Tracking visual rhythms: a concert of sensory and motor simulation

Aaron Kaltenmaier[1,2,4], Quirin Gehmacher[1,2], Peter Kok[2], Matthew H. Davis[3], Clare Press[1,2]

[1]Department of Experimental Psychology, UCL, UK

[2]Functional Imaging Laboratory, UCL, UK

[3]MRC Cognition and Brain Sciences Unit, University of Cambridge, UK

[4]Lead contact, aaron.kaltenmaier.22@ucl.ac.uk

ABSTRACT

Neural oscillations have been proposed to model external temporal structure by phase-coupling to environmental rhythms, thereby supporting adaptive perception. However, there is little evidence supporting these theories, particularly in the visual domain, and the underlying mechanisms remain unclear. Using MEG and a new empirical approach we addressed this issue. Participants attended 1.3 and 2 Hz visual displays of rotating Gabors and judged either the timing or content of these events. We show behaviourally-relevant rate-specific phase-coupling in motor structures to – and beyond – the visual rhythm specifically when judging temporal features of the display. We subsequently devised a rate-specific decoding measure to show that visual structures track anticipated, temporally-precise content regardless of task. This sensory simulation predicted the temporal tracking in motor structures. We consequently propose a mechanism by which automatic, temporally-specific sensory simulation yields an information envelope read out by motor areas when estimating temporal characteristics in our environment.

INTRODUCTION

Relevant sensory information is often embedded within predictable temporal structure. Music and speech provide prominent examples of quasi-rhythmic sensory streams that require temporal parsing before we can appreciate their content [1]. One popular view that has emerged from research into how the brain accomplishes this task is that neural oscillations temporally align with sensory temporal patterns to support perceptual processing [2-4]. Given the example of music, spectral peaks in neural activity thereby flexibly shift depending on the beat meter, such that moments of maximal neural sensitivity become aligned with likely peaks in sensory information [5]. This account has been supported by a range of 'oscillatory tracking' evidence - greater neural phase consistency in the frequency band corresponding to the presented information [6-8]. Recent advances have additionally been made to dissociate these supposedly oscillatory tracking signatures from evoked responses to the presented events, in part by showing that they are still detectable in periods following the sensory stimulation [9-11]. The degree of such neural tracking has in turn been shown to determine our ability to perceive the input stream [9, 10], supporting the idea that these mechanisms optimize perception of our continuously fluctuating world.

The visual domain, in contrast, has typically investigated how theta (and alpha) band oscillations shape visual perception independently of external dynamics [12-18]. As a result, the comparatively small literature that has focused on how qualitatively distinct low-frequency visual rhythms are tracked in the brain [19, 20] is yet to address the supporting mechanisms including whether true oscillatory tracking signatures exist independently of evoked potentials. Notably, the human visual system appears to be tuned more towards spatially – rather than temporally – precise computation, which has led to proposals that oscillatory tracking may indeed operate differently in vision than it does in audition [21]. However, rhythms, and temporal structure in general, shape our visual environment much like its auditory counterpart, especially as a consequence of interacting with our environment – like when typing on a keyboard or walking along a road [22]. It is therefore crucial to determine whether insights from auditory research can be extended to the visual domain. Should a similar mechanism appear to support rhythmic tracking across modalities, this would underline the need for accounts that combine intrinsic and externally-based sources of rhythm to determine sampling [23].

As well as asking whether true oscillatory tracking can be found in vision, we further examined the mechanisms that may support it. First, although it appears likely from audition that attention modulates the level of tracking [3], it remains unclear whether this is a function of the stimulus stream being relevant to the task or explicit prediction of its temporal structure [9, 24]. For example, in audition, oscillatory tracking enables the resolution of competition between several co-occurring auditory streams – like in so-called 'cocktail party' situations [25] – but it is unclear whether the modulation is driven by attention to the sensory stream of interest as a whole or specifically to its temporal features. A distinction between the two is important in order to show which tracking signatures subserve specifically *temporal* prediction [3, 9, 25].

Second, we were interested in determining the representational content of the tracking response as there is evidence that neural templates of anticipated content appear – and therefore potentially change dynamically over time – without the overall signal amplitude necessarily increasing [26-28]. For example, some research suggests that activating one representation may suppress others, perhaps via lateral inhibition [29, 30]. If oscillatory tracking mechanisms like entrainment encoded temporal predictions in a feature-specific manner, we would therefore not expect whole-scale neural activity to become coupled to the external rhythm but rather activated representations of the changing feature that constitutes the rhythm. If this were the case, examinations of temporal tracking may be informed via content-based,

rather than purely amplitude-based, measures, and the literature concerning predictive processing may yield informative insights for mechanisms supporting oscillatory processing.

To this end, we conducted a magnetoencephalography (MEG) study in which participants attended to rhythmic visual sequences of rotating Gabor patches and subsequently judged the orientation or timing of a delayed probe. Using a rate-specific phase-locking analysis, we show that the motor system tracks visual rhythms, even when they are no longer concurrently present. Importantly, this maintained tracking response only appears in the timing task and predicts perception of temporal features of the probe stimulus. Using time-resolved decoding analyses, we further show temporally-specific visual simulation of rhythmic predicted sensory outcomes regardless of task, but whose extent predicts motor tracking when deriving time. These results show that true rhythmic tracking extends to the visual modality and provide neural evidence that this capacity is related to internal sensory simulation of perceptual content. On a larger scale, our findings highlight the need for temporal tracking accounts that consider how we combine environmental dynamics with intrinsically-generated rhythmic processes, and that the study of temporal tracking can take logic from the predictive processing literature to move understanding forwards.

RESULTS

Participants successfully maintain representations of stimulus orientation and timing

Thirty human participants observed a series of 7-8 flashing Gabor patches, whose onsets were separated by 500 or 750ms - creating rhythmic displays at 2 and 1.33Hz respectively. From a random starting point, the orientations progressed in steps of 15, 30 or 45 degrees in either the clockwise or counterclockwise direction. The final Gabor of each trial's stimulus series was surrounded by a black circle at its perimeter to signal the start of a 'Maintenance' period. This Maintenance period lasted three cycles of the stimulus presentation rate, during which participants were shown a fixation cross. At the end of each trial, a final Gabor patch served as the probe for participants' responses to judge orientation or timing.

In orientation blocks, participants judged whether the probe Gabor had the orientation that would have immediately followed the final orientation of the sequence. In timing blocks, they judged whether probes appeared at an 'on-beat' timepoint if the preceding rhythm had continued. On both dimensions and orthogonally to each other, probes were 'off' on 22% of trials. The stimuli could be either half a frequency cycle too early or too late and rotated either one step size too little or too far relative to the orientation that should have been predicted (described above). Stimuli were identical regardless of block, but what differed was the judgement. The timing task therefore directly required participants to maintain the rhythm and, given the contrast with the orientation task, allowed us to determine what aspect of rhythmic tracking relies on explicit attention to temporal properties of stimulus appearance. Participants were able to perform both tasks at both stimulation rates with above-chance sensitivity (one-sample *t*-test of *d-primes* against 0: all $p<0.001$).

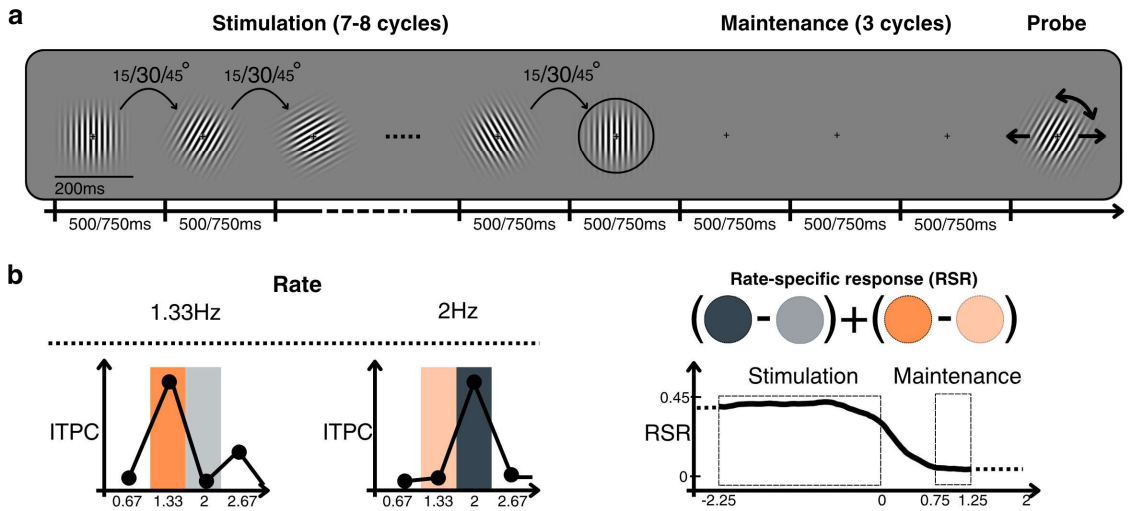

Figure 1. Design and conceptual illustration of rate specific response (RSR) analysis. **a** Participants observed rhythmic sequences of rotating Gabor patches before judging a delayed probe's orientation or timing. **b** RSR was determined by contrasting intertrial phase consistencies (ITPCs) of stimulated and non-stimulated rates. Stimulation and Maintenance windows were extracted from the RSR timecourse to determine oscillatory tracking both during and beyond ongoing visual stimulation. Note that the x-axis represents time in seconds relative to the onset of the final stimulation Gabor.

