scientific reports

OPEN

The order of multisensory associative sequences is reinstated as context feature during successful recognition

Marike Christiane Maack, Jan Ostrowski & Michael Rose ✉

The ability of the human brain to encode and recognize sequential information from different sensory modalities is key to memory formation. The sequence in which these modalities are presented during encoding critically affects recognition. This study investigates the encoding of sensory modality sequences and its neural impact on recognition using multivariate pattern analysis (MVPA) of oscillatory EEG activity. We examined the reinstatement of multisensory episode-specific sequences in $n = 32$ participants who encoded sound-image associations (e.g., the image of a ship with the sound of a frog). Images and sounds were natural scenes and 2-second real-life sounds, presented sequentially during encoding. During recognition, stimulus pairs were presented simultaneously, and classification was used to test whether the modality sequence order could be decoded as a contextual feature in memory. Oscillatory results identified a distinct neural signature during successful retrieval, associated with the original modality sequence. Furthermore, MVPA successfully decoded neural patterns of different modality sequences, hinting at specific memory traces. These findings suggest that the sequence in which sensory modalities are encoded forms a neural signature, affecting later recognition. This study provides novel insights into the relationship between modality encoding and recognition, with broad implications for cognitive neuroscience and memory research.

The ability to remember episodes from the past is a cornerstone of human memory. Episodes consist of multiple features that may stem from different modalities (e.g., visual, auditory), reaching us in specific sequences. In line, it has been shown that humans encode not only the semantic content but also the temporal order (sequence) of features, which is crucial for recalling the flow of a past episode[1,2]. The capacity to encode and retrieve sequential information allows us to mentally rebuild the dynamic structure of events, highlighting the role of temporal context in episodic memory[1,3]. Recalling the order that features of an episode were originally encoded in, is based on sequential reinstatement[4].

Episodic memory relies on the integration of contextual information during encoding, with the hippocampus playing a key role in binding event features such as sensory modality and spatial-temporal context[5–7]. Prior research has investigated how we remember the temporal order of events, including the role of unimodal cueing[8–10]. Evidence suggests that episodic memory involves temporal compression and event segmentation, where the hippocampus supports memory organization by structuring event sequences and contextual boundaries[9,11]. Temporal compression refers to the tendency of episodic memory to condense events during recall, influenced by event segmentation and goal-directed actions[12–15]. Event segmentation, in turn, affects how temporal order is remembered, as events chunked at perceptual boundaries enhance object-context binding but may reduce precise temporal order memory[9,16]. The hippocampus further supports the encoding and retrieval of event sequences, integrating spatial and temporal contexts essential for remembering event order[17,18]. In line, episodic memory retrieval is shaped by the availability of contextual information at encoding, with reinstatement of encoding context enhancing recognition in providing characteristic cues that mitigate interference[19–21]. Sensory modality sequences, as part of contextual information, contribute to the organization of memory representations[22,23]. The hippocampus is crucial for integrating these contextual elements, facilitating recognition by reactivating modality-based associations rather than strictly reconstructing event sequences in order[18,24,25].

While previous studies have explored unimodal cueing and temporal order memory, the current study specifically investigates whether the modality sequence acts as a contextual feature that influences encoding and

Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, Building W34, 20248 Hamburg, Germany. ✉email: rose@uke.de

recognition rather than sequential recall per se. Contextual cues, including sensory modality sequences, can mitigate interference effects and improve memory performance by reinstating elements of the original encoding context[21,26,27]. Conversely, mismatches between encoding and retrieval contexts may lead to competition between overlapping memory traces, impairing recall[4,28,29]. Both physical and mental reinstatement of contextual information facilitate episodic retrieval, with mental reconstruction yielding comparable benefits to direct environmental cues[30,31]. Furthermore, cognitive control plays a role in sustaining contextual reinstatement, as individuals with higher working memory capacity are better able to maintain contextual associations through strategic memory processes[32].

Beyond unimodal sequences, results from the animal, as well as human studies, suggest that sequential information plays a critical role in recalling multisensory episodes, where encoded features from different sensory modalities (e.g., an image-sound pair) are reinstated during retrieval[33–36]. Multisensory features enable the brain to store and retrieve information across various perceptual domains, enhancing our ability to recall past experiences and make informed decisions[37–39]. As such, the human memory has evolved to function optimally under multisensory conditions[40]. Interestingly, multisensory memories can be cued by unimodal features upon remembering, suggesting that multisensory encoding enhances subsequent recognition[39,41]. This means that one modality may serve as a cue to retrieve the complete episode, even if it includes features from different modalities[42]. It is, however, unclear so far whether and how one modality (i.e., auditory) relates to the reinstatement of another (i.e., visual). The Scene Reconstruction Theory suggests that especially the hippocampus helps to reconstruct memories by integrating various sensory details associated with an event[4]. This integration is supposed to allow richer, and more detailed, memory formation, as the brain can draw upon multiple sensory inputs to create a coherent narrative of the event[11,43].

While neural plasticity allows multisensory learning[44], understanding the neural reinstatement of sequential information across sensory modalities remains crucial for elucidating the broader mechanisms of remembering. Recall generally involves the activation and reconstruction of neural pathways tied to previously encoded details[45]. This process is influenced by retrieval cues and familiarity with the material[46]. Beyond sequential and multisensory reinstatement, the recollection of contextual details surrounding past events, such as where and when they occurred, significantly contribute to the liveliness and specificity of the memory representation[47–50]. The Context Maintenance and Retrieval (CMR) model here provides a framework to understand how the brain organizes memories around contextual and temporal cues, facilitating the accurate retrieval of episode-specific feature sequences[51,52], propelling the reconstruction of the correct order and updating of associations. In line, previous research has highlighted that the process of episodic remembering involves not only recalling specific items but also reinstating the contextual details of the original event[53–55]. These studies have demonstrated that the success of memory retrieval is closely associated with the reactivation of the encoding-related memory trace[52,56–58]. Interestingly, it has been shown that during memory reactivation, not only episode-specific features but also contextual features, that were not directly related to the current memory task, are reinstated[29,59–65]. However, the precise neural mechanisms underlying this sequential reinstatement, especially in multisensory contexts, remain elusive.

EEG is a powerful tool to track these reinstatement processes including multisensory episode-specific feature sequences. Here, especially multivariate temporal-pattern analysis has emerged as the gold-standard to examine how neural activity during retrieval reflects (sequential) reinstatement. Moreover, it represents a tool for investigating the role of context reinstatement in memory processes, revealing the (beneficial) effects of reinstating neural encoding patterns in memory retrieval[30,64]. Accordingly, multivariate pattern analysis (MVPA) has been used to decode oscillatory activity patterns during memory retrieval, successfully classifying specific neural signatures tied to remembering[30,64]. Here, low-frequency activity (e.g., beta (13–30 Hz) and theta (3–7 Hz) oscillations) have been shown to be particularly important for episodic memory processes, facilitating successful retrieval[66–70]. Multisensory inputs from different modalities as well as their sequential encoding thereby enrich the formation of stable memory traces[39,41]. As multivariate approaches have shown that neural pattern reinstatement is indeed associated with episode-specific feature sequences, this study aims to investigate the oscillatory mechanisms underlying the retrieval of sequential information in human memory, focusing on whether the modality sequence in which information was presented during encoding is reinstated during recognition. Specifically, we utilize EEG and MVPA to classify the neural patterns associated with modality sequence reinstatement, providing new insights into the role of oscillatory activity in organizing and retrieving sequential memories across different sensory modalities. Importantly, the current study aims to demonstrate how modality sequence functions as a contextual feature during encoding and recognition (but not temporal reinstatement in the sense of reactivating the sequential order during retrieval; i.e., first visual, second auditory). This design offers a novel perspective on context memory in representing a sharp contrast to previous designs, which employed a parallel presentation of modalities during encoding[71,72].

Methods
Participants
Thirty-six healthy participants were recruited for the experiment. Data from four participants were excluded due to a high number of missing trials ($n=1$) and poor behavioral performance ($n=3$). Outliers in terms of memory performance (d') were identified and excluded if they exceeded ± 3 absolute deviations from the median (MAD;[73]). Therefore, the final sample included $n=32$ participants (19 females, 52.77% female) with a mean age of 24.25 years (SD = 3.34), ranging from 18 to 33 years. All participants had normal or corrected-to-normal vision and hearing ability and reported no neurological or psychiatric diseases. They gave written informed consent and received financial reimbursement for participating in the study. The Hamburg Medical Council ethics committee (PV5893) approved this investigation. We confirm that all research was performed in accordance with relevant guidelines and regulations.

Task and procedure

We implemented an explicit sequential associative memory task consisting of an encoding task, a short intermission, and a subsequent recognition task. In order to measure associative multisensory memory, participants were presented sequential image-sound pairs during the encoding task. The images (resolution: 640 × 480 pixels; 24-bit color depth) and sounds (length: 2 s; bitrate: 1411kBit/s) were randomly selected from an internal stimulus database, presenting real-life objects, animals, and landscapes. The individual stimuli were paired pseudo-randomly (semantically congruent stimuli pairs were excluded). Each pair was presented once during the encoding task. In each encoding trial, the stimulus pairs were presented sequentially. Each trial started with a 1-second *modality cue* indicating whether a visual or an auditory stimulus would be presented as the first pair component (Fig. 1). The *modality cue* was represented as an icon, cueing the following modality (image or tone). Additionally, the icon involved a red number, stating whether the next stimulus would be the first or second stimulus of the pair. The *modality cue* was followed by a red fixation cross for 2 s, which served as a visual cue for the upcoming stimulus. Afterwards, the first stimulus was presented, followed by a 500 ms Inter-Stimulus-Interval (ISI). Complementary to the first stimulus presentation, a second *modality cue* indicated whether a visual or an auditory stimulus would be presented as the second stimulus. Again, a red fixation cross served as a visual cue for the upcoming stimulus followed by the announced stimulus and the inter-trial interval (ITI). The ITI was jittered between 3 s and 5 s. Trials were counterbalanced for *modality-order*, such that the encoding task consisted of the same number of visual-auditory (VA) as auditory-visual (AV) trials. We applied a modality cueing procedure, including both modality-specific cues and numerical indicators, to clearly signal the upcoming auditory and visual pairings, thereby facilitating robust and explicit encoding of the stimulus associations. All participants were explicitly instructed to memorize the stimulus pairs, and not to focus on the individual images or sounds. This ensured that subsequent memory would later reflect associative memory, but not item memory. We divided the experiment into three consecutive blocks, each consisting of an encoding and subsequent recognition task. During each encoding task, 47 stimulus pairs were presented. After the encoding task, a short 3-minute intermission followed, during which participants were asked to count down aloud from 100 (115 and 125 in the second and third run, respectively) in steps of 7 (9 and 13 in the second and third run, respectively). In the subsequent recognition task, participants were presented with the 47 stimulus pairs shown during the previous encoding phase, as well as 47 new pairs. Both components of a pair were presented in parallel, in sharp contrast to the encoding task, where the components were shown sequentially. New pairs consisted of the same individual components that rendered the pairs from the encoding task, but were shuffled to create 47 new pairs. The participants were asked to indicate via button press whether the presented pairs were already known from the encoding task or not. Stimulus presentation lasted for 2 s, and no cue was used. The inter-trial interval was fixed to 5 s. The recognition task was followed by a short break of 3 to 5 min, followed by the encoding task of the next run. Across the three blocks, 141 stimulus pairs were presented during encoding, and 282 were presented during recognition in total.

EEG data acquisition

EEG data were collected using a 64-channel Ag/AgCl active electrode system (ActiCap64; BrainProducts, Gilching, Germany), arranged in accordance with the extended 10–20 system[74]. Sixty electrodes were positioned at the most central scalp locations. To facilitate offline artifact removal, a bidirectional bipolar electrooculogram (EOG) was concurrently recorded using the remaining four electrodes. These bipolar EOG electrode pairs were positioned above and below the left eye, as well as at the lateral ends of the bicanthal plane. FCz served as the reference electrode for data acquisition, while the ground electrode was situated at position Iz. Signals were digitized at a sampling rate of 500 Hz and was amplified with a low cut-off frequency of 0.53 (0.3 s time constant). Impedances were maintained below 10 kΩ throughout the recording session.

