Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision

Lixiang Chen[1,*], Radoslaw M Cichy[1,#], Daniel Kaiser[2,3,*,#]

[1]*Department of Education and Psychology, Freie Universität Berlin, Germany*

[2]*Mathematical Institute, Department of Mathematics and Computer Science, Physics, Geography, Justus-Liebig-University Gießen, Germany*

[3]*Center for Mind, Brain and Behavior (CMBB), Philipps-University Marburg and Justus-Liebig-University Gießen, Germany*

*<u>*Correspondence to:*</u>

Lixiang Chen

lixiang.chen@fu-berlin.de

or

Daniel Kaiser

danielkaiser.net@gmail.com

[#]*These authors contributed equally.*

Abstract

During naturalistic vision, the brain generates coherent percepts by integrating sensory inputs scattered across the visual field. Here, we asked whether this integration process is mediated by rhythmic cortical feedback. In EEG and fMRI experiments, we experimentally manipulated integrative processing by changing the spatiotemporal coherence of naturalistic videos presented across visual hemifields. Our EEG data revealed that information about incoherent videos is coded in feedforward-related gamma activity while information about coherent videos is coded in feedback-related alpha activity, indicating that integration is indeed mediated by rhythmic activity. Our fMRI data identified scene-selective cortex and hMT as likely sources of this feedback. Analytically combining our EEG and fMRI data further revealed that feedback-related representations in the alpha band shape the earliest stages of visual processing in cortex. Together, our findings indicate that the construction of coherent visual experiences relies on cortical feedback rhythms that fully traverse the visual hierarchy.

Keywords: cortical alpha rhythms; top-down modulation; multivariate pattern analysis; EEG-fMRI fusion; real-world vision

Introduction

We consciously experience our visual surroundings as a coherent whole that is phenomenally unified across space (Block, 2007; Cohen et al., 2016). In our visual system, however, inputs are initially transformed into a spatially fragmented mosaic of local signals that lack integration. How does the brain integrate this fragmented information across the subsequent visual processing cascade to mediate unified perception?

Classic hierarchical theories of vision posit that integration is solved during feedforward processing (Riesenhuber and Poggio, 1999; DiCarlo and Cox, 2007). On this view, integration is hard-wired into the visual system: Local representations of specific features are integrated into more global representations of meaningful visual contents through hierarchical convergence over features distributed across visual space.

More recent theories instead posit that visual integration is achieved through complex interactions between feedforward information flow and dynamic top-down feedback (Rao and Ballard, 1999; Friston, 2005; Bastos et al., 2012). On this view, feedback information flow from downstream adaptively guides the integration of visual information in upstream regions. Such a conceptualization is anatomically plausible, as well as behaviorally adaptive, as higher-order regions can flexibly adjust current whether or not stimuli are integrated through the visual system's abundant top-down connections (Salin and Bullier, 1995; Lamme et al., 1998; Markov et al., 2013).

However, the proposed interactions between feedforward and feedback information pose a critical challenge: Feedforward and feedback information needs to be multiplexed across the visual hierarchy to avoid unwanted interferences through spurious interactions of these signals. Previous studies propose that neural systems meet this challenge by routing feedforward and feedback information in different neural frequency channels: high-frequency gamma (31–70 Hz) rhythms may mediate feedforward propagation, whereas low-frequency alpha (8–12 Hz) and beta (13–30 Hz) rhythms carry predictive feedback to upstream areas (van Kerkoerle et al., 2014; Bastos et al., 2015; Fries, 2015; Michalareas et al., 2016).

Here, we set out to test the hypothesis that rhythmic coding acts as a mechanism mediating coherent visual perception. We used a novel experimental paradigm that manipulated the degree to which stimuli could be integrated across space through the spatiotemporal coherence of naturalistic videos shown in the two visual hemifields. Combining EEG and fMRI measurements, we show that when inputs are integrated into a coherent percept, cortical alpha dynamics carry stimulus-specific feedback from high-level visual cortex to early visual cortex. Our results show that spatial integration of naturalistic visual inputs is mediated by feedback dynamics that traverse the visual hierarchy in low-frequency alpha rhythms.

Results

We experimentally mimicked the spatially distributed nature of naturalistic inputs by presenting eight 3-second naturalistic videos (Fig. 1A) through two circular apertures right and left of fixation (diameter: 6° visual angle, minimal distance to fixation: 2.64°). To assess spatial integration in a controlled way, we varied how the videos were presented through these apertures (Fig. 1B): In the right- or left-only condition, the video was shown only through one of the apertures, providing a baseline for processing inputs from one hemifield, without the need for spatial integration across hemifields. In the coherent condition, the same original video was shown through both apertures. Here, the input had the spatiotemporal statistics of a unified scene expected in the real world and could thus be readily integrated into a coherent unitary percept. In the incoherent condition, by contrast, the videos shown through the two apertures stemmed from

two different videos (see fig. S1). Here, the input did not have the spatiotemporal real-world statistics of a unified scene, and thus could not be readily integrated. Contrasting brain activity for the coherent and incoherent condition thus reveals neural signatures of spatial integration into unified percepts across visual space.

Participants viewed the video stimuli in separate EEG (n = 48) and fMRI (n = 36) recording sessions. Participants performed an unrelated central task (Fig. 1C) to ensure fixation and to allow us to probe integration processes in the absence of explicit task demands.

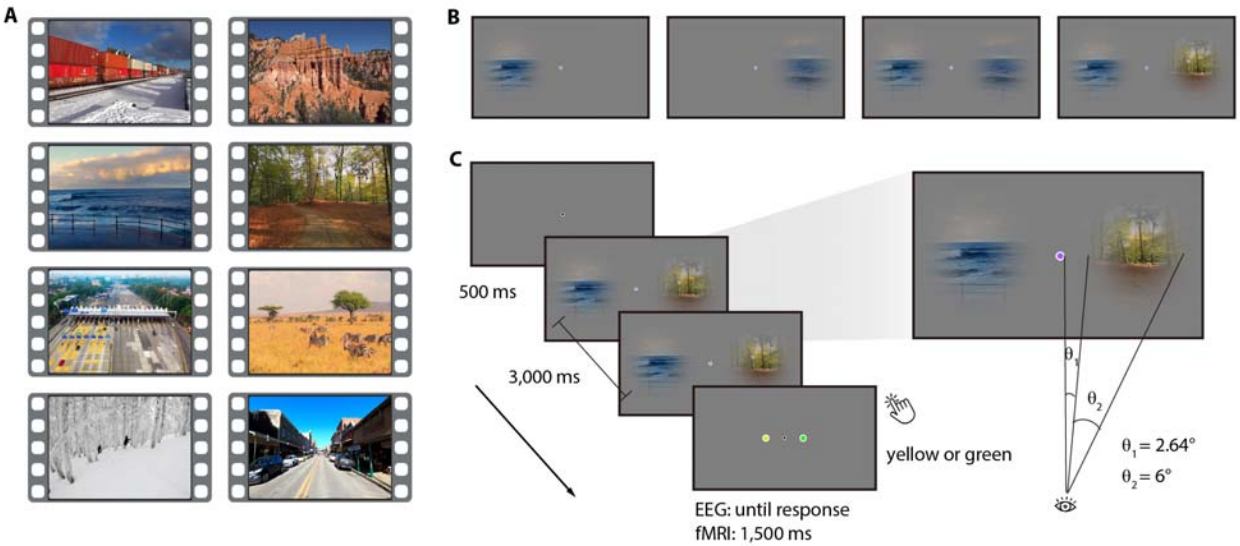

Fig. 1. Stimuli and Experimental design. A) Snapshots from the eight videos used. **B)** In the experiment, videos were either presented through one aperture in the right or left visual field, or through both apertures in a coherent or incoherent way. **C)** During the video presentation, the color of the fixation dot changed periodically (every 200 ms). Participants reported whether a green or yellow fixation dot was included in the sequence.

Harnessing the complementary frequency resolution and spatial resolution of our EEG and fMRI recordings, we then delineated how inputs that either can or cannot be integrated into a coherent percept are represented in rhythmic neural activity and regional activity across the visual hierarchy. Specifically, we decoded between the eight different video stimuli in each of the four conditions from frequency-resolved EEG sensor patterns (Xie et al., 2020; Kaiser, 2022) and from spatially resolved fMRI multi-voxel patterns (Haynes, 2015).

Rhythmic brain dynamics mediate integration across visual space

Our first key analysis determined how the feedforward and feedback information flows involved in the processing and integrating visual information across space are multiplexed in rhythmic codes. We hypothesized that conditions not affording integration lead to neural coding in feedforward-related gamma activity (van Kerkoerle et al., 2014; Michalareas et al., 2016), whereas conditions that allow for spatiotemporal integration lead to coding in feedback-related alpha/beta activity (van Kerkoerle et al., 2014; Michalareas et al., 2016).

To test this hypothesis, we decoded the video stimuli from spectrally resolved EEG signals, aggregated within the alpha, beta, and gamma frequency bands, during the whole stimulus duration (Fig. 2A; see Materials and Methods for details). Our findings supported our hypothesis. We observed that incoherent video stimuli, as well as single video stimuli, were decodable only from the gamma frequency band (all $t(47) > 3.41$, $p < 0.001$; Fig. 2B and 2C). By stark contrast, coherent video stimuli were decodable only from the alpha frequency band ($t(47) = 5.43$, $p <$

0.001; Fig. 2C). Comparing the pattern of decoding performance across frequency bands revealed that incoherent video stimuli were better decodable than coherent stimuli from gamma responses ($t(47) = 3.04$, $p = 0.004$) and coherent stimuli were better decoded than incoherent stimuli from alpha responses ($t(47) = 2.32$, $p = 0.025$; interaction: $F(2, 94) = 7.47$, $p < 0.001$; Fig. 2C). The observed effects also held when analyzing the data continuously across frequency space rather than aggregated in pre-defined frequency bands (see fig. S2) and were not found trivially in the evoked broadband responses (see fig. S3). We also analyzed the theta (4–7 Hz) and high-gamma bands (71–100 Hz) using the decoding analysis. For the theta band, we did not find any significant decoding (see fig. S2 and fig. S4). The results from the high-gamma band were highly similar to the results obtained for the lower-gamma frequency range (see fig. S2 and fig. S4). In addition, we conducted both univariate and decoding analyses on time- and frequency-resolved responses, but neither of these analyses revealed any differences between the coherent and incoherent conditions (see fig. S5), indicating a lack of statistical power for resolving the data both in time and frequency. Together, our results demonstrate the multiplexing of visual information in rhythmic information flows. When no integration across hemifields was required, visual feedforward activity is carried by gamma rhythms. When spatiotemporally coherent inputs allow for integration, integration-related feedback activity is carried by alpha rhythms.

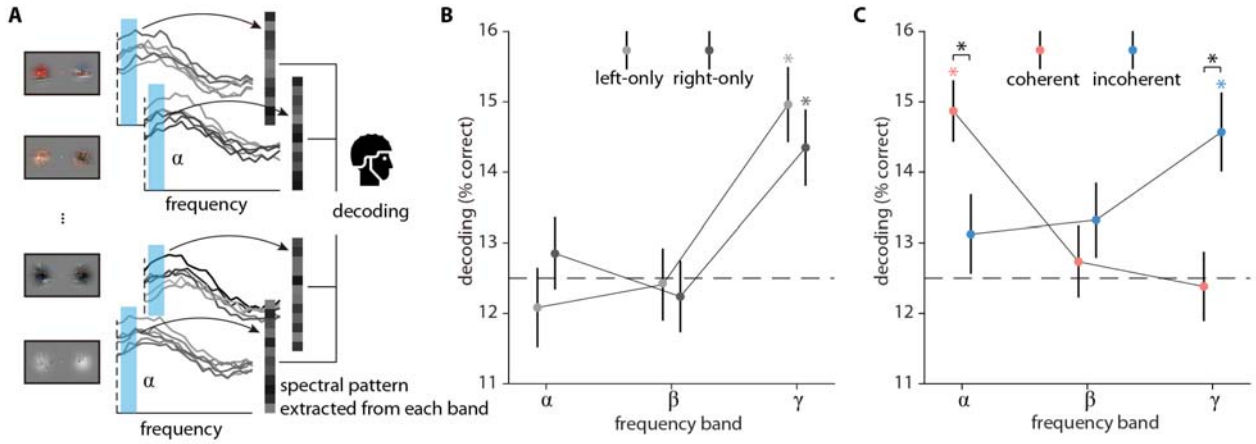

Fig. 2. EEG decoding analysis. A) Frequency-resolved EEG decoding analysis. In each condition, we used eight-way decoding to classify the video stimuli from patterns of spectral EEG power across electrodes, separately for each frequency band (α, β, γ). **B, C)** Results of EEG frequency-resolved decoding analysis. The incoherent and single video stimuli were decodable from γ responses whereas the coherent stimuli were decodable from α responses, suggesting a switch from dominant feedforward processing to the recruitment of cortical feedback. Error bars represent standard errors. *: $p < 0.05$ (FDR-corrected).