Motor regions support timing judgements through rate-specific tracking beyond ongoing visual stimulation

Like similar work in the auditory domain, we first operationalized oscillatory tracking as inter-trial phase consistency (ITPC) [9]. ITPC leverages that all trials – within each of our stimulation rates – have identically timed stimuli. If neural oscillations became phase-aligned with the sensory rhythm as proposed, this would yield large ITPC in specifically the neural frequency band that matches the stimulation rate. To capture this rate-specificity driven by stimulation rates, we contrasted ITPC between the two rates of stimulation to create a rate-specific response (RSR) that precisely captures how much more ITPC, and therefore tracking, is apparent at the trial's stimulated compared to non-stimulated frequency (Figure 1b) [9]. During the Stimulation window, participants showed a scalp-wide RSR in both the timing (cluster-corrected, $p < 0.001$; summed $t = 3118$; 272 sensors) and orientation task (cluster-corrected, $p < 0.001$; summed $t = 3086$; 272 sensors). Both conditions revealed a similar topography that spanned all 272 planar-transformed MEG sensors and was, as expected, maximal at occipital sensors (Figure 2a, left panel, top and middle). No cluster was found for the difference between the two tasks in the Stimulation window (Figure 2a, left panel, bottom).

To distinguish oscillatory tracking from evoked responses during rhythmic stimulation, we tested for an RSR in the absence of visual stimuli during the Maintenance window between the rhythmic sequence and the probe (Figure 1a, b). We found a cluster of significant RSR in the timing task that had a central topography spanning frontal, parietal and temporal sensors bilaterally, cluster-corrected, $p < 0.001$; summed $t = 239.426$; 86 sensors (Figure 2a, right panel, top). Meanwhile, no significant cluster was found in the orientation task (Figure 2a, right panel, middle). Comparing the two tasks directly, a central cluster emerged in which maintained RSR values were significantly higher in the timing than in the orientation task, cluster-corrected, $p = 0.019$; summed $t = 44.93$; 17 sensors (Figure 2a, right panel, bottom). All sensors belonging to this task difference cluster were also part of the cluster found in the timing task alone. Using an LCMV beamformer in combination with a probabilistic atlas to localise this tracking signal ([31],

see Methods), we found the maintained timing task cluster to be maximal in the right paracentral lobule and supplementary motor area (Figure 2b).

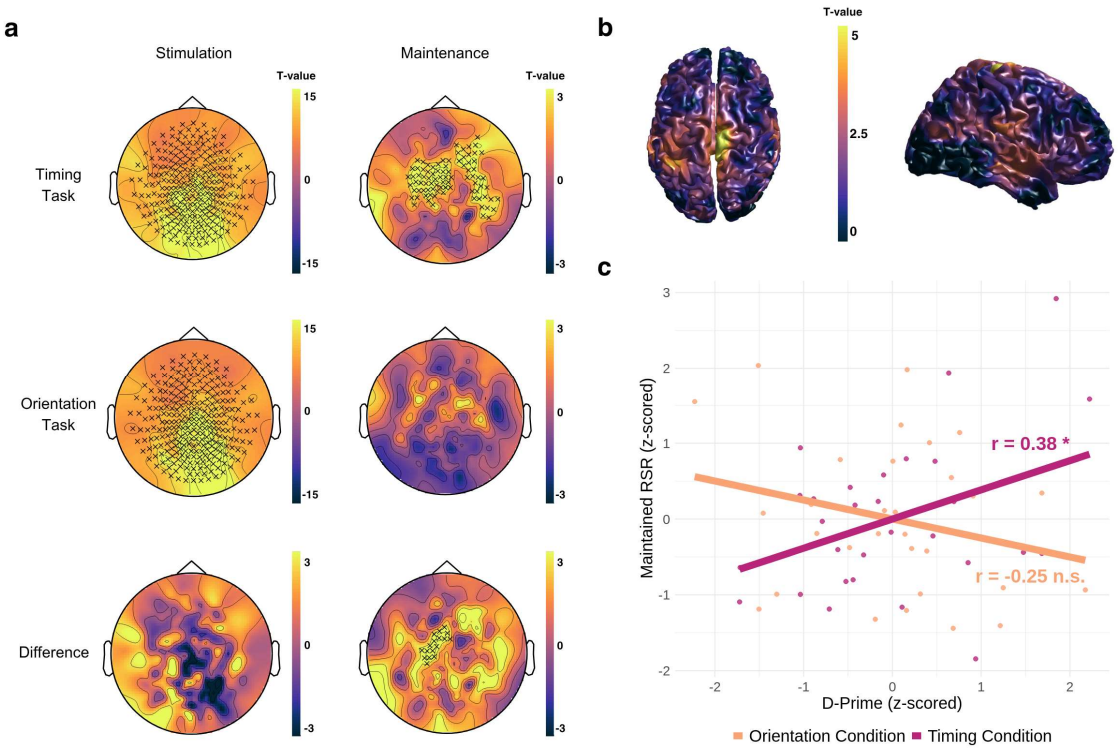

Figure 2. RSR results. **a** Permutation test t-value topographies for the RSR during the Stimulation and Maintenance windows, separately for the timing task, the orientation task and the difference between the two. Sensors marked with a cross belonged to the significant cluster found for a given condition. **b** Same as **a**'s timing task Maintenance window but source-localised using an LCMV beamformer. **c** Correlation between averaged activity across the sensors belonging to the significant cluster in **b-II** and task performance, separately for the orientation and timing task.

To establish the perceptual relevance of this signal, we next extracted participant-specific maintained RSR values from this cluster and averaged over its sensors to create a single value per participant. We did so separately for each task and then correlated participants' RSRs with their behavioural task performance (d-prime). We found a significant correlation between participants' maintained RSR responses and their timing task performance, $r = 0.385$, $p = 0.036$, while no such correlation was found for the orientation task, $r = -0.251$, $p = 0.182$ (Figure 2c) - in line with the absence of a maintained RSR in this condition. Comparing the bootstrapped z-distributions of the two correlation values indeed revealed a large difference between the two, Cohen's $d = 2.884$ [32].

In sum, these results suggest that explicit tracking of rhythmic structure engages motor areas that become phase-locked to the tracked rhythm in a frequency-specific manner. This tracking maintains the external temporal structure beyond its presentation in a manner that predicts the detection of timing discrepancies.

Temporally- and feature-specific sensory simulation appears regardless of task but scales rate-specific temporal tracking responses

We next investigated the presence of neural representations associated with a continuation of the rhythmically-rotating Gabors (Figure 3a). Should tracking mechanisms align neural

sensitivity with external temporal structure in a feature-specific way, we would expect representations of specifically the currently expected stimulus content – rather than overall neural amplitude – to change in a rate-specific manner. To this end, we trained an inverted encoding model (IEM) to convert MEG sensor-level activity into evidence for the orientation spectrum (see Methods) [26, 33]. We first confirmed that the IEM could decode truly presented Gabor orientations with high fidelity and that overt gaze position, measured using eye-tracking (see Methods), could not account for decoding results (Figure S1b, d).

Since we were interested in decoding evidence reflective of a mechanism of rhythm maintenance, we again focused our analysis on the Maintenance window between the final sequence Gabor and the probe onset. We focused on specifically those orientations that would have appeared in each of the three 'empty' cycles of the maintenance window if the sequence had continued. We formalized our predictions about when those orientations should be represented most strongly in theoretical signed difference matrices (SDMs) and correlated these with empirical counterparts filled using the decoding traces from the IEM (Figure 3c, also see Methods). The resulting correlations thus represent the fit between participants' data and our predictions about temporally-specific content representation.

