From an evolutionary standpoint, this sensitivity may serve the quick detection of predators or environmental hazards that threaten survival. However, in everyday life, we typically encounter gradual transitions, such as when scanning a scene, moving through space, or as objects move through our visual field. As a result, category-diagnostic visual information unfolds gradually in naturalistic settings, rather than appearing all at once as it does with abrupt stimulus onsets. This discrepancy raises a central question: to what extent are the temporal decoding profiles reported in conventional, onset-locked paradigms a consequence of the abrupt artificial onsets that elicit them? And how might those temporal dynamics differ when visual information emerges gradually, as it often does in naturalistic viewing?

To investigate this question, we developed a paradigm in which one natural scene gradually transitioned into another without a discrete visual onset. To create these transitions, we used generative AI to seamlessly blend images from different scene categories into composite panoramas. We then generated video stimuli by panning a square aperture smoothly across these panoramas, such that the visible content gradually shifted from one scene to the next (Fig. 1*A*). We recorded EEG while participants viewed these videos and compared the resulting neural representations with those elicited by scenes presented abruptly following a grayscale screen (Fig. 1*C*). Thus, the two conditions contrasted what we refer to as a naturalistic onset, in which scene information emerged gradually during continuous visual stimulation, with an artificial onset, in which a scene appeared abruptly in the absence of preceding scene content. Because these conditions differed in both onset dynamics and the presence of preexisting visual information, the design did not isolate the contribution of either factor. Instead, it allowed us to compare neural scene representations across two distinct presentation regimes: one involving continuous visual change and the other involving abrupt scene appearance.

We found that these presentation regimes were associated with distinct temporal decoding profiles. Scenes presented with artificial onsets produced a sharp, transient peak in decoding accuracy, whereas scenes presented with naturalistic onsets showed a slower rise to a sustained and markedly lower level of accuracy. Representations observed around the transient decoding peak generalized to representations in the naturalistic-onset condition across a broad and temporally nonspecific range. This pattern suggests that the two conditions did not exhibit closely matched temporal decoding dynamics, while nevertheless sharing a broadly similar underlying representational format. Future studies incorporating additional control conditions will be needed to determine the relative contributions of onset dynamics, preexisting visual stimulation, and other factors such as representational competition, adaptation, memory demands, and attentional capture to the observed differences in temporal decoding profiles. Taken together, these findings demonstrate that temporally resolved EEG decoding is sensitive to presentation context. This sensitivity should be

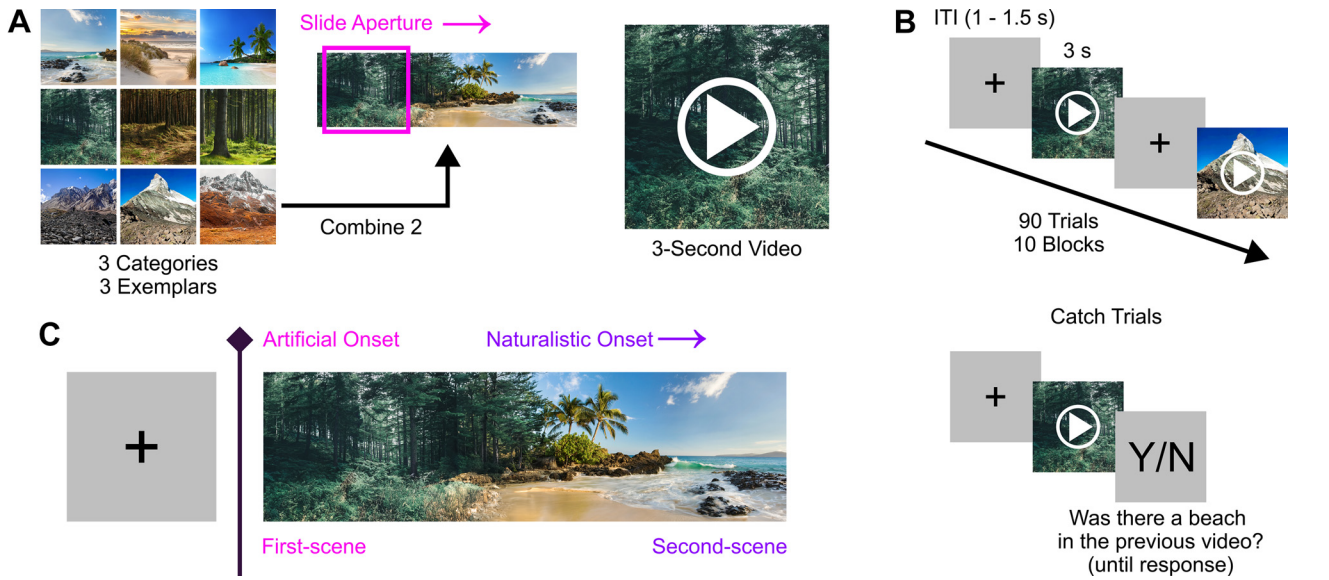

Figure 1. Experimental design. *A*: scene transition videos were created by combining two scenes into a wide panorama and moving a square aperture across it. *B*: trial sequence. Each trial began with a 1.5–2 s randomly jittered intertrial interval, followed by a 3-s transition video, after which the task advanced automatically to the next trial. On a random 20% of trials, a catch prompt appeared after the video, asking whether a specified scene category had been present in the preceding video. *C*: since each transition video was preceded by a blank screen, the first scene appeared with an abrupt artificial onset, whereas the second scene gradually entered view as the video progressed.

taken into account when designing experiments and interpreting differences between conventional trial-based paradigms and paradigms involving continuous, dynamically changing stimulation, particularly as the field increasingly adopts the latter.