The observation of a frequency-specific channel for feedback information underlying spatial integration immediately poses two questions: (1) Where does this feedback originate from? and (2) Where is this feedback heading? We used fMRI recordings to answer these two questions in turn.

Scene-selective cortex is the source of integration-related feedback

To reveal the source of the feedback, we evaluated how representations across visual cortex differ between stimuli that can or cannot be integrated across space (Fig. 3A). We reasoned that regions capable of exerting integration-related feedback should show stronger representations of spatiotemporally coherent inputs that can be integrated, compared to incoherent inputs that do not. Scene-selective areas in visual cortex are a strong contender for the source of this feedback, as

they have been previously linked to the spatial integration of coherent scene information (Mannion et al., 2014; Kaiser and Cichy, 2021).

To test this assertion, we decoded the video stimuli from multi-voxel patterns in a set of three early visual cortex regions, V1, V2, and V3, one motion-selective region, hMT/V5, and three scene-selective regions, the occipital place area (OPA), the medial place area (MPA), and the parahippocampal place area (PPA).

In a first step we decoded between the single video stimuli, and found information in early visual cortex (V1, V2, V3) and scene-selective cortex (OPA, MPA, PPA) only when video stimuli were shown in the hemifield contralateral to the region investigated (all $t(35) > 3.75$, $p < 0.001$; Fig. 3B). This implies that any stronger decoding for coherent, compared to incoherent, video stimuli can only be driven by the interaction of ipsilateral and contralateral inputs, rather than by the ipsilateral input alone. On this interpretative backdrop, we next decoded coherent and incoherent video stimuli. Both were decodable in each of the seven regions (all $t(35) > 4.43$, $p < 0.001$; Fig. 3C). Critically, coherent video stimuli were only better decoded than incoherent stimuli in the MPA ($t(35) = 3.61$, $p < 0.001$; Fig. 3C) and PPA ($t(35) = 3.32$, $p = 0.002$; Fig. 3C). A similar trend was found in hMT ($t(35) = 1.73$, $p = 0.092$; Fig. 3C). In hMT, MPA, and PPA, coherent video stimuli were also better decoded than contralateral single video stimuli (all $t(35) > 2.99$, $p < 0.005$). Similar results were found in the whole-brain searchlight decoding analysis. We found significant decoding for single video stimuli across the visual cortex in the contralateral hemisphere (see fig. S6) as well as significant decoding across the visual cortex for both coherent (Fig. 3D) and incoherent stimuli (Fig. 3E). The differences between coherent and incoherent conditions were only found in locations overlapping – or close to – scene-selective cortex and hMT (Fig. 3F). Given the involvement of motion-selective hMT in integrating visual information, we also tested whether differences in motion coherence (operationalized as motion energy and motion direction) contribute to the integration effects observed here. When assessing differences between videos with high and low motion coherence across hemifields, however, we did not find qualitatively similar effects to our main analyses (see fig. S7), suggesting that motion coherence is not the main driver of the integration effects.

Together, these results show that scene-selective cortex and hMT aggregate spatiotemporally coherent information across hemifields, suggesting these regions as likely generators of feedback signals guiding visual integration.

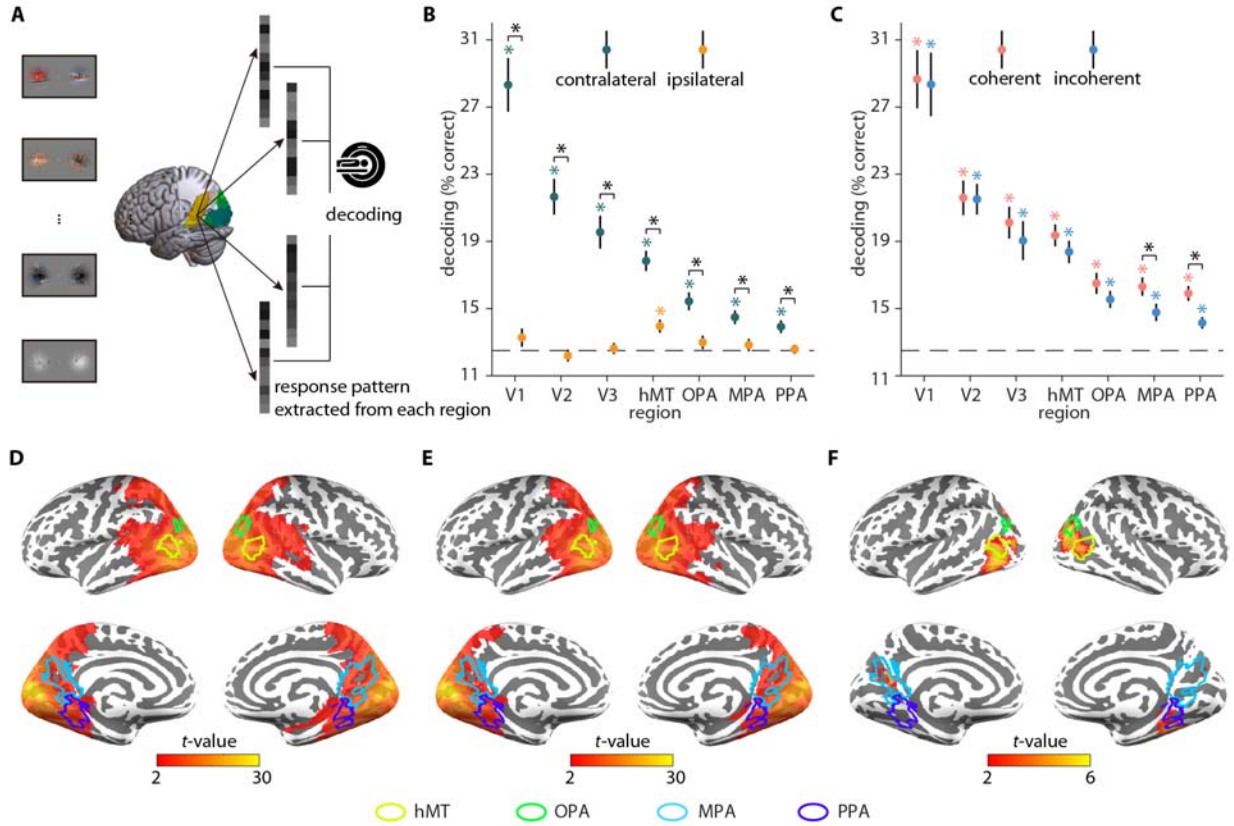

Fig. 3. fMRI decoding analysis. A) fMRI decoding analysis in regions of interest (ROIs). In each condition, we used eight-way decoding to classify the video stimuli on response patterns in each ROI (V1, V2, V3, hMT, OPA, MPA, PPA). **B)** Results of fMRI decoding analysis for the right- and left-only conditions. Single video stimuli were decodable in regions contralateral to the stimulation. In the ipsilateral hemisphere, they were only decodable in hMT, but not in early visual cortex (V1, V2, V3) and scene-selective cortex (OPA, MPA, PPA). **C)** Results of fMRI decoding analysis for the coherent and incoherent conditions. Video stimuli were decodable in both conditions in each of the seven regions. Coherent stimuli were decoded better than incoherent stimuli in scene-selective cortex (MPA and PPA). **D)** Results of fMRI searchlight decoding analysis for the coherent condition. Coherent stimuli were decodable across visual cortex. **E)** Results of fMRI searchlight decoding analysis for the incoherent condition. Incoherent stimuli were decodable across visual cortex. **F)** Significant differences between coherent and incoherent conditions in fMRI searchlight decoding analysis. Significant differences between the coherent and incoherent conditions were observed in locations overlapping – or close to – scene-selective cortex and hMT. Taken together, the results suggested that scene-selective cortex and hMT integrate dynamic information across visual hemifields. Error bars represent standard errors. *: $p < 0.05$ (FDR-corrected).

Integration-related feedback traverses the visual hierarchy

Finally, we determined where the feedback-related alpha rhythms are localized in brain space. We were particularly interested in whether integration-related feedback traverses the visual hierarchy up to the earliest stages of visual processing (Clavagnier et al., 2004; Muckli and Petro, 2013). To investigate this, we performed an EEG/fMRI fusion analysis (Cichy et al., 2014; Cichy and Oliva, 2020) that directly links spectral representations in the EEG with spatial representations in the fMRI. To link representations across modalities, we first computed representational similarities between all video stimuli using pair-wise decoding analyses and then correlated the similarities obtained from EEG alpha responses and fMRI activations across the seven visual regions (Fig. 4A). Here, we focused on the crucial comparison of regional representations (fMRI) and alpha-frequency representations (EEG) between the coherent and the incoherent conditions. Our fMRI decoding analyses for the single video stimuli demonstrates that V1 only receives sensory

information from the contralateral visual field. As feedforward inputs from the contralateral visual field are identical across both conditions, any stronger correspondence between regional representations and alpha-frequency representations in the coherent condition can unequivocally be attributed to feedback from higher-order systems, which have access to both ipsi- and contralateral input. We found that representations in the alpha band were more strongly related to representations in the coherent condition than in the incoherent condition in V1 ($t(35) = 3.37$, $p = 0.001$; Fig. 4B). A similar trend emerged in V2 ($t(35) = 2.32$, $p_{uncorrected} = 0.025$; Fig. 4B) and V3 ($t(35) = 2.15$, $p_{uncorrected} = 0.036$; Fig. 4B), but not in hMT ($t(35) = -0.28$, $p = 0.783$; Fig. 4B) and scene-selective cortex (OPA: $t(35) = 0.94$, $p = 0.351$; MPA: $t(35) = 0.005$, $p = 0.996$; PPA: $t(35) = -1.64$, $p = 0.108$; Fig. 4B). The correspondence between alpha-band representations in the EEG and activity in early visual cortex persisted after we controlled for motion coherence in the fusion analysis (see fig. S8), suggesting that the effect was not solely attributable to coherent patterns of motion. By contrast, no such correspondences were found between beta/gamma EEG responses and regional fMRI activations (see fig. S9). The results of the fusion analysis show that when inputs are spatiotemporally coherent and can be integrated into a unified percept, feedback-related alpha rhythms are found at the earliest stages of visual processing in cortex.