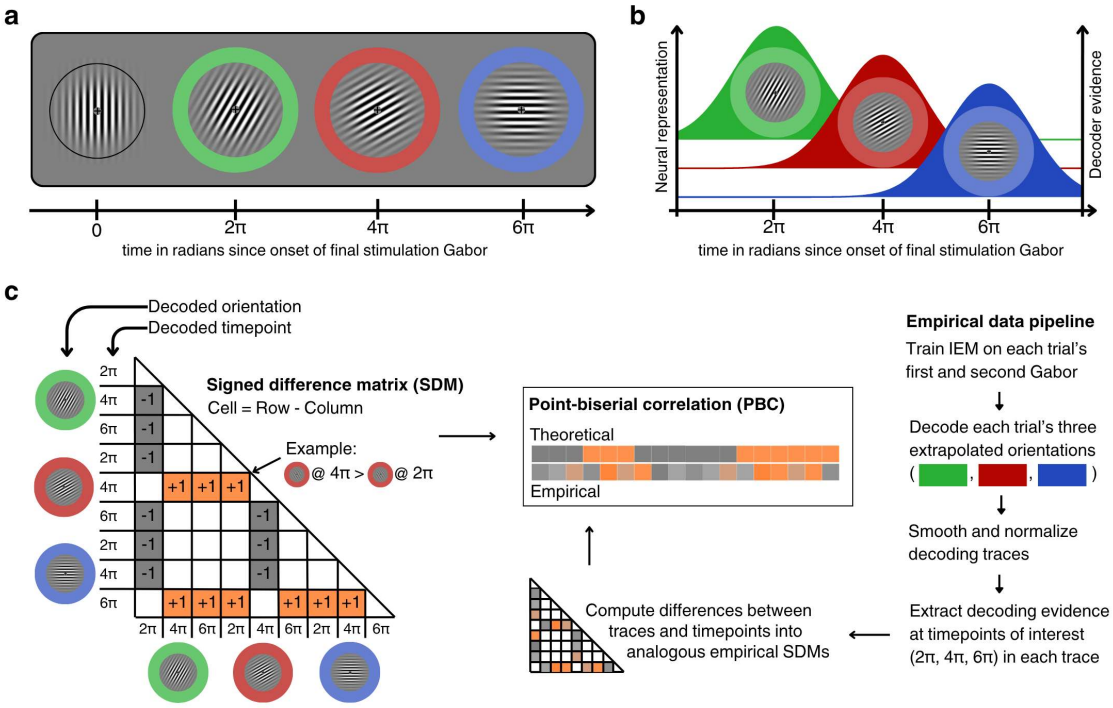

Figure 3. Decoding analysis. **a** Extrapolating each trial's rhythmic sequence into the Maintenance window yields one predicted orientation for each of the window's three 'empty' cycles. Relative to the onset of the sequence's final Gabor, these cycles begin at 2, 4 and 6π rate-specific radians (i.e. if the rate is 2Hz, 2π = 500ms). **b** Neural representation of each of the three predicted orientations should be maximal at their respective predicted timepoint of appearance and therefore temporally-distinct from each other. We use time-resolved decoding (inverted encoding model, IEM) of Gabor orientation to capture these representations. **c** We organise our predictions about the decoding output in signed difference matrices (SDMs) in which each cell contains the binarised (+/- 1) predicted difference between the row and column element. Specifically, each presumed decoding peak (i.e. green orientation @ 2π) is compared to all other 'non-peak' decoding timepoints (i.e. green orientation @ 4π, red orientation @ 2π). Comparisons for which we made no prediction (i.e. green orientation @ 2π vs. red orientation @ 4π, white cells in SDM) are masked out when the matrix is flattened into a vector before correlation with SDMs filled with participants' true decoding output from the IEM.

Comparing participants' fits to a label-permuted baseline (Figure S1a, also see Methods), we found that participants showed such sensory 'simulation' in both the orientation, $t(29) = 2.186$, $p = 0.037$, and timing task, $t(29) = 2.685$, $p = 0.012$, without a significant difference between the two, $t(29) = 0.342$, $p = 0.735$ (Figure 4a & b). Source-localising the trained weight matrix of our IEM, we found the origin of this signal in occipital sources and specifically maximal along the left calcarine sulcus (Figure 4c, see S1c for sensor-level). This measure of sensory simulation additionally correlated with the motor-based RSR signature in the timing task, $r = 0.43$, $p = 0.018$, but not the orientation task, $r = -0.128$, $p = 0.501$ (effect size of difference, $d = 3.532$, Figure 4d). Further underlining that this represents a direct link with rhythmic tracking and not simply general attentiveness, we did not find an analogous correlation between the simulation and the presumably ERP-dominated occipital RSR during ongoing visual stimulation in the timing task, $r = -0.112$, $p = 0.557$, nor in the orientation task, $r = 0.075$, $p = 0.695$.

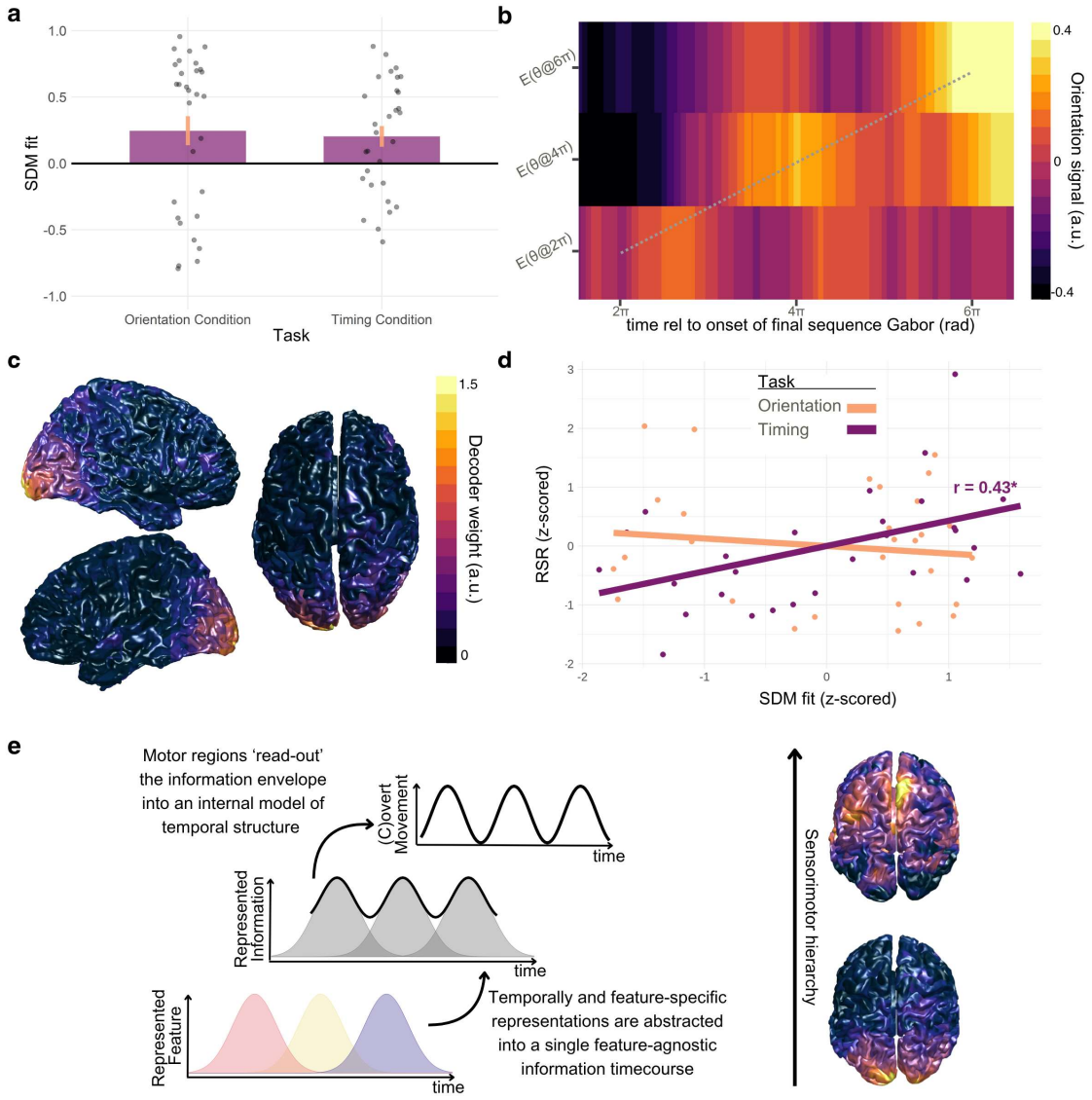

Figure 4. Sensory simulation results. **a** Participant-wise SDM fits, separate for each task. **b** Maintenance window decoding traces of the three expected orientations at each cycle of the Maintenance window, averaged over all participants and tasks for illustration purposes. Dotted grey line shows predicted trajectory as quantified in the theoretical SDM. **c** Source-localised inverted encoding model (IEM) weights reveal occipital origin. **d** Correlations between SDM fits and RSR, separately for each task. Each task's

RSR was additionally estimated separately for the two windows of interest by averaging over the 86 sensors with the strongest RSR in each respective window. Note that the choice of 86 sensors was based on matching the number of sensors contributing to the significant cluster in the timing task's Maintenance window cluster. **e** Potential mechanism based on our findings. Activity in sensory processing areas represents predicted stimuli in a temporally and feature-specific manner which yields an envelope of predicted, information-dense timepoints. Motor regions maintain this envelope through covert or overt movement and use it to flexibly coordinate temporally-sensitive processing.