METHOD

Participants

Thirty-five healthy adults (19 females, 16 males; mean age: 26.9 yr, SD = 5.09 yr) with normal or corrected-to-normal vision participated in the experiment. Participants provided written informed consent prior to the experiment session and received monetary compensation. The study protocol was approved by the general ethical committee of Justus Liebig University Giessen. All experimental protocols were in accordance with the Declaration of Helsinki.

Stimuli

The stimuli comprised videos depicting natural scene transitions between three categories: beach, forest, and mountain. For each category, we selected three distinct exemplar images, yielding a total of nine base images. We created composite panoramic scenes by combining all possible pairs of exemplars on a 3,500 × 1,000 pixel canvas. In each composite, we placed different scene exemplars on the left and right edges of the canvas. We filled the central (initially empty) region using Adobe Photoshop's (version 25) generative fill tools to blend the two edge images into a naturalistic transition. When needed, we provided prompts to guide the generation, removing artifacts and ensuring physically plausible transitions. We then converted each panoramic image into a 3-s video stimulus (60 frames per second), in which a square aperture smoothly panned from the left to the right edge, simulating a naturalistic scene transition (Fig. 1*A*).

We generated 81 unique videos based on all pairwise combinations of the nine exemplars (9×9). Each condition appeared 10 times, except for the nine conditions in which the first and second scenes were identical, which were presented 20 times to balance overall scene appearance. This yielded a total of 900 total trials per participant.

Procedure

We recorded 64-channel EEG while participants viewed the scene transition videos on a 32-in. Gigabyte AORUS FI32Q monitor (resolution: 2,560 × 1,440 pixels; refresh rate: 120 Hz). Participants sat 60 cm from the screen using a chinrest to maintain consistent viewing distance and minimize head movement. We instructed participants to maintain their gaze on a fixation cross in the center of each video.

Trials began with a fixation cross presented on a gray background for a jittered duration between 1 and 1.5 s (in 100-ms increments), followed by the 3-s video subtending $10^\circ \times 10^\circ$ of visual angle on the same gray background (Fig. 1*B*). On 20% of trials, a response prompt appeared immediately after the video, requiring a yes/no response regarding the presence of a specified scene category (e.g., "Was there a mountain in the previous video?") with equal likelihood for each answer. The prompt remained on screen until response, and no feedback was provided. Participants were performing very well on this task (mean accuracy = 88.5%, SD = 11.7%). On the remaining 80% of trials, the next trial began immediately after the video. Trials were organized into 10 experimental blocks, with participants taking self-paced breaks between blocks.

EEG Acquisition and Preprocessing

We recorded EEG using a 64-channel ActiCAP system with an ActiCHamp amplifier (Brain Products GmbH, Gilching, Germany) at a 1,000 Hz sampling rate. Electrodes were positioned according to the international 10-10 system. Preprocessing and subsequent analyses were performed using MNE-Python (16). Data were epoched from −1 to 3 s relative to stimulus onset and band-stop filtered at 48–52 Hz to remove line noise. We then re-referenced the data to the average across all electrodes, downsampled it to 200 Hz, and applied baseline correction using the −0.5 to 0 s prestimulus interval. Noisy channels were identified and removed via visual inspection. Blink-related artifacts were removed using independent component analysis (ICA).

Time-Resolved Decoding

To examine the temporal dynamics of visual representations during artificial versus naturalistic scene onsets, we trained classifiers at each time point to decode labels corresponding to the first and second scene categories (forest, beach, and mountain) in each transition. Separate classifiers were trained to decode the first- and second-scene category labels from the same time-series data, allowing us to track the evolving representational content in EEG signals as visual stimuli changed over time. In Table 1, we provide an overview of the different decoding analyses we performed. These analyses are described in detail in the following sections.

We excluded trials in which the first and second scenes were the same exemplar. Although these conditions were originally included as potential baseline trials, we did not use them in the decoding analyses because decoding would produce identical results for the two scene labels, making first-scene and second-scene representations impossible to dissociate.

Classification was performed using linear discriminant analysis (LDA) separately at each time point, allowing us to track changes in decoding accuracy across time. We decoded among three scene categories, separately for first-scene labels and second-scene labels, using 10-fold cross-validation. Folds were balanced at both the category level and the individual video (scene transition) level by splitting trials by repetition: each of the 10 repetitions of each unique scene transition served once as the test set while the remaining repetitions served as the training set. Because the stimulus set was balanced with respect to category exemplars, each fold contained the same number of trials per category as well as the same number of repetitions for each individual scene transition. Classification performance was quantified as the mean accuracy across folds (chance = 33.3%).

To identify time points at which classification performance was significantly above chance, we used a cluster-based permutation test with threshold-free cluster enhancement (TFCE; 17) as implemented in MNE-Python, generating the null distribution via 10,000 sign-flip permutations (18).

Table 1. *Overview of decoding analyses*

Analysis	Train Labels	Test Labels	Interpretation
Time-resolved decoding (Fig. 2*A*)			
First-scene decoding	1st scene	1st scene	Time course of representations with artificial onsets.
Second-scene decoding	2nd scene	2nd scene	Time course of representations with gradual, naturalistic emergence.
Temporal cross-generalization (Fig. 2*B*)			
First-to-second*	1st scene	2nd scene	Shared representational format across time.
Second-to-first*	2nd scene	1st scene	Same but in the reverse generalization direction.
Within-scene temporal generalization (Fig. 2*C*)			
First-scene	1st scene	1st scene	Temporal stability of first-scene representations.
Second-scene	2nd scene	2nd scene	Temporal stability of second-scene representations.

Each stimulus comprised a transition between two scene categories, such that the same trial-wise EEG time series could be labeled according to either the first-scene or second-scene category. The interpretation of each analysis depended on which label set was used for classifier training and testing. *Results from these analyses were averaged to obtain a bidirectional estimation of shared representational format.

