Replay of Stimulus Specific Temporal Patterns during Associative Memory Formation

Hanslmayr, Simon; Bowman, Howard; Michelmann, Sebastian

Link to publication on Research at Birmingham portal

Replay of Stimulus Specific Temporal Patterns during Associative Memory Formation

Sebastian Michelmann[1], Howard Bowman[1, 2], Simon Hanslmayr[1]

1. University of Birmingham, School of Psychology; 2. University of Kent, School of Computing

Abstract

Forming a memory often entails the association of recent experience with present events. This recent experience is usually an information rich and dynamic representation of the world around us. We here show that associating a static cue with a previously shown dynamic stimulus, yields a detectable, dynamic representation of this stimulus. We further implicate this representation in the decrease of low-frequency power (~4-30 Hz) in the ongoing electroencephalogram (EEG), which is a well-known correlate of successful memory formation. The reappearance of content specific patterns in desynchronizing brain oscillations was observed in two sensory domains, i.e. in a visual and in an auditory condition. Together with previous results, these data suggest a mechanism that generalizes across domains and processes, in which the decrease in oscillatory power allows for the dynamic representation of information in ongoing brain oscillations.

Introduction

Not everything we associate in our memory occurs at the same time. When our favorite football player is seeing the red card, for instance, we are able to bring this together with the events we just witnessed a few seconds before. Later, we are naturally able to recall all relevant information leading to the red card. In order to successfully make this association, our brain has to accomplish two things. First, it has to keep track of the past and maintain a representation of the events in the ongoing football match and second, form memories in which past events are connected to the red card. Processes during the encoding phase that will determine our ability to later remember events can be investigated with the so-called subsequent memory paradigm (Paller & Wagner, 2002; Wagner, Koutstaal, & Schacter, 1999).

Subsequent memory effects refer to neural activity which distinguishes remembered from not remembered items at the time of encoding and are well documented in M/EEG and fMRI, showing involvement of cortical as well as medial temporal lobe regions (e.g. Long, Burke, & Kahana, 2014; Otten, Quayle, Akram, Ditewig, & Rugg, 2006). Concerning M/EEG power, decreases in low frequency (<40 Hz) brain dynamics have repeatedly and consistently been related to successful memory formation (Hanslmayr & Staudigl, 2014).

It has recently been proposed that cortical power decreases in the alpha/beta frequency range allow for a rich representation of memory content, since a desynchronized system has more flexibility to code information over a system of high synchrony. We call this view, the information via desynchronization framework (Hanslmayr, Staudigl, & Fellner, 2012). Confirming this idea, we have shown that sustained power decreases in the alpha band at approximately 8 Hz, contain item specific information about the remembered content, when subjects successfully replay dynamic stimuli (i.e. video and sound clips) from memory (Michelmann, Bowman, & Hanslmayr, 2016). In this study, we provided direct evidence that power decreases are involved in the representation of stimulus specific information (Hanslmayr et al., 2012). Moreover these results are well in line with numerous studies showing that perception is not continuous but rather is rhythmically sampled at a frequency of ~7-8 Hz (Hanslmayr, Volberg, Wimber, Dalal, & Greenlee, 2013; Landau & Fries, 2012; VanRullen, Carlson, & Cavanagh, 2007). These outcomes therefore indicate that rhythmic patterns from the perception of dynamic stimuli can reappear during internally guided retrieval processes, in the absence of the stimuli themselves (Michelmann et al., 2016). Accordingly, these prior findings also suggest the possibility that the replay of temporal patterns can be observed in a situation where dynamic stimuli have to be associated with a static cue.

To address this question, we here analyze the data during the encoding phase from a previous dataset (Michelmann et al., 2016). The paradigm required subjects to associate a dynamic stimulus with a static word that was used as a cue in the later retrieval phase. Importantly, during encoding the perception of the dynamic stimulus and the presentation of the word-cue was temporally separated, i.e. in every trial, one out of four dynamic stimuli was followed by a unique word-cue (Fig. 1, a-b). In a visual condition, these dynamic stimuli consisted of four short video-clips, in an auditory session four short sound clips were used. In a later retrieval block, participants were presented with the word-cue and were tested whether they remembered the associated video/sound clip.

We hypothesize that, in order to associate the word-cue with the dynamic stimulus, subjects maintain (i.e. replay) a sensory representation of the dynamic stimulus, which is why we refer to this phase as the "maintenance and association" phase. Using temporal pattern similarity analysis, we should therefore be able to detect the replay of these patterns during the "maintenance and association" phase, i.e. when the association between a word and the sound/movie is formed. In accordance with the information via desynchronization framework, we should observe stronger decreases for later remembered versus later not remembered items. This subsequent memory effect should be most evident in the frequency band that codes for the representation of the dynamic stimulus, i.e. 8 Hz as per our previous findings. Moreover, if power decreases enable a richer representation of the perceptual content, we should already observe stronger power decreases for later remembered compared to later not remembered stimuli during the perception of the dynamic stimulus.
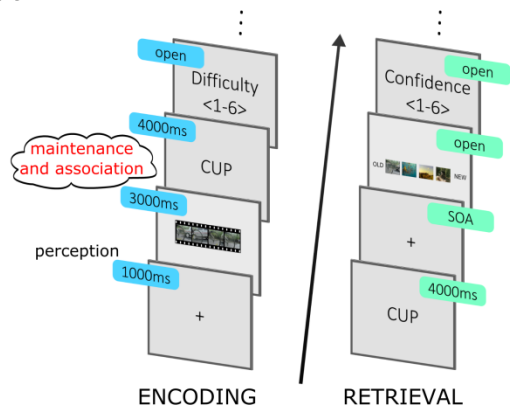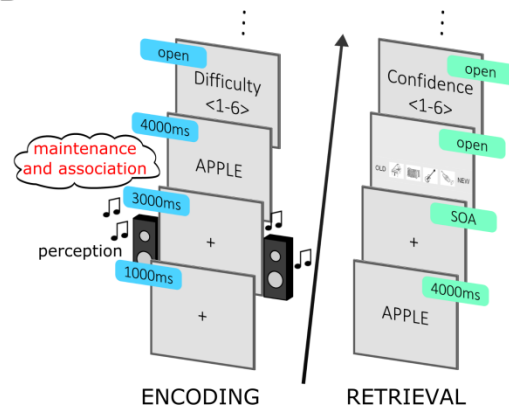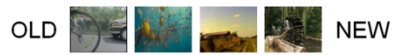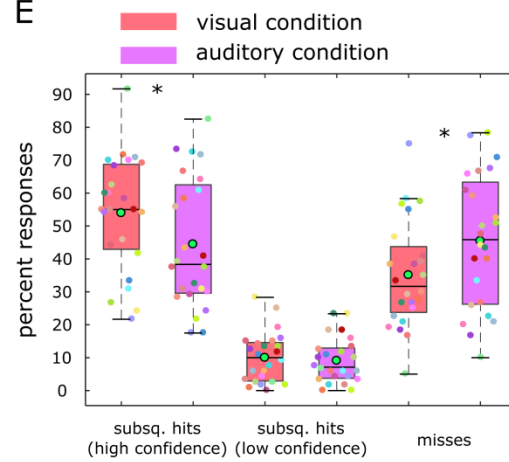
Materials and Methods