These findings suggest an association between behaviourally-relevant temporal tracking in motor regions and ongoing sensory simulation of the tracked sequence in early visual areas. Note that the presence of mental simulation in the orientation task, in addition to the timing task, indicates that this process is not specific to temporal tracking but rather appears regardless of task.

DISCUSSION

We demonstrate phase-coupling of motor cortex activity that tracks the temporal structure of low frequency visual rhythms specifically when participants performed explicit temporal judgements. Visual regions meanwhile dynamically represent the concurrently predicted content of the rhythm in a temporally-precise manner across both tasks, but also in a manner that is associated with the motoric tracking response when deriving time. We thus show neural representation of visual temporal structure that consists of two concurrently unfolding components. On the one hand, the visual system automatically simulates the anticipated content of predictable events at their inferred timepoint of occurrence. Elsewhere, the motor system leverages this simulation by effectively maintaining an amplitude readout of the inferred information envelope when time is relevant to the task at hand (Figure 4e).

Interestingly, the involvement of motor regions, and specifically the supplementary motor area (SMA), in temporal tracking is in line with recent findings in the auditory domain [34-40]. Therefore, the visual domain may similarly capitalize on predictable temporal structure because the underlying mechanism is orchestrated by a centralised motor hub. Our proposed role of the relationship between motor and sensory processing is also consistent with recent work in which monkeys were trained to track a left-to-right moving visual metronome, where neural components reflecting spatial and tempo tracking were dissociated [41]. Local field potential power fluctuations were predominantly determined by tempo in SMA (beta-band), but by space in more posterior regions like V4 (alpha and beta-bands). Our findings further support this tuned, multi-site internal simulation architecture, and additionally show that rather than simply operating in parallel, sensory simulation and temporal tracking modulate each other in the human brain. Importantly, our findings also directly link internal simulation to delta-band phase locking, which forms a bridge with recent theories of human temporal tracking.

The involvement of the motor system in temporal tracking has led to elegant proposals of simulated action as the basis for temporal predictions [34-40]. These commonly state that the motor system takes its central role since it orchestrates sensory predictions as the consequence of (simulated) actions through efference copies [34, 37]. However, our results clearly show that prediction-based sensory simulation appears regardless of motor-based temporal tracking and is therefore unlikely to be its result. Although a true test of directional causality must be the aim of future – potentially interventional – research, our findings are instead more consistent with a new possibility that we propose. That is, the seemingly causally-inverse interpretation that sensory systems continuously represent predictions of future sensory states without relying on the motor system, but that these representations are read out by the motor system only if we must directly extract temporal information (Figure 4e, also see [41]). This interpretation is consistent with behavioural work showing a dependence of temporal predictions on explicit spatial predictions [42], as well as recent high-field fMRI work showing

that duration and rate-tuned neural responses arise as processing ascends the visual processing hierarchy and are presumably derived from early, feature-tuned visual responses [43]. At the top of said hierarchy, we suggest that motor regions maintain the predicted information envelope as a potential mental model of external temporal structure. Under this view, temporal predictions are simply derivatives of this simulated information envelope, and sensory predictions are their necessary bedrock.

Following this interpretation, the success of temporal tracking scales with the presence of temporally-precise predictions of the stimulus feature that uniquely changes with each cycle of the tracked rhythm. In our case, the temporal prediction of one frequency cycle coincided with the predictable change in orientation angle that came along with it. In the previously mentioned study by de Lafuente and colleagues it was the spatial location of the visual metronome [44]. The sensory predictions upon which the brain potentially bases temporal predictions thus seem to be feature-specific rather than constituting overall stimulus presence [45]. Whether that is a requirement or not must be addressed in future studies but it raises the question of how temporal predictions are formed in situations that do not allow for feature-specific sensory predictions. Following this notion, the link between temporal tracking and feature-specific sensory predictions could potentially offer new views on certain clinical conditions. For example, it may provide an explanation for the reduced phase-synchronization to speech in those with dyslexia [46, 47]. The communication of phonological representations from primary auditory to higher-order areas has been found to be impeded and so the motor system may not be able to derive the temporal structure needed for phase-synchronization to occur [48]. More generally, sufficiently precise sensory predictions being a prerequisite for effective temporal tracking could reveal an underutilized explanation for why clinical conditions associated with aberrant prediction strengths such as schizophrenia are also characterized by altered time perception [49].

The interpretation of temporal predictions as derived from content-based sensory predictions does not conflict with the idea that the motor system influences sensory processing too. This has been the focus of the literature on auditory temporal tracking which has suggested that the motor system's propensity to synchronize overt and covert action to sensory events leads to improved perception [37]. Our finding that motor-based temporal tracking supports performance in visual timing judgements is consistent with these effects. Sensory predictions scaffolding motor-based temporal predictions which in turn shape sensory processing instead support the view of an action-perception-loop. Theories like active sensing outline how movement not only passively reflects temporal structure but also reorients sensors to sample the environment in a maximally informative way [50, 51].

In line with these theories, our findings may also suggest that manipulating the time course of sensory predictions should consequently affect the rate at which we overtly sample the environment. For example, focusing on specifically ocular movement, recent accounts have highlighted the overlap between typical saccade rates and visual theta (~ 4Hz) rhythms that are found when stimulation is not itself changing at that particular rate [52], with most suggesting that this is due to the common, fixed theta rate functioning of the frontoparietal network [53]. However, as we recently proposed in a theoretical piece [23], this rate could instead reflect a 'stubborn' visual prior, for example resulting from a lifetime of saccading at approximately that rate when exploring our environment. Once reliable temporal structure appears besides this canonical rate, such as through sensory rhythms in our study, the rate of this rhythm and all its corollary effects may flexibly shift to retain a maximally informative sensory sampling routine in the current environment. We therefore suggested that the resulting perceptual rhythm may reflect a dynamically-updated, precision-weighted combination of various estimates of temporal information, some learned over a lifetime and some more recently [23, 54]. Future research could investigate this proposal by testing whether overwriting the theta default using

reliably predictable structure leads to systematic changes in the commonly probed oscillatory profile of behavioural performance.

These new ideas show the potential explanatory power of bringing logic from the predictive processing literature to the understanding of how we represent the temporal characteristics of our environment (see also [55]). The current study's findings further showcase the strength of this cross-pollination at the empirical level. Much debate has recently focused on distinguishing the oscillatory phenomenon of entrainment from other neural signatures. Quite rightly, many effects attributed to entrainment are argued to provide poor evidence because they are measured during periodic stimulation, and such measures could therefore simply reflect periodic evoked responses. Isolating oscillatory tracking at the rate of stimulation in a period afterwards, as we have done, has been proposed as a solution to revealing the nature of the underlying mechanism. Our findings in visual regions, namely the lack of overall amplitude changes tracking the rhythm but instead activated representations of specifically predicted content, suggest that an exclusive focus on overall sensory neural amplitude changes may be misleading. Notably, this suggestion is in line with past findings that auditory entrainment may be tonotopically-specific [7, 56]. More broadly, it reflects the value of oscillatory tracking and predictive processing research taking inspiration from each other's insights – both in terms of mechanistic explanations and empirical methodology – to facilitate progress.

In sum, this study provides novel insights into how the human brain tracks temporal structure in its environment. Complementing past research in the auditory domain, we found rate-specific phase coupling between delta oscillations in the motor system and a presented visual rhythm - supporting the interpretation of motor regions as a timing hub when required to extract temporal features from our environment. Importantly, this tracking response was maintained beyond the offset of the visual rhythm and influenced participants' sensitivity to timing violations. Most interestingly, this tracking response was scaled by the degree to which participants extrapolated the predictable visual sequence by means of feature and temporally-specific visual simulation, which occurred regardless of task. These findings suggest a coordinated interaction of sensory and motor simulation for representing the intricate temporal relations of our environment.