DNN-Extracted Scene Onsets

The second scenes in our transitions lacked a discrete onset, as they were gradually blended in from the first scene. As a result, the point at which the second scene became visually recognizable, and thus capable of eliciting a distinct onset response, likely varied across stimuli. This variability in the perceptual emergence of the second scene could smear the temporal decoding profile for the second scenes, given that these scenes would asynchronously pass through the visual hierarchy.

To account for this, we used a deep neural network (DNN) trained on scene recognition (VGG16-Places365; 19) to estimate the perceptual emergence of the second scene for each transition individually. To verify that the network captured the scene categories represented in our stimuli, we used it to classify the nine original scene images from which the scene transitions were generated, and observed 100% Top-1 accuracy. We then extracted, for each frame of a transition, embeddings from the output layer of the DNN, which were 365-dimensional vectors containing probabilities for each scene category included in the Places365 training data set. Using these probability vectors, we calculated the cosine similarity between each frame and the final frame of the scene transition, which corresponded to the fully revealed second scene. We then fit a sigmoid function to the sequence of cosine similarities across frames to model the emergence of the second scene.

From each fitted sigmoid, we identified three threshold points corresponding to 25%, 50%, and 75% similarity to the second scene as candidate onsets. In addition, as an alternative onset-estimation approach, we also identified, for each trial, the first frame at which the DNN probability assigned to the second-scene label exceeded that assigned to the first-scene label (i.e., decision boundary). These estimated onsets were then used to re-align the EEG data on a trial-by-trial basis, such that each stimulus was assigned its own estimated onset and corresponding temporal shift. We then recalculated second-scene decoding accuracy using the aligned data.

This approach allowed us to test whether aligning trials based on DNN-estimated perceptual onsets would reveal a more pronounced decoding peak and improve classification performance compared with analyses using unaligned data.

Temporal Generalization

We conducted a temporal generalization analysis (2) to investigate whether neural representations of scenes with artificial onsets (first-scene representations) generalize to those of scenes with naturalistic onsets (second-scene representations). To this end, we trained classifiers at each EEG time point using one set of labels (first-scene or second-scene labels) and tested them across all time points using the other set of labels. We then averaged decoding performance across the two directions of cross-generalization (i.e., first-to-second and second-to-first scene generalization). If second-scene representations share the same neural code as first-scene representations, we would expect a cluster of significant generalization at earlier time points for first-scene representations and later time points for second-scene representations (Fig. 2*B*, *top-left quadrant*), when the first- and second-scene labels are most diagnostic.

In addition to this cross-decoding analysis, we also examined temporal generalization within first- and second-scenes. For these complementary analyses, classifiers were trained and tested across all time points using only first-scene labels, and then separately using only second-scene labels. This enabled us to characterize the temporal stability of representations for both scenes and to assess whether any cross-decoding effects could be explained by the temporal generalization structure observed within each scene. Within-scene temporal generalization matrices were computed using the same preprocessing pipeline, classifier parameters, and cross-validation procedure as in the cross-decoding analysis.

Temporal generalization decoding was performed using a linear discriminant analysis (LDA) classifier across time. As in the time-resolved decoding analyses, a leave-one-out cross-validation approach was used, where each trial was excluded once as a test set, and the classifier was trained on the remaining trials. EEG data were resampled to 100 Hz before analysis for computational efficiency. The resulting accuracy matrices (time x time) were smoothed separately for each participant using Gaussian smoothing with a square kernel (SD: 2, kernel size: 9×9 time points).

To assess statistical significance, we again applied cluster-based permutation testing with threshold-free cluster enhancement (TFCE; 10,000 permutations) to identify time points with decoding accuracy significantly above chance.