EEG preprocessing and time-frequency decomposition

The acquired EEG data were preprocessed offline using the FieldTrip toolbox[75] in MATLAB (Release 2022a, The MathWorks Inc., Natick, Massachusetts, USA). For each participant, the encoding and the recognition task were analyzed separately. For the recognition task, epochs were extracted from −2500 ms to 3500 ms relative to stimulus onset, resulting in a trial duration of 5 s. A high-pass filter at 0.5 Hz was applied to remove extreme low-frequency fluctuations. The data were visually inspected, and trials containing artifacts, such as high-frequency noise indicating muscular activity or spikes resembling poor electrode connections, were removed. Independent Component Analysis (ICA) was used to identify components corresponding to blinks and other ocular activity, and the data were corrected accordingly. On average, 4.8 components (SD = 3.31) were removed per subject. The data were then visually inspected again, and trials with artefacts were excluded. The remaining trials for each subject were split into correct old (remembered; mean = 83.11; SD = 26.40), incorrect old (forgotten; mean = 46.94; SD = 24.08), correctly rejected (mean = 112.46; SD = 17.05) and false alarm (mean = 18.60; SD = 14.78) trials. The correct old trials were additionally split for the "visual-auditory" (VA) and "auditory-visual" (AV) conditions according to modality order in the encoding task, resulting in a similar number of trials in each condition (44 trials on average in each condition; AV: mean = 44.75, SD = 10.79; VA: mean = 44.41, SD = 11.02). Time-frequency decomposition was performed from −1 to 2 s relative to stimulus onset, covering the 1 to 40 Hz frequency range. This was achieved using a multitaper convolution approach with a sliding Hanning window of 500 ms and a 100 ms step size. Although a 500 ms window has an intrinsic frequency resolution of 2 Hz (1/0.5 s), we applied zero-padding to the maximum trial length, which interpolated the spectrum to a 1 Hz grid, thereby effectively balancing temporal precision with frequency resolution. No baseline correction was applied for the recognition tasks, as the primary focus was on within-subject differences in oscillatory power between the remembered and forgotten pairs. We did not apply baseline correction to preserve potential pre-stimulus effects, which have been implicated in learning and memory processes. Additionally, baseline correction assumes a proportional scaling
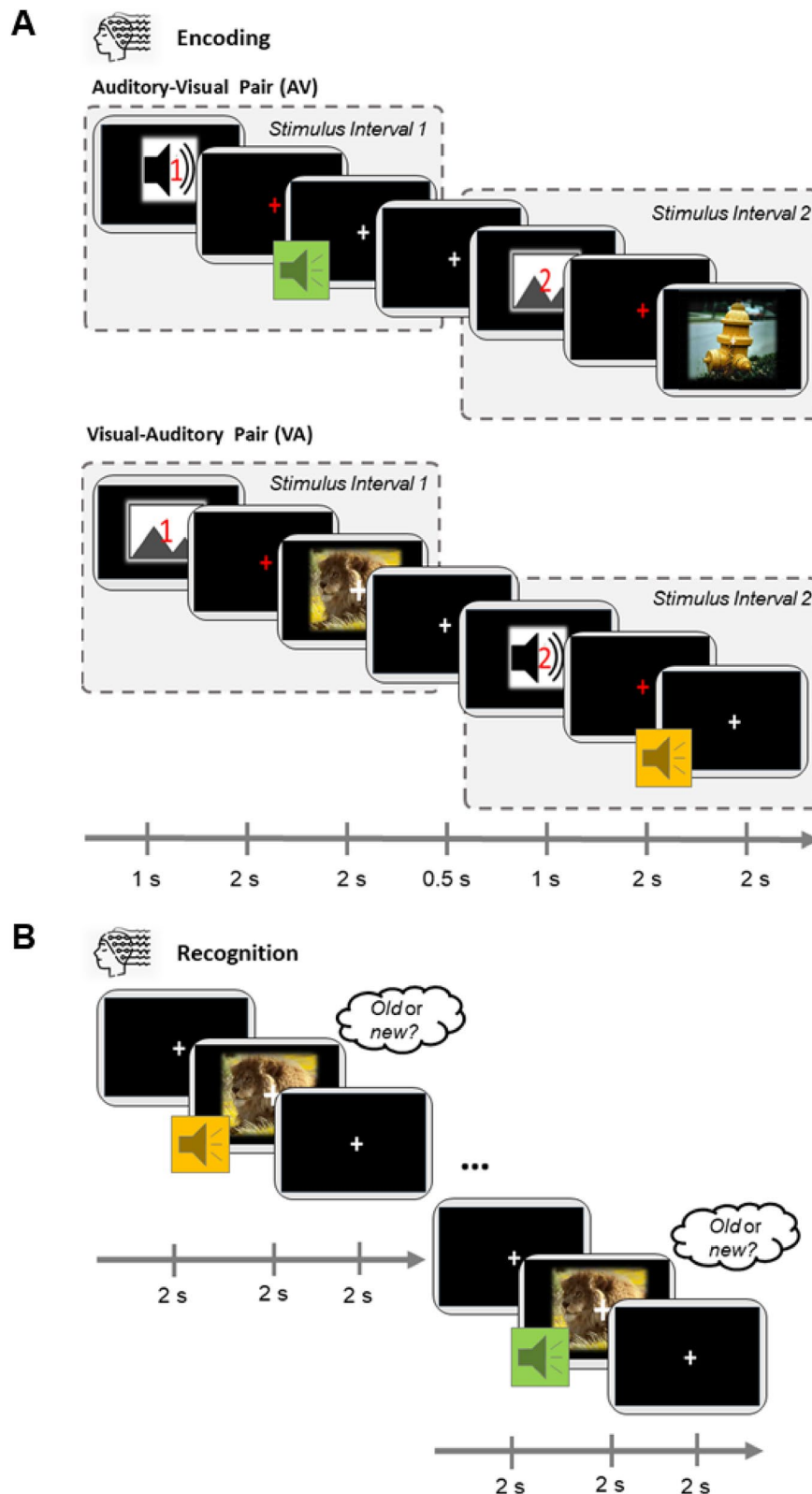
of oscillatory and non-oscillatory (1/f) activity, an assumption that may not always hold. As a result, the reported effects should be interpreted in terms of relative power changes rather than absolute polarity shifts. Additionally, for the data of the recognition task, the resulting frequency spectra (electrode × frequency × time) from stimulus onset onward, reflecting the processing phase, were used to predict the modality order in the encoding task for each participant and condition. Only trials with successfully remembered items were included in the analysis, while pre-stimulus activity was not included. For the multivariate analyses, the Matlab toolbox for classification and regression of multi-dimensional data (MVPA-Light;[76] was used. A support vector machine (SVM) with a

◄**Fig. 1.** Schematic overview of the encoding task in both sequence variations as well as the following recognition task. (**A**) The two different modality order sequences in the encoding task. Stimuli were presented sequentially. The trial started with a cue, indicating which stimulus modality is presented first followed by a red fixation cross. The red fixation cross indicated that a stimulus would be presented in the next 2 s. Afterwards, a complex natural image was presented. The presentation of the second stimulus of the pair followed the same procedure. We utilized two different encoding sequence orders: either the sound was present first and the image afterwards (auditory-visual; AV-condition or the image was presented first and the sound afterwards (visual-auditory; VA-condition). For the analysis of the encoding task, the task was subdivided into Stimulus Intervals 1 and 2, indicated by the dashed boxes. (**B**) One example trial from the recognition task. Unlike the encoding phase, stimulus pairs were presented simultaneously. Participants had to identify whether the pair **had** been presented during encoding (old) or not (new) within 4 s. The recognition task pairs were composed of the stimuli from the encoding resulting in new and old pairs.

k = 5-fold cross-validation was used for classification with a five-time repetition. The classifier was trained on all electrodes using single-trial frequency spectra.

The preprocessing routine and the time-frequency analysis for the encoding task were the same as for the recognition task, with deviations during epoching. The deviation in epoching procedure resulted from the difference in presentation mode. While sound-image associations were presented simultaneously in the recognition task, the associations were presented sequentially in the encoding task. Epochs were extracted from the first modality cue until 2 s after the onset of the second stimulus, resulting in a trial duration of 10.5 s. Using ICA, on average 4.8 independent components were rejected from each individual data set (SD = 2.61). After preprocessing, the remaining trials for each subject were split into a "visual-auditory" and "auditory-visual" condition according to modality order in the encoding task, resulting in a similar number of trials in each condition. These trials were then used in the subsequent analyses of later remembered and forgotten trials (55 trials on average in each condition). After time-frequency decomposition, data were averaged separately for remembered and forgotten trials for each participant.

Statistical analysis

In the recognition task, the percentages of remembered old pairs (hits), correctly rejected new pairs, forgotten old pairs, and falsely remembered new pairs (false alarms) were extracted. In order to index memory performance, we utilized the sensitivity measure d', which is the difference between the z-transformed hit and false alarm rates[77–79]. A one-sample t-test against zero was conducted to probe associative memory formation with d' as dependent variable. Additionally, a repeated-measures ANOVA was used to analyze reaction times during the recognition task, with the within-subject factors *Pairing* (old vs. new) and *Correctness* (correct: remembered or correct rejection vs. incorrect: forgotten or false alarm).

Statistical analysis of the time-frequency EEG data acquired during the recognition phase of the experiment was conducted to explore the potential effect of remembered and forgotten trials within the low-frequency spectrum. This analysis was further differentiated by modality-independent sequence order during encoding, as well as sequential auditory-visual and visual-auditory presentations. Employing a non-parametric permutation testing approach with cluster-based correction for multiple comparisons, as implemented in the FieldTrip toolbox[75], we statistically compared time-frequency representations corresponding to remembered trials against those of forgotten trials. The cluster-based permutation test defines a cluster as a set of contiguous significant points in a three-dimensional space comprising electrode location (spatial dimension), frequency, and time. A significant point is identified when the test statistic at a specific electrode, frequency, and time point surpasses a predefined threshold (p < .01, uncorrected). These significant points are then grouped into clusters based on their adjacency. This means that a cluster can extend across multiple electrodes, frequency bands, and time points, rather than being restricted to a single dimension. We used the electrode neighbourhood structure defined by Fieldtrip function ft_prepare_neighbours to determine adjacency, ensuring that spatially close electrodes are considered neighbors. In the frequency and time dimensions, adjacency is defined by consecutive frequency bins and time points. To be classified as a cluster, an effect needed to span at least two neighbouring electrodes., preventing isolated effects in one electrode from being classified as clusters. Multiple comparison correction was applied using the cluster-based permutation test (*cfg.correctm* = 'cluster', *cfg.method* = 'montecarlo'), which controls the family-wise error rate (FWER). This means that while individual points initially pass a cluster-forming threshold (p < .01, uncorrected), the final significance of a cluster is determined via a permutation-based correction (p < .05 two-tailed, cluster-corrected). Thus, only clusters that survived this multiple comparison correction are reported as significant. Subsequently, Monte Carlo method was utilized to generate a distribution of t-values[80].

The main focus of the study was to probe neuronal reinstatement of the stimulus modality order (visual-auditory, auditory-visual) during subsequent recognition via MVPA. To investigate this effect, we employed MVPA in the time-frequency domain of the recognition task. The SVM classifier distinguished between the two different encoding stimulus modality orders (AV/VA) based on the EEG data from the recognition task in which the previously encoded associations were presented simultaneously. To comprehensively capture neural processing during the recognition phase, we applied MVPA to the entire a priori defined dataset, analyzing the full trial period during recognition from 0 to 2 s relative to stimulus onset, across the 1–40 Hz frequency range and all 60 electrodes. Classification accuracy was assessed using a single–subject k = 5-fold cross-validation procedure, ensuring that model training and testing were performed on separate data splits within each subject to reduce overfitting and improve generalization. We conducted the MVPA on remembered trials of the
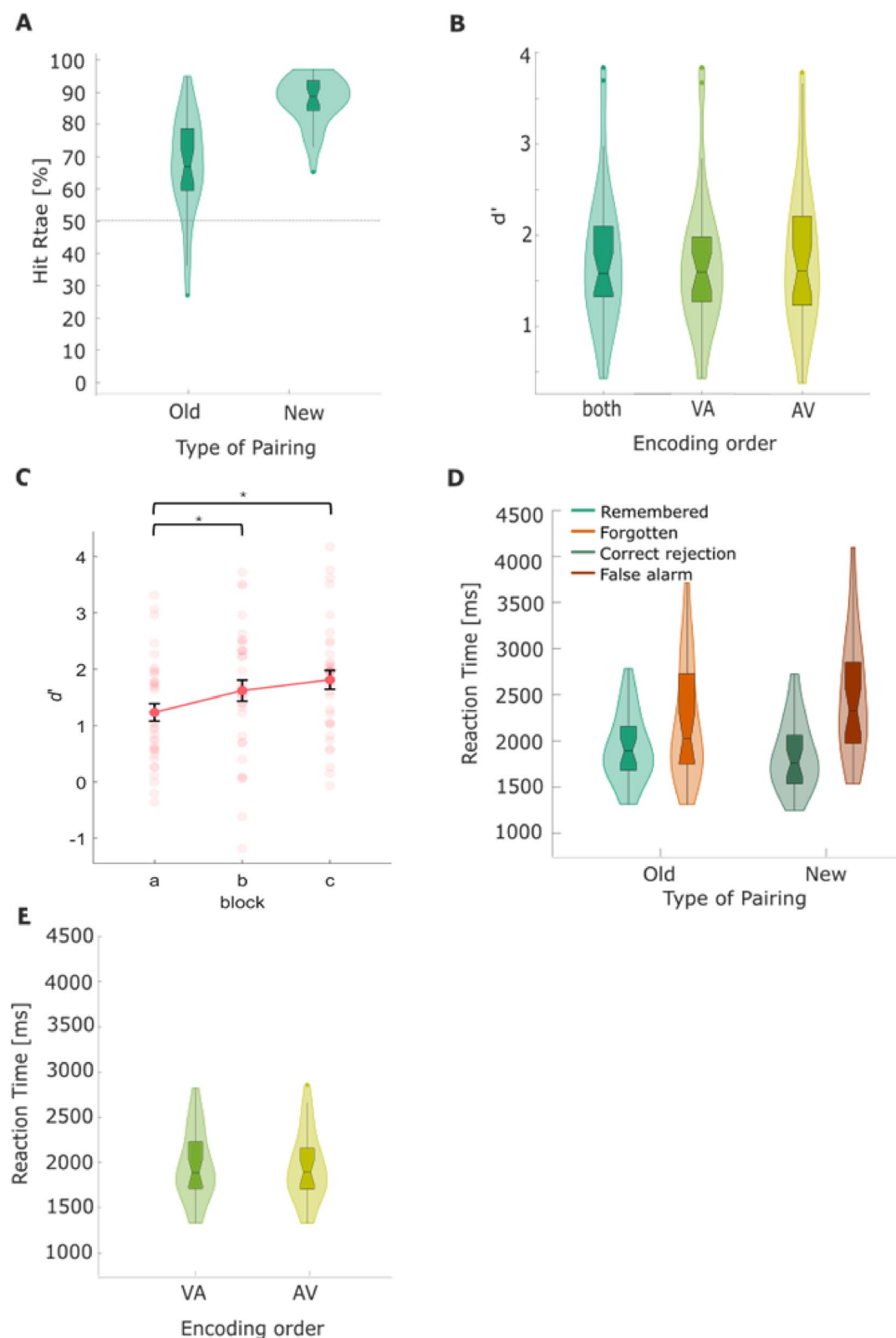
recognition task to classify the two different modality orders in which the stimulus pairs were presented during the encoding task (*Modality Sequence Classifier*). To assess the statistical significance of the accuracy achieved by the *Modality Sequence Classifier*, we conducted a one-sample *t*-test (one-tailed) comparing the overall classifier accuracy against chance-level (50%) The classifier accuracy was determined by averaging the individual classifier performance within the entire analysis window. This methodology enabled us to assess whether the performance of the *Modality Sequence Classifier* significantly exceeds the chance classification level, thus providing insight into the presence of meaningful patterns associated with memory retrieval as opposed to random classification. Subsequently, to probe statistically significant accuracy of the *Modality Sequence Classifier* across the entire