MATERIALS AND METHODS

Participants

Thirty-one participants were screened for MEG eligibility and gave informed consent in line with the UCL ethics committee (protocol 3090/004). Participants were recruited through the local SONA subject pool, reimbursed in either credit or cash (£10/hour) and required to have normal or corrected-to-normal vision. One participant was excluded from analyses as they did not complete the experiment, leaving a sample of 30 participants (24 females; mean age = 24.81, SD = 4.39 years; 30 being our preregistered sample size: https://osf.io/8ptsz).

Stimuli and Task

Gabor patches were created using the Psychophysics Toolbox implemented in MATLAB, with 12° of visual angle diameter and a spatial frequency of 0.4 per visual angle [57, 58]. Annuli were created by overlaying a background-coloured (grey) circle on its centre that spanned 0.43° of visual angle and contained a black fixation cross of 0.35° of visual angle. Each Gabor was presented for 200ms each and the first in the sequence started at one of the 12 possible orientations that result from equally dividing the 0-165° space using steps of 15°.

On 78% of trials within each block, this probe appeared in time with the preceding rhythm: exactly four frequency cycles from the onset of the final entrainer of the flashing series. On the remaining 22% of trials, it appeared either half a frequency cycle too early or too late (counterbalanced across trials within each block). Furthermore, and independently of this timing manipulation, on 78% of the trials, the probe showed the result of adding the fixed amount of rotation from the preceding series to its final patch. On the other 22% of trials, the probe showed either the same orientation as the final patch of the series or two steps added to it. Before each of the blocks, participants were told to either judge timing or orientation through text on the screen. In the former, they were asked to press a button when the probe did not appear in line with the preceding rhythm. In the latter, they were asked to press the button when the probe did not show the correct (next following) orientation. They could respond as soon as the probe appeared and had 1500 ms to do so. The next trial started approximately 1250 ms (SD = 300 ms) after the end of the previous trial's response window. There were three blocks of each stimulation rate and task combination, with presentation order approximately balanced across participants.

All combinations of levels of the other trial variables – number of entrainers, step size, direction and starting angle of orientation rotation as well as direction of deviant probes (forward or backward in either rotation or time) – were balanced across trials within a participant and presented in a randomised order.

The study consisted of two testing sessions. The first was a behavioural practice session of one hour in which participants learned and practised both tasks. Given sufficient performance in both tasks as well as MEG eligibility, participants were invited for a scanning session on a different day. If, due to scheduling constraints, the scan occurred more than a week after behavioural practice, participants practiced both tasks again for a combined half hour before entering the scanner. In the scanner, participants completed three blocks of 36 trials for each rate and task combination resulting in 108 trials per condition. We instructed participants to stay still citing MEG data quality issues as the reason but also to prevent overt movement along with the stimuli. Participants were monitored during scanning using live video transmission. Following the scan, participants completed a finger tapping task designed to measure their natural motor rates. This measure was however recorded for a different research question and was not analysed for the current study.

MEG acquisition and preprocessing

MEG data acquisition occurred in a magnetically shieled room using a 272-channel CTF MEG system with axial gradiometers sampling at 600Hz. A third-order gradient compensation algorithm was applied online for noise reduction. Fiducial coils on the nasion as well as on the left and right preauricular were used to monitor the participant's head position during data acquisition. A photodiode at the bottom left of the screen was used to temporally align the neural signal with the on-screen stimulus presentation. Alongside MEG acquisition, an EyeLink 1000 infrared tracker (SR Research Ltd.) was used to record eye movements.

We further preprocessed the MEG data using the Fieldtrip toolbox as implemented in MATLAB [59]. Visual artifact rejection was carried out using trial-wise MEG sensor variance as the selection criterion for further manual inspection. Trials were removed when they showed excessive artifacts. The eye-tracking signal was further used to flag trials in which blinks occurred during stimulus presentation. The resulting data were then temporarily downsampled to 200Hz and a 1-40Hz bandpass-filter (two-pass Butterworth) was applied before conducting an independent components analysis [60]. The first 50 components' topographies and time courses were visually inspected seeking cardiac and eye-related artifacts, whose components were removed. The resulting projections were applied to the original, not filtered nor downsampled, data.

Behavioural measures

Behavioural analyses were conducted in MATLAB and focused on the signal detection theoretic (SDT) measure of d-prime [61]. It is calculated as the difference between participants' normalized hit and false alarm rates within a condition and reflects their sensitivity for detecting oddballs in each task.

Rate-specific intertrial phase consistency

All MEG-related analyses were conducted in MATLAB using the Fieldtrip toolbox. For the following analyses concerning phase consistency, we used custom scripts adapted from [9]. We preregistered the analyses mentioned in this section together with the planned sample size, experimental variables and hypothesis on Open Science Framework (https://osf.io/jm6er) before starting data collection. Inspired by past work, we used rate-specific inter-trial phase consistency (ITPC) as a measure of oscillatory tracking. When stimulated at a particular frequency, the neural phase of that frequency should be aligned with the timepoints of stimulus appearance. Making use of the fact that all trials in our design within a frequency follow the same time course, this means that, at a given timepoint, a tracked frequency should show a higher ITPC than the non-tracked frequency. To directly make this comparison, we used the rate-specific response (RSR) proposed by [9]. It quantifies how much more ITPC there is to a given trial's stimulation frequency (e.g. 2Hz) compared to the other, non-stimulated frequency (1.3Hz). This approach controls for any non-experimental, spectral differences between the stimulation rates.

To calculate ITPC, preprocessed data were low-pass filtered at 20Hz (two-pass Butterworth) to mitigate the influence of movement-related artifacts. The axial gradiometer data were then subjected to a planar transform, yielding two components for each of the 272 sensors. ITPC calculation was carried out on each of the resulting 544 components separately, after which the two components belonging to each sensor were averaged to restore the original gradiometer structure. We used a sliding-window fast fourier transform (FFT) to transform the data into the frequency domain. The window slid in steps of 20ms with its width set to 1500ms to strike a balance between sufficient resolution to distinguish between our stimulation rates 1.33Hz (2 cycles in 1500ms) and 2Hz (3 cycles), and temporal specificity. The phase angle was extracted for both frequencies of interest for each of the timepoints upon which the FFT window was centred. The complex mean of the resulting trial-by-trial phase was then computed separately

over all trials of each of the four (rate by task) conditions before taking the absolute value to yield each condition's ITPC. To further account for non-oscillatory 1/f-components that exist in the frequency spectrum of neural data and that could bias the RSR, we fitted and subtracted a 1/f curve from the ITPC values for each participant, sensor and condition separately before calculating RSR based on the residuals [9].

Since the number of Gabors in each flashing sequence was the same for both stimulation frequencies – 7 or 8 – the trials in the respective conditions were of different lengths. To align the two, we centred trials of both frequencies to the onset of their final Gabor of the sequence. Then, the longer 1.3Hz trial time courses were cut to the length of the 2Hz trials which allowed us to calculate RSR values for each sensor, task, participant and timepoint upon which the FFT was centred. The result of this approach can be seen in the following definition of our windows of interest where the earliest and latest timepoints resemble the bottleneck imposed by the length of the shorter 2Hz trials.

To distinguish between oscillatory tracking during and after the offset of external stimulation, we defined two analysis windows of interest. The first window ('Stimulation') spanned from 2250ms before each sequence's final Gabor to its onset, noting that these timepoints represent the centre of the respective FFT window during phase extraction. The second analysis window ('Maintenance') took the maximal length of data within the three cycles of the Maintenance period while avoiding contamination from evoked responses stemming from the offset of stimulation sequence and the onset of the probe. The FFT window centres therefore spanned from 750ms to 1250ms after the onset of the final Gabor. Within each analysis window, RSR values were averaged so that task and window-wise results from all participants could be statistically evaluated in the following way.

Cluster-based permutation dependent-samples t-tests were used to both compare the two tasks to one another and to evaluate whether each condition's individual RSR is significantly different (two-sided) from a baseline of 0. Previous work supports this analysis approach by showing that RSR values are normally distributed and that a surrogate distribution based on adding random phase values to the FFT output before calculating ITPC is indeed centred around 0 [9]. We used a Monte-Carlo algorithm with 5000 permutations for each planar-transformed MEG sensor separately. Clusters were formed by neighbouring (minimum 2) sensors which individually reached significance (alpha = 0.05). The resulting cluster-level statistic represents the sum of the t-values of the sensors within the cluster and is tested for significance (cluster-alpha = 0.05).

Source localisation

In order to outline the source of the RSR signal, as well as for visualization purposes, we conducted source localisation by using a template approach as implemented in Fieldtrip. Using participants' fiducial measurements and the MEG sensor structure, we warped a standard Fieldtrip single shell head model and a standard Fieldtrip source model (8mm dipole spacing) into participant-specific approximations. We estimated separate spatial filters for the Stimulation and Maintenance windows described above using LCMV-beamformers after combining the data from the two stimulation rates [62]. We applied the spatial filters to FFT-transformed single-trial data before extracting the phase and calculating ITPC and RSR as in the sensor-level data described above. We used the AAL atlas implemented in Fieldtrip for probabilistic anatomical labeling [31].