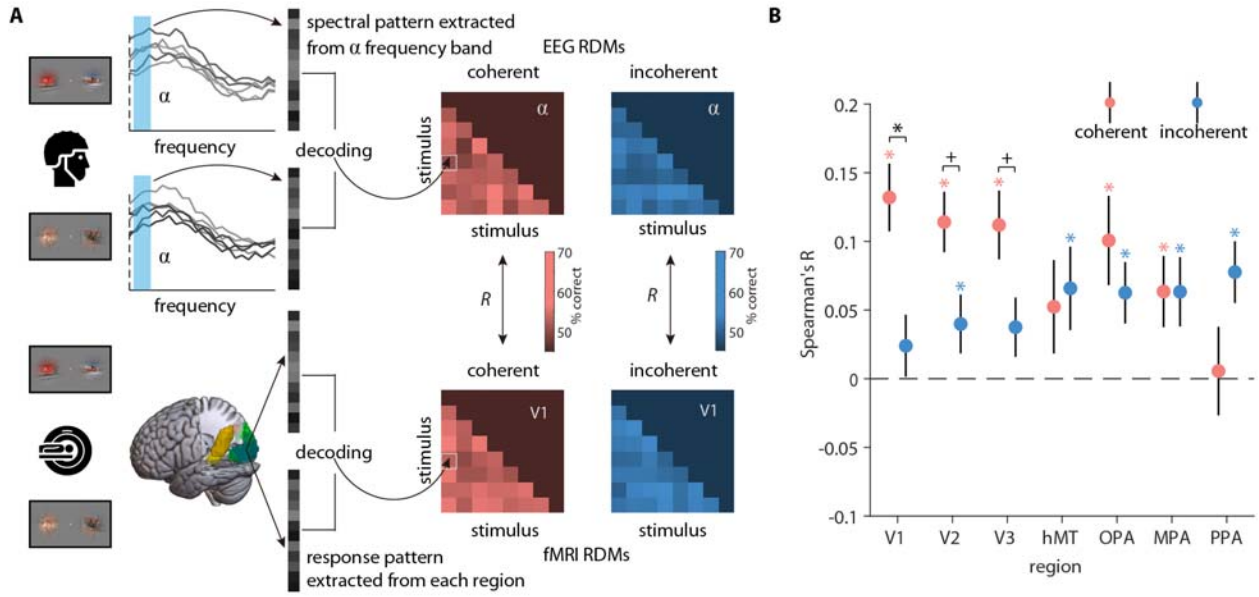

Fig. 4. EEG-fMRI fusion analysis. A) For each condition, EEG representational dissimilarity matrices (RDMs) for each frequency band (α, β, γ) and fMRI RDMs for each region of interest (V1, V2, V3, hMT, OPA, MPA, PPA) were first obtained using pair-wise decoding analyses. To assess correspondences between spectral and regional representations, we calculated *Spearman*-correlations between the participant-specific EEG RDMs in each frequency band and the group-averaged fMRI RDMs in each region, separately for each condition. **B)** Results of EEG-fMRI fusion analysis in the α band. Representations in the α band corresponded more strongly with representations in V1 (with a similar trend in V2 and V3) when the videos were presented coherently, rather than incoherently. No correspondences were found between the β and γ bands and regional activity (see fig. S9). Error bars represent standard errors. *: $p < 0.05$ (FDR-corrected), $^{+}$: $p < 0.05$ (uncorrected).

Discussion

Our findings demonstrate that the spatial integration of naturalistic inputs integral to mediating coherent perception is achieved by cortical feedback: Only when spatiotemporally coherent inputs can be integrated into a coherent whole, stimulus-specific information was coded in feedback-related alpha activity. We further show that scene-selective cortex and hMT interactively process information across visual space, highlighting them as likely sources of integration-related

feedback. Finally, we reveal that integration-related alpha dynamics are linked to representations in early visual cortex, indicating that integration is accompanied by feedback that traverses the whole cortical visual hierarchy from top to bottom. Together, our results promote an active conceptualization of the visual system, where concurrent feedforward and feedback information flows are critical for establishing coherent naturalistic vision.

Our finding that feedback reaches all the way to initial stages of visual processing supports the emerging notion that early visual cortex receives various types of stimulus-specific feedback, such as during mental imagery (Winlove et al., 2018; Ragni et al., 2021), in cross-modal perception (Vetter et al., 2014, 2020), and during the interpolation of missing contextual information (Smith and Muckli, 2010; Muckli et al., 2015). Further supporting the interpretation of such signals as long-range feedback, recent animal studies have found that contextual signals in V1 are substantially delayed in time, compared to feedforward processing (Keller et al., 2020; Papale et al., 2022; Kirchberger et al., 2023). Such feedback processes may utilize the spatial resolution of V1 as a flexible sketchpad mechanism (Dehaene and Cohen, 2007; Williams et al., 2008) for recreating detailed feature mappings that are inferred from global context.

Our fMRI data identify scene-selective areas in the anterior ventral temporal cortex (the MPA and the PPA) and motion-selective area hMT as probable sources of the feedback to early visual cortex. These regions exhibited stronger representations for spatiotemporally coherent stimuli placed in the two hemifields. Scene-selective cortex is a logical candidate for a source of the feedback: Scene-selective regions are sensitive to the typical spatial configuration of scene stimuli (Mannion et al., 2014; Bilalić et al., 2019; Kaiser et al., 2020; Kaiser and Cichy, 2021), allowing them to create feedback signals that carry information about whether and how stimuli need to be integrated at lower levels of the visual hierarchy. Such feedback signals may stem from adaptively comparing contralateral feedforward information with ipsilateral information from interhemispheric connections. In the incoherent condition, the ipsilateral information received from the other hemisphere does not match with typical real-world regularities and may thus not trigger integration. Conversely, when stimuli are coherent, interhemispheric transfer of information may be critical for facilitating integration across visual fields. This idea is consistent with previous studies showing increased inter-hemispheric connectivity when object or word information needs to be integrated across visual hemifields (Mima et al., 2001; Stephan et al., 2007). Motion-selective hMT is also a conceivable candidate for integration-related feedback. The region not only showed enhanced representations for spatiotemporally coherent stimuli, but also had representations for both contralateral and ipsilateral stimuli. The hMT's sensitivity to motion (Watson et al., 1993; Tootell et al., 1995), and its bilateral visual representation (Cohen et al., 2019) make it suited for integrating coherent motion patterns across hemifields. Although speculative at this point, scene-selective and motion-selective cortical areas may jointly generate adaptive feedback signals that combine information about coherent scene content (analyzed in MPA and PPA) and coherent motion patterns (hMT). Future studies need to map out cortico-cortical connectivity during spatial integration to test this idea.

Our results inform theories about the functional role of alpha rhythms in cortex. Alpha is often considered an idling rhythm (Pfurtscheller et al., 1996; Romei et al., 2008), a neural correlate of active suppression (Jensen and Mazaheri, 2010; Haegens et al., 2011; Clayton et al., 2015), or a correlate of working memory maintenance (Jokisch and Jensen, 2007; de Vries et al., 2020). More recently, alpha rhythms were associated with an active role in cortical feedback processing (Bastos et al., 2015, 2020; Michalareas et al., 2016; Clayton et al., 2018). Our results highlight that alpha dynamics not only modulate feedforward processing, but encode stimulus-specific information. Our findings thus invite a new conceptualization of alpha dynamics, where alpha rhythms are critically involved in routing feedback-related information across the visual cortical hierarchy (Clark, 2013; Xie et al., 2020; Kaiser, 2022; Hu and Yu, 2023). An important remaining

question is whether the feedback itself traverses in alpha rhythms or whether the feedback initiates upstream representations that themselves fluctuate in alpha rhythms (Alamia and VanRullen, 2019; Lozano-Soldevilla and VanRullen, 2019). The absence of a correspondence of alpha-band representations and regional activity in scene-selective cortex may suggest that it is not the feedback itself that is rhythmic, but alpha dynamics in scene-selective cortex may also be weaker – or to some extent initiated for both coherent and incoherent stimuli. More studies are needed to dissociate between the rhythmic nature of cortical feedback and the representations it instills in early visual cortex.

The increased involvement of alpha rhythms in coding the coherent visual stimuli was accompanied by an absence of concurrent representations in the gamma band. A potential reason for the absence of decoding from feedforward-related gamma activity in the coherent condition is that feedforward representations were efficiently suppressed by accurate top-down predictions (Rao and Ballard, 1999). In our experiment, the video stimuli were presented for a relatively long time, without many rapid or unexpected visual events, potentially silencing feedforward propagation in the gamma range. We also did not find a correspondence between gamma dynamics and regional fMRI activity. Despite a general difficulty in linking high-frequency EEG activity and fMRI signals, another reason may be that, unlike the alpha dynamics, the gamma dynamics were relatively broadband and did not reflect a distinct neural rhythm (see fig. S2).

Our findings can be linked to theories of predictive processing that view neural information processing as a dynamic exchange of sensory feedforward signals and predictions stemming from higher-order areas of cortex (Rao and Ballard, 1999; Friston, 2005; Clark, 2013). On this view, feedback signals arising during stimulus integration are conceptualized as predictions about sensory input derived from spatially and temporally coherent contralateral input. In our paradigm, feedback signals can be conceptualized as predictions for the contralateral input generated from the spatiotemporally coherent ipsilateral input. A challenge for predictive coding theories is that it requires a strict separation of the feedforward sensory input and the predictive feedback. Our results indicate a compelling solution through multiplexing of feedforward and feedback information in dedicated frequency-specific channels (van Kerkoerle et al., 2014; Fries, 2015; Michalareas et al., 2016; Bastos et al., 2020) in human cortex. It will be interesting to see whether similar frequency-specific correlates of predictive processing are unveiled in other brain systems in the future.

Our findings further pave the way towards researching integration processes under various task demands. In our experiments, we engaged participants in an unrelated fixation task, capitalizing on automatically triggered integration processes for spatiotemporally coherent stimuli. Such automatic integration is well in line with phenomenological experience: The coherent video stimuli – but not the incoherent stimuli – strongly appear as coherent visual events that happen behind an occlude. Future studies should investigate how integration effects vary when participants are required to engage with the stimuli that need to be integrated. It will be particularly interesting to see whether tasks that require more global or local analysis are related to different degrees of integration, and different rhythmic codes in the brain. Such future studies could also set out to determine the critical features that enable integration and what is integrated. As our incoherent stimuli were designed to be incoherent along many different dimensions (e.g., low- and mid-level visual features, categorical content, and motion patterns), a comprehensive mapping of how these dimensions independently contribute to integration is needed. Future studies could thereby delineate how integration phenomenologically depends on the coherence of these candidate visual features.

More generally, our results have general implications for understanding and modelling feedforward and feedback information flows in neural systems. Processes like stimulus integration that are classically conceptualized as solvable in pure feedforward cascades may be

more dynamic than previously thought. Discoveries like ours arbitrate between competing theories that either stress the power of feedforward hierarchies (Riesenhuber and Poggio, 1999; DiCarlo and Cox, 2007) or emphasize the critical role of feedback processes (Rao and Ballard, 1999; Friston, 2005). They further motivate novel approaches to computational modelling that capture the visual system's abundant feedback connectivity (Kietzmann et al., 2019; Kreiman and Serre, 2020; Lindsay, 2021).

Taken together, our results reveal feedback in the alpha frequency range across the visual hierarchy as a key mechanism for integrating spatiotemporally coherent information in naturalistic vision. This strongly supports an active conceptualization of the visual system, where top-down projections are critical for the construction of coherent and unified visual percepts from fragmented sensory information.

Materials & Methods

Participants

Forty-eight healthy adults (gender: 12 M/36 F, age: 21.1 ± 3.8 years) participated in the EEG experiment, and another thirty-six (gender: 17 M/19 F, age: 27.3 ± 2.5 years) participated in the fMRI experiment. Sample size resulted from convenience sampling, with the goal of exceeding n = 34 in both experiments (i.e., exceeding 80% power for detecting a medium effect size of d = 0.5 in a two-sided *t*-test). All participants had normal or corrected-to-normal vision, and had no history of neurological/psychiatric disorders. They all signed informed consent before the experiment, and they were compensated for their time with partial course credit or cash. The EEG protocol was approved by the ethical committee of the Department of Psychology at the University of York, and the fMRI protocol was approved by the ethical committee of the Department of Psychology at Freie Universität Berlin. All experimental protocols were in accordance with the Declaration of Helsinki.

Stimuli and design

The stimuli and design were identical for the EEG and fMRI experiments unless stated otherwise.