Participants

24 healthy, right-handed subjects (18 female and 6 male) participated in this study. 7 further participants were tested, or partly tested, but could not be analyzed due to poor memory performance (N=2), misunderstanding of instructions (N=2), and poor quality of EEG-recording and technical failure (N=3). All participants had normal or corrected-to normal vision. The average age of the sample was 23.38 (s.d. = 3.08) years. Participants were native English speakers (20), bilingual speakers (2) or had lived for more than 8 years in the UK (2). Ethical approval was granted by the University of Birmingham Research Ethics Committee, complying with the Declaration of Helsinki. Participants provided informed consent and were given a financial compensation of 24£ or course-credit for participating in the study.

Material and experimental set up

The cues amounted to 360 words that were downloaded from the MRC Psycholinguistic Database (Coltheart, 1981). Stimulus material consisted of 4 video clips and 4 sound clips in the visual and auditory session respectively. All clips were 3 seconds long; videos showed colored neutral sceneries with an inherent temporal dynamic, sounds were short musical samples, each played by a distinct instrument. In both sessions, a clip was associated with 30 different words. 60 words were reserved for the distractor trials. Those words were not associated with a clip but only shown as new words during memory retrieval.12 additional words were used for instruction and practice of the task. For presentation, words were assigned to the clips or to distractors in a pseudorandom procedure, such that they were balanced for Kucera-Francis written frequency (mean = 23.41, s.d. = 11.21), concreteness

(mean = 571, s.d. = 36), imageability (mean = 563.7 s.d. = 43.86), number of syllables (mean = 1.55, s.d. = 0.61) and number of letters (mean = 5.39, s.d. = 1.24). Furthermore lists were balanced for word-frequencies taken from SUBTLEXus (Brysbaert & New, 2009). Specifically, "Subtlwf" was employed (mean = 20.67, s.d. = 27.16). The order of presentation was also randomized, assuring that neither the clips and their associates, nor distractor words were presented more than 3 times in a row or in temporal clusters. The presentation of visual content was realized on a 15.6 inch CRT-monitor (Taxan ergovision 735 TC0 99) at a distance of approximately 50 centimeters from the subjects eyes. The monitor refreshed at a rate of 75 Hz. On a screen size of 1280 x 1024 pixels, the video clips appeared in the dimension of 360 pixels in width and 288 pixels in height. 'Arial' was chosen as the general text-font, but font-size was larger during presentation of word-cues (48) than during instructions (26). In order to reduce the contrast, white text (rgb: 255, 255, 255) was presented against a grey background (rgb: 128, 128, 128). Auditory stimuli were presented using a speaker system (SONY SRS-SP1000). The 2 speakers were positioned at a distance of approximately 1.5 meters in front of the subject with 60 centimeters of distance between the speakers.

Procedure

Upon informed consent and after being set up with the EEG-system, participants were presented with the instructions on the screen. Half of the subjects started with the auditory session, the others were assigned to undertake the visual task first. Both sessions consisted of a learning block, a distractor block and a test block. The sessions were identical in terms of instructions and timing and differed only in the stimulus material that was used. During instruction, the stimulus material was first presented for familiarization and then used in

combination with the example words to practice the task. Instructions and practice rounds were completed in both sessions.

As a way to enhance memory performance, participants were encouraged to use memory strategies. The suggestion was to imagine the word in a vivid interaction with the material content, yet the choice of strategy remained with the subject. In the learning block, 120 clip-word sequences were presented. Each sequence started with a fixation cross that was presented in the center of the screen for 1 second, and then the video-clip played for 3 seconds. In the auditory condition, the fixation cross stayed on the screen and the sound-clip played for 3 seconds. Immediately after the clip, a word cue was presented in the center for 4 seconds, giving the subject time to learn the association. After that, an instruction requested participants to subjectively rate on a 6 point scale how easy the association between the clip and the word was. After a press on the space bar, this scale was shown. Equidistant categories were anchored with the labels "very easy" and "very hard"; those labels were displayed at both ends above the scale. Participants used six response buttons to rate the current association (see: Fig. 1). In total each video-clip and each sound-clip was shown 30 times. Each word cue was unique, therefore targets were shown only a single time during encoding and retrieval, distractors were only shown a single time during retrieval.

In the distractor block, subjects engaged in a short unrelated working memory task, namely they counted down in steps of 13, beginning from 408 or 402 respectively. After 1 minute the distractor task ended. Following a short self-paced break, subjects refreshed the instructions on the retrieval block.

In this retrieval block, either a cue or a distractor word (i.e. a new word) was presented upon a button press on the space bar. Subjects were instructed to try to vividly replay the content

of the corresponding video-clip or sound-clip in their mind upon presentation of the cue. The word stayed on the screen for 4 seconds, giving the subject the opportunity to replay the memory. Finally, a fixation cross was presented for a varying time window between 250 and 750 milliseconds to account for movement and preparatory artifacts, before the response options appeared on the screen.