Orientation decoding

We used an inverted encoding model (IEM) to decode sensory representations, specifically an implementation used in previous work that was able to decode predicted grating orientations in

the absence of ongoing stimulus presentation . We will provide a condensed description of the IEM here but for a full implementation, see [26] and [33], from which we used custom MATLAB scripts.

At its core, the approach includes two steps. First, a forward model is trained to model the MEG sensor-level activity evoked by different Gabor orientations. Specifically, this model consists of 32 equally-spaced channels whose Von Mises tuning curves cover the entire orientation spectrum and can be linearly combined to reproduce any single orientation. Once trained, the transition from orientation to MEG data is described by a weight matrix that captures each MEG sensor's 'orientation tuning' while accounting for the noise covariance of neighbouring sensors. Note that we examined this weight matrix to probe which sensors contribute to the decoder (Figure 4c & S1c). In the second step of the IEM, this filter is applied in the opposite direction to create an inverse model that converts MEG data into evidence for the orientation spectrum.

In this study, we used the first two Gabors of each trial's flashing sequence to train the forward model. We chose the first two Gabors because their orientation could not be predicted; after the second Gabor is presented one can anticipate subsequent Gabor orientations because the rotation direction and stepsize for that trial have already been determined. We extracted two 500ms epochs from each preprocessed trial, starting at the respective onsets of the first and second Gabor of the sequence. We also applied a 40Hz low-pass filter with a filter order of 6, in line with past research [26]. This process yielded an average of around 840 (SD = 34) training examples (~ 70 per orientation class) per participant. The model was trained at each training data timepoint, separated by 5ms. While doing so, the data were averaged within windows of 29.2ms using a sliding window approach to improve the signal-noise-ratio at each training timepoint [26]. The resulting time-resolved filter matrix was then applied to our test data.

We again focused our analysis on the window between the final entrainer and the probe onset. Specifically, we extracted the epoch between each trial sequence's final Gabor and 1000ms after the probe onset and again applied a 40Hz low-pass filter with a filter order of 6. We estimated the orientation channel responses of these test data at each timepoint, separated by 5ms and averaged within a window of 29.2ms - which yielded a 4D matrix (training time X testing time X model channel X trial) containing the estimated channel responses during the window of interest for each trial. Channel responses were consequently shifted so that the channel centred on the orientation of interest (see below) occupied the position of the 0°-centred channel. This allowed averaging over trials to result in a decoding output of the orientation of interest which was consequently transformed into a scalar projection that was used as the measure of decoding performance. Trial-averaging was done separately for each condition (orientation/timing) and rate (1.3Hz/2Hz).

We determined four orientations of interest for every trial. Namely, based on each trial's orientation sequence parameters (starting orientation, rotation direction and rotation step size), we determined the three orientations that would have appeared in the three cycles of the Maintenance window if the sequence had continued. These could subsequently be used to establish whether participants simulated this continuation. The fourth orientation of interest was the presented probe orientation, which was used to establish the fidelity of the decoding approach (Figure S1b). We focused decoding analyses exclusively on non-oddball trials since unexpected presentations or omissions could infringe on the analysis window or lead to divergent responses.

Rate-specific sensory simulation analysis

Within the decoding output, we selected a training time window from 120 to 160ms after stimulus onset, based on past research showing optimal decoding sensitivity for similar stimuli and presentation durations in this window, and averaged the decoding output within this

window [26]. We first tested the approach's fidelity using the probes presented in both tasks to establish that subtle orientation signals can be decoded. Then, the decoding output of each trial was narrowed to the extrapolated orientations and the time between the onsets of the final sequence Gabor and the probe. This resulted in windows of 2000 or 3000ms for decoding output with 2 or 1.3Hz stimulation rate respectively. This yielded three decoding traces for each trial, one for each extrapolated orientation. To focus on delta-rate representation changes and to improve the signal-to-noise ratio, we next smoothed these traces by averaging within a 150ms sliding window. Then, to avoid the analysis being contaminated by temporally-consistent biases towards a particular orientation, for example towards the orientation of the final Gabor of the sequence or the probe, we normalized each decoding trace by z-scoring. This allowed us to examine relative temporal changes in representations of the extrapolated orientations in an unbiased manner.

We next defined three timepoints of interest within each decoding trace as the predicted onsets of the extrapolated gratings – one at the beginning of each empty cycle of the Maintenance window. For example, for 2Hz trials, this corresponded to the timepoints 500, 1000 and 1500ms after the onset of the final Gabor of the flashing sequence. We will call these timepoints 2π, 4π and 6π (interval from the final Gabor onset in radians) and their associated orientations A, B and C. For each stimulation rate separately, we extracted the momentary decoding evidence for each combination of timepoint and decoded orientation before averaging over both stimulation rates within each timepoint. Note that the latter is possible since the timepoints of interest were chosen based on the common scale of radians.

We assessed whether the decoding evidence resembled the profile of rate and feature-specific mental simulation by using signed difference matrices (SDMs) in which the nine datapoints are related to one another in a pairwise manner. Specifically, we first created a theoretical SDM that captured the predicted differences between the datapoints if participants indeed showed a sensory simulation profile. Within each pair / cell of the matrix, a positive value (1) value was assigned if the row element was predicted to be larger than the column element and a negative value (-1) for the reverse. For example, orientation A should show higher decoding evidence at timepoint 2π than at 4π. Combinations for which we had no prediction, for example orientation A at 2π vs orientation B at 4π, received a 0 but were masked later anyway (see below).

Having quantified the model, we filled corresponding empirical SDMs for each participant with the pairwise differences between the momentary decoding evidence as described above. We did so for each task separately. To assess how well each participant's empirical matrices were captured by the model, we flattened the off-diagonal triangle of each matrix into an array [63]. Then, like previously mentioned, we masked out the values of the resulting array for which we had no empirical prediction (zeros in the theoretical SDM). This produced two 18-item arrays, one binary and theoretical and the other continuous and empirical. Finally, we calculated the point-biserial correlation coefficient between these two arrays for each participant and task individually. The resulting fit value between -1 and 1 should be interpreted as the extent to which the model predictions of an internal simulation mechanism capture the observed data in each participant.

To determine whether such sensory simulation occurred at the sample level, we compared the fit values against a permutation-based baseline using two-sided paired-sample t-tests. This baseline was based on a label shuffling procedure in which, before filling each participant's empirical SDM, we shuffled the labels of the three decoding traces. Although this only allows for five permutations outside of the true labelling, we elected against also shuffling the measured timepoints since this would artificially destroy temporal structure [64]. Decoding traces were consequently smoothed and normalized as usual before extracting the timepoints of interest to fill the empirical SDMs. For each participant, we correlated these permutation-filled empirical

SDMs with the theoretical SDM before averaging over the permutations. This procedure represents specifically the null hypothesis that sensory simulation is not specific to the predicted feature-sequence.

Orientation decoding source localisation

To determine the neural sources of the IEM-based decoding of Gabor orientations, we focused on the weight matrix of the forward model that maps MEG sensor activity onto the model's hypothetical orientation channels for each training timepoint (see above). We used LCMV beamforming as we did for the ITPC analyses but this time created the spatial filter based on the epochs that served as training data for the IEM. We multiplied each training timepoint's IEM weight matrix with the resulting source filter before taking the absolute value of the product and averaging over the model's hypothetical orientation channels. Finally, we averaged each voxel's model weight within our IEM training window of interest. To account for the centre of head bias that arises during source localisation, we repeated the above steps using a sensor-wise shuffled version of the IEM weight matrix so that the source filter was multiplied with a permuted weight matrix. Specifically, we ran 100 repetitions of this procedure, averaged the result and consequently z-scored the true weight matrix against this shuffled baseline. Since the permuted baseline still contains the centre of head bias after multiplication with the spatial filter but no longer any meaningful weights, this procedure results in a bias-free source-localisation of the IEM weight matrix.

Gaze-based orientation decoding

To investigate whether varying Gabor orientation angles are associated with systematic differences in gaze position, we conducted a linear discriminant analysis (LDA) on the time windows that served as training epochs for the neural IEM (see above). Following the logic introduced by [65], we divided our orientation set (0:15:165) into a near-cardinal (0 degrees +/- 15 and 90 degrees +/- 15) and a near-oblique (45 degrees +/- 15 and 135 degrees +/- 15) subset. This allowed us to conduct two separate two-way classifications between orthogonal classes before averaging the two decoding accuracies into a single performance measure.