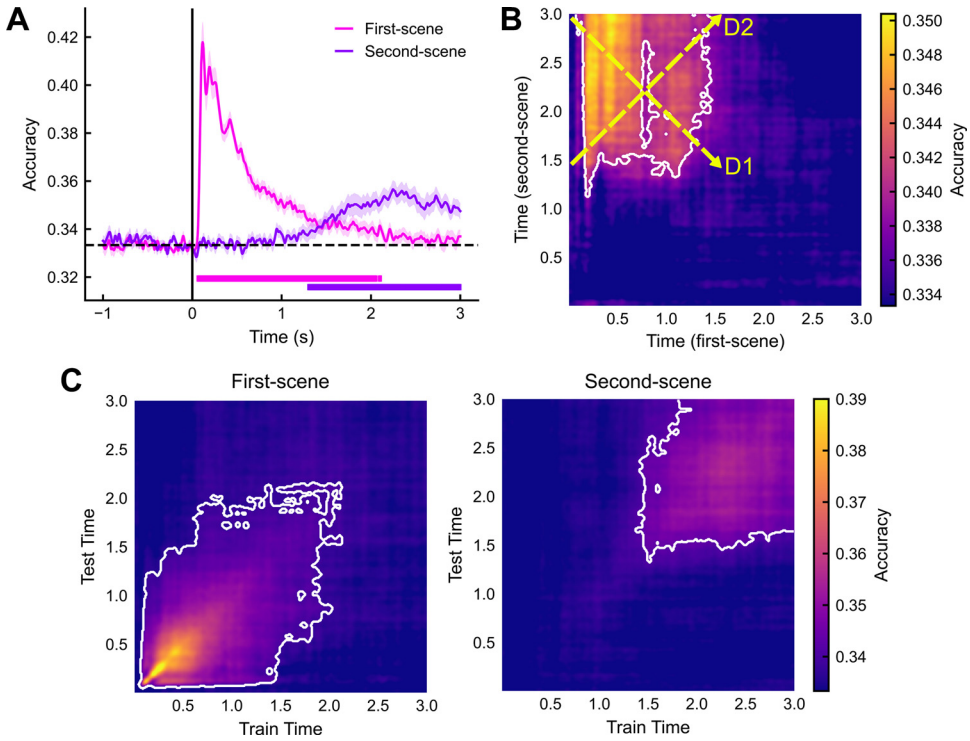

Figure 2. Decoding results. *A*: time-resolved category decoding accuracy ($n = 35$ participants) for the first scene (artificial onset) and second scene (naturalistic onset). Category information for the first scene emerged rapidly and yielded multiple discernible peaks, whereas category information for the second scene increased gradually without a distinct peak structure, indicating markedly different temporal dynamics of visual representation under artificial vs. naturalistic onsets. Horizontal bars: $P < 0.05$, shaded areas: means ± SE. *B*: temporal cross-generalization across first- and second-scene category labels. Classifiers were trained and tested across scenes in both directions (first-scene to second-scene and second-scene to first-scene) using separate sets of trials in 10-fold cross-validation, and results were averaged across directions at all time points. Temporal generalization revealed a broad, temporally diffuse cross-decoding pattern rather than a forward (D2) or reversed off-diagonal (D1) structure. This might suggest that first- and second-scene representations do not share a common hierarchical processing sequence, and that second-scene representations instead stabilize rapidly with minimal temporal evolution. D1: Expected diagonal generalization pattern if the classifier relied on frame-to-frame similarities. D2: Expected diagonal generalization pattern if decoding were driven by correspondence between hierarchical processing stages. Outlined areas: Significant clusters ($P < 0.05$). *C*: temporal generalization analyses for first (*left*) and second scenes (*right*). Classifiers were trained on either first- or second-scene labels and tested across all time points in an independent dataset. Decoding for the first scenes was most pronounced along the diagonal of the matrix, consistent with hierarchical processing, whereas decoding for the second scenes was not.

RESULTS

Time-Resolved Decoding

To examine the temporal dynamics of visual representations under artificial versus naturalistic scene onsets, we trained classifiers to decode the category labels of both the first and second scenes in each transition.

First-scene category decoding was significantly above chance between 60 ms and 2.06 s, peaking at 41.8% accuracy around 115 ms (Fig. 2*A*). In contrast, second-scene decoding did not show a distinct peak. Instead, decoding became significant after 1.3 s, gradually ramping up without a prominent peak, and reaching a maximum accuracy of 35.6%. This gradual buildup suggests a more temporally diffuse emergence of category information for scenes with naturalistic onsets.

Temporal Generalization

To investigate whether first- and second-scene representations share common encoding dynamics, we performed a temporal generalization analysis in which classifiers were trained at each time point on one set of category labels (first-scene or second-scene labels) and tested across all time points on the other set of labels. We used 10-fold cross-validation to divide the data into training and test sets, and generalization performance was averaged across folds. The first-to-second and second-to-first temporal generalization matrices were computed separately and then averaged to assess the extent to which first- and second-scene representations shared a common representational format. Classifier performance was quantified using decoding accuracy, with a chance level of 1/3 (≈33.3%). Statistical significance was assessed using a threshold-Free Cluster Enhancement (TFCE; 10,000 permutations).

Overall, there was robust temporal cross-generalization ($P < 0.05$; Fig. 2*B*) from 150–1,450 ms (first-scene) to 1,500–3,000 ms (second-scene). Surprisingly, we observed a temporally unspecific, broad generalization pattern, rather than more pronounced decoding along an off-diagonal in the generalization matrix. Given that the first and second scenes in the transition videos were mirrored, and thus time-reversed, versions of one another, a classifier driven primarily by frame-to-frame similarity would therefore be expected to yield a generalization pattern along a reversed off-diagonal (e.g., the first frame of one scene resembling the last frame of the other; Fig. 2*B*, D1). Conversely, if both scenes engage the

same hierarchical processing cascade once they appear, their representations would become more similar when corresponding processing stages align in time, yielding the opposite off-diagonal pattern (Fig. 2*B*, D2).

The broad, temporally diffuse generalization we observed instead suggests two possible interpretations. One is that the classifiers tracked relatively stable scene-category representations throughout the scene transitions. The other is that classifiers trained separately on first- and second-scene labels relied, at least in part, on neural representations of low-level visual features that differed systematically across scene categories and were shared within categories, rather than on fully developed high-level scene representations. Under this second interpretation, broad cross-generalization could arise because both classifiers were sensitive to representational dimensions available early in the visual processing hierarchy, which may help explain why classifiers trained on first-scene labels at around 150 ms generalized broadly to second-scene representations, despite such early latencies not typically being associated with fully developed high-level scene representations (20, 21).

In either case, the broad temporal cross-generalization pattern, rather than a diagonal one, argues against a shared temporal processing cascade in which different levels of stimulus features are processed at distinct latencies, akin to a hierarchical processing sequence. Instead, the within-scene temporal generalization results suggest that second-scene representations rapidly reach a relatively stable code, with little evidence of subsequent temporal evolution, whereas first-scene representations show a clear diagonal generalization pattern consistent with hierarchical processing (Fig. 2*C*). Although these findings are consistent with the possibility that hierarchical visual processing engages visual information differently across stimulus-presentation contexts, additional studies using multi-level categorical stimuli, together with control conditions that disentangle the various sensory and neural influences contributing to the observed effects, are needed to determine whether and how hierarchical processing unfolds during continuous visual input.

There was also a brief time window around the 720–850 ms interval during which first- and second-scene representations failed to generalize to one another. This may indicate a transient change in the representational format and/or content of first-scene representations during this period, causing them to diverge from the shared format otherwise suggested by the broad generalization profile. The basis of this brief disruption in temporal generalization remains unclear, as does whether it reflects a general phenomenon or is specific to our stimulus set.