Eight short video clips (3 s each; Fig. 1A) depicting everyday situations (e.g., a train driving past the observer; a view of red mountains; a view of waves crashing on the coast; first-person perspective of walking along a forest path; an aerial view of a motorway toll station; a group of zebras grazing on a prairie; first-person perspective of skiing in the forest; first-person perspective of walking along a street) were used in the experiments. During the experiment, these original videos were presented on the screen through circular apertures right and left of central fixation. We manipulated four experimental conditions (right-only, left-only, coherent, and incoherent) by presenting the original videos in different ways (Fig. 1B). In the right- or left-only condition, we presented the videos through right or left aperture only. We also showed two matching segments from the same video in the coherent condition, while showed segments from two different videos in the incoherent condition, through both apertures. In the incoherent condition, the eight original videos were yoked into eight fixed pairs (see fig. S1), and each video was always only shown with its paired video. Thus, there were a total of 32 unique video stimuli (8 for each condition). The diameter of each aperture was 6° visual angle, and the shortest distance between the stimulation and central fixation was 2.64° visual angle. The borders of the apertures were slightly smoothed. The central fixation dot subtended 0.44° visual angle. We selected our videos to be diverse in content (e.g., natural versus manmade) and motion (e.g., natural versus camera motion). This was done to maximize the contrast between the coherent and incoherent video stimuli. For assessing motion, we quantified the motion energy for each video stimulus using Motion Energy Analysis (MEA) software (https://psync.ch/mea/). We did not find any significant

between-condition differences (comparison on the means of motion energy: right- vs left-only, $t(7) = 0.36$, $p = 0.728$, coherent vs incoherent, $t(7) = 0.03$, $p = 0.976$; comparison on the SDs of motion energy: right- vs left-only, $t(7) = 1.08$, $p = 0.316$, coherent vs incoherent, $t(7) = 1.13$, $p = 0.294$). Although this suggests there was no difference in overall motion between conditions, there are many other candidates for critical differences between conditions, which will have to be evaluated in future studies.

The experiments were controlled through MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). In each trial, a fixation dot was first shown for 500 ms, after which a unique video stimulus was displayed for 3,000 ms. During the video stimulus playback, the color of the fixation changed periodically (every 200 ms) and turned either green or yellow at a single random point in the sequence (but never the first or last point). After every trial, a response screen prompted participants to report whether a green or yellow fixation dot was included in the sequence. Participants were instructed to keep central fixation during the sequence so they would be able to solve this task accurately. In both experiments, participants performed the color discrimination with high accuracy (EEG: 93.28 ± 1.65% correct; fMRI: 91.44 ± 1.37% correct), indicating that they indeed focused their attention on the central task. There were no significant differences in behavioral accuracy and response time (RT) between the coherent and incoherent conditions in both the EEG and fMRI experiments (Accuracy-EEG, $t(47) = 1.07$, $p = 0.29$; Accuracy-fMRI, $t(35) = 0.20$, $p = 0.85$; RT-fMRI, $t(35) = 0.41$, $p = 0.69$). Note that response times were not recorded in the EEG experiment. The mean accuracy and RT for each condition in both experiments are listed in table S1. In the EEG experiment, the next trial started once the participant's response was received. In the fMRI experiment, the response screen stayed on the screen for 1,500 ms, irrespective of participants' response time. An example trial is shown in Fig. 1C.

In the EEG experiment, each of the 32 unique stimuli was presented 20 times, resulting in a total of 640 trials, which were presented in random order. In the fMRI experiment, participants performed 10 identical runs. In each run, each unique stimulus was presented twice, in random order. Across the 10 runs, this also resulted in a total of 640 trials. The extensive repetition of the incoherent combinations may lead to some learning of the inconsistent stimuli in our experiment (den Ouden et al., 2010). However, such learning would if anything lead to an underestimation of the effects: Learning of the incoherent combination would ultimately also lead to an integration of the incoherent stimuli and thus create similar – albeit weaker – neural signatures of integration as found in the coherent condition.

To make sure that our intuition about the coherence of video stimuli in the coherent and incoherent conditions is valid, we conducted an additional behavioral experiment on 10 participants (gender: 4 M/6 F, age: 24.8 ± 2.5 years). In the experiment, we presented each of the coherent and incoherent video stimuli once. After each trial, we asked the participants to rate the degree of unified perception of the stimulus on a 1-5 scale. We found that the rating of coherent stimuli was higher than the rating of incoherent stimuli (mean ratings for coherent stimuli: 4.1–4.6, mean ratings for incoherent stimuli: 1.2–1.6; $t(9) = 36.66$, $p < 0.001$), showing that the coherent stimuli were indeed rated as more coherent than the incoherent ones.

To assess the general fixation stability for our paradigm, we collected additional eye-tracking data from six new participants (see fig. S10 for details). We calculated the mean and standard deviation (SD) of the horizontal and vertical eye movement across time (0–3 s) in each trial and then averaged the mean and SD values across trials separately for each condition. For all participants, we found means of eye movement lower than 0.3 degrees, and SDs of eye movement lower than 0.2 degrees, indicating stable central fixation (see fig. S10A). In addition, participants did not disengage from fixating after the target color was presented (see fig. S10B).

EEG recording and preprocessing

EEG signals were recorded using an ANT Waveguard 64-channel system and a TSMi REFA amplifier, with a sample rate of 1000 Hz. The electrodes were arranged according to the standard 10-10 system. EEG data preprocessing was performed using FieldTrip (Oostenveld et al., 2011). The data were first band-stop filtered to remove 50Hz line noise and then band-pass filtered between 1 Hz and 100 Hz. The filtered data were epoched from -500 ms to 4,000 ms relative to the onset of the stimulus, re-referenced to the average over the entire head, downsampled to 250 Hz, and baseline corrected by subtracting the mean pre-stimulus signal for each trial. After that, noisy channels and trials were removed by visual inspection, and the removed channels (2.71 ± 0.19 channels) were interpolated by the mean signals of their neighboring channels. Blinks and eye movement artifacts were removed using independent component analysis (ICA) and visual inspection of the resulting components.

EEG power spectrum analysis

Spectral analysis was performed using Fieldtrip. Powerspectra were estimated between 8 and 70 Hz (from alpha to gamma range), from 0 to 3,000 ms (i.e., the period of stimulus presentation) on the preprocessed EEG data, separately for each trial and each channel. A single taper with a Hanning window was used for the alpha band (8–12 Hz, in steps of 1 Hz) and the beta band (13–30 Hz, in steps of 2 Hz), and the discrete prolate spheroidal sequences (DPSS) multitaper method with ±8 Hz smoothing was used for the gamma band (31–70 Hz, in steps of 2 Hz).

EEG decoding analysis

To investigate whether the dynamic integration of information across the visual field is mediated by oscillatory activity, we performed multivariate decoding analysis using CoSMoMVPA (Oosterhof et al., 2016) and LIBSVM (Chang and Lin, 2011). In this analysis, we decoded between the eight video stimuli using patterns of spectral power across channels, separately for each frequency band (alpha, beta, and gamma) and each condition. Specifically, for each frequency band, we extracted the power of the frequencies included in that band (e.g., 8–12 Hz for the alpha band) across all channels from the power spectra, and then used the resulting patterns across channels and frequencies to classify the eight video stimuli in each condition. For all classifications, we used linear support vector machine (SVM) classifiers to discriminate the eight stimuli in a ten-fold cross-validation scheme. For each classification, the data were allocated to 10 folds randomly, and then an SVM classifier was trained on data from 9 folds and tested on data from the left-out fold. The classification was done repeatedly until every fold was left out once, and accuracies were averaged across these repetitions. The amount of data in the training set was always balanced across stimuli. For each classification, a maximum of 144 trials (some trials were removed during preprocessing) were included in the training set (18 trials for each stimulus) and 16 trials were used for testing (2 trials for each stimulus). Before classification, principal component analysis (PCA) was applied to reduce the dimensionality of the data (Chen et al., 2022). Specifically, for each classification, PCA was performed on the training data, and the PCA solution was projected onto the testing data. For each PCA, we selected the set of components that explained 99% of the variance of the training data. As a result, we obtained decoding accuracies for each frequency band and each condition, which indicated how well the video stimuli were represented in frequency-specific neural activity. We first used a one-sample *t*-test to investigate whether the video stimuli could be decoded in each condition and each frequency band. We also performed a 2 condition (coherent, incoherent) × 3 frequency (alpha, beta, gamma) two-way ANOVA, and post-hoc paired *t*-tests (FDR-corrected across frequencies; $p_{\text{corrected}} < 0.05$) to compare the decoding differences between coherent and incoherent conditions separately for each frequency band. The comparisons of right- and left-only conditions were conducted using the same approaches. To track where the effects appeared across a continuous frequency space,

we also decoded between the eight stimuli at each frequency from 8 to 70 Hz using a sliding window approach with a 5-frequency resolution (see fig. S2).

fMRI recording and processing

MRI data were acquired using a 3T Siemens Prisma scanner (Siemens, Erlangen, Germany) equipped with a 64-channel head coil. T2*-weighted BOLD images were obtained using a multiband gradient-echo echo-planar imaging (EPI) sequence with the following parameters: multiband factor = 3, TR = 1,500 ms, TE = 33 ms, FOV = 204 × 204 mm^2, voxel size = 2.5 × 2.5 × 2.5 mm^3, 70° flip angle, 57 slices, 10% interslice gap. Field maps were also obtained with a double-echo gradient echo field map sequence (TR = 545 ms, TE1/TE2 = 4.92 ms/7.38 ms) to correct for distortion in EPI. In addition, a high-resolution 3D T1-weighted image was collected for each participant (MPRAGE, TR = 1,900 ms, TE = 2.52 ms, TI = 900 ms, 256 × 256 matrix, 1 × 1 × 1 mm^3 voxel, 176 slices).

MRI data were preprocessed using MATLAB and SPM12 (https://www.fil.ion.ucl.ac.uk/spm/). Functional data were first corrected for geometric distortion with the SPM FieldMap toolbox (Hutton et al., 2002) and realigned for motion correction. In addition, individual participants' structural images were coregistered to the mean realigned functional image, and transformation parameters to MNI standard space (as well as inverse transformation parameters) were estimated.

The GLMsingle Toolbox (Prince et al., 2022) was used to estimate the fMRI responses to the stimulus in each trial based on realigned fMRI data. To improve the accuracy of trial-wise beta estimations, a three-stage procedure was used, including identifying an optimal hemodynamic response function (HRF) for each voxel from a library of 20 HRFs, denoising data-driven nuisance components identified by cross-validated PCA, and applying fractional ridge regression to regularize the beta estimation on a single voxel basis. The resulting single-trial betas were used for further decoding analyses.

fMRI regions of interest definition

fMRI analyses were focused on seven regions of interest (ROIs). We defined three scene-selective areas – occipital place area (OPA; also termed transverse occipital sulcus (Grill-Spector, 2003; Dilks et al., 2013)), medial place area (MPA; also termed retrosplenial cortex (Maguire, 2001; Epstein and Baker, 2019)), and parahippocampal place area (PPA (Epstein et al., 1999)) – from a group functional atlas (Julian et al., 2012); and three early visual areas – V1, V2, and V3, as well as motion-selective hMT/V5, – from a probabilistic functional atlas (Rosenke et al., 2021). All ROIs were defined in MNI space and separately for each hemisphere, and then transformed into individual-participant space using the inverse normalization parameters estimated during preprocessing.

fMRI ROI decoding analysis

To investigate how the video stimuli were processed in different visual regions, we performed multivariate decoding analysis using CoSMoMVPA and LIBSVM. For each ROI, we used the beta values across all voxels included in the region to decode between the eight video stimuli, separately for each condition. Leave-one-run-out cross-validation and PCA were used to conduct SVM classifications. For each classification, there were 144 trials (18 for each stimulus) in the training set and 16 trials (2 for each stimulus) in the testing set. For each participant, we obtained a 4 condition × 14 ROI (7 ROIs by 2 hemispheres) decoding matrix. Results were averaged across hemispheres, as we consistently found no significant inter-hemispheric differences (condition × hemisphere and condition × region × hemisphere interaction effects) in a 2 condition (coherent, incoherent) × 7 region (V1, V2, V3, hMT, OPA, MPA, PPA) × 2 hemisphere (left, right) three-way ANOVA test. We first tested whether the video stimuli were decodable in each condition and each region using one-sample *t*-tests (FDR-corrected across regions; $p_{corrected} < 0.05$). To further

investigate the integration effect, we used paired *t*-tests to compare the decoding difference between coherent and incoherent conditions in different regions. For the right- and left-only conditions, we averaged the decoding results in a contralateral versus ipsilateral fashion (e.g., left stimulus, right brain region was averaged with right stimulus, left brain region to obtain the contralateral decoding performance).

fMRI searchlight decoding analysis

To further investigate the whole-brain representation of video stimuli, we performed searchlight decoding analyses using CoSMoMVPA and LIBSVM. The single-trial beta maps in the native space were first transformed into the MNI space using the normalization parameters estimated during preprocessing. For the searchlight analysis, we defined a sphere with a radius of five voxels around a given voxel and then used the beta values of the voxels within this sphere to classify the eight video stimuli in each condition. For the left- and right-only conditions, the decoding analysis was performed separately for each hemisphere. The decoding parameters were identical to the ROI decoding analysis. The resulting searchlight maps were subsequently smoothed with a Gaussian kernel (FWHM = 6 mm). To investigate how well the video stimuli in each condition were represented across the whole brain, we used one-sample *t*-tests to compare decoding accuracies against chance separately for each condition (GRF correction, voxel-level $p < 0.005$, cluster-extent $p < 0.05$). To investigate the integration effect in the whole brain, we used paired *t*-tests to compare the differences in decoding accuracy between coherent and incoherent conditions and performed multiple comparisons correction within the voxels showing significant decoding for either coherent or incoherent stimuli (GRF correction, voxel-level $p < 0.005$, cluster-extent $p < 0.05$).