The response options consisted of 6 options. 4 small screen shots of the videos or 4 black and white pictures of the featured instruments were presented in equidistant small squares of 30x30 pixels. Additionally, the options "new" and "old" were displayed in the form of text at the most left and most right position (see: Fig. 1 C-D). Subjects could now either indicate the target (video/sound) they just replayed, by pressing the button corresponding to that clip. Instead, subjects could also indicate that the word was a distractor by pressing the button corresponding to the option "new", or they would simply indicate that they remembered the word, but could not remember the clip it was associated with. In this last scenario, subjects would press the button corresponding to "old". The positions of "old" and "new" at the ends, as well as the permutation of the 4 target positions in the middle, were counterbalanced across participants. Finally, after making a decision, a six point rating scale was presented on which subjects could rate the confidence in their response. Now a scale with equidistant categories was presented ranging from "guess" to "very sure". An additional possibility was to press "F2" in case of an accidental wrong button press. In this case, the whole trial was discarded from analysis. Following the retrieval block, individual electrode positions were logged allowing for a break of approximately 30 minutes before beginning the second session. During encoding, participants performed 120 trials in the visual and in the auditory condition respectively. Each stimulus was therefore associated with a unique word-cue 30 times. During retrieval

participants performed a total of 180 trials in each condition that consisted of the 120 unique word-cues from encoding and an additional 60 distractor words. The whole experiment therefore comprised of 240 encoding and 360 retrieval trials.

Data Collection

The recording of behavioral responses and the presentation of instructions and stimuli were realized using Psychophysics Toolbox Version 3 (Brainard, 1997) with MATLAB 2014b (MathWorks) running under Windows 7, 64 Bit version on a desktop computer. Response buttons were "s, d, f, j, k, l" on a standard "QWERTY" layout. Buttons were highlighted and corresponded spatially to the response options on the screen, so participants did not have to memorize the keys. To this end, the shape of corresponding fingers was also displayed under the scale. To proceed, participants used the space bar during the experiment. Physiological responses were measured with 128 sintered Ag/AgCl active electrodes, using a BioSemi Active-Two amplifier, the signal was recorded at 1024 Hz sampling rate on a second computer via ActiView recording software, provided by the manufacturer (BioSemi, Amsterdam, Netherlands). Electrode positions were logged with a Polhemus FASTRAK device (Colchester, VT, USA) in combination with Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011) implemented in MATLAB.

Preprocessing

The data was preprocessed using the Fieldtrip toolbox for EEG/MEG-analysis (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data was cut into trial-segments from 2.5 seconds pre-stimulus to 7 seconds after the onset of the dynamic stimulus. The linear trend was removed from each trial and a baseline correction was applied based on the whole trial. Trials were then

downsampled to 512 Hz and a band-stop filter was applied at 48-52, 58-62, 98-102 and 118-122 Hz to reduce line noise at 50 Hz and noise at 60 Hz; additionally a low-pass filter at 140 Hz was applied. After visual inspection for coarse artifacts, an independent component analysis was computed. Eye-blink artifacts and eventual heartbeat/pulse artifacts were removed, bad channels were interpolated and the data was referenced to average. Finally the data was inspected visually and trials that still contained artifacts were removed manually. In the auditory condition, on average 12.95% of trials were excluded during preprocessing (s.d. = 4.71%, min = 4.17%, max = 25.83%), in the visual condition, excluded trials amounted to an average of 14.31% (s.d. = 4.49%, min = 5.83%, max = 25.00%). There was no difference in the amount of trials that went into the similarity analysis after preprocessing. Neither in the visual condition (mean = 26.13, 25.00, 26.13 and 25.58 trials, s.d. = 2.09, 1.91, 2.29 and 2.04 trials, ANOVA: $F_{(2.96, 68.15)} = 2.037$, $p = 0.12$) nor in the auditory condition (mean = 26.04, 26.25, 25.71 and 26.46 trials, s.d. = 1.92, 1.92, 2.12, 2.06 trials, ANOVA: $F_{(2.68, 61.54)} = 0.906$, $p = 0.434$).

Behavioral analysis

For behavioral analysis, correct trials were defined as those in which the target was correctly identified. The confidence rating of the response was considered as high if a rating of 5 or 6 was selected. Misses were defined as trials in which a cue-word was incorrectly identified as a new word, the wrong clip was selected, or the response "old" was given to indicate recognition of the word without remembering the target video or sound it was associated with.

Power analysis

Oscillatory power was determined by multiplying the Fourier-transformed data with a complex Morlet wavelet of 6 cycles. Raw power was defined as the squared amplitude of the complex Fourier spectrum and estimated for every 4th sampling point (i.e. sampling rate of 128 Hz). For the contrast of subsequent hits and subsequent misses, a baseline was computed as the average power between -1 and 7 seconds of all trials within the contrast (Long et al., 2014). Every trial was then normalized by subtracting the baseline and subsequently dividing by the baseline (activity$_{tf}$ − baseline$_f$)/baseline$_f$, where t indexes time and f indexes frequency. The relative power was calculated for all frequencies between 2 and 30 Hz.

Phase pattern analysis during perception and "maintenance and association"

While participants learned the associations in the encoding block, they repeatedly perceived (saw/heard) the same dynamic stimulus. Content-specific properties could consequently be identified if they were shared by trials of the same content but not by trials of a different content. Hence, content-specific phase during perception was assessed by contrasting the phase similarity between pairs of trials in which the same content was presented, with the phase similarity of an equal number of trial pairs that were of different content. For each pair of trials, the cosine of the absolute angular distance was then computed and finally averaged across all (same or different) combinations [29]. The average similarity value for same and different combinations was subjected to statistical testing across subjects at every time point, at every electrode and in every frequency of interest; this contrast embodies content specific phase patterns during perception.

Participants also repeatedly associated the same dynamic stimuli (one of four videos/sounds) with a different word cue. Therefore the temporal pattern during perception of the dynamic stimulus could also be compared to the temporal pattern during association. This way the

similarity between combinations of same content (e.g. watching video A, associating video A to a cue) could be compared to the similarity between combinations of different content (e.g. watching video A, associating video B to a cue). Notably, excluding within trial combinations eliminates the potential confound of temporal autocorrelation.

In the end, the phase similarity between combinations of same content was contrasted with the phase similarity between trials of different content. This contrast reveals phase patterns that are specific to the dynamic stimulus which subjects associated with the cue. Phase similarity between perception and association was assessed across time points, consequently comparing temporal patterns. To this end a time window from perception was used as in sliding window and a measure of phase coherence over time was computed (see below and Fig. 3 A-B).

To maximize the signal to noise ratio, the following restrictions were applied: The tested frequency was 8 Hertz, following our previous results and hypotheses (Michelmann et al., 2016); The time-window during perception was centered on the cluster in which phase patterns were most reliably content specific during encoding (i.e. the cluster with the lowest p-value). Finally, all possible combinations of trials were used regardless of subsequent memory performance.