Fig. 2. Behavioral results from the recognition task. (**A**) Hit-rate in the recognition task for old and newly rearranged pairs. Within the boxplots, the horizontal lines indicate the median of the subset, while the notch around the median represents its 95% confidence interval. The upper and lower edges indicate quartiles 1 and 3. (**B**) The distribution of memory performance (*d'*) overall, as well as separated between the two modality-order sequences during the encoding task (VA = visual-auditory, AV = auditory-visual). (**C**) The block effect for memory performance (d'). Asterisks indicate significant differences between blocks. $p < .05$(*). (**D**) Distribution of reaction times for the respective response categories from the recognition task, split for remembered, forgotten, correct rejection and false alarm trials. (**E**) Distribution of reaction times for the respective response categories from the recognition task, split between the two modality-order sequences during the encoding task (VA = visual-auditory, AV = auditory-visual) for the remembered trials.

analysis window, we performed a cluster-based permutation t-test, comparing the accuracy values from the *Modality Sequence Classifier* to chance level (50%). To further examine frequency-specific effects, accuracy values were averaged within predefined frequency bands: theta (3–7 Hz), alpha (8–13 Hz), low beta (13–21 Hz), high beta (22–32 Hz), beta (13–32 Hz), and gamma (32–40 Hz). We determined the electrode with the highest mean t-value for each band and computed individual mean accuracy values at these electrodes.

Furthermore, a correlational analysis was conducted to examine the relationship between the oscillatory power contrast between remembered and forgotten trials during encoding task and d' (memory performance) estimates from the recognition task. This analysis is expected to yield valuable insights into multisensory processing during associative learning and memory. Drawing on previous research[81,82], we expected power differences to correlate with the memory performance within the low-frequency bands during the presentation of the first and second stimuli. Specifically, these power differences were expected to vary according to the modality sequence order (AV vs. VA condition). To accomplish this objective, we employed a non-parametric cluster-based permutation technique for the correlation analysis to control for alpha-error inflation. Here, we calculated the neuronal activity power differences (remembered > forgotten; AV/VA condition) and correlated these with the behavioural measure d'. Cluster-based permutation t-tests were conducted for each time-frequency data point across channels and participants. Significant differences between conditions ($p < .05$) resulted in adjacent data points being grouped into clusters based on temporal, spatial, and spectral criteria.

Results
Successful acquisition of sequentially encoded multisensory associative pairs
Participants completed one recognition task after each encoding task, which consisted of 47 multisensory associative pairs (Fig. 1). In the recognition task, participants were presented pairs of images and sounds in parallel, which were sequentially presented during previous encoding. This task included previously presented pairs (*old*) and tested memory specificity by presenting newly formed pairs, consisting of old stimuli elements (*new*). Participants had to indicate whether they remembered the presented stimulus pair from the encoding task (old) or whether it was a newly rearranged pair, consisting of an old image with a sound previously paired with another image (new). Overall, participants performed very well in remembering *old* pairs, with an average hit rate of M(SD) = 67.19 (± 15.29%; Fig. 2A). Although the *new* pairs consisted of images and sounds from encoding that were now rearranged, the false alarm rate was low (M(SD) = 12.12 (±7.68%)). Accordingly, signal detection theory-based analysis confirmed robust learning, expressed by an average associative d' of M(SD) = 1.76 (±0.76) independent of stimulus modality order ($t_{(30)} = 12.69$, $p < .001$, *cohen's d* = 0.59). The d' estimates of auditory-visual (AV; M(SD) = 1.74 ± 0.76) and visual-auditory stimuli (VA; M(SD) = 1.77 ± 0.77) did not differ significantly ($t_{(30)} = -0.14$, $p = .884$, see Fig. 2B). Taken together, behavioral results confirmed the successful acquisition of sequentially encoded multisensory associative pairs, with no differences in performance due to modality order.

Our study was intentionally designed as an explicit learning and memory paradigm, where participants were specifically instructed to remember the pairs. Given the experimental structure, it was expected that participants adapt their strategies over time, mainly as they were aware that a recognition test followed each block. Crucially, due to the design, these improvements were likely to occur consistently across participants, regardless of their overall performance. Therefore, this effect should not be seen as a systematic bias but rather as an inherent characteristic of explicit learning. To assess whether behavioural performance changed across blocks, we conducted a repeated measures ANOVA on the performance measure d', which revealed a significant main effect of Block, $F_{(2, 70)} = 8.92$, $p < .001$, $\eta^2 = 0.05$. Given that Mauchly's test indicated a violation of sphericity ($W = 0.92$, $p = .239$), we applied the Greenhouse-Geisser correction ($\varepsilon = 0.925$), yielding a corrected significance value of $p < .001$. Post hoc tests further revealed that performance significantly differed between Block A and Block B ($p = .025$) as well as Block A and Block C ($p = .002$; Fig. 2C).

Next, we compared participants' reaction times during recognition as an index of memory confidence[83,84]. A two-way repeated-measures ANOVA revealed a significant main effect of the factor Correctness ($F_{(1,31)} = 69.84$, $p < .001$, $\eta^2 = 0.69$; Fig. 2D), indicating faster responses during remembered pairs, compared to forgotten. Furthermore, this analysis revealed a main effect of Pairing ($F_{(1,31)} = 9.24$, $p = .005$, $\eta^2 = 0.23$; Fig. 2D), indicating faster responses to old pairs, compared to new. Critically, we observed a significant Pairing × Correctness interaction ($F_{(1,31)} = 15.56$, $p = .001$, $\eta^2 = 0.33$; see Table 1). The post-hoc t-test revealed a significant decrease in reaction times for correctly recognised old pairs ($t_{(126)} = -0.90$, $p < .001$, *Cohen's d* = −0.44), indicating an increase in memory confidence in light of correctly retrieved associations compared to forgotten. The reaction time corresponding to auditory-visual (AV; RT = 1944.6 ± 381.7 ms) and visual-auditory stimuli (VA; RT = 1959.4 ± 385.9 ms) did not differ significantly ($t_{(62)} = 0.15$, $p = .877$, see Fig. 2E). Interestingly, participants correctly rejected new pairs significantly faster compared to mistakenly categorizing them as old (false alarm;

Effect	F(df)	p	Partial η²
Pairing (Old vs. New)	9.23(1,31)	0.005	0.23
Correctness (Correct vs. Incorrect)	69.84(1,31)	<0.001	0.69
Pairing × Correctness	15.56(1,31)	<0.001	0.33

Table 1. Repeated-Measures ANOVA results for reaction times.

Comparison	t(df)	p	p_{corr}
Remembered vs. forgotten	-2.26(62)	0.027	0.162
Remembered vs. correct rejected	0.81(62)	0.421	1
Remembered vs. false alarm	-4.30(62)	<0.001***	<0.001***
Correct rejected vs. forgotten	-2.83(62)	0.006**	0.036*
Correct rejected vs. false alarm	-4.88(62)	<0.001***	<0.001***
Forgotten vs. false alarm	-1.63(62)	0.110	0.660

Table 2. Post-hoc t-test results for reaction times. p_{corr} relates to p-values after Bonferroni correction. $p < .05$(*), $p < .01$(**), $p < .001$ (***).

$t_{(62)} = -4.88$, $p < .001$), supporting the idea that decision confidence might influence RTs in recognition memory. Also, the response in correctly rejected trials were significantly faster as compared to response in forgotten trials, $t_{(62)} = -2.84$, $p = .006$, suggesting that correctly detecting novel information is easier than failing to recognize old pairs. Responses to forgotten trials did not significantly differ in reaction times as compared to false alarm trials ($t_{(62)} = -1.63$, $p = .110$; Table 2).

Successful recognition of multisensory associations relies on alpha/beta oscillations

In the next step, we investigated the oscillatory power differences between remembered and forgotten multisensory pairs within the recognition task, which presented the previously sequentially encoded pairs simultaneously. Here, participants had to indicate whether the presented pair was previously shown in sequential order during encoding. In the following analysis, we split the recognition trials according to their modality-sequence during encoding (VA/AV), and conducted time-frequency analyses. In VA associations (remembered > forgotten) we observed a significant negative cluster of oscillatory activity covering the high theta to low beta frequency range (0.8–1.2 s after stimulus onset; 7–23 Hz; negative cluster: $p < .003$, SD = 0.001). This indicates that remembered pairs, which were represented in a VA sequence order during encoding, were associated with alpha and beta power during memory retrieval as compared to forgotten pairs. This effect was primarily driven by activity differences in frontotemporal and lateral-occipital areas (0.8–1.2 s after stimulus onset, 7–23 Hz; Fig. 3A). Interestingly, AV trials showed a different pattern, including a negative cluster in the theta and alpha range (0.5–1.9 s; 6–13 Hz; negative cluster 1: $p < .002$, SD = 0.009; negative cluster 2: $p < .045$, SD = 0.005), indicating that remembered pairs, which were represented in an AV sequence order during encoding, were associated with theta and alpha power during memory retrieval as compared to forgotten pairs. This effect was primarily driven by activity differences in parietal-occipital areas (0.5–1.9 s after stimulus onset, 6–13 Hz; Fig. 3B). The results indicate differential processes concerning the oscillatory processing of modality-sequences during recognition and were used to restrict the following MVPA analysis. When analyzing both conditions (combining AV and VA) together as an independent modality-sequence condition, we observed similar significant neuronal activity effects, further reinforcing the underlying processing patterns across modalities. The results revealed a negative cluster covering the theta (median cluster size = 10), alpha (median cluster size = 27), and beta (median cluster size = 18) bands, occurring 0.9 to 1.8 s after stimulus onset (negative cluster: $p < .001$, SD = 0.004; see Fig. 3C). This indicates that remembered pairs were associated with lower theta, alpha, and beta power during memory retrieval as compared to forgotten pairs. This effect was primarily driven by activity differences in frontotemporal and lateral-central areas (1.0 to 1.7 s after stimulus onset, 3–7 Hz) and in lateral parietal regions (0.9 to 1.7 s after stimulus onset, 8–18 Hz; see Fig. 3C).

Modality-sequences are reinstated as context-features during recognition

The neuronal signature within the recognition task indicated differential processing of the stimulus modality order, which we hypothesized to also be expressed as context-specific features of the underlying memory trace (i.e. the temporal sequence of the stimulus from different modalities). Accordingly, our main hypothesis stated that the neural signature during the recognition task would reflect the modality sequence in which the pairs were presented during the encoding task. To probe this effect, we employed MVPA, moving beyond the univariate comparisons of oscillatory power. The *Modality Sequence Classifier* distinguished between the two different encoding stimulus modality orders (AV/VA) based on the EEG data from the recognition task in which the previously encoded associations were presented simultaneously. Classification accuracy was assessed using a cross-validation (k = 5) procedure. The overall mean classifier performance for the *Modality Sequence Classifier* was 52.26%, which significantly exceeded the chance level of 50% ($t_{(31)} = 4.28$, $p < .001$, *cohen's d* = 0.76).

For the more detailed classification analyses, we calculated the classification accuracy over all electrodes for each frequency band within the significant data points, as well as the average classification accuracy for the electrode with the maximum mean t-value of the entire time within the specific frequency range. The cluster-based permutation test comparing MVPA accuracy values for remembered trials when decoding the two modality orders against chance level (50%) revealed eight significant positive clusters (most prominent cluster: $p < .001$, cluster-level statistic $t = 29.069$, SD = 0.0003, CI range = 0.0006). This cluster encompassed 45.302 data points out of a total of 50.400 (60 channels × 40 frequency × 21 time), covering approximately 89.88% of the analyzed search space. This significant cluster extended from stimulus onset to 2 s post-stimulus, spanning frequencies from 1 to 40 Hz, and was distributed across all electrodes, with the lowest representation in P7 (84.28%) and the highest in FC4 (93.69%). The remaining clusters had p-values ranging from 0.017 to 0.047, with their respective cluster statistics and confidence intervals indicating robust effects across multiple frequency bands and time points. No significant negative clusters were detected.