EEG-fMRI fusion with representational similarity analysis

To investigate the relationship between the frequency-specific effects obtained in the EEG and the spatial mapping obtained in the fMRI, we performed EEG-fMRI fusion analysis (Cichy et al., 2014; Cichy and Oliva, 2020). This analysis can be used to compare neural representations of stimuli as characterized by EEG and fMRI data to reveal how the representations correspond across brain space and spectral signatures. Specifically, we first calculated representational dissimilarity matrices (RDMs) using pair-wise decoding analysis for EEG and fMRI data, respectively. For the EEG power spectra, in each frequency band, we decoded between each pair of eight video stimuli using the oscillatory power of the frequencies included in the frequency band, separately for each condition; for the fMRI data, in each ROI, we classified each pair of eight stimuli using the response patterns of the region, separately for each condition. Decoding parameters were otherwise identical to the 8-way decoding analyses (see above). In each condition, we obtained a participant-specific EEG RDM (8 stimuli × 8 stimuli) in each frequency band and a participant-specific fMRI RDM (8 stimuli × 8 stimuli) in each ROI. Next, we calculated the similarity between EEG and fMRI RDMs for each condition; this was done by correlating all lower off-diagonal entries between the EEG and fMRI RDMs (the diagonal was always left out). To increase the signal-to-noise ratio, we first averaged fMRI RDMs across participants and then calculated the *Spearman*-correlation between the averaged fMRI RDM for each ROI with the participant-specific EEG RDM for each frequency. As a result, we obtained a 4 condition × 3 frequency × 14 ROI fusion matrix for each EEG participant. For the coherent and incoherent conditions, the results were averaged across hemispheres, as no condition × hemisphere, no condition × region × hemisphere, and no condition × frequency × hemisphere interaction effects were found in a 2 condition (coherent, incoherent) × 7 region (V1, V2, V3, hMT, OPA, MPA, PPA) × 2 hemisphere (left, right) × 3 frequency (alpha, beta, gamma) four-way ANOVA test. We first used one-sample *t*-tests to test the fusion effect in each condition (FDR-corrected across regions; $p_{\text{corrected}} < 0.05$) and each frequency-region combination, and then used a 2 condition × 3 frequency × 7 region three-way ANOVA to compare the frequency-region

correspondence between coherent and incoherent conditions. As we found a significant condition × frequency × region interaction effect, we further performed a 2 condition × 7 region ANOVA and paired *t*-tests (FDR-corrected across regions; $p_{corrected} < 0.05$) to compare frequency-region correspondence between coherent and incoherent conditions separately for each frequency. For the right- and left-only conditions, we averaged the fusion results across two conditions separately for contralateral and ipsilateral presentations and then compared contralateral and ipsilateral presentations using the same approaches we used for the comparisons of coherent and incoherent conditions (see fig. S9).

References

Alamia A, VanRullen R (2019) Alpha oscillations and traveling waves: Signatures of predictive coding? PLOS Biol 17:e3000487.

Bastos AM, Lundqvist M, Waite AS, Kopell N, Miller EK (2020) Layer and rhythm specificity for predictive routing. Proc Natl Acad Sci 117:31459–31469.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76:695–711.

Bastos AM, Vezoli J, Bosman CA, Schoffelen J-M, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. Neuron 85:390–401.

Bennett L, Melchers B, Proppe B (2020) Curta: A General-purpose High-Performance Computer at ZEDAT, Freie Universität Berlin. Available at: http://dx.doi.org/10.17169/refubium-26754.

Bilalić M, Lindig T, Turella L (2019) Parsing rooms: the role of the PPA and RSC in perceiving object relations and spatial layout. Brain Struct Funct 224:2505–2524.

Block N (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. Behav Brain Sci 30:481–499.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol TIST 2:1–27.

Chen L, Cichy RM, Kaiser D (2022) Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. Cereb Cortex 32:3553–3567.

Cichy RM, Oliva A (2020) A M/EEG-fMRI Fusion Primer: Resolving Human Brain Responses in Space and Time. Neuron 107:772–781.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455–462.

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204.

Clavagnier S, Falchier A, Kennedy H (2004) Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. Cogn Affect Behav Neurosci 4:117–126.

Clayton MS, Yeung N, Cohen Kadosh R (2015) The roles of cortical oscillations in sustained attention. Trends Cogn Sci 19:188–195.

Clayton MS, Yeung N, Cohen Kadosh R (2018) The many characters of visual alpha oscillations. Eur J Neurosci 48:2498–2508.

Cohen D, Goddard E, Mullen KT (2019) Reevaluating hMT+ and hV4 functional specialization for motion and static contrast using fMRI-guided repetitive transcranial magnetic stimulation. J Vis 19:11.

Cohen MA, Dennett DC, Kanwisher N (2016) What is the Bandwidth of Perceptual Experience? Trends Cogn Sci 20:324–335.

de Vries IEJ, Slagter HA, Olivers CNL (2020) Oscillatory Control over Representational States in Working Memory. Trends Cogn Sci 24:150–162.

Dehaene S, Cohen L (2007) Cultural Recycling of Cortical Maps. Neuron 56:384–398.

den Ouden HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal Prediction Error Modulates Cortical Coupling. J Neurosci 30:3210–3219.

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cogn Sci 11:333–341.

Dilks DD, Julian JB, Paunov AM, Kanwisher N (2013) The Occipital Place Area Is Causally and Selectively Involved in Scene Perception. J Neurosci 33:1331–1336.

Epstein R, Harris A, Stanley D, Kanwisher N (1999) The Parahippocampal Place Area: Recognition, Navigation, or Encoding? Neuron 23:115–125.

Epstein RA, Baker CI (2019) Scene Perception in the Human Brain. Annu Rev Vis Sci 5:373–397.

Fries P (2015) Rhythms for Cognition: Communication through Coherence. Neuron 88:220–235.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc B Biol Sci 360:815–836.

Grill-Spector K (2003) The neural basis of object perception. Curr Opin Neurobiol 13:159–166.

Haegens S, Nácher V, Luna R, Romo R, Jensen O (2011) α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. Proc Natl Acad Sci 108:19377–19382.

Haynes J-D (2015) A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. Neuron 87:257–270.

Hu Y, Yu Q (2023) Spatiotemporal dynamics of self-generated imagery reveal a reverse cortical hierarchy from cue-induced imagery. BioRxiv:2023.01.25.525474.

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002) Image distortion correction in fMRI: a quantitative evaluation. NeuroImage 16:217–240.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4.

Jokisch D, Jensen O (2007) Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. J Neurosci 27:3244–3251.

Julian JB, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally defining regions of interest in the ventral visual pathway. NeuroImage 60:2357–2364.

Kaiser D (2022) Spectral brain signatures of aesthetic natural perception in the α and β frequency bands. J Neurophysiol 128:1501–1505.

Kaiser D, Cichy RM (2021) Parts and wholes in scene processing. J Cogn Neurosci 34:4–15.

Kaiser D, Häberle G, Cichy RM (2020) Cortical sensitivity to natural scene structure. Hum Brain Mapp 41:1286–1295.

Keller AJ, Roth MM, Scanziani M (2020) Feedback generates a second receptive field in neurons of the visual cortex. Nature 582:545–549.

Kietzmann TC, McClure P, Kriegeskorte N (2019) Deep Neural Networks in Computational Neuroscience. In: Oxford Research Encyclopedia of Neuroscience. Oxford University Press.

Kirchberger L, Mukherjee S, Self MW, Roelfsema PR (2023) Contextual drive of neuronal responses in mouse V1 in the absence of feedforward input. Sci Adv 9:eadd2498.

Kreiman G, Serre T (2020) Beyond the feedforward sweep: feedback computations in the visual cortex. Ann N Y Acad Sci 1464:222–241.

Lamme VA, Supèr H, Spekreijse H (1998) Feedforward, horizontal, and feedback processing in the visual cortex. Curr Opin Neurobiol 8:529–535.

Lindsay GW (2021) Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. J Cogn Neurosci 33:2017–2031.

Lozano-Soldevilla D, VanRullen R (2019) The Hidden Spatial Dimension of Alpha: 10-Hz Perceptual Echoes Propagate as Periodic Traveling Waves in the Human Brain. Cell Rep 26:374-380.e4.

Maguire E (2001) The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. Scand J Psychol 42:225–238.

Mannion DJ, Kersten DJ, Olman CA (2014) Regions of mid-level human visual cortex sensitive to the global coherence of local image patches. J Cogn Neurosci 26:1764–1774.

Markov NT, Ercsey-Ravasz M, Van Essen DC, Knoblauch K, Toroczkai Z, Kennedy H (2013) Cortical high-density counterstream architectures. Science 342:1238406.

Michalareas G, Vezoli J, van Pelt S, Schoffelen J-M, Kennedy H, Fries P (2016) Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. Neuron 89:384–397.

Mima T, Oluwatimilehin T, Hiraoka T, Hallett M (2001) Transient Interhemispheric Neuronal Synchrony Correlates with Object Recognition. J Neurosci 21:3942–3948.

Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E (2015) Contextual Feedback to Superficial Layers of V1. Curr Biol 25:2690–2695.

Muckli L, Petro LS (2013) Network interactions: non-geniculate input to V1. Curr Opin Neurobiol 23:195–201.

Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Oosterhof NN, Connolly AC, Haxby JV (2016) CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Front Neuroinformatics 10:27.

Papale P, Wang F, Morgan AT, Chen X, Gilhuis A, Petro LS, Muckli L, Roelfsema PR, Self MW (2022) Feedback brings scene information to the representation of occluded image regions in area V1 of monkeys and humans. BioRxiv:2022.11.21.517305.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442.

Pfurtscheller G, Stancák A, Neuper Ch (1996) Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review. Int J Psychophysiol 24:39–46.

Prince JS, Charest I, Kurzawski JW, Pyles JA, Tarr MJ, Kay KN (2022) Improving the accuracy of single-trial fMRI response estimates using GLMsingle. eLife 11:e77599.

Ragni F, Lingnau A, Turella L (2021) Decoding category and familiarity information during visual imagery. NeuroImage 241:118428.

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025.

Romei V, Brodbeck V, Michel C, Amedi A, Pascual-Leone A, Thut G (2008) Spontaneous fluctuations in posterior α-band EEG activity reflect variability in excitability of human visual areas. Cereb Cortex 18:2010–2018.

Rosenke M, van Hoof R, van den Hurk J, Grill-Spector K, Goebel R (2021) A probabilistic functional atlas of human occipito-temporal visual cortex. Cereb Cortex 31:603–619.

Salin PA, Bullier J (1995) Corticocortical connections in the visual system: structure and function. Physiol Rev 75:107–154.

Smith FW, Muckli L (2010) Nonstimulated early visual areas carry information about surrounding context. Proc Natl Acad Sci 107:20099–20103.

Stephan KE, Marshall JC, Penny WD, Friston KJ, Fink GR (2007) Interhemispheric Integration of Visual Processing during Task-Driven Lateralization. J Neurosci 27:3512–3522.

Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW (1995) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J Neurosci 15:3215–3230.

van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M-A, Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci 111:14332–14341.