The measure to assess phase similarity between two windows was the Single-trial Phase Locking Value (S-PLV) (Lachaux et al., 2000; Mormann, Lehnertz, David, & E. Elger, 2000). This measure defines similarity between two windows (x and y) as the constancy of phase angle difference over time, where n denotes the width of the window and φ is the phase:

$$SPLV = \left| n^{-1} \sum_{t=1}^{n} e^{i(\varphi x t - \varphi y t)} \right|$$

S-PLV assesses the phase coherence between two time windows and has the advantage of increased robustness for noisy data at the expense of temporal resolution.

(Lachaux et al., 2000) suggest to compute the S-PLV over 6-10 cycles of a frequency for a good signal to noise ratio, for our purposes, S-PLV was applied to a time window of 8 cycles, which resulted in a 1 second window for 8 Hertz. Phase values were extracted by multiplying the Fourier-transformed data with a complex Morlet Wavelet of 6 cycles. Phase-values were then downsampled to 64 Hz. The similarity measure was computed for every pair of trials in the combinations of same content and in the combinations of different content. Importantly, the sliding window approach that was used accounts for the non-time-locked nature of the data (temporal patterns could be present anywhere in the "maintenance and association" interval). This resulted in a time course of similarity for the combinations of same and of different content.

The difference in this similarity was first averaged across the whole "maintenance and association" episode (between 3500 and 7000ms) and then statistically tested across subjects with a random permutation procedure (see: statistical analyses) based on clusters of summed t-values across electrodes (Maris & Oostenveld, 2007). In a second test, the time courses at every electrode were compared with a series of t-tests and subsequently tested with a cluster-based random permutation procedure, where clusters were summed across electrodes and time (see also: statistical analyses, below). Additionally, a control frequency was tested, namely

6 Hz, based on the results from the power analysis. Time windows were defined accordingly for this frequency as 8 cycles around the center of the most reliable cluster during perception.

Statistical analyses

Behavioral performance

Behavioral results were compared between the auditory and visual condition with a series of paired t-tests. P-values were compared against a Bonferroni-corrected threshold (Bland & Altman, 1995), however no specific hypothesis was tested. Performance differences between stimuli were tested with a repeated measures ANOVA with the factor "stimulus" (1, 2, 3 and 4) under Greenhouse-Geisser correction.

Decreases in power

To test for differences in baseline corrected power, a paired t-test was first computed for every time point and frequency at every channel. For multiple-comparison-correction, a random permutation procedure was applied (Maris & Oostenveld, 2007). This procedure sums up neighboring t-values above a cluster forming threshold and compares the resulting clusters' sizes to the distribution of the maximal cluster sums that are derived, when condition labels are randomly swapped with the Monte-Carlo method. The minimum number of neighboring channels to be considered a cluster was specified with 3, which attenuates the impact of spatially high frequency noise; neighboring electrodes were derived via the triangulation method of the Fieldtrip toolbox (http://www.fieldtriptoolbox.org/). The clusters were summed across time, frequency and channels, then labels were permuted 1000 times; thresholding of the clusters as well as the testing of the null hypothesis was addressed with an a priori defined threshold for single-sided testing (alpha level of 0.05). To identify frequencies with a reliable

power difference, a paired samples t-test was computed for every frequency on the average power difference across all channels and across the whole time window of interest.

Phase similarity during perception of the dynamic stimulus

Phase similarity during perception was tested in the same way as power. A series of paired t-tests was computed to contrast the average similarity of combinations of same content with the average similarity of combinations of different content. T-values for every frequency band, electrode and time point were then corrected for multiple comparisons in an unrestricted cluster-based permutation approach. The cluster permutation compared again the sums of t-values across frequency, electrodes and time against the distribution of these clusters derived via the Monte-Carlo method (as described above). Later, the frequency 8 Hertz was tested separately with the same cluster permutation in order to identify a temporospatial cluster, in which 8 Hertz phase could differentiate content particularly well.

Phase similarity between perception and "maintenance and association"

The similarity between the time-window during perception and the maintenance-episode was tested for differences between combinations of same and combinations of different content. As mentioned above, in a first step, the average difference between 3.5 and 7 seconds was contrasted with a paired t-test on every electrode to test for a general effect. Multiple comparisons correction was done again with the following procedure: The labels of the average similarity for same and different content combinations were randomly swapped for each participant 1000 times, therefore eliminating the association between condition and observed similarity. Then a t-test was computed between random conditions on the average similarities (3.5 and 7 seconds) in the same way as with the real conditions. Subsequently, t-

values were summed across neighboring electrodes for the random labels and for the real conditions, however only those t-values that exceeded the critical threshold (i.e. a t-value corresponding to alpha = 0.05, defined a priori) were considered. This way 1000 cluster-sums were derived from the random labels, which formed the null-distribution. The real cluster-sum under correct labels was compared against this distribution. The p-value of the whole statistical test was derived as the ratio of random cluster-sums that were bigger than the observed cluster-sum. In a second step, a paired t-test was not only computed for every electrode, but also at every time point during the "maintenance and association" interval. Differences were again tested with a cluster based permutation approach. Now clusters were formed by summation of the thresholded t-values across neighboring electrodes and neighboring time points. They were subsequently compared against the distribution of these clusters for 1000 random label permutations in exactly the same procedure described above.

Results

Behavioral performance

In the visual session, participants remembered on average 53.92% (standard deviation [s.d.] = 17.56%) of the video clips with high confidence (rating > 4), and they further remembered 9.97% (s.d. = 7.62%) of the clips with low confidence (Fig. 1 E). In the auditory session, 44.44% (s.d. = 19.8%) of the audio clips were subsequently remembered with high confidence, which was significantly less than in the visual condition (t_{23} = -2.81, p < 0.01). An additional 9.06% (s.d. = 6.9%) of the audio clips were remembered with low confidence. In accordance, the number of subsequent misses was significantly lower in the visual session (mean 35.07%, s.d. = 16.43%) than in the auditory session (45.45%, s.d. = 20.27%, t_{23} = -3.33, p < 0.01). The total

amount of subsequent misses in the auditory condition (45.45%, s.d. = 20.27%) comprised of 17.22% (s.d. = 12.98%) "old" responses, 14.24% (s.d. = 10.42%) erroneous selections of "new" and 13.99% (s.d. = 9.69%) selections of the wrong sound-clip. In the visual condition, subsequent misses (35.07%, s.d. = 16.43%) comprised of 11.28% (s.d. = 9.97%) "old" responses, 14.37% (s.d. = 11.74%) erroneous selections of "new" and 9.41% (s.d. = 6.33%) selections of the wrong video-clip.