To further interpret the decoding performance, average accuracy values were computed over the significant time points identified by the cluster-based permutation test, focusing on time points and electrodes with the highest mean t-values. To identify the electrodes exhibiting the strongest effects, we computed the mean t-value for each electrode by averaging all significant t-values ($p < .05$) across frequency bands and time points. This metric served as an index of the relative effect size at each electrode, highlighting regions that consistently demonstrated robust neural discrimination effects in the cluster-based permutation test. By focusing on electrodes with the highest mean t-values, we aimed to characterize the spatial distribution of the most pronounced neural decoding effects during the recognition task. The overall mean accuracy across all frequencies (1–40 Hz) was 52.26%, with the highest mean t-value observed at electrode F4 ($t = 4.18$). The mean individual accuracy at this electrode was 52.32%, while the highest individual accuracy value reached 62.38% at 0.8 s and 30 Hz. The maximal t-value was reached at F4 ($t = 6.58$).
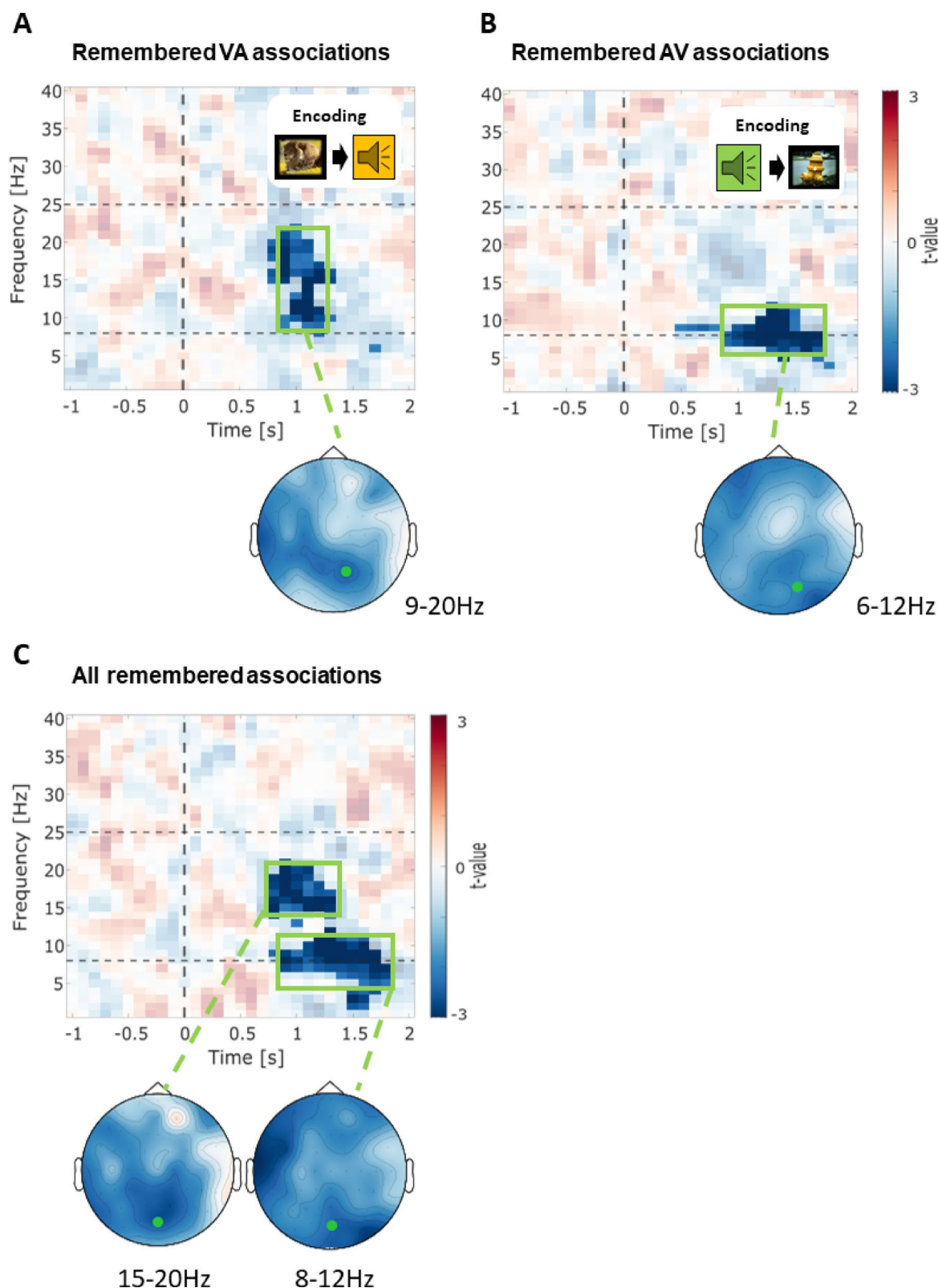
When examining specific frequency bands, the theta range (3–7 Hz) yielded the highest mean accuracy of 52.49%, with the strongest effect at electrode AF7 ($t = 4.22$), and an individual peak accuracy of 58.80% at 0.4 s and 7 Hz. The maximal t-value was reached at AF7 ($t = 6.29$). In the alpha band (8–13 Hz), the mean accuracy was 52.40%, with the highest t-value recorded at electrode Fpz ($t = 3.92$), and an individual maximum accuracy of 58.67% at 0.1 s and 11 Hz. The maximal t-value was reached at F4 ($t = 6.58$). The low beta range (13–21 Hz) showed a mean accuracy of 52.29%, with the most significant effect at electrode Fpz ($t = 3.80$) and a peak accuracy of 58.63% at 0.4 s and 21 Hz. The maximal t-value was reached at FC2 ($t = 5.93$). Similarly, the high beta range (22–32 Hz) demonstrated a mean accuracy of 52.22%, with the strongest effect at electrode P1 ($t = 4.27$) and a peak accuracy of 58.29% at 1.9 s and 24 Hz. The maximal t-value was reached at CP2 ($t = 6.31$).

When considering the full beta range (13–32 Hz), the mean accuracy was 52.24%, with the highest t-value at P1 ($t = 4.07$) and an individual maximum accuracy of 61.00% at 1.9 s and 24 Hz. The maximal t-value was reached at CP2 ($t = 6.31$). The mean accuracy in the gamma band (32–40 Hz) was 52.23%, with the effect at electrode P2 ($t = 4.69$) and an individual peak accuracy of 52.98% at 0.1 s and 33 Hz. The maximal t-value was reached at PO8 ($t = 6.18$). All results are summarised in Table 3. These findings suggest that decoding performance was significantly above chance level across multiple time points, frequencies, and electrode locations, with particularly strong effects in the theta, alpha, and beta bands.

To further explore the spatial distribution of significant effects, we visualized topographical maps of t-values obtained from the cluster-based permutation test for different frequency bands (theta, alpha, low beta, high beta, and gamma) over the trial time course (Fig. 4). Interestingly, the higher beta and gamma frequency range (21–40 Hz) exhibited pronounced discriminative power between the two conditions during the entire stimulus presentation over the centro-parietal electrodes, specifically strongest at the beginning (0.4 to 1.3 s) and the end (1.7 to 2.0 s). Additionally, also early (0 to 1.0 s to stimulus onset) centro-frontal electrode cluster in the lower frequency range (3 to 20 Hz) showed a significant classification performance. The results demonstrate a convergence between the neural activity patterns during successful recognition and retrospective discrimination between the modality order of the sequential encoding during recognition. In sum, these findings highlight the ability to encode and differentiate VA from AV sequences during the retrieval task, as evidenced by the distinct neural patterns observed in the EEG data during subsequent recognition. In addition to the averaged accuracy values from electrodes with the highest mean t-values, time-resolved decoding accuracy (MVPA) for remembered trials is shown at exemplary electrodes across different frequency bands (Fig. 5).

Successful recognition of multisensory associations relies on low-frequency oscillations during encoding

As an explorative analysis, we analyzed the dynamics between the neuronal activity from the encoding task and the recognition performance. Thus, we asked whether specific oscillations during encoding propel successful memory formation. First, we computed the differences in oscillatory power between later remembered and forgotten trials separately for the presentation of the first and second stimulus of each pair. Therefore, we focused on the 2-s period before and after each stimulus onset for the VA and AV conditions, resulting in two analysis time windows (Stimulus Interval 1 and Stimulus Interval 2; Fig. 1). Initially, we analyzed oscillatory data independent of the stimulus modality order, dividing Stimulus intervals 1 (SI1) and 2 (SI2). The differential time-frequency spectra between remembered and forgotten pairs were computed in each participant. These difference values were correlated with associative d' values from the recognition task. Results revealed significant correlations within SI2 for both encoding sequence conditions. During the visual stimulus of AV pairs, two clusters of significant negative correlation between associative d' and average power differences (remembered > forgotten) were revealed at multiple electrodes in the parietal and central region in the alpha band (pre-stimulus positive cluster: $p < .042$, SD < 0.005; post-stimulus negative cluster: $p < .013$, SD = 0.003, Fig. 6A) at multiple electrodes in the central-parietal region (Fig. 6A) within the alpha range. In VA pairs, however, during the presentation of

A Remembered VA associations

B Remembered AV associations

9–20Hz

6–12Hz

C All remembered associations

15–20Hz   8–12Hz

the auditory stimulus, we observed a significant positive post-stimulus correlation ($p < .006$, SD = 0.002; Fig. 6B) and negative pre-stimulus correlation ($p < .018$, SD = 0.003) at multiple electrodes in the parietal-occipital region (Fig. 6B) within the alpha and beta range.

Discussion

Successful retrieval of events is strongly bound to the context of encoding[55,85,86]. While it is well established that context can be reflected as the surrounding environment[31,87,88], it can also be represented differentially, i.e. as the sequence of episode-specific features. In line, context feature reinstatement (of sequential information)

◀**Fig. 3**. Memory effects on time-frequency power before and during recognition for electrodes with maximal t-value for corresponding encoding stimulus modality order. (**A**) Time-frequency plot of the statistical comparison of REMEMBERED > FORGOTTEN oscillatory power at P2 for recognized associations, which were presented in the visual-auditory sequence during encoding. Opaque data points show a significant difference at $p < .05$ (corrected). The lower panel shows the topographical distribution within the significant cluster during stimulus presentation in the theta-alpha range (0.9 to 1.2 s; left). The green marker illustrates P2. (**B**) Time-frequency plot of the statistical comparison of REMEMBERED > FORGOTTEN oscillatory power at PO4 for recognized associations, which were presented in the auditory-visual sequence during encoding. Opaque data points show a significant difference at $p < .05$ (corrected). The lower panel shows the topographical distribution within the significant cluster during stimulus in the alpha/beta range (0.8 to 1.4 s). The green marker illustrates P2. (**C**) Top: Time-frequency plot of the statistical comparison of REMEMBERED > FORGOTTEN oscillatory power at POz, independent of modality order presentation during encoding. The vertical line marks the stimulus onset, and the horizontal lines mark the frequency bins alpha and beta (8 Hz and 25 Hz). Opaque data points show a significant difference at $p < 0.5$ (corrected). Negative t-values signify higher power in FORGOTTEN trials. The lower panel shows the topographical distribution within the significant cluster during stimulus in the beta range (0.8 to 1.3 s; left) and the significant cluster in the alpha range (0.9 to 1.8 s, right). The marker illustrates POz.

| Frequency range [Hz]	Mean Accuracy [%]	Peak mean t-value Electrode	Time [s]	Frequency [Hz]	Peak mean t-value Max. [%]
Overall [1 40]	52.26	F4	0.8	30	62.38
Theta [3 7]	52.49	AF7	0.4	7	58.80
Alpha [8 13]	52.40	Fpz	0.1	11	58.67
Beta [13 21]	52.29	Fpz	0.4	21	58.63
Beta [22 32]	52.22	P1	1.9	24	58.29
Beta [13 32]	52.24	P1	1.9	24	61.00
Gamma [32 40]	52.23	P2	0.1	33	52.98

Table 3. Average Accuracy for each frequency range and at the electrode with the maximummean t-value.

has been shown to be of central importance for memory encoding and retrieval processes[51,88–91]. The role of contextual reinstatement of multisensory features that stem, i.e. from the auditory and visual domain, remains so far unexplored. Here we aimed to shed light on the oscillatory mechanisms underlying the recognition of sequentially encoded multisensory episodes. Our findings show that modality sequences are (incidentally) encoded within the memory trace and serve as a context feature that drives recognition based on theta, alpha and beta frequency pattern reinstatement.

It is well established that context reinstatement plays a crucial role in memory retrieval, allowing the brain to access the temporal and environmental cues associated with past events[51,52]. However, previous research has largely focused on unimodal stimuli, rendering our understanding about how complex multisensory sequences are encoded and retrieved incomplete. In the current study, participants acquired and retrieved the image-sound/sound-image associations overall very well, with recognition performance as measured by d' being comparable to similar study designs[72,92–94]. Importantly, recognition performance did not differ when comparing AV (auditory-visual) vs. VA (visual-auditory) pairs, suggesting that the order of features itself did not influence memory formation. While we did not test for incidental stimulus acquisition, several studies report reinstatement of encoding specific features in memory tasks[29,95,96].