Vetter P, Bola Ł, Reich L, Bennett M, Muckli L, Amedi A (2020) Decoding Natural Sounds in Early “Visual” Cortex of Congenitally Blind Individuals. Curr Biol 30:3039-3044.e2.

Vetter P, Smith FW, Muckli L (2014) Decoding Sound and Imagery Content in Early Visual Cortex. Curr Biol 24:1256–1262.

Watson JD, Myers R, Frackowiak RS, Hajnal JV, Woods RP, Mazziotta JC, Shipp S, Zeki S (1993) Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. Cereb Cortex 3:79–94.

Williams MA, Baker CI, Op de Beeck HP, Mok Shim W, Dang S, Triantafyllou C, Kanwisher N (2008) Feedback of visual object information to foveal retinotopic cortex. Nat Neurosci 11:1439–1445.

Winlove CIP, Milton F, Ranson J, Fulford J, MacKisack M, Macpherson F, Zeman A (2018) The neural correlates of visual imagery: A co-ordinate-based meta-analysis. Cortex 105:4–25.

Xie S, Kaiser D, Cichy RM (2020) Visual imagery and perception share neural representations in the alpha frequency band. Curr Biol 30:2621-2627.e5.

Acknowledgments

We thank Daniela Marinova and Alex Carter for their help in EEG data acquisition. We would also like to thank the HPC Service of ZEDAT, Freie Universität Berlin (Bennett et al., 2020), for computing time.

Funding:

L.C. is supported by a PhD stipend from the China Scholarship Council (CSC). R.M.C is supported by the Deutsche Forschungsgemeinschaft (DFG; CI241/1-1, CI241/3-1, CI241/7-1) and by a European Research Council (ERC) starting grant (ERC-2018-STG 803370). D.K. is supported by the DFG (SFB/TRR135 – INST162/567-1), an ERC starting grant (ERC-2022-STG 101076057), and "The Adaptive Mind" funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art.

Author contributions:

Conceptualization: L.C. and D.K.; Methodology: L.C. and D.K.; Software: L.C. and D.K.; Formal analysis: L.C.; Investigation: L.C. and D.K.; Resources: L.C. and D.K.; Data curation: L.C.; Writing–original draft: L.C. and D.K.; Writing–review & editing: L.C., R.M.C., and D.K.; Visualization: L.C.; Supervision: R.M.C. and D.K.; Project administration: R.M.C. and D.K.; Funding acquisition: R.M.C. and D.K.

Competing interests:

The authors declare that they have no competing interest.

Data and materials availability:

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Raw data used in the analyses are available at the following OSF repository: https://doi.org/10.17605/osf.io/57ryb. Processed data and code used in the analyses are available at the following OSF repository: https://doi.org/10.17605/osf.io/ca9rt.

Supplementary Materials for

Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision

Lixiang Chen *et al.*

*Corresponding author: lixiang.chen@fu-berlin.de (L.C.) or danielkaiser.net@gmail.com (D.K.)

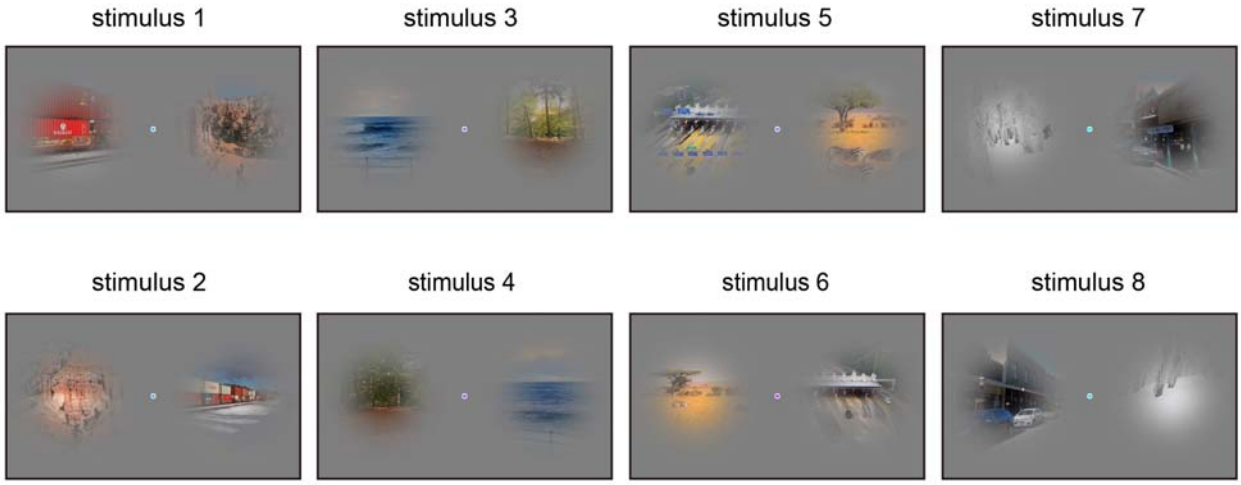

Fig. S1.

Still images from the incoherent video stimuli.

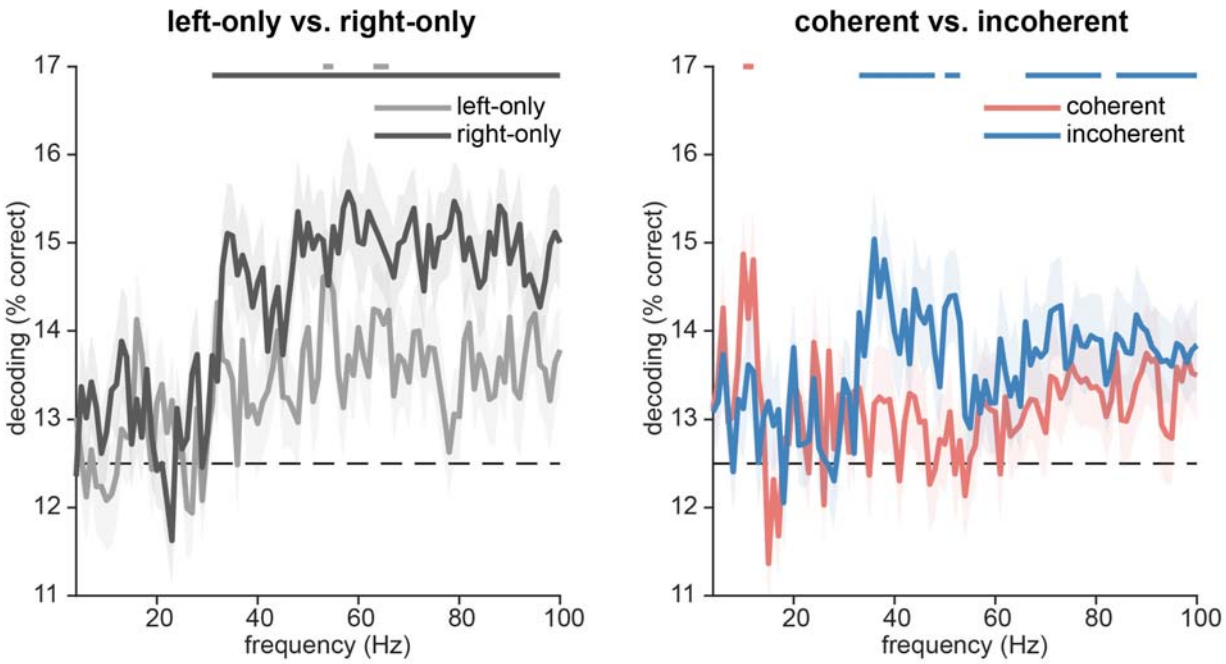

Fig. S2.

EEG frequency-resolved decoding analysis on spectral power patterns using sliding windows. We decoded between the eight video stimuli at each frequency from 4 to 100 Hz using a sliding window approach with a 5-frequency resolution, separately for each condition. The incoherent and single video stimuli were decodable from the γ frequency band, whereas coherent stimuli were decodable from the α frequency band. Line markers denote significant above-chance decoding ($p < 0.05$; FDR-corrected).

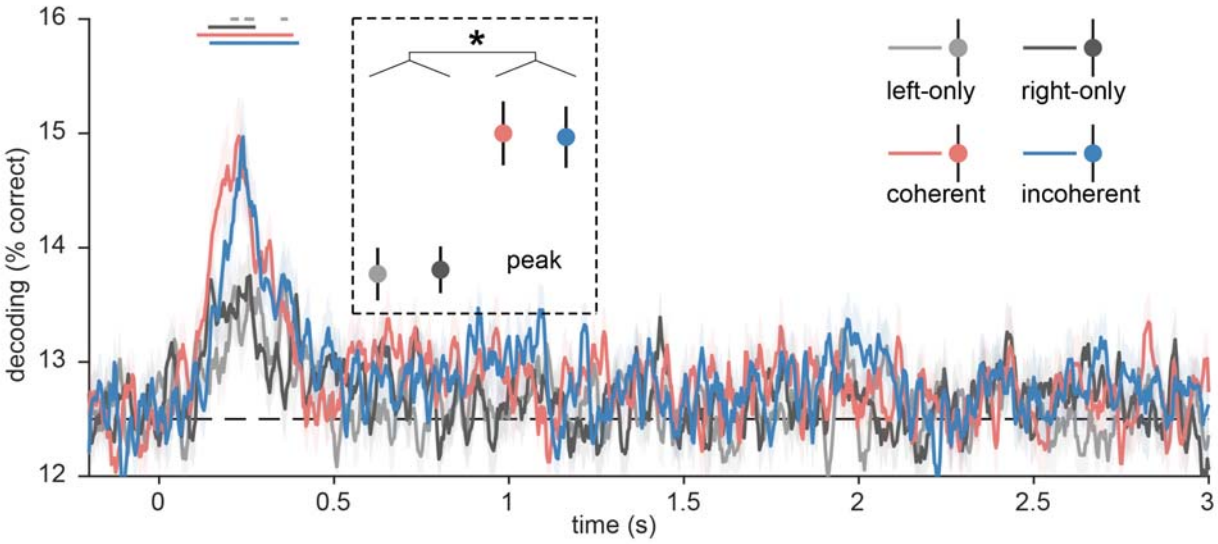

Fig. S3.

EEG time-resolved decoding on evoked response patterns. We performed decoding analysis on time-resolved broadband responses across channels to discriminate the eight video stimuli at each time from -200 ms to 3,000 ms relative to the onset of the stimulus, separately for each condition. The obtained decoding timeseries for each condition were smoothed by the moving average algorithm (6 time points). We extracted the peak decoding accuracy for each condition and then compared the decoding difference between conditions using paired t-tests. The results revealed a sustained representation of the video stimuli across the first 500 ms of processing, with stronger peak responses to two video conditions (coherent/incoherent) than to single video conditions (right-/left-only), but no differences between the coherent and incoherent conditions. Decoding onsets did not differ between the coherent and incoherent video stimuli (permutation test, $p = 0.176$). Error bars represent standard errors. Line markers denote significant above-chance decoding ($p < 0.05$; FDR-corrected). *: $p < 0.05$.

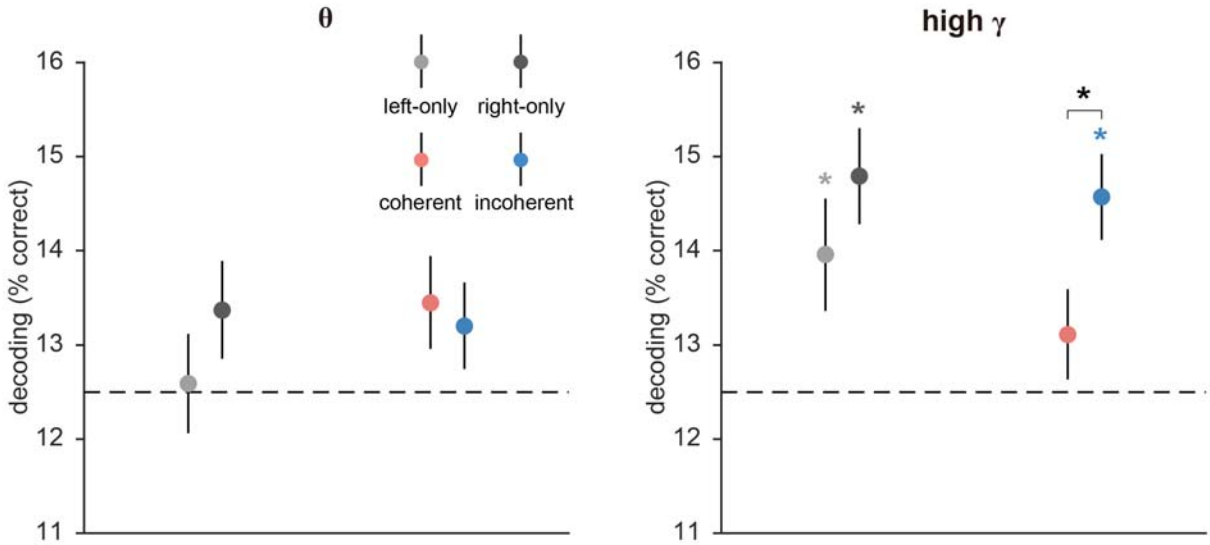

Fig. S4.