In the visual condition, there was no difference in the amount of trials that were remembered. The four videos were remembered with a probability of 61.94%, 63.61%, 62.08% and 68.06% (ANOVA: $F_{(2.39, 54.86)}$ = 2.250, p = 0.106, s.d. = 20.29%, 18.04%, 16.93% and 18.49%) for high confidence there was a probability of 50.28%, 54.17%, 55.14%, 56.11% (ANOVA: $F_{(2.42, 55.56)}$ = 1.438, p = 0.245), s.d. = 20.45%, 19.91%, 17.28%, 21.14%) to remember each video.

There was however a difference in the amount of trials per stimulus that were remembered in the auditory condition. The four sounds were remembered with a probability of 46.53% 61.81% 60.28% and 45.42% (ANOVA: $F_{(2.66, 61.11)}$ = 10.277, $p < 0.001$, s.d. = 24.93%, 20.59%, 20.83% and 26.70%). For high confidence there was a probability of 37.22%, 54.17%, 50.42% and 35.97%, (ANOVA: $F_{(2.67, 61.36)}$ = 11.342, $p < 0.001$, s.d. = 25.32%, 23.33%, 19.49%, 23.26%) to remember the different sounds.

Successful memory encoding is associated with low frequency power decreases in the visual and auditory condition

To find correlates of successful memory encoding, the oscillatory power between subsequently remembered (hits) and subsequently not remembered (misses) items was compared. Specifically, we contrasted trials for which associations were subsequently remembered with

high confidence, with trials in which the associations were subsequently not remembered correctly. In this analysis, only those datasets were used, in which a minimum of 15 trials remained for hits or misses after preprocessing (N=18). Two crucial episodes for successful memory encoding were tested separately: (i) the time interval in which the dynamic stimulus was actually perceived (0 to 3 seconds) and (ii) the "maintenance and association" interval (3 to 7 seconds), in which the memory formation would be expected to have taken place. In the time interval from 0 to 3 seconds, a small cluster of power decreases was associated with successful memory in the visual condition; it displayed a trend towards significance ($p < 0.07$, Fig. 2A, left). Likewise, in the auditory condition a similar cluster of power decreases appeared ($p = 0.047$, Fig. 2B, left).
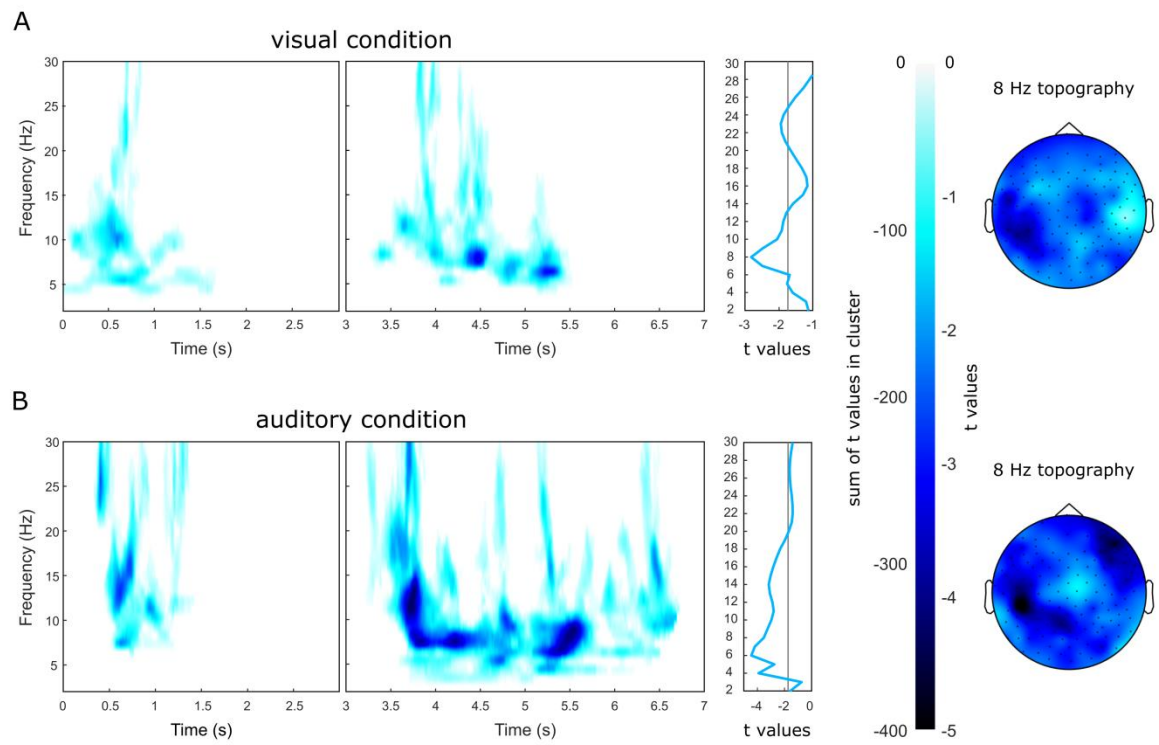
During the "maintenance and association" interval (3 to 7 seconds), substantially reduced power in the lower frequencies (<30 Hz) was observed for subsequent hits compared to subsequent misses (Fig. 2, middle) in both conditions. In the visual condition, a broad cluster emerged where power was significantly lower when tested against random permutations ($p = 0.031$, Fig. 2A, middle). Likewise a broad cluster of significant power decreases appeared in the "maintenance and association" interval of the auditory condition ($p < 0.003$, Fig. 2B, middle).

To identify frequencies that robustly exhibited lower oscillatory power for successful memory encoding, the power during the "maintenance and association" interval was averaged across all electrodes and time points and differences were subjected to a t-test. Following our previous results (Michelmann et al., 2016), we expected the strongest power decreases in both conditions to peak at 8 Hz. Indeed, a clear peak at 8 Hz was observed in the visual condition ($t_{17} = -2.82$, $p < 0.01$, Fig. 2A, middle). In the auditory condition, however, a peak was observed

at 6 Hz (t_{17} = -4.45, p < 0.001, Fig. 2B, middle), yet power decreases also extended to 8 Hz (t_{17} = -3.53, p = 0.001).