MVPA of EEG data revealed distinct neural signatures depending on the modality sequence presented during encoding, even though overall memory performance for both conditions was the same. This suggests that the brain encodes the order of multisensory episodes as part of the contextual memory trace, which aids in the retrieval process. This finding is consistent with prior work demonstrating context-specific temporal patterns during both encoding and retrieval processes[97]. Furthermore, it aligns with the Context Maintenance and Retrieval (CMR) model, which states that temporal and contextual features of episodes are essential components of the memory trace[52]. Our study extends this model by showing that modality order, as a contextual feature, can be decoded from oscillatory activity during memory retrieval.

Our multivariate results from EEG recordings suggest that decoding performance went significantly above chance level across multiple time points, frequencies, and electrode locations, with particularly strong effects in the theta, alpha, and beta bands. Theta oscillations have been associated with the binding of information into coherent memory traces and are crucial for organizing sequentially ordered working memory items[60,68,98]. These oscillations are thought to coordinate neural activity across different brain regions, facilitating the binding of sensory inputs into a coherent memory trace[97–99], acting as the "glue"[100]. Alpha oscillations have been generally related to the inhibition of irrelevant information and are involved in processing incoming information relevant to memory[101,102]. Decreases in alpha power during memory tasks have been associated with enhanced memory performance, particularly in semantic encoding tasks[103,104], while beta oscillations have been linked to memory formation, with elevated pre-stimulus beta power associated with successful memory encoding[105,106]. This activity is thought to reflect a memory-promoting state, possibly moderated by attentional or inhibitory processes[105]. In sum, theta, alpha, and beta oscillations play distinct yet interconnected roles in
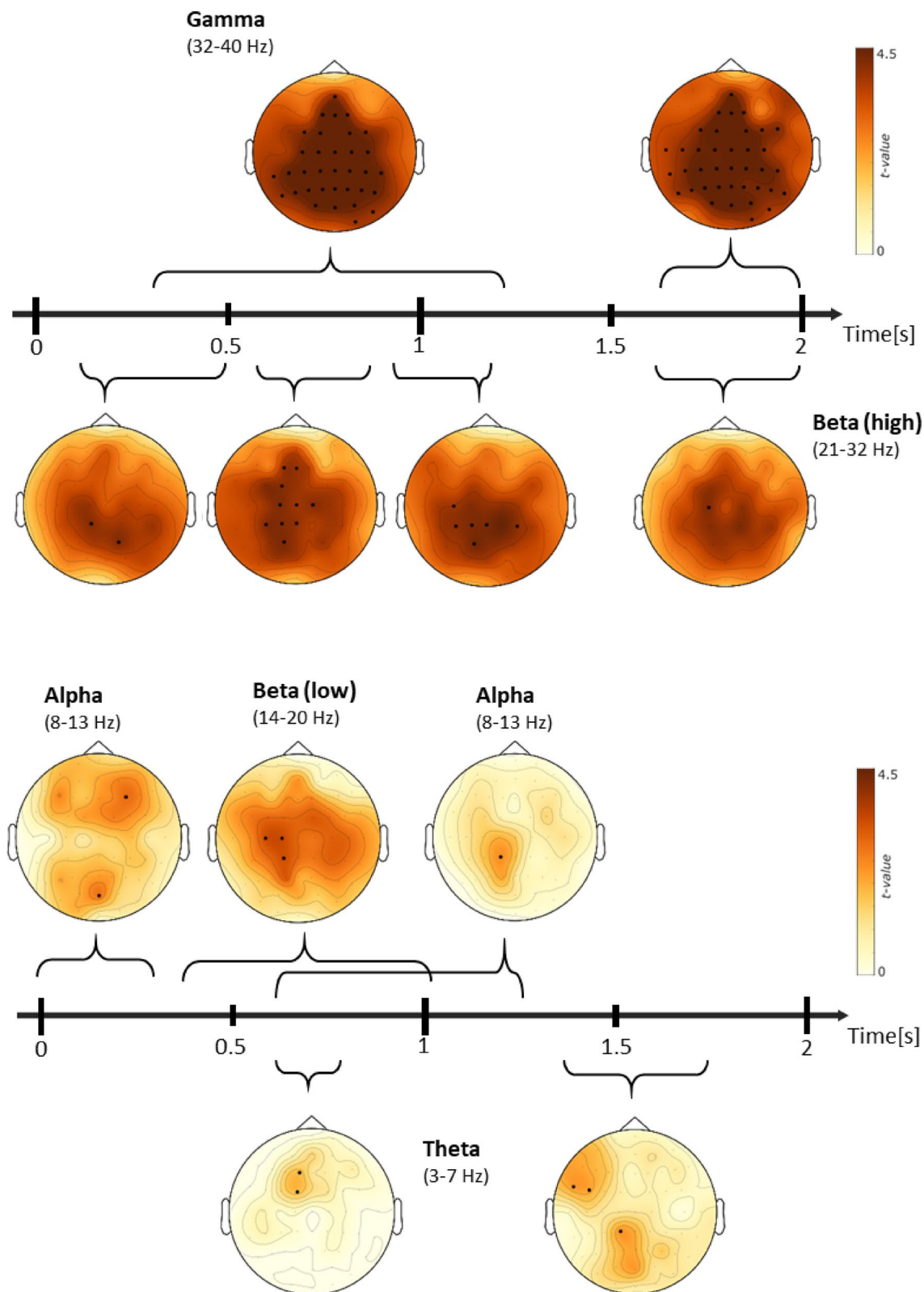
memory processing[107,108]. Taken together, the fact that the MVPA analysis revealed significant classification of modality order across different oscillatory bands provides strong evidence that sequential information in the form of a context feature is retained and reinstated during retrieval rather than merely reflecting general associative activation. Thus, our results suggest that encoding processes are sequence-specific, with VA pairs potentially engaging greater anticipatory processing due to the nature of auditory stimulus processing. One may speculate, that the reinstatement of the observed oscillatory patterns may facilitate the synchronization of neural activity across sensory processing regions, ensuring that the original modality sequence is represented during recognition within the specific memory trace, which has been suggested by several human studies[4,11,82,109,110].

◄ **Fig. 4**. Topographical plots of *t*-values from the cluster-based permutation test comparing MVPA accuracy for remembered trials against the chance level of 50%. The plots illustrate the distribution of significant effects across the scalp for different frequency bands (theta, alpha, low beta, high beta, and gamma) over the trial time course. Highlighted electrodes indicate regions with the highest number of neighbouring significant data points), reflecting areas with the strongest decoding effects. The gamma band showed significant effects between 400–1300 ms and 1700–2000 ms, with strong activations in centro-posterior regions. The high beta band displayed significant clusters in multiple time windows (100–500 ms, 600–800 ms, 900–1200 ms, and 1700–2000 ms), predominantly in centro-parietal areas. The low beta band exhibited significant effects between 400–1000 ms, mainly over central electrodes. The alpha band showed spatially distributed effects in early (0–300 ms) and later (600–1300 ms) time windows, particularly in partial-occipital and fronto-central sites. The theta band revealed significant clusters in two discrete time windows (600–700 ms and 1400–1700 ms), with the strongest effects observed in frontal and central regions. The results suggest that significant decoding effects are not uniformly distributed across the scalp but are concentrated in specific electrode regions, particularly in central and parietal areas in the beta and gamma range and frontal areas in the theta and alpha range.

Considering the potential cognitive processes elicited by the experimental design, one could argue that the initial stimulus might evoke visual or auditory mental imagery, creating an expectation of the second stimulus. In fact, there is evidence showing that mental imagery could also influence associative memory retrieval by engaging both modality-specific and modality-independent neural networks[111,112], thereby aiding in overcoming potential modality mismatches during encoding induced by incongruent association pairs[113]. This process involves the activation of sensory-specific regions, such as the visual and auditory cortices, alongside a modality-independent core network, including the default mode network, which supports imagery across different sensory domains[111,112,114]. The overlap between brain regions involved in mental imagery and those supporting retrieval suggests that successful retrieval relies on the same neural mechanisms that facilitate imagery[111,115]. Furthermore, encoding specificity plays a critical role in remembering, as the reactivation of encoding-related neural patterns benefits retrieval when there is a match between encoding and retrieval modalities but can impair memory under mismatch conditions[116]. However, individuals can flexibly employ mental imagery to compensate for mismatches, generating and maintaining mental representations even when encoding involves incongruent audiovisual information[117]. Moreover, imagery-based strategies, such as integrating items into interactive mental images, have been shown to enhance associative memory, emphasizing the functional significance of mental imagery in retrieval processes[118].