DNN-Extracted Scene Onsets

Because the second scenes were blended in gradually rather than appearing with a discrete onset, variability in their emergence likely introduced temporal jitter across transitions, which could smear the onset-locked neural response and obscure a clear peak in second-scene decoding accuracy. Therefore, we realigned signals to DNN-estimated perceptual onsets for second scenes to test whether this would improve decoding performance and/or recover a peak structure for second-scene decoding. We recalculated second-scene decoding accuracy after shifting trials to each of the four realignment thresholds (decision boundary, and 25%, 50%, and 75% similarity to the second scene). We conducted separate repeated-measures ANOVAs on the average decoding accuracy within the first 500 ms following the DNN-estimated onset of second-scene emergence, with realignment threshold (unaligned, 25%, 50%, 75%, decision boundary) as a within-subjects factor. We selected this 500-ms time window because we anticipated the strongest initial responses during this interval. This also ensured that offset responses occurring at the end of the videos were excluded from all realigned signals.

The analysis did not reveal a significant main effect of realignment threshold [$F(4,136) = 1.25$, $P = 0.294$, $\eta_p^2 = 0.035$]. None of the realigned conditions produced a qualitatively different temporal decoding pattern, and none yielded a distinct peak structure in second-scene decoding (Fig. 3). Instead, the decoding time courses remained comparably diffuse across realignment thresholds. Thus, DNN-based realignment could not recover any resemblance to the sharp onset-related peaks observed for first-scene decoding.

DISCUSSION

We investigated how decoding dynamics in M/EEG are shaped by the timing and manner in which category-diagnostic visual information becomes available. Specifically, we compared two presentation regimes: abrupt scene presentation following a grayscale screen (artificial onset) and gradual scene emergence during smooth panoramic transitions (naturalistic onset). Because these conditions differed in both onset dynamics and the presence of preexisting visual information, the present design does not isolate the contribution of either factor, but the observed differences provide a useful starting point for understanding how decoding dynamics depend on stimulus-presentation context.

We observed three key results. First, artificial onsets reproduced the well-established pattern of a pronounced early decoding peak at ~60 ms (peak accuracy = 41.8%), followed by a decay that approached, but remained above, chance (Fig. 2*A*). Second, naturalistic onsets lacked a discrete peak structure, instead showing a slower ramp to a lower, sustained decoding level (peak accuracy = 35.6%). Third, time-generalization analyses revealed that representations at the onset-locked peak generalized to naturalistic-onset scenes, indicating a shared representational code despite distinct temporal profiles (Fig. 2*B*). Together, these findings demonstrate that M/EEG decoding dynamics are sensitive to presentation context, highlighting that temporal decoding profiles can differ substantially across presentation regimes even when they share a common representational format.

Unlike classical ERP approaches, which emphasize transient, stimulus-locked peaks, multivariate decoding methods provide a richer characterization of how information unfolds continuously over time. By tracing representational dynamics rather than discrete components, decoding approaches offer a principled way to investigate the seamless evolution of category information as perceptual inputs change. Our results show that decoding dynamics differ substantially across stimulus-presentation contexts. In particular, the abrupt artificial-onset condition widely used in traditional trial-based paradigms yielded a different decoding profile from the naturalistic-onset condition, in which category-