EEG decoding analysis on spectral power patterns separately for the theta (4–7 Hz) and high-gamma (71–100 Hz) frequency bands. For each frequency band (theta and high-gamma), we extracted the power of the frequencies included in that band across all channels from the power spectra, and then used the resulting patterns across channels and frequencies to classify the eight video stimuli in each condition. In the theta band, we did not find any significant above-chance decoding. In the high-gamma band, we found significant above-chance decoding for both single video stimuli and incoherent stimuli, but not for coherent stimuli. As in the 31-70 Hz gamma range, the incoherent stimuli were also decoded better than coherent stimuli. *: $p < 0.05$.

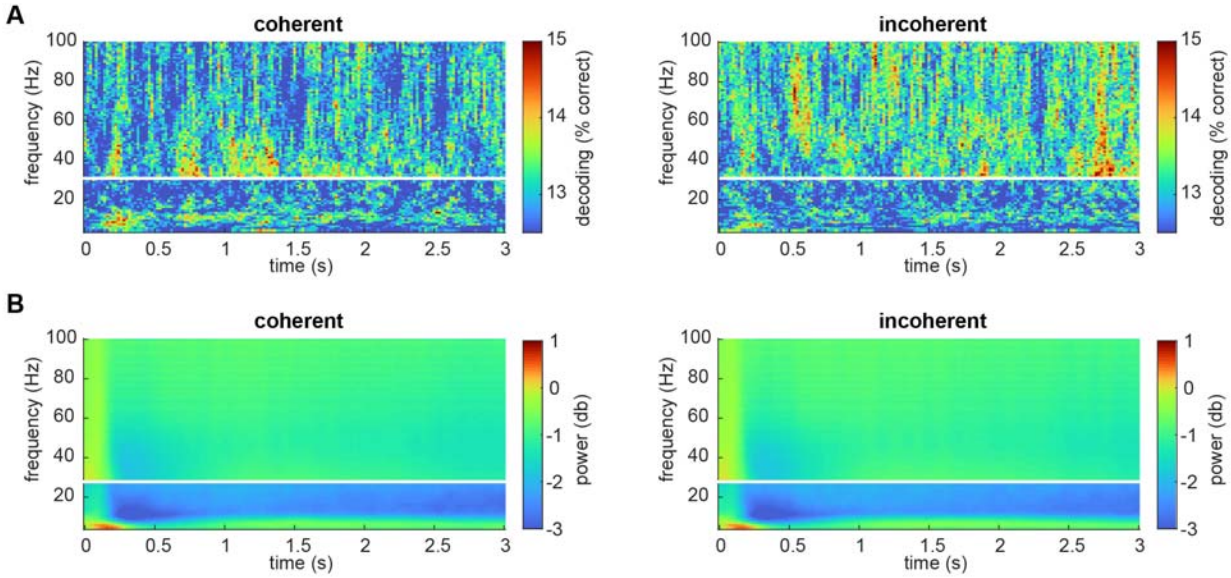

Fig. S5.

EEG time-frequency analysis. We first performed a time-frequency analysis to estimate the power at each frequency (4–100 Hz) and each time point (0–3 s) separately for each channel. As in the powerspectrum analysis (see Materials and Methods for details), we used single tapers for the low frequencies (4–30 Hz), and multitapers for high frequencies (31–100 Hz). **A)** We performed the decoding analysis at each time-frequency combination. We did not find significant differences in decoding performance between the coherent and incoherent conditions ($p < 0.05$; FDR-corrected; top two panels). The results suggested that we had insufficient statistical power for concurrently resolving the data across time and frequency, given the signal-to-noise ratio of our data. **B)** We transformed the power values to dB relative to the baseline to obtain the event-related spectral perturbation (ERSP). No significant differences were found in ERSP between the coherent and incoherent conditions ($p < 0.05$; FDR-corrected; bottom two panels). This suggests that the shift in representation from gamma to alpha dynamics is not accompanied by large-scale changes in the univariate spectral power over visual cortex.

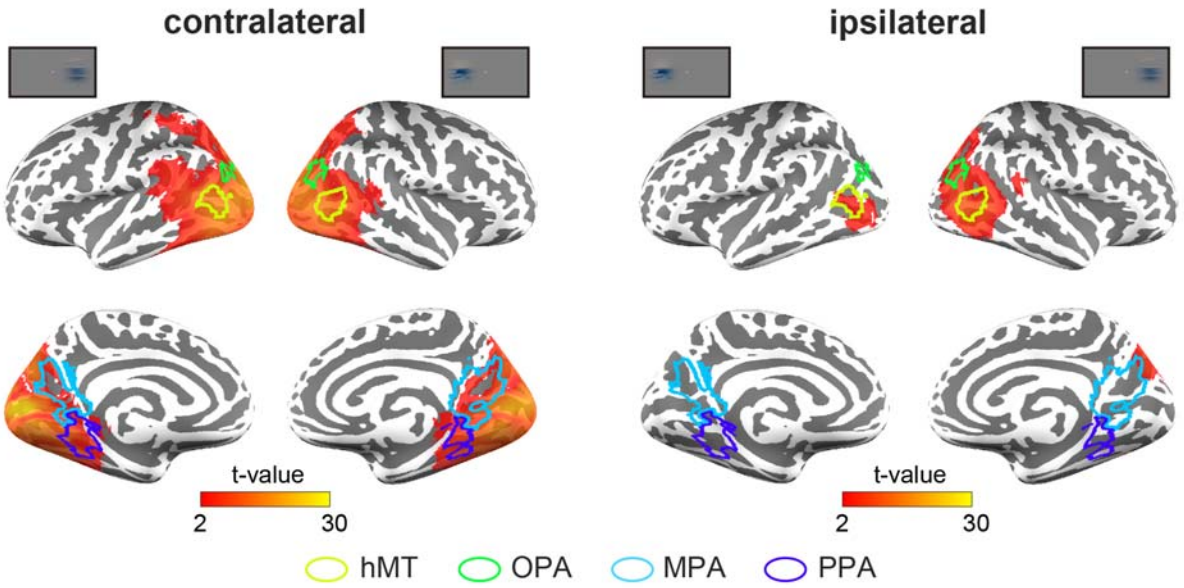

Fig. S6.

fMRI searchlight decoding analysis for right- and left-only conditions. Similar to the ROI decoding results, single video stimuli were decodable across the visual cortex in the contralateral hemisphere. In the ipsilateral hemisphere, the stimuli were primarily decodable in the parietal lobe including hMT. Multiple comparison correction was performed using GRF (voxel-level $p < 0.005$, cluster-extent $p < 0.05$).

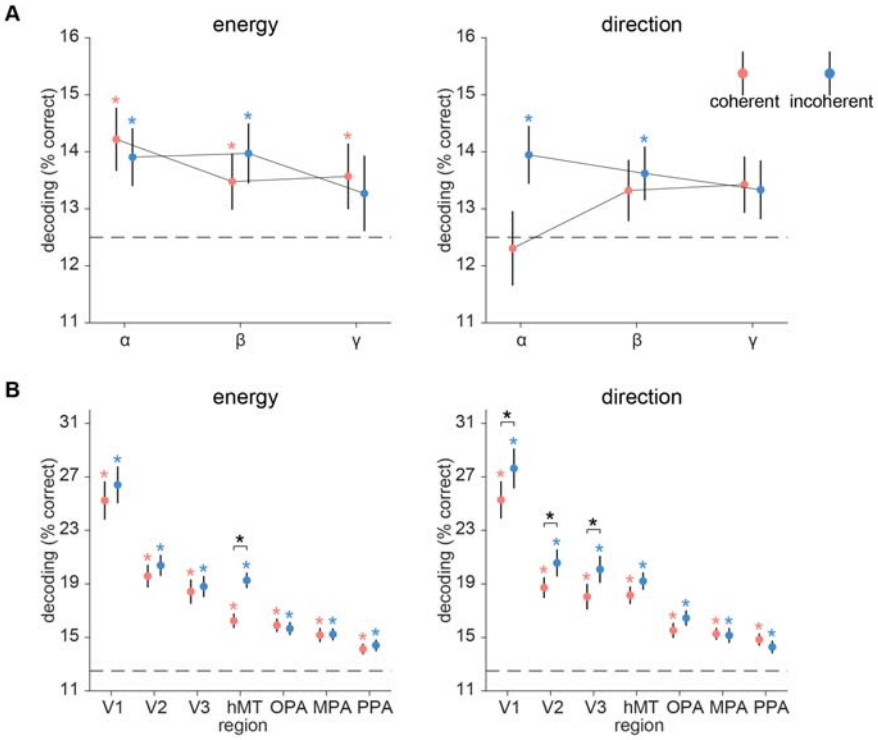

Fig. S7.

EEG and fMRI decoding analyses (grouping stimuli based on motion coherence). For each of the coherent and incoherent video stimuli, we first quantified inter-frame motion energy using the Motion Energy Analysis (MEA) software (https://psync.ch/mea/) separately for the left and right apertures, and then estimated its motion energy coherence by calculating the Pearson correlation between the left and right motion energy time-series. We split the stimulus set into two halves based on the values of motion energy coherence values. Eight stimuli that are more coherent in motion energy were grouped into the motion coherent group and the other 8 stimuli were grouped into the motion incoherent group. In addition, we also grouped the stimuli based on motion direction coherence. For each stimulus, we estimated inter-frame optical flow using the Computer Vision Toolbox implemented in MATLAB and estimated its motion direction coherence by calculating the Pearson correlation between the mean motion direction time-series of the left and right apertures. Equivalent as described above, based on the values of motion direction coherence we split half the stimuli into a motion direction coherent and a motion direction incoherent group. **A)** We performed EEG frequency-resolved decoding analysis between the 8 stimuli in each frequency band (alpha, beta, gamma), separately for the newly formed coherent and incoherent groups. For the motion energy grouping, we found significant decoding for coherent stimuli in the alpha, beta, and gamma frequency bands, and significant decoding for incoherent stimuli in the alpha and beta frequency bands. For the motion direction grouping, the incoherent stimuli were decodable in the alpha and beta bands. **B)** We performed fMRI ROI decoding analysis to classify the stimuli in each ROI (V1, V2, V3, hMT, OPA, MPA, PPA), separately for the newly formed coherent and incoherent groups. For the motion energy grouping, both coherent and incoherent stimuli were decodable in all seven ROIs, and the incoherent stimuli were decoded better than coherent stimuli in the hMT. For the motion direction grouping, both coherent and incoherent stimuli were decodable in all the ROIs, and the incoherent stimuli were more decodable than coherent stimuli in V1, V2, and V3. *: $p < 0.05$ (FDR-corrected).