While these findings demonstrate that mental imagery is connected to memory encoding and retrieval, prior research also suggests that multisensory, sequential encoding can enhance memory through encoding variability, introducing competition effects that alter retrieval dynamics[28,29]. This interpretation gains support from our univariate oscillatory findings, suggesting a modality-specific influence of alpha and beta oscillations during the encoding of sequentially presented audiovisual stimuli. Specifically, for auditory-visual (AV) pairs, increased pre-stimulus alpha and beta power before the visual stimulus and a subsequent decrease during stimulus presentation may indicate a preparatory state followed by active sensory processing. In contrast, the pattern is reversed for visual-auditory (VA) pairs: pre-stimulus decreases in alpha and beta before the auditory stimulus, and increased power during its presentation suggests a shift in processing demands between modalities. Previous research has linked decreased alpha-band activity in the prefrontal and occipital cortex to successful visual encoding, indicating that lower alpha power facilitates visual information processing and enhances memory formation[119–122]. Similarly, increased pre-stimulus beta power has been associated with improved memory formation, potentially reflecting attentional or inhibitory processes that aid in binding stimulus components[105,123]. The observed pre-stimulus increases in beta power in AV pairs may, therefore, indicate an anticipatory mechanism supporting visual encoding, whereas the decrease in VA pairs might reflect a shift in sensory dominance from vision to audition. Alpha and beta oscillations have further been implicated in the processing of expectations and prediction errors. An alpha-to-beta desynchronization (ERD) has been linked to expected stimulus valence, suggesting that these frequency bands contribute to prediction mechanisms that influence encoding efficiency[124]. In audiovisual tasks, alpha oscillations modulate sensory processing and attention, influencing the temporal integration of stimuli[125]. However, alpha activity does not solely predict auditory stimulus detection consistently due to its interaction with broadband neural activity[126]. Beta oscillations, on the other hand, are associated with top-down control processes and enhance memory formation across sensory modalities, including auditory processing[105]. Taken together, these findings suggest that alpha and beta oscillations in sequential encoding are modality-dependent rather than purely memory-driven. The observed pre-stimulus shifts in power may reflect preparatory mechanisms that optimize encoding by modulating attention and sensory processing demands across modalities. Specifically, lower alpha power in occipital-parietal regions has been associated with improved perceptual sensitivity and the enhancement of stimulus representations[127], supporting the idea that modality-dependent oscillatory changes may reflect the differential engagement of sensory and integrative processes during retrieval. Finally, theta oscillations play a critical role in cross-modal binding, supporting associative memory by synchronizing neural activity across sensory regions[128]. Given the angular gyrus' role in multimodal integration[129] and the contribution of multisensory cues to episodic retrieval[130–132], it is likely that power changes across frequency bands reflect both modality-specific processing and memory-related mechanisms in an interactive manner. Thus, the observed oscillatory dynamics align with well-established mechanisms of sensory reactivation, cortical excitability, and associative memory retrieval.
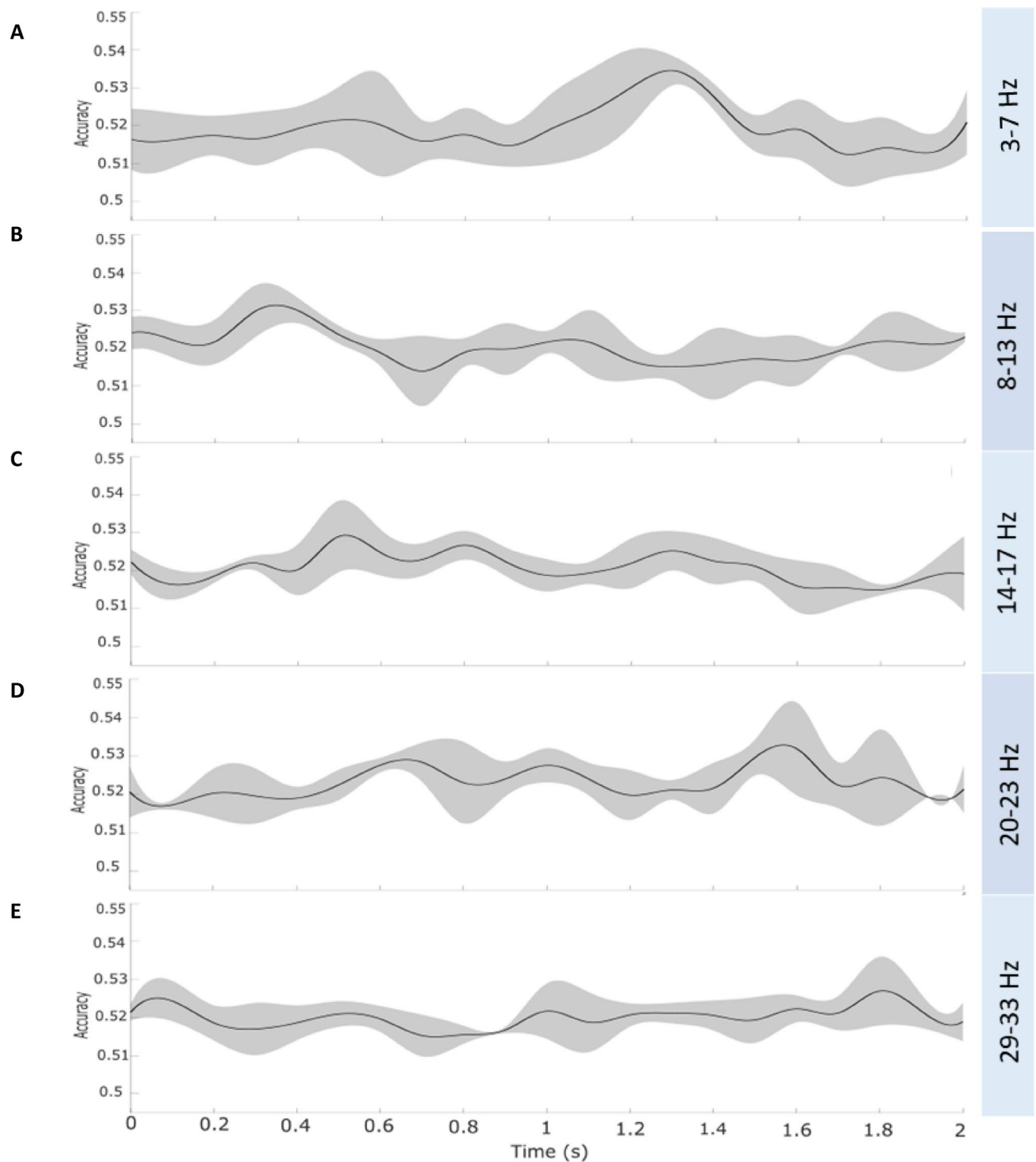
Fig. 5. Full-time courses of accuracy values from exemplary electrodes for each frequency range. Time-resolved decoding accuracy (MVPA) for remembered trials to classify modality order from 0 to 2 s relative to stimulus onset is shown at exemplary electrodes. Accuracy values, averaged within the respective frequency bands, range from 51–54%. (**A**) Theta (3–7 Hz) at Cz exhibits a subtle peak around 1.3 s. (**B**) Alpha (8–13 Hz) at F4 shows a slight peak around 0.3 s. (**C**) Decoding accuracy in the low beta (14–17 Hz) at Fz, (**D**) high beta (20–23 Hz) at C3, and (**E**) gamma (29–33 Hz) at F4 bands fluctuates over time but did not reveal distinct classification peaks. Caution should be exercised, as the averaging of accuracy estimates over multiple time points and across participants may have contributed to an overall reduction in classification performance.
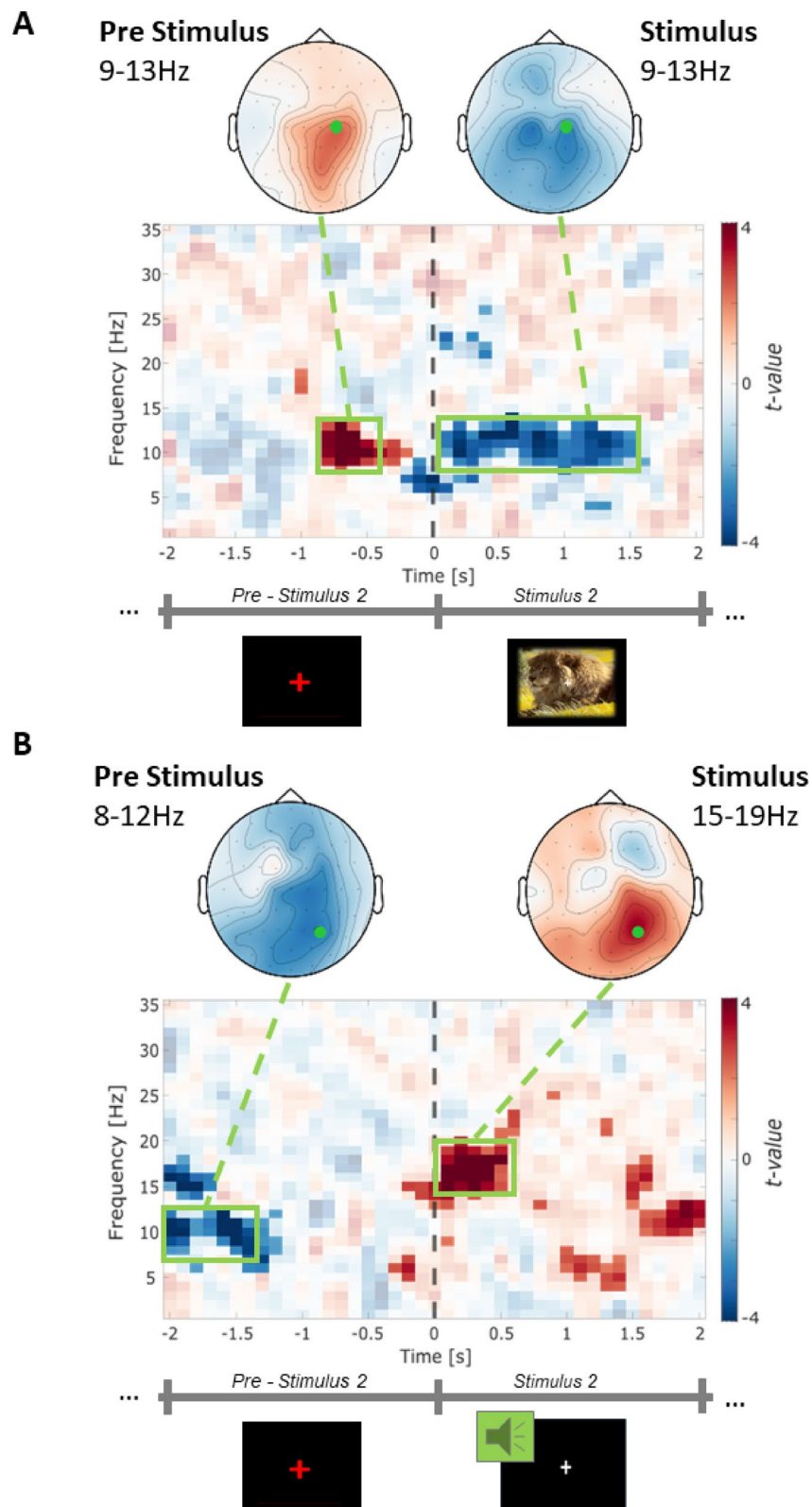
The role of modality order in retrieval is further supported by the finding that above-chance decoding of modality order from EEG activity during retrieval indicates its integration into the memory trace. If one modality had dominated encoding, successful decoding of modality order would not be expected. Prior research supports the idea that encoding modality order contributes to retrieval by providing structured cues that facilitate reconstruction of past experiences[133]. Contextual information is a well-established component of episodic memory, with hippocampal mechanisms playing a crucial role in binding sensory details into coherent memory representations[22,85,103,134,135]. Our study extends this body of work by demonstrating that modality order, as a contextual feature, can be decoded from oscillatory activity during retrieval. While our data do not conclusively establish a causal link between modality order encoding and retrieval success, they provide novel

Fig. 6. Correlation between the power difference of remembered vs forgotten trials in Stimulus interval 2 and the memory performance (d') from the recognition task. (**A**) The correlation between the power differences and the memory performance as measured by d' during visual stimuli presentation (AV pair) at C2. Topographical distribution within the significant cluster in the pre stimulus (-0.8 to -0.1 s; left) and the significant cluster in the stimulus presentation (0 to 1.5 s) in the alpha range (9 to 13 Hz, right) during visual stimuli presentation. The marker illustrates C2. (**B**) The correlation between the power differences and the memory performance as measured by d' during auditory stimulus presentation (VA pair) at P4. Topographical distributions are shown for the pre-stimulus cluster (-2 to -1.3 s; 8–12 Hz) and stimulus presentation cluster (0–0.5 s; 15–19 Hz). The green marker illustrates P4.

insights into the neural dynamics supporting multisensory sequential memory. Although no behavioural data confirm explicit retrieval of order information, the significant decoding results indicate that modality order was included in the memory trace and reinstated during retrieval. Future studies should explore the extent to which such reinstatement contributes to explicit order memory and whether implicit representations influence retrieval performance.

Finally, we observed modality effects in parietal locations during both, encoding, and recognition tasks. This might reflect multisensory association processes[136,137], as the parietal cortex is crucial for integrating information from various sensory modalities[138–141]. Interestingly, our findings align well with recent work around multisensory processing, which shows that information from different sensory modalities is integrated within several cortical regions (e.g. the parietal lobe;[139,141,142]). While the intraparietal sulcus (IPS) is known to process multisensory information[143–145], the angular gyrus has been shown to be centrally involved in binding information into coherent narratives[131]. In our results, classifier accuracy derived from MVPA was highest at centro-parietal electrodes (e.g., CP1), covering the superior parietal lobe. This might suggest a distinct role of the superior parietal cortex in multisensory sequential reinstatement processes, which supports the hypothesis that the neural systems for sequence encoding and multisensory integration are closely linked to facilitate the binding of presented items, thereby forming an episode.

In conclusion, our study provides new insights into the neural mechanisms underlying multisensory memory retrieval. The findings of this study yield important implications for our understanding of memory processes. First, the (incidental) encoding of modality sequence order as a context feature suggests that the brain actively integrates temporal and sensory information during memory formation. This has important implications for models of episodic memory, particularly those that emphasize the role of context[52]. Our findings suggest that the neural mechanisms underlying context-feature retrieval are not limited to unimodal tasks[4,51,64,146]. Instead, they extend to more complex multisensory episodes, during which we encode and integrate sequences of different sensory modalities. Crucially, the sequence of modalities as a contextual feature within the memory trace directly affects memory retrieval, cognitive control, and learning processes[147–150]. Prior research suggests that different sensory modalities contribute uniquely to encoding and retrieval mechanisms, with auditory and visual sequences influencing attentional engagement and memory consolidation in distinct ways (e.g.,[151,152]. Auditory sequences, for instance, have been linked to more durable temporal structuring, while visual sequences often benefit from spatial organization[71]. Recognizing modality sequence as a contextual feature allows us to investigate how the structure of sensory input shapes memory representations rather than focusing solely on content-based associations. Understanding these effects can help optimize learning and memory strategies by leveraging the strengths of different modalities. This might offer insights into educational and rehabilitative applications where multimodal integration plays a key role. Our findings provide new insights into how the brain encodes and retrieves complex episodic memories, particularly those that involve multisensory information.

Data availability
The raw EEG and behavioral data underlying our findings have been uploaded to an open repository (https://www.fdr.uni-hamburg.de/; https://www.fdr.uni-hamburg.de/record/17120) for accessibility.

References
1. Dannenberg, H., Alexander, A. S., Robinson, J. C. & Hasselmo, M. E. The role of hierarchical dynamical functions in coding for episodic memory and cognition. *J. Cogn. Neurosci.* **31**, 1271–1289 (2019).
2. Sheldon, S., Farb, N., Palombo, D. J. & Levine, B. Intrinsic medial Temporal lobe connectivity relates to individual differences in episodic autobiographical remembering. *Cortex* **74**, 206–216 (2016).
3. Sadeh, S. & Clopath, C. Theory of neuronal perturbome in cortical networks. *Proc. Natl. Acad. Sci.* **117**, 26966–26976 (2020).
4. Kim, J. S. & Lee, S. A. Hippocampal orchestration of associative and sequential memory networks for episodic retrieval. *Cell. Rep.* **42**, 112989 (2023).
5. Yonelinas, A. P., Ranganath, C., Ekstrom, A. D. & Wiltgen, B. J. A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nat. Rev. Neurosci.* **20**, 364–375 (2019).
6. Pronier, É., Morici, J. F. & Girardeau, G. The role of the hippocampus in the consolidation of emotional memories during sleep. *Trends Neurosci.* **46**, 912–925 (2023).
7. Cooper, R. A. & Ritchey, M. Progression from Feature-Specific brain activity to hippocampal binding during episodic encoding. *J. Neurosci.* **40**, 1701–1709 (2020).
8. Crystal, J. D. Temporal foundations of episodic memory. *Learn. Behav.* **52**, 35–50 (2024).
9. Clewett, D. & Davachi, L. The ebb and flow of experience determines the Temporal structure of memory. *Curr. Opin. Behav. Sci.* **17**, 186–193 (2017).