For the visual condition, the power decreases at 8 Hz displayed a broad topography with a parietal maximum over the left hemisphere (Fig. 2A, right). Decreases in 8 Hz power were similarly broadly distributed in the auditory condition, with maxima over left parietal and right frontal regions (Fig. 2B, right).

Together, these results confirm the fundamental role of decreases in low frequency oscillatory power for the successful formation of memory.

Temporal patterns are content specific during perception and can be detected during "maintenance and association"

The detection of content specific temporal patterns during the "maintenance and association" period necessitates that the dynamic stimuli themselves elicit temporally distinct neural responses. To address this, we first compared the pairwise phase consistency (PPC) (Vinck, van Wingerden, Womelsdorf, Fries, & Pennartz, 2010) between trials in which the same dynamic stimulus was perceived with the PPC between trials of different content. These findings of content specific phase during perception were previously published (Michelmann et al., 2016).

Oscillatory phase of the neural responses was specific to the dynamic stimuli in two broad clusters in the visual ($p < 0.001$, $p = 0.003$, Fig 3C) and one broad cluster in the auditory condition ($p < 0.001$, Fig 3G), confirming prior reports that the content of dynamic stimuli is tracked by the phase of low frequency oscillations (Ng, Logothetis, & Kayser, 2013). Vitally, both clusters included 8 Hz which was the oscillation for which we hypothesized to detect the reappearance of temporal patterns in the "maintenance and association" period.

We now identified periods during perception in which the time courses at 8 Hz were maximally content specific by restricting the statistical test to 8 Hz only and selecting the cluster in which content could most reliably be differentiated during perception (i.e., the cluster with the lowest p-value). In the visual condition, this cluster extended from -152 ms to 564 ms ($p < 0.001$). Note that post-stimulus effects are smeared temporally into the pre-stimulus interval because of the wavelet decomposition. The most reliable cluster of content specificity in the auditory condition extended from 22 ms to 871 ms ($p = 0.002$). A further cluster in the visual condition was observed between 2,650 ms and 3,300 ms ($p = 0.016$). In the auditory condition

further clusters emerged between 1,818 ms and 2,627 ms (p = 0.003) and between 1,203 ms and 1,504 ms (p = 0.047) indicating that in both modalities early and later time windows showed content specific temporal patterns.

For the 8 Hz oscillation, a 1-second wide window was now centered on the cluster that most reliably distinguished content during perception (i.e. at 206 ms in the visual condition and 446 ms in the auditory condition, Fig 3 A, E). In a sliding window approach, a measure of phase coherence (S-PLV (Lachaux et al., 2000)) was then computed between this window and every 1-second-wide window between 3 and 7 seconds during the "maintenance and association" period (see Fig. 3 A-B). This was done for all available trials from encoding, regardless of subsequent memory performance. For practical reasons, at the end of the trial the window was slid out back into the pre-stimulus interval (zero padding could be an alternative but more intricate approach). This time course of similarity (phase coherence) was now computed for trial-combinations comprising perception and association of the same stimulus and for trial-combinations of perception and association of different content. Importantly, the combinations of same content were never built within a trial, assuring a balancing of temporal autocorrelation between same and different combinations. Furthermore, by changing only the pairing between trials and using the same trials to form pairs of same and different content, it was ensured that the signal to noise ratio was the same in both conditions. In a first test, we subjected the average similarity across time to a t-test, contrasting same and different combinations at every electrode. A cluster-based permutation revealed a significant cluster in the visual condition (p < 0.001), but not in the auditory condition. In a follow-up test, we repeated the t-test for every time-point at every electrode and summed clusters across time and electrodes. A permutation test revealed 2 clusters of significant differences in the visual

condition ($p < 0.001$, $p = 0.035$, Fig. 3 B, D). The first cluster was located over left-frontal regions and extended from 4.8 to 5.41 seconds after stimulus onset (i.e. 1.8 to 2.41 seconds after the start of the "maintenance and association" phase). The second cluster was located over parietal and occipital areas, extending from 4.97 to 5.34 seconds (1.97 to 2.34 seconds of the "maintenance and association" phase, Fig. 3 D right). We applied the same approach to the auditory condition; a cluster ($p = 0.047$) emerged over right-frontal regions extending from 4.11 to 4.44 seconds after stimulus onset (1.11 to 1.44 seconds of the "maintenance and association" phase), even though strictly interpreted, this cluster does not exceed a corrected alpha threshold (Fig. 3 F, H right). Finally we also tested the frequency of 6 Hz, which showed the most reliable power decrease in the auditory condition, however no effects were found.
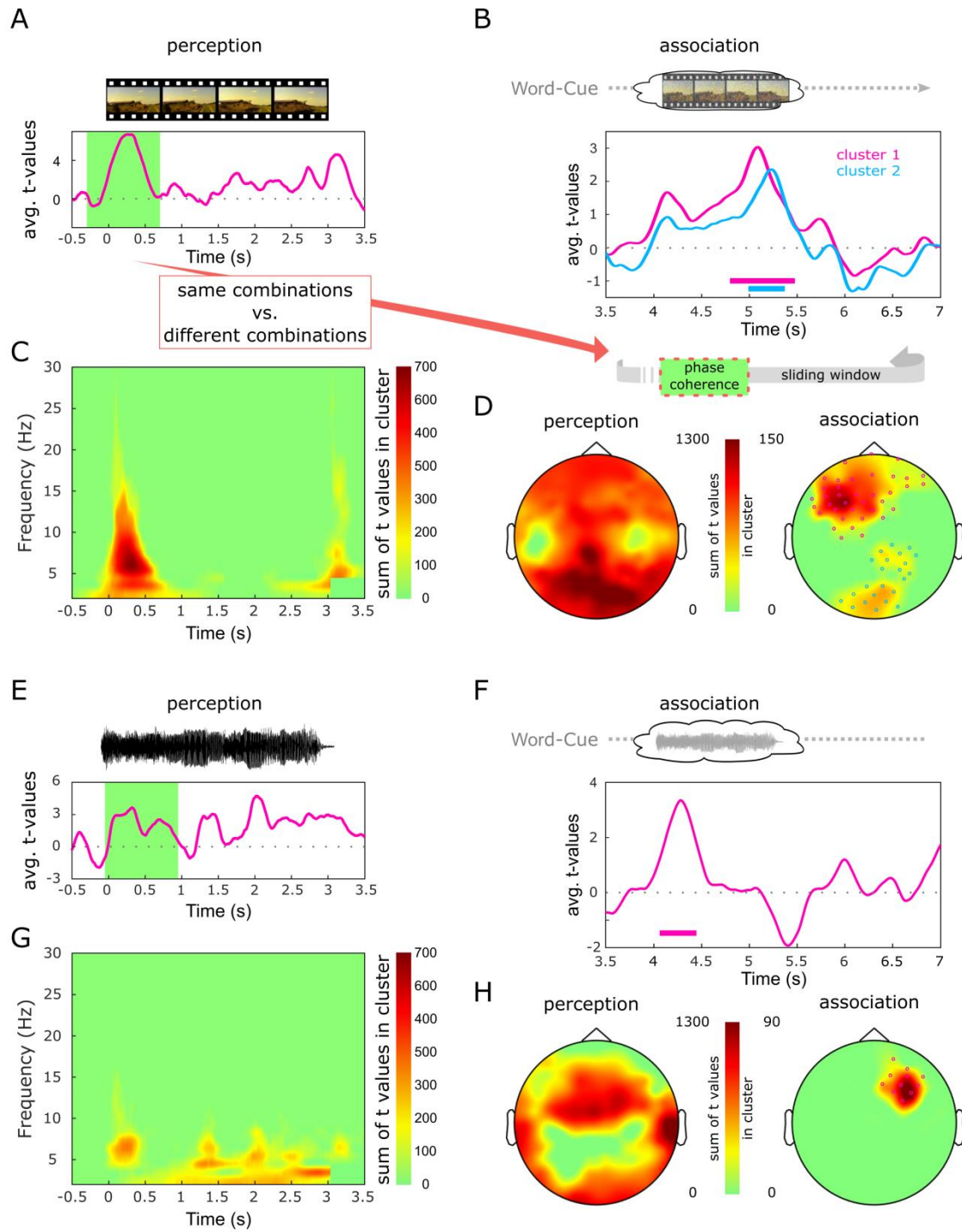
Discussion