As mentioned, we focused on the epochs of data that were used to train the neural IEM since the presence of above-chance gaze-based decoding would reveal whether any of the sensory simulation results could, at least in part, be explained by overt gaze position. Eye-tracking data could not be collected for three participants since their head required positioning at a height inside the MEG helmet that did not allow capture of the eyes by the eye-tracker. We preprocessed the raw gaze data of the remaining 27 participants into epochs of 500ms exactly like for the neural IEM besides electing not to demean the gaze data in order to preserve their true location on the screen. We removed epochs in which blinks occurred (M = 29) which, on average, yielded approximately 811 trials (SD = 41) per participant for the following analyses.

To balance the training and testing data in terms of instances of our 12 orientation classes, we determined the class with the least instances and downsampled all other classes to that number (M = 67). We then used a leave-one-out cross validation (LOOCV) approach to train and test the LDAs on this data. For each of 20 LOOCV blocks, 95% of each classes' downsampled instances contributed to the training data and the remaining 5% to the testing data. We then trained and tested the two (ordinal and oblique) classifiers using MATLAB's built-in 'classify' function in a time-resolved manner, training and testing on each pairwise combination of the 95 training timepoints. Within each classifier, accuracy was operationalized as the proportion of correctly classified testing data examples before averaging over the two classifiers as described above. This was done for each of the 20 LOOCV blocks and repeated five times for each participant to avoid spurious results based on which instances of the majority classes were included or excluded during downsampling. Averaging over these repetitions resulted in 95

(training time) x 95 (testing time) generalization matrices of gaze-based decoding accuracy for each participant.

We extracted the diagonal of each participant's generalization matrices where training and testing occurs at the same timepoint and used two-sided cluster-based permutation t-tests (10000 permutations) against the chance baseline of 0.5 to assess whether and when presented orientation could be significantly decoded from gaze data. Like in the RSR analyses, clusters were formed by, here temporally, neighbouring (minimum 2) data points which individually reached significance (alpha = 0.05). The resulting cluster-level statistic represents the sum of the t-values of the timepoints within the cluster and is tested for significance (cluster-alpha = 0.05).

SUPPLEMENTARIES

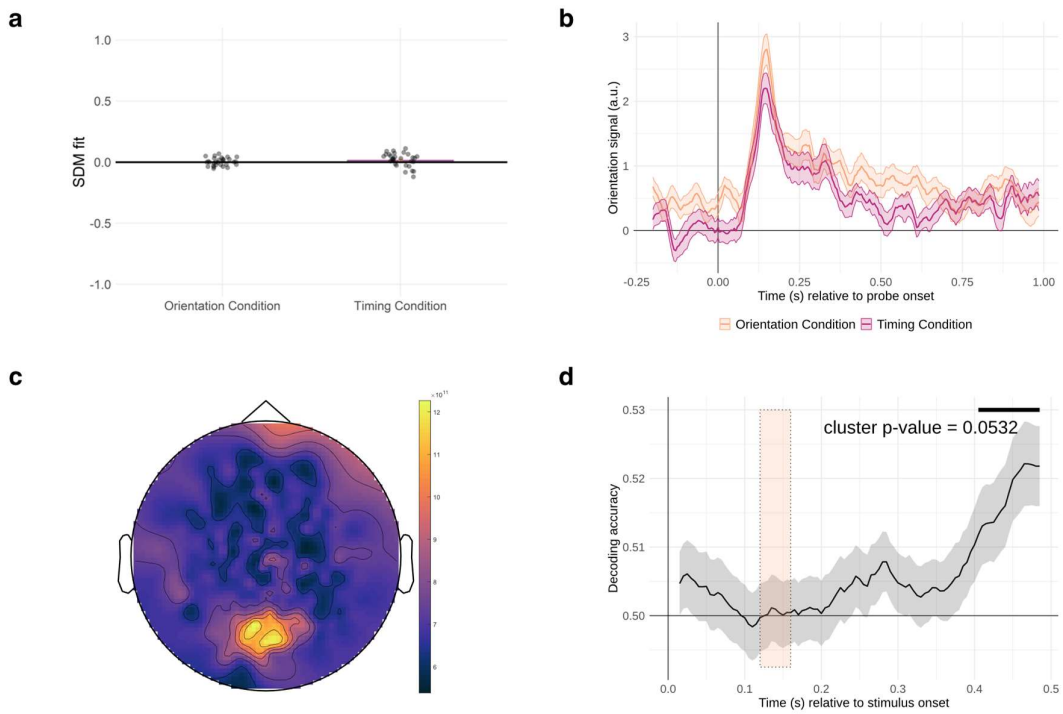

Figure S1. Decoding control analyses. **a** Same as 3b but for the label-shuffled permutation baseline. **b** Trained IEM successfully decodes probe orientations in both the orientation and timing task condition. **c** Same as 3a but before source localization. The scale refers to the absolute weight of each sensor, averaged over the 32 theoretical orientation channels of the IEM. **d** Within-training data orientation decoding based on gaze data. Orange outline shows the training window that was used for neural IEM decoding. Black bar indicates the late near-significant cluster which lasts from 405 to 485ms after stimulus onset as participants anticipate the next following sequence Gabor. Note that this cluster, besides not reaching statistical significance, is clearly separated in time from the training time window based on which the neural effects are based. We therefore take this as evidence that our decoding findings cannot be explained by overt gaze.

References

1. Patel, A.D., *Language, music, syntax and the brain*. Nat Neurosci, 2003. **6**.
2. Obleser, J. and C. Kayser, *Neural Entrainment and Attentional Selection in the Listening Brain*. Trends Cogn Sci, 2019. **23**(11): p. 913-926.
3. Lakatos, P., J. Gross, and G. Thut, *A New Unifying Account of the Roles of Neuronal Entrainment*. Curr Biol, 2019. **29**(18): p. R890-R905.
4. Haegens, S. and E. Zion Golumbic, *Rhythmic facilitation of sensory processing: A critical review*. Neurosci Biobehav Rev, 2018. **86**: p. 150-165.
5. Doelling, K.B. and D. Poeppel, *Cortical entrainment to music and its modulation by expertise*. Proc Natl Acad Sci U S A, 2015. **112**(45): p. E6233-42.
6. Lakatos, P., et al., *Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection*. Science, 2008. **320**.
7. Lakatos, P., et al., *The spectrotemporal filter mechanism of auditory selective attention*. Neuron, 2013. **77**(4): p. 750-61.
8. Luo, H. and D. Poeppel, *Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex*. Neuron, 2007. **54**(6): p. 1001-10.
9. van Bree, S., et al., *Sustained neural rhythms reveal endogenous oscillations supporting speech perception*. PLoS Biol, 2021. **19**(2): p. e3001142.
10. Kosem, A., et al., *Neural Entrainment Determines the Words We Hear*. Curr Biol, 2018. **28**(18): p. 2867-2875 e3.
11. Bouwer, F.L., et al., *A Silent Disco: Differential Effects of Beat-based and Pattern-based Temporal Expectations on Persistent Entrainment of Low-frequency Neural Oscillations*. J Cogn Neurosci, 2023. **35**(6): p. 990-1020.
12. Gray, M.J., et al., *Oscillatory recruitment of bilateral visual cortex during spatial attention to competing rhythmic inputs*. J Neurosci, 2015. **35**(14): p. 5489-503.
13. Dugue, L. and R. VanRullen, *Transcranial Magnetic Stimulation Reveals Intrinsic Perceptual and Attentional Rhythms*. Front Neurosci, 2017. **11**: p. 154.
14. Michel, R., L. Dugue, and N.A. Busch, *Distinct contributions of alpha and theta rhythms to perceptual and attentional sampling*. Eur J Neurosci, 2022. **55**(11-12): p. 3025-3039.
15. Fiebelkorn, I.C. and S. Kastner, *A Rhythmic Theory of Attention*. Trends Cogn Sci, 2019. **23**(2): p. 87-101.
16. Helfrich, R.F., et al., *Neural Mechanisms of Sustained Attention Are Rhythmic*. Neuron, 2018. **99**(4): p. 854-865 e5.
17. Landau, A.N., *Neuroscience: A Mechanism for Rhythmic Sampling in Vision*. Curr Biol, 2018. **28**(15): p. R830-R832.
18. Re, D., G. Karvat, and A.N. Landau, *Attentional Sampling between Eye Channels*. J Cogn Neurosci, 2023. **35**(8): p. 1350-1360.
19. Gomez-Ramirez, M., et al., *Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: a human electrocorticographic investigation*. J Neurosci, 2011. **31**(50): p. 18556-67.
20. Rohenkohl, G. and A.C. Nobre, *alpha oscillations related to anticipatory attention follow temporal expectations*. J Neurosci, 2011. **31**(40): p. 14076-84.
21. Zoefel, B. and R. VanRullen, *Oscillatory Mechanisms of Stimulus Processing and Selection in the Visual and Auditory Systems: State-of-the-Art, Speculations and Suggestions*. Front Neurosci, 2017. **11**: p. 296.
22. Davidson, M.J., F.A.J. Verstraten, and D. Alais, *Walking modulates visual detection performance according to stride cycle phase*. Nat Commun, 2024. **15**(1): p. 2027.
23. Kaltenmaier, A., M.H. Davis, and C. Press, *Fixed and flexible perceptual rhythms*. Trends Cogn Sci, 2025.
24. Bouwer, F.L., *Neural Entrainment to Auditory Rhythms: Automatic or Top-Down Driven?* J Neurosci, 2022. **42**(11): p. 2146-2148.