10. Horner, A. J., Bisby, J. A., Wang, A., Bogus, K. & Burgess, N. The role of Spatial boundaries in shaping long-term event representations. *Cognition* **154**, 151–164 (2016).
11. Eichenbaum, H. Time (and space) in the hippocampus. *Curr. Opin. Behav. Sci.* **17**, 65–70 (2017).
12. Leroy, N., Majerus, S. & D'Argembeau, A. Working memory capacity for continuous events: the root of Temporal compression in episodic memory? *Cognition* **247**, 105789 (2024).
13. Jeunehomme, O., Folville, A., Stawarczyk, D., Van Der Linden, M. & D'Argembeau, A. Temporal compression in episodic memory for real-life events. *Memory* **26**, 759–770 (2018).
14. Jeunehomme, O. & D'Argembeau, A. Event segmentation and the Temporal compression of experience in episodic memory. *Psychol. Res.* **84**, 481–490 (2020).
15. D'Argembeau, A., Jeunehomme, O. & Stawarczyk, D. Slices of the past: how events are temporally compressed in episodic memory. *Memory* **30**, 43–48 (2022).
16. Riegel, M., Granja, D., Amer, T., Vuilleumier, P. & Rimmele, U. Opposite effects of emotion and event segmentation on Temporal order memory and object-context binding. *Cogn. Emot.* **39**, 117–135 (2025).
17. Thavabalasingam, S., O'Neil, E. B. & Lee, A. C. H. Multivoxel pattern similarity suggests the integration of Temporal duration in hippocampal event sequence representations. *NeuroImage* **178**, 136–146 (2018).
18. Liu, C., Ye, Z., Chen, C., Axmacher, N. & Xue, G. Hippocampal representations of event structure and Temporal context during episodic Temporal order memory. *Cereb. Cortex.* **32**, 1520–1534 (2022).
19. Rey, L. et al. Episodic memory and recognition are influenced by cues' sensory.
20. Tamminen, J. & Mebude, M. Reinstatement of odour context cues veridical memories but not false memories. *Memory* **27**, 575–579 (2019).
21. Sorokowska, A., Nord, M., Stefańczyk, M. M. & Larsson, M. Odor-based context-dependent memory: influence of olfactory cues on declarative and nondeclarative memory indices. *Learn. Mem.* **29**, 136–141 (2022).
22. Marks, W. D., Yokose, J., Kitamura, T. & Ogawa, S. K. Neuronal ensembles organize activity to generate contextual memory. *Front. Behav. Neurosci.* **16**, 805132 (2022).
23. Noyce, A. L., Cestero, N., Shinn-Cunningham, B. G. & Somers, D. C. Short-term memory stores organized by information domain. *Atten. Percept. Psychophys.* **78**, 960–970 (2016).
24. Dimsdale-Zucker, H. R. et al. Representations of complex contexts: A role for Hippocampus. *J. Cogn. Neurosci.* **35**, 90–110 (2022).
25. Bird, C. M. The role of the hippocampus in recognition memory. *Cortex* **93**, 155–165 (2017).
26. Robin, J & Moscovitch, M. Familiar real-world Spatial cues provide memory benefits in older and younger adults. *Psychol. Aging.* **32**, 210–219 (2017).
27. Wälti, M. J., Woolley, D. G. & Wenderoth, N. Assessing rhythmic visual entrainment and reinstatement of brain oscillations to modulate memory performance. *Front. Behav. Neurosci.* **14**, 118 (2020).
28. Zhang, M. & Hupbach, A. The effects of variable encoding contexts on item and source recognition. *Mem. Cognit.* **51**, 391–403 (2023).
29. Bramão, I., Jiang, J., Wagner, A. D. & Johansson, M. Encoding contexts are incidentally reinstated during competitive retrieval and track the Temporal dynamics of memory interference. *Cereb. Cortex.* **32**, 5020–5035 (2022).
30. Bramão, I. & Johansson, M. Neural Pattern Classification Tracks Transfer-Appropriate Processing in Episodic Memory. *eneuro* 5, ENEURO.0251-18.2018 (2018).
31. Wheeler, R. L. & Gabbert, F. Using Self-Generated cues to facilitate recall: A narrative review. *Front. Psychol.* **8**, 1830 (2017).
32. Unsworth, N. & Spillers, G. J. Variation in working memory capacity and episodic recall: the contributions of strategic encoding and contextual retrieval. *Psychon Bull. Rev.* **17**, 200–205 (2010).
33. Versace, R. et al. Act-In: an integrated view of memory mechanisms. *J. Cogn. Psychol.* **26**, 280–306 (2014).
34. Duarte, S. E., Ghetti, S. & Geng, J. J. Object memory is multisensory: Task-irrelevant sounds improve recollection. *Psychon Bull. Rev.* **30**, 652–665 (2023).
35. Versace, R., Labeye, É., Badard, G. & Rose, M. The contents of long-term memory and the emergence of knowledge. *Eur. J. Cogn. Psychol.* **21**, 522–560 (2009).
36. Juan, C. et al. The variability of multisensory processes of natural stimuli in human and non-human primates in a detection task. *PLOS ONE.* **12**, e0172480 (2017).
37. Lehmann, S. & Murray, M. M. The role of multisensory memories in unisensory object discrimination. *Cogn. Brain Res.* **24**, 326–334 (2005).
38. Moran, J. K. et al. Multisensory processing can compensate for Top-Down attention deficits in schizophrenia. *Cereb. Cortex.* **31**, 5536–5548 (2021).
39. Thelen, A., Talsma, D. & Murray, M. M. Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition* **138**, 148–160 (2015).
40. Murray, C. A. & Shams, L. Crossmodal interactions in human learning and memory. *Front. Hum. Neurosci.* **17**, 1181760 (2023).
41. Tovar, D. A., Murray, M. M. & Wallace, M. T. Selective enhancement of object representations through multisensory integration. *J. Neurosci.* **40**, 5604–5615 (2020).
42. Benini, E., Koch, I., Mayr, S., Frings, C. & Philipp, A. M. Contextual features of the cue enter episodic bindings in task switching. *J. Cogn.* **5**, 29 (2022).
43. Karacsony, S. & Abela, M. R. L. Stimulating sense memories for people living with dementia using the Namaste care programme: what works, how and why? *J. Clin. Nurs.* **31**, 1921–1932 (2022).
44. Shams, L. & Seitz, A. R. Benefits of multisensory learning. *Trends Cogn. Sci.* **12**, 411–417 (2008).
45. Dudai, Y. Molecular bases of long-term memories: a question of persistence. *Curr. Opin. Neurobiol.* **12**, 211–216 (2002).
46. Sridhar, S., Khamaj, A. & Asthana, M. K. Cognitive neuroscience perspective on memory: overview and summary. *Front. Hum. Neurosci.* **17**, 1217093 (2023).
47. Brainerd, C. J., Gomes, C. F. A. & Nakamura, K. Dual recollection in episodic memory. *J. Exp. Psychol. Gen.* **144**, 816–843 (2015).
48. Chang, M. et al. Spatial context scaffolds long-term episodic richness of weaker real-world autobiographical memories in both older and younger adults. *Memory* **32**, 431–448 (2024).
49. Dobbins, I. G. & Wagner, A. D. Domain-general and Domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cereb. Cortex.* **15**, 1768–1778 (2005).
50. King, D. R., De Chastelaine, M., Elward, R. L., Wang, T. H. & Rugg, M. D. Recollection-Related increases in functional connectivity predict individual differences in memory accuracy. *J. Neurosci.* **35**, 1763–1772 (2015).
51. Lohnas, L. J., Healey, M. K. & Davachi, L. Neural Temporal context reinstatement of event structure during memory recall. *J. Exp. Psychol. Gen.* **152**, 1840–1872 (2023).
52. Polyn, S. M., Norman, K. A. & Kahana, M. J. A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* **116**, 129–156 (2009).
53. Clewett, D., DuBrow, S. & Davachi, L. Transcending time in the brain: how event memories are constructed from experience. *Hippocampus* **29**, 162–183 (2019).
54. Davachi, L. & DuBrow, S. How the hippocampus preserves order: the role of prediction and context. *Trends Cogn. Sci.* **19**, 92–99 (2015).
55. DuBrow, S. & Davachi, L. Temporal binding within and across events. *Neurobiol. Learn. Mem.* **134**, 107–114 (2016).

56. Gordon, A. M., Rissman, J., Kiani, R. & Wagner, A. D. Cortical reinstatement mediates the relationship between Content-Specific encoding activity and subsequent recollection decisions. *Cereb. Cortex*. **24**, 3350–3364 (2014).
57. Jafarpour, A., Horner, A. J., Fuentemilla, L., Penny, W. D. & Duzel, E. Decoding oscillatory representations and mechanisms in memory. *Neuropsychologia* **51**, 772–780 (2013).
58. Staresina, B. P. & Davachi, L. Mind the Gap: binding experiences across space and time in the human Hippocampus. *Neuron* **63**, 267–276 (2009).
59. Folkerts, S., Rutishauser, U. & Howard, M. W. Human episodic memory retrieval is accompanied by a neural contiguity effect. *J. Neurosci.* **38**, 4200–4211 (2018).
60. Herweg, N. A., Solomon, E. A. & Kahana, M. J. Theta oscillations in human memory. *Trends Cogn. Sci.* **24**, 208–227 (2020).
61. Koen, J. D. & Rugg, M. D. Memory reactivation predicts resistance to retroactive interference: evidence from multivariate classification and pattern similarity analyses. *J. Neurosci.* **36**, 4389–4399 (2016).
62. Leiker, E. K. & Johnson, J. D. Pattern reactivation co-varies with activity in the core recollection network during source memory. *Neuropsychologia* **75**, 88–98 (2015).
63. Miller, J. E., Carlson, L. A. & McAuley, J. D. When what you hear influences when you see: listening to an auditory rhythm influences the Temporal allocation of visual attention. *Psychol. Sci.* **24**, 11–18 (2013).
64. Staudigl, T., Vollmar, C., Noachtar, S. & Hanslmayr, S. Temporal-Pattern similarity analysis reveals the beneficial and detrimental effects of context reinstatement on human memory. *J. Neurosci.* **35**, 5373–5384 (2015).
65. Yu, J. Y., Liu, D. F., Loback, A., Grossrubatscher, I. & Frank, L. M. Specific hippocampal representations are linked to generalized cortical representations in memory. *Nat. Commun.* **9**, 2209 (2018).
66. Ratcliff, R., Sederberg, P. B., Smith, T. A. & Childers, R. A single trial analysis of EEG in recognition memory: tracking the neural correlates of memory strength. *Neuropsychologia* **93**, 128–141 (2016).
67. Rugg, M. D. & Wilding, E. L. Retrieval processing and episodic memory. *Trends Cogn. Sci.* **4**, 108–115 (2000).
68. Staudigl, T., Hanslmayr, S. & Bäuml, K. H. T. Theta oscillations reflect the dynamics of interference in episodic memory retrieval. *J. Neurosci.* **30**, 11356–11362 (2010).
69. Tan, R. J., Rugg, M. D. & Lega, B. C. Direct brain recordings identify hippocampal and cortical networks that distinguish successful versus failed episodic memory retrieval. *Neuropsychologia* **147**, 107595 (2020).
70. Waldhauser, G. T., Braun, V. & Hanslmayr, S. Episodic memory retrieval functionally relies on very rapid reactivation of sensory information. *J. Neurosci.* **36**, 251–260 (2016).
71. Cowan, N. & Guitard, D. Encoding colors and tones into working memory concurrently: A developmental investigation. *Dev. Sci.* **27**, e13552 (2024).
72. Ostrowski, J. & Rose, M. Increases in pre-stimulus theta and alpha oscillations precede successful encoding of crossmodal associations. *Sci. Rep.* **14**, 7895 (2024).
73. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **49**, 764–766 (2013).
74. Klem, G. H., Lüders, H. O. & Elger, C. *The ten-twenty Electrode System of the International Federation* (The International Federation of Clinical Neurophysiology, 1999).
75. Oostenveld, R., Fries, P., Maris, E., Schoffelen, J. M. & FieldTrip Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* 1–9 (2011). (2011).
76. Treder, M. S. & MVPA-Light A classification and regression toolbox for Multi-Dimensional data. *Front. Neurosci.* **14**, 289 (2020).
77. Artuso, C., Palladino, P. & Belacchi, C. Sensitivity detection in memory recognition: interference control as index of taxonomic memory development? *Memory* **28**, 187–195 (2020).
78. Haatveit, B. C. et al. The validity of *d* prime as a working memory index: results from the Bergen *n*-back task. *J. Clin. Exp. Neuropsychol.* **32**, 871–880 (2010).
79. Shermohammed, M., Davidow, J. Y., Somerville, L. H. & Murty, V. P. Stress impacts the fidelity but not strength of emotional memories. *Brain Cogn.* **133**, 33–41 (2019).
80. Marinari, E. Optimized Monte Carlo methods. in Advances in Computer Simulation (eds Kertész, J. & Kondor, I.) vol. 501 50–81 (Springer Berlin Heidelberg, (1998).
81. Kalafatovich, J., Lee, M. & Lee, S. W. Decoding declarative memory process for predicting memory retrieval based on source localization. *PLOS ONE.* **17**, e0274101 (2022).
82. Osipova, D. et al. Theta and gamma oscillations predict encoding and retrieval of declarative memory. *J. Neurosci.* **26**, 7523–7531 (2006).
83. Sikora-Wachowicz, B. et al. False recognition in Short-Term Memory – Age-Differences in confidence. *Front. Psychol.* **10**, 2785 (2019).
84. Weidemann, C. T. & Kahana, M. J. Assessing recognition memory using confidence ratings and response times. *R Soc. Open. Sci.* **3**, 150670 (2016).
85. Staudigl, T. & Hanslmayr, S. Theta oscillations at encoding mediate the Context-Dependent nature of human episodic memory. *Curr. Biol.* **23**, 1101–1106 (2013).
86. Sols, I., DuBrow, S., Davachi, L. & Fuentemilla, L. Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in Long-Term memory. *Curr. Biol.* **27**, 3499–3504e4 (2017).
87. Schwabe, L. & Wolf, O. T. The context counts: congruent learning and testing environments prevent memory retrieval impairment following stress. *Cogn. Affect. Behav. Neurosci.* **9**, 229–236 (2009).
88. Bramão, I., Karlsson, A. & Johansson, M. Mental reinstatement of encoding context improves episodic remembering. *Cortex* **94**, 15–26 (2017).
89. Lu, Y., Wang, C., Chen, C. & Xue, G. Spatiotemporal neural pattern similarity supports episodic memory. *Curr. Biol.* **25**, 780–785 (2015).
90. Sutterer, D. W., Foster, J. J., Serences, J. T., Vogel, E. K. & Awh, E. Alpha-band oscillations track the retrieval of precise Spatial representations from long-term memory. *J. Neurophysiol.* **122**, 539–551 (2019).
91. Xiao, X. et al. Transformed neural pattern reinstatement during episodic memory retrieval. *J. Neurosci.* **37**, 2986–2998 (2017).
92. Jablonowski, J. & Rose, M. The functional dissociation of posterior parietal regions during multimodal memory formation. *Hum. Brain Mapp.* **43**, 3469–3485 (2022).
93. Marian, V., Hayakawa, S. & Schroeder, S. R. Cross-Modal interaction between auditory and visual input impacts memory retrieval. *Front. Neurosci.* **15**, 661477 (2021).
94. Pillai, A. S., Gilbert, J. R. & Horwitz, B. Early sensory cortex is activated in the absence of explicit input during crossmodal item retrieval: evidence from MEG. *Behav. Brain Res.* **238**, 265–272 (2013).
95. Healey, M. K. Temporal contiguity in incidentally encoded memories. *J. Mem. Lang.* **102**, 28–40 (2018).
96. Mundorf, A. M. D., Lazarus, L. T. T., Uitvlugt, M. G. & Healey, M. K. A test of retrieved context theory: dynamics of recall after incidental encoding. *J. Exp. Psychol. Learn. Mem. Cogn.* **47**, 1264–1287 (2021).
97. Michelmann, S., Bowman, H. & Hanslmayr, S. The Temporal signature of memories: identification of a general mechanism for dynamic memory replay in humans. *PLOS Biol.* **14**, e1002528 (2016).
98. Rudoler, J. H., Herweg, N. A. & Kahana, M. J. Hippocampal Theta and episodic memory. *J. Neurosci.* **43**, 613–620 (2023).
99. Karakaş, S. A review of theta Oscillation and its functional correlates. *Int. J. Psychophysiol.* **157**, 82–99 (2020).