For most of the memories that we form during the day, we rely on rich and dynamic ongoing representations of the world around us. At a later point, we then associate these representations with distinct events. Both of these properties of our natural experience are rarely captured in experiments that investigate episodic memory. First, most studies use non-information rich stimuli to study memory, like words or pictures, and second material for association is usually presented simultaneously.

In this study, we used a memory task that can mimic memory in a more naturalistic scenario: an ongoing representation of an information rich, dynamic stimulus is maintained in working memory, in order to be associated with a subsequent event. In one session, subjects repeatedly watched one out of four short video clips, which was immediately followed by a unique word-cue. In a second session, subjects listened to one out of four sound clips, which they subsequently associated with a cue (Fig. 1). In order to form an association, participants had to maintain a representation of the video/sound clip in working memory.

Investigating the correlates of subsequent memory, we found broad and sustained decreases in ongoing oscillatory power to be associated with successful memory formation. These power decreases were particularly strong while subjects maintained dynamic representations in working memory, namely while they formed the association. Importantly, we found that these power decreases carried stimulus specific information in their temporal pattern of activity. Specifically, the phase of an 8 Hz frequency, which we previously linked to content representation (Michelmann et al., 2016) and where power decreases were strongest in the visual condition, was modulated in a stimulus specific way.

These results form part of converging evidence for a general mechanism, in which desynchronization of brain oscillations in the cortex, indicated by power decreases, allows for the rich representation of information (Hanslmayr et al., 2012). Specifically, the decrease in oscillatory strength, which also signifies a release from inhibition (Haegens, Nacher, Luna, Romo, & Jensen, 2011; Klimesch, Sauseng, & Hanslmayr, 2007), renders the oscillation less stationary, i.e. less predictable. In mathematical terms, this decrease of predictability means an increase in the amount of information that can be coded (Hanslmayr et al., 2012; Shannon & Weaver, 1949). During perception we observed content specific patterns over sensory and frontal electrodes, we then tracked these patterns on the same electrodes. When we previously observed reappearing patterns during episodic memory reinstatement, oscillatory patterns were localized in sensory-specific areas (Michelmann et al., 2016). In contrast, the pattern reappearance observed in this analysis displayed a different, i.e. more frontal topography. Speculatively, the difference in observed topography may be due to different task demands during the maintenance and association period that require, for instance, working memory processes (e.g. Goldman-Rakic, 1995). The generalization of the desynchronization-mechanism across different processes is further complemented by its generalization across modalities; namely, in this study as well as in previous results, we observed oscillatory patterns in desynchronizing brain dynamics for visual and auditory stimuli.

Finally, the frequency band of 7-8 Hz has been previously implicated in the rhythmic sampling of perceptual content (Hanslmayr et al., 2013; Landau & Fries, 2012; VanRullen et al., 2007). These studies integrate well with our findings and suggest that the 8 Hz frequency temporally organizes the representations of stimulus specific information during perception, episodic

memory reinstatement and episodic memory formation and that decreases in oscillatory power allow these temporal patterns to resurface.

Our results moreover inform current debates about the neural mechanisms underlying working memory. While some studies have previously shown that content specific activity patterns can be decoded during working memory maintenance (Fuentemilla, Penny, Cashdollar, Bunzeck, & Düzel, 2010; Jafarpour, Penny, Barnes, Knight, & Duzel, 2017), other studies suggest that representations in working memory may not always be maintained online, but rather latently stored in synaptic weights or even via more complex mechanisms (Stokes, 2015). Those representations can then reemerge when they become task relevant, or they can be evoked experimentally by either 'pinging' them with unspecific input (Wolff, Jochim, Akyürek, & Stokes, 2017) or by stimulating transcranially with a magnetic pulse (Rose et al., 2016). Hence, an important insight from the here presented study is that during the formation of an association with a previously shown dynamic stimulus, a detectable representation of that stimulus reappears. The method that we used in order to observe these stimulus patterns was specifically tailored to the detection of patterns that are dynamic in nature, i.e. it is suited to detect (in)consistencies in phase resets over time. It needs to be acknowledged that parts of the patterns that we track may be strongly driven by evoked responses to the stimulus onset during perception. Importantly, however, our method makes it possible to detect these patterns, when their exact time point of reappearance can be unknown and variable between trials. This is very relevant for studies that investigate working memory maintenance because patterns that are involved in the online maintenance of representations in prefrontal cortex and parietal cortex of nonhuman primates, have been found to be highly dynamic (Crowe, Averbeck, & Chafee, 2010; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008). An open

question that remains beyond the scope of this study however, is whether content specific temporal patterns can also be elicited by static stimuli and whether our method can help to detect them.