25. Morillon, B. and C.E. Schroeder, *Neuronal oscillations as a mechanistic substrate of auditory temporal prediction*. Ann N Y Acad Sci, 2015. **1337**(1): p. 26-31.
26. Kok, P., P. Mostert, and F.P. de Lange, *Prior expectations induce prestimulus sensory templates*. Proc Natl Acad Sci U S A, 2017. **114**(39): p. 10473-10478.
27. Myers, N.E., et al., *Testing sensory evidence against mnemonic templates*. Elife, 2015. **4**: p. e09000.
28. Blom, T., et al., *Predictions drive neural representations of visual events ahead of incoming sensory information*. Proc Natl Acad Sci U S A, 2020. **117**(13): p. 7510-7515.
29. Kok, P., J.F. Jehee, and F.P. de Lange, *Less is more: expectation sharpens representations in the primary visual cortex*. Neuron, 2012. **75**(2): p. 265-70.
30. Yon, D., et al., *Action sharpens sensory representations of expected outcomes*. Nat Commun, 2018. **9**(1): p. 4288.
31. Tzourio-Mazoyer, N., et al., *Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain*. Neuroimage, 2002. **15**(1): p. 273-89.
32. Di Plinio, S., *Testing the Magnitude of Correlations Across Experimental Conditions*. Front Psychol, 2022. **13**: p. 860213.
33. Mostert, P., P. Kok, and F.P. de Lange, *Dissociating sensory from decision processes in human perceptual decision making*. Sci Rep, 2015. **5**: p. 18253.
34. Cannon, J.J. and A.D. Patel, *How Beat Perception Co-opts Motor Neurophysiology*. Trends Cogn Sci, 2021. **25**(2): p. 137-150.
35. Coull, J.T., *Discrete Neuroanatomical Substrates for Generating and Updating Temporal Expectations*, in *Space, Time and Number in the Brain*. 2011. p. 87-101.
36. Arnal, L.H., *Predicting "When" Using the Motor System's Beta-Band Oscillations*. Front Hum Neurosci, 2012. **6**: p. 225.
37. Morillon, B. and S. Baillet, *Motor origin of temporal predictions in auditory attention*. Proc Natl Acad Sci U S A, 2017. **114**(42): p. E8913-E8921.
38. Patel, A.D. and J.R. Iversen, *The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis*. Front Syst Neurosci, 2014. **8**: p. 57.
39. Ross, J.M., J.R. Iversen, and R. Balasubramaniam, *Motor simulation theories of musical beat perception*. Neurocase, 2016. **22**(6): p. 558-565.
40. Rimmele, J.M., et al., *Proactive Sensing of Periodic and Aperiodic Auditory Patterns*. Trends Cogn Sci, 2018. **22**(10): p. 870-882.
41. Boettcher, S.E.P., et al., *Output planning at the input stage in visual working memory*. Sci Adv, 2021. **7**.
42. Rohenkohl, G., et al., *Combining spatial and temporal expectations to improve visual perception*. J Vis, 2014. **14**(4).
43. Hendrikx, E., et al., *Visual timing-tuned responses in human association cortices and response dynamics in early visual cortex*. Nat Commun, 2022. **13**(1): p. 3952.
44. de Lafuente, V., et al., *Keeping time and rhythm by internal simulation of sensory stimuli and behavioral actions*. Sci Adv, 2024. **10**.
45. Haarsma, J., et al., *Expectations about presence enhance the influence of content-specific expectations on low-level orientation judgements*. Cognition, 2025. **254**: p. 105961.
46. Molinaro, N., et al., *Out-of-synchrony speech entrainment in developmental dyslexia*. Hum Brain Mapp, 2016. **37**(8): p. 2767-83.
47. Jimenez-Bravo, M., V. Marrero, and A. Benitez-Burraco, *An oscillopathic approach to developmental dyslexia: From genes to speech processing*. Behav Brain Res, 2017. **329**: p. 84-95.

48. Boets, B., et al., *Intact But Less Accessible Phonetic Representations in Adults with Dyslexia*. Science, 2013. **342**.
49. Amadeo, M.B., et al., *Time in schizophrenia: a link between psychopathology, psychophysics and technology*. Transl Psychiatry, 2022. **12**(1): p. 331.
50. Leszczynski, M. and C.E. Schroeder, *The Role of Neuronal Oscillations in Visual Active Sensing*. Front Integr Neurosci, 2019. **13**: p. 32.
51. Schroeder, C.E., et al., *Dynamics of Active Sensing and perceptual selection*. Curr Opin Neurobiol, 2010. **20**(2): p. 172-6.
52. Hogendoorn, H., *Voluntary Saccadic Eye Movements Ride the Attentional Rhythm*. J Cogn Neurosci, 2016. **28**(10): p. 1625-35.
53. Anil Meera, A., et al., *Reclaiming saliency: Rhythmic precision-modulated action and perception*. Front Neurorobot, 2022. **16**: p. 896229.
54. Sotiropoulos, G., A.R. Seitz, and P. Series, *Changing expectations about speed alters perceived motion direction*. Curr Biol, 2011. **21**(21): p. R883-4.
55. Cannon, J., *Expectancy-based rhythmic entrainment as continuous Bayesian inference*. PLoS Comput Biol, 2021. **17**(6): p. e1009025.
56. L'Hermite, S. and B. Zoefel, *Rhythmic Entrainment Echoes in Auditory Perception*. J Neurosci, 2023. **43**(39): p. 6667-6678.
57. Kleiner, M., et al., *What's new in Psychtoolbox-3*. Perception, 2007. **36**.
58. Inc., T.M., *MATLAB*. 2024.
59. Oostenveld, R., et al., *FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data*. Comput Intell Neurosci, 2011. **2011**: p. 156869.
60. Ferrante, O., et al., *FLUX: A pipeline for MEG analysis*. Neuroimage, 2022. **253**: p. 119047.
61. Stanislaw, H. and N. Todorov, *Calculation of signal detection theory measures*. Behavior Research Methods, Instruments & Computers, 1999. **31**.
62. Van Veen, B.D., et al., *Localization of Brain Electrical Activity via Linearly Constrained Minimum Variance Spatial Filtering*. IEEE Trans. Biomed. Eng., 1997. **44**(9).
63. Kriegeskorte, N., M. Mur, and P. Bandettini, *Representational similarity analysis - connecting the branches of systems neuroscience*. Front Syst Neurosci, 2008. **2**: p. 4.
64. Brookshire, G., *Putative rhythms in attentional switching can be explained by aperiodic temporal structure*. Nat Hum Behav, 2022. **6**(9): p. 1280-1291.
65. Rademaker, R.L., C. Chunharas, and J.T. Serences, *Coexisting representations of sensory and mnemonic information in human visual cortex*. Nat Neurosci, 2019. **22**(8): p. 1336-1344.

Acknowledgments

The authors would like to thank the FIL's imaging support team, particularly Dimitra Moraiti and Daniel Bates, as well as Dorottya Hetenyi for MEG scanning assistance. This work was supported by a Leverhulme Trust project grant (RPG-2022-358) and European Research Council (ERC) consolidator grant (101001592) under the European Union's Horizon 2020 research and innovation programme, both awarded to CP. The Wellcome Centre for Human Neuroimaging was supported by core funding from the Wellcome Trust (203147/Z/16/Z).

Author contributions

Conceptualization, A.K., Q.G., P.K., M.H.D, C.P.; data curation, A.K.; formal analysis, A.K.; funding acquisition, C.P.; investigation, A.K.; methodology, A.K., Q.G., P.K., M.H.D, C.P.; project administration, A.K., C.P.; software, A.K.; resources, C.P.; validation, A.K.; visualization, A.K.; writing – original draft, A.K.; writing – review and editing, A.K., Q.G., P.K., M.H.D, C.P.