100. Clouter, A., Shapiro, K. L. & Hanslmayr, S. Theta phase synchronization is the glue that binds human associative memory. *Curr. Biol.* **27**, 3143–3148e6 (2017).
101. Schroeder, S. C. Y., Ball, F. & Busch, N. A. The role of alpha oscillations in distractor Inhibition during memory retention. *Eur. J. Neurosci.* **48**, 2516–2526 (2018).
102. Hall, M. C. et al. Oscillatory activity in bilateral prefrontal cortices is altered by distractor strength during working memory processing. *NeuroImage* **301**, 120878 (2024).
103. Griffiths, B. J., Martín-Buro, M. C., Staresina, B. P. & Hanslmayr, S. Disentangling neocortical alpha/beta and hippocampal theta/gamma oscillations in human episodic memory formation. *NeuroImage* **242**, 118454 (2021).
104. Vogelsang, D. A., Gruber, M., Bergström, Z. M., Ranganath, C. & Simons, J. S. Alpha oscillations during incidental encoding predict subsequent memory for new foil information. *J. Cogn. Neurosci.* **30**, 667–679 (2018).
105. Scholz, S., Schneider, S. L. & Rose, M. Differential effects of ongoing EEG beta and theta power on memory formation. *PLOS ONE.* **12**, e0171913 (2017).
106. Winterling, S. L., Shields, S. M. & Rose, M. Reduced memory-related ongoing oscillatory activity in healthy older adults. *Neurobiol. Aging.* **79**, 1–10 (2019).
107. Elmers, J., Yu, S., Talebi, N., Prochnow, A. & Beste, C. Neurophysiological effective network connectivity supports a threshold-dependent management of dynamic working memory gating. *iScience* **27**, 109521 (2024).
108. Siebenhühner, F., Wang, S. H., Palva, J. M. & Palva, S. Cross-frequency synchronization connects networks of fast and slow oscillations during visual working memory maintenance. *eLife* **5**, e13451 (2016).
109. Engel, A. K. & Singer, W. Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* **5**, 16–25 (2001).
110. Pennartz, C. M. A. Identification and integration of sensory modalities: neural basis and relation to consciousness. *Conscious. Cogn.* **18**, 718–739 (2009).
111. Huijbers, W., Pennartz, C. M. A., Rubin, D. C. & Daselaar, S. M. Imagery and retrieval of auditory and visual information: neural correlates of successful and unsuccessful performance. *Neuropsychologia* **49**, 1730–1740 (2011).
112. Caenegem et al. Multisensory approach in mental imagery: ALE meta-analyses comparing motor, visual and auditory imagery. *Neurosci Biobehav Rev* **167**.
113. Tian, X., Zarate, J. M. & Poeppel, D. Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex* **77**, 1–12 (2016).
114. Daselaar, S. M., Porat, Y., Huijbers, W. & Pennartz, C. M. A. Modality-specific and modality-independent components of the human imagery system. *NeuroImage* **52**, 677–685 (2010).
115. Spagna, A. et al. Visual mental imagery: evidence for a heterarchical neural architecture. *Phys. Life Rev.* **48**, 113–131 (2024).
116. Staudigl, T. & Hanslmayr, S. Reactivation of neural patterns during memory reinstatement supports encoding specificity. *Cogn. Neurosci.* **10**, 175–185 (2019).
117. Umar, H., Mast, F. W., Cacchione, T. & Martarelli, C. S. The prioritization of visuo-spatial associations during mental imagery. *Cogn. Process.* **22**, 227–237 (2021).
118. Sahadevan, S. S., Chen, Y. Y. & Jeremy, B. C. Imagery-based strategies for memory for associations. *Memory* **29**:10, 1275–1295 .
119. Yin, Q. et al. Direct brain recordings reveal occipital cortex involvement in memory development. *Neuropsychologia* **148**, 107625 (2020).
120. Xie, Y., Feng, Z., Xu, Y., Bian, C. & Li, M. The different Oscillation patterns of alpha band in the early and later stages of working memory maintenance. *Neurosci. Lett.* **633**, 220–226 (2016).
121. Staudigl, T., Hartl, E., Noachtar, S., Doeller, C. F. & Jensen, O. Saccades are phase-locked to alpha oscillations in the occipital and medial Temporal lobe during successful memory encoding. *PLOS Biol.* **15**, e2003404 (2017).
122. Romei, V. et al. Spontaneous fluctuations in posterior -Band EEG activity reflect variability in excitability of human visual areas. *Cereb. Cortex.* **18**, 2010–2018 (2008).
123. Schneider, S. L. & Rose, M. Intention to encode boosts memory-related pre-stimulus EEG beta power. *NeuroImage* **125**, 978–987 (2016).
124. Strube, A., Rose, M., Fazeli, S. & Büchel, C. Alpha-to-beta- and gamma-band activity reflect predictive coding in affective visual processing. *Sci. Rep.* **11**, 23492 (2021).
125. Ronconi, L., Busch, N. A. & Melcher, D. Alpha-band sensory entrainment alters the duration of Temporal windows in visual perception. *Sci. Rep.* **8**, 11810 (2018).
126. Cunningham, E., Zimnicki, C. & Beck, D. M. The influence of prestimulus 1/f-Like versus Alpha-Band activity on subjective awareness of auditory and visual stimuli. *J. Neurosci.* **43**, 6447–6459 (2023).
127. Zhou, Y. J., Iemi, L., Schoffelen, J. M., De Lange, F. P. & Haegens, S. Alpha oscillations shape sensory representation and perceptual sensitivity. *J. Neurosci.* **41**, 9581–9592 (2021).
128. Wang, D., Clouter, A., Chen, Q., Shapiro, K. L. & Hanslmayr, S. Single-Trial phase entrainment of Theta oscillations in sensory regions predicts human associative memory performance. *J. Neurosci.* **38**, 6299–6309 (2018).
129. Yazar, Y., Bergström, Z. M. & Simons, J. S. Reduced multimodal integration of memory features following continuous theta burst stimulation of angular gyrus. *Brain Stimulat.* **10**, 624–629 (2017).
130. Tibon, R., Fuhrmann, D., Levy, D. A., Simons, J. S. & Henson, R. N. Multimodal integration and vividness in the angular gyrus during episodic encoding and retrieval. *J. Neurosci.* **39**, 4365–4374 (2019).
131. Grob, A. M., Heinbockel, H., Milivojevic, B., Doeller, C. & Schwabe, L. Causal role of the angular gyrus in insight-driven memory reconfiguration. *Preprint At.* https://doi.org/10.7554/eLife.91033.1 (2023).
132. Bonnici, H. M., Richter, F. R., Yazar, Y. & Simons, J. S. Multimodal feature integration in the angular gyrus during episodic and semantic retrieval. *J. Neurosci.* **36**, 5462–5471 (2016).
133. Pierce, B. H. & Gallo, D. A. Encoding modality can affect memory accuracy via retrieval orientation. *J. Exp. Psychol. Learn. Mem. Cogn.* **37**, 516–521 (2011).
134. Thakral, P. P., Benoit, R. G. & Schacter, D. L. Characterizing the role of the hippocampus during episodic simulation and encoding. *Hippocampus* **27**, 1275–1284 (2017).
135. Foudil, S. A., Pleche, C. & Macaluso, E. Memory for spatio-temporal contextual details during the retrieval of naturalistic episodes. *Sci. Rep.* **11**, 14577 (2021).
136. Rohe, T. & Noppeney, U. Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Curr. Biol.* **26**, 509–514 (2016).
137. Sereno, M. I. & Huang, R. S. Multisensory maps in parietal cortex. *Curr. Opin. Neurobiol.* **24**, 39–46 (2014).
138. Andersen, R. A. Multimodal integration for the representation of space in the posterior parietal cortex. (1997).
139. Kayser, C. & Logothetis, N. K. Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* **212**, 121–132 (2007).
140. Lin, J. Sensory inputs guiding cognitive behaviors and decision making. *Highlights Sci. Eng. Technol.* **74**, 1399–1404 (2023).
141. Porada, D. K., Regenbogen, C., Freiherr, J., Seubert, J. & Lundström, J. N. Trimodal processing of complex stimuli in inferior parietal cortex is modality-independent. *Cortex* **139**, 198–210 (2021).
142. Lewis, J. W. & Van Essen, D. C. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137 (2000).
143. Anderson, K. L., Rajagovindan, R., Ghacibeh, G. A., Meador, K. J. & Ding, M. Theta oscillations mediate interaction between prefrontal cortex and medial Temporal lobe in human memory. *Cereb. Cortex.* **20**, 1604–1612 (2010).

144. Regenbogen, C. et al. The intraparietal sulcus governs multisensory integration of audiovisual information based on task difficulty. *Hum. Brain Mapp.* **39**, 1313–1326 (2018).
145. Sours, C. et al. Structural and functional integrity of the intraparietal sulcus in moderate and severe traumatic brain injury. *J. Neurotrauma.* **34**, 1473–1481 (2017).
146. DuBrow, S. & Davachi, L. The influence of context boundaries on memory for the sequential order of events. *J. Exp. Psychol. Gen.* **142**, 1277–1286 (2013).
147. Schmidt, M., Frings, C. & Tempel, T. Context-Dependent memory of motor sequences. *J. Cogn.* **4**, 15 (2021).
148. Kelber, P., Mackenzie, I. G. & Mittelstädt, V. Cognitive control in cross-modal contexts: abstract feature transitions of task-related but not task-unrelated stimuli modulate the congruency sequence effect. *J. Exp. Psychol. Learn. Mem. Cogn.* **50**, 902–919 (2024).
149. Pazdera, J. K. & Kahana, M. J. Modality effects in free recall: A retrieved-context account. *J. Exp. Psychol. Learn. Mem. Cogn.* **49**, 866–888 (2023).
150. Song, G. & Tan, X. Real-world Cross-modal retrieval via sequential learning. *IEEE Trans. Multimed.* **23**, 1708–1721 (2021).
151. Zhu, B. et al. Multiple interactive memory representations underlie the induction of false memory. *Proc. Natl. Acad. Sci.* **116**, 3466–3475 (2019).
152. Harrison, N. R. & Woodhouse, R. Modulation of auditory Spatial attention by visual emotional cues: differential effects of attentional engagement and disengagement for pleasant and unpleasant cues. *Cogn. Process.* **17**, 205–211 (2016).

Acknowledgements

Author contributions

M.R. and M.M. designed the study. M.M. performed data acquisition, and J.O. provided parts of the scripts for the behavioural and univariate analysis analysis. M.M. and M.R. analyzed the data. M.R. acquired funding, conceptualized, and supervised the project. M.M. and M.R. wrote the original manuscript. M.M., J.O., and M.R. reviewed and edited the paper and approved the final manuscript.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.