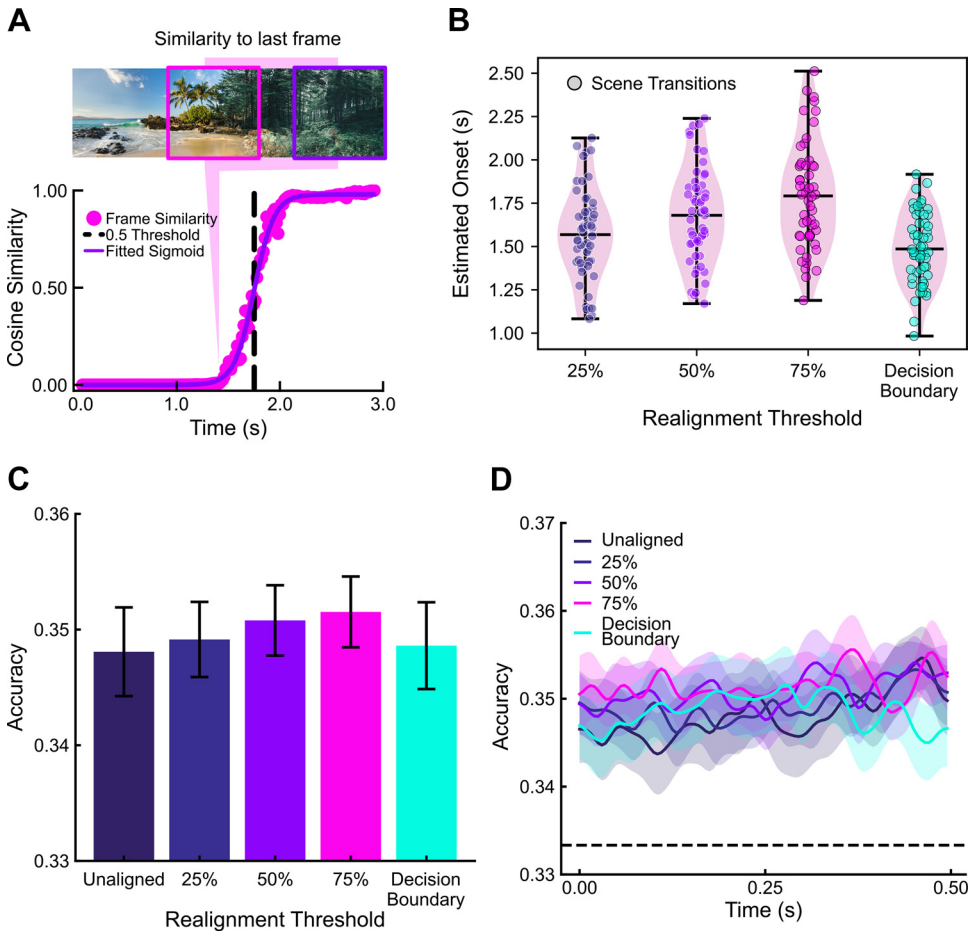

Figure 3. Deep neural network (DNN)-extracted scene onsets. *A*: example condition showing frame-wise similarity to the second scene (i.e., the last frame of the video; dots) with a fitted sigmoid (line). Transition-specific onsets of second-scene emergence were estimated either as the first frame at which the fit crosses each threshold (25%, 50%, 75%), or as the first frame at which the probability of the second-scene label exceeded that of the first-scene label (decision boundary). Trials were then realigned to these estimated onset times. *B*: distribution of estimated onset times across thresholds (25%, 50%, 75%, decision boundary) for all scene transitions. *C*: mean second-scene decoding accuracy evaluated for the 500-ms window after the DNN-extracted onset for each alignment condition (25%, 50%, 75%, decision boundary) and for the unaligned baseline (stimulus-onset-locked signals cropped at 1.5–2 s). None of the realignment thresholds significantly improved the mean decoding accuracy within the 500-ms time window following the DNN-extracted onsets. Error bars: means ± SE. *D*: time-resolved decoding accuracy aligned to each threshold (and unaligned baseline). None of the aligned conditions produces a distinct second-scene peak; trajectories remain relatively flat across thresholds, consistent with the absence of a sharp onset-locked response.

diagnostic visual information emerged gradually during continuous visual input. This raises the possibility that a narrow focus on early, onset-locked decoding peaks may underutilize one of the core advantages of multivariate approaches over ERPs and limit the generalizability of resulting interpretations to real-world visual processing.

Reassuringly, our findings show that representations evoked by artificial onsets generalized robustly to those elicited by naturalistic onsets. This demonstrates that although abrupt visual onsets modulate the temporal profile and apparent strength of decoding, they do not fundamentally change the underlying representational format. In this sense, the core neural code appears stable across different modes of stimulus presentation. As a result, for studies primarily aimed at establishing whether specific stimulus information is represented in the brain, irrespective of the precise temporal dynamics, the choice between artificial and naturalistic onsets is unlikely to alter the essential interpretive conclusion.

However, our results also raise the question of how well the dynamics observed in trial-based paradigms capture neural processing in natural vision, where category-diagnostic information often unfolds within continuous visual input rather than appearing as discrete visual onsets. Many M/EEG decoding studies with artificial stimulus onsets have reported latencies for the emergence of representations at different levels (e.g., see Refs. 4, 22, and 23). Although these benchmarks have been valuable for mapping processing under abrupt-transient stimulation, their exact latencies may not translate to naturalistic viewing. In everyday vision, incoming information is temporally correlated, and predictive mechanisms can preactivate and sustain representations before diagnostic sensory evidence arrives (24, 25). From a predictive-processing perspective, higher-level categorical expectations may at times precede and shape lower-level sensory representations (24, 25), in contrast to the strictly bottom-up timeline implied by onset-locked designs. Consequently, onset-locked latency estimates should be treated as design-contingent markers rather than general laws of visual processing, and our understanding of processing order should be supplemented with paradigms that incorporate continuous, naturalistic input.

At the same time, one might argue that abrupt changes in visual input are not entirely unnatural, because each saccade brings a new region of the scene into foveal vision, effectively creating a sequence of quasi-onset events. Consistent with this, combined eye-tracking and EEG studies have identified robust saccade-locked responses during free viewing, including fixation-related components and category-sensitive effects that closely resemble those reported in classic, fixation-enforced paradigms (e.g., see Refs. 26 and 27). Moreover, recent evidence suggests that fixation-related potentials can exhibit faster temporal dynamics than stimulus-onset-locked activity (28), and that stimulus-related responses are more tightly aligned with saccade onsets than with subsequent fixation onsets (29); consistent with top-down, action-oriented emergence of visual representations

that is already underway before the new fixation is established. Together, these findings point to a sophisticated coordination of goal-directed behavior, predictive processing, and peripheral vision in shaping visual representations, and underscore the need for future decoding work that explicitly models saccade-locked dynamics during free viewing to assess how far onset-based paradigms generalize to natural vision. Although the gradual emergence of scenes in our paradigm did not elicit strong transient responses, instead giving rise to more temporally diffuse and sustained representations, transient visual activity may still play a major role in natural vision through self-generated abrupt onsets caused by saccades. Accordingly, our naturalistic onset condition approximates one feature of natural vision, namely the progressive accumulation of category-diagnostic information. At the same time, it does not capture the dynamics of natural viewing driven by sequences of eye movements. Our results are therefore most directly informative for paradigms that restrict eye movements and reinforce the importance of saccade-locked analyses by revealing markedly different temporal dynamics under gradual, continuous visual stimulation.

What causes the marked differences in temporal profiles for artificial versus naturalistic onsets? Our data do not arbitrate among mechanisms, but we outline three not mutually exclusive accounts. First, artificial onsets may recruit neurons tuned to transient, flicker-like input (14, 15), briefly boosting signal-to-noise ratio and producing the sharp decoding peak, with responses then decaying. The absence of such a sharp decoding peak would also align with the lower, sustained accuracy for naturalistic onsets. Second, visual onsets are highly salient and can trigger bottom-up attentional capture (12, 13), thereby enhancing visual representations immediately after a sudden appearance. Third, an artificial onset may reset the phase of ongoing oscillations (30, 31), synchronizing population activity across trials during the transient window and improving temporal alignment. This phase-resetting account is also consistent with the "pinging" effect, in which an abrupt task-irrelevant stimulation impulse temporarily restores decoding accuracy for working memory items (32, 33). With the present data, we cannot determine the relative contributions of these mechanisms and therefore do not prioritize one account over the others.