An interesting question that arises from our results is whether the (re-)appearance of temporal patterns is functionally relevant for the successful formation of memories. We could demonstrate subsequent memory effects for power decreases here, because a minimum of 15 trials per condition can yield stable power estimates (Hanslmayr, Spitzer, & Bäuml, 2009). We could further link power decreases to the presence of content specific temporal patterns; however because the trial count of forgotten associations for most of the subjects was too low for stable similarity estimates, it is not clear whether these patterns are functionally involved in memory formation. Specifically, the present study was designed to produce a sufficient number of remembered trials and we consequently could not contrast stimulus-specific temporal patterns between remembered and forgotten associations. Repeating this study in a longer and more adaptive design, could therefore allow for the contrast of patterns during successful and unsuccessful memory formation.

Additionally, future studies should address whether content-specific temporal patterns are causally involved in memory formation, either by disrupting content specific temporal patterns and therefore tampering with memory formation or even by artificially introducing spurious patterns to cause forged associations.

References

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ : British Medical Journal*, *310*(6973), 170. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2548561/

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Crowe, D. A., Averbeck, B. B., & Chafee, M. V. (2010). Rapid Sequences of Population Activity Patterns Dynamically Encode Task-Critical Spatial Information in Parietal Cortex. *Journal of Neuroscience*, *30*(35), 11640–11653. https://doi.org/10.1523/JNEUROSCI.0954-10.2010

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Düzel, E. (2010). Theta-Coupled Periodic Replay in Working Memory. *Current Biology*, *20*(7), 606–612. https://doi.org/10.1016/j.cub.2010.01.057

Goldman-Rakic, P. . (1995). Cellular basis of working memory. *Neuron*, *14*(3), 477–485. https://doi.org/10.1016/0896-6273(95)90304-6

Haegens, S., Nacher, V., Luna, R., Romo, R., & Jensen, O. (2011). -Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences*, *108*(48), 19377–19382. https://doi.org/10.1073/pnas.1117190108

Hanslmayr, S., Spitzer, B., & Bäuml, K. H. (2009). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral Cortex*, *19*(7), 1631–1640. https://doi.org/10.1093/cercor/bhn197

Hanslmayr, S., & Staudigl, T. (2014). How brain oscillations form memories — A processing based perspective on oscillatory subsequent memory effects. *NeuroImage*, *85*, 648–655. https://doi.org/10.1016/j.neuroimage.2013.05.121

Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in Human Neuroscience*, *6*(April), 1–12. https://doi.org/10.3389/fnhum.2012.00074

Hanslmayr, S., Volberg, G., Wimber, M., Dalal, S. S., & Greenlee, M. W. (2013). Prestimulus oscillatory phase at 7 Hz gates cortical information flow and visual perception. *Current Biology*, *23*(22), 2273–2278. https://doi.org/10.1016/j.cub.2013.09.020

Jafarpour, A., Penny, W., Barnes, G., Knight, R. T., & Duzel, E. (2017). Working Memory Replay Prioritizes Weakly Attended Events. *Eneuro*, *4*(4), ENEURO.0171-17.2017. https://doi.org/10.1523/ENEURO.0171-17.2017

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88. https://doi.org/10.1016/j.brainresrev.2006.06.003

Lachaux, J.-P., Rodriguez, E., Le Van Quyen, M., Lutz, A., Martinerie, J., & Varela, F. J. (2000). Studying single-trials of phase-synchronous activity in the brain. *International Journal of Bifurcation and Chaos*, *10*(10), 2429–2439. Retrieved from http://cepa.info/2083

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, *22*(11), 1000–1004. https://doi.org/10.1016/j.cub.2012.03.054

Long, N. M., Burke, J. F., & Kahana, M. J. (2014). Subsequent memory effect in intracranial and scalp EEG. *NeuroImage*, *84*, 488–494. https://doi.org/10.1016/j.neuroimage.2013.08.052

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex. *Journal of Neurophysiology*, *100*(3), 1407–1419. https://doi.org/10.1152/jn.90248.2008

Michelmann, S., Bowman, H., & Hanslmayr, S. (2016). The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans. *PLoS Biology*, *14*(8). https://doi.org/10.1371/journal.pbio.1002528

Mormann, F., Lehnertz, K., David, P., & E. Elger, C. (2000). Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D: Nonlinear Phenomena*, *144*(3–4), 358–369. https://doi.org/10.1016/S0167-2789(00)00087-7

Ng, B. S. W., Logothetis, N. K., & Kayser, C. (2013). EEG Phase Patterns Reflect the Selectivity of Neural Firing. *Cerebral Cortex*, *23*(2), 389–398. https://doi.org/10.1093/cercor/bhs031

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Otten, L. J., Quayle, A. H., Akram, S., Ditewig, T. A., & Rugg, M. D. (2006). Brain activity before an event predicts later recollection. *Nature Neuroscience*, *9*(4), 489–491. https://doi.org/10.1038/nn1663

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*(6316), 1136–1139. https://doi.org/10.1126/science.aah7011

Shannon, C. E. C., & Weaver, W. (1949). *The mathematical theory of information*. *Urbana University of Illinois Press* (Vol. 97).

Stokes, M. G. (2015). "Activity-silent" working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405. https://doi.org/10.1016/j.tics.2015.05.004

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A User-Friendly

Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, *2011*, 1–13. https://doi.org/10.1155/2011/879716

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19204–19209. https://doi.org/10.1073/pnas.0707316104

Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., & Pennartz, C. M. a. (2010). The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. *NeuroImage*, *51*, 112–122. https://doi.org/10.1016/j.neuroimage.2010.01.073

Wagner, A. D., Koutstaal, W., & Schacter, D. L. (1999). When encoding yields remembering: Insights from event-related neuroimaging. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *354*(1387), 1307–1324. https://doi.org/10.1098/rstb.1999.0481

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871. https://doi.org/10.1038/nn.4546