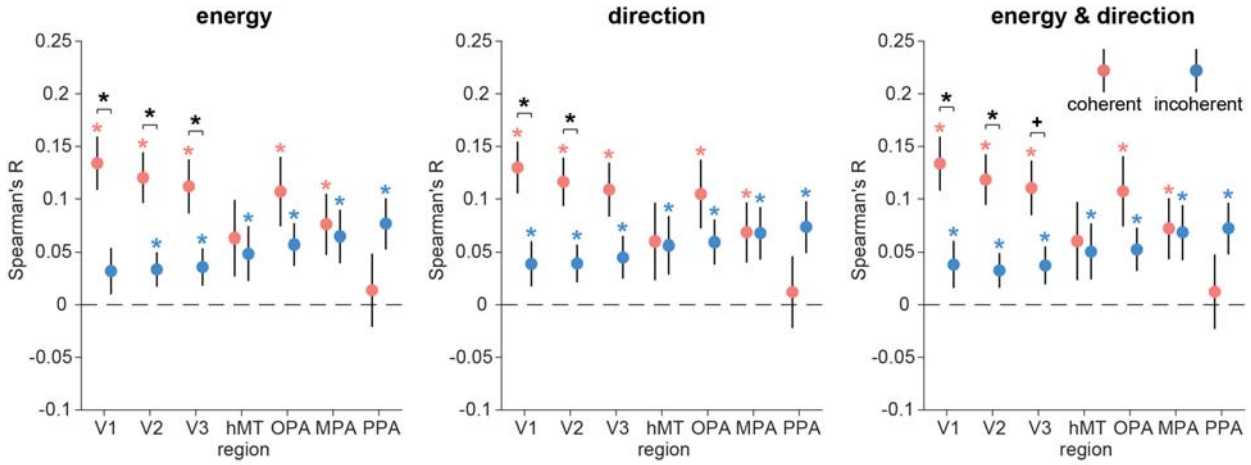

Fig. S8.

EEG-fMRI fusion analysis while controlling for motion coherence. For each of the coherent and incoherent video stimuli, we first quantified the inter-frame motion energy using the Motion Energy Analysis (MEA) software (https://psync.ch/mea/) separately for the left and right apertures, and then estimated its motion energy coherence by calculating the Pearson correlation between the left and right motion energy time-series. To construct the motion representational dissimilarity matrices (RDMs), we used the absolute difference in motion energy coherence as the distance between stimuli separately for coherent and incoherent conditions. In addition, we also constructed a motion RDM based on motion direction coherence. Specifically, for each stimulus, we estimated the inter-frame optical flow using the Computer Vision Toolbox (in MATLAB) and estimated its motion direction coherence by calculating the Pearson correlation between the mean motion direction time-series of the left and right apertures. Next, we constructed a motion direction RDM based on the absolute difference in motion direction coherence between stimuli separately for the coherent and incoherent conditions. In the EEG-fMRI fusion analysis, we calculated partial correlations between the participant-specific EEG RDMs in each frequency band (α, β, γ) and the group-averaged fMRI RDMs in each region (V1, V2, V3, hMT, OPA, MPA, PPA) that control for the motion energy RDM, motion direction RDM, or both motion RDMs, separately for the coherent and incoherent conditions. Similar to the main fusion results, we found that representations in the alpha band corresponded more strongly with representations in the early visual cortex when the videos were presented coherently, rather than incoherently, in all three analyses. Error bars represent standard errors. *: $p < 0.05$ (FDR-corrected), $^{+}$: $p < 0.05$ (uncorrected).

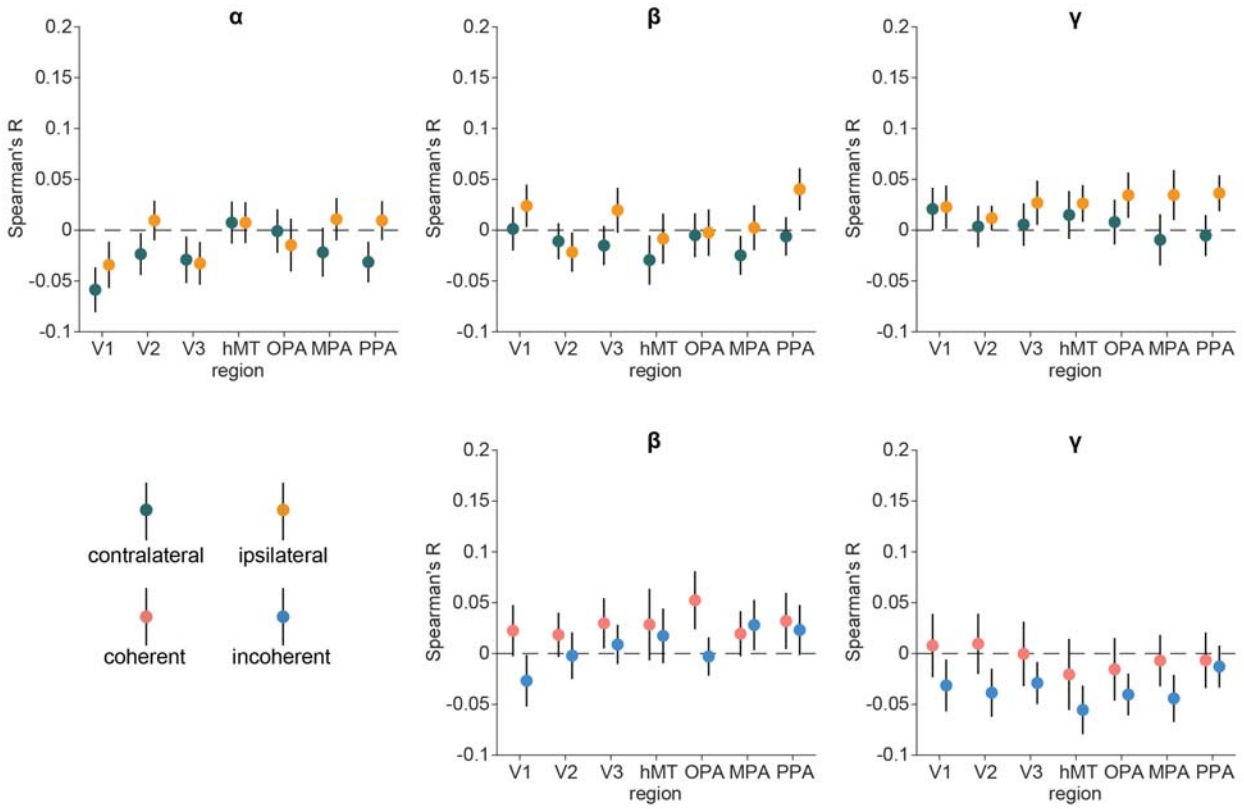

Fig. S9.

EEG-fMRI fusion analysis separately for each frequency band. For each condition, EEG representational dissimilarity matrices (RDMs) for each frequency band (α, β, γ) and fMRI RDMs for each region of interest (V1, V2, V3, hMT, OPA, MPA, PPA) were first obtained using pair-wise decoding analyses. To assess correspondences between spectral and regional representations, we calculated *Spearman*-correlations between the participant-specific EEG RDMs in each frequency band and the group-averaged fMRI RDMs in each region, separately for each condition. For the right- and left-only conditions, there was no significant correspondence between EEG responses in each frequency band and fMRI activations in each region, either for contralateral or ipsilateral presentations. For the coherent and incoherent conditions, there were no significant correspondences between β/γ responses and fMRI activations. Error bars represent standard errors.

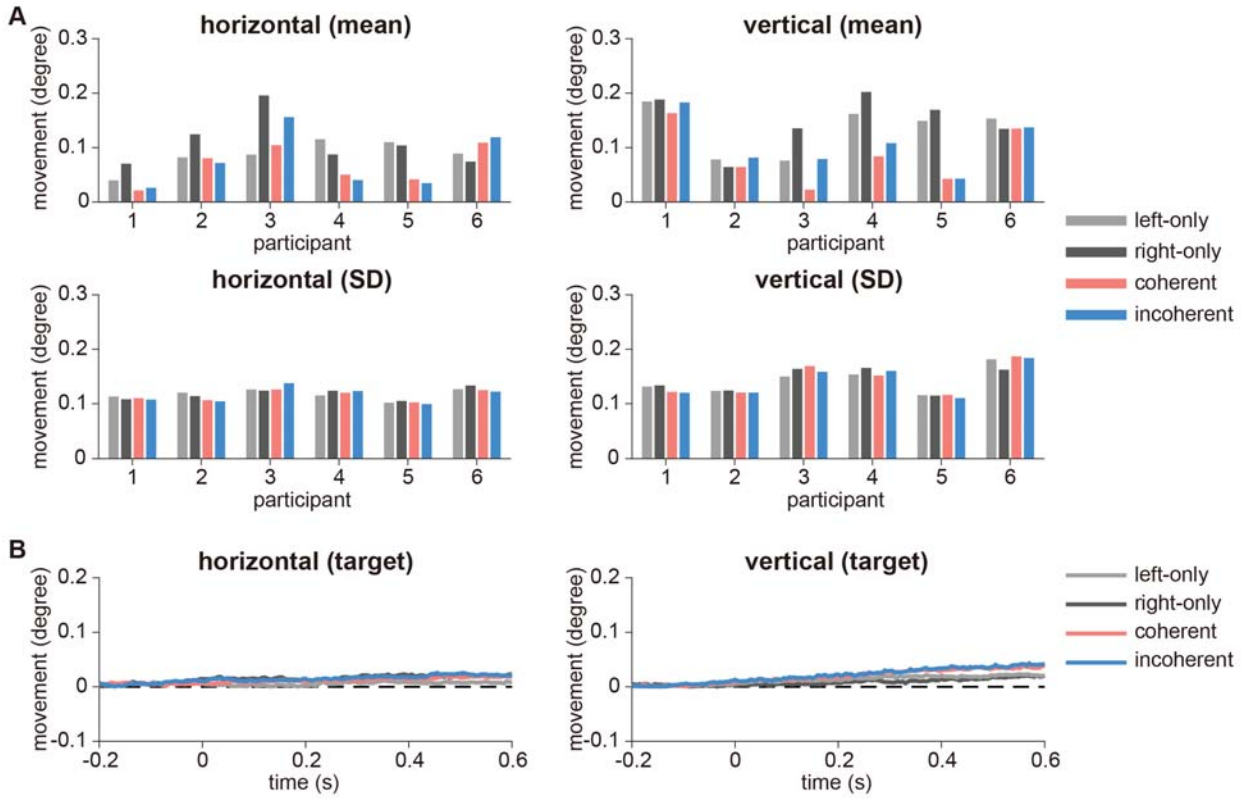

Fig. S10.

Eye-tracking data. Six participants (gender: 2 M/4 F, age: 26.5 ± 1.2 years) took part in the eye-tracking experiment using the same paradigm and the same stimuli as in the EEG and fMRI experiments. Eye movements were recorded monocularly (left eye) with an Eyelink 1000 Tower Mount (SR Research Ltd., Mississauga, Ontario, Canada) using the Psychophysics and Eyelink Toolbox extensions at 1000 Hz. Eye tracking data were segmented into epochs from -0.5 to 3.5 s relative to the onset of the stimulus, downsampled to 200 Hz, and baseline corrected. The data were then transformed from their original screen coordinate units (pixels) to visual angle units (degrees). **A)** To check fixation patterns during video presentation, we calculated the mean and standard deviation (SD) of the horizontal and vertical eye movement across time (0–3 s) in each trial and then averaged the mean and SD values across trials separately for each condition. For all participants, we found means of eye movement lower than 0.3 degrees (top two panels), and SDs of eye movement lower than 0.2 degrees (middle two panels), indicating stable central fixation. **B)** To determine whether eye movements occurred once the fixation color was detected, we extracted eye tracking data from -200 to 600 ms relative to the onset of the target color. We found no significant eye movement deviations after the target was presented, indicating that participants did not disengage from fixation after the target was presented.

Table S1.

Behavioral accuracy and response time for the color discrimination task in both EEG and fMRI experiments. Notes: Means ± standard errors (SE); N.A., not available.

	Left-only	Right-only	Coherent	Incoherent
Accuracy (EEG)	93.27 ± 1.70%	93.27 ± 1.58%	93.13 ± 1.66%	93.45 ± 1.71%
Accuracy (fMRI)	91.58 ± 1.18%	91.09 ± 1.19%	91.58 ± 1.24%	91.51 ± 1.21%
Response Time (EEG)	N.A.	N.A.	N.A.	N.A.
Response Time (fMRI)	521.9 ± 19.3 ms	521.7 ± 18.6 ms	525.2 ± 19.4 ms	526.8 ± 18.5 ms