The present results also highlight some important considerations for future work. First, the perceptual onset of second scenes likely varied across transitions and across repetitions as participants became familiar with the sequences, which can blur time-resolved decoding estimates and flatten transient peaks. We attempted to mitigate this by realigning trials to DNN-estimated perceptual onsets, but residual across-repetition variability may still have contributed to the absence of a distinct peak structure. Future studies could use temporally unconstrained decoding approaches based on Hidden Markov Models to characterize distinct representational states associated with dynamically changing stimulus categories (34, 35). Second, our stimulus set was relatively small. With only three exemplars per category and correlations among within-category exemplars, classifiers may have relied primarily on the EEG representations of low-level visual features rather than semantic content. As a result, our findings are not well-suited for comparing the temporal dynamics of high-level category representations across onset regimes, although they still elucidate how artificial onsets shape the temporal profile of low-level visual decoding. Third, in our design, the visual system was already adapted to ongoing stimulation by the time the second scenes appeared. This may have reduced overall response magnitudes for the second scene and thereby limited decoding performance. However, our primary interest was in the temporal structure and cross-generalization of decoding rather than absolute signal strength, so adaptation is unlikely to undermine our main conclusions. Finally, the scene categories in our design were always task-relevant, which might have influenced the temporal dynamics we report through task-related memory effects or representational competition.

Taken together, our results show that the artificial-onset condition was associated with a brief, pronounced peak in decodable scene information, whereas the naturalistic-onset condition produced a lower and more sustained decoding profile. Despite these distinct temporal profiles, cross-condition generalization was consistent with a broadly shared underlying representational format. Because the two conditions differed in both onset dynamics and the presence of preexisting visual information, these differences cannot be attributed specifically to onset type. Instead, they demonstrate that time-resolved decoding measures are sensitive to the broader stimulus-presentation context. Accordingly, onset-locked latency estimates should be interpreted as potentially design-dependent rather than as invariant markers of visual processing, and inferences about processing order would benefit from comparison across a wider range of presentation regimes and appropriate control conditions. More broadly, our findings motivate further investigation of how M/EEG decoding dynamics vary between conventional trial-based paradigms and paradigms involving continuous, dynamically changing visual input.

DATA AVAILABILITY

All data and code are publicly available: https://doi.org/10.5281/zenodo.18326079.

ACKNOWLEDGMENTS

We thank Rico Stecher and the students in the Neural Computation I course for help with experiment setup and data collection. We further thank Malaika Alphonsus, Tanja John, Pietra Pacheco Alves, and Nasibeh Babaei for assistance in data collection.

GRANTS

This work was supported by the DFG Grant KA4683/5-1; under Project Number 518483074 and under Germany's Excellence Strategy (EXC 3066/1 "The Adaptive Mind") under Project Number 533717223.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

I.D., M.E., and D.K. conceived and designed research; I.D. and M.E. performed experiments; I.D. and D.K. analyzed data; I.D.,

M.E., and D.K. interpreted results of experiments; I.D. prepared figures; I.D. and D.K. drafted manuscript; I.D., M.E., and D.K. edited and revised manuscript; I.D., M.E., and D.K. approved final version of manuscript.

REFERENCES

1. **Grootswagers T**, **Wardle SG**, **Carlson TA.** Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J Cogn Neurosci* 29: 677–697, 2017. doi:10.1162/jocn_a_01068.
2. **King JR**, **Dehaene S.** Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn Sci* 18: 203–210, 2014. doi:10.1016/j.tics.2014.01.002.
3. **Carlson T**, **Tovar DA**, **Alink A**, **Kriegeskorte N.** Representational dynamics of object vision: the first 1000 ms. *J Vis* 13: 1, 2013. doi:10.1167/13.10.1.
4. **Cichy RM**, **Pantazis D**, **Oliva A.** Resolving human object recognition in space and time. *Nat Neurosci* 17: 455–462, 2014. doi:10.1038/nn.3635.
5. **Kaiser D**, **Azzalini DC**, **Peelen MV.** Shape-independent object category responses revealed by MEG and fMRI decoding. *J Neurophysiol* 115: 2246–2250, 2016. doi:10.1152/jn.01074.2015.
6. **Kaiser D**, **Oosterhof NN**, **Peelen MV.** The neural dynamics of attentional selection in natural scenes. *J Neurosci* 36: 10522–10528, 2016. doi:10.1523/JNEUROSCI.1385-16.2016.
7. **Contini EW**, **Wardle SG**, **Carlson TA.** Decoding the time-course of object recognition in the human brain: from visual features to categorical decisions. *Neuropsychologia* 105: 165–176, 2017. doi:10.1016/j.neuropsychologia.2017.02.013.
8. **Wardle SG**, **Baker C.** Recent advances in understanding object recognition in the human brain: deep neural networks, temporal dynamics, and context. *F1000Res* 9: F1000 Faculty Rev-590, 2020. doi:10.12688/f1000research.22296.1.
9. **Kaiser D**, **Moeskops MM**, **Cichy RM.** Typical retinotopic locations impact the time course of object coding. *NeuroImage* 176: 372–379, 2018. doi:10.1016/j.neuroimage.2018.05.006.
10. **Cole GG**, **Kentridge RW**, **Heywood CA**, **Cole GG.** Visual salience in the change detection paradigm: the special role of object onset. *J Exp Psychol Hum Percept Perform* 30: 464–477, 2004. doi:10.1037/0096-1523.30.3.464.
11. **Van Pelt J**, **Lowe BG**, **Robinson JE**, **Donaldson MJ**, **Johnston P**, **Yamamoto N.** An event-related potential study of onset primacy in visual change detection. *Atten Percept Psychophys* 87: 1219–1229, 2025. doi:10.3758/s13414-025-03027-4.
12. **Yantis S**, **Jonides J.** Abrupt visual onsets and selective attention: evidence from visual search. *J Exp Psychol Hum Percept Perform* 10: 601–621, 1984. doi:10.1037//0096-1523.10.5.601.
13. **Yantis S**, **Jonides J.** Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *J Exp Psychol Hum Percept Perform* 16: 121–134, 1990. doi:10.1037//0096-1523.16.1.121.
14. **Todd JT**, **Gelder PV.** Implications of a transient-sustained dichotomy for the measurement of human performance. *J Exp Psychol Hum Percept Perform* 5: 625–638, 1979. doi:10.1037//0096-1523.5.4.625.
15. **Tolhurst DJ.** Sustained and transient channels in human vision. *Vision Res* 15: 1151–1155, 1975. doi:10.1016/0042-6989(75)90014-0.
16. **Gramfort A**, **Luessi M**, **Larson E**, **Engemann DA**, **Strohmeier D**, **Brodbeck C**, **Goj R**, **Jas M**, **Brooks T**, **Parkkonen L**, **Hämäläinen M.** MEG and EEG data analysis with MNE-Python. *Front Neurosci* 7: 267, 2013. doi:10.3389/fnins.2013.00267.
17. **Smith SM**, **Nichols TE.** Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* 44: 83–98, 2009. doi:10.1016/j.neuroimage.2008.03.061.
18. **Maris E**, **Oostenveld R.** Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164: 177–190, 2007. doi:10.1016/j.jneumeth.2007.03.024.
19. **Zhou B**, **Lapedriza A**, **Khosla A**, **Oliva A**, **Torralba A.** Places: a 10 million image database for scene recognition. *IEEE Trans Pattern Anal Mach Intell* 40: 1452–1464, 2018. doi:10.1109/tpami.2017.2723009.
20. **Kaiser D**, **Cichy RM.** Parts and wholes in scene processing. *J Cogn Neurosci* 34: 4–15, 2021. doi:10.1162/jocn_a_01788.
21. **Harel A**, **Groen IIA**, **Kravitz DJ**, **Deouell LY**, **Baker CI.** The temporal dynamics of scene processing: a multifaceted eeg investigation. *eNeuro* 3: ENEURO.0139-16.2016, 2016. doi:10.1523/ENEURO.0139-16.2016.
22. **Cichy RM**, **Pantazis D**, **Oliva A.** Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cereb Cortex* 26: 3563–3579, 2016. doi:10.1093/cercor/bhw135.
23. **Grootswagers T**, **Robinson AK**, **Carlson TA.** The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage* 188: 668–679, 2019. doi:10.1016/j.neuroimage.2018.12.046.
24. **Battistoni E**, **Stein T**, **Peelen MV.** Preparatory attention in visual cortex. *Ann NY Acad Sci* 1396: 92–107, 2017. doi:10.1111/nyas.13320.
25. **Peelen MV**, **Berlot E**, **de Lange FP.** Predictive processing of scenes and objects. *Nat Rev Psychol* 3: 13–26, 2024. doi:10.1038/s44159-023-00254-0.
26. **Dimigen O**, **Ehinger BV.** Regression-based analysis of combined EEG and eye-tracking data: theory and applications. *J Vis* 21: 3, 2021. doi:10.1167/jov.21.1.3.
27. **Gert AL**, **Ehinger BV**, **Timm S**, **Kietzmann TC**, **König P.** WildLab: a naturalistic free viewing experiment reveals previously unknown electroencephalography signatures of face processing. *Eur J Neurosci* 56: 6022–6038, 2022. doi:10.1111/ejn.15824.
28. **Auerbach-Asch CR**, **Gal V**, **Oded W**, **Leon YD.** Decoding object categories from EEG during free viewing reveals early information evolution compared to passive viewing (Preprint). *bioRxiv*, 2023. http://biorxiv.org/lookup/doi/10.1101/2023.06.28.546397 [2025 Jul 2].
29. **Amme C**, **Sulewski P**, **Spaak E**, **Hebart MN**, **König P**, **Kietzmann TC.** Saccade onset, not fixation onset, best explains early responses across the human visual cortex during naturalistic vision (Preprint). *bioRxiv*, 2024. http://biorxiv.org/lookup/doi/10.1101/2024.10.25.620167 [2025 Dec 23].
30. **Hanslmayr S**, **Klimesch W**, **Sauseng P**, **Gruber W**, **Doppelmayr M**, **Freunberger R**, **Pecherstorfer T**, **Birbaumer N.** Alpha phase reset contributes to the generation of ERPs. *Cereb Cortex* 17: 1–8, 2007. doi:10.1093/cercor/bhj129.
31. **Makeig S**, **Westerfield M**, **Jung TP**, **Enghoff S**, **Townsend J**, **Courchesne E**, **Sejnowski TJ.** Dynamic brain sources of visual evoked responses. *Science* 295: 690–694, 2002 [Erratum in *Science* 295: 1466, 2002]. doi:10.1126/science.1066168.
32. **Kandemir G**, **Wilhelm SA**, **Axmacher N**, **Akyürek EG.** Maintenance of color memoranda in activity-quiescent working memory states: evidence from impulse perturbation. *iScience* 27: 109565, 2024. doi:10.1016/j.isci.2024.109565.
33. **Wolff MJ**, **Ding J**, **Myers NE**, **Stokes MG.** Revealing hidden states in visual working memory using electroencephalography. *Front Syst Neurosci* 9: 123, 2015. doi:10.3389/fnsys.2015.00123.
34. **Vidaurre D**, **Myers NE**, **Stokes M**, **Nobre AC**, **Woolrich MW.** Temporally unconstrained decoding reveals consistent but time-varying stages of stimulus processing. *Cereb Cortex* 29: 863–874, 2019. doi:10.1093/cercor/bhy290.
35. **Vidaurre D**, **Masaracchia L**, **Larsen NY**, **Ruijters LRPT**, **Alonso S**, **Ahrends C**, **Woolrich MW.** The Gaussian-linear hidden Markov model: a Python package. *Imaging Neurosci (Camb)* 3: imag_a_00460, 2025. doi:10.1162/imag_a_00460.